Big Data Analytics in Innovation Context

Abdullah Hamadi

B.Sc., M.Sc.

Submitted to Swansea University in fulfilment of the requirements for the Degree of

Doctor of Philosophy

Swansea University

2022

Abstract

The potential of big data analytics (BDA) in enabling improvements in business processes has urged researchers and practitioners to understand whether and under what combination of conditions, such novel technologies can support innovation, competitive advantage, and firm performance. Although the extent of research in this area is substantial, research on the influence of the BDA components such as effective use of big data analytics tools and big data management on innovation process (IP), competitive advantage (CA), and financial performance (FP) must be conducted. This research examines the impact of effective use of BDA tools and big data management on the innovation process, competitive advantage and financial performance based on the resource-based view (RBV). Based on this conceptualisation, the current study examines the relationship between the BDA components (effective use of BDA tools and big data management) on the innovation process, competitive advantage, and financial performance.

A questionnaire was designed to measure the influence of the effective use of big data analytics tools and big data management on the innovation process and its drivers on competitive advantage and financial performance. Quantitative data collected using an online survey method, and (n= 174) samples were gathered from top managers working in firms operating in the United Kingdom and the United States of America. The hypotheses and model fitness were tested in SPSS, AMOS v.26 using structural equation modelling (SEM).

The findings indicate that applying the significant data analytics components (effective use of BDA tools and big data management) has a significant positive impact on the innovation process capabilities (β = .317, p = 0.000 and β = .490, p = 0.000), competitive advantage (β = .322, p = 0.000 and β = .298, p = 0.000), and financial performance (β = .188, p = 0.000 and β = .444 and p = 0.000). There is also a significant positive impact of the innovation process on competitive advantage (β = .485, p = 0.000), and the competitive advantage led to a financial performance (β = .397, p = 0.000). Statistical findings also show that the innovation process as mediator has a significant positive impact on financial performance (β = .333, p = 0.000). Surprisingly, firm mediators such as age (β = -.002 and p = 0.000), size (β = -.003 with p = 0.000), type of industry (β = -.008 with p = 0.000), and environment turbulence (β = -.001 with p = 0.000), were found not to positively impact a firm's financial performance.

This work provides an original contribution to knowledge by extending and validating a new model and the factors affecting the innovation process, competitive advantage, and financial performance.

Declarations

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed **Abdullah Hamadi**

Date **30/03/2022**

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged through citation for which a bibliography is appended.

Signed **Abdullah Hamadi**

Date **30/03/2022**

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed **Abdullah Hamadi**

Date **30/03/2022**

Acknowledgements

While this thesis remains of my own work, its completion would have been impossible without the support of the following individuals.

Firstly, I am indebted to my primary supervisor, Dr Nick Hajli, for his support, valuable insights, and commitment to our meetings, where he was accessible at any time. Thank you so much for your constructive feedback, which has assisted me at every stage. I am incredibly grateful for Dr Fahad Ibrahim's, my second supervisor, and his support to complete the methodology and data analysis chapters. Whereas Dr Fahad joined my supervisors' team the last year, he had a huge impact on my work. I really appreciate his support. I would like to thank Dr John Mulyata, my second supervisor, for his help and participation in this work.

Secondly, thanks must go to my PhD colleagues for the lengthy lunchtime chats that have kept me on track and for encouragement and feedback throughout my PhD. I am so lucky to have worked alongside such supportive colleagues at Swansea University.

Ultimately, I would not have achieved this without the unconditional support and love from my children, Esam, Haleh and Hadil, who always believed in me and inspired and supported me throughout every chapter of this PhD journey.

List of abbreviations

AGFI	Adjusted Goodness of Fit Index
AIP	American Institute of Physics
AVE	Average Variance Extracted
BD	Big Data
BDA	Big Data Analytics
BDACs	Big Data Analytics Capabilities
BPR	Business Process Reengineering
CAO	Chief Analytics Officers
CDO	Chief Data Officers
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CITC	Corrected item-total Correlation
CLF	Common Latent Factor
CMB	Common Methods Bias
CMV	Common Method Bias
CR	Composite Reliability
CRM	Customer Relationship Management
CRM	Customer Relationship Management
DVC	Dynamic Capabilities View
EFA	Exploratory Factor Analysis
ERP	Enterprise Resource Planning
ERP	Enterprise Resource Planning
EU	European Union
FAME	Financial Analysis Made Easy
GFI	Goodness of Fit Index
GLS	General Least Square
GOF	Good of Fit
IDC	Industrial Development Corporation
IFI	Incremental Fit Indices
IoT	Internet of Things
IPV	Information Processing View
IS	Information System
IT	Information Technology
KBV	Knowledge-Based View
KMO	Kaiser- Meyer-Olkin
LS	Least Squares
ML	Maximum Likelihood
NFI	Normed Fit Index
NSIs	National Statistical Institutes
OMB	Office of Management and Budget
OSTP	Office of Science and Technology Policy
PA	Parallel Analysis
PCLOSE	*p* of Close Fit
PFI	Parsimony Fit Indices
PGFI	Parsimony Goodness of Fit Index

PNFI	Parsimony Normed Fit Index
R&D	Research and Development
RBT	Resources-Based Theory
RBT	Resources-Based Theory
RMSEA	Root Mean Square Error of Approximation
SEM	Structural Equation Modelling
SEQ	Structural Equation Questionnaire
SMEs	Small, and Medium Enterprises
SRMR	Standardized Root Mean Square Residual
TLI	Tucker Lewis Index
TQM	Total Quality Management
TQM	Total Quality Management
UK	United Kingdom
US	United State
VIF	Variance Inflation Factor

Table of Contents

List of Figures

List of Tables

Chapter One: Introduction

This chapter introduces the proposed research study. The chapter outlines the research background, including the potential big data analytics (BDA) in creating value and a competitive advantage for businesses and presents the border research problems around BDA and its impact on innovation, competitive advantage, and firm performance. The chapter then identifies research gaps and presents the research aim and objectives. The chapter identifies the research questions and provides a brief outline of the research approach proposed to accomplish the research aim and objectives. The chapter highlights the potential contributions of this research to both theory and practice and presents the research scope. Finally, the thesis structure is followed by a summary of this chapter.

1.1 Research Background

Firms must innovate in answer to altering customer needs and the opportunities presented by new technology and changing marketplaces (Baregheh et al., 2009; Duan et al., 2020). According to scholars (e.g., Cooper et al. (1994); Ottum & Moore (1997); Porter & Millar (1985); (Rehm & Goel, 2015), information is a vital asset that can support firms to develop innovation and achieve competitive advantage. Therefore, the emergence of big data (BD) and its capacity to generate actionable insights has fuelled the attention of managers, decision-makers, and academics toward BDA (Davenport et al., 2012; Hao et al., 2019; Mikalef & Krogstie, 2020). Some firms operating in a highly competitive business environment have exploited the phenomenon of BD to improve their position among competitors (Hao et al., 2019; Lee et al., 2014). Obtaining insights through BD will provide firms with agility in answering changes in the externally disturbed environment, predicting market and consumer currents, and anticipating the future (Erevelles et al., 2016). However, as firms make frantic efforts to develop new ideas to satisfy the ever-dynamic needs of their customers and create a competitive advantage, the question of whether such strategic efforts will be successful lingers, necessitating the requirement for suitable tools to tackle the problem (Hajli et al., 2020).

On examining the potential benefits of BDA, the research began by looking at the challenges of BDA adoption and the complementary resources that firms must develop to be able to increase their BDA investment (Mikalef & Krogstie, 2018, 2020; Vidgen et al., 2017). It is now broadly accepted that to receive any benefit from BDA, firms must recognise the influencing factors alongside BD that might lead to harnessing the insight, which can be manifested as new ideas for innovation (Abbasi et al., 2016; Conboy et al., 2020).

BD is relatively a new concept which defined from different perspectives. For example, Özköse et al. (2015) defined it as *"a high volume, high velocity, and/or high variety of information assets requiring new processing forms to enable enhanced decision making, insight discovery, and process optimisation"*. In addition, Hashem et al. (2015) described BD as "*a cluster of methods and technologies in which new forms are integrated to unfold hidden values in diverse, complex, and high-volume data sets*". Moreover, Johnson (2012) mentioned that BD is: "extremely large sets of data related to consumer behaviour, social network posts, geotagging, sensor outputs". Data refer to those elements taken and extracted through observations, computations, experiences, and record keeping, considered a vital resource in the modern world (Kitchin, 2014). It is defined *"techniques and technologies that make handling data at extreme scale affordable"* (Xu et al., 2016, p. 1562). Also, BD is defined as "*the amount of data just beyond technology's capability to store, manage and process efficiently"* (Kaisler et al., 2013, p. 995). This definition refers to an unmanageable amount of data. According to the National Security Agency, the Internet processes 1826 petabytes (PB) of data per day (Jaseena & David, 2014). Data generation is increasing daily; in 2018, the amount of data produced daily was 2.5 quintillion bytes (Hariri et al., 2019). It has risen from gigabyte GB and terabyte TB to exabyte (EB) and zettabytes ZB (Hou et al., 2020). Data generation reached 16 (ZB) in 2017 and 40 in 2020 (Dash et al., 2019). What is appealing here is the value or the benefits that firms can be gained from the BD, where 70% of managers believed that big data is considered an opportunity, while 30% believed it brings them problems (Russom, 2011). According to the International Data Corporation report cited in Lutfi et al. (2022) the global market of BD was worth USD 66.2 billion in 2020 and is forecasted to reach USD 157.2 in average annual growth until 2026. Furthermore, scholars such as Brown et al. (2011) and Chen et al. (2014), BDA has a significant influence on business activities, and it has shifted management, businesses, and research. In addition, it can lead to dramatic cost reductions, significant improvements in the time needed to complete a computing duty, and new service and product offerings, promoting internal business decisions (Davenport & Dyché, 2013). To highlight the importance of the digital age and BD, authors such as Parviainen et al. (2017) stated that the effect of the Digital Age is equal to the Industrial Revolution and BD is the new oil (Brownlow et al., 2015). BD is also described as petrol in-ground; its values must be extracted to take advantage of it. Therefore, the importance of big data lies in the information and insights that can be extracted from it (McAfee et al., 2012). Moreover, Matt et al. (2015) argued that the potential benefits of digitisation vary and include growth in sales or productivity, innovation in value creation, as well as new forms of interaction with customers and others. In recent years, "born digital" promisors' digital users such as Facebook, Google, and Amazon have grown into mighty giants (Sebastian et al., 2017). The Digital Age and BD revolution have already begun, and firms are being urged to convert their activities to catch up with the technological current.

However, many firms still hesitate to adopt BD technology due to a shortage of experienced workers, data, finances, and technical infrastructure (Demirkan et al., 2016). Moreover, it is clear that BD itself does not have value and adopting BD alone will not profit firms unless they have the internal analytic expertise, business methods, and technology to gather data and transform the thoughts into an actionable plan (Chase, 2013). Therefore, achieving benefits from BDA requires significant changes and improvements of technological infrastructure, business applications, business processes, change in the business model, and using new methods to derive knowledge from data (Buhl et al., 2013; Vis, 2013). All these changes in fundamental infrastructures of firms and new requirements need substantial investment, and they will be a significant challenge for most businesses. Additionally, to these challenges, the most critical issue remains the dearth of knowledge about how firms can transform the potential of BD into business value (Mikalef et al., 2019a).

1.2 Research Problem

Undoubtedly, if done correctly, BDA can have significant effects on stimulating business insights and revealing valuable information. There is enough evidence showing that firms such as Google, IBM, and Apple already jumped on the BD bandwagon and thus are very optimistic about their business potentials and gaining a large share of the industry (Kwon et al., 2014). In addition, big data capabilities (BDACs) and associated technological developments and their use in the context of innovation can be an essential pivot for developing timely business solutions, enhancing innovation, and meeting customers' needs (Cuevas, 2018; Marcos-Cuevas et al., 2016; Shirazi et al., 2021). Although many studies have validated the significant impact of BDA on firm performance (Al-Sai et al., 2020; Anwar et al., 2018; Upadhyay & Kumar, 2020), many firms still are not confident about whether to adopt BDA for various reasons

(Ramadan et al., 2020). For example, they might lack a sufficient understanding of or capabilities to realize its prerequisites or integrate BDA with their existing processes and systems to extract value (Maroufkhani et al., 2019). In addition, even some of those firms that have adopted BDA are still struggling with gaining value from big data. It might be a problem of misinterpreting correlations identified from big data or misleading patterns in data, because using such patterns may lead to decisions without the potential for improvement or making unwise decisions (McAfee et al., 2012; Niebel et al., 2019). That is why using BDA may not guarantee a sustainable, positive impact on firm performance, which means that this phenomenon still requires more empirical research on how BDA can support sustainable innovation, enhance value, and improve firm performance (Hao et al., 2019; Ramadan et al., 2020). Hence, analysing the impact of BDA has become an essential topic for managers and decision-makers who are wondering how it can be used to enhance the innovation process, create a competitive advantage, and increase the financial performance as well as for academics which seeks to describe the phenomenon, its implications, and even its direction and future scope.

1.3 Research Gap

BDA is increasingly being endorsed for its potentially crucial role in providing benefits in businesses. Researchers (e.g., Barchiesi and Colladon (2021); Chen et al. (2015); Constantiou and Kallinikos (2015); Zeng and Glaister (2018)) have agreed on the importance of BDA in creating value for businesses and scholars such as Akter et al. (2016); Dahiya et al. (2021); Gupta and George (2016) covered the role of BDA as a source of competitive advantage. Similarly, Jha et al. (2020); Maheshwari et al. (2021) investigated the effects of big data in supply chain management and also Smys and Raj (2019); Wang and Hajli (2017) highlighted the success of BDA in the healthcare industry. On the other hand, scholars (e.g., Cooper et al. (1994); Ottum and Moore (1997); Rehm and Goel (2015)) argue that information is an essential asset in helping firms develop innovation. However, BD altered the sources and type of information available to decision-makers in the firms (Niebel et al., 2019). In addition, some recent studies (e.g., Hao et al. (2019); Mikalef, Krogstie, et al. (2020); Niebel et al. (2019)) have emerged to highlight the association of BDA and innovation. However, these studies differ somewhat from the subject of the current research, where, for example, Hao et al. (2019) investigate big data and BDA capability on sustainable innovation performance, while at the current study, the context of BDA has been break down into two main components (effective use of BDA tools and BD management).

Therefore, there is a lack of theory linking BDA to innovation, and thus also a lack in practical guidance for managers (Duan et al., 2020; Hao et al., 2019). In addition, there are critical gaps in understanding how, where, and when businesses can leverage BD as an ability and a valuable resource to create innovation success in the active marketplaces (Barton & Court, 2012; Johnson et al., 2017; Kiron et al., 2012; Tambe, 2014). Specifically, understanding of the ways and factors in which BDA operates and how it influences the innovation process, competitive advantage, and financial performance remains scarce and needs further empirical research (Duan et al., 2020; Hao et al., 2019).

On the other hand, many firms have invested in BDA technologies in the hope of enhancing their innovation and competitive advantage. Evidence from a survey shows that 67% of managers surveyed believe that using data analytics has created at least a moderate competitive advantage, and 61% of them mentioned that data analytics improved their firms' ability to be innovative (Kiron et al., 2012). However, Choi et al. (2021) stated that 80% of firms that applied BDA would not achieve satisfactory business results. Scholars (e.g., Groves et al. (2016); Murdoch and Detsky (2013)) explained that the probable reason for these failures is to lack of understanding of the economic potential of big data analytics. Therefore, the large proportion of firms still struggling to enhance innovation or improve their processes through their BDA investments. This has ignited a broad debate regarding what BDA resources are most critical to develop (Günther et al., 2017; Marr, 2016; Mikalef & Krogstie, 2020). On the other hand, some scholars, such as Grover et al. (2018), justified that firms have focused on technological issues and ignored their requirements. At the same time, other authors supported the idea that BDA capabilities require development based on the innovation process needs (Bouncken et al., 2018; Liu et al., 2018; Mikalef & Krogstie, 2020). For example, the analysis of extensive quantities of data and the need to take value out of customer behaviours demand processing methods beyond traditional statistical purposes (De Mauro et al., 2015) required staff with specific skills. This perspective emphasises the significance of complementary organisational factors when using BDA technology toward innovation (Mikalef & Krogstie, 2020; Recker & Mendling, 2016; Schmiedel et al., 2015). Furthermore, scholars (e.g., (Mikalef & Krogstie, 2020; Schmiedel et al., 2015; Trkman, 2010)) argued that using BDA to enhance the innovation process requires identifying the resources that create business value and the context in which they are most relevant, which requires that firms must consider other aspects to improve their innovation process and thus gain a competitive advantage (Upadhyay & Kumar, 2020).

The issue of situational factors (e.g., firm age, size, type of industry, and environmental turbulence) and their impact on firm growth and performance has attracted the attention of researchers (see, e.g., Babirye et al., 2014; Bentzen et al., 2012; Palestrini, 2007; Evans,1987; Barba Navaretti et al. 2014; Carr, 2010). For example, questions regarding whether larger firms are superior in performance to smaller firms, or vice-versa, and whether older firms are prominent in version to younger firms, or vice-versa, need to address (Majumdar, 1997). In addition, the environment factor refers to understanding the advantage of using BD in enhancing the firm performance. This factor may facilitate the improvement of the BDA capability and create a new business model (Adrian et al., 2017).

Therefore, determining the capabilities and components of BDA, organisational, and situational factors may provide more actionable guidance for practitioners, outlining a variety of paths that they can follow for enhancing the innovation process capabilities and competitive advantage of firms.

1.4 Research Aim and Questions

Based on the research gap identified within Section 1.3, the research aim of this study is to examine the impact of BDA on firms' innovation process, competitive advantage, and financial performance. To achieve the aim of the research, the following key questions were developed:

1. Does effective use of BDA influence the innovation process of a firm?
2. Does BDA's role impact competitive advantage?
3. How does the innovation process lead to improved competitive advantage and performance?
4. To what extent do the situational factors (e.g., environmental turbulence, type of industry, firm age, size) impact the relationship between competitive advantage and firm performance?

1.5 Research Approach

Big data is a mature research stream supported by several theories (Schilke, 2014; Wang et al., 2019) detailed in Chapter Three. Several of these theories, such as the resource-based view (RBV) and dynamic capability view are employed to explore the impact of BDA on innovation, competitive advantage, and financial performance. Hence, this study seeks to develop and validate an existing theory by proposing and examining the research hypotheses instead of building a new theory. According to Straub et al. (2004), the survey method is the favourite approach for statistically testing hypotheses; therefore, it is broadly used by scholars of information technology, BD, and marketing to investigate the relationships between variables. The philosophies and supported causes for this selection are explained in the Methodology chapter, which also describes the study procedure illustrated in Figure 1.1 in more detail.

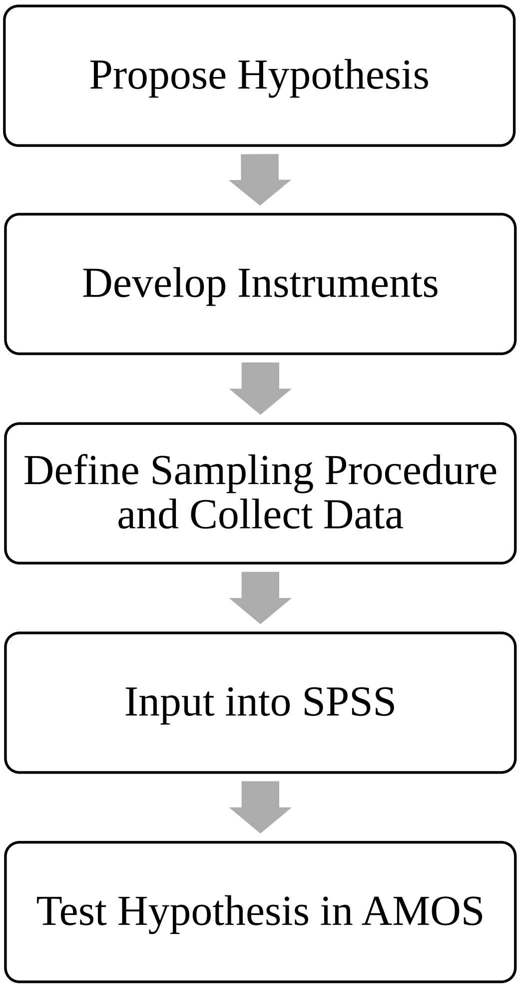


Figure 1.1
Quantitative Research Procedure Used by this Thesis

1.6 Potential Contribution

According to Lee and Baskerville (2003) *"Whether research is conducted quantitatively or qualitatively, there is only one scientifically acceptable way to establish a theory's generalizability to a new setting: it is for the theory to survive an empirical test in that setting"* (p.241).

This study crosses the BDA research interface to generate theoretical contributions and practical implications. The present study is distinguished from the existing literature in several ways. It will be the first study to explore factors affecting innovation context through testing an original conceptual model. This research also validates the effective use of BDA tools and BD management and the innovation process, which will provide more practical benefits. The outcomes of this study will be helpful and valuable for managers and decision-makers of different businesses and academic researchers in the fields of information systems, BDA, innovation, and firm performance.

1.7 Research Scope

The subject area of this study will focus on components such as effective use of BDA tools which include skills and abilities data aggregation, data analysis, data interpretation, and BD management, as factors that can enhance the innovation process, competitive advantage, and financial performance. There is no debate that the area of innovation is a broad subject and has many ramifications some of which will be addressed in the following chapters, but the focus will be on the innovation process. In addition, the impact of the innovation process as a mediator on competitive advantage and financial performance will be analysed. Further, the influence of competitive advantage as a mediator between the innovation process and financial performance will be addressed.

The data will be collected through the managers, executives, and seniors of different business sectors working in firms operating in the United Kingdom and United States.

1.8 Thesis Structure

This thesis's structure is presented as follows. This chapter has introduced the research background, research problem, research gap, aim and objectives, research questions, research approach, research potential contribution, and research scope. The thesis has six additional chapters.

Chapter Two reviews the existing literature regarding BDA, innovation, competitive advantage, and financial performance. Analysing such studies is essential as findings related to these technologies will be relevant to the BD, BDA, innovation process, competitive advantage, and financial performance.

Chapter Three provides the details of developing a conceptual model to explore the influence of BDA on innovation context. First, it provides a detailed discussion on selecting an appropriate theoretical lens to examine BDA components. The chapter then identifies appropriate constructs, formulates hypotheses, and finally develops a conceptual model for the research.

The methodology employed for the research and the philosophy behind decisions is detailed in Chapter Four. First, it reviews the most suitable philosophical position and research procedure to adopt and analyse the data. Then, after justifying a quantitative research method, the chapter develops the research instrument items, sampling and data gathering method. After that, the chapter describes the methods used for pre-testing and pilot testing questionnaire instruments and overcoming procedures for errors. Finally, Chapter Four concludes by discussing the procedure of preliminary data analysis and structural equation modelling, model identification and confirmatory factor analysis.

Results of this study are presented in Chapter Five. First, the respondents' characteristics and contextual information are reported, followed by the results of data screening and preparation. After that, multivariate analyses of research variables are illustrated. Finally, the outcome of structural equation modelling analysis and tests are presented.

The results of the hypothesis of this study will be discussed along with the results in similar studies in Chapter Six, which also highlights the link between research questions, research findings, and research implications.

Chapter Seven concludes this study and highlights the research overall and conclusions of each chapter. In addition, it will discuss the theoretical and practical contributions of this study and refer to limitations that faced this research and future research directions.

1.9 Chapter Summery

This introductory chapter has provided an outline of the research in this thesis. By providing the context of BDA, the background of this study was initially provided. Then the research problem discussed the managers' and decision-makers' need to understand the factors influencing innovation to gain a competitive advantage and improve financial performance. After identifying the research gap, the aims and objectives were explained. After that, the research questions were sated, and a suitable research approach was selected and justified. Finally, after highlighting the potential contributions of this research, the scopes of research and research structure were discussed as well. The next chapter will review the existing literature related to BDA in innovation, competitive advantage, and financial performance.

Chapter Two: Literature Review

2.1 Introduction

This chapter provides a critical overview to understand better the literature on big data analytics (BDA), innovation, competitive advantage, and financial performance through previous studies, books, and journal articles. The literature review overviews big data (BD) and BDA. It addresses the adoption and adaption of BDA by various sectors focuses on the importance of BD. The chapter then discusses sector perspective use of BDA, sector perspectives of challenges in using BDA, and innovation. Finally, the chapter covers competitive advantage and firm performance.

2.2 Overview of Big Data

The globalisation and widespread dissemination of social media and the diversity of devices linkable to the internet has created a massive amount of available data (De Mauro et al., 2016). The sharing and use of these data by media, academia, and industries are some of the motives that have contributed to the exponential development of data (Ward & Barker, 2013). In addition, the emergence of social media platforms such as Facebook and Twitter have contributed to the large amount of data doubling every 2 years. For example, Google now processes more than 40,000 searches every second or 3.5 billion searches per day (Hariri et al., 2019). Similarly, Facebook stores and analyses more than about 30 petabytes (PB) of user-generated data. Such large amounts of data constitute “big data”(Dash et al., 2019).

Scholars such as Chen and Zhang (2014) contend that BD can create a revolution in many areas such as business, public administration, and scientific research. Nonetheless, from a business perspective, the massive amount of data created and to be created represents the behaviour of consumers’ perceptions and marketers who translate those insights to a competitive market advantage (Erevelles et al., 2016). However, to obtain insights and views from BD, managers need to adopt a mindset different from the “small data” perspective of the past (Brownlow et al., 2015). Furthermore, BD has changed the adoption mode in operation businesses, investigations, and management. Similarly, BDA can be applied to make various business decisions depending on the kinds of business, industry, and sector.

According to Muhtaroğlu et al. (2013), businesses employ BDA for two principal reasons. First, firms use BDA for analytic goals, such as obtaining valuable insights about their trades and helping in higher-level decision-making. Second, BDA empowers the development of applications and real-time services that leverage enormous amounts of electronic data to offer clients value (e.g., intelligent services, productivity, and entertainment) that would not be achievable without the availability of such data. Based on these valuable insights, businesses are increasingly facing challenges in dealing with and exploiting BD to their supremacy (Chen et al., 2012). In other words, the most significant challenge for firms is identifying a colossal volume of data and drawing out useful information from them to plan and act in the future (Rajaraman & Ullman, 2011). Moreover, BD allows companies to change traditional business to make more accurate forecasts, more certain decisions, and somewhat rely on intuition and experience (Berner et al., 2014).

Over time, as firms and industries become more digital and rely on information, communication, and connectivity functionality, BDA which strategy will be the business strategy (Bharadwaj et al., 2013; Sambamurthy et al., 2003). However, companies gradually rely on information technologies, knowledge and communication technologies, BD and practical analysis may increase firms' agility (Sambamurthy et al., 2003). However, companies aiming to better use data need to focus on training staff to efficiently manage data properly and incorporate them into decision-making processes (Buhl et al., 2013). The lack of skilled and professional staff in BDA emerged as a significant challenge for companies (Sagiroglu & Sinanc, 2013) as far as; in a Harvard Business Review article, Davenport and Patil (2012) described a BD analyser as "a hybrid of data hacker, analyst, communicator, and trusted adviser" and described the work of BD experts as "the sexiest job of the 21st century" (Debortoli et al., 2014). In addition to the above, other challenges such as organisational structure and culture will arise for companies that intend to benefit from the advantage of BD. These issues were emphasised by (Bughin et al., 2010), who believed that applying experimentation and BD as crucial elements of the management decision-making process need new abilities and organisational and cultural change.

2.2.1 Big Data Definitions

BD originally meant the amount of data that could not be processed (efficiently) by standard database methods and tools (Chen et al., 2013). These data may be created by humans or machines (Singh et al., 2015). According to McMahon (2019), BD refers to pervasive and complex information and data that traditional software applications cannot deal with. There are several definitions of BD. In this regard, Wamba et al. (2017) two sets of BD definitions for some of those researchers in a table. The first set of definitions is based on V concepts (i.e., volume, velocity, variety, veracity, and value). The second set of definitions emphasises different aspects of the concept of BD, such as storage, analysis, and the main characteristics of BD. Table 2.1 illustrates different BD definitions.

Table 2.1
Big Data Definitions

Definitions	Source
BD is the amount of data beyond the ability of technology to store, manage, and process efficiently.	Manyika et al. (2011)
BD technologies are new generation technologies and architectures designed to extract value from multivariate high volume data sets efficiently by providing high speed capturing, discovering, and analysing.	Gantz & Reinsel (2011)
BD is cluster of methods and technologies in which new forms are integrated to unfold hidden values in diverse, complex, and high-volume data sets	Hashem et al. (2015)
BD is a high volume, high velocity, and/or high variety of information assets requiring new processing forms to enable enhanced decision making, insight discovery, and process optimisation.	(Özköse et al., 2015)
BD is “datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse.”	(Manyika et al., 2011)
“Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.”	(Gantz & Reinsel, 2011, p. 6)
BD is: “extremely large sets of data related to consumer behaviour, social network posts, geotagging, sensor outputs”.	(Johnson, 2012, p. 21)

In the current study, BD is a *“high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimisation”* (Gärtner & Hiebl, 2017, p. 3; Özköse et al., 2015). This definition refers to several elements of the BD concept:

- BD characteristics (Vs) will be explained in detail in the following sections and sub-sections. These sections will provide a general overview and enhance the background of the current study.

- It emphasises that BD requires a new technique of processing. These parts will be highlighted and discussed in deep by explaining the effective use of BDA tools, BD management, and innovation process as they are the fundamental factors of the study's research conception model.
- Insight discovery is the primary purpose of BDA.

By analysing BD, the professionals and talented staff are supposed to discover new ideas. These innovative ideas could be a new approach, process, services, or products that can offer a competitive advantage and enhance the firm performance.

2.2.2 Big Data Characteristics

Chen et al. (2013) highlighted Volume, Velocity, Variety (3Vs) as the main features of BD. The traditional data processing techniques are not suitable to deal with such characteristics. Most researchers generally accept this classification (Van Rijmenam, 2014). However, other researchers, such as Demchenko et al. (2013), and Katal et al. (2013), pointed out other two components of BD: Value and Veracity. All these features together are known as the 5Vs (see Figure 2.1). Uddin and Gupta (2014) added another two BD Vs Validity and Volatility. Van Rijmenam (2014) categorised 7Vs as; velocity, variety, volume, veracity, variability, visualisation, and value.

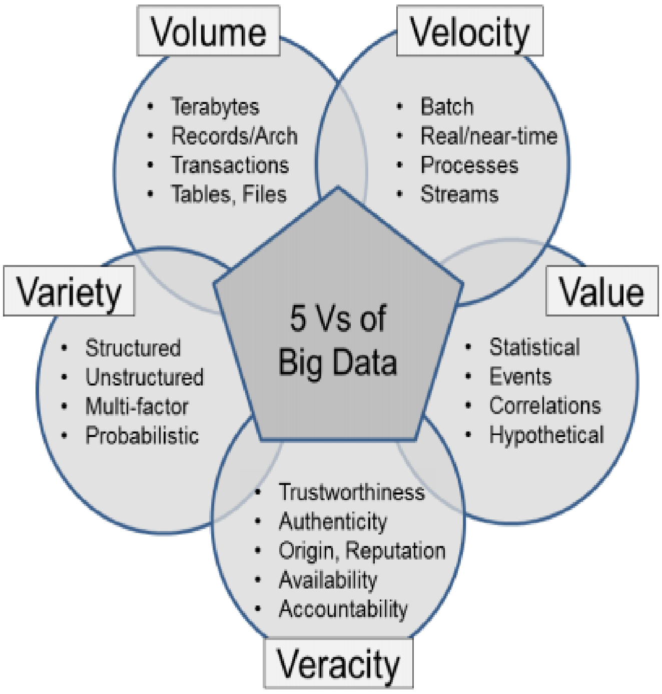


Figure 2.1: Big data 5Vs
Source: Demchenko et al. (2013); Katal et al. (2013)

Volume is the most significant and vital feature of BD, and it has a remarkable influence on the scalability of algorithms (Saboo et al., 2016). It is currently measured in terabytes (TB), petabytes (Echambadi et al.), exabytes (Manyika et al.), or zettabytes (ZB) (Kitchin, 2013; Zaslavsky et al., 2013). One petabyte is equal to 20 million ordinary filing cabinets of text, and it imposes further and specific requirements to all conventional technologies and tools currently used (Demchenko et al., 2013; Erevelles et al., 2016). The volume presents an immediate challenge to traditional IT. Many firms already have massive amounts of archived data in the form of logs but cannot process that data (Anuradha, 2015). In 2012, about 2.5 exabytes (1 Ex-abyte = 1.000.000 Terabytes) of data were produced each day, and that volume is doubling every 40 months or so. The digital universe's capacity in 2013 was measured at 4.4 zettabytes. Each zettabyte is equal to 250 billion DVDs (Erevelles et al., 2016). The increased rate of using smartphones and the internet is driving the wave of the digital revolution, which leads to an increase in the generation of BD for firms (Behl et al., 2019). The benefit obtained from the capacity to process enormous amounts of information is the principal attraction of BDA (Bello-Orgaz et al., 2016; McAfee et al., 2012). This matter caused too many companies to move toward exploration and innovation activities; therefore, BD volume will be playing a vital role and likely widely and effectively used (Johnson et al., 2017).

The velocity, the velocity of BD is often generated at high speed, including data generated from various sources by arrays of sensors or multiple events, and need to be processed in real-time, near real-time (Katal et al., 2013; Kitchin, 2013). Data processing in real time or near real time plays a critical role in identifying and catching frauds in firms' transactions, and a tsunami was heading toward a nuclear power plant (Chopra & Madan, 2015). In such situations, real-time analytics can help prioritise the most critical problems earlier to avoid more damages caused (Qadir et al., 2016). According to Kitchin and McArdle (2016), there are two types of velocity related to BD (a) frequency of generation, and (b) frequency of handling, recording, and publishing. Real-time or nearly real time processing makes information usable at a higher frequency (Saboo et al., 2016). This feature can provide a firm with an opportunity to be much more agile than its competitors (McAfee et al., 2012). Thus, data's enhanced volume and velocity mean that

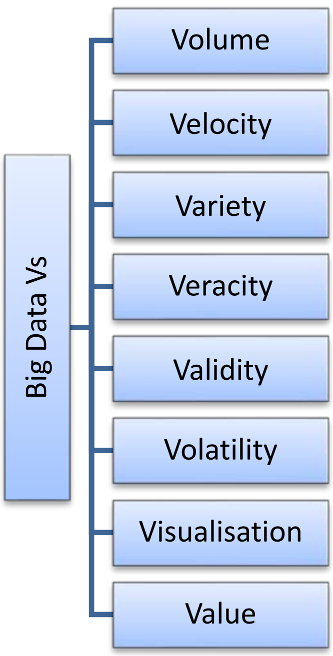


Figure 2.2: Big Data 8Vs
Sources: Adapted from Chen et al. (2013); Gupta & George (2016); Katal et al. (2013)

companies will be required to develop continuous and speedy methods for collecting, analysing, and interpreting data. These works' ideas can be connected with production or service applications and processes to empower continuous processing (Davenport et al., 2012).

The variety of data collected will probably take one of the three forms of structured, semi-structured, or unstructured (Kitchin, 2013). "Variety" refers to the complexity of BD and information and semantic models behind these data, which take various formats of digital data, including photos, email, and text documents (Power, 2014).

By "veracity," scholars mean determining the accuracy of the data (Uddin & Gupta, 2014), and the veracity dimension of BD involves two features: data consistency (or certainty) and data trustworthiness (Demchenko et al., 2013). Indeed, BD veracity guarantees that the data used are trustworthy and guarded against illegal changes and modifications. The data must be reliable during the whole of their life cycle, from their collection from trusted sources to their processing on secure computing devices and protected and reliable' storage.

There is a critical question about the validity of BD, especially that collected from social media platforms like Facebook (Panger, 2016). The validity concerns the truth and accuracy of data concerning the proposed usage (Uddin & Gupta, 2014).

BD volatility calls back to the retention policy of structured data that firms achieve each day in their businesses. Once the retention duration expires, that duration can immediately destroy the data (Uddin & Gupta, 2014).

BD is challenging both computationally and perceptually, and as BD grows, it is necessary to increase the ability to visualise such data (Wickham, 2013). "Virtualization" is defined as "a process of resource sharing and isolation of underlying hardware to increase computer resource utilization, efficiency, and scalability" (Hashem et al., 2015). Scholars argued that visualisation is the most challengeable aspect of BD; however, it has been recommended as a helpful tool to spread BD (Teras & Raghunathan, 2016) because it helps to make that massive amount of data understandable and meaningful in a way that is clear to read and comprehend through the making data visual (McCosker & Wilken, 2014).

A significant characteristic of the data values is described by the gain obtained from collected data, and it is firmly linked to the data's volume and variety. In each industry, senior managers wonder whether they are capturing total value from the enormous quantities of what they already hold within their firms (LaValle et al., 2011). According to Uddin and Gupta (2014) classification, value is the seventh special V, and it is the desired result or worth derived from exploring and processing BD pressing (Assunção et al., 2015).

Uddin and Gupta (2014) highlighted six common challenges due to BD characteristics:

- Heterogeneity: some amount of information heterogeneity in the population can lead to more efficient outcomes, in terms of the system throughput or social welfare, than information homogeneity (Hu et al., 2018).
- Inconsistency and incompleteness: BD increasingly contain information provided by gradually diverse sources of variable reliability.
- Scale: the first thing to consider about BD is its size.
- Timeliness: As data rise in volume and many samples are not economically viable, real-time techniques to summarise and filter what is to be stored are required.
- Privacy and data ownership: The privacy of data is another significant concern and one that grows in the situation of BD.
- The human perspective, Visualisation, and collaboration: To fully reach the potential of BD; consider scale not just for the system but also from the perspective of humans required.

Similarly, Wu et al. (2013) identified four characteristics of BD. They named them the HACE theorem: (H) huge data with heterogeneity and diversity; (A) autonomous sources with distributed and decentralized control dimensionality; and (CE) complex and evolving relationships.

2.2.3 Classification of Big Data

The nature of BD can be known better by separating BD into categories; therefore, scholars such as Özköse et al. (2015) divided BD into five classes regarding their characteristics: Data Sources (Web and Social, Machine, Sensing, Transactions and the internet of thing), content format (structured, semi-structured, and unstructured) data storage (document-oriented, column-oriented, graph-based and key-value), data staging (cleaning, normalization, and transformation), and data processing (batch and real-time). In addition, Hashem et al. (2015) stated another classification, as shown in Figure 2.3.

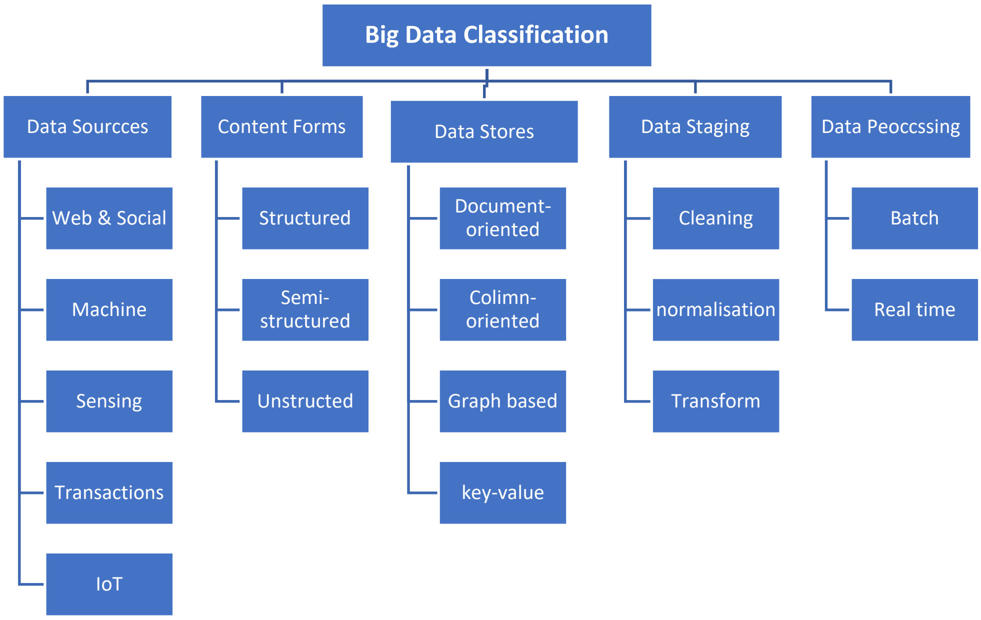


Figure 2.3: Big Data Classification
Source: Hashem et al. (2015)

2.2.3.1 Big Data Sources

BD is a combination of various kinds of granular data (Yaqoob et al., 2016), which may be generated by people or devices. People generate data such documents, images, emails, videos, and posts on Twitter or Facebook. Data can also be produced by machines, such as with sensor logs, click logs, weblogs, email logs, cameras, and computers (Hashem et al., 2016). Significant BD resources are shopping transaction recordings, social media, customer sentiments, cell phone GPS signals, clickstream data and sensor data (Katal et al., 2013). In other words, BD is also an envelope for various kinds of granular data. Scholars classified them in different ways (Singh et al., 2015). Verhoef, Kooge and Walk (2016) is shown in Figure 2.4.

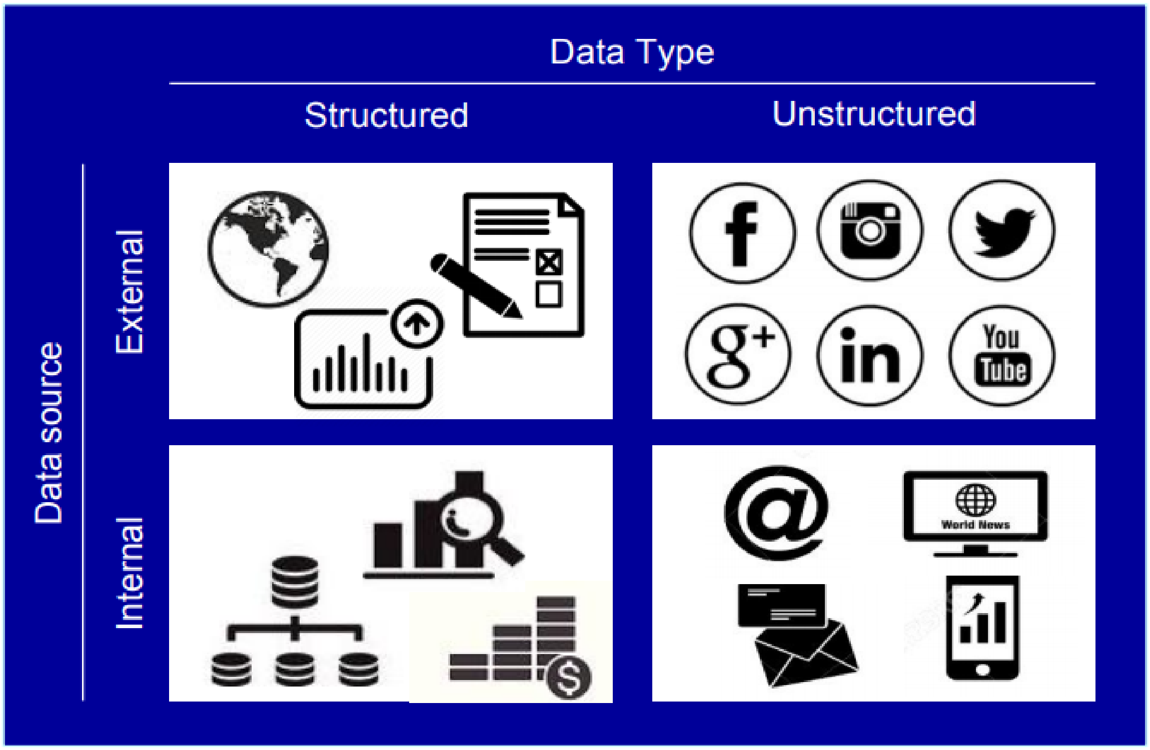


Figure 2.4: Dimensions of Big Data
Source: Verhoef et al. (2016)

George et al. (2014) listed five primary sources of high-volume data (1) public data, (2) private data, (3) data exhaust, (4) community data, and (5) self-quantification data. Hashem et al. (2015) classified BD as (a) web and social, (b) machine, (c), sensing, (d) transactions, and (e) the internet of things (IoT). Similarly, Baesens et al. (2016) argued that BD in today's globally connected networked economies arise from five primary sources: (1) Large-scale enterprise systems; (2) Online social graphs; (3) Mobile devices; (4) internet-of-things; (5) Open data/public data. Table 2.2 shows BD classifications based on different views of scholars.

Table 2.2
BD Classifications Based on Different Studies

BD sources classifications based on different studies		
(George et al., 2014)	(Raghupathi & Raghupathi, 2014)	(Hashem et al., 2015)
Public Data	Internal	Web and Social
Private Data	External	Machine
Data Exhaust	Multiple Forms	Sensing
Community Data	Multiple Locations	Transactions
Self-Quantification Data	Multiple Applications	IoT

Web and Social Media, according to Gärtner and Hiebl (2017) and Gärtner and Hiebl (2017), by 2020, there will be 26 billion devices on the planet, more than three devices on average per user, which will increase communication through social media. In addition, all kinds of robots are increasing daily, and the number of sensors is expected to reach 30 trillion by 2030, corresponding to several thousand per inhabitant of the planet (Bonnaud, 2020). Consequently, the connected objects need more and more connections, transmissions, servers and data centres. This evolution has a significant consequence on data generation. Social media is the source of information created via URL to share or switch ideas and information in virtual societies and networks, such as blogs and microblogs, LinkedIn, Facebook, and Twitter, and Myspace, which generate a great deal of data every day (Han et al., 2015; Hashem et al., 2015). Therefore, social media and consumer sentiments have become a most important source of BD. In addition, Thus, the IoT data will be the most critical sector of BD (Chen & Zhang, 2014), and devices like smartphones and machine-to-machine or thing-to-thing communication will be the principal places for further IoT growth (Vermesan et al., 2011), that means the IoT is one of the most promising resources of BD expansion.

Naturally, such valuable information and data whet many firms' appetite to monitor customers' comments and postings about their products and services. Many firms have begun to use this information to modify decisions and create more value due to increased service levels or developing their products (Katal et al., 2013). Social media users are interacting with friends, viewing videos and photos, and finding brands and businesses. More than 50% of online buyers interact with a retailer on social media networking such as Twitter, Facebook, and LinkedIn (Rapp et al., 2013). Such communication through social media caused a fundamental change in the trade landscape and redefined commercial communication over distribution channels and with the consumers (Rapp et al., 2013).

Machine-Generated. According scholars such as Narasimhan and Bhuvaneshwari (2014) machine-data are information automatically created from software or hardware such as computers, medical devices, GPS, and airplanes. In contrast, human-created data consists of social media platforms such as Twitter and Facebook (Hashem et al., 2015; Narasimhan & Bhuvaneshwari, 2014).

Data Sensing. Due to sensor technology improvements, sensors are becoming more robust, affordable, and small, stimulating many scale deployments. As a result, now we have a massive quantity of sensors already established, and it is anticipated that the quantity will multiply in following years. Sequentially, these sensors will generate BD (Sundmaeker et al., 2010). In addition, virtual sensors can contribute a property of knowledge for relevant application areas, including disaster management, customer opinions analysis, smart cities, and bio-surveillance (Perera et al., 2015). However, sensors failure probably passes inaccurate and wrong data to decision-making algorithms, which will cause inaccurate outcomes (Lee et al., 2014).

Transactions. Computers and intelligent devices are involved in most business transactions currently, and these smart systems are creating a record of customers' transactions (Varian, 2010). In this regard, Sobolevsky et al. (2014) stated that many customer behaviour styles can be discovered by analysing social media posts, records of mobile calls, smart cards, and GPS.

"Internet of Things (IoT). The pervasiveness of various smart objects (e.g., smartphones, radio-frequency identification, sensors, tags, actuators) that can communicate with each other and collaborate with their neighbours to achieve the general goals, called the internet of things" (Wortmann & Flüchter, 2015). This growing availability of sensor-enabled, linked devices provides businesses with inclusive information assets from which it is likely to create new business models, boost businesses processes and minimize costs and risks (Wortmann & Flüchter, 2015).

2.2.3.2 Content Format

BD can take many forms, such as organised, semi-structured and unorganised and have the potential to be extracted for useful information (Balachandran & Prasad, 2017).

Structured Data. Structured data come in an appropriate form and PostgreSQL databases formation (Santoso, 2017). Conventional modelling has been described to interpret these data, such as spreadsheets and databases (Sharma et al., 2016). Enterprise resource planning (ERP) and customer relationship management (CRM) applications are designed to deal with data in organisations in business respect. ERP is a system responsible for managing the organizations' requirements, and it makes handling all the firm's tasks and activities possible for managers. As a software solution, CRM incorporates several applications, such as marketing, sales and customer services (Maleki & Anand, 2008). Structured data are estimated to account for about 10% of all BD (Nambiar et al., 2013).

Semi-Structured. Semi-structured data are structured but not in a manner conforming to the formal structures of data models. These data are often a dynamic mixture of data and metadata before they are introduced to the user (Sharma et al., 2016). These data are also estimated to comprise about 10% of all BD, while most BD, 80%, is unstructured, meaning it cannot be stored in a row-and-column format (Nambiar et al., 2013).

Unstructured Data. Unstructured data are mixed and variable and come in several formats, such as text, images, documents, videos, and social media. Facebook, Twitter, LinkedIn, and Instagram offer an incredible range of users, and those users are generating an astounding rate of unstructured data in every second (Santoso, 2017; Subramaniyaswamy et al., 2015). Unstructured data are generated faster, with an exponential growth of 60% per year (Hou et al., 2020). They are a new, comparatively untapped source of penetration, and their analytics can reveal significant interrelationships that were earlier hard or impossible to discover (Das & Kumar, 2013).

2.2.3.3 Big Data Storages

BD storage concerns *"the storage and management of large-scale datasets while achieving reliability and availability of data accessing"*(Siddiqa et al., 2017, p. 2221). A perfect BD storage system will allow the storage of a massive amount of data; cope with incredible speeds of random read and write paths; flexibly and efficiently deal with a variety of data forms; support structured, semi-structured, and unstructured data; and consider data privacy (Strohbach et al., 2016, p. 119). The massive amount of BD requires storage platforms with tremendous abilities to be distributed, scalable, flexible, and error-tolerant (Kambatla et al., 2014). According to Industrial Development Corporation (IDC) and EMC Corporation cited in Khan et al. (2014), the volume of data created in 2020 will be 44 times higher 40 zettabytes (ZB) than in 2009. This increment rate is supposed to continue at 50-60% yearly, and 85% of this data amount is generated by enterprises (Surbakti et al., 2020). To store this quantity of data in one place, hard disc drives need vast storage capacities. Data must be to keep safe from attackers, and BD stores are tightly controlled. Access to data should be by entry permits to those concerned only, and this issue needs to apply a robust security system (Sagiroglu & Sinanc, 2013). Using traditional structured query language SQL cannot deal with the sheer volume of data, and relational database management systems cannot accommodate such a mass of data. Data diversity imposes new requirements for data storage and the design of new databases tools and techniques, which should be dynamically adapted to data coordination, particularly expansion down and up. Considering the variety of data requires different storage or database systems, scholars such as Hashem et al. (2015) highlighted four types of storage document: orientated, column-oriented, graph-based, and key-value. Similarly, Agrawal et al. (2011) and Kim et al. (2014) named Hadoop as an open-source BD processing platform. It is the most applied technology to managing storage and access, the ability of parallel processing in high speed for the massive amount of data distributed file system (HDFS) is suitable for XML files, and relational database management systems (RDBMS), and not only SQL (NoSQL). SMEs can probably not use Hadoop easily, as using Hadoop needs specialists staff who are most probably not available in such firms (Kim et al., 2014). The essential components of Hadoop are that (a) its file system HDFS allows easy access to data dispersed over various machines without coping with the complexity inherent to their scattered nature, and (b) its MapReduce programming model is designed to implement distributed efficiency and parallel algorithms (Sagiroglu & Sinanc, 2013). In addition, the NoSQL database provides a mechanism for storage and retrieval of data modelled in means other than the tabular relations used in relational databases. Furthermore, graph DBMS should be combined to obtain excellent efficiency for a data warehouse in cloud platforms.

2.2.3.4 Data Cleaning

Databases usually include unnecessary data (Guyon et al., 1996), and BD most probably contains an incredible amount of dirty data, with possible errors, incompleteness, or differential accuracy (O'Leary, 2013). Dirty data boost operational expense because time and other sources are used, identifying and fixing errors (Redman, 1998). This low-data quality an estimated cost to U.S. businesses of $600 billion in 2002 (Eckerson, 2002). Data cleaning is the method of recognising and identifying dirty data (errors and incoherence) and repairing and fixing them to improve the quality of the data (Rahm & Do, 2000). Data cleaning plays an essential role in data management and data analytics, and it is still undergoing agile development. In addition, the importance of cleaning data is crystallised when we know that more than 30% of the data is dirty (Tang, 2014). It is one of the sustained difficulties in data analytics, and failure to the data will result in inaccurate analytics and therefore unreliable decisions (Chu et al., 2016). The duty of recognizing and fixing data errors usually needs manually investigating data, which can quickly become expensive and time-consuming (Krishnan et al., 2015).

Data cleansing is essential for consequent analysis, as it improves analysis accuracy (Rahm & Do, 2000). The process of detecting incomplete, inaccurate, unrelated data and then modifying and clearing them to improve their quality is called the data cleansing technique (Van den Broeck et al., 2005). Hu et al. (2014) identified five essential and complementary steps as a general framework for data cleansing: (a) recognising and determining types of error; (b) defining and assessing error types; (c) searching and detecting error instances; (d) fixing the errors; notarising error instances and error forms; and (e) adjusting data entry techniques to decrease future errors. Normalisation is a method of structuring database schema to minimise the redundancy of data, and transformation is a process of transforming data into a suitable form of analysing (Hashem et al., 2015).

2.3 Big Data Analytics

Analytics is not only an information technology, rather it is an enabler of organisational innovation, processes design, strategy formulation, scenario planning, risk mitigation, and performance efficiencies in manufacturing and service delivery(George & Lin, 2017), and it becomes a movement (Ribarsky et al., 2014). Therefore, in BD literature, many definitions of BDA have been presented. For instance, Wamba et al. (2015) defined BDA as a "holistic process to manage, process and analyse 5 Vs (i.e., volume, variety, velocity, veracity and value) to create actionable insights for sustained competitive advantages" (p. 236). In addition, Russom (2011) states that BDA is "advance techniques operate big data"(p. 6), and Ghasemaghaei et al. (2018) described BDA as "tools and processes often applied to large and disperse datasets for obtaining meaningful insights, has received much attention in IS research given its capacity to improve organizational performance"(p. 103).

2.3.1 Big Data Analytics and Value Creation

According to the McKinsey Global Institute (2011); as cited in Kubina et al. (2015), there are five ways of how BD create value:

- It can help create transparency via new accessibility.
- It empowers firms to experiment. For instance, tests for process changes could generate and analyse massive amounts of data from these experiments to recognise potential performance improvements.
- It makes for a more detailed customer segmentation for customising procedures and setting up specific services.

- Some hidden risks or hidden correlations could be clarified through BDA. Their identification will lead to better decision-making by managers.
- New business models, goods, and services, and even existing ones, could improve via BD. Data on how to use services and goods can upgrade and improve new versions of goods and services.

Therefore, BD as a resource that is supposed to feed into the innovation process requires tools and methods that can be applied to analyse and extract patterns from the enormous unstructured data (Dey et al., 2019; Najafabadi et al., 2015; Polyakova et al., 2019).

The necessity of using BD as a valuable resource to support the innovation process and value creation is building BDA capabilities in firms. Therefore, companies need to invest in developing their capabilities, including management and the ability to analyse data. The investment should be in three essential infrastructure elements: BD assets, human skills, and analytics portfolios (Grover et al., 2018).

Saleem et al. (2020) implemented a study to investigate how BDA influence Chinese SMEs' performance. This study used BDA (predictive and prescriptive) through technology innovation (product and process) as the basic model and structural, empirically tested using data from 312 responses by Chinese SMEs. The authors indicated that both predictive and prescriptive BDA provide product and process innovation in SMEs business activities. They also discovered that product and process innovation results in the improvement of SMEs performance. Also, technological innovation (product and process innovation) mediates heavily between BDA (predictive and prescriptive) and SMEs' performance.

Dubey et al. (2019) explored the influence of BD and predictive analytics and firm performance. The authors examined the relationship between institutional factors such as (coercive pressure, normative pressure, mimetic pressure) and resources of the firm (data connectivity, technology, primary resources) and human skills on the operational performance under the moderating influence of BD-driven culture through 195 samples from Indian factories. The authors found that institutional pressures have significant effects on the selection of tangible resources. Also, normative, and mimetic forces have a substantial impact on human skills. BD culture has significant and positive moderating influences on the ways leading from tangible resources/staff skills to BD predictive analytics. While a considerable amount of literature has been published to identify the influence of using BD and BDA on decision-making, competitive advantage, and firm performance, it seems that researchers have not addressed the value created from BD and through the innovation process.

The current study, summarised the most related studies that assessment BDA on firm performance, as shown in Table 2.3.

Table 2.3
Summary of Studies That Examined the Effect of BDA on Innovation and Firm Performance.

Studies	Author/s	BDA Capabilities Assessed	Assessed Item	Major findings
How to improve firm performance using BDA capability and business strategy alignment?	(Akter et al., 2016)	BDA capabilities (BDA management capability, BDA technology capability, BDA talent capability)	Firm performance	Confirming the value of the higher-order BDA capabilities model's complexity conceptualisation and its influence on firm performance. and the results also clarify the crucial moderating effect of analytics capability–business strategy alignment on the BDAC–FPER relationship
BDA and firm performance: Effects of dynamic capabilities	(Wamba et al., 2017)	BDA Business Analytics Capabilities: BDA Infrastructure Flexibility, BDA management Capabilities, BDA Personnel Expertise Capabilities	Firm performance (financial performance and market performance)	The results confirm the value of the hierarchical BDAC model's entanglement conceptualisation, which has both direct and indirect influences on FPER. The results also confirm the powerful mediating position of PODC in developing insights and improving FPER.
Exploring the relationship between BDA capability and competitive performance: The mediating roles of dynamic and operational capabilities	(Mikalef, Krogstie, et al., 2020)	BD Predictive Analytics: Resources & Capabilities (Managerial Skills, Technical Skills)	Dynamic capabilities (Marketing capabilities and Technological capabilities) Competitive performance	Strong BDA capabilities can support firms to create a competitive advantage. This influence is indirect but completely mediated by dynamic capabilities, which strive a positive and notable influence on two kinds of operational capabilities (technological and marketing capabilities).
Examining the interplay between BDA and contextual factors in driving process innovation capabilities	(Mikalef & Krogstie, 2020)	BDA resources (data, technology, basic resources, technical skills, managerial skills, organisational learning, data-driven culture)	Process innovation capabilities (incremental, radical)	Under various combinations of contextual factors, BDA resources' importance varies, and with specific configurations, high levels of process innovation capabilities achievable.
Assessing the impact of BD on firm innovation performance: BD is not always better data	(Ghasemaghaei & Calic, 2020)	BD Characteristics (Variety Volume, Velocity)	Innovation performance (Innovation efficacy, Innovation Efficiency) and Firm Performance (Financial returns, customer perspective, operational excellence)	Data variety and velocity positively influence firm innovation performance, but data volume has no notable effect. Also, data velocity practices a more significant role in enhancing firm innovation performance than other BD characteristics.

2.3.2 Big Data Analytics Process

The processing and analysis of BD pose a fundamental challenge (Sandryhaila & Moura, 2014) as BD resources are varied and complex. In addition, applying BDA needs collaboration between a company and its different departments to create a smooth and easy activity. Obtaining valuable information and patterns from massive volumes of data needs innovative methods that efficiently process the massive volume of data and use their structure and scalable analysis algorithms to provide timely outcomes (Chopra & Madan, 2015; Sandryhaila & Moura, 2014). Data analyses results from digitisation and refer to mining activities and analysing data to enhance productivity (Stieneker, 2018). In summary, "Information is the oil of the 21st century, and analytics is the combustion engine" (Chaudhari & Patel, 2017, p. 595) , as, Senior Vice President of Gartner Research Peter Sondergaard, famously declared. As the different kinds and amounts of BD need various diverse analysis methods (Huang et al., 2015), the methods and techniques of data analytics have changed during in recent years from the traditional analytics 1.0 measurement to business intelligence and analytics 2.0, and business intelligence and analytics 3.0 (Abbasi et al., 2016). Indeed, to realise the enormous potential of BDA, not only will a firm's IT structure require changing, but almost every unit within a firm will also experience modifications (Davenport et al., 2010).

According to Ikeda and Marshall (2016), the most successful innovative organisations created appropriate innovation environments in the following three areas:

- Organisation: creating impact from innovation sources, opening up innovation processes, and building dedicated innovation teams;
- Culture: putting innovation at the organisation's core, building an innovation atmosphere, and prioritising agility as a significant capacity;
- Processes: building ideation stages and competencies; ensuring an innovation budgeting stream and applying quantitative metrics to assess innovation.

Firms need to employ extensive analytical skills to benefit from user-generated data analysis's innovation and competitive advantages (To & Lai, 2015). According to the IBM Tech Trends Report, business analytics was identified as one of the four primary technology currents in the 2010s. Another survey by Bloomberg Businessweek in 2011 demonstrates that 97% of firms with revenues exceeding $100 million were found to use some method of business analytics (Chen et al., 2012). BDA is defined as a holistic approach to management, processing and analysing the five Vs of data-related dimensions (volume, variety, velocity, veracity, and value) to generate actionable ideas for carrying continued value, measuring performance, and launching competitive advantages.

The smarter businesses such as Amazon, Google, and eBay are embedding analytics to transform data into insight and action (LaValle et al., 2011). Wymbs (2016) defined “data analytics” as “the scientific process of transforming data into insight for making better decisions”(p. 62). Companies require workers with essential and specific skills to analyse BD to detect hidden patterns, unknown communications, correlations, insights, and other useful information (Balachandran & Prasad, 2017). Such valuable information can bring companies competitive advantages over competitors and lead to business benefits such as more effective marketing and revenue growth BD opportunities.

Todays, businesses need employees to know how to do significant tasks such as data collection, data analysis, and data interpreter through software tools (McMahon, 2019). According to LaValle et al. (2011), MIT Sloan Management Review, and the IBM Institute for Business Value studied about 3,000 executives, managers, and analysis. Top-performance businesses are using analytics five times more than low-performance businesses. In contracts, the most significant barrier to inclusive analytics adoption is the lack of understanding of how-to use analytics to grow the business, according to about 40% of participants. These facts will lead to investigating the types of skills or abilities that data analytics require. To and Lai (2015) classified staff’s BDA abilities into three kinds or levels:

- Specialists who run analytic models and algorithms gain the output and exhibit the results in a way that organisational managers can interpret and act.
- Experts who develop advanced models and use them to answer business problems; and
- Scientists who manage teams of specialists and experts to create unique and innovative approaches for analysing data and creating actionable and quantitative solutions to promote organisational strategies.

According to Kaisler et al. (2013), the BD process includes two main stages: data management and data analytics (see Table 2.4).

Table 2.4
Big Data Process

Big Data Process	
Data Management	Data Analytics
• Acquisition and Recording • Extraction, Cleaning and Annotation • Integration, Aggregation and Representations	• Modelling and Analysis • Interpretation

Source: Adopted from Kaisler et al. (2013)

Scholars such as Grover et al. (2018); Kaisler et al. (2013) highlighted three types of BDA that probably lead to various kinds of decisions and optimisation models. They include: (a) the descriptive analysis that reports on the past, (b) the predictive analysis develops models based on past data for future forecasts, and (c) the prescriptive analysis uses models to identify optimal behaviours and activity. BDA has a rising insistence on prescriptive analytics. To assess and enhance their services level and reduce expenses, businesses use the prescriptive analytics method working through the optimisation and randomised testing (Naganathan, 2018).

2.3.3 Big Data Analytics Tools

Drawing from several disciplines such as computer science, applied mathematics, and economics, various techniques and software tools have been developed for aggregation and analytics BD (Manyika et al., 2011). These tools, technologies, and infrastructure include mobile devices, social media, automated identification technologies. The phrase "software in the BD concept" refers to both infrastructure and analytics software (Kraska, 2013). In infrastructure software, data will contain storage, retrieve, transfer, and processing information (Otero & Peter, 2014). The volume of data increases daily as data creation is at 2.5 quintillion bytes every day, and due to the diversity of sources and forms, it becomes more complex; therefore, enterprises need to re-review and re-evaluate their approach to in collecting, storing, managing and analytics of data (Ularu et al., 2012). Generally, BDA tools according to the whole analysis process are divided into four categories; (a) programming models, (b) data gathering, processing, and warehousing, (c) extraction and monitoring, and (d) management, modelling, and analytics (Frizzo-Barker et al., 2016).

The following section and tables will summarise different tools, techniques, and software required for data collecting, storing, and analysing, and transforming BD. Tables 2.5 illustrates techniques, and the section presents tools and software required for BDA.

Table 2.5
Big Data Analytics Techniques

No	Techniques	Application	Reference
1	Association rule learning	This is a commonly applied technique for discovering exciting relationships hidden in a data set.	(Ban et al., 2015)
2	Cluster	Cluster analysis is a technique that can identify structures or homogeneous groups in the database.	(Agrawal et al., 2005; Ng & Han, 1994; Rakesh et al., 1998)
3	Data mining	Data mining technology has emerged as a means for identifying patterns from BD.	(Sowmya & Suneetha, 2017)
4	Genetic algorithms	"Genetic Algorithm is a nature-inspired heuristic approach used for solving search based and optimisation problems."	(Hans et al., 2015)
5	Machine learning	Machine-learning techniques apply to cluster existing data streams and classify clusters.	(McGregor et al., 2004)
6	Natural language processing (NLP)	Natural language processing technique deals with the issues of analyses, understands, and translates written documents and spoken languages, and uses natural languages to communicate with computers.	(Gudivada et al., 2015)
7	Network analysis	Social network analysis is prepared to extract and process data from social media platforms such as Facebook and Twitter and the results are reported as visualisations forms.	(Chang, 2018)
8	Pattern recognition	Pattern recognition is a technique that extracts features and other insights from BD.	(Zerdoumi et al., 2018)

Table 2.6
Big Data Tools

BD Tools		
Storage Tools	**Support Technologies**	**Computing Tools**
HBase	JSON	Hadoop Map Reduce
Apache Hive	SQL and NoSQL	Cloudera
Cassandra	RESTful	IBM Netezza
Neo4j	Machine-to-Machine	Apache Giraph
		Spark
		Scala
		Tableau

Source: Adapted from Almeida (2017)

2.3.4 Big Data Management

As BD and its analytics tools develop, they will change long-standing ideas about the value and the nature of experience and management practice (McAfee et al., 2012). Managers today are increasingly basing their decisions on real-time insight generated from BD and are directing a growing number of initiatives in this direction (Constantiou & Kallinikos, 2015). Therefore, database management systems (DBMS) have appeared as a flexible and cost-effective solution to deal with data (Vargas-Solar et al., 2017). However, firms will not benefit from the transition to using BD unless they can effectively manage the five areas that are practically important in the process: leadership, talent management, technology, decision-making, and company culture (McAfee et al., 2012). In this regard, a survey of about 600 global managers indicated that most firms are still studying how to handle BD at the firm level (Johnson et al., 2017). The survey also found that companies with a top executive responsible for data administration showed better financial performance than did their peers (Johnson et al., 2017).

2.4 Sector Adaption and Adoption of Big Data

According to Rogers (2003), adopting innovation by organisations is a sequential process, beginning with awareness and understanding of the innovation, resulting in the introduction and execution of a process, product, or practice in the adopting organisation (Hameed et al., 2012). BD adoption is described as a process that permits innovation to change the infrastructure to improve productivity and foretell risk to meet customers' requirements more efficiently (Baig et al., 2019; Günther et al., 2017) and improve firm performance (Saleem et al., 2020). Like IT, the BD adoption process has been divided into three stages: intention to adopt, adoption decision, and implementation (Chen et al., 2015). Table 2.7 summarises the BD adoption process.

Table 2.7
IT Adoption Stages

Initiation	Adoption decision	Implementation
Awareness of innovation	Adoption decision	Acquisition of innovation
Attitude formation of innovation	Resources allocation for implementation	User acceptance of innovation
Proposal for adoption		Actual use of innovation

Source: Adopted from Hameed et al. (2012)

Deciding on adopting BDA as a new technology launch in firms is of utmost importance because of its associated complexities, mainly at the beginning of the diffusion phase (Sun et al., 2018). Although BD has been around for years, it emerged as practical option in 2014 at its hype's peak. For example, Lufthansa Airways took 3 years with the *Value Discovery* process before adopting BDA. Value discovery consists of the three stages of the innovation process, use case development, and strategic development planning, which requires analysing cost-benefit, talent management, and securing decision (Chen et al., 2016). BD deployment was limited, and failures abounded, which caused most organisations to hesitate its adoption and take more time to understand its role and capabilities (Chen et al., 2016). A worldwide survey showed that 43% of organisations gained little or no BDA profit (Côrte-Real et al., 2019). The adoption rate of BDA by organisations, like any new technology, is subject to five determinants: relative advantage, complexity, adaptability, observability, and trialability (Rogers, 2003). Two of the most critical obstacles to adopting and implementing BDA are its complexity and lack of BD management standardisation. These issues are primarily manifested in SMEs and organisations suffering from a shortage of abilities and knowledge (Ardagna et al., 2016). Today, most BD adopters are large, multinational firms, yet these firms still have difficulties integrating BD into their organisation's culture. There are still questions such as what BD is, who uses it, its benefits and risks, which need to be further addressed (Frizzo-Barker et al., 2016).

In contrast, the opportunities and benefits that BDA will provide to different disciplines have attracted significant attention globally from government funding agencies, industries and academia (Zeng & Lusch, 2013). BDA technology adopters were found to have an advantage compared to their competitors by 6% in profitability and 5% in productivity, which has inevitably inspired many companies to spend massively on BDA technology (McAfee et al., 2012). These leading organisations' hallmark is their adherence to distinct strategies in innovation organisation, process, and culture (Ikeda & Marshall, 2016). Kristina McElheran, assistant professor of strategy at the University of Toronto, noted, "Those big early adopters got an early benefit" (Ransbotham & Kiron, 2017, p. 5). Netflix is an example of early adopters who use the data analytics method to model and optimise customers' experience. In this, each client's performance and profile are first analysed, and then the most suitable and rated films are suggested for each user separately.

Chen et al. (2016) divided firms into four groups in terms of BD adoption: (a) not adopting, (b) experimented but not adopting, (c) not yet deployed, and (d) deployed. The category of not adopting groups (experimented or not) refers to companies that are unable to transfer to deployment, and this may be due to several reasons such as (a) a mismatch with the business model, (b) a shortage of organisation capability (Business-IT alignment and robust firm Innovation Process), (c) the "innovator's dilemma" (the belief that the current systems are appropriate and it's best to avoid the risk), (d) the inability to detect innovative use cases, (e) disappointment due to exaggerated and impractical expectations generated by hype, and (f) an inability to act with the complexity related to shifts in the system paradigm. Later, these limitations will be discussed in detail as challenges or barriers to adopting and using BD by organisations. In addition, some scholars have developed models that include the factors that will affect the success or defeat of adoption and implementation of BD and the essential factors they require to consider when deciding to enforce BD.

2.4.1 The Enablers and Inhibitors of Adaption and Adoption of Big Data

Many organisations are hesitant to adopt and use BD effectively as the adoption process is complex and includes risks. For instance, some firms are not sure that BD benefits outweigh their costs; therefore, identifying enablers and inhibitors that might affect the adoption of BD became necessary and can support decision-makers in firms (Sejahtera et al., 2018). In this section, the main enablers of BDA will be identified as critical to the success of adoption and use. Similar to barriers, enabling factors may differ from one sector to another and from one activity to another (Yin & Kaynak, 2015). According to the recent work of Li, Xing, et al. (2019), there are three categories of enablers in smart companies. First, the organisation includes a commitment from top management, business-IT strategic alignment, and the practical structure of the organisation. Second, the technology combines hardware and software. Third, the external environment consists of market competition, the requirement for integration with the supplier, and wide usage of social media.

The main factors that can affect the rate of adoption of BD by firms can be extracted into three areas: technical, organisational, and environmental, as Malaka & Brown (2015) classified BD adoption challenges in the communication sector in technology, organisation, and environment categories. Similarly, Baig et al. (2019) classified these factors into four categories: technical with 11 factors; organisational with 16 elements; environmental with eight aspects; and innovational with seven factors.

However, it seems easier to overcome the barriers to adopting BD for large firms than SMEs due to the potential of these companies. The large firms that adopted BDA in different business sectors are making notable progress in their purchaser relationships, goods choice and improvement and consequent profitability (Coleman et al., 2016). Moreover, data-driven companies have been expected to have an output and productivity of 5-6% higher than similar organisations that are not utilising the data-driven process (Brownlow et al., 2015). In this regard, Brown et al. (2011) stated that BD may well become a novel kind of corporate asset that will cut beyond business units and capacity much as a strong brand does, representing an essential competition source.

Meanwhile, SMEs have proven to be slow in adopting new technology for BDA. As a result, they are at risk of being left behind (Coleman et al., 2016).

2.4.2 Applications of Big Data Analytics

BD commonly points to a heterogeneous set of business applications that work on massive amounts of data (Costa et al., 2012). Although the business sector is leading the development of BD applications, the public sector has begun to devise a vision to help support real-time decision-making from data characterised by fast growth from different sources (Kim et al., 2014). Based on that, in the literature on BD, scholars referred to many BD applications that cover several areas. For example, BD applications have emerged in scientific disciplines such as atmospheric science, astronomy, medicine, genomics, biology, chemistry, and other complicated and interdisciplinary scientific research (Chen & Zhang, 2014). In addition, it emerged in other areas such as technology, health, and smart cities (Fan & Bifet, 2013). Also, Li et al. (2015) listed electronic commerce, telecommunication, astronomy, computer Science, health care, government, and financial trading as examples supported by BD (see Table 2.8).

Table 2.8
Application of BDA Based on Different Studies

No	Author/s year	Areas
1	(Fan & Bifet, 2013)	Business, Technology, Health, and Smart cities
2	(Chen & Zhang, 2014)	Atmospheric science, Astronomy, Medicine, Genomics, Biology, Chemistry,
3	(Li et al., 2015)	Electronic Commerce, Telecommunication, Astronomy, Computer Science, Health Care, Government, and Financial Trading
4	(Özköse et al., 2015)	The automotive industry, High technology and industry, Oil and gas, Telecommunication sector, medical field, Retail industry, Packaged consumer products, Media and show business, Transport, and travel sector, Financial, banking and insurance services, Online and social media services, Public or government services, Research and education, Health care services, Law enforcement and defense industry

Sectors mentioned above included most commercial and non-commercial activities; therefore, it has become more apparent that BD will be vital in most sectors and activities in the future. As shown in Table 2.8, Özköse et al. (2015) mentioned a long list of different sectors and industries that have already started applying BDA. Hence, it has become apparent that BD has spread widely and serves large sectors. This issue leads the current study to highlight the importance of BD and its analytics.

2.5 Importance of Big Data

Talking about the importance of BD should be considered in two sectors. First, the business sector has been at the forefront of exploring the importance of BD and investing heavily in it to create value. Second, the public sector recognised the importance of BD to provide better services to citizens. BD's principal importance consists of the circumstances to use a massive volume of data of verity types (Dong & Srivastava, 2013). Indeed, the importance of BD is more manifested when in August 2010, the White House, Office of Management and Budget (OMB), and the Office of Science and Technology Policy (OSTP) announced that BD is a national challenge and priority besides with healthcare and national security, American Institute of Physics (AIP), 2010 cited in (Kaisler et al., 2013). This issue followed an investment of more than $200 million to develop new BD tools and techniques by different departments of the United States in 2012. To further highlight the importance of BD, Manyika et al. (2011: cited in (Zaslavsky et al., 2013), summarised some statistics as follows:

- $300 billion potential annual value to US health care;
- €250 billion potential annual value to Europe's public sector administration;
- $600 billion potential annual consumer surplus from using personal location data globally;
- 60% potential increase in retailers' operating margins possible with BD;
- 140,000–190,000 deep analytical talent positions; and
- 1.5 million more data-savvy managers needed to take full advantage of BD in the United States.

It is worth mentioning here that these figures refer to 2010. Todays, the amount of investment, benefits and skills staff required to deal with BDA far exceeds these and significant witnessing amazing facts the importance of BD and the role that can be played in most sectors. According to Naganathan (2018), BDA and business revenues are prospected to grow from $122 billion in 2015 to $187 billion in 2019, which means more than half in five years.

2.6 Sector Perspectives Use of BDA

Industries and organisations recognise the capabilities and potential of BDA in addressing business challenges and innovation creativity. Therefore, BDA adoption has become inevitable for those innovation-seeking organisations (Marshall et al., 2015). BD can shift the business model through growth and cost-reducing possibilities in most sectors, and the revolution of BD has touched approximately every area of the global economy (Jokonya, 2015). Scholars argue that economic transformation, creating a novel surge of productivity growth and consumer surplus, can happen through the effective use of BD (Manyika et al., 2011).

In contrast, some experimental studies have shown conflicting results, so how investing in BD can increase firm performance has not yet been resolved. For instance, Ghasemaghaei and Calic (2020) argued that BD is not always better data. The authors examined the influence of the main characteristics of BD (Volume, velocity, and variety) on innovation performance (efficacy and efficiency), which finally influence firm performance (operational excellence, financial returns, and customer perspective). Results confirmed that data variety and velocity positively influence firm innovation performance, but data volume has no notable effect. Also, data velocity practices a more significant role in enhancing firm innovation performance than other BD characteristics.

Furthermore, Hao et al. (2019) conducted an empirical study that included 1,109 data-driven innovation projects to examine how BD and BDAC affect innovation success in China and the United States and highlight a trade-off among BD and BDAC and that the optimal balance of BD is relying on levels of BDAC. For example, BD shows an inverted U-shape relationship with sales growth and gross margin in the US; BDA's more robust capability leads to a more significant impact on sales growth and gross margin. In China, promoting the BDA in the event of low BD resource leads to the growth of profits and gross margin up to a certain point and beyond that point, impeding the innovation process.

According to Ikeda and Marshall (2016), the level of sophistication of firms in their successful adoption and use of BDA in terms of innovation and creating competitive advantage is classified into three maturity levels. The first is analytical innovators, which refers to firms that enjoy a culture of analytics; their decision-making is based on data; their innovative ideas and strategic insights rely on analytics. Leader organisations are greeting information of the ideation process over the organisation, and it is estimated they more likely use 23 of BD and

%79 of analytics to recognise innovation opportunities. The second is analytical practitioners, who have no difficulty with suitable data accessing. They are working to be more data-dependent, but they are still beginners in data analytics. The third is analytically challenged, which are firms that still depend on management intuition rather than data for decision-making. They are still facing challenges to quality data accessing and have a shortage of skilled in data management.

However, Coleman et al. (2016) argued that firms need a pre-evaluation of their maturity about the effective use of BD in daily activities. This pre-assessment of the firms' willingness and maturity toward adopting and using data can be made systematically via a maturity model. Using this method, the firm's current situation regarding BD will be known and diagnosed, of which a developing and implementation strategy of adopting BD can be adequately designed. There is a variety of maturity models named (IBM, IDC, SAP, TDWI, HIMSS) that can provide an assessment of the firm maturity about the adoption of BD according to a subset of the following dimensions:

- Business strategy.
- Data management.
- The existence of specialised.
- Staff and analytical skills.
- Technological infrastructure.
- The level of enterprise adoption.
- Leadership and corporate culture; and
- Data governance.

In any case, a firm's level of maturity is a function of that firm's ability to overcome the adoption challenges mentioned above. For example, the BD revolution began in the pioneering firms. When it became clear that adopting it can be beneficial, other significant companies adopted it in various sectors, in which (Lin, 2014) stressed that BDA would help firms achieve perceived benefits, leading to more of its adoption.

2.6.1 Big Data Adoption in Businesses

In business, BD’s principal purpose is to obtain profits by providing products or services, improving competitive advantage, and pleasing consumers and other stakeholders by creating new value (Kim et al., 2014). BD is also envisioned as the next frontier of innovation, competition and productivity (Yang et al., 2017). Davenport (2014) has helped shed light on

BD's business implication to understand the business environment through the means better that BD is supposed to provide. For example, Yadegaridehkordi et al. (2020), examined the BDA has a significant impact on firm performance in the hotel industry. Furthermore, Marr (2015) explained in his book how 45 successful businesses such as Amazon, Walmart, Shell, and Microsoft used BDA to find solutions to their business problems and how they got extraordinary results. For instance, Amazon, the Seattle-based e-commerce firm, currently uses BDA to anticipate the customers' behaviours so that the products may be transported to them before deciding to purchase (Marr, 2015). Another example of BD adoption is using sensors by some trucking companies and railroads. Transportation companies have added multiple sensors to each of their fleet vehicles and train cars. The BD streams from sensors enable companies to manage mobile assets more efficiently, deliver goods to clients more predictably, identify non-compliant operations, and spot trucks that require maintenance (Russom, 2013).

Undoubtedly, BD has tremendous benefits for companies, and it will play an essential role in the fourth generation of the industrial revolution. According to a survey by Mckinsey (as cited in Yin and Kaynak (2015), intelligent companies, through adopting BDA, could reduce up to 50% of their efforts in decision-making and product development, where the customers preferences via the BDA process are determined, and the results will be used in the manufacture and shaping the future products. However, BDA adoption in the manufacturing sector is still in its infancy (Retrialisca & Chotijah, 2020; Yadegaridehkordi et al., 2018). For instance, only 15% out of 250 selected companies in India claimed they were adopting or using BD, and this rate is less than 10% in Korean firms (Park et al., 2015; Yadegaridehkordi et al., 2018). The rate of BD adoption in the service sector, such as banking, insurance, finance, e-commerce, telecommunication, and IT companies, is more significant than in other sectors (Chen et al., 2012).

2.6.2 Government Sector

Governments are being to use BD widely to recognise and analyse problems while making data available to citizens (Bertot et al., 2014). The use of BD by the government leads to unleashing innovation inside and outside the public sector and allowing other parties such as commercial businesses, non-profits, and individuals to use BDA tools, which enhances the opportunities for creating added value in these sectors (Manyika et al., 2011). The government sector has set the primary goals of BD, supporting domestic peace, achieving sustainable improvement, guarding citizens' fundamental rights, and improving the common welfare and economic growth (Kim et al., 2014). The operational costs of governments can be reduced by analysing and using BD. Governments can adopt some methods and practices from the private sector to improve their services and outcomes. For example, In the United States, The Cloud First initiative was launched to close approximately one thousand data centres and move 79 services to the Cloud during 2015 (Al-Fares et al., 2008). Scholars argued that the government could use BD to increase its efficiency via automating and redesigning while its effectiveness through data segmentation and information transparency (Joseph & Johnson, 2013).

2.6.3 Big Data Analytics and Smart Cities

To create novel commercial opportunities in the private sector, BD can be a fascinating data source for official statistics (Daas et al., 2015). As the critical duty of National Statistical Institutes (NSIs), providing official statistics used by policymakers and other vital players in society, BD is expected to significantly influence NSIs (Struijs et al., 2014). Concerning smart cities, there is increasing optimism that states' use of BD-the enormous volume of digital data gathered from multiple sources will make government policymaking further citizen-focused by taking citizens' favourites into account more perfectly than earlier (Hong et al., 2019). An example of smart cities is IBM's "Smarter Cities" initiative, which proposes to help city authorities locate their cities as charming within decreasing traffic jams, developing the waste-collecting, and better managing water (Etzion & Aragon-Correa, 2016).

2.6.4 Big Data Analytics and Health Care

Many sectors such as banking and retail have already adopted BD, but traditionally, the health care industry has delayed other industries in using BD (Ferranti et al., 2010; Fihn et al., 2014; Groves et al., 2016). However, the health data amount is assumed to increase dramatically in the following years, improving and addressing new challenges in the quality of health care services. In human welfare and health care, BDA provide the enormous potential for prognostic interventions, forming lifestyle and behaviour, minimising expenses, and sustainability of the health care foundation. (Archenaa & Anita, 2015; Kambatla et al., 2014). Recently, the developing availability of health care data in the form of medical files, applications and expense data, R&D data from pharmaceutical firms and other kinds of medicinal content has begun to appear in new kinds of BD applications. The applications present the analysis and aggregate of health care data as a service to third parties (Muhtaroğlu et al., 2013). In the health care sector, BD refer to information about the patients such as lab reports, physician notes, X-Ray reports, diet regime, case history, names of doctors and nurses' therapists, medicine and its expiry date, and national health register data (Archenaa & Anita, 2015). By collecting and performing BD, the different health care sectors can increase their competitiveness in a market estimated to be worth more than $10 billion by 2020 (Manyika et al., 2011). Application of BDA in the health care industry would seem to be one of the areas to have most profited from BDA (Garapati & Garapati, 2018).

Wang and Hajli (2017) investigated how BDA capabilities can be developed and the possible advantages of these capabilities in the health care sector. This study included 109 case descriptions, including 63 healthcare institutions. This study's results regarding the relationship between BDAC and organisations' benefits confirmed five dimensions of potential benefits: operational benefits, speed to decision capability, infrastructure benefits, managerial benefits, and strategic benefits.

2.6.5 Examples of Big Data Applications

To enrich this study further, it is necessary to refer to concrete examples of BD use. Analytics in some sectors and highlight the extent of its impact on creating added value for those businesses (see Figure 2.5).

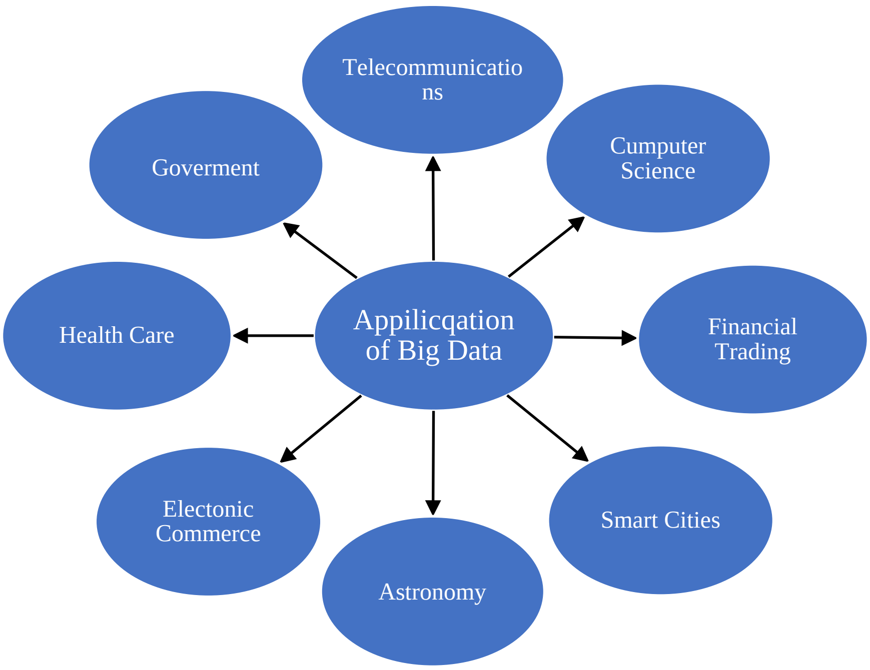


Figure 2.5: BD Applications Based on Different Sectors

Source: Adapted from Katal et al. (2013); Özköse et al. (2015); Van Rijmenam (2014); Wang and Hajli (2017)

2.7 Sector Perspectives of Challenges in the Use of BDA

A review of BDA literature shows that BD’s term is typically accompanied by two words, “opportunities and challenges,” which means that firms that hope to obtain benefits from BD will face several challenges that have to be sorted out. The difficulties include developing even better BD approaches and analytics to manage and leverage BD to achieve business value (Baesens et al., 2016). The most crucial factor is organisational readiness for change (Sun et al., 2018). These difficulties, which will be discussed at length, were the main factors in not adopting BDA in firms. BD researchers have highlighted the various barriers of non-adoption BDA. A study examining the organisational and technological management practices in the North of America confirmed that most organisations are not prepared to adopt big data to improve their performance (McAfee et al., 2012).

Scholars (e.g., Alharthi et al. (2017); Ardagna et al. (2016); Bizer et al. (2012); Kaisler et al. (2013); Sivarajah et al. (2017)) indicated several factors that could be cause for not adopting BD, and they classified them from different perspectives. According to the type of activity and the firm's capabilities, these barriers naturally could be diverse from sector to sector (Yin & Kaynak, 2015).

2.7.1 Barriers to the Adoption of BDA

Several scholars, such as Alharthi et al. (2017); Ardagna et al. (2016); Russom (2011); Sivarajah et al. (2017); Willetts et al. (2020), discussed the barriers and challenges of using BDA from different perspectives. In general, the barriers to not adopting BDA can be classified into three areas: technological, organisational, and environmental (Alharthi et al., 2017). A recent study classified the barriers of not adopting BDA into five sectors: business, environmental, human, organisational, and technological (Willetts et al., 2020). Similarly, Russom (2011) identified more than 14 types of barriers; the first five in importance are (a) inadequate staffing or skills for BDA, (b) overall cost, (c) shortage of business sponsorship, (d) the challenge of architecting a BD analytic system, and (e) current database software lacking in-database analytics. Moreover, Kaisler et al. (2013) highlighted storage, processing and management as three main data analysis challenges. They emphasise that data processing is the most critical challenge, which requires staff with high and specific skills. They used two metaphors to show the importance of processing BD as "finding the needle in the haystack and turning straw into gold." Further, Ardagna et al. (2016) named management and regulatory as the main barriers of BDA adoption, particularly in SMEs. Moreover, Sivarajah et al. (2017) stated three types of challenges for BDA.

- Data challenges are a group of challenges related to characterising (Vs) of data itself. Volume, velocity, variety, variability, veracity, visualisation, and value. In this regard, some researchers classified BD’s Vs as characterises, but some others classified them as challenges.

- Process challenges include all those challenges encountered while proceeding with the data, data aggregation and integration, analysis and modelling, data interpretation, data acquisition and warehousing, data mining, and cleansing.
- Management challenges are manifested in tackling privacy, security, data governance, lack of understanding, and a lack of analysing data skills.

Also, Alharthi et al. (2017) divided obstacles of BDA into three sections:

- Technical obstacles that encompass foundations readiness and complexity of data;
- Human limitations include a shortage of skills and privacy; and
- Organisational obstacles, which refer to organisational culture.

By highlighting these barriers in different sectors, scholars tried to reveal the obstacles while providing appropriate recommendations or solutions to overcoming them. Table 2.9 presents the main challenges facing firms in adapting and adopting BDA.

Table 2.9
Summary of BDA Adoption Challenges/Barriers Based on Previous Studies

BDA Challenge/Barriers	Source(s)
Technological: Encompass foundations readiness and complexity of data, Data scalability, Data Quality, Lack of techniques and procedures	(Alharthi et al., 2017; Arunachalam et al., 2018)
Human limitations: Shortage of data analytics specialists, scientists, experts, and privacy	(Alharthi et al., 2017; To & Lai, 2015)
Management challenges: Privacy, security, data governance and lack of skills related to understanding and analysing data	(Sivarajah et al., 2017)
In the supply chain: Data Quality	(Arunachalam et al., 2018)
Organizational: Culture, Time-consuming, Insufficient resources, Privacy and security concerns, Behavioural issues, Issues with Return on Investment, Lack of skills.	(Arunachalam et al., 2018; Sivarajah et al., 2017)
Barriers in SMEs: Lack of understanding; Dominance of domain specialists; Cultural barriers and intrinsic conservatism; Shortage of in-house data analytic expertise; Bottlenecks in the labour market; Lack of business cases; Shortage of useful and affordable consulting and business analytics services; non-transparent software market; Lack of intuitive software; Lack of management and organisational models; Concerns on data security, Concerns on data protection and data privacy; Different venture concept; Financial barriers.	(Coleman et al., 2016)
Environmental: Competitive pressure, market turbulence, government support, laws, and policy, ethical concerns, digital risks, regulatory issues, shortage of common standards	(Willetts et al., 2020)
Innovation: Relative advantage, cost of adoption, shortage of skilled workers	(Frenkel, 2003; Galia & Legros, 2004)
BD characteristics (Vs): Processing, data aggregation and integration, analysis and modelling, data interpretation, data acquisition and warehousing, data mining and cleansing.	(Kaisler et al., 2013; Sivarajah et al., 2017; To & Lai, 2015)

BDA Challenge/Barriers	Source(s)
Regulatory: The concern of violating data use and sharing regulations, high costs of obtaining legal permits.	(Alharthi et al., 2017)
Management: Complexity and non- standardisation of BD	(Ardagna et al., 2016)
Barriers in organizations: Culture and Managerial Organisation-wide barriers: Lack of understanding and strategic planning, Lack of top management commitment, Lack of collaboration and alignment among organisational departments, failure to identify BD analytical needs in smart factories, People barriers: Lack of qualified and experienced consultants, Lack of in-house data scientists, Lack of trust in BDA results, User resistance caused by changes in job roles and skills.	(Zhang & Lam, 2019)
Barriers in Smart Factories: Technical and data barriers: Immature CPS and IoT development, Lack of integrated and consistent BD set, Poor BD management, Increasing information security threats.	(Li et al., 2019)
Barriers in Agriculture: data governance, asymmetric market information, agriculture policy, agri-environmental policy	(Weersink et al., 2018)

2.8 Innovation

Innovation is the primary determinant of company performance (Damanpour, 1991; Zhong & Nieminen, 2015), and outstanding innovation may lead to sustainable competitive advantage (Ahuja et al., 2008; Hess & Rothaermel, 2011). However, the nature of innovation has shifted and usually involves teamwork and collaboration, and it happens most often in an open and economic environment (Ikeda & Marshall, 2016). "BD is considered a new technology innovation application for firms (Sun et al., 2020). "Innovation" refers to outcomes of the research and development sector's activities performed by chemists, engineers or material researchers (Geissdoerfer & Weerdmeester, 2019). The idea of innovation as a necessary competitive tool for long-term benefit and survival of companies primarily in dynamic markets is well known to those concerned (Deshpandé et al., 1993; Jiménez-Jiménez & Sanz-Valle, 2011). Innovation is embedded in the organisational structures, processes, services and products within a company (Gunday et al., 2011). An aspect that distinguishes organizations from each other is the innovation and the extent to which they relate to the innovation process (Fichman, 2001). Strictly speaking, the meaning of innovation is the successful launching of a novel or developed product, process or service to the market (Hobday, 2005). It refers to a socially acceptable change (Knight, 1967), a primary driver of commercial development and playing a vital role in the race at state and company levels.

Moreover, regarding the importance of innovation, some scholars considered it the industrial religion of the late twentieth century (Baer & Frese, 2003). Also, a fruitful innovation is expected to grow a company's market share (Blundell et al., 1995). The concept of innovation arises from the idea of newness, of *jamais vu*, somewhat never seen earlier (Bomsel & Le Blanc, 2004). Most innovation definitions suggest an idea that innovation involves adopting a novel idea or behaviour (Eveleens, 2010; Jiménez-Jiménez & Sanz-Valle, 2011). Furthermore, the first description of innovation made by the German economist and political scientist Schumpeter defined it as "the driving force for development" (Atalay et al., 2013).

A wide range of companies has embraced a variety of activities that aim to improve the production process and services, such as Total Quality Management (TQM), Business Process Reengineering (Dutta & Bose), Simultaneous Engineering, Lean Production, or Just-in-Time Production to enhance creativity and competitive advantage (Baer & Frese, 2003; Sharma et al., 2016). However, innovation is a term for products and processes and marketing and organisation (Gunday et al., 2011), and there are several types of innovations. By examining the leading firms and what they do concerning innovation, and how they consistently overseeding their peers, it has been found that pioneer firms are distinguished in three critical characterises: (a) Build an enterprise that promotes innovation, (b) Establish a culture that reinforces innovation, and (c) Design processes that empower innovation (Ikeda & Marshall, 2016).

2.8.1 Types of Innovations

Scholars reported different types and many subcategories of innovation in management and business literature. For example, in early innovation studies, Schumpeter (1934) identified three different types of innovation: new goods/services and new techniques of production, new resources of supply, the use of new marketplaces and new rules to organise business (Gunday et al., 2011). Later on, Knight (1967) introduced four categories of innovation, product or service innovations, production-process innovation, organisational-structure innovation, and people innovations. According to Baregheh et al. (2009), innovation can be a product, service, process, and technology. Recently, Tidd and Bessant (2018) classified innovations as the 4Ps, for product innovation, process innovation, position, and paradigm innovation.

Another categorisation of innovation divided innovation into closed innovation and open innovation. Close innovation means that companies focus on internal sources and abilities to develop new technologies and apply them in their products or services, while the open innovation, which has been developed over the past decades, means that many businesses across industries now obtain a significant volume of their technologies from external sources (Lichtenthaler, 2009). In this context, open innovation defines innovation processes in which the borders of the firm are porous. As the previous studies showed, depending on the purpose of the study, the innovation can take various forms, such as organizational innovation, product innovation, process innovation, and marketing innovation (Cascio, 2011). Table 2.10 shows the type of innovation and definitions.

Table 2.10
Types of Innovation and Definitions

Types	**Definitions**
Product innovation	Changes in the things (products/services) to better ones that an organisation offers (Hauser et al., 2005).
Process innovation	Changes in how they are created and delivered; new or significantly improved methods in the production or manufacturing process (Fagerberg, 2004)
Position innovation	Changes in the context in which the products/services are introduced (Bessant & Tidd, 2007)
Paradigm innovation	Changes in the underlying mental models frame what the organisation does (Bessant & Tidd, 2007)

Types	Definitions
Organizational innovation	It is adopting an idea or behaviour that is novel to the business (Damanpour, 1991)
Marketing innovation	“The implementation of a new marketing method involving significant changes in product design or packaging, product placement, product promotion, or pricing” (Mortensen & Bloch, 2005, p. 49)
Closed innovation	Focused on internal sources and abilities to develop new technologies and apply them in their products or services (Lichtenthaler, 2009)
Open innovation	Innovation processes, in which the borders of the firm are poriferous (Lichtenthaler, 2009)

2.8.2 Innovation Process

BDA is considered a vital driver of the innovation process and marketing (Dong & Yang, 2020), productivity, and growth (Niebel et al., 2019). The innovation process is supposed to bring several profits to the business and help a business achieve a competitive advantage (Baer & Frese, 2003). Competitive differentiation and business model innovation can frequently be obtained by process innovation and supporting sources, organisation and data (Weinman, 2015). Thus, firms frequently make substantial attempts to innovate their processes and goods to obtain revenue and keep or enhance profit margins (Amit & Zott, 2012). However, adopting the innovation process in a firm does not happen easily. Researchers such as Frenkel (2003); Galia and Legros (2004); Hadjimanolis (1999); Martinez and Briz (2000) have found that one of the main barriers of innovation process and innovation is a shortage of skilled workers.

Furthermore, Wymbs (2016) argued there were 532,337 postings for Data Analytics in the United States in 2013, as the *Burning Glass Survey* reported. The McKinsey Global Institute reports that the United States alone will face a lack of 140,000 to 190,000 people with high analytical skills, as well as a shortage of 1.5 million data-savvy managers with the expertise to analyse BD to make effective decisions in 2018 (Chen et al., 2012). Also, it is also estimated that by 2020, 5 million jobs will be gone, old skills of employees will become obsolete, and new ones will be needed (Sousa & Rocha, 2019). It is foreseeable that there will be another heated competition about human resources in BD developments in different countries, regardless of whether they are developed countries or developing countries (Chen & Zhang, 2014). However, they did not specify the types of workers' skills. There are several abilities and skills such as effective use of data aggregation, effective use of data interpretation tools, effective use of data analytics, and effective use of data management (Hajli et al., 2020). Overall, if users of BDA do not have the appropriate analytical skills, achieving benefits, improved-decision making, and innovation within firms may be hampered (Waller & Fawcett, 2013).

"Innovation process" is defined as the development and choice of ideas for innovation and transformation of these ideas into innovation (Eveleens, 2010). According to Gunday et al. (2011), the innovation process is the enforcement of a new or noticeably improved production or delivery process. This process contains massive changes in methods, equipment, hardware and software. These changes are often costly and time-consuming, needing a significant investment in everything from study and improvement to specialised sources (Amit & Zott, 2012). However, it can be expected to decrease production or transfer unit costs, improve quality, or produce or deliver new or significantly upgraded outcomes (Bloch, 2007).

Innovation process literature shows that scholars have arisen some points regarding the adoption of the innovation process in the organisations, for instance, Hashi and Stojčić (2013) stated that the first essential step of the innovation process is innovation throughput, which involves the transformation of innovation efforts into innovation productivity (Hashi & Stojčić, 2013). In this regard, Pisano (2015) identified two dimensions of innovation. (a) the degree to which it engages a technology change and (b) the degree of engagement of a business model shift. Furthermore, Hashi and Stojčić (2013) believed in two essential stages required to establish the firm's innovation process: (a) the decision to innovate and (b) the amount of investment in innovation. According to Desouza et al. (2009), there are four stages for the innovation process that are most common: (a)generation and mobilisation, (b) advocacy and screening, (c) experimentation, commercialisation, (d) and diffusion and implementation (see Figure 2.6).

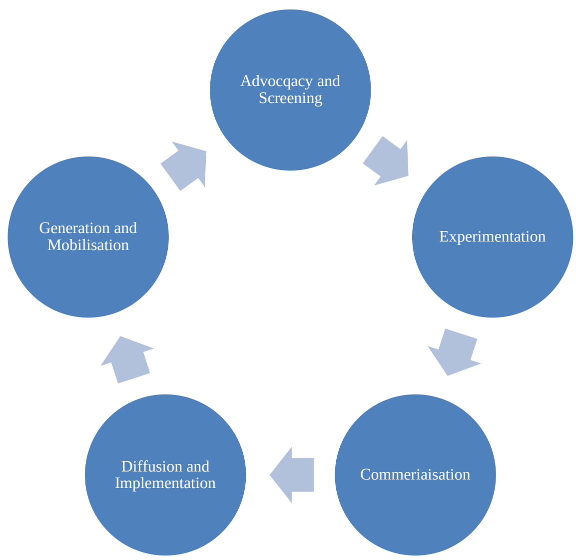


Figure 2.6: The Innovation Process
Sources: Adopted from Desouza et al. (2009); Katal et al. (2013); Uddin & Gupta, (2014)

Furthermore, Tidd and Bessant (2020) stated that the core innovation process revolves around four major activities:

- Searching: scanning the external and internal environment to find insights, opportunities, and threats.
- Selecting: determining the best insight to act according to the strategic objective.
- Implementing: converting the initial idea to a specific service and launching it in the market.
- Capturing value: learning from experience and supporting the adoption of future innovation.
- There are new techniques and approaches, such as dashboards and simulations, to transform insights into actions. These techniques can create value by generating results that can be easily understood and dealt with (LaValle et al., 2011).

However, Lichtenthaler (2017) said that although innovation's significance for value creation has been acknowledged, workers' activities and capabilities that firms pledge remain unclear. Indeed, the application of BDA remains incomplete if it is not used to enhance the innovation process in the firms (Ramadan et al., 2020).

2.8.3 Rules of BDA in the Innovation Process

BDA technology plays an essential role in organisations with strong analytics abilities, not just in innovating existing operations but also in new services, products, processes, and whole business models (Ransbotham & Kiron, 2017). It has been identified as the "next big thing for innovation" (Côrte-Real et al., 2017). Scholars identified eight core processes as tools that can support the innovation process execution effectively:

- problem definition;
- information gathering;
- information organisation;
- conceptual combination;
- idea generation;
- idea evaluation;
- implementation planning; and
- solution monitoring.

According to Mariani and Wamba (2020) in respect of using BDA in customer goods company innovation, the innovation process requires nine stages:

- data generation through experiment;
- data collection and streaming;
- data quality assurances;
- data integration;
- data preparation;
- data analysis;
- data visualisation and reporting;
- data storage; and
- data use for firm's innovation process.

BDA is not only to create the idea of innovation, but also it includes whole the processes starting from the idea to the implementation and marketing (Koman et al., 2018).

2.8.4 Innovation Process Models

Innovation process models have proliferated since the 1950s, intending to explore or guide innovation within industrial companies (Hobday, 2005). It is defined as an implementation to analyse technology transfer decisions (Rice & Rogers, 1980). Researchers such as Cooper (1990); Du Preez and Louw (2008) have categorised the innovation process models into different generations. For instance, Robertson (1967) said the innovation process consists of four stages: participation of the problem, the setting of the stage, the act of insight, and the critical revision. Furthermore, the stage-gate system is a conceptual and operational model that demonstrates the creativity of a new product from idea to launch, and it includes five gates and five stages:

- preliminary assessment;
- detailed investigation (business case);
- preparation;
- development;
- testing and validation; and
- full production and market launch (Cooper, 1990).

Historically, the innovation process models have been categorised into six generations, starting from simple linear models to more interactive and complex (Du Preez & Louw, 2008). Figure 2.7 summarise the innovation process models and their features.

Complexity	Innovation Process Models	Features
Simple Linear	First and second generation (The 1950s - early 1970s)	Describing innovation as either being pulled by market needs or driven by technology and science
	Third generation (The mid-1970s -1980s)	The Collaborative Innovation (CI) process-developed at United Technologies Research Centre (UTRC) - focused on the conceptual design phase of product innovation
	Fourth generation (Late 1980s-early1990s)	The interactive approach, which views the innovation process as similar activities across organisational functions
	Fifth generation/network models (Post-1990)	Originated in the 1990s and attempted to explain the complexity of the innovation process
Complex	Sixth generation	Can be called open innovation models

Figure 2.7: Generation of Innovation Process models

Source: Adapted from (Rogers, 1996; Rothwell, 1994)

2.8.4.1The First Generation

First-generation innovation process refers to the post-World War II cycle that saw unprecedented economic growth in many sectors. Many of the new industries, mainly based on new technological opportunities, appeared at this stage. These developments have also created new employment opportunities, and rising prosperity increased purchasing power and the demand for new goods and services. Under such conditions, the industrial innovation process was generally seen as a linear progression from scientific discovery to the marketplace through technological development in firms. This first technology push concept of innovation assumed that "more R&D in" resulted in "more successful new product out" (Rothwell, 1994). Thus, the first generation is shown in Figure 2.8.

Figure 2.8: Technology Push - First Generation

Source: Rothwell (1994)

2.8.4.2 Second Generation R&D

In the 1960-1970s, R&D departments started to link with the rest of the company function sectors. This new approach at the project level identified the benefit of linked insight required to complete goods development. Extended dependence fostered proactive cooperation and interaction across technology sectors and within the other departments in across-factory units, showing more focus on the market. Technology systems were fundamentally databased, dependent upon statistical analysis and synthesis. The project was acknowledged as a managed asset. In both first and second-generation, the focus was upon client retention (Rogers, 1996). See Figure 2.9.

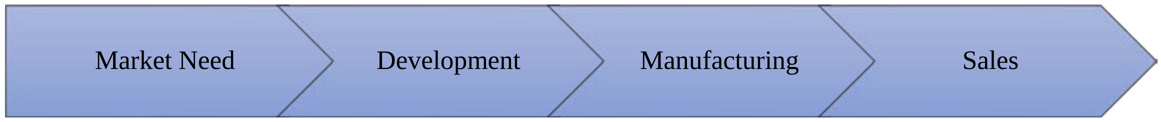


Figure 2.9: Demand-Pull Model Second Generation

Source: Rogers (1996)

2.8.4.3 Third Generation R&D

In third-generation R&D (mid-1970s-1980s), executive managers and R&D managers acted as partners to combine their insights in deciding what to do and why and when given each business's requirements and the organisation. They realistically evaluated expenses, benefits, and risk/remuneration, and they fulfilled these factors within a portfolio of R&D exercises that best met the enterprise's goals. In the third generation, companies took a comprehensive look at the scope of their R&D activities. They were managing their R&D activities in a way that combined with the rest of the firm's sectors to improve the sense of partnership among R&D managers and their general manages (Roussel et al., 1991).

2.8.4.4 Fourth Generation R&D

The fourth-generation R&D (early 1980s-1990s) department is concerned with three crucial administrative requirements, creativity, network links and knowledge exploitation (Liyanage et al., 1999).

2.8.4.5 Fifth Generation/Network models R&D

From 1990, the fifth-generation module emphasised the learning that goes on within and among companies, suggesting that innovation generally and principally is a shared networking process (Hobday, 2005).

2.8.4.6 Sixth Generation R&D

Figure 2.10: The Cambridge Business Model Innovation Process

Source: Geissdoerfer et al. (2017)

According to Geissdoerfer et al. (2017), knowing the resources that feed the innovation process is vital for at least three reasons.

- The innovation process itself is a significant source to improve productivity. Understanding factors and the different characteristics that cause firms may drive more specific knowledge regarding economic development resources.
- The innovation process can permit companies to gain a competitive advantage; consequently, a more attractive understanding of process innovation supports a higher appreciation of factors companies began to achieve and keep a competitive advantage.
- The innovation process is a crucial factor in government innovation strategy and examining the different elements that evoke process innovations explains the tools that encourage expanded private-sector innovation.

In addition, Sergeeva and Trifilova (2018) tackled the innovation process through to the role of storytelling and identified four stages of an organisation's innovation process: the generation of innovative ideas made by the employee; development of innovating ideas made by senior management; implementations of innovating ideas (events, forums, and digital platforms); and promotion of innovative ideas (clients, investors, and public). Before that, Acklin (2010) summarised the innovation process in four stages: idea generation, idea selection, concept development, and implementation. However, "implementation is the critical gateway between the decision to adopt the innovation and the routine use of the innovation"(Klein & Sorra, 1996, p. 1057). Table 2.11 summarises obstacles of innovation and innovation process in different business activities.

Table 2.11
Summary of the Obstacles to Implementing Innovation/Innovation Process in Various Businesses.

Authors	Studies	Findings
(Hadjimanolis, 1999)	Barriers to innovation for SMEs in small, less-developed countries (Cyprus)	They identified two types of barriers (external and internal). Staffs' skills shortage is the second important item of innovation barriers by 71.4%.
(Martinez & Briz, 2000)	Innovation in the Spanish Food and Drink industry	Lack of skilled staff was the fourth essential item hampering innovation activities for 12 factors by 1.769 of 3 scales.
(Frenkel, 2003)	Barriers and limitations in the development of industrial innovation in the region	The lack of high skilled workers was found as a significant barrier to developing new products and processes.
(Galia & Legros, 2004)	Complementarities between abstracts to innovation: evidence from France	Lack of personnel skilled as a barrier to innovation activities is the fifth factor out of six in postponed projects and the fourth factor out of eight abandoned projects.
(Madrid-Guijarro et al., 2009)	Barriers to innovation among Spanish manufacturing (SMEs)	Fifteen kinds of barriers were divided into groups (external & internal).

According to Hadjimanolis (1999), who tested less-developed Cyprus countries, identified two kinds of innovation barriers: (a) external obstacles that can be divided into demand, supply, and environment-related and (b) internal obstacles that can be divided into a lack of internal funds, technical expertise or management time, culture and out-of-date accountancy, the top manager's attitude to risk, and employee resistance to innovation. He also found that staff's skills shortage is an important item of innovation barriers by 71.4%. In this regard, (Galia & Legros, 2004) investigated obstacles to French companies' innovation due to 1772 postponed and abandoned projects. Their findings showed that the lack of skilled people is the second central abstract of the postponed project.

Moreover, Frenkel (2003) used a personal interview to identify barriers and limitations in developing industrial innovation in northern Israel in 211 industrial firms; the lack of highly skilled workers was seen as a significant barrier to developing new products and processes.

What is more, Martinez and Briz (2000) identified 12 items as the main obstacles to the innovation process, which are classified into economic factors and innovation potential; the outcomes of the study indicated that in addition to other factors, lack of skilled staff could increase risks of the innovation process.

Acklin (2010) highlighted some of the barriers of innovation capabilities on the businesses: lack of experiences and awareness, lack of confidence and certainty in the worth, lack of knowledge in obtaining specialises assistances, and limitations on ambition and risk-taking. Further, Madrid-Guijarro et al. (2009) studied barriers to innovation among 294 Spanish SMEs with high capabilities of innovation and fast-growing industries. The authors classified innovation barriers into 15 items and divided them into two groups: (a) internal barriers, which include lack of financial resources, insufficient human resources, weak financial position, and high cost and risk; and (b) external barriers, which include turbulence, lack of external partners' opportunities, lack of information, and government support. Nevertheless, the most barriers to innovation to importance included high cost, innovation cost challenging to control, insufficient government support, economic turbulence, lack of qualification.

2.9 Competitive Advantage

The primary and logical concept of the competitive advantage philosophy is value creation (Barney, 1991). Therefore, studying firms' competitive advantage and the resources that could provide a sustained competitive advantage has been the focus of serious attention to many pieces of research since the 1980s (Liu, 2013). These studies led to several theories, such as the theory of competitive company behaviour and resource- advantage, which has lately been articulated in management, marketing, and socioeconomics (Hunt, 1997). In addition, the Resource-Based View (RBV) research focused on building theoretical and practical relations between different source types and the improvement of sustained competitive advantages (Kwon et al., 2014).

2.9.1 Big data and Competitive Advantage

Over the past 2 decades, competitive advantage in companies has been affected by IT, digitalisation, and BD. Businesses are increasingly involved in competitive dynamics approved or caused by IT (Pavlou & El Sawy, 2010). According to Porter and Millar (1985), IT affects competition in three critical ways. First, it changes industry infrastructure and sequences the play rules. Then it builds a competitive advantage by creating firms with new methods to better their competitors. Finally, it reproduces completely new businesses, usually from within a company's existing processes. As rare and valuable resources can provide or increase the competitive advantage, BD as valuable resources can play a vital role in competitive advantage. Therefore, a data burst is frequently becoming a requirement for a business to continue competing and is a new twist to the old proverb, "Knowledge is power" (Brownlow et al., 2015).

Today, BD aggregation and analytics are speedily becoming a modern edge of competitive differentiation (Bughin et al., 2010). For BD to become a competitive advantage for a company, one has to take into account the broader context in which the data were created, where companies often collect data without a thorough understanding of it (Charles & Gherman, 2013). Scholars such as Brown et al. (2011) argued that BD probably becomes a novel corporate asset, which cuts across business units and functions much as a strong brand does, becoming a fundamental source for competition. It could increase the business competition by modifying processes, altering corporate ecosystems, and supporting innovation (Brown et al., 2011).

2.9.2 Resources of Competitive Advantage

Barney (1991) pointed out five potential resources which could lead to competitive advantage: product and process technology, protected and regulation marketplaces, access to financial sources, economic of scales, and workforce and how it is managing. Similarly (Grant, 1991) highlighted six primary probably resources that can lead to competitive advantage: financial resources, human resources, physical resources, technological resources, organisational resources, and reputation. However, innovation and organisational learning can play critical roles in increasing companies' competitive advantage (Jiménez-Jiménez & Sanz-Valle, 2011). Innovation can be a source of competition in a firm, so three critical factors to consider are awareness of the importance of innovation and consequences of action, motivation to act, and capacity and ability to act (Chen, 1996). Therefore, the framework and method of resources can significantly influence firm heterogeneity and sustainable competitiveness suggested that the framework and method of resource selection significantly influence firm heterogeneity and sustainable competitiveness. In this context, Pisano (2015) argued that technological innovation is a great builder of economic value and an engine of competitive advantage.

BD management can be another source of competitive advantage as it is similar to IT management, where some scholars stated that managing IT capacity can create distinction and give businesses a competitive advantage (Bhatt & Grover, 2005). Some evidence from the banking sector confirmed that using BD has increased their competitive advantage (Brownlow et al., 2015). In addition, BDA can improve the dynamic and adaptive capability through consumer insights obtained from big data, facilitating value creation in various marketing activities (Erevelles et al., 2016). In other words, value is created because of improved decision-making enabled by BD. This created value may result in a sustainable or temporary competitive advantage. Table 2. 12 summarises some of the studies using BD to obtain a competitive advantage and their findings.

Table 2. 12
Summary some of Previous Studies on Competitive Advantage

Author/s	**Study**	**Findings**
(Oliver, 1997)	Sustainable competitive advantage	The framework and method of resource selection have a significant influence on firm heterogeneity and sustainable competitive.
(Jiménez-Jiménez & Sanz-Valle, 2011)	Innovation, organizational learning, and performance	Testing relationships using SEM with data from 451Spanish firms. The findings show that both variables-organisational learning and innovation-contribute positively to business performance and that organisational learning affects innovation.
(Brownlow et al., 2015)	Data and analytics - Data-driven Business Models: A Blueprint for Innovation	Testing seven key competitive advantages (shortened supply chain, expansion, consolidation, processing speed, differentiation, and brand) by six questions in five sectors of data-driven business (telecommunications, retail, publishing, Insurance, and Finance)
(Shan et al., 2019)	BDA adaptation and enterprises' competitive advantages: the perspective of dynamic capability and resource-based theories	The results confirm that firms can obtain competitive advantages by BDA practice.

2.10 Firm Performance

According to Baker and Sinkula (2005), firm performance is a multidimensional contract consisting of more than just financial performance. It refers to the firm's ability to meet the shareholders' expectations, improve sales, retain customers, and its own needs for continuity and survival (Al-Alak & Tarabieh, 2011; Mithas et al., 2011; Tippins & Sohi, 2003). Inman et al. (2011) stated that firm performance contains two dimensions: financial and marketing performance at the corporation level. Therefore, improving the firm's performance needs to be at the heart of any successful command transformation into the market-oriented activities (Estrin, 2002).

A substantial number of studies in the BD literature have been conducted to explore the relationship between BDA and firm performance, considering the financial and operational performance (Yasmin et al., 2020). Scholars such as Akter et al. (2016); Mikalef et al. (2019a, 2019b); Ren et al. (2017b); Wamba et al. (2017) investigated different factors that may influence the innovation process and firm's performance. These studies confirmed that BDA could improve firm performance through creative and innovative capabilities (Kwon et al., 2014). For example, Ren et al. (2017b) examined three factors of BDA System Quality, BDA Information Quality, and BDA Business Value on firm performance. They found that system quality explains 70% of the variance of information quality, 74% of the variance of business value and 76% of the variance of firm performance. Hence, the current study will examine the effective use of BDA tools and BD management, innovation process, and competitive advantage on firm performance.

2.10.1 Financial Performance

The most significant priority for firms is achieving a higher growth performance concerning their financials (Gupta et al., 2020). Financial performance refers to the firm's ability to improve profitability and return on investment. It could result for several reasons. For example, the ability of a firm to introduce new services or products, enter new markets and have a more significant market share more than competitors (Raguseo & Vitari, 2018). In order to enhance financial performance, companies face increasing pressures (Hull & Rothenberg, 2008). BDA could be used to maximise a firm's financial performance. BDA assistant firms apply advanced analytical skills to gain property information from BD, enabling firms to obtain higher operational efficiency (Ali et al., 2020). A sample example of the effect of BDA is reducing maintenance costs by accurately determining equipment downtime (Bucur, 2015).

2.11 Chapter Summary

This chapter has reviewed existing literature regarding the concept of BDA, innovation, the adaption, and adoption of BDA by different sectors. It also highlighted how BD stream is putting pressure on firms to adopt it. The literature review revealed that most of the research conducted on BDA tended to focus on the importance of BDA and investigating its impact on value-creating in different sectors such as healthcare, value chain, and firm performance.

However, little research has looked at BDA in an innovation context, what are the components of factors that may lead to successful BDA, and then how do these analytics affect the innovation process, competitive advantage, and financial performance, suggesting a literature gap. This gap in the literature provides a fertile opportunity to conduct research. Consequently, the justification presents the impetus to conduct this research, attempting to fill the gap and contribute to a high body of knowledge (Gandomi & Haider, 2015).

The next chapter will discuss theories and relationships explored in existing relevant research to develop a conceptual model and its justification and formulate hypotheses, which form the basis for a framework to conduct this experimental research.

Chapter Three:
Conceptual Model and Hypotheses Development

3.1 Introduction

The previous chapter reviewed the existing literature on BDA and identified the focus of the research, which allowed this chapter to develop a conceptual model and formulate the hypotheses. The conceptual model developed is based on several previous studies. In addition, two components that can impact the innovation process (effective use of BDA tools and BD management) have been selected along with, and competitive advantage. Considering the assumptions of this study, hypotheses have been defined and formulated. This chapter also will discuss and present the theoretical background that was chosen to support this research.

3.2 Theoretical Background

Reviewing theoretical frameworks used to explore BDA and innovation in related areas may provide an appropriate framework for studying BDA in an innovation context. Therefore, highlighting the different theories that have guided previous studies might help build a suitable framework to explore components of BDA to enhance the innovation process ability, obtain a competitive advantage, and firms' financial performance.

Previous studies in BD and its analytics have used theories like Resource-Based View (RBV), innovation diffusion, and network theory, which discuss the benefits of applying the information system (Ghasemaghaei & Calic, 2019). Furthermore, Wang et al (2019) named three groups of studies that tried to explore the influence of BDA on business growth: the lens of resources-based theory (RBT), the knowledge-based view (KBV) and the information processing view (IPV). These approaches offer selective insights into the phenomenon of BD, and they provide a framework to examine BD processes, relationships and resources (Braganza et al., 2017). Hence, several theories will be discussed in detail in the following subsections with an emphasis on those underlying this study.

3.2.1 Resource-Based View (RBV)

RBV has been broadly recognised as one of the most notable theories used to explain how companies obtain and maintain competitive advantage due to their own or have under their control resources (Barney, 2001). According to the assumptions in the RBV, a firm is seen as a set of valuable tangible and intangible resources, which can be merged to create a competitive advantage (Peteraf, 1993). Akter et al. (2016) and DeSarbo et al. (2007) stated that the fundamental stone and fundamental components of the RBV are resources and capabilities, and it focuses on attributes of resources/capabilities and the relationship with competitive advantage and performance of the firm (DeSarbo et al., 2007). RBV explains how to turn resources into a sustainable competitive advantage (Barney, 1991; Barney, 2000). The notion of the firm's resources was further divided to include resource picking and capability building, two different forces primary to the RBV(Mikalef et al., 2017). “Resources” refers to tangible and intangible assets such as workers, technology, and capital, while the “capabilities” refers to subsets of resources that are not transferable and objective to boost the efficiency of other sources (Makadok, 1999). IT scholars usually use RBV to support their studies (Bhatt & Grover, 2005) as IT is a resource that can build uniqueness and grant firms a competitive advantage (Bharadwaj, 2000). Reviewing the BD literature as a type of IT reveals that most researchers have applied RBV theory in terms of supporting their studies regarding BDA and its components. For instance, in health care (Wang & Byrd, 2017; Wang & Hajli, 2017), manufacturing (Cao & Duan, 2014; Popovič et al., 2018), retail (Wamba et al., 2017), and other domains (Akter et al., 2016; Gupta & George, 2016; Kwon et al., 2014; Ren et al., 2017a). In addition, Kozlenkova et al. (2014) said that RBV theory had been used in marketing research continuously, where scholars counted 173 direct implementations of RBV in the marketing literature from the early 1990s to 2012. Another estimate was that implementations of RBV have risen by more than 500% in the last decade, which presents its importance as a structure for describing and forecasting the basis of its competitive advantage and performance (Barney et al., 2011; Kozlenkova et al., 2014). Thus, based on the RBV views, a company can achieve the full competitive advantage according to its capabilities and resources when they are rare, valuable, and hard to imitate (Barney & Hesterly, 2009). Moreover, the richness of the firm in terms of the resources will increase the ability and agility of the company to use new resources (Kwon et al., 2014).

Likewise, the central hypothesis in the IT capability concept is that although resources can be replicated, a firm's distinctive abilities cannot be easily aggregated via the markets and can therefore be a source of sustained competition (Lu & Ramamurthy, 2011). Regarding resources, there are two essential assertions on which the RBV is based. First is resource heterogeneity, which means the capabilities and resources owned by competition firms may differ, and second is resource immobility, which mean the differentiation of resources may be long listing (Peteraf & Barney, 2003).

Previous studies have suggested the existence of relationships between the use of BDA tools and firm performance (Evans & Lindner, 2012). For instance, Bharadwaj et al. (2013) suggest that processing large amounts of data from various sources helps firms to improve their decision quality and decision-making efficiency and better make their business strategies which could directly improve their performance. Similarly, Wang & Hajli (2017) argue that prescriptive analytics help firms make thorough decisions. In the current study, the effective use of BDA tools is contextualised as physical resources. While effective BDA management is considered human capital. Furthermore, as in the context of BD, "data" by itself is considered a critical firm resource in affecting firm outcomes (Ghasemaghaei et al., 2017). Based on the RBV, BDA tools as tangible resources that could improve firm performance (Ghasemaghaei, 2018). The level of effective use of BDA tools and its depth of analysis could provide descriptive, predictive analysis and simulate ideas (Delen, 2013), which in turn can enhance the innovation process, competitive advantage, and firm performance.

Furthermore, previous IT capability research has suggested technical and managerial skills as critical dimensions of human resources for IT (Bharadwaj, 2000; Chae et al., 2014). In addition, RBV implies that data quality management is an intangible firm resource (Kwon et al., 2014). In this study, human capital (BDA management) has been contextualized as the management's analytical and managerial skills in extracting insights and ideas from data analytics. BDA management is highly firm-specific and is developed over time by individuals (Dubey et al., 2019). Moreover, mutual trust and a good working relationship between BD managers and other functional managers will likely enhance superior human BD skills, which may be difficult to copy by other organizations (Dubey et al., 2019). Therefore, according to RBV, the BDA management could be considered a resource that can render competitive advantages to firm.

Hence, it is vital to understand the role of BDA management in the innovation process, competitive advantage, and firm performance. Therefore, this study assumed the components of (effective use of BDA tools and BDA management) as valuable and rare resources that can enhance the BDACs, which provides insights that can feed the innovation process stage and create a competitive advantage to improve financial performance. Therefore, the selection of the RBV as the underlying theoretical framework is considered suitable for this research. Table 3.1 shows the tangible and intangible resources of firms in terms of BDA.

Table 3.1
Tangible and Intangible Resources Factors

Factors		**Sources**
Tangible Resources		
Data	Data standardization Data openness Data quality Data privacy & security	(Gupta & George, 2016; Kim & Park, 2017; Popovič et al., 2018)
Technology	Data aggregation & processing Data storage	(Chen et al., 2015; Gupta & George, 2016; Kim & Park, 2017)
Basic resources	Basic resources Investment	(Gupta & George, 2016; Kim & Park, 2017; Popovič et al., 2018)
Intangible Resources		
Data-driven culture	Data-driven culture	(Cao & Duan, 2014; Dutta & Bose, 2015; Gupta & George, 2016; Janssen et al., 2017; Wang & Byrd, 2017)
The intensity of organizational learning	The intensity of organizational learning	(Gupta & George, 2016)
Perceived benefits	Perceived benefit of external data usage	(Kwon et al., 2014)

Source: Adopted from Adrian et al. (2018)

3.2.2 Dynamic Capability View

The Dynamic Capabilities View (DCV) has appeared as one of the most critical theoretical viewpoints in the research of technology and strategic management in the recent decade (Schilke, 2014). Extending the resource-based view of the firm, the DCV efforts to describe how a firm keeps a competitive advantage in turbulent environments (Eisenhardt & Martin, 2000). In addition, the DCV quotes a company's ability to regeneration itself in the cope challenges of an ever-changing environment (Eisenhardt & Martin, 2000; Teece et al., 1997). During the last decade, there have been significant attempts in defining and distinguishing the boundaries and conditions that discriminate dynamic capabilities theory (Mikalef et al., 2017). According to, Teece (2007) there are three of the foundations of dynamic capabilities includes:

- *"Analytical systems (and individual capabilities) to learn and to sense, filter, shape, and calibrate opportunities."*
- *"Enterprise structures, procedures, designs, and incentives for seizing opportunities."*
- *"Continuous alignment and realignment of specific tangible and intangible assets" (p. 1326).*

The outcomes of the DCV are the conversion in firm capabilities and companion resources (Easterby-Smith & Prieto, 2008). It must be mentioned that some scholars pointed out that the RBV and dynamic capabilities theories are tautological, routines for learning routines and fuzzy, and they wonder if the capabilities in question are in place. Basically, they ask, Is it real? Does it create a competitive advantage for the firm (Kraaijenbrink et al., 2010; Priem & Butler, 2001; Williamson, 1999; Winter, 2003)? This criticism appears to be based on the claim that competitive advantage can only be confirmed after the firm achieves that advantage in most resource-based studies, and then the acquired advantage is attached to the resources and capabilities without taking into account other variables (Prescott, 2014).

The definitions of DCV also have evolved as the research evolves around them, so various definitions can be found. For instance, Helfat and Peteraf (2009) defined the DCV as: “the capacity of an organisation to purposefully create, extend, and modify its resource base’(p. 4). Further, Teece et al. (1997) stated that DCV operates on “organizational skills, resources, and practical competencies” (p. 515), and it reconfigures capabilities and resources to accommodate changing requirements to gain innovation (Kim et al., 2015), enhancing the competitive advantage (Ambrosini et al., 2009; Kozlenkova et al., 2014). Therefore, according to the DCV, some companies that can change their resources and capabilities to cope with environmental changes may prosper (Eisenhardt & Martin, 2000; Teece et al., 1997). In other words, the DCV relates to the company's ability to sense, seize, and modify to create and apply the company's external and internal competencies and resources while answering environmental changes (Helfat & Peteraf, 2009). Researchers are interested in DCV because it shows the route to a competitive advantage in a turbulent environment. Although the RBV focuses on resources (tangible and intangible) and operational capabilities, the DCV concentrates on meaningful modification of its current resource base (Akter et al., 2020). Teece et al. (1997) suggest that DCV, through reconfiguring and renewing capabilities and assets of a company can serve as a potential means to overcome some of RBV’s weaknesses to continue providing benefits and maintaining the competitive advantage. In a turbulent environment, firms need to keep pace with market changes and newness. Therefore, DCV can affect the firm's strategy by abandoning those procedures and processes that no longer provide an advantage and re-integrate resources into new capabilities that can provide a competitive advantage (Kogut & Zander, 1992; Teece et al., 1997). Hence, the DCV can also be underlying the current study approach, which is looking for those firms that will abandon the traditional data system that can no longer compete and use BDA with new capabilities combined to enhance its innovation process and gain a competitive advantage.

3.3 Conceptual Model and Hypotheses Development

The current study developed the conceptual model proposed based on Duan et al. (2020); Hao et al. (2019); Mikalef, Boura, et al. (2020); Mikalef and Krogstie (2020); Ramadan et al. (2020); Wang and Hajli (2017) and offers 13 hypotheses with several moderators and mediators. Figure.3.1 shows the conceptual model developed to examine the architectural relationships between components (effective use of BDA tools and big data management) on the innovation process, competitive advantage, and financial performance. As shown in Figure .3.1, the proposed model includes four stages. First, the effective use of BDA tools and BD management will be examined on the innovation process, competitive advantage, and financial performance. Then, the influence of the innovation process will be examined on competitive advantage and financial performance. In the third step, the impact of competitive advantage will be tested on financial performance. Finally, the effects of the moderator variable (environment turbulence) and control variables (firm size, firm age, and industry type) will be tested on financial performance. The theoretical background and related previous studies will be discussed in more detail in the following sections.

3.3.1 Related Previous Studies

Scholars examined several factors that can enhance the capabilities of BDA. For instance, Wang and Hajli (2017) tested the impact of data aggregation, data processing, data visualisation on business value in the health care industry. Similarly, Shan et al (2019) examined the effect of IT technology resources, IT relationship resources, IT resources and competitive advantage. Furthermore, Mikalef and Krogstie (2020) investigated the effect of managerial skills and technical skills on the competitive performance of Norway firms. Finally, Akter et al. (2016) classified three principal dimensions that display BDA capability: BDA management capability, BDA technology capability and BDA talent capability. However, the effect of BDA on innovation context needs empirical studies. Hence, the current study investigates the influence of BDA on the innovation process and competitive advantage. Table 3.2 summarises the typologies of BDA resources and assess items of similar previous studies.

Table 3.2
Typologies of BDA Resources and Capability

Source	BDAC	Assessed Item
(Akter et al., 2016)	BDA management capability, BDA technology capability, BDA talent capability	Firm performance
(Wamba et al., 2017)	BDA business analytics capabilities: BDA infrastructure flexibility, BDA management capabilities, BDA personnel expertise capabilities	Firm performance (financial performance and market performance)
(Wang & Hajli, 2017)	Data aggregation, Data processing, Data visualisation	Multidimensional benefit (IT infrastructure benefits, Operational benefits, Organisational benefits, Managerial benefits, Strategic benefits)
(Mandal, 2018)	BDA technical knowledge, BDA technology management knowledge, BDA business knowledge, BDA relational knowledge	Supply chain agility
(Shan et al., 2019)	BDA adoption: (IT technology resources, IT relationship resources, Idle resources)	Competitive advantage
(Mikalef & Krogstie, 2020)	BDA resources (data, technology, basic resources, technical skills, managerial skills, organisational learning, data-driven culture); environment (dynamism, heterogeneity, hostility) organisational factors (organisation size, industry)	Process innovation capabilities (incremental, radical)
(Mikalef & Krogstie, 2020)	BD Predictive analytics: resources and capabilities (managerial skills, technical skills)	Competitive performance

Table 3.3
Summary of BDA Studies and Theories Used

Studies examined the effect of BDA on competitive advantage							
Source	**BDAC/Factors/ Elements**	**Mediating factors**	**Assessed Item**	**Approach**	**Theory Used**	**Major findings**	**County**
(Mikalef & Krogstie, 2020)	BD predictive analytics: resources & capabilities (managerial skills, technical skills)	Dynamic capabilities (Marketing capabilities and Technological capabilities)	Competitive performance	Empirical	RBV	A strong BDAC can help organisations to build a competitive advantage. The effect of BDAC comes through the mediation of dynamic capabilities that have a positive influence on operational capabilities (marketing and technology)	Norway
(Shan et al., 2019)	BDA adoption: (IT technology resources, IT relationship resources, Idle resources)	Strategy flexibility, Compatibility, IT technology capabilities	Competitive advantage	Imperial	RTB and DCTR	The results confirm that firms can obtain competitive advantages by BDA practice.	China
(Mikalef & Krogstie, 2020)	BDA resources (data, technology, basic resources, technical skills, managerial skills, organisational learning, data-driven culture).	Direct	Process innovation capabilities (incremental, radical)	Imperial	RBV	Under various combinations of contextual factors, BDA resources' importance varies, and with specific configurations, high levels of process innovation capabilities are achievable.	Norway
(Mikalef et al., 2019b)	BDAC: tangible (data, basic resources, and technology), human skills (managerial and technical skills), and intangible (data-driven culture and	Dynamic capabilities	Incremental innovation and radical innovation	Imperial	RBV	Results indicated that the dynamic capabilities fully mediate the impact on both radical and incremental innovation. At the high environment heterogeneity condition, the effect of	Greece

Studies examined the effect of BDA on competitive advantage							
Source	**BDAC/Factors/ Elements**	**Mediating factors**	**Assessed Item**	**Approach**	**Theory Used**	**Major findings**	**County**
	organisational learning)					BDACs will be enhanced dynamic capabilities, which in turn enhanced the incremental innovation. Whilst the dynamic high environment will lead to amplifying the dynamic capabilities of incremental innovation capabilities.	
(Akter et al., 2016)	BDA capabilities (BDA management capability, BDA technology capability, BDA talent capability)	Direct	Firm performance	Empirical	RBT	BDAC was found to have a positive relationship with all the initial domains (BDA talent capacity, BDA management capacity and BDA technology capacity)	United States
(Wamba et al., 2017)	BDA Infrastructure Flexibility, BDA management Capabilities, BDA Personnel Expertise Capabilities	BDA Business Analytics Capabilities and Process-Oriented Dynamic Capabilities	Firm performance financial performance and market performance)	Empirical	RBV	Personnel and infrastructure capacities were found relatively more significant than management capability.	China
(Ghasemaghaei & Calic, 2020)	BD characteristics (variety volume, velocity)	Innovation performance (Innovation efficacy, Innovation Efficiency)	Firm Performance (Financial returns, customer perspective, operational excellence)	Empirical	Organizational learning theory	Data variety and velocity positively influence firm innovation performance, but data volume has no notable effect. Also, data velocity practices a more significant role in enhancing firm innovation performance than other BD characteristics.	United States

3.3.2 Big Data Analytics Capabilities (BDACs)

To reach the purpose of this research which seeks to explain the BDA capability profile and its potential benefits for the innovation process and obtain the competitive advantage, building the conceptual model and supporting the hypothesises are needed. It is necessary to understand BDA' architecture, components, and functionalities (Wang, Kung, & Byrd, 2018). BDACs refer to the ability to gather (Simon, 2013), and manage an enormous variety of disparate data (Hurwitz et al., 2013), use resources (Cosic et al., 2012) and analytics to plan (Trkman et al., 2010), gain speed in business insight (Wixom et al., 2013), make targeted investments, improve cost reduction (LaValle et al., 2011), and operation optimisation to perform a business analytics task (Hurwitz et al., 2013). Scholars such as Wang, Kung and Byrd (2018); Wang, Kung, Wang, et al. (2018) have argued that the implementation of big BDACs could affect businesses positively in many ways, such as real-time decision making, managerial, operational, strategic and organisation benefits. Therefore, several studies were inducted to identify and analyse the affecting elements and factors of BDA implementation outcomes. Through a review of BD, the literature illustrates several kinds of BDACs. They can be categorised into six groups that include management capabilities (Akter et al., 2016; Wamba et al., 2017), organisational capability (Chen et al., 2015; Dutta & Bose, 2015; Janssen et al., 2017; Koronios et al., 2014; Popovič et al., 2018), infrastructure flexibility/technology capability (Akter et al., 2016; Dutta & Bose, 2015; Gupta & George, 2016; Koronios et al., 2014; Popovič et al., 2018; Wamba et al., 2017), talent capability/ personnel expertise (Akter et al., 2016; Dutta & Bose, 2015; Gupta & George, 2016; Koronios et al., 2014; Popovič et al., 2018; Wamba et al., 2017), information processing capability (Cao & Duan, 2014; Dutta & Bose, 2015; Wang & Hajli, 2017; Wang, Kung, & Byrd, 2018), and other capabilities (Wamba et al., 2017; Wang & Byrd, 2017). Table 3.4 classified factors of BDACs into six groups which are summarised from different studies.

Table 3.4
Summary of BDACs factors

Factors		References
Management capability	BDA planning Investment decision-making	(Akter et al., 2016; Wamba et al., 2017)
Organisational capability	Collaboration, BDA strategy Information strategy Top management support	(Chen et al., 2015; Dutta & Bose, 2015; Janssen et al., 2017; Koronios et al., 2014; Popovič et al., 2018)
Infrastructure Flexibility/Technology capability	Infrastructure flexibility Process integration and standardization	(Akter et al., 2016; Dutta & Bose, 2015; Janssen et al., 2017; Koronios et al., 2014; Popovič et al., 2018; Wamba et al., 2017; Wang & Byrd, 2017)

Factors		References
Talent capability/ Personnel Expertise	Technical knowledge Technology management capability Relational knowledge Managerial skills	(Akter et al., 2016; Dutta & Bose, 2015; Gupta & George, 2016; Koronios et al., 2014; Popovič et al., 2018; Wamba et al., 2017)
Information processing capability	Analytical capability Patterns of care Unstructured data analytical capability Decision support capability Traceability capability	(Cao & Duan, 2014; Dutta & Bose, 2015; Wang & Hajli, 2017; Wang, Kung, & Byrd, 2018)
Other capabilities	Process-oriented dynamic capabilities	(Wamba et al., 2017; Wang & Byrd, 2017)

Source: Adopted from Adrian et al. (2018)

3.3.2.1 Staff Skills/People Capabilities

Capability building is not limited to merely pooling resources but also involves complex coordination patterns between employees and between employees and other resources (Bharadwaj, 2000). According to Grant (1991), there is a distinction between capabilities and resources. Resources can be classified into three categories: tangible, intangible, and personnel-based resources. Tangible resources are physical assets such as equipment, raw materials, and stocks and a firm's financial capital. Capabilities are subgroups of sources that are non-transferable (Makadok, 1999), tangible and intangible processes (Akter et al., 2016); these special kinds of resources are targeted to enhance other resources' efficiency (Morgan et al., 2009). According to the RBV, firm performance achievement depends on its capabilities of managing all critical resources effectively. Moreover, Galia and Legros (2004) highlighted that highly skilled personnel and scientific experts are essential since they constitute innovation requirements.

Furthermore, Wymbs (2016) argued there were 532,337 postings for data analytics in the United States in 2013, as the *Burning Glass Survey* reported. The McKinsey Global Institute reported that the United States alone will face a lack of 140,000 to 190,000 people with high analytical skills, as well as a shortage of 1.5 million data-savvy managers with the expertise to analyse BD to make effective decisions in 2018 (Chen et al., 2012). Also, it is also estimated that by 2020, 5 million jobs will be gone, old skills of employees will become obsolete, and new ones will be needed (Sousa & Rocha, 2019). It is foreseeable that there will be another heated competition over human resources in BD developments in different countries, regardless of whether developed countries or developing countries (Chen & Zhang, 2014). Effectively utilising BDA outcomes requires skills such as critical thinking and adequate interpretation of those results by managers and employees alike. Thus, firms must either hire individuals with those skills or provide the necessary training to upgrade their skills (Wang, Kung, & Byrd, 2018).

BD and BDA literature reviews showed that many researchers have touched on this topic from different perspectives, trying to measure its effect on decision making, competitive advantage, and firm performance.

3.3.3 Conceptual Model

As mentioned earlier, this research aims to measure effective use of BDA tools and BD management on the innovation process, competitive advantage, and firm performance. Therefore, it is essential to propose a suitable model for context that conceptually reflects the relationships between different constructs. The proposed research model constitutes the key constructs including effective use of BDA tools, BD management, innovation processes, competitive advantage, and financial performance, as well as moderate variables such as environment turbulence, firm size, firm age, and industry type. In addition, each construct of the proposed model presents a hypothesis. The arrows show relationships between constructs. The proposed conceptual model includes 13 hypotheses, as they are shown in Figure 3.1.

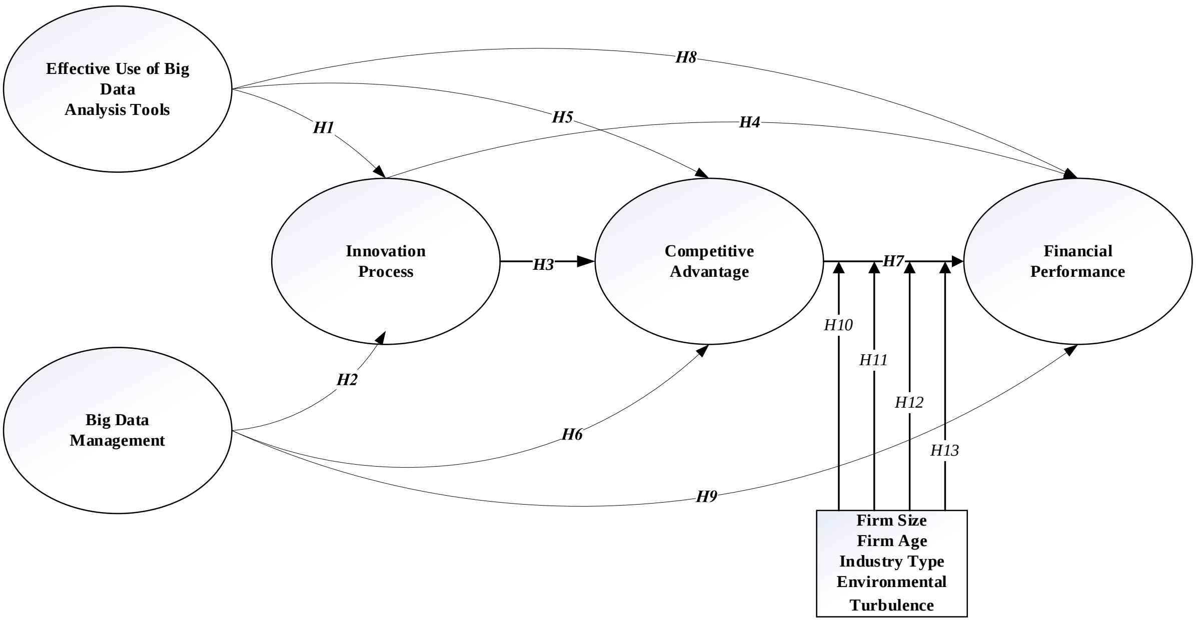


Figure 3.1: Proposed Research Conceptual Model

3.3.3.1 Constructs and Definitions

According to different scholars, the definitions of all constructs came after the figure of the proposed conceptual model directly, as shown in Table 3.5.

Table 3.5
Constructs and Definitions

Construct	Definition	Source(s)
Effective Use of BDA Tools	Effective data analysis refers to the deployment and uses BDA tools to improve operational efficiency, obtain new profits streams, and achieve a competitive advantage over business competitors. BDA can be explained as a process consisting of several phases, including data aggregation, data analysis, data interpretation, and data applications.	(Sivarajah et al., 2017; Xu & Shi, 2015)
BD Management	BD management is an emerging discipline that includes the implementing of methods and software tools and platforms, including storage, pre-processing, and processing.	(Siddiqa et al., 2016)
Innovation Process	The innovative activities depend on the variety and form of its relation to internal sources of a firm, such as data, knowledge, technologies, practices, staff, and financial resources.	(Data, 2005)
Competitive Advantage	Creation of "more economic value than the marginal (breakeven) competitor in its product market."	(Peteraf & Barney, 2003, p. 314)
Firm Performance	Firm performance is the firm's ability to gain and retain customers and to improve sales, profitability, and return on investment (ROI).	(Mithas et al., 2011; Tippins & Sohi, 2003; Wamba et al., 2017)

3.3.4 Effective Use of BDA Tools

The rise of the BD age asks for more efficient and effective data analysis tools to achieve a competitive advantage (Ianni et al., 2020). In this respect, scholars highlighted different stages that can be used to enhance the effectiveness of BDA. This process includes several phases such as data aggregation, data analysis, data interpretation, and data applications (Xu & Shi, 2015).

3.3.4.1 Effective Use of BD Aggregation Tools

Data aggregation is the first step of the data life cycle, and it is a pivotal step as the first form of data is semi-structured and unstructured (Xu & Shi, 2015). Although BD may not be meaningful in small amounts and unstructured form, it is the aggregation of data that leads to finding new patterns and meanings of scattered data (Couldry & Powell, 2014; Khan et al., 2014). Data aggregation is as the systematic gathering of sensed data from various sensors to be eventually transmitted to the central station for processing and eliminating redundancy (Boubiche et al., 2018; Rajagopalan & Varshney, 2006), which requires synchronize outside data sources and distributed big data platforms with the internal infrastructures of an organization (Oussous et al., 2018). Compared with the past, data now can generally be collected calmly, without much effort and energy or even knowledge on the part of those being recorded (Brownlow et al., 2015). Indeed, the concept of data aggregation is beyond just data collecting from internal and external resources, and this term refers to some other tasks involved in this context. Data aggregation includes data collecting, data integration, data cleaning (e.g., removing irrelevant documents, removing irrelevant terms), data reduction, data transformation, data discretization, and data storage (Debortoli et al., 2014; Raghupathi & Raghupathi, 2014; Zhang et al., 2003). Gathering, storing, and analysing BD is not an end in itself for businesses; they are involved in producing real business value (Hartmann et al., 2014).

The aim of processing data collected is the amount of data while maintaining data information (Zhang et al., 2013). However, the processing of data aggregation depends on the sources and objectives may face several challenges. For example, a significant problem in data aggregation is how reports can be directed to the same node; thus, the reports can be merged (Yu et al., 2008). Scholars highlighted four accessible sources: the internet, Mobile phones, Social media, and sensors in collecting large-scale data (Huang et al., 2015). Chen, Mao and Liu (2014) stated four data aggregation methods: Long files, Sensing, Methods for acquiring network data, and Libpcap-based packet capture technology. However, obtaining big data by organisations does not necessarily mean their effective use (LaValle et al., 2011). According to Merino et al (2016), the effective use of BDA refers to effectively using BD technologies that include processing the raw data, storage, analytics, and data management.

3.3.4.2 Effective Use of BDA Tools

The right IT supports firms with the information needed for organisational tasks such as improving a product/service and helping to make the right decisions at the right time (Trieu, 2017). The speed of transaction completion and decision-making depends on the speed of technological techniques used un the firm. The higher the rate, the greater the amount of generated data (Prescott, 2014). Scholars argue that BDA staff's skills are the essential component in enabling companies to leverage the advantages of BDA (Davenport & Patil, 2012). As the different kinds and amounts of BD need various diverse analysis methods (Huang et al., 2015), the methods and techniques of data analytics have changed during the last years from the traditional analytics 1.0 measurement to business intelligence and analytics 2.0, and business intelligence and analytics 3.0 (Abbasi et al., 2016). Indeed, to obtain the enormous potential of BDA, not only will a firm's IT structure require changing, but almost every unit within a firm will also experience modifications (Davenport et al., 2010).

3.3.4.3 Effective Use of BD Interpretation Tools

Another critical challenge of BD is developing algorithms and solutions capable of interpreting and interacting with a massive amount of data into valuable information and meaningful knowledge (Colace et al., 2018; Di Martino et al., 2014). BD required novel tools to deal with BDVs (volume, velocity, and variety) to obtain reliable and accurate information, as 80% of the collected data is in an unstructured format (Gupta & George, 2016). The net outcome of interpretation is often the formulation of views interpreting the base data (Jagadish et al., 2014). Furthermore, the interpretation of data concerns many possible errors. For example, computer systems can have bugs, models usually have assumptions, and decisions can be based on inaccurate data (Jagadish et al., 2014). Data visualisation is an excellent method to interpret and explain data due to converting raw data into meaningful visual information such as information graphics, statistics, and scientific visualisation (Zhong et al., 2016). Therefore, based on the current findings, it is proposed that:

H1: There is a relationship between the effective use of BDA tools and the innovation process.

H5: Effective use of BDA tools benefits competitive advantage.

H8: There is a positive relationship between effective use of BDA tools and financial performance

3.3.5 BD Management

Big data management is an emerging discipline that includes implementing methods and software tools and platforms, including storage, pre-processing, and processing (Siddiqa et al., 2016). BD management is one of the most critical factors contributing to the innovation process and creating a competitive advantage for companies (McAfee et al., 2012). Thus, it is imperative to have a BD management strategy to manage the IT and processes during the data life cycle (Kung et al., 2015). For the importance of BDA, leading organisations have established two new senior executives' positions, the Chief Analytics Officers (Assunção et al.) and Chief Data Officers (CDO), to deal with and manage BDA. Creating BD strategies is the responsibility of the CDO, while the CAO finds effective ways to use analytics across the firm. The essential requirements of those managers' duties are data and related technology. Management perspective in terms of dealing with BD problems includes data aggregation and analysis domains (Geczy, 2014). Firms must also enable the internal capability to manage data quality and enhance data usage experience at a company through focusing on two variables of data quality: data consistency and data completeness (Kwon et al., 2014). BD transforms business by serving as a cross-functional capability that allows managers to align strategies and make decisions following market demands. However, this mission will not ease the massive quantity of data flow, the great variety of heterogeneous data, and the velocity of growing data, and the data may not be satisfactory in terms of quality (Saha & Srivastava, 2014). As data become reachable and cheaper, the role of complements to data becomes more important. Thus, BD scientists and other professionals skilled at working with large quantities of information become the most crucial from data itself (McAfee et al., 2012). Nevertheless, the staff who deal with BD requires substantial and creative IT abilities. They also need to close to processes and productions within companies, which means they require different organizing compared with analytical workers in the past (Davenport et al., 2012).

Further, according to Powell and Dent-Micallef (1997), the wide variety in competitive advantage and economic profits that firms obtain from information technology pauses on a management difference and not a technical difference. In addition, McAfee et al. (2012) stated that managers involved in managing a BD transformation could begin with two simple methods. First, they can get in the habit of asking, “What do the data say?” Second, they can let themselves be overruled by the data. Therefore, high-performing teams recognise their biases, disagree constructively, synthesise opposing viewpoints, and learn better and with more agility than others do. Comparative learning rates are essential because the capacity to learn quicker than competitors is sometimes considered the single source of sustainable competitive advantage (Liebowitz, 2013). According to Buhl et al. (2013) data value must be as a firm asset must be understood. To maximise this goods’ utility, data management must ensure high data quality as a basis for any initiative or innovation. In terms of exercising this vital role, data management is supposed to have proficiency, experience, and full supervision of technical and managerial skills.

Meanwhile, several studies in recent years argued that there is a shortage of knowledge in BD management to fully exploit its potentials to develop new insights and to contribute to diverse socio-economic (Bello-Orgaz et al., 2016; Chen & Zhang, 2014; Mikalef et al., 2017).

Furthermore, Shamim et al. (2019) assessed management functions concerning BD and highlighted four management roles to enhance BD decision-making: leadership focus on BD, talent management for BD, technology for BD, talent management for BD, technology for BD, and organisational culture of BD. This study focuses on two skills that may positively affect the effective management of BDA: technical skills and managerial skills. Each skill will be evaluated with relevant questions from participating managers of firms operating in the United States and United Kingdom.

3.3.5.1Technical Skills

BD's technical management skills refer to the knowledge required to work and operate new forms of technology to extract intelligence and value from BD. Some of these skills include abilities in machine learning, data cleaning, data extraction, and statistical analysis and knowing programming models such as MapReduce (Uddin & Gupta, 2014). Scholars (e.g., Mata et al. (1995); McAfee et al. (2012) argued that managerial IT skills are rare and firm-specific, and it is always a fundamental component of a BD strategy. Therefore, it is likely to serve as a source of sustained competitive advantage. Scholars argued that managerial IT skills are rare and specific. IT is always a fundamental component of a BD strategy, and so is likely to serve as a source of sustained competitive advantage.

3.3.5.2 Managerial Skills

Using BD in and of itself does not necessarily yield significant benefits unless many issues are addressed, including related managerial challenges (Shamim et al., 2019). Scholars such as LaValle et al. (2011) argued that managerial challenges are more significant compared with technical issues. Firms can improve technical skills by training their current staff or recruiting new talent; however, managerial skills are highly company-special and develop over time by individuals working in the same firm. These skills are improved due to strong interpersonal bonds between corporate staff working in the same sections or different departments (Uddin & Gupta, 2014). The management of BD must have a strong tie with managers in charge of goods and services (Davenport & Patil, 2012). Effectively using BD at scale may lead to competitive advantage for companies (Brown et al., 2011).

BD is a treasured resource for the firm because it can present a glimpse into customers' minds. However, the use of BD for a competitive advantage depends mainly on the capabilities and understanding of management to use that data (Kamioka & Tapanainen, 2014). Hence, based on the current findings, it is proposed that:

H2: BD management has a positive impact on the innovation process.

H6: There is a positive relationship between BD management and competitive advantage.

H9: BD management impacts financial performance positively.

3.3.6 Innovation Process

Firm innovation activities depend on the variety and form of its relation to internal sources such as data, knowledge, technologies, practices, staff, and financial resources (Data, 2005). The innovation activities could happen in the product, process, and organisation, and even in different forms such as radical innovation and incremental innovation (McDermott & O'connor, 2002); however, this study focuses on the innovation process as the central area. The power of BDA as a tool lies in its ability and transformative power to be used in the innovation process (George et al., 2014).

In contrast to most BDA studies, Ghasemaghaei and Calic (2019) performed a study to examine the influence of BD characteristics (volume, velocity, variety, veracity) on firm innovation competency mediated through generating data-driven insights (predictive, descriptive, and prescriptive), that included 280 managers. They confirmed that each BD characteristic has a different impact on improving firm performance, and by knowing each feature's effect, directors can enhance the firm's innovation competency. Furthermore, Mikalef et al. (2019b) examined the impact of BDACs through three factors: tangible resources, human skills, and intangible resources on two types of innovation (incremental and radical) with dynamic ability as a mediating factor. The study involved 175 chief information officers and IT managers of Greek companies. The authors found that dynamic capabilities fully mediate the influence of both radical and incremental innovation. In the event of high environmental heterogeneity, BDACs will enhance the impact of dynamic capabilities, which in turn have fostered incremental innovation. In contrast, a highly dynamic environment will amplify the dynamic capabilities of incremental innovation capabilities. Therefore, based on the current findings, it is stated that:

H3: There is a positive relationship between the innovation process and competitive advantage.

H4: The innovation process has a positive influence on financial performance.

3.3.7 Competitive Advantage and BDA

Scholars have identified BDACs and resources as a potential strong foundation to promote firm performance (Al-Sai et al., 2020). The competitive advantage refers to implementing a value creation strategy that is not performed simultaneously by any current or potential competitors (Barney, 1991). Successful competitive strategies lead to superior profitability in a firm, and the significant impact of competitive advantage will be reflected in financial and market performance. However, to obtain a competitive advantage, a firm must establish and activate many resources and capabilities at the firm level. In this case, BD, by itself, is not sufficient to create the capabilities of BDA.

Since IT could directly and positively impact the competitive advantage, and its capabilities can be developed through employment and training (Shan et al., 2019); therefore, BDA capabilities can be a critical source of competitive advantage throughout all kinds of organisations. Hence, based on the current findings, it is proposed that:

H6: Competitive advantage positively influences financial performance.

3.3.8 Variables Moderates

Based on the nature of the business and capabilities, companies are classified into different categories. For example, sometimes, the firm's internal structure may restrict the development or upgrading of technological advancements. But, it is often observed that firms that produce goods with a short life cycle have great flexibility in adopting a new technological advantage (Fatorachian & Kazemi, 2018).

This study identified the control variables as firm size, firm age, and activity sector. Dunne et al. (1988, 1989); as cited in Audretsch (1995)). Firm size positively links with company survival. Information regarding a firm’s external environment, such as market opportunities, change in technology, and government policy, impact managers’ adoption of innovation as a strategy to better meet consumer needs and help make the company more competitive. Therefore, based on the current findings, it is stated that:

H11: Firm size moderates the relationship between competitive advantage and firm financial performance.

H12: Firm age moderates the relationship between competitive advantage and firm financial performance.

H13: Industry type moderates the relationship between competitive advantage and firm financial performance.

3.3.9 Environment Turbulence

Companies facing environmental turbulence must to renew themselves (Dierickx & Cool, 1989). The business environment discusses managerial subjects in the economic, political, social, technology, and competitive environments of business (Suikki et al., 2006), and environmental turbulence explains the condition of uncertainty or unpredictability as a result of changes in consumer performance and technology (Pavlou & El Sawy, 2006). Business environment turbulence has been divided into market turbulence and technological turbulence (Eisenhardt & Martin, 2000). Environmental turbulence has often been recognised as significant and unexpected situations and challenges facing organisations that require new solutions (Cameron et al., 1987; Pavlou & El Sawy, 2010). In such turbulent environment conditions, flexibility and relentlessly renewing knowledge innovation helps the firms to cope with the external environment disorders. The flexibility assumed to be of significant importance relates to a firm’s capability to adapt to change and/or employ opportunities emerging from environmental changes and can be considered a firm's specific ability or a source (Dreyer & Grønhaug, 2004). Therefore, companies require BD approaches and capabilities, sufficient knowledge about external and internal environments to adapt, survive, and succeed over time (Rachinger et al., 2019; Santos et al., 2017). Scholars thus considered that management during the BD age became vital. For example, Grant (1996) stated that efficient and effective management of knowledge is required under conditions of environmental turbulence. In addition, Chong (2004) stated that a practical approach to improving organisational crisis preparation is strategic management, which is considered a critical contemporary component.

Furthermore, Candi et al. (2013) argued that innovation management has to consider the dynamics of the environment. An active business operating in an unstable and turbulent environment is an organisation that is more flexible, adaptable, and innovative. Innovations to be successful need a practical focus on the external environment, as companies often activate external environments’ influence on opportunities and constraints on innovation (Tsai & Yang, 2014).

3.3.9.1 Technological Turbulence

The extent of technological change and its unpredictability is referred to as technological turbulence, and the speed of technology change will make the technological knowledge of the company obsolete. Managers should be aware of whether technology in their sector is changing rapidly and whether these changes involve opportunities or threats (Fernández et al., 2010). When companies observe technological turbulence in the environment, they are expected to adapt to these external pressures by selecting more manageable project planning and other flexible project blueprints (Candi et al., 2013). Competition is a strong indicator of innovation adoption, and the higher the pressure of the competitors, the greater the potential of innovation

(To & Ngai, 2006). Competitive pressures are defined as external pressures that drive companies to make changes and adopt new technologies (Obal, 2017). In other words, the competitors' pressure is the amount of stress that a company feels from its competitors within a similar industry (To & Ngai, 2006).

3.3.9.2 Market Turbulence

Market turbulence concerns the degree of diversity in consumer preference and goods demand (Jaworski & Kohli, 1993), which speedily causes the obsolescence of contemporary market knowledge in a company. In turbulent environmental conditions, flexibly accommodating environmental change and relentlessly renewing knowledge bases is the best way to sustain competitive advantage. The pioneers in technology adopters have a chance to be a first-mover advantage in a specific industry. In a competitive market, competitors try to influence consumers using BD strategies. The use of BD innovation by pioneers puts pressure on other actors, which leads them to follow the leaders. Usually, in the face of fierce competition, the firm tries to use the latest technology to strengthen its competitive advantage (Zhu et al., 2006). The business environment has become more turbulent and ferocious, which force firms in the same industry to imitate their pioneers to achieve parity and avoid market pressures. Intense competition is likely to push firms to take advantage of and benefits from BDA. Therefore, environmental turbulence could affect firms looking to gain a competitive advantage in the market by adopting big data analytics as a new technology strategy to overtake their competitors' performance. Thus, the following hypothesis is proposed.

H10: Environmental turbulence moderate the relationship between the competitive advantage and firm financial performance.

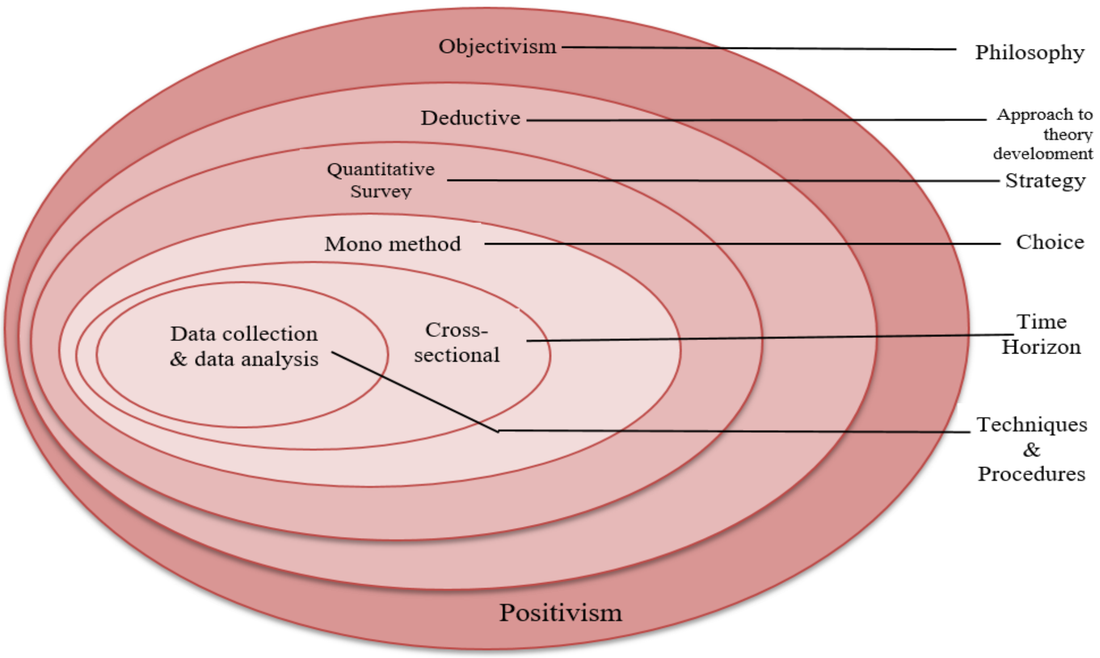
3.4 Chapter Summary

This chapter has developed a conceptual model to examine the impact of two fundamental components (effective use of BDA and BD management) on the innovation process, competitive advantage, and financial performance. In addition, it indicated theories RBV, DCV that support the current research and offered 13 hypotheses based on the conceptual framework to achieve the research's objective and answer the questions raised

Chapter Four: Research Methodology

4.1 Introduction

The current chapter discusses the most suitable research method to evaluate the conceptual model and related hypotheses developed in the former chapters to address the research questions. This chapter outlines the current research's philosophical stance and methodological applications. First, the chapter investigates various assumptions of the research philosophy and approach that will lead to the research methodologies and strategies. Second, the chapter discusses the survey methods, sampling techniques, sample size, questionnaire design, data collection process, and validity and reliability issues. Finally, the last section of the chapter covers the statistical data analysis and techniques that have been used to analyse the data for the current research. The research framework and methodology will be discussed considering the research process onion illustrated in Figure 4.1.



4.2. Research Philosophy

Figure 4.1 Research Process Onion
Source: Adopted from (Saunders, 2007, p. 130)

According to Collis (2003), research philosophy is "*the progress of scientific practice based on people's philosophies and assumptions about the world and the nature of knowledge*" (p. 46). Likewise, Saunders et al. (2007) referred to research philosophy as "*a system of beliefs and assumptions about the development of knowledge*"(p. 130). It is also known as an outline that guides how an investigation should be conducted based on the nature and evolution of knowledge (Bryman, 2016; Collis & Hussey, 2013; Saunders et al., 2007). Hence, the research philosophy aims to address questions such as:

- What distinguishes between science and nonscience?
- Scientists should employ what methods?
- How can we understand that a description is scientifically correct? (Newton-Smith & Newton-Smith, 2000).

To answer such questions, the researcher should consider several important assumptions that influence a researcher's stance (Saunders, 2009). This has guided the development of different research paradigms, a collection of logically related assumptions, concepts, or propositions that determine the action and view of the researcher about the world and the nature of knowledge (Creswell, 2009). Ontology, Epistemology, and Axiology are three fundamental assumptions of a paradigm. Ontological assumption wants to question the nature of reality, whereas epistemological assumption tries to understand what counts as knowledge, and axiological assumption questions the role of values in research (Bryman, 2016; Collis & Hussey, 2013; Saunders et al., 2007).

4.2.1 Ontology Approach

Ontology is the study of being (e.g., What is the nature of reality?), and it is concerned with what can be known about the world in terms of whether the realities of the social world are perceived as objective or subjective (Burrell & Morgan, 2017; Saunders, 2007, p. 108). Hence, it can be concluded that objectivism and subjectivism are two aspects of ontology. Objectivism confirms that social phenomena have a separate existence from social actors, which means that the researcher will not be influenced by the views or beliefs of the respondents. In contrast, subjectivism, or "understanding the meanings that individuals attach to social phenomena" (Saunders et al., 2007, p. 108), turns around the opinion that social phenomena are generated from participants' opinions and consequent actions (Orlikowski & Baroudi, 1991; Saunders, 2009). It means that the participants' views can create or at least adjust the researcher's beliefs. Thus, the subjectivist deems reality as the result of social interaction between the respondents. Table 4.1 summarises the fundamental differences between subjective and objective nature of research.

Table 4.1
Differences between Subjective and Objective Nature of Research

	Subjectivity	Objectivity
View of Human behaviour	Situational, dynamic, social, and personal	Regular and anticipated
Most common research objective	Discover, explore, and construct	Describe, define, and predict
Focus	Wide-angle lens; studies the dimension of phenomena	Narrow-angle lens; examines a particular hypothesis
Nature of reality	Multiple realities; subjective	Single reality; objective
View of social entities	Social phenomena are social activities of social actors	Social phenomena have a separate existence from social actors.
Research approach	Discuss interpretations of the subjective world	Apply objective techniques to gain the truth
Researcher role	Researcher and their preferences may be known to the respondents in the research, and respondents' characteristics may be known to the researcher	The researcher and their biases are unknown to the respondents in the research, and the characteristics of the respondents are intentionally hidden from the researcher. (double-blind studies)
Results	Specific results that are limiting generalisation conclusions	Generalisable conclusions, which are applicable to other populations

Sources: Adopted from Orlikowski and Baroudi (1991); Sarantakos (2012); Saunders (2015b).

In the current study, the objective approach appears suitable to answer the research questions and obtain the research objectives because this research aims to examine BDA hypotheses and a conceptual model. The objective viewpoint in this research investigates the impact of effective data analysis and effective data management on the innovation process, competitive advantage, and firm performance.

4.2.2 Axiology Approach

Axiology is interested in the role of values and ethics in research and the researcher's stance (Saunders, 2009; Wahyuni, 2012). According to scholars (e.g., Bell and Bryman (2007); Collis and Hussey (2009)) , axiology has two different perspectives: value-free or value-laden involvement, and the researcher should adopt one of them. The value-free viewpoint implies that the researcher's values do not affect the research process and interpreting the results (Saunders, 2009). In contrast, the value-involvement view reflects the researcher's values within the research process and interpretation of the findings (Collis & Hussey, 2009). However, Heron (1996) argued that values guide human actions; thus, their values may influence researchers' work. According to Saunders (2009), researchers should thus consider such value related issues throughout the research process. In the current study, data will be collected using the online questionnaire, which involves minimum interaction with the respondents and so entails less bias. Consequently, the chance of the researcher’s values playing a role in this phase is relatively minimised. Therefore, the research is considered value-free, as the researcher is supposed to be independent and should be able to analyse the phenomena without being biased by it or influencing it.

4.2.3 Epistemology Approach

Epistemologists are interested in the nature and sorts of knowledge and considers two critical issues: what knowledge is and how to achieve valid knowledge (Hirschheim, 1985). According to Saunders (2015a) there are three kinds of epistemology: positivism, interpretivism and critical realism. Positivism claims that the social world reflects the natural world (Collis & Hussey, 2009). The reality is separate from the individual who observes, which can be explained through an objective perspective (Levin, 1991; Süzen & Mamur, 2014). The positivists' perfect research method includes field experiments, laboratory experiments, and surveys (Bhattacherjee, 2012; Weber, 2004).

Interpretivists argue that the social world differs from the natural world because they are creating meaning, and their social world cannot be investigated in the same way as natural phenomena (Saunders, 2015b). On this basis, interpretivists have stated that social realities need subjective interpretation (Bell & Bryman, 2007). Thus, they are trying to keep an independent and objective position (Collis & Hussey, 2009; Weber, 2004). Scholars such as Hudson and Ozanne (1988); Orlikowski and Baroudi (1991) highlighted some criticisms regarding this approach, including bias that falls within researchers' interpretations and negligence in interpreting historical changes. They also addressed the inability to generalise the results because the conclusions are contaminated with personal values and perceptions, which also weakens the reliability and representativeness of the data to some extent. Furthermore, this perspective cannot consider the circumstances in which the respondents' calculations about actions and intentions are inconsistent with their real behaviour (Orlikowski & Baroudi, 1991). Therefore, there is a difference in terms of research in social science and natural science (Saunders, 2016). Interpretivism focuses on creating new and rich insights into social phenomena with a focus on richness and complexity; therefore, according to the above explanations and comparison with the positivism approach, it seems unsuitable for this research.

The third epistemological paradigm is critical realism, which differs from positivism and interpretivism, and it sees reality as the most important philosophical consideration (Bhattacherjee, 2012). This paradigm observes reality as independent and external, which cannot be accessed through knowledge and observation (Saunders, 2015a). According to Saunders (2015a), critical realism requires two essential stages of understanding the world: (a) the feeling and events that we are going through and (b) the mental process to think about the experience to figure out the cause. In critical realism research, it is recommended that the researcher should be attentive to how social and cultural background may affect the research and wipe out such biases and advance research objectively as much as possible (Saunders, 2015a). The critical realism approach is well known for focusing on economic factors with standard assumptions (Orlikowski & Baroudi, 1991). This approach demands a historical economic account to make sense of the outcomes generated; applying this approach would take much longer than other choices. However, this is essential for doctoral study as completion is limited.

The current research mainly focuses on the effective use of BDA tools and BD management aspects to create an innovation process, competitive advantage, and performance. Hence, an objective stance would justify the relationship between these factors, so adopting the interpretivism approach does not seem appropriate for this research (Saunders, 2016). As previously discussed, the positive approach tests theory to improve an adaptive understanding of social phenomena (Straub et al., 2004). It is concerned with validity, rigor, and the possibility of repetition (Orlikowski & Baroudi, 1991). Also, the positivist approach considers authenticity, including what objectively exists, and estimates that numbers and figures can measure knowledge (Hughes & Sharrock, 2016). Considering that BDA is an emerging field, and further research would be required to create a holistic understanding, positivist research could contribute to the advancement of this field by establishing a foundation for replicability and comparability. Hence, considering the above discussion, the positivist approach will be an appropriate choice to lead this research. Table 4.2 illustrates common types of research paradigms and distinguishing characteristics.

Table 4.2
Summary of Common Types of Research Paradigms and Distinguishing Characteristics

Issue	Positivism	Interpretivism	Pragmatism
Ontology	Naive realism Reality is real and apprehensible. objective	Relativism Multiple local and specific subjective	Reality is constantly renegotiated, debated, interpreted out of the social situation; objective and exists independently of the human mind
Epistemology	Dualist/objectivist Researcher is dependent of subject objective knowledge; Subject, object distinction gap between research and subject	Subjective point of view. (Knower and known are inseparable).	Researchers are trying to find the means and change is the fundamental purpose
Axiology	Value-freedom Facts must be separated from values	Inquiry is value bound	Values play a large role in interpreting results
Methodological	Deductive approach; Strict procedures	Hermeneutical/dialectical. impossible to differentiate fully causes and effects. inductive reasoning. Through the qualitative method, time and context-free, generalisation is rather desirable not possible	Thoughtful/dialectical eclecticism and pluralism of methods and perspectives. determine what works and solves individual and social problems
Methods	Quantitative which includes sampling measurement and scaling, statistical analysis, questionnaire	Qualitative, which includes interviews, observation, case study, life history, focus group, and narrative.	Quantitative+ Qualitative methods. In addition, data mining, expert review, usability testing, physical prototype, and like

Source: Adapted from Danaee Fard (2012); Saunders (2009, 2015b); Tashakkori et al. (1998)

4.3 Research Approach

The research approach concerns how theory supports the research process, or more specifically, it concerns the theoretical foundations of research (Saunders, 2009). Researchers adopted two main methodological approaches, deductive and inductive (Bhattacherjee, 2012). Deductive research is "*a study with a conceptual and theatrical structure is developed with is then tested by empirical observation; thus, particular instances are deductive from general influences*"(Collis & Hussey, 2013, p. 7). The deductive approach uses the questionnaire to generate an understanding of observations, which enables the researcher to analyse the different opinions of participants within experimental data. In contrast, the inductive approach is *"a study in which theory is developed from the observation of empirical reality; thus, general inferences are induced from particular instances"* (Collis, 2003, p. 15). In addition, Bhattacherjee (2012) and Bryman and Bell (2007) said that the inductive approach collects data to create a theory to discover phenomena. Therefore, the deductive method is more suitable with the positivist approach, whereas the induction method is affected by interpretive epistemology (Bhattacherjee, 2012). Based on this approach, the proposed theories can be scrutinised using statistical analysis methods (Collis & Hussey, 2009).

Table 4.3 summarises the key differences between deductive and inductive approaches. It illustrates that the deductive approach is utilised to enable facts to be quantified. Additionally, the deductive approach is distinguished by causality, generalisation, interest in measurement and replication (Bryman & Bell, 2007). According to Creswell (2003) the choice of a research approach depends on several criteria such as richness of literature on the research subject, nature of the audience, and time constraints. The deductive approach will be beneficial if a richness of literature supports developing a theoretical framework and hypotheses. Therefore, based on the suggested aspects of choosing a research approach, the current study goes appropriately with the deductive approach. Adopting this approach assists in explaining the causal relationships among constructs of the research, as a deductive approach involves the formation of hypotheses that are subjected to examination through the quantitative method, whereas the inductive approach is not concerned with testing hypotheses (Punch, 2013; Sarantakos, 2005). At first, this deductive research goes through a literature review of the prior studies and existing theories related to the research topic to support conceptualising a conceptual model to investigate the relationship between BDA, innovation process, competitive advantage, and firm performance. Then, the study develops the constructs that generate the research hypotheses and scales for measuring each research variable depending on existing literature on the investigated phenomena. Finally, data are gathered and analysed with a highly structured methodology, through which the proposed model and hypotheses are confirmed or modified.

Table 4.3
Differences Between Deductive and Inductive Approaches

Deduction Emphasises	Induction Emphasises
• Scientific principles • Moving from theory to data • The need to explain causal relationships between variables • The collection of quantitative data • The application of controls to ensure validity of data • The operationalisation of concepts to ensure clarity of definition • A highly structured approach • Researcher independence of what is being researched • The necessity to select samples of sufficient size to generalise conclusions	• Gaining an understanding of the meanings humans attach to events • A close understanding of the research context • A more flexible structure to permit changes of research emphasis as the research progresses • A realisation that the researcher is part of the research process • Less concerned with the need to generalise

Source: Saunders (2009, p. 127)

4.4 Research Methodology

Research methodology is a research strategy that interprets the research paradigm into guiding principles that assist the researcher in carrying the research (Sarantakos, 2005, p. 30). It is an action strategy that connects methods to results, controls research selection, and determines the proper research methods (Creswell, 2003, p. 5). Two main research methodologies in terms of the nature of data used in research are quantitative and qualitative (Venkatesh et al., 2013). Quantitative research typically utilises an objective ontological position, positivist epistemological orientation, and a deductive approach. In contrast, the qualitative research method is more adapted for a subjective ontology position and an inductive approach (Bell et al., 2018). This method is considered as more situatable when research needs numbers/figures to draw research conclusions. Therefore, the quantitative research method is more appropriate for examining statements or hypotheses as in confirming investigations. Looking at data analysis, qualitative data analysis refers to the content analysis, while quantities data analysis refers to numerical statistics (Allwood, 2012). Table 4: illustrates the differences between quantitative and qualitative methodologies.

Table 4.4 explains that the quantitative methodology adopts the objective viewpoint, while the qualitative methodology is based on the subjective viewpoint (Neuman, 2006, p. 153). According to Bryman and Bell (2007), a quantitative methodology is value-free and unbiased by the researcher. In contrast, the qualitative methodology is characterised as value-laden and biased, and it gives the researcher a significant role in interpreting the research findings (Collis, 2003; Collis & Hussey, 2013; Creswell, 2014). Furthermore, a quantitative methodology is proper for the positivistic paradigm. In contrast, an interpretivist paradigm is mentioned to be more suitable for a qualitative methodology (Bryman & Bell, 2007, p. 28). Therefore, a quantitative methodology utilises the mathematical models and statistics for analysis, and its outcomes can be generalised and apply as a guide for prediction, explanation and understanding (Bryman & Bell, 2007). According to Sarantakos (2005, p. 134), a methodology's selection is subjected to several factors such as research paradigm, compatibility with the theoretical research aims, research objects, general goal of the study, nature of anticipated results, and a suitable research approach.

Table 4.4
Differences Between Quantitative and Qualitative Methodologies

Issue	Quantitative	Qualitative
Ontological orientation	Objectivism	Constructionism
Epistemological orientation	Natural science model, in particular positivism	Interpretivism
Axiology	Value-free and unbiased	Value-laden and biased
Aim	Quantitative/numerical description, causal explanation, and prediction. Offer generalizable findings providing a representation of objective outsider viewpoint of populations	Qualitative/subjective description, empathetic understanding, and Exploration; offer particularistic findings; provision of insider viewpoints
Principal orientation to the role of theory in relation to research	Deductive; testing of theory	Inductive; generation of theory
Research Design	Independent of the knower	Not independent of the knower
Data	Questionnaires, surveys, and systematic measurement involving numbers	Observation, in-depth interviews, document analysis, and focus groups
Data Analysis	Mathematical models and statistics analysis	Textual, sometimes graphical, or pictorial form

Source: Adapted from Bryman and Bell (2007); Castellan (2010); Collis and Hussey (2013); Creswell (2014); Guba and Lincoln (1994)

Since this research adopts the positivistic stance through depending on theory testing, a quantitative methodology will be applied to obtain the study objectives. The literature review helps develop a conceptual research model to test the influence of effective use of BDA tools and BD management on the innovation process, competitive advantage, and financial performance. Adopting the quantitative approach will assist in establishing the impact factors of BDA that lead to influence the innovation process, competitive advantage, and financial performance. In addition, several studies have used a quantitative approach in BDA research (e.g., (Akter et al., 2016; Ghasemaghaei & Calic, 2020; Mikalef et al., 2019a; Mikalef, Boura, et al., 2020; Mikalef & Krogstie, 2020; Shan et al., 2019; Wamba et al., 2017)). Further, this approach will guarantee the replicability of this study, as the research will be unbiased and free of the researcher's beliefs, values, or any other features. Also, due to the measurable nature of this approach, the reliability and validity of the research can be confirmed. Hence, following the previous studies and considering inherent advantages, this study adopts a quantitative approach as a more appropriate method through which outcomes can be generalised to the broader population.

4.5 Research Strategy

Research strategy is *"a general plan of how the researcher is going to answer the research questions"* (Saunders, 2007, p. 131). It provides powerful guidance to support the researcher in organising, implementing, and governing the research. Many strategies can be adopted in research, and no strategy is preferred and be better than others (Saunders, 2009). However, the research strategy is chosen to fit with the study's assumptions, objectives, and characteristics and how it empowers the researcher to address the research questions and obtain the research goals (Collis & Hussey, 2013; Saunders, 2007). Research strategies may be appropriate either to the deductive approach, the inductive approach, or both inductive and deductive. The following Table 4.5 summarises some of these strategies.

Table 4.5
The Different Research Strategies

Strategy	Definition	Advantage	Disadvantage
Experimental	*"An experiment is a study involving intervention by the researcher beyond the required for measurement. The usual intervention is to manipulate some variable in a setting and observe how it affects the participants or subjects being studied"* (Cooper et al., 2006, p. 302).	• The ability to manipulate the independent variable and thus measure the change in the dependent variable. • The ability to use control group. • It can be easily replicated.	• The participants' perceptions may be affected by surrounding environment. • There are problems regarding generalizing the results of this strategy.

Strategy	Definition	Advantage	Disadvantage
Case Studies	*"An empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident"* (Yin, 2009, p. 13).	• One of the effective strategies that can help in gaining depth understanding of the context. • An effective strategy to find answers about most research questions.	• It is often used in exploratory and explanatory research. • There is a difficulty of generalizing the research findings
Survey	*"research technique in which information is gathered from a sample of people by use of a questionnaire; a method of data collection based on communication with a representative sample of individuals"* (Zikmund, 2003, p. 175).	• The ability to administer; quick; inexpensive; easy to code, analysis and interpret; and provide an accurate means of assessing information about the target population. • The ability to collect a large sample of data. • The ability to reduce the variability in the findings due to using fixed questions.	• Individuals may refuse to participate; tend to give false answers and tend to answer in a certain direction (response bias). • Question's structure, forming and wording may affect respondents' answers. • The problem of systematic error.
Action theory	*"Action research is a design that simultaneously combines action to bring about change in a setting and research to increase and/or develop understanding on the part of the researcher, client group, etc. about that social system to develop knowledge"* (Tharenou et al., 2007, p. 89).	• Action research focuses on learning and bringing about change in a social system. • It provides deeper understanding of system processes is sought, as well as comprehension of what was not understood before, to contribute to knowledge.	• Requires close cooperation between the researcher and the objects. • This strategy is time consuming in building relationship with researched objects. • The difficulty of generalizing the research findings.

Strategy	Definition	Advantage	Disadvantage
Ethnography	*"It is a distinct form of qualitative data collection that seeks to understand how social and cultural influences affect people's behaviour and experiences"*(Hair et al., 2007, p. 184).	• It provides rich information about human, organisation, social, and aspects. • One can use audio and visual recording. • The questioning and observation method are combined to understand the behaviour of respondents.	• This strategy is time consuming in collecting data, interpreting, and reporting findings. • Besides the difficulty of selecting the research object(s), it requires building a high degree of trust between the researcher and the researched.

The current research applies the survey method as a strategy. In management and business studies, the survey is practiced broadly as a strategy for collecting and analysing quantitative data (Collis & Hussey, 2013; Sekaran & Bougie, 2016). Studies based on the surveys are apportionable on exploratory scopes and predictive theory (Straub et al., 2004). In addition, the survey method is considered economical in gathering a large amount of data in a short period, which can be easily standardised and compared. Surveys operating standardised questions allow the data to be easily collected and analysed using quantitative methods, and when used with an appropriate sampling technique, the research outcomes can be generalised to a more significant population (Neuman, 2007, p. 166). Therefore, the current study used the survey method as a strategy because this strategy is proper to identify the impact of BDA on the innovation process, competitive advantage, and performance.

4.6 Time Horizon

The time element is one of the most critical issues in drawing up the research plan, which must be considered. In this respect, scholars discussed two different research plans: cross-sectional or from the same participants at a different time, (i.e., longitudinal; (Bhattacherjee, 2012)). A cross-sectional study is designed to gather data about a phenomenon, situation, or problem in various contexts at a specific point in time (Sekaran & Bougie, 2016). Also, a cross-sectional approach is used to collect data on more than one case and gather a body of quantitative or quantifiable data connected with two or more variables tested to detect relationship patterns (Bryman & Bell, 2007). Several scholars (e.g., Anderson and Narus (1990); Ganesan (1994); Merlo et al. (2006); Ulaga and Eggert (2001) suggested that cross-sectional studies are the most used in social studies efforts to generalise research findings.

In contrast, a longitudinal approach studies a phenomenon or people at more than one single time to address the research question. An example of a longitudinal study is tracing the sales patterns of a particular good in several separate areas in every quarter for the next 2 years, which requires data collecting for several seasons at different times (Sekaran & Bougie, 2016). The nature of longitudinal research is time-consuming, costly, and vulnerable to intervening events and participant attrition; therefore, because most cross-sectional questionnaires are filled out by one participant, this kind of research is believed to be precise and less prone to common methods bias CMB and causal inference (Rindfleisch et al., 2008). The current study is cross-sectional. Given that the study embraces a correlational survey that investigates the effects of effective use of data analytics tools and BD management on the innovation process, competitive advantage, and financial performance at a single point in time, the cross-sectional approach is suitable for the study. Furthermore, thesis research is always accompanied by many limitations. Therefore, to avoid the limitations such as time frame, associated cost, and viability of analysing the massive amount of data, the cross-sectional approach is most proper for the study.

4.7 Data Collection Methods

Undoubtedly, the suitable selection of data collection methods is essential to achieving the research aims (Collis, 2003). The data collection method refers to a systematic process of information collecting and measuring targeted variables (Ghauri et al., 2010). Different scholars (e.g., Axinn and Pearce (2006), Hair (2009), Sekaran and Bougie (2016), highlighted several methods of data collecting includes interviews, surveys, observations, focus groups, and archives. However, the most common quantitative research method is the survey (Easterby-Smith & Prieto, 2008; Ghauri et al., 2010). Table 4.6: summarises the advantage and disadvantages of some of the most important data collection methods.

Table 4.6
Advantage and Disadvantage of Different Data Collection Methods

Data collection method	Advantage	Disadvantage
Personal interviews	• Can establish rapport and motivate respondents. • Can clarify the questions, clear doubts, add new questions. • Can read nonverbal cues. • Can use virtual aids to clarify points. • Can obtain rich data	• Takes personal time. • It costs more when a wide geographic region is covered. • Respondents may be concerned about the confidentiality of information given. Interviewers need to be trained. • Can introduce interviewer bias.

Data collection method	Advantage	Disadvantage
Telephone interviews	• Less costly and speedier than personal interviews. • Can reach a wide geographic area. • Greater anonymity than personal interviews.	• Nonverbal cues cannot be read. • Interviews will have to be kept short. • Obsolete telephone numbers could be contacted, and unlisted ones omitted from the sample.
Personally, administrated questionnaires	• Can establish reports and motivate sample units. • Doubts can be clarified. • Less expensive when administered to groups of respondents. • Almost 100% response rate ensured. • The anonymity of respondents is high.	• Organisations maybe unwilling to give up company time for the survey using groups of participants assembled for the purpose.
Mail questionnaires	• Anonymity is high. • A wide geographic region can be reached. • Token gifts can be enclosed to seek compliance. • Sample units can be allowed to have time to complete the questionnaire at convenience and can be used with web surveys.	• The response rate is almost always low. • It cannot clarify questions. Follow-up procedures for nonresponse are necessary.
Electronic questionnaires	• Can reach globally. • Very inexpensive. • Fast delivery. • People can respond at their convenience.	• Computer literacy is a must. • Access must be available • Respondents must be willing to complete the questionnaire.

Source: Adopted from Hair (2009); Montash et al. (2014); Sekaran and Bougie (2016)

The selection of data collection methods depends on the researcher's available sources and how the method can best create the requested information (Peterson, 2000). For example, the personal interviews method cannot be used in the current study because it is costly and time-consuming, and the target population is distributed in different regions. In addition, the telephone interviews method was a challenge due to the unavailability of the direct phone numbers of the target participants. Since the target respondents of this study were top managers of firms, it was not easy to reach them through the phone. Table 4.6 compares the pros and cons of data collection methods; it turned out that the best way to collect data was the online survey method. The online survey, a specific form of an email survey, has been broadly used and has gained popularity since the early 1990s. The widespread use of the internet and communication technologies has massively facilitated the extensive use of web-based surveys. Besides considering associated cost, time limitation, and geographical issues, the online survey method was considered suitable for the context of this study and ideally suited the participants' personalities. Furthermore, all respondents are professional internet users, well-educated, and had access to the survey. Thus, the weaknesses of the web surveys did not influence the data gathering in this study. Hence, such an adopted method is well compatible with the context of this research.

4.7.1 Survey Method

The survey method is typically the most popular tactic applied in the quantitative study (Easterby-Smith & Prieto, 2008; Ghauri et al., 2010). In general, researchers have deemed surveys as reliable and similarly easy to explain and understand (Saunders, 2015b). Furthermore, it uses standardised questions that allow the data to be collected easily to analyse with quantitative methods (Neuman, 2007, p. 166). In contrast to other data collection methods, surveying is considered a more interactive procedure that allows the participants to be aware of the context in which they are being questioned. It can be completed by respondents directly or by the interviewer (Brace, 2018). In the self-administered method, the questionnaire is sent by mail, internet, fax, or drop off/pick up, whereas the interviewer method requires face-to-face or telephone contact with participants (Sekaran & Bougie, 2016). As discussed in Table 4.5, all methods have advantages and disadvantages (Bryman & Bell, 2007).

The appropriate selection of data collection methods depends on the available resources and how best the selected method can generate the required data (Peterson, 2000). Personal interviews were not used because of the geographical distribution because the target population was in the American and European countries and limitations of time and money. Although the FAME database provides the telephone numbers of companies located in the UK, it does not cover all the target population. Also, because the target population of this survey are firms' managers, reaching them would have been difficult, if not impossible. Telephone interviews were thus also ruled out. Considering the abovementioned, the most proper data collection method for this research was the online questionnaire, where it became straightforward to send the survey link to participants through email.

4.7.2 Survey Distribution

There are several ways to distribute questionnaires. The questionnaire can be distributed by post or electronically distributed through the internet, email, and social media (Bhattacherjee, 2012; Fowler Jr & Cosenza, 2009; McLafferty, 2003). Today, the electronic method is more common and preferred as it allows a fast and wide distribution of the survey at a low cost compared with other methods (Bradley, 2007; Collis & Hussey, 2009). According to Deutskens et al. (2004), the participants' rate is expected to improve the online questionnaire's wide geographical distribution. Some scholars (e.g., Bhattacherjee (2012); Evans and Mathur (2005), have argued that relying on data collection only through an electronic method may exclude a group of people who do not have access to the internet and practice; the participants are limited to the internet user. Because the sample of the study includes business managers from different countries and the survey is about BDA and management, it was logical to assume the participants could take part in the survey through online platforms. In addition, using an online survey enables the researcher to transfer the data collected directly into the statistical software for analysis aids to decrease non-sampling error (Churchill & Iacobucci, 2006). Hence, considering the convenience of the participants and the benefits of the research, an online survey was used in this study.

Table 4.7
Data Collection Phases

Phases	Number	Source/participants
Questionnaire validation	6 samples	Professors/Managers
Pilot - study	21 samples	Top managers
First wave	78 samples	FAME database and other companies
Second wave	96 samples	Via Qualtrics
Totals	**(78 + 96) = 174 samples**	

4.7.3 Target Population

The research population can be defined as the whole group of cases or elements that share features or have common characteristics that a researcher could examine (Bhattacherjee, 2012; Sekaran & Bougie, 2016). Assessing the whole population is impossible because of its size and the lack of research resources; therefore, researchers choose a representative sample to assess the total population. Moreover, to get a suitable representative sample group, it is initially necessary to identify the research population. Like many BD and BDA studies (e.g., (Gupta & George, 2016; LaValle et al., 2011; Marshall et al., 2015; Niebel et al., 2019)) the sample for the current study was gathered from firm managers such as executive managers, senior staff, and other managerial positions. Therefore, this research focused on the firms' managers as participants who take managerial positions in different industries such as agriculture, automotive, energy, high tech, and communication, financial and insurance, health care, real estate, and service. However, any person can fill out an online survey (Kaye & Johnson, 1999), and respondent identification procedures are considered a breach of the respondent's privacy (Smith, 1997). This issue is one of the disadvantages of the electronic questionnaire, which can lead to incorrect responses and skewed results. Hence, following scholars' suggestions, the current study applied outliers' assessment, normality test and multivariate assessment (Mahalanolobi's distance method) to minimise online questionnaire's weaknesses. The sample frame refers to an accessible part of the population that (Bhattacherjee, 2012). The principal purpose of choosing a sample is to ability the generalise results to the entire population (Mellinger & Hanson, 2016). The selected sample must be large enough to represent the entire study population to ensure that the data collected is likely to provide reasoned assumptions and convincing conclusions (Aggarwal, 2011). The firms were chosen from developed countries because they are supposed to be more advanced and their managers familiar with BDA technology; therefore, the current study focused on firms located in the UK and the United States. A total of 55 samples were collected from America and 98 from European countries. Noteworthy, 57 samples out of the 98 related to the UK and Ireland went through the FAME database.

4.7.4 Sampling Technique

The next step after identifying the sample frame is to determine the sampling technique. To guarantee similar characteristics of the whole population, choosing a proper sampling technique is crucial for research. According to Bhattacherjee (2012), it is essential to pick a sample so that it represents the population, and the results of the study can be generalised to that population. Scholars categories sampling techniques into two broad groups: probability and non-probability (Bhattacherjee, 2012; Mellinger & Hanson, 2016). Probability sampling or random sampling is described as *"a sampling which permits every single item from the universe to have an equal chance of presence in the sample"* (Etikan & Bala, 2017, p. 2). In comparison, non-probability sampling refers to *"a sampling procedure that will not bid a basis for any opinion of probability that elements in the universe will have a chance to be included in the study sample"* (Etikan & Bala, 2017, p. 2). The main difference between the two sampling methods is that in probability sampling, each member of the population has an equal chance of selection, whereas in non-probability sampling, each member of the population's likelihood of choice is unknown (Bhattacherjee, 2012; Zikmund et al., 2013). Daniel (2011) pointed out that non-random sampling is cost-effective, time-saving, and useful when the population is unlimited; however, it is likely biased toward the sample, more limited generalizability, and the inability to represent the population and make statistical inferences. Instead, random (probability) sampling does not have the limitations of nonprobability sampling. In addition, it relies on the fundamental theory of normal distributions, which offers more objective findings that display the features of the population (Creswell, 2014). Table 4.8

summarises the main categories and characteristics of probability and non-probability techniques.

Table 4.8
Sampling Technique Categories

Different types of Probability (random) Sampling:
• Simple random sampling: every member of the population has an equal probability of being elected. • Systematic sampling: members are selected regular from equal intervals of an ordered list. • Stratified sampling: the population is divided into categorised according to homogeneous and non-overlapping (strata), then random choices are made from per category. • Cluster sampling: the target population is divided into clusters, some clusters are chosen randomly, and then the selected clusters are measured. • Matched-pairs sampling: two sub-groups from one population are compared based on a particular criterion • Multi-stage sampling: combine single-stage methods to manage multi-stage sampling
Different types of Non-probability Sampling:
• Convenience sampling: also called (opportunity or accidental sampling) the sample is taken from a part of people, who reachable, readily available, or appropriate. • Quota sampling: easy access to the population sample using the quota sample (sex, race). • Expert sampling: participants are selected in a non-random manner based on their specialisation in the subject being studied. • Snowball sampling: contact with a small number of individuals will lead to other groups.

Source: Adapted from: Bhattacherjee (2012); Cochran (2007); Collis and Hussey (2013); Etikan and Bala (2017); Taherdoost (2016)

As mentioned in Table 4.8, scholars (e.g., (Collis & Hussey, 2013; Ghauri et al., 2010; Saunders, 2015b)), highlighted six probability sampling techniques, including simple random sampling, systematic sampling, stratified sampling, cluster sampling or multi-stage cluster sampling that can be conducted to collect data. Alternatively, non-probability sampling can be conducted via four techniques: convenience sampling (opportunity or accidental sampling), snowball, quota, expert sampling (Bhattacherjee, 2012; Bryman, 2016; Collis & Hussey, 2013). The current study used the random sampling technique to collect the data from the targeted population. The aimed population is firm's managers that located in European countries and United States.

4.7.5 Sample Size

According to Lenth (2001), determining the sample size is considered one of the researcher's most critical decisions. Indeed, selecting the proper sample size assists in obtaining more generalisable results from the sample to the targeted population and empowers the researcher to apply the suitable statistical analysis method, which promises more reliable and valid outcomes (Hair et al., 2006). Conversely, failure to select the appropriate sample may affect the credibility of the study results (Westland, 2010). Scholars such as Kline (2005) and Hair et al. (2010) stated that the choice of sample size depends mainly on the statistical analysis method. The more complex the statistical method, the larger the sample size required.

The study selected the structural equation modelling (SEM) technique for the statistical analysis to test the proposed hypotheses. Hence, it is necessary to determine the sample size corresponding with this technique. According to Kline (2011), around 200 cases are suitable for adapting the SEM method. In the same context, other scholars (e.g., (MacCallum et al., 1996)) stated that the determination of sample size should be based upon various factors such as the number of parameters, model complexity, the estimation technique applied, degrees of freedom, statistical power, and/or the volume of missing data. For instance, Harrington (2009) proposed that "less than 100" is a small sample but acceptable for straightforward models, whereas samples from 100–200 would be appropriate for the models that are not very complex, and over 200 will fit most models. Reviewing some of the studies of BDA that used the quantitative method, it was clear that the size of the samples varied from one study to the next, and the samples size range from 175 (Mikalef et al., 2019a), 232 (Gupta & George, 2016), 500 (Côrte-Real et al., 2017). However, some scholars (Chin, 1998; Westland, 2010) pointed out an ad hoc rule that each indicator needs at least 10 observations for sufficiency sample sizes. Since the current study's conceptual model has five indicators, therefore; the research samples should be at least 50. This study's actual samples size is 143, approximately 24 samples for each indicator, which seems adequate for the current study.

4.8 Questionnaire Design

The questionnaire is one of the principal tools for obtaining primary data (Collis, 2003). Therefore, arranging and presenting the questions correctly are essential to prevent confusion and bias or suspicions about the respondents' research (Bradley, 2013). In addition, questionnaire design influences the reliability, validity, and response rate (Saunders, 2015b). Therefore, to maintain the quality of the questionnaire, several issues should be considered. First, the questionnaire should not contain double-barrelled questions or ambiguity and should be well organised (Brace, 2018). Second, relevance (avoiding unnecessary information) and accuracy (reliable and valid information) should be considered in the questionnaire design.

(Zikmund et al., 2013). Third, it should be simple, straightforward, reliable and valid (Neuman, 2006). Finally, it should provide appropriate response scales to ensure the participants of understand their options (Saris & Gallhofer, 2007). Considering such issues, the questionnaire for the current study was designed in three sections. The first part of the questionnaire focused on opening questions concerning age, gender, and education levels, which were easy and quick to answer. Following a logical and natural order, the second section accommodated the questions measuring the constructs of hypothesised model drawn in Chapter three (Acharya, 2010). The last part of the questionnaire was devoted to an open question to receive the opinions and recommendations of the participants. Items measuring a specific construct were grouped together. Careful writing of questions is as important as deciding on their order and style (Bryman & Bell, 2007); therefore, questions were made simple, clear, and understandable. Following the scholars' recommendations, such as Acharya (2010); Bhattacherjee (2012); Bryman and Bell (2007), the phrases were made appropriate for the respondents by avoiding questions that participants may not genuinely answer or feel offended to. Also, biased words and double-barrelled and double-negative questions were carefully removed from the questionnaire. The final version of the questionnaire can be seen in detail in Appendix 4.1. After the questionnaire was formulated, it was necessary to conduct further tests before the final survey. Therefore, compliance was tested for the guidelines through pre-test and pilot tests and based on the results, some necessary adjustments were made (Neuman, 2006).

4.8.1 Pre-Testing the Questionnaire

Pre-testing is used to ensure the accuracy, validity, and reliability of questions (Johnson & Rapp, 2010). The researcher needs to ensure that the aimed participants understand the questions and proposed response options and complete the survey accurately (Perneger et al., 2015). Therefore, validation of the questionnaire is necessary after its development and before its distribution to collect final data (Saunders, 2015b). This study adopted two steps to verify the questionnaire validity: expert evaluation and pilot testing (Acharya, 2010). First, as advised by Saunders (2015b), a questionnaire draft was sent to three top managers and three academics for a critical review. The academics conducting the initial questionnaire had extensive experience in questionnaire design and improvement. However, experts were asked to check several points, such as the validity of the content, whether the questions measure what was intended to measure, and ascertaining the questionnaire structure, such as language and layout (Hair et al., 2010; Saunders, 2016). Based on the comments and suggestions the reviewers, some modifications were made, such as improving clarity, simplifying vagueness of the words, and removing duplications. Some questions were also reconsidered, as they were reduced from 78 to 64 questions. Finally, the modified questionnaire was sent for pilot testing.

4.8.2 Pilot Testing the Questionnaire

According to Boudreau et al. (2001), a pilot study is a preliminary assessment of probable problems and finding out whether the questionnaire may provide the data necessary to achieve the research objectives. In addition, the validity and reliability of the questionnaire can also be assessed in this stage of the questionnaire development (Brace, 2018). Therefore, scholars (e.g., Saunders (2015b) and Malhotra et al. (2006)) stated that conducting a pilot test may increase the research likelihood of success. The measurement items for this research were developed by adopting validated measures in previous studies and modifying them to fit the current research context. Following different scholars' recommendations, such as, Saunders (2015b) and Malhotra et al. (2006) the pilot study of this research was conducted by sending the questionnaire to 30 managers, 21 of whom completed the questionnaire. The pilot test revealed that the participants could understand the items appropriately and complete the questionnaire within an average of 20 minutes. However, some minor issues were identified in a few of the items, and slight modifications were made, which reduced the length of the questionnaire to 12-15 instead of 15-20 minutes of completion time.

4.8.3 Improving Response Rates

Response rate varies exceptionally from one survey to another (Ilieva et al., 2002). To increase the response rate, several strategies were used. First, a covering letter describing the goal of the study was created (see Appendix 4.1). This included the information of the author and his first supervisor and their affiliation. Furthermore, the assurance of confidentiality, privacy, and anonymity of the participants was provided (Bryman & Bell, 2007). Second, follow-up emails were sent after one week, and with all the strategies taken to increase participation, the results were not encouraging. One of the reasons for such a low response rate could be that companies' online security system does not allow such emails to reach the inbox of the targeted participant. However, some of the participants responded with an apology email stating that they were bound by the company regulations, preventing them from participating in such a kind of survey. Since the response rate was inadequate, the study has used several extra steps to increase the response rate. First, the researcher sent a second reminder email. Second, the researcher changed the strategy to make the issue more personalised such as using personal email addresses of directors, managers, vice presidents, and senior staff to send invitations to participate in the survey. In this way, the researcher used the managers' names -

to make the emails to be more personalised, instead of “Dear Participant” in the cover letters. Thus, the researcher searched companies’ websites and sent emails directly to managers instead of general emails of companies such as info, service, office, sale, and admin. However, just 84 samples were collected by using these methods, which was insufficient to conduct data analysis techniques. Therefore, the researcher discussed this issue with supervisors, and based on their suggestions, the study used a third-party agency to collect the data to complete the research. In this regard, some data collection agencies were checked through the information available on their websites and sent emails to ensure their work quality and reliability. In the end, the choice fell on Qualtrics as it has a partnership with Swansea University. Overall, due to the researcher’s effort and Qualtrics agency, 174 samples were collected.

4.8.4 Scale Development

Identifying the constructs to be measured is the first phase in organising measurement scales. Following this, it is essential to operationalise the constructs into assessable variables. According to Hinkin (1998), The utmost challenge in conducting social science research is proving the accuracy of the experimental measurements for each latent variable under investigation. Hence, researchers can accumulate knowledge and ensure comparability between studies using previously validated instruments (Boudreau et al., 2001). Thus, the study used a survey instrument drawing upon a comprehensive literature review, as illustrated in Table 4.8.

Effective use of BDA tools variable’s items was developed from three data analysis dimensions: effective data aggregation, effective use of data analysis tools, and effective data interpretation. The items related to effective use of data express the idea about data collection sources, ease of accessibility, and data gathering process, whereas effective use of data- related items reflects identifying business insights and trends, analysing data in real time, and translating outcomes into innovative ideas and improving products or services. Scales to measure effective data analysis constructs had previously been validated through two domains: data aggregation tools and effective use of data analysis tools by Hajli et al. (2020); Narver et al. (2004); Shirazi et al. (2021).

Scales of BD management were developed from two dimensions of technical and managerial skills. The items related to technical skills of management use of data express the idea about managers' ability to coordinate BD-related activities in ways that support the team of the innovation process and other functional managers, whereas managerial skills of data management related items reflect employment, training, and cooperation with other fictional departments, supplies, and customers. Scales to measure this construct were validated by Hajli et al. (2020); Narver et al. (2004); Shirazi et al. (2021).

According to Goldsmith and Hofacker (1991), researchers measure innovativeness to investigate the relationship between this construct and other variables; this research attempted to explore the influence of effective data analysis and effective data management on the innovation process. Analysis sources such as De Luca and Atuahene-Gima (2007); Jiménez-Jiménez and Sanz-Valle (2011); Narver et al. (2004) led to developing a scale to measure the innovation process. After checking their scales, the researcher believed that a combination of scales would be most suitable for measuring the innovation process in this research.

Competitive advantage was evaluated in terms of whether the innovation process, effective data analysis, and effective data management help create a competitive advantage for the firm by becoming unique in its products/services, reducing product/service cost, and sales growth. Scales to measure competitive advantage constructs had previously been validated by Côrte-Real et al. (2017); Shan et al. (2019); Zhou et al. (2009).

Firm performance was operationalised by measuring profitability, customer retention, sales growth, return on investment, overall financial performance (Jiménez-Jiménez & Sanz-Valle, 2011; Liu et al., 2013). These measures are representative of potential value that can be realised as a result of BDA Gupta and George (2016); Wamba et al. (2017).

Scales for environmental scanning were directly adopted from Hajli et al. (2020); Jiménez-Jiménez and Sanz-Valle (2011), describing how environmental turbulence can affect a firm in terms of change in the industry, technology, products/services, markets, and customer performance. The measures for effective data analysis, effective data management, innovation process, competitive advantage, and firm performance used in the current study are shown in Table 4.9.

Table 4.9
Measurement Items and Sources

Construct	No	Item	Source
Effective data analysis	16	Collect data from external sources and various customer relationship management systems.	(Hajli et al., 2020; Narver et al., 2004; Shirazi et al., 2021)
	17	Make customer records and transactions consistent, visible, and easily accessible for further analysis.	
	18	We integrate data from multiple internal sources into a data warehouse or mart for easy access.	
	19	Identify essential business insights and trends to improve the innovation process.	
	20	Predict product /service patents in response to customers' needs.	
	21	Analyse data in near real or real-time that allows responses to unexpected market threats.	
	22	Analyse social media data to understand current trends from a large population	
	23	Provide systematic and comprehensive reporting to help recognise available opportunities for product/service improvement.	
	24	Support data visualisation that enables users to interpret results easily.	
	25	Provide near real or real-time reporting for the innovation process.	
BD management	26	We provide BDA training to our employees	(Hajli et al., 2020; Narver et al., 2004; Shirazi et al., 2021)
	27	We hire new employees that already have BDA skills	
	28	Our BDA staff has the right skills to accomplish their jobs successfully	
	29	Our BDA staff has a suitable education to fulfil their jobs	
	30	Our BDA staff holds suitable work experience to accomplish their jobs successfully	
	31	Our BDA staff is well trained	
	32	Our BDA managers understand and appreciate the business needs of other functional managers, suppliers, and customers.	
	33	Our BDA managers can work with team of innovation process, functional managers, suppliers, and customers to determine opportunities that BD might bring to our business.	
	34	Our BDA managers can coordinate BD-related activities in ways that support the team of innovation process and other functional managers.	
	35	Our BDA managers can anticipate the future business needs of functional managers, suppliers, and customers.	
	36	Our BDA managers have a good sense of where and when to apply BD.	
	37	Our BDA managers can understand and evaluate the output extracted from BD.	
Innovation process	38	Increase in our technology competitiveness	

Construct	No	Item	Source
	39	The speed of adopting the latest technology	(De Luca & Atuahene-Gima, 2007; Jiménez-Jiménez & Sanz-Valle, 2011; Narver et al., 2004; Prajogo, 2006; Slater & Narver, 2000)
	40	The updated-ness or novelty of technology used in the process	
	41	The rate of change in process, techniques, and technology	
	42	The number of changes in process introduced	
	43	Pioneer disposition to introduce new process	
	44	Clever response to new processes introduced by other companies in the same sector	
	45	Better than our competitors in reducing the cost of product/ service	
Competitive advantage	48	Our competitors cannot copy us easily	(Côrte-Real et al., 2017; Shan et al., 2019; Zhou et al., 2009)
	49	Our company become unique in its products/services, and nobody can offer it	
	50	No one can copy our brand name easily	
	51	Our advantages are optimised in the firm and not in individuals - no one can copy us by stealing our staffs away from us	
	52	No one can copy our business procedures, methods, and culture	
	53	Reduce the cost of our product/service	
	54	Increase delivery dependability	
	55	Increase speed to reach markets	
Firm performance	56	Customer retention	(Jiménez-Jiménez & Sanz-Valle, 2011; Liu et al., 2013; Wamba et al., 2017; Wang & Wang, 2012)
	57	Sales growth	
	58	Return on investment	
	59	Overall financial performance	
Environmental turbulence	10	The environment in our product/service area is continuously changing.	(Hajli et al., 2020; Jiménez-Jiménez & Sanz-Valle, 2011)
	11	Environment changes in our industry are very difficult to forecast	
	12	The product/service technology is changing rapidly in our industry.	
	13	Technological breakthroughs provide big opportunities in our industry.	
	14	In our kind of business, customers' product/service performance change a lot over time.	
	15	Marketing practices in our product/service area are constantly changing. New product/service introductions are very frequent.	

4.8.5 Response Formats

It is essential to ensure that the data are received in a way that can be easily analysed. According to Bhattacherjee (2012), there are five different formats for responding in structured surveying: nominal, dichotomous, interval level, ordinal, and continuous.

- Nominal response formats give participants more than two options without any total order "United Kingdom/United States/Europe."
- The dichotomous format is used for category questions such as "male/female".
- The ordinal format offers participants two or more options in answering questions, and it is suitable for questions such as 1-5 years, 6 -10 years, and 11-15 years.
- "Scale" format used for 5-point Likert scale questions that present the main variables (effective data analysis, effective data management, innovation process, competitive advantage, and firm performance).
- "Text" format for an open-ended question that is used for any other suggestions and comments.

Table 4.10 presents the variables, question type, and question formats used in the current study in more detail. Except for the last question, all the questions used in the final questionnaire comprised closed-ended questions.

Table 4.10
Questions / Responses Format

Variable	Question Type	Question Formats	Responses
Age	Quantity	Ordinal	<30; 31-40; 41-50; 51-60; 60 +
Gender	Category	Nominal	Male; Female; Others
Education Level			High School; University Graduate; Postgraduate; PhD.
Job Title			Manager; Executive; Senior staff; Vice; Other managerial positions
Been in Position			Less than a year; 1-5 years; 6-10 years; 11-15 years; 16-20 years; 20+ years
Activity Sector			Agriculture; Automotive; Energy; High Tech and communication; Financial and Insurance; Health Care; Real Estate; Services; Other

Variable	Question Type	Question Formats	Responses
Country			UK; United States; EU; Other
Number of Employees		Ordinal	<200; 200~500; 500~1000; 1000~3000; 3000~5000
Age of the Company			Fewer than 5 years; 5-10 years; 10-15 years; 15-20 years; 20+ years
Effective Use of BDA Tools	Five-point Likert scale	Scale	1 = Poorly developed, 2 = Slightly developed, 3 = Averagely developed, 4 = Developed, 5 = Well developed
BD management			
Innovation process			
Competitive advantage			
Firm performance			
Environmental turbulence			
General Question	Open-ended	text	Free- text answer

4.9 Validity and Reliability of the Questionnaire

The terms validity and reliability appear to be synonyms, but they have quite distinct meanings in evaluating conceptual scales (Bryman & Bell, 2007, p. 157). According to Saunders (2015b), in terms of maximising the reliability and validity of a study, many steps should be considered. The measurement tools should contain a clear layout, relevant questions, clear explanations, pre-testing, pilot-testing, and methodical management. Reliability is primarily involved with issues of consistency of measures, and there are three different meanings of reliability: stability, internal reliability, and inter-observer consistency (Bryman & Bell, 2007, p. 158). Therefore, it is essential to ensure the consistency of the tool in developing the questionnaire, which in turn helps to collect accurate and credible data.

The validity measurement issue concerns whether the items related to a variable measure the concept correctly (Bell et al., 2018, p. 174). For instance, the ability of a research instrument to measure what it is designed to measure (Punch, 2013, p. 100). According to Malhotra et al. (2006), there are three approaches to validating a questionnaire: content validity, criterion-related validity, and construct validity. Content or face validity refers to the length to which the questionnaire presents sufficient coverage of the analytical questions (Malhotra et al., 2006). Construct validity is the extent to which measured items reflect the latent theoretical construct those items are designed to measure (Hair et al., 2006). In other words, content validity is a subjective but systematic evaluation of the content that reflects a theory or a latent construct. In contrast, construct validity measures the anticipated performance of a measurement scale concerning other chosen variables as meaningful criteria (Malhotra et al., 2006). The content validity of the survey was measured through professionals in the relevant field and via the pilot-testing phase. Further, construct validity has been measured via convergent and discriminant validity by conducting a factorial analysis (Malhotra et al., 2006). An in-depth discussion of the construct validity is presented in the subsequent data analysis chapter.

4.10 Errors Emanating from the Research Procedures

According to Bell et al. (2018, p. 196), there are four resources of errors in social survey research. This first is sampling error, which occurs because it is incredibly unusual that one will end up with a truly representative sample, even when probability sampling is applied. Second is that non-sampling relates errors, which can happen when the sampling frame is inaccurate, or some members of the sample refused to cooperate. Third, data collection error is caused by poor question-wording in self-completion survey or flaws in the administration of study instruments. Finally, data processing errors may arise from incorrect data management, particularly errors in the coding of responses. Sampling errors could broadly be controlled by choosing the appropriate sampling design method, methodology, and sample size (Groves, 2004). Moreover, with high sample size, the effect of sampling error is minimised; thus, in this study, a sustained attempt was made to increase the sample size as much as possible (Bryman & Bell, 2007). In addition, non-sampling errors appear due to different causes right from the primary step when the survey is being designed to the final phase when data are analysed. They may take various forms,(e.g., non-response, observation errors (Banda, 2003)). Non-response errors fail to measure some of the sample units, while observation errors could happen due to inadequate question expressions, errors in measurement scales, or data processing and analysis techniques; such errors are more challenging to treat than non-response errors (Churchill & Iacobucci, 2006). However, adopting pre-validated scales from existing literature can minimise observation errors. Hence, to minimise observation errors, this study has adopted measurement scales from existing literature which has been illustrated (illustrated in Table 4.10). Also, the pre-tests and pilot studies confirmed respondents' understanding of the questions' wording.

Furthermore, data processing errors were reduced by carefully entering data from paper surveys and using an online platform for collecting, which enabled the researcher to export the data collected from the electronic surveys directly into the statistical software package for analysis (Brace, 2018). Finally, to avoid data processing errors, statistical software such as SPSS and AMOS was conducted most appropriately manner to facilitate the elimination of errors in data analysis (Brace, 2018).

4.11 Ethical Issues

According to Fowler Jr and Cosenza (2009), research may involve the intersection of people's right to privacy and intellectual freedom and the search for knowledge. Therefore, to deal with ethical considerations, this research obtained approval from Swansea University's Ethics Commission. Also, this study considered issues related to ethical guidelines (Saunders, 2016; Sekaran & Bougie, 2016). In the cover letter, the researcher provided sufficient information regarding the study's objectives and contribution and detailed information about the researcher, the institute, and voluntary participation. The participant did not disclose their identity in research; therefore, there was no personal risk about participating in this investigation (Sekaran & Bougie, 2016). Third, the researcher informed participants that the companies' names would not be published at any phase of the research due to confidentiality. Fourth, because participation in this study was voluntary, the participant could opt out of filling out the questionnaire at any time. Finally, the researcher suggested an opportunity for the firm to obtain access to the summary of the study outcomes. Consequently, this research followed ethical guidelines or required societal criteria of behaviour (Sekaran & Bougie, 2016).

4.12 Statistical Data Analysis

The main aim of data analysis *is "getting a feel for the data, testing the goodness of data, and testing the hypotheses developed for the research"* (Sekaran & Bougie, 2002, p. 36). Data analysis may take a simple form such as descriptive statistics or more complex forms as discriminant analysis, factors analysis, regression and structural modelling (Straub et al., 2004). The current research analyses collected data via a set of univariate and multivariate statistical techniques. The univariate analysis deals with issues, such as analysing descriptive statistics, missing values, outliers, data normality, multicollinearity, homoscedasticity, and non-response bias tests (independent t-test, ANOVA). In contrast, the multivariate analysis involves exploratory factor analysis (EFA), reliability, sample adequacy, confirmatory factor analysis (CFA), discriminant validity, convergent validity, and structural equation modelling (SEM).

4.12.1 Preliminary Data Analysis

After cleaning, coding, and screening the data, the study performed different univariate analyses to understand the data better and to measure the tendency (e.g., mean, median, mode), dispersion (e.g., range, quartiles) and spread of a single variable, e.g., variance, standard deviation (Hair et al., 2010). At first, the study explored missing data to increase reliability, validity, and generalisability (Tabachnick et al., 2007). Second, data normality and outliers were identified using the Mahalanobis distance (D2/df, where df=the number of variables) and skewness, Kurtosis z-values, and Shapiro–Wilk and Kolmogorov-Smirnov, Histograms, Normal Q-Q plots, and Box plots techniques to understand the normal distribution of data. Third, the study used multicollinearity and homoscedasticity tests using the residual scatterplots and Levene's Test for Equality of Variances. Also, the tolerance and Variance Inflation Factor (VIF) was used to identify multicollinearity issues in the dataset. Fourth, an independent sample t-test was run to assess non-response bias of the respondents. This assessment aims to determine the extent to which non-respondents were different from the respondents to the survey (Dillman, 2000). Finally, a variety of descriptive statistics were used, such as the mode, mean, range, standard deviation, and variance, to present and summarise the data and describe variances for every single variable.

4.12.2 Structural Equation Modelling

Structural equation modelling (SEM) was applied in this research where it acted as a suitable statistical analysis to confirm the conceptual model and test the hypotheses of the research (Byrne, 2013; Hair et al., 2010). SEM becomes a fundamental practice in marketing research (Bagozzi & Yi, 1988; Hulland, 1999) because it can provide scholars with comprehensive means for evaluating and adjusting theoretical models (Anderson & Gerbing, 1988). SEM can analyse related questions in a single analysis by modelling the relationships between independent and dependent constructs. In addition, it is a robust set of multivariate statistical analysis techniques that support comprehensive and simultaneous assessments of all relationships for a complex and multidimensional phenomenon (Byrne, 2013). Consequently, most of the Information System (IS) and BD studies used SEM techniques. As a multivariate procedure in data analysis, structural equation modelling contains many issues that have taken up a great deal in scholarly discussions (Byrne, 2013; Tabachnick et al., 2007). Scholar such as Byrne (2013) highlighted several advantages to use SEM. One of SEM's primary privileges is that it can study the relationships among latent constructs indicated by multiple measures. It is also applicable to experimental and non-experimental data and cross-sectional and longitudinal data (Lei & Wu, 2007).

Table 4.11
The Significant Differences between SEM and other Multivariate

SEM methodology
Confirmatory approach:
It provides explicit estimates of the error variance parameters
SEM uses both observed and unobserved variables in data analysis
It is an easy and widely used method to examine both direct and indirect impacts between constructs in one shot.
Other multivariate procedures:
Exploratory approach:
They are incapable of either assessing or correcting for measurement error (e.g., regression or the general linear model)
Previous data analysis methods rely only on observation measurements.
They cannot measure the indirect influence among model relationships.

Source: Adopted from Byrne (2013, p. 3)

SEM software supports three different levels of statistics: individual path, overall model, and modification (Gefen et al., 2000). According to Anderson and Gebing (1988) that SEM analysis can take a two-step approach. First, assess the measurement model and then the structural model. The measurement model (confirmatory factor analysis) describes the latent variables or constructs that the model will utilise and allocates to each observed variable, while the structural model defines the causal relationship between variables (Hair et al., 2010; Iacobucci, 2009). In other words, the measurement model aims to evaluate convergent and distinct validity, whereas the structural model evaluates nominal validity. Thus, SEM uses two techniques: full information and limited information (Schumaker & Lomax, 2004).

4.12.2.1 Model Identification

Model identification is an essential step that should be done before parameter estimation to evaluate the hypothesised model (Schumacker & Lomax, 2010). Identifying a model refers to the possibility of finding adequate estimates for every parameter with an unknown value in the model (e.g., factor loadings, correlation). Model identification has to do with the difference between the number of variables and the number of parameters that need to be estimated by the model (Meyers et al., 2006). This analysis can be done via the chi-square (x^2) value and associated degrees of freedom. According to Hair et al. (2010), applying the chi-square (x^2) technique for model identifying could result in three levels. First, in the "under-identified" level, the model's degrees of freedom are negative (- *df*), and the parameter estimates are not trustworthy. Second, the "just-identified" model means the result is equal to unknown elements (*df = 0*). Third, the "over-identified" model happens when df > 0, and the number of known factors is more significant than the number of unknown parameters. The estimated parameter numbers should not be more than the number observed within the model to obtain an over-identified model (Byrne, 2010).

4.12.2.2 Confirmatory Factor Analysis (CFA)

Factor analysis is a statistical method used to explore a small set of unobserved variables that can obtain covariance among a more extensive set of observed variables (Albright & Park, 2009). According to Marshall et al. (2015), there are two kinds of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). CFA was used mainly for two reasons. The first was to examine whether the adopted and developed measures were suitable for the population examined in this research (Harrington, 2009). The second was to examine the significance of the theoretical measurement model and whether the sample data proved the proposed model and its validity (Schumacker & Lomax, 2010).

CFA is a theory-testing and hypothesis-driven procedure (Albright & Park, 2009; Roberts, 1999) that aims to confirm the specified set of factors. Similarly, Hair et al. (2010) and Harrington (2009) stated that the primary objective of CFA measures the validity of theoretical measurements proposed model for assuring whether the data collected confirms the model that is proper for the population investigated in the research. Therefore, the CFA measurement model's adequacy is evaluated through the following steps:

4.12.2.2.1 Model Estimation

One of CFA's and SEM's key objectives is to achieve estimates such as factor loadings, covariances, variances, and errors for each parameter and path in the conceptual model. Several methods or estimations procedures can be used for model estimation, such as the maximum likelihood (ML), least squares (LS), unweighted least squares, generalised least square, and asymptotic distribution-free (ADF) (Schumacker & Lomax, 2004; Weston & Gore Jr, 2006). The ML method is an estimation method for model parameters, and it *"aims to find the parameter values that make the observed data most likely (or conversely) maximise the likelihood of the parameters given the data"*(Brown, 2006, p. 73). Harrington (2009) highlighted two distinct reasons for using the ML method; first, it can present standard error for each parameter used to calculate the p-value. Second, *"its fitting function is used to calculate many goodness-of-fit indices"* (p. 29).

More specifically, the current study used the ML estimation method for the following reasons:

- The ML has preferable and acceptable asymptotic characteristics such as lowest level variance and impartiality, and it is scale-free (Hair et al., 2010; Schumacker & Lomax, 2004).
- The ML is known as one of the most common estimation methods (Hair et al., 2010; Iacobucci, 2010).
- The ML is more potent with observations that are slight non-normality (quietly peaked or skewed) (Bagozzi, 2010). In addition, according to Lei and Lomax (2005), the ML estimation method is superior to the general least square (GLS) estimation method in the case of small to moderate deviation from normality.
- The ML estimation method proved to be more sensitive to model error estimation than the ordinary least square (OLS)(Fan et al., 1999).

4.12.2.2.2 Model Evaluation

After estimating the parameters, the following step is to evaluate the data appropriateness for the model. This step asks whether the observed sample data support the theoretical model (Schumacker & Lomax, 2010). The evaluation process can be done on three levels: the overall measurement model, the individual construct measurement model, and the structural model. The relationships among the observed variables and the latent variable will be specified by the measurement model, which will be done using CFA (Hair et al., 2010). A non-significant chi-square x^2 indicates that the two matrices are similar. In other words, significant x^2 indicates that the two metrics are different, which points to a problem with the model fit. However, using x^2 statistics for model assessment might be misleading in at least two ways (Shah & Goldstein, 2006). The overall model fit is assessed by selecting several fit indicators such as absolute, incremental, or parsimonious fit indices (Hair et al., 2010).

Absolute Fit Measures discover how well a model fits the sample data (Hu & Bentler, 1998) or, as Meyers et al. (2006) stated, *"how well the correlation/covariance of the hypothesised model fits the correlation/covariance of the actual or observed data"*(p. 558).

Hair et al. (2010) and Najafabadi et al. (2015) stated that it determines the model's suitability to the sample data. Chi-square (x^2) is used to measure essential goodness-of-fit to determine the variations among the observed and estimated metrics; therefore, if the chi-square value indicates as non-significant, two matrices are similar, and the model fits. In contrast, if the chi-square value is significant, it indicates the difference between the two metrics and the model's problem (Hair et al., 2010). However, the chi-square test is affected by the sample's size, and it could be misleading by rejecting the model in two ways. First, when the model is rejected (error I), it could happen when the sample size is greater than 200, and with a small sample, it might lack power and may be accepted as poor models (error II) (Byrne, 2010; Hair et al., 2010; Schumacker & Lomax, 2010). It has been found that the (x^2) statistic is highly sensitive to violations of the assumption of normality (Jöreskog & Sörbom, 1986).

Researchers (e.g., Chin and Todd (1995) have been advised that the chi-square value, regardless of whether it is significant or not, should be stated. To solve the problem of chi-square, utilising (x^2/df ratio) to minimise the sample size's impact is recommended (MacCallum et al., 1996). Nevertheless, it is suggested that the ratio x^2 to the degree of freedom should be less than value 3.0 (Chin & Todd, 1995; Hair et al., 2010). To overcome the chi-square (x^2) indicator's limitations and weaknesses, researchers can use other statistical indicators in respect of testing goodness of fit (Hu & Bentler, 1998), and scholars can use several other indexes that are also included in the category of absolute fit indices (Hair et al., 2010).

The Goodness of Fit Index (GFI) and its Adjusted Goodness of Fit Index (AGFI) are the absolute fit indexes that compare the proposed model with the absence of any model, and they are based on the degrees of freedom and define the ratio of variance (Schumacker & Lomax, 2010; Tabachnick et al., 2007). According to Hu and Bentler (1995), if the GFI value is close to 1, the model is a good fit; however, the GFI should be equal to or bigger than 0.90 for a model to be accepted, and the AGFI value should exceed .80 (Gefen et al., 2000). Nevertheless, many scholars stated that GFI is becoming less used due to the sensitivity of sample size and the model's misidentification (Gefen et al., 2000; Hair et al., 2010; Hu & Bentler, 1998). Another absolute fit measure is the Root Mean Square Error of Approximation (RMSEA), which has been identified as one of the most useful and broadly applied indicators in covariance structure modelling (Hair et al., 2010). RMSEA *"estimates the lack of fit in a model compared to a perfect (saturated) model"* (Tabachnick et al., 2007, p. 717). Dissimilar to chi-square (x^2), RMSEA has been widely accepted as a Good of Fit (GOF) indicator because of its advantages in avoiding the rejection of models with too much sampling and/or too many variables (Hair et al., 2010). According to Hair et al. (2010), if there are more than 30 variables within the model and the sample size is more than 250, it is recommended that a cut-off value should be 0.07. However, some scholars such as MacCallum and Browne (1993) and Jöreskog and Sörbom (1993) advised that values ≤ 0.08 indicate a proper and adequate model fit. Furthermore, Hu and Bentler (1999) stated that RMSEA ≤ 0.06 is a good model fit.

Incremental fit indices evaluate the fit of the proposed model compared to the baseline or (null) model (Hair et al., 2010; Hu & Bentler, 1998). The Comparative Fit Index (CFI) is one of the operational indices. CFI assumes that all unobserved variables are uncorrelated and compare the sample covariance matrix to the null model. When the CFI value is higher than 0.90, it is usually accepted as a well-fitting model (Bentler & Stein, 1992; Hu & Bentler, 1999). The Normed Fit Index (NFI) is another incremental fit measure that compares x^2 of the proposed model with x^2 of the null model (Liyanage et al., 1999). Again, the NFI values should be above 0.90 to be accepted (Chin & Todd, 1995; Meyers et al., 2006).

Parsimony fit indices (PFI), or adjusted fit measures, such as the parsimony normed Fit Index (PNFI) and Parsimony Goodness of Fit Index (PGFI), have been developed mainly to provide information about the best model among several competing models, concerning its complexity by considering the number of parameters in the evaluated model (Hair et al., 2010). Consequently, though these indices do not help validate a model, they are beneficial for comparing alternative models with various complexity levels (Byrne, 2010; Hair et al., 2010). PNFI improves the incremental fit index of NFI via multiplying its value by the parsimony rate (Hair et al., 2010). Therefore, the values of PNFI can be used to compare various models, considering the model complexity degree (Byrne, 2010). Both PNFI and PCFI have values ranging between 0 -1, which are much lower than those accepted based on the normed indices. However, if the values of PNFI and PCFI are more than 0.50, it indicates a good fit (Byrne, 2010; Meyers et al., 2006; Mulaik et al., 1989).

AMOS software provides several GOF indicators. Deciding which to report has been an issue of disagreement among the scholars; however, they agreed that researchers do not need to report all GOF (Byrne, 2010; Hair et al., 2010; Hu & Bentler, 1999; Tabachnick et al., 2007). For instance, Hair et al. (2010) stated that if varying complexity models are compared, in addition to chi-square (x^2) with associated df, at least one absolute index, one incremental index, and one parsimony index should be reported. The PNFI indicator is also appropriate in the case of comparing models of different complexity. More clearly, to evaluate model fit, indicators such as chi-square (x^2) with the associated degree of freedom, CFI and the RMSEA report adequate information (Meyers et al., 2006). Furthermore, Meyers et al. (2006) and Hair et al. (2010) stated that the researcher must present at least three fit tests to reflect various criteria, including one incremental, one absolute, and one parsimonious. In this regard, Tabachnick et al. (2007) mentioned that RMSEA and CFI are the most appropriate indicators that have been used frequently, particularly RMSEA, supporting the researcher's purpose to do power calculations. Moreover, Iacobucci (2010) and MacKenzie et al. (2011) recommended using CFI, SRMR and RMSEA for indicating a well-fitting model. Also, Echambadi et al. (2006) stated that in respect to selecting the best model, fit indices, factor loadings, and structural paths should all be applied. Given the above and mentioned recommendations, this study used the three fit measures to evaluate the model fit: absolute, incremental, and parsimonious. Table 4.12 summarises model fit statistics that will be used in the current study (Segars & Grover, 1993).

Table: 4.12 Summary of Model Fit Statistics

Model fit indices	Acceptable values	Source(s)
Absolute Fit Measures:		
Chi-square (X^2)	> 0.05	(Bagozzi, 2010; Boyle et al., 1995; Hu & Bentler, 1999)
χ^2/df	< 3	(Bentler, 1990; Chau, 1997; Chin & Todd, 1995; Hair et al., 2010; Kline, 2015)
RMSEA	≤ 0.07	(Hair et al., 2010; MacCallum et al., 1996; Meyers et al., 2006; Sharma et al., 2005)
SRMR	< 0.08	(Hair et al., 2010; Hu & Bentler, 1995)
AGFI	≥ 0.80	(Chau, 1997; Gefen et al., 2000; Hair et al., 2010; Segars & Grover, 1993; Straub et al., 2004)
Incremental fit indices:		
CFI	≥ 0.90	(Bentler & Stein, 1992; Hair et al., 2010; Hu & Bentler, 1995; Iacobucci, 2010)
IFI	≥ 0.90	(R. Kline, 2011; Schumacker & Lomax, 2010)
TLI	≥ 0.90	(Bentler, 1990; Hu & Bentler, 1995, 1998; Sharma et al., 2005)
NFI	≥ 0.90	(Chin & Todd, 1995; Tabachnick et al., 2007)
Parsimony fit indices:		
PNFI	> 0.50	(Byrne, 2010; Chow & Chan, 2008; Hair et al., 2010; Meyers et al., 2006)
PCFI	> 0.50	(Byrne, 2010; Chow & Chan, 2008; Hair et al., 2010; Meyers et al., 2006)

4.13 Chapter Summary

This chapter outlined and discussed the research process (philosophy, approach, strategy, methodology, time horizon, sampling method, data collection process, and data analysis techniques). Different views on research design were presented to elucidate the assumptions that underlie the research methodology. Furthermore, the reasons and justifications for selecting the procedures, methods, and techniques chosen were explained. Since this research aimed to test research hypotheses and conceptual models, an objective ontology position, positivist epistemological paradigm, and deductive approach have been selected to support this study. First, the target population was determined, then the quantitative method with questionnaire techniques was chosen for data collection. Also, the geographical dispersion of participants and other advantages of an online questionnaire was selected to distribute the survey. The sampling technique and sample size were also discussed. Moreover, the questionnaire design stages and appropriate techniques for conducting statistical analysis have been explained.

Chapter Five: Data Analysis

5.1 Introduction

The purpose of statistical data analysis is to answer research questions, objectives, test hypotheses, and conceptual models. This chapter describes and analyses the data obtained from managers UK and U.S. firms, helping to respond to research purposes, related questions, and hypotheses. The task of analysis is to transform large, complex, and even incomprehensible data sets into logical units, patterns, and justifications of research-related issues. To do this, first, the data were prepared and cleaned from issues such as missing values, outliers, non-response bias, normality, linearity, homoscedasticity, and multicollinearity. Next, a conformity factor analysis (CFA) was performed following the first step to validate the measurement model. Finally, structural equation modelling was used to assess the relationships between variables to validate the proposed research model.

5.2 Preliminary data analysis

Statistical rigour is essential for proceeding with any data analysis. Therefore, data must be evaluated through the standard statistical procedures before undertaking structural equation modelling (SEM) to ensure that it meets the needs of the multivariate techniques (Hair et al., 2006; Kline, 2011). Furthermore, preparing, cleaning, and treating the data enhances the reliability and validity of the outcomes (Tabachnick et al., 2007). In addition, the multivariate analysis may particularly suffer from the effects of missing data, outliers, and violations of normality assumptions (Hair et al., 2010). Hence, in the current study, the necessary steps of data screening and preparation such as assessing missing value, outliers, normality, homoscedasticity and linearity, non-response bias were checked to ensure that the data were ready for conducting the primary statistical analyses. The following sections will discuss the preliminary data analysis techniques performed in SPSS v.26. SPSS is a software used for data entry and to perform the preliminary examinations of data, and AMOS has been used as an add-on module for this package to perform SEM. The analyses included testing missing data, outliers, homoscedasticity, normality, and non-response bias tests.

5.2.1 Sample Summary

The current study uses a two-step data collection process. First, 2,500 questionnaire links were distributed via email to the listed companies at the Financial Analysis Made Easy (FAME) Database across the United Kingdom (UK) and Ireland. The FAME Database includes details of more than 176,000 firms registered in the UK and Ireland (FAME, 2022). Firms had been chosen randomly, where every firm had an equal chance of being selected from among the long list of firms. First, 176,000 firms were divided into 2,500, equal to about 70. Then the selection was done on numbered firms 70, 140, 210, and so on. The questionnaire link was sent with a cover letter to the email addresses of the elected firms. After a reminder and several attempts, a total of 84 samples was collected, which is inadequate for the research purpose. Therefore, in the second phase, the current study used Qualtrics, an online platform, where 120 questionnaires were distributed, and 90 completed questionnaires were received. In total, 2,620 questionnaires were distributed in total, and 174 responses were obtained from both the data collection stages. After data screening and preparation process, the number of final useable responses was 143. The rate of overall response was very low (5.46%) because the response rate in the first data collection stage was extremely low (2.52%), which ultimately affected the overall response rate. However, considering the second stage of data collection, the response rate was reasonable (66.7%). Table 5.1 highlights the distribution of the survey responses. There may be several reasons. First, a massive number of received electronic responses indicated that the emails were not delivered to the concerned party, meaning that firms may be using software that may not allow the questionnaire emails to pass through. Second, it is possible that firms' information provided by FAME is not valid or is not up to date.

Table 5.1
Response Rates

Distribution	Data collection stages	Totals distribute	Total collected response	Useable response	Usable response rate (%)
Online	1st Stage: FAME Database	2,500	84	63	2.52
Online	2nd Stage: Online questionnaire through Qualtrics	120	90	80	66.7
Total	Overall	2,620	174	143	5.46

5.2.2: Respondents' characteristics and contextual information

After screening the data and ensuring its suitability for further data analysis, this section analyses the demographic characteristics of the respondents who have completed the survey. The survey predominantly targeted top managers in the firms as they are more knowledgeable about BDA technology. Therefore, in terms of gender, there were more men representing management positions (71.9%) than (28.1%) women. These outcomes are a line with the other studies (e.g., Wamba et al. (2017); Waqas et al. (2021). Women traditionally are underrepresented in leading positions in many countries (Commission, 2013). However, the lack of women's representation in managerial positions (28.1%) in the targeted firms may not directly affect the results of the current study. In addition, most of the respondents (41.2%) were aged between 31–40 years and followed by the second majority (33.3%) aged 41–50 years, which shows that most of the respondents were middle- aged. Furthermore, most respondents held postgraduate (45.8%), followed by university graduates (21.6%); 19% of respondents had a PhD. These rates are reasonable when the research is targeted the top managers of firms. Also, nearly half of the respondents (47.7%) were managers, 35.9% were executives, 7.2% were senior managers, and the others held other managerial positions. Additionally, regarding job duration, the most majority (31.4%) of respondents had been in their position for more than 1 year, and the second majority (28.1%) had been in their position for more than 5 years. In contrast, just 17.0% had been in their position for less than 1 year. The results show that the managers in the survey represented five firms in terms of size-the percentage of participants from small firms, 32.7. Managers from medium firms also participated at 18.3% and from the slightly large with 19%, and larger firms with 15%. Finally, the percentage of participants from huge firms reached 10%. These results indicate that the survey included managers from all types of firms in terms of size, where the findings based on their answers can provide a broad theoretical and practical contribution.

Moreover, the survey covered different firms in terms of their age, where nearly half of them were 10–20 years old, and about 18.3% were between 5–10 years, and the same proportion held for firms more than 20 years old. Furthermore, the responses received came from companies of a diverse industry background such as agriculture, automotive, energy, high tech, and communication, financial and insurance, health care, real estate, and service. However, the high tech and communication business had the highest proportion (35.9%), followed by financial and insurance with 15%, and the lowest participant rates were for the health care sector (.07%) and agriculture (1.3%). Since the questionnaire was distributed electronically, the researcher hoped that it would cover the largest area in terms of geography so that results would be more comprehensive; thus, the survey included firms in different countries. However, most of the despondences came from the United Kingdom (37.3%) and the United States (35.9%).

Table 5.2
Characteristics of Research Sample (N=143)

Demographics	Variable Item	Response	Percentage (%)
Gender	Male	110	71.9
	Female	43	28.1
Age	<30	20	13.1
	31-40	63	41.2
	41-50	51	33.3
	51-60	16	10.5
	61+	3	02.0
Education	High School	21	13.7
	University graduate	33	21.5
	Postgraduate	70	45.8
	PhD	29	19.0
Position	Executive	55	35.9
	Manager	73	47.7
	Senior Staff	11	07.2
	Vice	9	05.9
	Regional Manager	1	00.7
	Other managerial position	4	02.6
Time Position	Less than a year	26	17.0
	1-5 years	48	31.4
	6-10 years	43	28.1
	11-15 years	19	12.4
	16-20 years	11	07.2
	20+ years	6	03.9
Firm Size	<200	50	32.7
	200-500	28	18.3
	500-1000	29	19.0
	1000-3000	23	15.0
	3000-5000	15	09.8
	More than 5000	8	05.2
Firm Age	Fewer than 5 years	19	12.4
	5-10 years	28	18.3
	10-15 years	33	21.6
	15-20 years	45	29.4
	More than 20 years	28	18.3
Activity Sector	Agriculture	2	01.3
	Automotive	6	03.9
	Energy	3	02.0
	High Tech and communication	55	35.9
	Financial and Insurance	23	15.0

Demographics	Variable Item	Response	Percentage (%)
	Health Care	1	00.7
	Real Estate	14	09.2
	Service	12	07.8
	Other	37	24.2
Country	UK	57	37.3
	US	55	35.9
	EU (European Union)	1	00.7
	Other	40	26.1

5.3: Data Screening and Preparation

Before statistical analysis of data, they should be screened to ensure that the data are acceptable in terms of completeness and quality. In addition, screening is crucial to guarantee that data are consistent, practical, and valid to analyse the proposed concept. The data screening and preparation was done in the current study by examining missing values, outliers, linearity, non-response bias, homoscedasticity, and multicollinearity. The potential issues of these stages and their treatment methods will be discussed in detail in the following sections.

5.3.1: Assessment of Missing Values

Missing values are a common issue in data collection via the survey method (Lepkowski et al., 1987; Tsikriktsis, 2005), and it could significantly impact reliability, validity, and generalisability (Tabachnick et al., 2007). According to Roth et al. (1999), the missing value might happen in missing whole instruments or missing just a few items. Also, there are two significant adverse effects of missing value: they may negatively affect statistical power, and missing values may cause biased estimates in different ways. Moreover, Tsikriktsis (2005) highlighted different reasons for missing data, such as mistakes in data entry, disclosure restrictions, failure to complete the whole survey, refusal to address sensitive questions, and having participants who do not know or have insufficient information to respond to the questions. Raymond and Roberts (1987) stated that missing data is inevitable; however, the amount of missing data and its effect on statistical analysis is debatable. For instance, Byrne (2010); Cohen and Cohen (1983) recommended that the ratio of 5% or even 10% of missing data is acceptable and is random. According to Little and Rubin (2019, p. 14) random missing value is critical, as it not only restrains the capacity of a researcher to interpret the observed phenomena but also directs to an inaccurate description that outcomes in partial statistical explanation and report (McKnight et al., 2007). In the current study, SPSS software was used to sort out the problem of missing values, and the univariate tests revealed that variables showed one case of .0057 missing values. The “Competitive Advantage” variable showed 12% non-random missing values.

In addition, the outcomes of Little’s MCAR test confirmed that 21 values were missing from one construct item, and one missing value was recognised as random. Non-random missing values are only related to the construct “competitive advantage.” The rate of non-random missing data was more than 12%; therefore, appropriate remedial action cannot be taken as the missing value is more than the acceptable rate of 10%, which can treat values using some techniques. Therefore, the non-random items of missing value were removed, whereas the Expectation-Maximization (EM) algorithm was used to support the imputation of random missing values (Dempster et al., 1977). Overall, after the data-cleaning process deleted 21 non-random cases, so the number of responses was reduced from 174 to 153 samples. (Further sample analysis reduced sample size to 143, see section 5.3.2: Assessment of outliers), p.122.

5.3.2: Assessment of Outliers

In research, outliers are described as “observations with a unique combination of characteristics identifiable as distinctly different from the other observations” (Hair et al., 2010). According to Gao et al. (2011), there are two kinds of outliers: univariate and multivariate. Univariate is an extreme value on a single variable, while multivariate is a state containing an unusual set of values in two or more variables (Kline, 2015). Scholars such as Meyers et al. (2006) and Tabachnick et al. (2007) described four reasons that can lead to outliers: error in sampling, error in data entry, wrong transcription of missing value, and observations that might happen within the reasonable range of values on each of the variables. Failure to address outliers’ issues may negatively affect the data analysis results in three forms: first, outliers may increase the variance of error and weaken the efficiency of statistical analysis. Second, the outliers can affect the normality distribution and alter the odds of forming types I and II errors if they were non-normally distributed. Third, outliers may lead to detrimental outcomes or biased assumptions that limit the generalisability of the study (Rasmussen, 1988; Schwager & Margolin, 1982). Therefore, it is essential to evaluate outliers through univariate, bivariate, and multivariate assessments to avoid extraordinary misrepresentations of parameter estimates or increased error rates.

To do this, scholars such as Hair et al. (2010); Tabachnick et al. (2007); Tinsley and Brown (2000) recommended several methods to assess and treat data outliers. For example, univariate outliers can be treated by converting the data values to standard scores (z-scores). At the same time, the Mahalanobis distance (D2/df, where df = the number of variables) method can be applied for outliers of multivariate level. The current study, which contains more than two variables, identified outliers using a normality test and multivariate assessment (Mahalanobis distance method). Following scholars' recommendations, the outliers were checked by converting the data values to standard scores z-score > 1.9 (p < 0.001). At first, the outlier's data remained to check their effect on the model; however, a significant impact was noted. Thus, to obtain reasonable outcomes, the extreme cases were removed as suggested by (Hair et al., 2010). The results of the Mahalanobis distance assessment shows that some of the responses are more than the targeted value. Table 5.3 shows the results of multivariate outliers, and those cases indicated as outliers were deleted. Overall, 143 out of 153 responses were considered useful for further data analysis.

Table 5.3
Identification on Multivariate Outliers

Multivariate Assessment				
Case Number	**Mahalanobis D^2**	***df***	**D^2/df**	**Significance**
116	116.2142	63	1.844669841	0.0000
153	113.5751	63	1.802779365	0.0000
100	104.0939	63	1.652284127	0.0000
58	97.3284	63	1.544895238	0.0002
65	97.2966	63	1.544390476	0.0002
115	96.6403	63	1.533973016	0.0002
129	95.971	63	1.523349206	0.0003
90	94.6861	63	1.502953968	0.0004
28	91.2933	63	1.4491	0.0008
Note: df = degree of freedom, The D^2/df value is approximately distributed as a *t* - value, *p* >. 001				

5.3.3: Assessment of Normality

Data normality is the essential assumption of the validity of all statistical analyses for any multivariate examination (Hair et al., 2010). According to Meyers et al. (2006) "*shape of a distribution of continuous variables in a multivariate analysis should correspond to a (univariate) normal distribution*"(p. 67). Also, Hair et al. (2010), stated that the severity of non-normal distribution is affected by two distinct dimensions: the sample size and the offending shape of the distribution. They stated that if the sample size is less than 50, a large deviation from normality may have a significant effect; however, if the sample size is over 200, the slight departure from normality can be ignored. The normality of data can be investigated through several tests. For example, the data distribution's shape can be measured by conducting the skewness and kurtosis statistics techniques (Hair et al., 2010; Thode, 2002). It is supposed that skewness determines whether the data distribution follows a normal bell-shaped curve and looks identical to the left and right of the centre point. The skewness and kurtosis z-values are considered two probability density functions, measure the "tailedness" of a kurtosis by computing a z- score for each variable. The skewness and kurtosis values show that each item's distribution is in the threshold range of ±1.96, which indicates it can accept a normal distribution. Also, Shapiro–Wilk and Kolmogorov-Smirnov tests were used for normality tests, where the p-value should be above 0.05. Furthermore, in the current study, histograms, normal Q-Q plots, and Box plots were used to observe the normal distribution of data. These tests showed that not all data scores were distributed normally. Since the sample of this study was less than 300, skewness and kurtosis were performed for the dataset, and a significant deviation from normality was observed (see Table 5.4). Therefore, the curve skewed to the left, which means the data were negatively skewed. In such a case, variables are often logarithmically transformed to achieve a normal or near-normal distribution (Broothaerts et al., 2020). Hence, the data set has been transformed through log10 to shrink the numerical range of data to make the data set close to normal distribution. Such transformation made the data pattern more interpretable and helped to meet the assumptions of inferential statistics. Table 5.4 illustrates the normality test outcomes for variables.

Table 5.4
Univariate Normality Test Results for Variables

Items	Mean	S.D.	Skewness		Kurtosis	
			Stat.	S. E	Stat.	S. E
Effective use BDA tools						
Collect data from external sources and various customer relationship management systems.	.272	.221	.053	.203	-1.17	.403
Identify essential business insights and trends to improve the innovation process.	.276	.209	-.088	.203	-1.07	.403
Analyse data in near real or real-time that allows responses to unexpected market threats.	.318	.223	-.231	.203	-1.10	.403
Provide systematic and comprehensive reporting to help recognise available opportunities for product/service improvement.	.271	.206	-.039	.203	-1.01	.403
Provide near real or real-time reporting for the innovation process.	.282	.217	-.080	.203	-1.19	.403
BD management						
Our BDA staff has a suitable education to fulfil their jobs.	.265	.215	.081	.203	-1.07	.403
Our BDA staff holds suitable work experience to accomplish their jobs successfully.	.268	.210	.002	.203	-1.06	.403
Our BDA staff is well trained.	.280	.213	-.001	.203	-1.00	.403
Our BDA managers can anticipate the future business needs of functional managers, suppliers, and customers.	.270	.214	.059	.203	-1.02	.403
Our BDA managers can understand and evaluate the output extracted from BD.	.262	.220	.166	.203	-1.03	.403
Innovation process						
The updated-ness or novelty of technology used in the process.	.270	.200	-.133	.203	-1.06	.403
The number of changes in process introduced.	.269	.214	-.023	.203	-1.20	.403
Pioneer disposition to introduce new process.	.300	.197	-.311	.203	-0.84	.403
Clever response to new processes introduced by other companies in the same sector.	.240	.210	.163	.203	-1.21	.403
Competitive advantage						
Our advantages are optimised in the firm and not in individuals; no one can copy us by stealing our staffs away from us.	.280	.211	-.027	.203	-1.00	.403
No one can copy our business procedures, methods, and culture.	.281	.219	.015	.203	-1.09	.403
Using big data analytics reduced the cost of our product/service.	.300	.205	-.236	.203	-0.93	.403
Using big data analytics increase delivery dependability.	.280	.198	-.065	.203	-0.73	.403
Firm performance (financial performance)						

Items	Mean	S.D.	Skewness		Kurtosis	
			Stat.	S. E	Stat.	S. E
During the last 3 years, using BDA has improved our customer retention relative to competitors.	.276	.238	.137	.203	-1.29	.403
During the last 3 years, using BDA has improved our sales growth relative to competitors.	.301	.219	-.083	.203	-1.04	.403
During the last 3 years, using BDA has improved our return on investment relative to competitors.	.315	.224	-.134	.203	-1.05	.403
Overall, during the last 3 years, using BDA has improved our financial performance relative to competitors.	.293	.229	-.003	.203	-1.17	.403
Overall environment turbulence						
In our kind of business, customers' product/service performance changes a lot over time.	.276	.217	.051	.203	-1.05	.403
The product/service technology is changing rapidly in our industry.	.268	.215	.065	.203	-1.11	.403

Note: S.D: Standard deviation; Stat: Statistic; S.E: Standard error; VIF: Variance inflation factor

5.3.4: Assessment of Homoscedasticity and Linearity

According to Hair et al. (2006), among the fundamental assumptions that must be fulfilled to achieve high-quality results of regression analysis are homoscedasticity and linearity (p. 186). Homoscedasticity occurs when "dependent variable(s) exhibit different equal levels of variance across the range of predictor variable(s)" Hair et al. (2010, p. 74)Hair et al. (2010, p. 74). Linearity refers to the degree to which the change in the dependent variable is associated with the independent variable (Tabachnick et al., 2007, pp. 162-163). Therefore, Levene's Test for Equality of Variances was conducted to ensure that the study has no significant deviations from homoscedasticity that can affect multivariate analyses. The test assesses the null hypothesis, which assumes that the variances of the populations from which different samples are drawn are equal (called homogeneity of variance or homoscedasticity). Hence, when the result of Levene's test is non-significant ($p > .05$), it can be assumed that variables are equal, and homogeneity of variances is present. Table 5.5 illustrates the analysis of homoscedasticity.

Table 5.5
Analysis of Homoscedasticity

Education vs Variables		
Variables	Levene Statistic	Sig.
Effective data analysis	1.364	0.256
Effective data management	0.748	0.525
Innovation process	2.090	0.104
Competitive advantage	2.151	0.097
Firm performance	1.123	0.342
Position vs Variables		
Variables	Levene Statistic	Sig.
Effective data analysis	0.662	0.619
Effective data management	0.667	0.616
Innovation process	0.074	0.988
Competitive advantage	1.874	0.118
Firm performance	0.575	0.681
Time Position vs Variables		
Variables	Levene Statistic	Sig.
Effective data analysis	0.327	0.896
Effective data management	1.127	0.349
Innovation process	1.329	0.256
Competitive advantage	0.483	0.788
Firm performance	0.925	0.467

5.3.5 Assessment of Multicollinearity

Examining the assumption of multicollinearity is considered a primary concern regarding the measurements used in structural equation modelling. The feature of multicollinearity is high values of the multiple correlation coefficient (Senawi et al., 2017), which refers to a state where the predictor variables themselves have 0.90 and above correlation (Tabachnick et al., 2007). The multicollinearity might lead to a substantial impact that can undermine the independent variables statistical significance (Hair, 2009). Therefore, the current study evaluated the degree of multicollinearity and determined its influence on the outcomes by applying the Tolerance and Variance Inflation Factor (VIF) (Hair et al., 2010). According to Pallant (2003), if a typical cut-off threshold value for VIF is 10 or even as low as 4 and the tolerance value is more than .10, the data set does not contain multicollinearity problems. The analysis in the current study illustrated that the VIF values ranged from 2.519 - 4.615, and tolerance values ranged from .217 - .397, which are within the acceptable range; therefore, it can be confirmed that the independent variables are free of collinearity problems.

5.3.6 Non-response Bias

Non-response bias refers to the difference between respondents' answers and the responses from non-respondents to the investigation questionnaire. It is a common problem in large-scale questionnaire studies (Mikalef, Boura, et al., 2020). One popular approach to tackling non-response is to analyse the demographics of the study respondents with demographics of the second wave of the target group. It is recommended to measure non-response bias by comparing the last quarter of participants to the rest of the sample (Armstrong & Overton, 1977). In the current study, paired sample t-tests through SPSS were employed to determine whether a significant difference existed. The sample of respondents was divided into early group (first phase) samples and late group (second phase) samples based on two waves of data collection because the first wave data were collected based on the FAME database and the second wave data were collected through the Qualtrics. The results of the tests found no significant difference at 95% confidence for any attitudinal items. Table 5.6 presents the results of the t-test.

Table 5.6

T-Test for Non-response Bias

Items	Groups	No	Mean	Std. Dev	*t*-Statistics	Sig.(2-tailed)
Collecting data from external sources and from various customer relationship management systems.	Early	63	.229	.189	-2.68	.320
	Late	80	.318	.209		
Providing systematic and comprehensive reporting to help recognise feasible opportunities for product/service improvement.	Early	63	.235	.184	-1.94	.891
	Late	80	.304	.214		
Providing near-real or real-time reporting for the innovation process.	Early	63	.251	.198	-149	.710
	Late	80	.309	.224		
Identifying important business insights and trends to improve the innovation process.	Early	63	.211	.193	-2.90	.379
	Late	80	.326	.225		
We integrate data from multiple internal sources into a data warehouse or mart for easy access.	Early	63	.305	.212	-.790	.629
	Late	80	.337	.225		
Our BDA staff has suitable education to fulfil their jobs.	Early	63	.204	.199	-3.57	.082
	Late	80	.318	.208		
Our BDA staff is well trained.	Early	63	.252	.186	-1.37	.986
	Late	80	.303	.226		
Our BDA managers can understand and evaluate the output extracted from BD.	Early	63	.222	.209	-1.95	.865
	Late	80	.296	.221		
Our BDA managers hold suitable work experience to accomplish their jobs successfully.	Early	63	.241	.203	-1.43	.254
	Late	80	.289	.208		
Our BDA managers are able to coordinate BD -related activities in ways that support other teams of innovation process, functional managers.	Early	63	.203	.200	-3.53	.896
	Late	80	.328	.199		
The number of changes in process introduced.	Early	63	.209	.198	-3.10	.553
	Late	80	.324	.205		
The updated-ness or novelty of technology used in the process.	Early	63	.235	.184	-2.12	.544
	Late	80	.306	.204		
Clever response to new process introduced by the companies in the same sector.	Early	63	.157	.189	-4.64	.267
	Late	80	.305	.198		
Pioneering disposition to introduce new process.	Early	63	.244	.190	-3.45	.825
	Late	80	.357	.173		

Items	Groups	No	Mean	Std. Dev	*t*-Statistics	Sig.(2-tailed)
Our company became unique in its products/services, and nobody can offer similar produce.	Early	63	.246	.201	-1.78	.589
	Late	80	.308	.208		
Our advantages are optimised in the firm and not in the individuals; no one can copy us by stealing our staff away from us.	Early	63	.278	.214	-1.55	.208
	Late	80	.329	.185		
No one can copy our business procedures, methods, and culture.	Early	63	.246	.183	-1.86	.547
	Late	80	.311	.197		
No one can copy our brand name easily.	Early	63	.247	.211	-1.84	.125
	Late	80	.310	.217		
Using BDA improved – during the last 3 years our customer retention.	Early	63	.243	.230	-1.49	.718
	Late	80	.308	.245		
Using BDA improved – during the last 3 years our sales growth.	Early	63	.296	.203	-.433	.790
	Late	80	.312	.232		
Using BDA improved – during the last 3 years our return on investment.	Early	63	.300	.205	-.850	.341
	Late	80	.331	.236		
Overall financial performance improved by using BDA.	Early	63	.258	.234	-1.80	.718
	Late	80	.331	.236		

5.4: Multivariate Analysis

To examine the proposed hypotheses, the current study performed multivariate analyses in three main phases: exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and structure equation modelling (SEM). EFA allows the researcher to explore the main dimensions to generate a theory or model from a relatively large set of latent constructs often represented by a set of items (Hair et al., 2010; Pallant, 2010, p. 90; Pett et al., 2003). According to scholars such as Brown (2006), it is the most extensively used statistical method that can be conducted to identify patterns in a set of variables. While it is a technique to identify the structure of underlying variables, it also summarises and reduces the dataset to a more manageable size which referring as much of the original information as possible (Hair et al., 2010). Thus, the current study used EFA through data summarisation and reduction to structure associated items into a single factor. In addition, CFA was used as a more precise method to estimate the factor structure of a set of observed variables to evaluate the hypothesised relationship between observed variables and their underlying latent constructs (Anderson & Gerbing, 1988; Pallant, 2010, p. 181; Tabachnick & Fidell, 2013). Moreover, the study used SEM to assess various interconnected dependence relationships and denote unobserved concepts in these relationships (Hair et al., 2010, p. 585). The following sections will explain the various stages of multivariate analysis and their findings.

5.4.1: Exploratory Factor Analysis

Exploratory factor analysis (EFA) was conducted individually for each of the variables considered in this investigation. This was done using Principal Component Factor Analysis with Orthogonal (Varimax) rotation to reduce the number of variables and excretion the underlying formation in the relationships between variables. The main goal of performing the principal component factor analysis is to obtain the essential information from the dataset, to draw a new set of orthogonal variables named principal components, and discover the pattern of similar variables, as it is considered one of the acceptable methods for extracting information (Hair et al., 2010; Norris & Lecavalier, 2010). In addition, Varimax rotation was performed to obtain a clear separation of the variables so that the resultant components are orthogonal to each other (Hair et al., 2010; Matsunaga, 2010; Tabachnick et al., 2007).

Furthermore, to assess the factorability of the data set, two statistical measures were used: Bartlett’s test of sphericity and the Kaiser- Meyer-Olkin (KMO) measure of sampling adequacy

(Kaiser & Rice, 1974). According to Pallant (2003), accepting factorability of the KMO index should be between 0 to 1, with 0.60 or higher suggested as the benchmark of satisfactory factor analysis. Also, Bartlett's Test of Sphericity describes the overall significance of whole correlations within a correlation matrix. In addition, three other conditions were considered to determine whether to retain the variables. First, the percentage of the total variance is recommended to be 60% or less for social science research (Hair et al., 2010). Second, factor loading was used to outline how decisively a factor explains a variable. The range of loadings can be from -1 to1. Whenever that loading is close to 1, regardless of negative or positive, shows the strong influence on the variable and vice versa, the loadings closer to (0) indicate that the factors weakly influence a variable. Finally, based on the sample size, factor loadings should be considered significant, and sample size less than 200 should consider (-30 to +30) as the benchmark loading value (Hair et al., 2010, p. 112; Williams et al., 2010). Given the sample size of N=143, the current research uses high (-60 to +60) loading values to have a strong measurement of the factors, which can adequately account for the variables (Hair et al., 2010, p. 115). The following sections will address the results of factors for each construct in detail.

5.4.1.1Exploratory Factor Analysis of Effective Use of Big Data Analytics Tools

The EFA for effective use of BDA tools extracted a one- factor solution, which accounts for 42% of the total variance. In addition, principal component factor analysis PCA shows retention of one-factor solution. Therefore, the current study retains the effective use of BDA tools as one factor. Bartlett's test of sphericity illustrates a statistically significant (sig = 0.00, $p < 0.05$) approximate chi-square of 929.469 with *df* 15. Furthermore, the KMO measure of sampling adequacy shows a value of .905, higher than the cut-off point of 0.60 as suggested by Pallant (2010, p. 183).

Out of 10 items, four were excluded due to low factor loadings and. The factor loadings for accepted items for effective data analysis tools range from .601 to .843. According to Ho (2006, p. 240) when the α value is more than .80, it indicates high reliability and internal consistency items. Also, Hair et al. (2006, p. 137) stated that even if the α value is more than .70, it indicates high reliability. Nevertheless, Cronbach's alpha result of effective use of BDA tools construct refers to a firm consistency and reliability (α = .924). Furthermore, each item's reliabilities (Cronbach's alpha) ranged from .907 to .921. Moreover, Corrected item-total Correlation (CITC) was employed to evaluate the internal consistency of the items by exposing the degree of correlation among each item and a scale score that excludes that item (Nunnally, 1994). The low correlation for an item shows that the item is not evaluating the same object that the rest of the scale is attempting to measure. The CITC between items ranges from .665 to .836 values in this research. These values exceed the CITC cut-off of 0.30 recommended by Pallant (2003, p. 100). Finally, Table 5.7 summarises the factor analysis, reliability analysis, and descriptive statistics for effective use of BDA tools.

Table 5.7
Statistical Summary – Descriptive Statistics, Factor Analysis, and Reliability Analysis of Effective Use of BDA Tools

Factors and Variables	**Descriptive Statistics**		**FL**	**Reliability**	
	M	**S.D.**	**1**	**CITC**	**α**
Effective use of BDA tools					**.924**
Our firm uses collecting data from external sources and from various customer relationship management systems tools effectively	.276	.208	.843	.770	.912
Our firm is providing systemic and comprehensive reporting tools to help recognize feasible opportunities for product/service improvement effectively	.271	.206	.839	.836	.907
Our firm is providing near-real or real-time reporting for the innovation process	.282	.217	.802	.793	.910
Our firm use analysing social media data tools to understand current trends from a large population.	.293	.239	.685	.665	.921
Our firm is identifying important business insights and trends to improve the innovation process effectively	.272	.221	.613	.759	.913
Our firm is analysing data in near-real or real-time that allows responses to unexpected market threats.	.318	.222	.601	.738	.915
% of Cumulative variance			**42.0%**		
Note: KMO = .905, Bartlett test of sphericity = 929.469 with df 15, significance = 0.000, FL = factor loadings					

5.4.1.2 Exploratory Factor Analysis of Big Data Management

Exploratory factor analysis for BD management revealed a two-factor solution based on the Eigenvalue rule and Parallel Analysis (PA), representing the first factor 37.62% and the second factor 35.33% of the total variance. In addition, the internal reliability of the factors also supports the two-factor solution for this EFA. Therefore, this study keeps two-factor solutions for the effective data management dimension: technical skills and managerial skills, satisfying the essential criteria (-30 to +30) as the benchmark loading value. Bartlett's test of sphericity illustrates a statistically significant (sig = 0.00, $p < 0.05$) approximate chi-square of 1140.857 with df 36. Furthermore, the KMO measure of sampling adequacy shows a value of 0.938, higher than the cut-off point of 0.60 as suggested by (Pallant, 2010, p. 183).

Three items out of 12 measured variables of BD management were excluded due to low factor loadings and cross-loadings. The factor loadings for accepted items within technical and managerial skills ranged from .622 to .856 and .614 to.709. The Cronbach's alpha result of two factors indicates strong consistency and reliability of BD management .952. Furthermore, the item's reliability (Cronbach's alpha) ranged from .943 - .950, exceeding the cut-off of point 0.60. Moreover, Corrected item-total Correlation (CITC) was employed to evaluate the internal consistency of the items by exposing the degree of correlation among each item and a scale score that excludes that item (Nunnally, 1994). In this research, the CITC between items ranges for technical skills from .730 - 855 and managerial skills from .779 - .845. These values exceed the CITC threshold of 0.30 recommended by Pallant (2003, p. 100). Table 5.8 displays the factor analysis, reliability analysis, and descriptive statistics for Bd management.

Table 5.8
Statistical Summary-Descriptive Statistics, Factor Analysis, and Reliability Analysis of BD Management

Factors and Variables	**Descriptive Statistics**		**FL**		**Reliability**	
	M	**S.D.**	**1**	**2**	**CITC**	**α**
Technical Skills						**.952**
Our BDA managers have a good sense of where and when to apply BD.	.273	.217	.856		.828	.945
Our BDA managers can coordinate BD-related activities in ways that support other teams of innovation process, functional managers.	.248	.221	.762		.855	.943
Our BDA staff has the right skills to accomplish their jobs successfully.	.277	.226	.760		.819	.945
We provide BDA training to our own employees.	.281	.250	.739		.753	.949
Our BDA managers understand and appreciate the business needs of other teams of innovation process, functional managers, suppliers, and customers.	.264	.222	.708		.843	.944
We hire new employees that already have the BDA skills.	.353	.222	.622		.730	.950
Managerial Skills						
Our BDA managers can understand and evaluate the output extracted from BD.	.261	.220		.709	.845	.944
Our BDA managers can work with innovation process team, functional managers, suppliers, and customers to determine opportunities that big data might bring to our business.	.279	.213		.627	.817	.945
Our BDA managers can anticipate the future business needs of innovation process's team, functional managers, suppliers, and customers.	.269	.214		.614	.779	.947
% of Cumulative variance			**37.62%**	**35.33%**		
Note: KMO = 0.938, Bartlett test of sphericity = 1140.857 with df 36, significance = 0.000, FL = factor loadings						

5.4.1.3 Exploratory factor Analysis of Innovation Process

The EFA for effective use of innovation process extracted a one- factor solution, which accounts for 34.3% of the total variance. In addition, principal component factor analysis PCA shows retention of one-factor solution. Therefore, this study keeps one-factor solutions for the innovation process, which satisfies the essential criteria (-30 to +30) as the benchmark loading value. Bartlett's test of sphericity illustrates a statistically significant (sig = 0.00, $p < 0.05$) approximate chi-square of 768.088 with *df* 28. Furthermore, the KMO measure of sampling adequacy shows a value of .932, higher than the cut-off point of 0.60, as suggested by Pallant (2010, p. 183).

Two items out of eight measured variables of innovation process were excluded due to low factor loadings. The factor loadings for accepted items within innovation process ranged from .638 - .875. The Cronbach's alpha result of one factor indicate strong consistency and reliability innovation process from (.920). Furthermore, the item's reliability (Cronbach's alpha) ranged from .943 - .950, exceeding the cut-off of point 0.60. Moreover, Corrected item-total Correlation (CITC) was employed to evaluate the internal consistency of the items by exposing the degree of correlation among each item and a scale score that excludes that item (Nunnally, 1994). In this research, the CITC between items ranges technological innovation from .741 - .817. These values exceed the CITC threshold of 0.30. Table 5.9 presents the innovation process's factor analysis, reliability analysis, and descriptive statistics.

Table 5.9
Statistical Summary – Descriptive Statistics, Factor Analysis, and Reliability Analysis of Innovation Process

Factors and Variables	**Descriptive Statistics**		**F L**	**Reliability**	
	M	**S.D.**	**1**	**CITC**	**α**
Innovation Process					**.920**
Our innovation process leads the number of changes in process introduced	.269	.213	.875	.817	.913
Our innovation process leads to increasing in our technology competitiveness.	.230	.199	.815	.754	.919
Our innovation process leads the updated-ness or novelty of technology used in the process	.270	.200	.783	.803	.914
Our innovation process leads clever response to new process introduced by the companies in the same sector.	.236	.210	.753	.768	.918
Our innovation process leads increasing the rate of change in process, technique, and technology	.265	.206	.690	.809	.914
Our innovation process leads speeding the adoption of the latest technology	.243	.213	.638	.741	.920
% of Cumulative variance			**34.27**		
KMO = 0.932, *Bartlett test of sphericity = 768.088 with df 28, significance = 0.000, FL = factor loadings*					

5.4.1.4 Exploratory factor Analysis of Competitive Advantage

The EFA for effective use of competitive advantage extracted a one-factor solution, which accounts for 39.78% of the total variance. In addition, the internal reliability of the factors also supports the one-factor solution for this EFA. Therefore, this study keeps one-factor solutions for the competitive advantage, which satisfies the essential criteria (-30 to +30) as the benchmark loading value. Bartlett's test of sphericity illustrates a statistically significant (sig = 0.000, $p < 0.05$) approximate chi-square of 641.353 with df 28. Furthermore, the KMO measure of sampling adequacy shows a value of .933, higher than the cut-off point of 0.60 as suggested by Pallant (2010, p. 183).

Five items out of 10 measured variables of competitive advantage were excluded due to low factor loadings. The factor loadings for accepted items within competitive advantage ranged from .661 to .812. The Cronbach's alpha result of one-factor indicate strong consistency and reliability (.898). Furthermore, the item's reliability (Cronbach's alpha) ranged from .871 to .894, exceeding the cut-off of point .60. Moreover, Corrected item-total Correlation (CITC) was employed to evaluate the internal consistency of the items by exposing the degree of correlation among each item and a scale score that excludes that item (Nunnally, 1994). In this research, the CITC between items ranges for competitive advantage from .733 to .774 values. These values exceed the CITC threshold of 0.30. Table 5.10 presents the competitive advantage's factor analysis, reliability analysis, and descriptive statistics.

Table 5.10
Statistical Summary – Descriptive Statistics, Factor Analysis, and Reliability Analysis of Competitive Advantage

Factors and Variables	**Descriptive Statistics**		**F L**	**Reliability**	
	M	**S.D.**	**1**	**CITC**	**α**
Competitive Advantage					**.933**
Our company become unique in its products/services, and nobody can offer similar produce.	.279	.211	.812	.733	.878
No one can copy our business procedures, methods, and culture.	.277	.198	.749	.738	.878
Our advantages are optimised in the firm and not in individuals - no one can copy us by stealing our staffs away from us.	.300	.205	.744	.777	.871
Reduce the cost of our product/service.	.261	.219	.716	.632	.894
No one can copy our brand name easily.	.280	.219	.661	.774	.872
% of Cumulative variance			**39.78**		
Note: KMO = 0.933, Bartlett test of sphericity = 641.353 with df 28, significance = 0.000, FL = factor loadings					

5.4.1.5 Exploratory factor Analysis of Financial Performance

The EFA for financial performance extracted a one-factor solution, which accounts for 39.97% of the total variance. In addition, the internal reliability of the factors also supports the one-factor solution for this EFA. Therefore, this study provides one-factor solutions for the financial performance satisfying the essential criteria (-30 to +30) as the benchmark loading value. Bartlett's test of sphericity illustrates a statistically significant (sig = 0.00, $p < 0.05$) approximate chi-square of 782.208 with df 28. Furthermore, the KMO measure of sampling adequacy shows a value of .904, higher than the cut-off point of 0.60, as suggested by Pallant (2010, p. 183).

Four items out of eight measured financial performance variables were excluded due to low factor loadings. The factor loadings for accepted items within financial performance ranged from .754 - .874. The Cronbach's alpha result of one-factor indicates strong consistency and reliability financial performance (.904). Furthermore, the item's reliability (Cronbach's alpha) ranged from .871 - .894, exceeding the cut-off of point 0.60. Moreover, CITC was employed to evaluate the internal consistency of the items by exposing the degree of correlation among each item and a scale score that excludes that item (Nunnally, 1994). In this research, the CITC between items ranges for financial performance from .780 to .809 values. These values exceed the CITC threshold of 0.30. Table 5.11 presents the financial performance factor analysis, reliability analysis, and descriptive statistics.

Table 5.11
Statistical Summary – Descriptive Statistics, Factor Analysis, and Reliability Analysis of Financial Performance

Factors and Variables	**Descriptive Statistics**		**F L**	**Reliability**	
	M	**S.D.**	**1**	**CITC**	**α**
Financial Performance					**.904**
Using BDA improved our return on investment during the last 3 years relative to competitors.	.315	.223	.874	.780	.900
Using BDA improved our financial performance during the last 3 years relative to competitors.	.292	.229	.791	.822	.891
Using BDA improved our sales growth during the last 3 years relative to competitors.	.300	.219	.767	.797	.896
Using BDA improved our customer retention during the last 3 years relative to competitors.	.275	.238	.754	.809	.894
% of Cumulative variance			**39.97**		
Note: KMO = 0.904, Bartlett test of sphericity = 782.208 with df 28, significance = 0.000, FL = factor loadings					

5.5: Structural Equation Modelling (SEM)

As mentioned in Chapter Four, SEM is used to analyse multiple and interrelated relationships between the constructs for model specification and development (Hair et al., 2010; Tabachnick et al., 2007). Therefore, after data are cleaned, screened, and achieving a better understanding, the next step is conducting statistical analysis through the SEM, allowing analysis of interrelated questions in a single analysis by simultaneously modelling relationships among independent and dependent variables. Applying SEM in the current study includes confirmatory factor analysis (CFA), regression, and path models. CFA is used to examine the one-dimensionality of research constructs by assessing the measurement model, and path models are implemented to assess the association between the proposed variables. The following sections present the outcomes of both the CFA and path analysis.

5.5.1 Confirmatory Factor Analysis

After the exploratory factor analysis (EFA), the next step is CFA, a tool that allows the researcher to confirm or refuse the preconceived theory (Hair et al., 2010). The CFA is a practice in conceptual parsimony as its initial purpose is to determine and outline latent constructs and statistically test its fit (Everitt, 2006; Russell, 2002). The CFA has been suggested as an accurate process to examine the validity and one-dimensionality of measurements and adjust the theoretical models (Anderson & Gerbing, 1988; Bentler, 1983; Jöreskog, 1978). In terms of assessing the adequacy of the measurement models in CFA, the current study employed a single index from each measure category, chi-square/df, Root-mean-square (RMR), root-mean-square error of approximation (RMSEA) and standardized root mean square residual (SRMR) for absolute fit measures, comparative fit index (CFI), Tucker Lewis index (TLI), and incremental fit indices (IFI) for incremental fit measures, and parsimony goodness of fit index (PGFI), and parsimony normed fit index (PNFI) for parsimony fit measures. The chi-square statistic supplies the most conventional fit index; it is called the "exact fit index"; therefore, to test the selected model, this test (x2/df) value less than 5, with a non-significant difference was employed (Bentler & Bonett, 1980; Hair et al., 2010; Hu & Bentler, 1999). The other model fit test, such as RMR, is considered to be less than 0.05 (Bagozzi & Yi, 1988; Hu & Bentler, 1999). The RMSEA is smaller than 0.08 (Schumacker & Lomax, 2010), and less than 0.08 for SRMR (Hu & Bentler, 1999). In addition, it is suggested that the values for additional model fit measures such as FI, TLI, and IFI, considered to be higher than 0.90 (Kline, 2005). Parsimony-adjusted fit criteria (PNFI and PCFI) higher than

0.5 indicate a good fit (Byrne, 2010; Tabachnick & Fidell, 2013). Table 5.12 presents the results of model fit indices that achieve the required thresholds for these measurements.

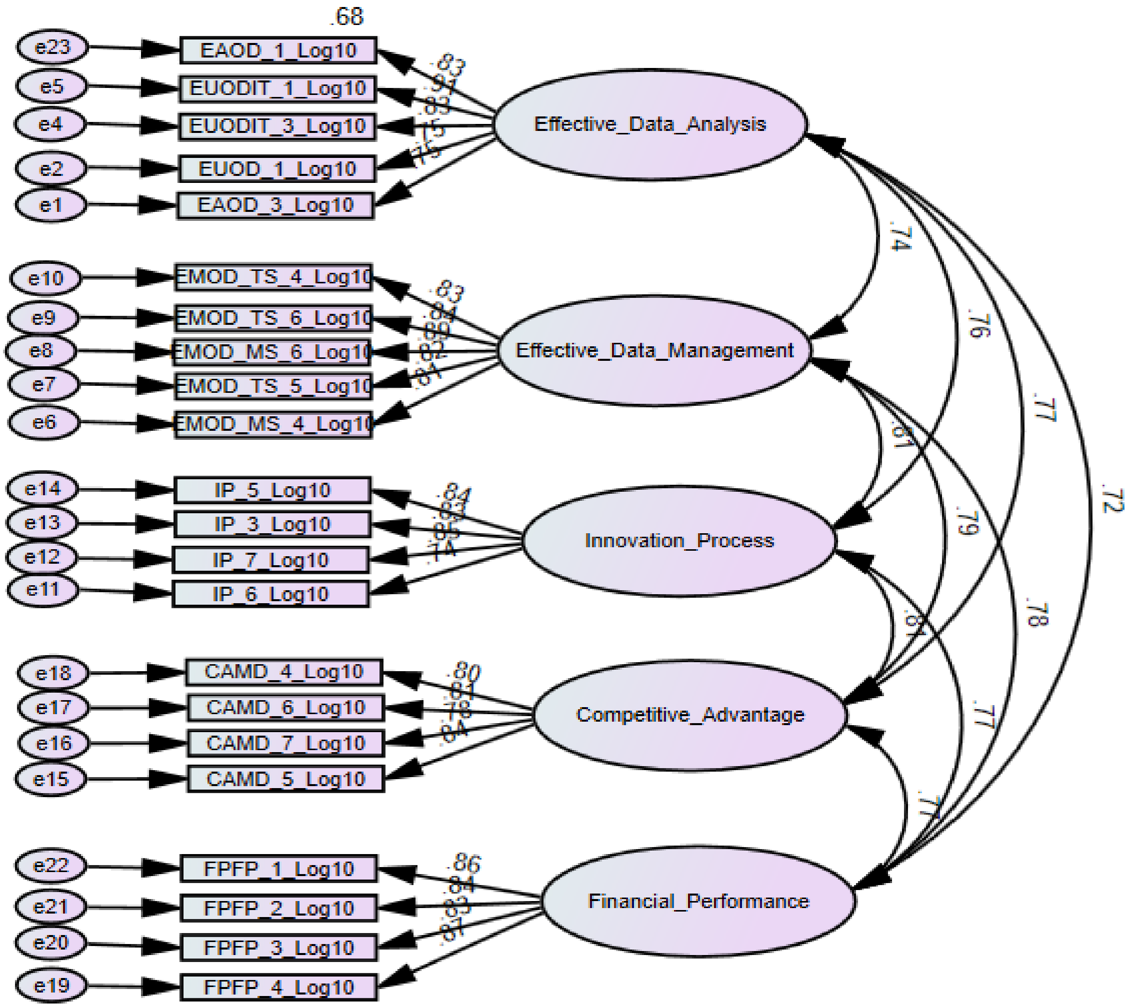
Table 5.12
Summary of Model Fit Indices

Fit Index	**Indices**	**Obtained Fit Indices**	**Suggested Fit Indices**	**Model Fit**
Absolute Fit Indices	X^2/df	1.82	≤5; *p*>.05	Excellent
	RMR	.002	≤.05	Excellent
	SRMR	.044	<0.08	Excellent
	RMSEA	.076	≤.08, fit well ≤.05 fit very well	Good Fit
Incremental Fit Indices	CFI	.937	≥.90	Excellent
	TLI	.924	≥.90	Excellent
	IFI	.938	≥.90	Excellent
Parsimony Fit Indices	PCFI	.807	>.50	Excellent
	PNFI	.750	>.50	Excellent

Note: RMR = root-mean-square residual; SRMR = standardized root mean square residual; RMSEA = root-mean-square error of approximation; CFI = comparative fit Index; TLI = Tucker Lewis index; IFI = incremental fit indices

As shown in Table 5.12, the results of the indices in absolute, incremental fit, and parsimony fit reached all the requirements for a good fit model. Regarding the threshold values, the outcomes present that the absolute fit measures are x2/df = 1.82, p < 0.05; RMR = .002, SRMR = .044 and RMSEA = .076 all meet the requirements. Furthermore, incremental fit measures also represent good model fit, where CFI = .937, TLI = .924, and IFI = .938, all the indices are higher than .90. Finally, parsimony fit measures also validate the goodness of fit for the model as PNFI = .750 and PCFI = .807 are greater than the cut-off value of 0.50.

Figure 5.1
Confirmatory Factor Analysis (CFA) Results for the Research Model



5.5.1.2 Discriminant Validity for Measurement Model

According to Anderson & Gerbing (1988) recommendation, the discriminant validity should be measured by calculating the inter-correlation between a pair of latent variables, which should be smaller than the square root of Average Variance Extracted (AVE) for each variable. Table 5.13 shows that all the factors of the measurement model support discriminant validity, and the square roots of AVE are greater than whole correlations among all the combinations and between any pair of those factors. Further, discriminant validity analysis was conducted to confirm that none of the correlation coefficients exceeds the cut-off value of 0.70. Furthermore, composite reliabilities (CRs), like Cronbach's alpha, were found greater than the cut-off of 0.70 as suggested by Hair et al. (2012) for each variable. Finally, the smallest value of MaxR(H) is 0.885, which means all values are more significant than the threshold of 0.70, confirming the discriminant validity of the measurement model. These results prove that all the requirements have been reached to validate the discriminant validity of the measurement model.

Table 5.13
Discriminant Validity Assessment for Measurement Model

Construct	CR	AVE	MSV	MaxR(H)	EUBDATs	BDM	IP	CA	FP
Effective Use of Big Data Analysis Tools (EUBDATs)	0.907	0.663	0.595	0.920	**0.814**				
Big Data Management (BDM)	0.922	0.702	0.662	0.926	0.736	**0.838**			
Innovation Process (IP)	0.889	0.668	0.662	0.894	0.764	0.814	**0.817**		
Competitive Advantage (CA)	0.883	0.654	0.649	0.885	0.771	0.788	0.806	**0.809**	
Financial Performance (FP)	0.911	0.718	0.607	0.912	0.723	0.779	0.773	0.772	**0.847**

Note: CR = Composite (construct) Reliability; AVE = Average Variance Extracted; MSV = Maximum Shared Variance; ASV = Average Shared Variance; EUBDATs= Effective Use of Big Data Analysis Tools; BDM= Big Data Management; IP= Innovation Process; CA= Competitive Advantage; FP= Financial Performance

5.5.1.3 Common Method Bias

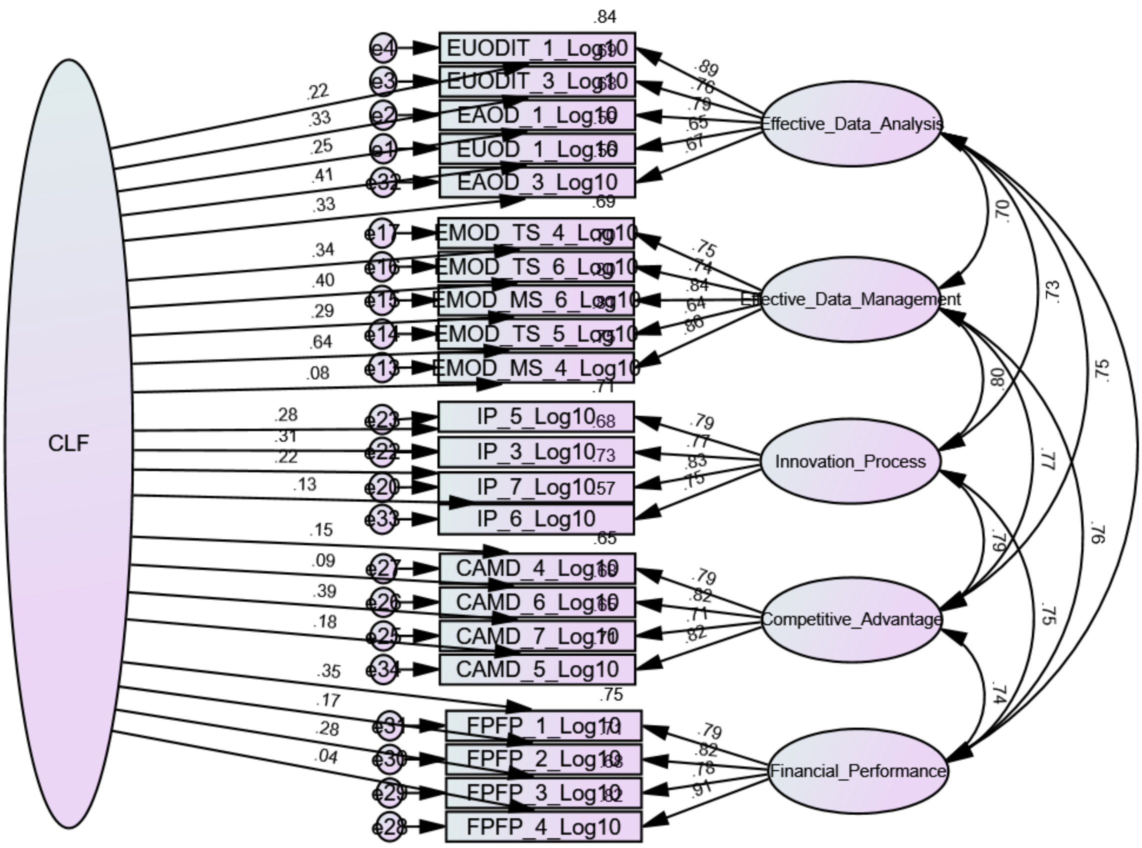
Common method variance (CMV) refers to the amount of spurious covariance shared among variables because of the common method used in collecting data (Podsakoff et al., 2003). In investigations based on surveys where all data are self-reported and gathered through a questionnaire during a period with a cross-sectional research design, data are likely susceptible to CMV (Lindell & Whitney, 2001). The CMV may lead to systematic measurement error and bias the estimates of the actual relationship among theoretical constructs. Scholars (e.g., Antonakis et al., 2010; Burton-Jones, 2009; Podsakoff et al., 2003) highlighted several potential causes, such as social desirability, knowledge deficiency, or consistency motif, for the CMV. Hence, scholars such as Chang et al. (2010); Malhotra et al. (2006); Podsakoff et al. (2003) recommended different methods named ex-ante and ex-post to avoid or correct the effects of CMV. According to the ex-ante approach, researchers should use different information resources to create dependent and independent variables in the design phase of the questionnaire to avoid or reduce CMV (Podsakoff et al., 2003). In addition, researchers should be applying different methodical preparations in designing and managing the questionnaire, such as ensuring the anonymity and confidentiality of the questionnaire to using various scales (Podsakoff et al., 2003). In contrast, researchers can apply the ex-post approaches once the research has been conducted. The ex-post refers to various statistical techniques that can assess and control the issue of CMB, such as the Harman one-factor test, common latent factor test, marker variable test (Burton-Jones, 2009; Chang et al., 2010; Malhotra et al., 2006; Podsakoff et al., 2003). Harman's single factor test was used to assess the effect of CMV through the EFA model. The results are shown in Table 5.14; the cumulative variance extracted was 57.19% which is more than 50%, indicating data may be suffering from CMB (Eichhorn, 2014).

Table 5.14
Harman Common Method Variance Analysis

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulate %	Total	% of Variance	Cumulate %
1	12.581	57.188	57.188	12.581	57.188	57.188
2	1.360	6.180	63.368			
3	1.095	4.979	68.347			

Hence, to mitigate the effect of CMB, the current study conducted the unmeasured common latent factor as a statistical remedy (Podsakoff et al., 2003; Richardson et al., 2009). If common method variance is largely responsible for the relationship among the variables, the one-factor model fit should be better than the second model where items are loaded according to their theoretical constructs (Chang et al., 2010; Podsakoff et al., 2003). The variances and loadings were positive and none of the loadings exceeded 1.0. Furthermore, as suggested by Cheung and Rensvold (2002) and Malhotra et al. (2013), the chi-square difference test was used to compare both models. The Chi-square difference test = 361.9 – 302.3 = 59.6 (Δdf = 199 - 177 = 22) was found to be significant at p-value = 0.05 (p = 0.000).

Figure 5.2
Common Latent Factor (CLF)



5.5.2 Hypothesis Testing (Path Analysis)

After demonstrating the fitness of the satisfactory measurement model, the next step is testing the structural model. Each path in the structural model between independent and dependent constructs, as shown in Table 5.15, presents a particular research hypothesis. The approval or rejection of a hypothesis depends on the direction and significance of the standardised coefficient *(β),* which indicates the percentage and nature of the association between endogenous and exogenous variables. The greater the values are, the higher is the joint explanatory power of the exogenous variables. Hypotheses are generally tested in the form of null hypotheses H0, where no statistical relationship exists among the examined variables on the significance level (p-value). In other words, H0 is either accepted or rejected based on the level of the p-value of the standardised coefficient of a research parameter. Three various levels of significance are employed in this research, 0.05 = acceptable significance, 0.00 = strong significance, and 0.001 = high significance. Table 5.15 presents the hypotheses findings according to associate sequences proposed in the conceptual research model. As shown in Table 5.15, most hypothesis results strongly supported the theoretical research model. More precisely, seven hypothetical relationships in the model were significantly supported, two of the hypotheses were found acceptable significant, and four were rejected. The following section will address the hypotheses results in detail. The analysis of path-coefficients (see Figure 5.5) revealed that effective use of BDA tools (β = .317, p = 0.000), BD management (β = .490, p = 0.000) were both significant predictors of the innovation process. In addition, the innovation process was found to be positive and significant on competitive advantage (β = .485, p = 0.000) and financial performance (β = .333, p = .012). Moreover, effective use of BDA tools and BD management were both significant positive predictors of competitive advantage with (β = .322, p = 0.000), (β = 0.298, p = 0.000) respectively. Therefore, our findings confirm that using BDA tools and BD management leads firms to innovate their products, processes, and solutions (Niebel et al., 2019). Our findings align with the work of Ramadan et al. (2020), who found that BDACs have a significant effect on innovation capability and that innovation capability enhances sustainable competitive advantage.

Furthermore, competitive advantage was found to positively influence financial performance (β = 0.397, p = 0.000). Also, the research outcomes provide the positive impacts of effective use of big data analysis tools and big data management on financial performance. These outcomes provide those managers interested in BDA with tools that can be used to analyse the impact of BDA solutions on firm performance (Raguseo & Vitari, 2018). Surprisingly, regarding the environment turbulence as a moderator, the statistical results indicate no relationships between environment turbulence and financial performance (β = -.002, p = .771). Similarly, the research outcomes indicate no significant effect of control variables firm size (β = -.003, p = .611), firm age (β = -.008, p = .233), and industry types (β = -.001 p = .791) on financial performance. In order to be surer, the relationships of moderators (environmental turbulence, firm size, firm age, and type of industry) have been tested on the model's variables (innovation process and competitive advantage), and no significant results were found. Therefore, it can be confirmed that moderators' factors have no significant impact on the whole model.

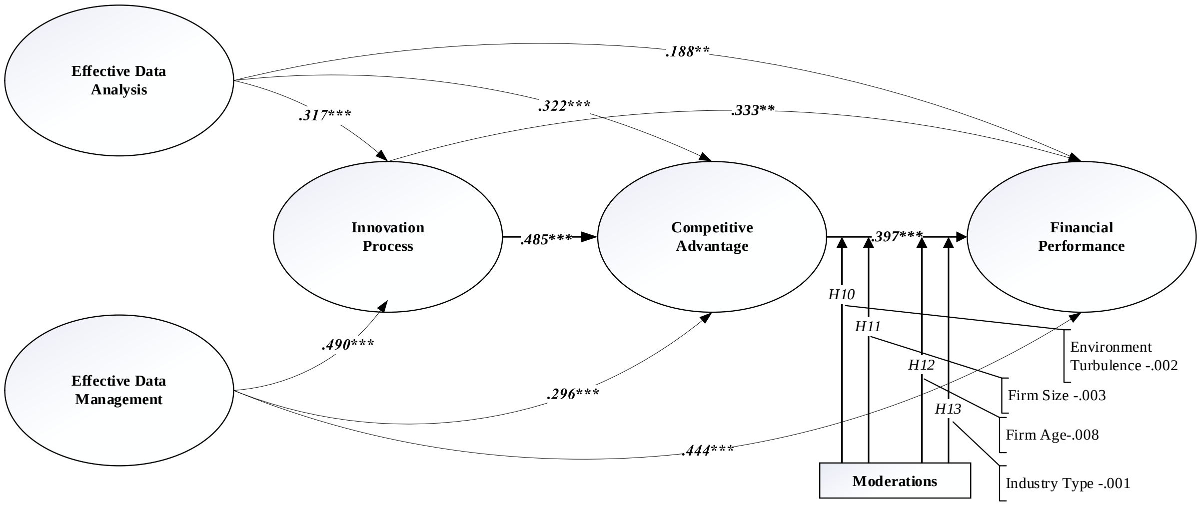
Table 5.15

Summary of Path Model Results and Hypothesis Testing

H	Path (relationship)			Standardised Regression Weight *(β)*	S. E	Critical ratio (CR)	*p*-value	Results
	Independent Variable	→	**Dependent Variable**					
H1	Effective Use of BDA Tools	→	Innovation Process	.317	.052	6.040	***	Supported
H2	BD Management	→	Innovation Process	.490	.050	9.761	***	Supported
H3	Innovation Process	→	Competitive Advantage	.485	.102	4.747	***	Supported
H4	Innovation Process	→	Financial Performance	.333	.132	2.511	.012	Supported
H5	Effective Use of BDA Tools	→	Competitive Advantage	.322	.072	4.489	***	Supported
H6	BD Management	→	Competitive Advantage	.298	.079	3.767	***	Supported
H7	Competitive Advantage	→	Financial Performance	.397	.096	4.147	***	Supported
H8	Effective Use of BDA Tools	→	Financial Performance	.188	.091	2.058	.040	Supported
H9	BD Management	→	Financial Performance	.444	.092	4.843	***	Supported
H10	Environment Turbulence	→	Financial Performance	-.002	.007	-.291	.771	Not Support
H11	Firm Size	→	Financial Performance	-.003	.006	-.508	.611	Not Support
H12	Firm Age	→	Financial Performance	-.008	.006	-1.193	.233	Not Support
H13	Industry Type	→	Financial Performance	-.001	.003	-.265	.791	Not Support

Figure 5.3

Validation of Research Framework



5.6 Chapter Summary

This chapter presented the results of data preparation stages, descriptive statistics, exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and structural equation modelling (SEM). First, this chapter presented the results of several tests and assessments such as (value missing, outliers, normality, homoscedasticity and linearity, multicollinearity, and Non-response bias) to ensure that data were ready for further statistical analysis. Then, it reported the results of exploratory factor analysis (EFA) for all structures. The EFA outcomes set reliable structures for effective use of BDA tools, BD management, innovation process, competitive advantage, and financial performance. Then, the CFA was used to evaluate and validate the research instrument. The CFA results proved that the measurements were suitable and reliable for the targeted sample population. After that, this chapter presented the SEM results, which confirmed the validation and fitness of the proposed research framework. Finally, the research hypotheses were tested. The hypotheses results will be discussed in more detail in the following chapter

Chapter Six: Discussion of the results

6.1 Introduction

As discussed in the first chapter, the mechanisms, and conditions of using BDA to enhance the innovation process to achieve a competitive advantage and financial performance remain underexplored in existing research. To address this gap, the current study developed a conceptional model and hypotheses on two main core aspects of BD: effective use of BDA tools and BD management. Based on the analysis of the previous chapter, this chapter interprets and discusses the research findings and attempts to provide relevant justification related to the study outcomes. Hence, along with the discussion, this chapter also provides support to the findings with the prior studies within the literature. The chapter aims to inspect and explain different hypothesized relationships between effective use of BDA tools, BD management, innovation process, competitive advantage, and financial performance.

6.2 Effective Use of BDA Tools

In the current study, effective BDA tools has been defined as a process comprising several phases, including data aggregation, data analysis, data interpretation, and data applications. Meanwhile, the innovation process is implementing a new or significantly improved production or delivery method, which includes significant changes in techniques, equipment, and/or software (Henao-García et al., 2021). In addition, BDA has raised expectations of being particularly beneficial for the firm's innovation process. Therefore, the major contribution of effectively BDA lies in the fact that it enables firms to direct their investment and deployments accordingly and focus on areas of higher benefit (Mikalef & Krogstie, 2020), increase their operational abilities such as marketing and technological (Niebel et al., 2019), and assist well-informed decision-making (Abbasi et al., 2016), and higher innovation capabilities (Niebel et al., 2019). Hence, in the proposed model of this research, the impact of effective use of BDA is examined on three factors: innovation process, competitive advantage, and financial performance.

6.2.1 Effective use of BDA tools and innovation process (H1)

BDA can change the innovation landscape by efficiently and effectively bridging consumer needs and products or services features (Johnson et al., 2017). However, a firm simultaneously must employ customer insights gathered from BD to redefine its marketing activities to implement innovation continuously (Erevelles et al., 2016; Story et al., 2011). Today, the R&D departments of giant companies like Rolls Royce, Caterpillar, Aerospace, and

Tesla use BD collected from sensors to plan their following product improvements and discover what components and features require to be developed or created (Blackburn et al., 2017). Therefore, H1has been formulated to investigate the impact of the effective use of BDA tools on the innovation process. The current research result supports H1, where the influence of effective use of BDA tools on the innovation process was positively significant $\beta = 0.317$, $P < 0.001$. The finding is consistent with previous research findings. For example, Mikalef, Boura, et al. (2020) found similar results, demonstrating that BDACs can significantly enhance firms' incremental and radical innovate capability. In addition, Niebel et al. (2019) stated that BDA is the core of the innovation process and can generate new innovative products and services.

Furthermore, Mikalef and Krogstie (2020) indicated that investing in BD and technological infrastructure to analyse and visualise data insight can direct a firm to strengthen capacity to create incremental innovation processes in the industry and radical innovation process in services. The current study's findings are also aligned with recent research conducted by Saleem et al. (2020), who found that BDA is positively and significantly related to SMEs' innovation process. However, Ghasemaghaei & Calic (2020) analysed the three main BD characteristics (i.e., data variety, velocity, and volume) on innovation performance, where data variety and velocity were shown to improve firm innovation performance, whereas data volume do not show any significant relationship. This means the amount of data as data characteristics may not be significant. The current study similarly suggests that data from external sources and various customer relationship management (variety), near real or real-time reporting (velocity) can identify essential business insights and trends to improve the innovation process. Hence, due to its high strategic and operational potential, BDA is considered to be a game changer, empowering business to improve efficiency and performance(Germann et al., 2014; Hooi et al., 2018).

6.2.2 Effective Use of BDA Tools and Competitive Advantage (H5)

A firm's competitive advantage is built on a set of strategic resources. Hence, the logic is that the firm's BDACs are valuable resources that may generate competitive advantages. However, recent literature (e.g., Côrte-Real et al. (2020)) highlighted that there is a shortage of theoretical-driven studies on how to use BD tools to gain a competitive advantage. Therefore, H5 was formulated to examine the impact of the effective use of BDA tools on competitive advantage. The outcome supports the claim that effective data analysis can enhance the firm's competitive advantage. The statistical analysis found a significant relationship between competitive advantage and effective data analysis tools (β = 0.322, P < 0.001). The current study's result is consistent with the literature. For instance, Côrte-Real et al. (2017) found that dynamic capability of a firm (e.g., organizational agility) is leveraged by BDA that positively affects the creation of competitive advantages. Côrte-Real et al. (2020) demonstrated that effective usage of BDA leads to achieving a competitive advantage. The current study's study also provide evidence that understanding current trends and insights through near-real or real-time data analysis can offer competitive advantage. A similar notion has been presented by Waller & Fawcett (2013), who stated that insights obtained from BDA can develop real-time business process monitoring and analysis and improve quality control. Moreover, Kiron et al. (2014) found that adopting BDA could play a crucial role in supporting operational and strategic decision-making, thus enhancing firm performance.

However, some scholars such as Dezdar and Ainin (2011); Hong and Kim (2002) stated that transforming the BDA' advantages into the firm's value is challenging. The challenges start from the superior level, where top-level management is required to embrace evidence-based decision making and encourage all firm staff to redefine their understanding of the BDA context (McAfee et al., 2012). In addition, readiness factors and firms' infrastructure are essential factors for BDA adoptions, so, developing IT infrastructures capacities may not necessarily improve firm performance (Hong & Kim, 2002). Using BDA capabilities and organizational factors (in terms of top management support, BDA strategy, financial resources, and people engagement) can both frustrate and facilitate the effective use of BDA. Hence, the advantages of BDA use do not translate into business benefits quickly and easily (Popovič et al., 2014).

However, different studies (e.g., Horng et al. (2022); Suoniemi et al. (2020)) provided the evidence that companies investing more resources in the development of BD, can enhance marketing capabilities which ultimately provide edge to competitive advantage and in turn leads to better firm performance. For instance, BD-driven customer insight has led Amazon.com to achieve the greatest competitive advantage in the market through minimization of costs and time spent on pricing, distribution, and communications with automated decision support engines (Day, 2011; Suoniemi et al., 2020). The current study's findings similarly suggest that firms that acquire appropriate BDA capabilities can indeed achieve competitive advantages. Hence, the underlining point is that the managers working with differentiated products and services, regardless of the industry, should consider BDA as a key component in building more efficient, effective, and innovative competitive advantage over their rivals.

6.2.3 Effective Use of BDA Tools and Financial Performance (H8)

BDA is broadly acknowledged to play a critical role in enhancing firm performance (Akter et al., 2016). Furthermore, the BD literature confirms a positive relationship between BDA and firm performance in improving prices and maximising profits (Davenport, 2006), sales growth, profitability, market share (Manyika et al., 2011), and return on investment (McAfee et al., 2012). In the current study, the effective use of BDA refers to activities such as data gathering, analysing, and interpreting effectively through suitable methods and software. Hence, H8 has been raised to test the effective use of BDA tools on financial performance. Findings show that effective use of big data analytics tools significantly positively impacts financial performance (β = 0.188, P < 0.001). Similar results have been documented in BD literature. A positive relationship of BDACs implementation was found not just on firm financial performance but also on other firms' performance elements such as operations and market performance. For instance, Singh and El-Kassar (2019) found that BD assimilation through acceptance and routinization routes enhances firms' sustainable performance. The current study also provides evidence that implementation of effective use of BDA tools will resulted into increase customer retention, sales growth, and return on investment. Furthermore, using BD improves operations performance, and firm performance makes a much more significant market share (Gupta et al., 2020). Similarly, Gupta and George (2016) found a significant, positive effect of BDA capability on market performance and operation performance. The same findings were made by Raguseo and Vitari (2018); Rahman and Zhao (2020); (Ren et al., 2017b); Wamba et al. (2018). While the current study's result is identical to previous studies, it is distinguished by focusing on the effective use of BDA tools as an independent factor and financial performance as a dependent variable. Therefore, according to the current literature findings and the current study's findings, the conclusion is that the application of BDA increases all forms of firm performance. Hence, these results may reassure decision-makers that investing in BD technology will pay off by boosting sales, customer retention, and return on investment.

6.3 Big Data Management

BD literature has paid little attention to big data's organizational and managerial aspects, emphasizing the technical aspects (Henao-García et al., 2021; Mikalef & Krogstie, 2018). Therefore, BD management and its conditions have recently become a debate topic in the management field (Caputo et al., 2020), where the importance of this issue has led to the emerging a new field of management named "BD management." BD management functions include implementing methods, complex architectures and sophisticated techniques and platforms, including storage, pre-processing, and processing (Manogaran et al., 2020). Furthermore, concerning the innovation process, BD management refers to understanding the factors that facilitate and inhibit the development of innovations. Hence, this research has proposed three hypotheses to investigate the effectiveness of BD management for the innovation process (Hypothesis 2), competitive advantage (Hypothesis 6), and financial performance (Hypothesis 9). The following sections will discuss the results of these three hypotheses.

6.3.1 BD Management and Innovation Process (H2)

Data management is at threat of missing the BD train (Labrinidis & Jagadish, 2012). One of the main challenges of adopting BDA in firms is BD management, which is believed to be a precondition of BDA(Qi, 2020). BD management needs to deal with BD characteristics (Vs), analysis processes, and even social concerns (Almeida & Calistru, 2013). In addition, there is a concern that managing and handling BD from multiple resources contaminate the analysis precision (Manogaran et al., 2020). Moreover, innovation through BD derives from skills, competencies, and knowledge of human resources through R&D and technology assessment files (Caputo et al., 2020; Hekkert et al., 2007). Therefore, understanding the impact of BDA on enabling innovation capabilities allows management to direct their investment and centre on the most significant essential areas. For example, employing professional staff can increase the abilities of BDA. Hence, to investigate the contributions of BD management, H2 was formulated to examine the impact of BD management on the innovation process. The current study found that BD management has a strong significant positive impact on the innovation process ($\beta = 0.490$, $P < 0.001$). Therefore, the current study can confirm that the adoption and practice of BD management in a firm will enhance innovation process capabilities. The research finding aligns with the current literature. For instance, Mardani et al. (2018) found that knowledge management dimensions (the production, integration, and application of knowledge) in a BD environment significantly improve innovation. In contrast, the current research has focused on BD management as a component of BDA skills (managerial and technical skills). In comparison, Caputo et al. (2020) identified that acquisition and exploitation link innovation and information management. Hence, the current findings provide evidence that BDA managers, through working with the innovation team, functional managers, suppliers, and customers, can determine opportunities that enhance the innovation process capabilities. Similarly, Waqas et al. (2021) found that BDA capability management positively impacts green innovation. Therefore, the current study's findings add a new contribution to the literature regarding the role of BD management on innovation. These findings will enhance the previous studies on the importance of management, particularly BD management. Now, the firm can learn about the positive of BD management on the innovation process. Specifically, those firms that are looking to invest in data analytics technology. In addition, the current study's findings also highlight the importance of BD management in an organisational chart, which might help the decision-makers create a new managerial position to deal with BD.

6.3.2 BD Management and Competitive Advantage (H6)

Suggestions from empirical literature about BD used for competitive advantage stress the abilities and understanding of the management regarding the application of BD LaValle et al. (2011); McAfee et al. (2012). It was emphasised that firms could achieve a sustainable competitive advantage through BD management (Akter et al., 2016). However, firms will not reap the full benefits of using BD unless they can manage change effectively in five areas (leadership, talent management, technology, decision making, and firm culture) (McAfee et al., 2012). BDA is crucial in advanced operations management, because it requires BDA techniques, identifying their strengths and weaknesses, and applying different BD methods such as techniques, strategies, and architectures (Choi et al., 2018). Therefore, H6 was formulated to investigate the influence of BD management on competitive advantage. The current study's results proved the positive effect of BD management on competitive advantage ($\beta = 0.298$, $P < 0.001$), verifying H6. The current study's findings are consistent with results of Waqas et al. (2021), who examined the management of BDACs on competitive advantage. However, they looked at green innovation and green human resources as mediating between BDACs management and competitive advantage, which is different from the current investigation. Similarly, Kwon et al. (2014) found that BD quality management and its experience positively impact the intention to adopt BDA in firms where intention is the primary step of BDA adoption. Accordingly, it can be suggested that firms need to adopt and practice BD management along with the technical side of BDA. The current study's findings also proved that BD management could enhance the competitive advantage capabilities by producing new technologies to reduce the cost of products/services.

Furthermore, it can be concluded that decision-makers should be aware that just investment in BDA technology is not enough. Perhaps the lack of attention to BD management was one of the reasons for not getting the desired results from adopting BDA technology for most of the firms that have already adopted it (White, 2019). Therefore, BD management should be considered an essential factor of BDA strategy.

6.3.3 BD Management and Financial Performance (H9)

The highest priority for organizations is to have higher growth performance concerning their financials (Rajaguru & Matanda, 2013). Yet, the literature provides evidence of a positive relationship between BDA and firm performance (Akter et al., 2016; Anwar et al., 2018; Serrato & Ramirez, 2017; Setia & Patel, 2013; Tambe, 2014). However, the rule of BD management is not discovered (Bello-Orgaz et al., 2016; Chen & Zhang, 2014; Mikalef et al., 2017). Therefore, in the current research, BD management is assumed to be a component of BDA that can play a vital role in enhancing BDADs. Hence, H9 was formulated to examine the impact of BD management on financial performance. The current study's findings support the hypothesis that BD management is positively associated with financial performance. The statistical results indicate a strong confirmation of the positive association between the two variables (β = 0.444, $P < 0.001$), which support H9. The current study's findings align with recent studies indicating that BD management has a positive relationship with financial performance (Anton et al., 2021; Gupta et al., 2020; Yasmin et al., 2020). For instance, Anton et al. (2021) examined BDA management capabilities' impact on firm performance directly and indirect through the operational performance as a mediator. In contrast, the current research examined the influence of BD management on financial performance directly and indirect via the innovation process and competitive advantage. The current study's findings provide evidence that BD management by hiring professional staff, anticipating future business needs, and dealing constructively with functional managers, suppliers, and customers could participate efficiently in improving firm financial performance. Similarly, Yasmin et al. (2020) found that BD management is the second factor (28.7%) that has a positive impact on firm performance after infrastructure with (45.96%). In comparison, the current research outcomes indicate that BD management has the highest (β = 0.444, $P < 0.001$) impact on financial performance compared with the other factors such as (effective use of BDA β = 0.188, $P <$ 0.001, the innovation process β = 0.333, $P < 0.05$, and competitive advantage β = 397, $P <$ 0.000). According to this robust finding, it can be suggested that technology is unusable if it is not accompanied by effective management, which can extract insight and take strategic decisions Gupta et al. (2020). Hence, these results confirm that BD management could play an essential role in enhancing financial performance. Consequently, the decision-makers can be confident regarding their investment's return in BDA.

6.4 Innovation Process

A great deal of literature has dealt with the effects of information and communication technology on the innovation and competitive advantage (e.g., (Gërguri-Rashiti et al., 2017; Higón, 2012; Neirotti & Pesce, 2018; Yunis et al., 2018; Zhang & Wang, 2019). This is because ICT has considerably changed the knowledge generation process of firms (Brynjolfsson & McAfee, 2014). Similarly, BDA is considered advanced information that could be an enabler for innovation. Meanwhile, innovation capability is the fundamental firms' strategic asset that could lead to competitive advantage (Erevelles et al., 2016; Ponta et al., 2020), and competitive advantage can be seen as "being better than others"(Bartosik-Purgat & Ratajczak-Mrożek, 2018). Therefore, the innovative activities depend on the variety and form of their relation to firms' internal resources such as data, knowledge, technologies, practices, staff, and financial resources. Hence, this research has proposed two hypotheses to investigate the influence of the innovation process on competitive advantage (H3) and financial performance (H4). The following sections will discuss the results of these two hypotheses.

6.4.1 Innovation Process and Competitive Advantage (H3)

According to scholars (e.g., Ramadan et al. (2020); Sutapa et al. (2017)), BDACs and innovation capabilities (IC) are of essential importance in enhancing competitive advantage, where IC indicates a firm's ability to introduce and define innovative ideas and deploy them in designing new products/services. Sutapa et al. (2017) demonstrated that innovation is vital in guaranteeing competitive advantage. In addition, scholars such as Erevelles et al. (2016); Sutapa et al. (2017) considered BDA to be a valuable resource that enhances innovation capabilities, which in turn leads to competitive advantage (Côrte-Real et al., 2017; Ghasemaghaei & Calic, 2020). Hence, in the current study, H3 was formulated to examine the impact of the innovation process on competitive advantage. The result shows that the innovation process has a significant positive relationship with competitive advantage (β = 0.485, $P < 0.001$). The current study's findings align with recent studies (e.g., Distanont and Khongmalai (2020); Niebel et al. (2019); Ramadan et al. (2020); Waqas et al. (2021)) indicating that innovation capabilities enhance competitive advantage. For instance, Distanont and Khongmalai (2020) found that market orientation positively influences innovation, which, in turn, has a positive impact on competitive advantage. However, the current study evaluated the innovation process as mediating between the effective use of BDA tools and BD management and competitive advantage. Similarly, Ramadan et al. (2020) found that innovation capabilities and BDACs positively affect sustainable competitive advantage.

Furthermore, Anwar et al. (2018) found that in addition to BDACs, BD personal capabilities, which refer to knowledge of technology, business, and management, also benefit competitive advantage. In contrast, the current study’s findings show that firms can improve their competitiveness by presenting new or improved products before competitors and thus increasing market share (Distanont & Khongmalai, 2020).

The current study provides evidence that usage of BDA can contribute to enhancing the competitive advantage directly and indirect through the innovation process. Therefore, firms with successful competitive advantage enjoy superior performance in turbulence markets (Batista et al., 2016).

6.4.2 Innovation Process and Financial Performance (H4)

Innovation is one of the interdisciplinary phenomena, and its role in a firm's survival has received wide attention from several academic disciplines (Lee et al., 2019; Zahra & Das, 1993). Therefore, much of the research has expanded its scope to include different types of innovation such as process, product, organizational, and marketing innovation and examined when their interrelationship is effective in improving firm performance (Rubera & Kirca, 2012; Rust et al., 2004; Srinivasan et al., 2009; Tellis et al., 2009). Further, according to Rios-Morales and Brennan (2009); Tsai et al. (2013), innovation capabilities can create, support, and maintain competitive advantage and performance. Therefore, the current study focused on the innovation process, where BDA can play a crucial role in enhancing financial performance (Bigliardi, 2013; Saleem et al., 2020). Although innovation has been considered to be a key to surpass competitors in the market, a considerable body of research still emphasizes specific types of innovation activities, such as process, product, and organisational innovation (Lee et al., 2019). Hence, H4 was formulated to investigate the influence of the innovation process on financial performance. The current study’s novel findings revealed that the innovation process significantly positive influences financial performance ($\beta = 0.333$, $P < 0.05$). This study result is consistent with recent studies (Caputo et al., 2020; Saleem et al., 2020; Waqas et al., 2021). For example, Saleem et al. (2020) found that process innovation and product innovation are positively mediating the BDA factors (predictive and prescriptive) and performance in China SMEs. However, the current study’s findings confirm the same results for all levels of firms in the UK and United States. Similarly, Agarwal et al. (2003) noted that adopting innovation strategies would impact the firms' economic performance because it accelerates their ability to perceive business environment changes and customers' expectations; thus, they can adjust their activities. Meanwhile, the current study's findings confirm that adopting the innovation process abilities can improve firms' financial performance by adopting the latest technology, introducing new methods, reducing the cost of products or services.

6.5 Competitive Advantage and Financial Performance (H7)

Since the principal goal of businesses is to obtain a greater level of financial profits, achieving a competitive advantage plays a significant role in reaching this aim (Ma, 2000). According to Ghasemaghaei and Calic (2019), the influence of BDA on firm performance could be mediated by intermediate variables. However, studies on the mediating role of competitive advantage between BDACs, the innovation process, and FP are rare (Anwar et al., 2018). In the current study, competitive advantage is mediating between two components of BDA (effective use of BDA and BD management) as the first level, the innovation process and financial performance. Therefore, H7 was formulated to examine the impact of competitive advantage on financial performance. The result revealed that competitive advantage has a significant positive impact on financial performance ($\beta = 397$, $P < 0.000$), which supported H7. The current study's findings are consistent with recent studies, such as Anwar et al. (2018); Ferreira and Fernandes (2017); Waqas et al. (2021). For instance, Waqas et al. (2021) found that competitive advantage is mediating positively between (green innovation and green human resources management) and environmental performance. However, in the current research, competitive advantage is mediating between the innovation process and financial performance. Furthermore, scholars such as Anwar et al. (2018) and Kuo et al. (2017) also found that competitive advantage has a significant positive relationship with firm performance. They confirm that firms focus on competitive strategy (differentiation-based and cost-based), often referred to as a competitive advantage (Salavou & Sergaki, 2013), to achieve superior performance. Likewise, the current study also provides evidence that a firm can be unique in its products/services and reduce the cost of products or services. Furthermore, firms with successful competitive strategies enjoy the superior performance in dynamic markets (Batista et al., 2016). Finally, the current finding confirms the importance of the innovation process in achieving competitive advantage and financial performance for firms operating in the UK and United States.

6.6 The Mediating Effect (H10, H11, H12, and H13)

The current study examined the impact of control variables (environmental turbulence H10, firm's size H11, age H12, and type of industry H13) on financial performance. The results show that the path coefficient is not significant. These findings agree with those of Akter et al. (2019), who showed the impact of control variables has nonsignificant on firm performance. In addition, the current study's finding is consistent with that of Calic and Ghasemaghaei (2021), who found that control variables such as firm size and industry type in the BD environment do not significantly influence corporate social performance (philanthropic, ethical-legal, and economic performance). Similar results were found by Mikalef, Krogstie, et al. (2020), who examined the influence of control variables on competitive performance and found all cases of control variables relationship with the dependent variable is nonsignificant.

H10 proposed to test the environmental turbulence as a moderating factor on financial performance. Surprisingly, the environmental turbulence did not emerge as a significant moderator factor that could affect the firm's performance; this finding is consistent with that of Vitari and Raguseo (2020), who tested environmental dynamism on financial performance and market performance found the same outcome. In contrast, Prescott (2014) found that environment turbulence due to advances in technology will influence the competitive advantage and make it shorter. In addition, Asad et al. (2021) found that environmental turbulence has a high impact on BDA and the performance of SMEs; however, the environmental turbulence factor fails to hold a significant impact on both the relationships between entrepreneurial orientation and performance of SMEs. Furthermore, Suoniemi et al. (2020) found that environmental turbulence does not directly influence structural model variables (pricing, product development, channel management, selling). However, market turbulence significantly impacts the relationship between market communication and firm performance.

Similarly, the control variables factors hypotheses H11, H12, and H13 (firm size, firm age, and type of industry) were found to have made no significant impact on financial performance. While in the literature some studies provided evidence that firm performance and BDA adoption can be sensitive to firm size and type, other studies suggested no such relationships between the variables. For instance, Raguseo et al. (2020) identified that firm size is a significant moderator of firm profitability when BDA is leveraged on. Additionally, Mikalef et al. (2019a) found that size-class of the firm can significantly contribute to a firm's performance, whereas Dong and Yang (2020) identified that adaptation of BDA is also dependent on different sizes of the firms. McDermott and Prajogo (2012) even found that firm size moderate the effect of innovation on overall performance. Alternatively, like the current study's finding, different studies also did not find any relationships between firm size and performance. For instance, Gupta and George (2016) found that firm size and industry type has non-significant relationship with market and operational performance. Moreover, similar results were found in studies (e.g., Duan et al. (2020); Wamba et al. (2018); Yu et al. (2013)), where they have found that variable such as firm size, firm type, and industry type do not affect firm performance and competitive advantages. One reason for such finding could be that larger firms can leverage on a wider pool of resources than the smaller firms. According to Yu et al. (2019) larger firms have more resources to use BDA capabilities and maintain long-lasting, rare, inimitable, and non-substitutable resources to achieve higher business performance than smaller firms. In the current study sample, more than half of the companies (51%) are by definition SMEs, and so might have fewer resources and capabilities to use the BDA capabilities, reflected in the non-significant impact of firm size on performance relationships. Existing literature has widely defined SMEs as enterprises with no more than 500 employees (e.g., Dong and Yang (2020); Lu and Beamish (2001); Van de Vrande et al. (2009)), hence the current study also used 500 employees as the cut-off to distinguish the firm size in the current study's sample. Furthermore, firms can have multiple levels of bureaucracy and their long decision chains result in slow reaction times (Lin, 2014), which can impede BDA adoption and its related performance. While such conflicting result requires further attention and investigation, the result indicated that impact of BDA capability on firm performance may not necessarily depends on the firm size or type.

Table 6.1
The link Between Research Questions, Research Findings and Research Implications

Research questions	Predictor specification	Findings	Implications
Whether effective use of BDA influence the innovation process of a firm?	Firms can enhance their innovation process capabilities by conducting two principal components (BD management and effective use of BDA tools).	The current study results indicated that BD management has a strong positive impact on the innovation process (β = 0.490, P < 0.001) and also the effective use of BDA tools has a significant positive significant impact on the innovation process (β = 0.317, P < 0.001).	According to findings, firms can enhance their innovation competency. In particular, the findings suggest that firms could discover patterns related to the past and current issues, therefore, accurately predict future trends and events, which leads to creating and implementing new ideas such as successfully developing new products or services.
Does BDA's role impact competitive advantage?	BDA can influence the competitive advantage direct and indirect through the innovation process as a mediator.	Our findings show that BDA, with its two components, has a direct and indirect positive impact on competitive advantage. The results indicate that BD management (β = 0.298, P < 0.001) and effective use of BDA tools (β = 0.322, P < 0.001) have a positive significance on competitive advantage. In addition, BDA positively impacts competitive advantage through the innovation process (β = 0.485, P < 0.001).	Managers can note that the BDA could play a vital role through the effective use of BDA tools and BD management to increase the ability of competitive advantage. In addition, firms can obtain better competitive advantage opportunities through the innovation process as a mediator.
Is the innovation process leading to improved competitive advantage and performance?	The innovation process can impact the competitive advantage and firm performance	According to the current study's findings, the innovation process has a strong significant positive relationship with competitive advantage (β = 0.485, P < 0.001). In addition, it has a significant positive influence on financial performance (β = 0.333, P < 0.05).	These results confirm that the way to be different from competitors and achieve a competitive advantage is to enhance the innovation process. In addition, the innovation process will increase the financial performance due to the sales growth and return on investment.

Research questions	Predictor specification	Findings	Implications
To what extent the situational factors (e.g., environmental turbulence, type of industry, firm age, size) impact the relationship between competitive advantage and firm performance?	Situational factors such as control variables (firm size, firm age, and type of industry) and moderating variables such as environmental turbulence may influence the relationships between competitive advantage and financial performance.	According to our findings, the control variables firm size (β= -.003), firm age (β= -.008), and type of industry (β= -.001) were found to have no significant impact on the relation between the competitive advantage and financial performance. A similar result was found for the environmental turbulence (β= -.002).	These results are unexpected; however, similar outcomes were found by the other studies. Therefore, these findings may change if this model is examined in a specific industry.

6.7 Chapter Summary

This chapter interpreted and discussed the different results of the proposed research model and hypotheses to deepen the current study's understanding of the factors that influence the innovation process, competitive advantage, and financial performance. The discussed outcomes indicated that the two factors of effective use of BDA tools and BD management have a significant positive impact on the innovation process, competitive advantage, and financial performance. Furthermore, this research confirmed that the effective use of BDA tools and BD management are essential components of BDA capabilities for firms willing to adopt and use BDA. The "effective use of BDA tools" component can support firms by collecting data from internal, external, and CRM systems resources to make customer records. These data are used for further analysis, resulting in comprehensive reports and data visualisation in real-time. According to these reports, managers can identify the critical business insights that can increase innovation capabilities.

Furthermore, the current study findings identified that "BD management" is another component of BDA. BD management can enhance the innovation process by working with the innovation team to understand the business needs, functional managers, suppliers, and customers. Also, it is required to provide suitable training for employees, hire new skilled staff, and ensure that BDA teams have the proper education, skills, and experiences to accomplish their jobs. The current study's findings emphasise the importance of BD management and adopting BDA without defining BD management for this proposal may lead to unacceptable results for firms.

Moreover, this study adds theoretical contribution via highlighting the importance of BD management and effective use of BDA tools and practical contribution through the actionable guidelines and workable model for firms seeking to achieve a better competitive position through BDA. The following chapter will discuss these contributions in detail.

Chapter Seven: Conclusions

7.1 Introduction

This final chapter presents a conclusion of this research. It commences through an overview of the thesis and a summary of each chapter (Section 7.2). Section 7.3 offers main contributions to knowledge. Section 7.4 outlines the limitations arising from this research and identifies potential avenues for future research to further develop understanding in this area. Section 7.5 provides concluding remarks.

7.2 Overview of Research

This study began with Chapter One, which included research context, research justification, aim and objectives, and the research questions. The chapter also specified the research motivation, scope, approach, and potential contributions. Next, the chapter briefly presented the new trend of BDA and its importance for businesses, specifically in innovation. Then, the chapter referred to potential opportunities by adopting the BD that firms may obtain and the challenges that they may be facing through the delay in joining this new movement. The BDA literature has suggested that few studies are available that have examined BDA in an innovation context.

Chapter Two reviewed existing literature associated with BD. The literature review concentrated on studies investigating BDA on innovation, competitive advantage, and firm performance. The relevant studies were identified from a keyword-based search in the Scopus database and Google Scholar, and the related articles and books were included in the review. Subsequent searches at later stages assisted in identifying and adding more relevant works referenced until 2021. The first part of Chapter Two highlighted the topics such as BD definitions, characteristics (Vs), classifications, sources of BD, content format, storage, and staging. The second section discussed BDA in terms of value creation, process, tools, and management. The third part looked at sector adaption and adoption of BD, the enablers, and inhibitors, and BD applications. The importance of BD was then discussed, followed by the sector perspective use of BDA such as BD adoption in businesses, government, smart cities, and health care. In addition, the chapter addressed the sector perspectives of challenges in the use of BDA and barriers to the adoption of BDA and innovation in terms of types, processes, rules of BDA in the innovation process, innovation process models. The chapter highlighted the competitive advantage in BD and resources and discussed firms' performance.

Chapter Three moved foremost to fulfil the second objective, highlighting the role of BDA components, such as effective use of BDA tools and its management in the innovation process, competitive advantage, and financial performance, through comprehensive review and propose a conceptual model. Different theories explain the adoption of BDA, innovation, and competitive advantage; therefore, a review has been made of the most popular theories of BD adoption, and the resource-based view and dynamic capability view theories were selected to support this research. In addition, the rationale for choosing these theories and how they can be the theoretical basis to support the adoption of BDA was explained. The proposed model is distinguished from models used in the previous BDA research in terms of the inclusion of effective use of BDA tools, BD management, and innovation process. Chapter Three described and justified the proposed model, followed by formulating corresponding hypotheses. A total of 13 hypotheses were proposed, which were backed by robust theoretical and argumentative discussions for supporting relationships assumed in each hypothesis. Chapter Three overall produced a framework for directing empirical work for this research.

At the next step, it was necessary to determine the most appropriate method framework for measuring and validating the proposed conceptual model, so Chapter Four began with reviewing and evaluating the fitness of the methodological approaches. It briefly explained research philosophy, three paradigm assumptions widely employed in business and management research: ontology, axiology, and epistemology. Since this study is concerned with BDA as a social phenomenon and aimed to test hypotheses and a conceptual model for this purpose, a positivist approach was deemed appropriate to lead this research. Chapter Four also covered selecting a suitable methodology to conduct empirical research. The chapter explained and justified a deductive selection method as the most proper approach to testing the formulated hypotheses. The chapter also explained how the survey questionnaire was designed and validated with the process of data gathering (including sampling frame, sample size, pilot testing). This chapter then gave the details for ethical considerations and how data was made fitting for use in SPSS and AMOS analysis tools. This chapter finally presented and justified statistical data analysis techniques (including using structural equation modelling (SEM) and tools suited for examining the proposed hypotheses and conceptual model.

Chapter Five illustrated the outcomes by examining the proposed hypotheses through the quantitative data. This chapter presented all the findings by using methods discussed in Chapter Four. Findings were generated using data from 174 samples, which is considered to be enough for robustly examining the proposed conceptual model and hypotheses. This chapter conducted necessary steps such as data cleaning, screening, and assessing data outliers before performing the main tests using SPSS, SEM, and AMOS. All tests were matched cut-off values suggested in the literature for gaining satisfactory reliability (using Cronbach composite, average variance extracted, maximum shared variance) and validity (using model fit indices from measurement model) in results. The relationships proposed using 13 hypotheses were examined applying the structural model, and the validity of that was evaluated using model fit indices. Structural model testing confirmed that data supported nine proposed relationships and presented a good fit with the theory.

Finally, Chapter Six provided a detailed discussion of the findings regarding validating the hypotheses and the conceptual model compared with alongside similar outcomes presented in previous studies. It was also highlighted that the proposed model is accepted and supported by existing studies on BDA. In addition, this chapter demonstrated how the proposed model and through the components of BDA (effective use of BDA tools and BD management) could enhance the innovation process ability, gain a competitive advantage, and increase the financial performance of firms, where these findings are considered a unique contribution in BDA studies. The chapter provided a detailed discussion of theoretical contributions and practical implications.

7.3 Research Contributions

Despite the growing interest in understanding the phenomenon of BDA and its impact on the different business environments of the firm (e.g., Abbasi et al. (2016); Chen et al. (2015); Constantiou and Kallinikos (2015); Müller et al. (2018); Raguseo and Vitari (2018)), there has been little empirical research into the potential effects of BDA on the innovation process, competitive advantage, and financial performance. Moreover, most scholarly studies regarding BDA (Beath et al., 2012; Chen et al., 2012; Demirkan & Delen, 2013; McAfee et al., 2012; Wamba et al., 2015) have focused on the technical side or system development rather than the managerial or strategic prospects. Therefore, there is a great potential to extend BD investigations to other research areas (Kwon et al., 2014; Wang & Hajli, 2017). The importance of this research's contribution is to provide empirical evidence verifying the existence of a direct and indirect and positive association between two complements (effective use of BDA tools and BD management) of BDA capabilities and innovation process, competitive advantage, and financial performance. From this perspective, this study contributes to broadening the understanding of how BDA impacts the innovation process, competitive advantage, and financial performance. This study extends BD research in several ways that will be discussed in theoretical and practical fields in the following sections.

7.3.1 Theoretical Contributions

From a theoretical point of view, the current study contributes to existing knowledge in several ways. A unique theoretical contribution of this research is its conceptualization model to draw business value from BDA in the context of innovation. The current study uses RBT and dynamic capability view (DCV) as the base to develop a conceptual research model and then validate it empirically in the context of innovation. Given the lack of models to explore the influence of BDA components (effective use of BDA tools and BD management) on enhancing innovation process, competitive advantage, and financial performance, the current research model underpins that BDA is a technical IT source that, can support firms in processing raw data into useful information to spot ideas that can enhance the innovation process, which in turn, improve competitive advantage and financial performance. The innovation process is supposed to be costly and risky (Wiklund & Shepherd, 2005); hence, efficient implementation of BDA reduces risk and supports the right decisions.

In addition, the currents study's model is different from the other models in several respects. For example, the model of BDA-enabled business value by Wang and Hajli (2017), focused on the health care industry while the current study's model can be applied in different sectors as its data were collected from different industries; thus, the current study's model is more generalizable. In addition, the previous model examined the impact of BDA on benefit dimensions (e.g., organisational, operational IT infrastructure, managerial and strategic benefits). In contrast, the current study's model investigated the impact of BDA on the innovation process, competitive advantage, and financial performance (e.g., Anwar et al. (2018); Ferraris et al. (2019); Ren et al. (2017b); Wang and Hajli (2017)) that introduce BDA benefits in some sectors. No other study has examined the BDA components on innovation, competitive advantage, and financial performance. Hence, the current study model can enhance the literature by identifying the (effective use of BDA tools and BD management) as components that increase the innovation process capabilities, reinforcing the competitive advantage and financial performance. However, considering business varieties and conditions, there is still room for improvement. Questions remain on how firms can create strategic flexibility and use analytics to draw strategic insights for innovation (George, 2017). For example, more specific variables can be added to the model tested in different sectors. Thus, more accurate results could be obtained, and sector differentiation could be precisely analysed

Further, these findings would support managers to understand and recognise the elements that empower BDA capability in their constant quest for competitive advantage over their competitors. Thus, it can help firms adopt BDA to obtain new ideas that could distinguish them from their competitors (Johnson et al., 2017). This research built upon the resource-based view and dynamic capabilities theory to stress the role of BDA capability as a valuable source. Thus, by answering the research questions and treating BD as valuable assets working side-by-side with other firms' capabilities, firms can enhance their ability of the innovation process, competitive advantage, and financial performance; as a result, theoretically, the current study provides significant contributions evidence-backed for the position of effective use of BDA tools and BD management.

Moreover, as the analysis result implies, a firm's BDA capability has a direct positive significant impact on innovation process practices and a direct positive significant influence on its competitive advantage and financial performance. Therefore, firms implementing BDA are more innovative than others that do not. This outcome is consistent with Mikalef et al. (2019b). Hence, these findings add new knowledge for BDA by offering compelling empirical proof on how the firm's BDA capability impacts and enhances these areas.

Fourth, this study extended the BDA literature by examining the innovation process as mediating the impact of effective use of BDA tools and BD management on competitive advantage. To the researcher's knowledge, there is no previous study that explores this mediation mechanism to predict competitive advantage and financial performance. In this way, the current study contributed to BD, competitive advantage, and financial performance literature by identifying the technological innovation process as an essential mediator that connects BDA (effective use of BDA tools and BD management), competitive advantage and financial performance. In addition, it highlighted the impact of controls (firm age, firm size, and industry type) and mediated factors (environmental turbulence) on financial performance.

7.3.2 Practical Contributions

From a practical perspective, the implementation of BDA among practitioners is still in its initial phase (Maroufkhani et al., 2019); therefore, the findings of this research can provide several insights to managers. First, the approved conceptual model has several practical implications for managers engaged in implementing BDA because it shows the path of relationship between factors and how the effective use of BDA tools and effective data management is directly linked to the innovation process, competitive advantage, and financial performance. The overarching implication to practitioners is that this model offers an advanced, unique, insightful, and essential framework and highlights the role of BD management and the impact of effective use of BDA tools.

Second, because the data were that supported this study's model was collected from different industries such as agriculture, automotive, energy, high tech, and communication, financial and insurance, health care, real estate, and service; therefore, this model is suitable to be applied in most sectors. However, the results may differ if the focus is on a particular industry. Therefore, model should be examined in different industries, where more accurate results could be achieved.

Third, by highlighting the importance and impact of effective use of BDA tools, this study has attempted to enlighten managers that firms need to have access to advanced technology (e.g., Hadoop, Apache Spark, Tableau) and data analysis experts to discover unseen relationships due to data collection, analysis, and interpreting as well as reporting data effectively, where insights can be extracted and accurately evaluated. More specifically, this study provides insights into the influence of effective use of BDA tools as an essential factor in enhancing innovation process abilities, competitive advantage, and financial performance.

Fourth, the current study explored the importance of BD management as a critical factor in dealing with BDA, as this study confirmed that BD management positively impacts the innovation process, competitive advantage, and financial performance. Similarly, Niebel et al. (2019) found that the relation between a firm's use of big data and the likelihood of the firm innovating is not contingent on general human capital; it is contingent on firms' investment in IT-specific knowledge and skills. This evidence can support decision-makers to restructure their organisation and create a new position for BD management. The new position needs both managerial and technical capabilities that help firms to collect, exchange, and analyse information on a real-time basis and make meaningful decisions needed to help the innovation process and competitive advantage (Wamba et al., 2017).

Fifth, according to the findings of this study, the effective use of BDA and BD management has a significant positive impact on the innovation process. Thus, BDA have the potential to empower innovation (Niebel et al., 2019). BDA provide new insights, which can create novel ideas and thoughts that can be developed across the innovation process and transferred as a new process, product, or service. It is worth to mentioned that analytics in the innovation process may take different forms considering the differentiations in businesses and their objectives. For example, George (2017) developed a simple 2×2 BDA taxonomy using two dimensions (analytics focus and innovation focus) to provide a stylistic model of the role of analytics in innovation. In contrast, our model identified (the effective use of BDA tools and BDA management) as two factors that affect the innovation process. In addition, as BDA has a strong direct and indirect positive impact on obtaining competitive advantage, firms could enhance their competitive position due applying BDA as a strategic approach. Moreover, these findings can help managers who are still hesitant to adopt BDA technology as a strategic approach to take steps forward to adapt BDA.

Sixth, the current study’s findings have another important managerial implication: both constructs of BD management and effective use of BDA tools positively impact financial performance; therefore, investment in BDA technology is assumed to increase the firm’s benefits.

Finally, the study outcomes yield empirical evidence allowing that BDA capabilities positively impact innovation process results and that there is great literature positively relating innovation process with the competitive advantage and firm performance. This means that the management should understand that attempts related to BDA and the innovation process must be considered under an action framework in which organisational processes, mechanisms, applications, and procedures are adjusted.

7.4 Limitation and Directions for Future Research

Even though this study represented a fruitful attempt regarding BDA and identified a meaningful relationship between effective use of BDA tools and BD management, the innovation process, competitive advantage, and financial performance, it suffers from other research several limitations. This research's main limitations are worth mentioning and necessary to consider in further research.

First, the generalisability of these results is subject to certain limitations, where the samples of this study were collected from firms that mainly located in the UK and United States, that were supposed to have the advanced infrastructure in terms of the internet network, data collection organisations, and the BDA skilled staff; therefore, caution must be exercised when generalising the results of this study to firms in developing or non-developing countries. It is recommended that future research in BDA and innovation encompasses firms in emerging markets in different countries.

Second, the current study targeted BDA in the innovation context through the innovation process. However, as discussed in section 2.10 of chapter two, the innovation domains include several types of innovation: product innovation, position innovation, paradigm innovation, organisational innovation, marketing innovation, close innovation, and open innovation. Hence, the current study just focused on the innovation process, and future research could determine the effectiveness of BDA on other aspects of innovation.

Third, this research has only considered two components of big BDA capabilities (effective use of BDA tools and BD management). However, BD literature referred to other components that may affect the innovation process, competitive advantage, and financial performance. Further experimental investigations are needed to investigate the other BDA capabilities components and their effect on innovation, competitive advantage, and financial performance.

Fourth, this research collected samples from different industries to examine the conceptual model's relationships. The results may be changed positively or negatively if data are collected from a specific sector. It is suggested that future research can examine the current study's model in different sectors separately, which may provide specific findings related to each sector. Hence, decision-makers in firms can adopt the outcomes more confidently according to each sector.

Fifth, another limitation lies in the nature of the survey because the research used a perception-based self-reported questionnaire survey, which is naturally linked with the common method variance (CMV) issue. Although the participants were managers and classified as high-level experts with sufficient knowledge and experience to assess cognitive states accurately, the possibility of bias remains, and an opinion-based survey may suffer from a certain degree of bias or inflated responses. Therefore, to treat data and minimise the severity of CMV, the current research applied methodological and procedural care such as Harman's one-factor approach and latent modelling approach to assess the risk of CMV as recommended by Bagozzi (2010); Podsakoff et al. (2003); Richardson et al. (2009). To avoid such problems, Rousseau (1990) advocated using multiple data collection methods to address the problem of common method variance. Future research might consider reproducing these results using a multiple informant approach for expanded generalizability and validation.

Finally, due to limited time and budget, the research is restricted by its cross-sectional design. Consequently, it is impossible to infer causality based on the findings, and instead, the causal relationships have been discussed based on the existing theoretical foundation. Although Casey and Krauss (2013) said that a cross-sectional study is suitable *"in the early stages of theory development to demonstrate relationships between variables"* (p.139), more work are needed to reproduce these findings and identify causality through conducting longitudinal research.

7.5 Concluding Remarks

This chapter presented a summary of the former chapters of this thesis and indicated that objectives set in the first chapter were accomplished. In addition, key findings and contributions to knowledge were outlined, and research limitations were identified in line with potential avenues for future research. To the best of the author's knowledge, this is the first study to explore the impact of BDA in the innovation context. In this way, the study adds valuable empirical results to current BD literature through the creation BDA model to identify factors affecting the innovation process, competitive advantage, and financial performance.

The conclusion of this research emphasizes the role of the effective use of BDA tools and BD management within the innovation process where it has been confirmed that BD can be a nurturing source for the innovation process and create a competitive advantage for firms through the effective use of BDA and BD management.

References

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2), I.

Acharya, B. (2010). Questionnaire design. *Central Department of Population Studies.*

Acklin, C. (2010). Design-Driven Innovation Process Model. *Design Management Journal*, *5*(1), 50-60.

Adrian, C., Abdullah, R., Atan, R., & Jusoh, Y. Y. (2017). Factors influencing to the implementation success of big data analytics: A systematic literature review. 2017 international conference on research and innovation in information systems (ICRIIS),

Aggarwal, A. (2011). Sampling issues in research methodologies. *Indian Journal of Medical Specialties*, *2*(2), 169-172.

Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (2005). Automatic subspace clustering of high dimensional data. *Data Mining and Knowledge Discovery*, *11*(1), 5-33.

Ahuja, G., Lampert, C. M., & Tandon, V. (2008). 1 moving beyond Schumpeter: management research on the determinants of technological innovation. *Academy of Management annals*, *2*(1), 1-98.

Akter, S., Gunasekaran, A., Wamba, S. F., Babu, M. M., & Hani, U. (2020). Reshaping competitive advantages with analytics capabilities in service systems. *Technological Forecasting and Social Change*, *159*, 120180.

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113-131.

Al-Alak, B. A., & Tarabieh, S. A. (2011). Gaining competitive advantage and organizational performance through customer orientation, innovation differentiation and market differentiation. *International Journal of Economics and Management Sciences*, *1*(5), 80-91.

Al-Fares, M., Loukissas, A., & Vahdat, A. (2008). A scalable, commodity data center network architecture. *ACM SIGCOMM computer communication review*, *38*(4), 63-74.

Al-Sai, Z. A., Abdullah, R., & Husin, M. H. (2020). Critical success factors for big data: a systematic literature review. *IEEE access*, *8*, 118940-118956.

Albright, J. J., & Park, H. M. (2009). Confirmatory factor analysis using amos, LISREL, Mplus, SAS/STAT CALIS.

Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*, *60*(3), 285-292.

Ali, Q., Salman, A., Yaacob, H., Zaini, Z., & Abdullah, R. (2020). Does Big Data Analytics Enhance Sustainability and Financial Performance? The Case of ASEAN Banks. *The Journal of Asian Finance, Economics, and Business*, *7*(7), 1-13.

Allwood, C. M. (2012). The distinction between qualitative and quantitative research methods is problematic. *Quality & Quantity*, *46*(5), 1417-1429.

Almeida, F., & Calistru, C. (2013). The main challenges and issues of big data management. *International Journal of Research Studies in Computing*, *2*(1), 11-20.

Almeida, F. L. (2017). Benefits, challenges and tools of big data management. *Journal of Systems Integration*, *8*(4), 12-20.

Ambrosini, V., Bowman, C., & Collier, N. (2009). Dynamic capabilities: An exploration of how firms renew their resource base. *British Journal of Management*, *20*, S9-S24.

Amit, R., & Zott, C. (2012). Creating value through business model innovation. *2012*.

Anderson, J. C., & Gerbing, D. W. (1988). Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological bulletin*, *103*(3), 411-423. https://doi.org/10.1037/0033-2909.103.3.411

Anderson, J. C., & Narus, J. A. (1990). A model of distributor firm and manufacturer firm working partnerships. *Journal of marketing*, *54*(1), 42-58.

Anton, E., Oesterreich, T. D., & Teuteberg, F. (2021). Understanding the Operational Value of Big Data Analytics Capabilities for Firm Performance: A Meta-Analytic Structural Equation Modeling Approach.

Anwar, M., Khan, S. Z., & Shah, S. Z. A. (2018). Big data capabilities and firm's performance: a mediating role of competitive advantage. *Journal of Information & Knowledge Management*, *17*(04), 1850045.

Archenaa, J., & Anita, E. M. (2015). A survey of big data analytics in healthcare and government. *Procedia Computer Science*, *50*, 408-413.

Ardagna, C. A., Ceravolo, P., & Damiani, E. (2016). *Big data analytics as-a-service: Issues and challenges* International conference on big data (big data), Washington DC,USA. https://doi.org/10.1109/BigData.201Ardagna6.7841029

Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of marketing research*, *14*(3), 396-402.

Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. *Transportation Research Part E: Logistics and Transportation Review*, *114*, 416-436.

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79*, 3-15.

Atalay, M., Anafarta, N., & Sarvan, F. (2013). The relationship between innovation and firm performance: An empirical evidence from Turkish automotive supplier industry. *Procedia-Social and Behavioral Sciences*, *75*, 226-235.

Axinn, W. G., & Pearce, L. D. (2006). *Mixed method data collection strategies*. Cambridge University Press.

Baer, M., & Frese, M. (2003). Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *24*(1), 45-68.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational Issues of Big Data and Analytics in Networked Business. *MIS quarterly*, *40*(4).

Bagozzi, R. P. (2010). Structural equation models are modelling tools with many ambiguities: Comments acknowledging the need for caution and humility in their use. *Journal of consumer psychology*, *20*(2), 208-214.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74-94.

Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2019). Big data adoption: State of the art and research challenges. *Information Processing & Management*, *56*(6), 102095.

Balachandran, B. M., & Prasad, S. (2017). Challenges and benefits of deploying big data analytics in the cloud for business intelligence. *Procedia Computer Science*, *112*, 1112-1122.

Ban, T., Eto, M., Guo, S., Inoue, D., Nakao, K., & Huang, R. (2015). A study on association rule mining of darknet big data. 2015 International Joint Conference on Neural Networks (IJCNN),

Banda, J. P. (2003). *Nonsampling errors in surveys* Expert Group Meeting to Review the Draft Handbook on Designing of Household Sample Surveys. United Nations Secretariat, New York, https://mdgs.un.org/unsd/demographic/meetings/egm/sampling_1203/docs/no_7.pdf

Barchiesi, M. A., & Colladon, A. F. (2021). Big data and big values: When companies need to rethink themselves. *Journal of Business Research*, *129*, 714-722.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management*, *17*(1), 99-120.

Barney, J. B. (2000). Firm resources and sustained competitive advantage. In *Economics Meets Sociology in Strategic Management* (pp. 203-227). Emerald Group Publishing Limited.

Barney, J. B. (2001). Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view. *Journal of management*, *27*(6), 643-650.

Barney, J. B., & Hesterly, W. S. (2009). *Strategic management and competitive advantage*. Pearson Education Upper Saddle River, NJ.

Barney, J. B., Ketchen Jr, D. J., & Wright, M. (2011). The future of resource-based theory: revitalization or decline? *Journal of management*, *37*(5), 1299-1315.

Barton, D., & Court, D. (2012). Making advanced analytics work for you. *Harvard Business Review*, *90*(10), 78-83.

Bartosik-Purgat, M., & Ratajczak-Mrożek, M. (2018). Big data analysis as a source of companies' competitive advantage: A review. *Entrepreneurial Business and Economics Review*, *6*(4), 197-215.

Batista, P. C. d. S., Lisboa, J. V. d. O., Augusto, M. G., & Almeida, F. E. B. d. (2016). Effectiveness of business strategies in Brazilian textile industry. *Revista de Administração (São Paulo)*, *51*, 225-239.

Beath, C., Becerra-Fernandez, I., Ross, J., & Short, J. (2012). Finding value in the information explosion. *MIT sloan management review*, *53*(4), 18.

Bell, E., & Bryman, A. (2007). The ethics of management research: an exploratory content analysis. *British Journal of Management*, *18*(1), 63-77.

Bell, E., Bryman, A., & Harley, B. (2018). *Business research methods*. Oxford university press.

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, *28*, 45-59.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika*, *48*(4), 493-517.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, *107*(2), 238.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588.

Bentler, P. M., & Stein, J. (1992). Structural equation models in medical research. *Statistical methods in medical research*, *1*(2), 159-181.

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information polity*, *19*(1, 2), 5-16.

Bessant, J., & Tidd, J. (2007). *Innovation and entrepreneurship*. John Wiley & Sons.

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. v. (2013). Digital business strategy: toward a next generation of insights. *MIS quarterly*, 471-482.

Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: an empirical investigation. *MIS quarterly*, 169-196.

Bhatt, G. D., & Grover, V. (2005). Types of information technology capabilities and their role in competitive advantage: An empirical study. *Journal of management information systems*, *22*(2), 253-277.

Bhattacherjee, A. (2012). Social science research: Principles, methods, and practices.

Bigliardi, B. (2013). The effect of innovation on financial performance: A research study involving SMEs. *Innovation*, *15*(2), 245-255.

Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of big data: four perspectives--four challenges. *ACM Sigmod Record*, *40*(4), 56-60.

Blackburn, M., Alexander, J., Legan, J. D., & Klabjan, D. (2017). Big Data and the Future of R&D Management: The rise of big data and big data analytics will have significant implications for R&D and innovation management in the next decade. *Research-technology management*, *60*(5), 43-51.

Bloch, C. (2007). Assessing recent developments in innovation measurement: the third edition of the Oslo Manual. *Science and Public Policy, 34*(1), 23-34.

Blundell, R., Griffith, R., & Reenen, J. V. (1995). Dynamic count data models of technological innovation. *The Economic Journal*, *105*(429), 333-344.

Bomsel, O., & Le Blanc, G. (2004). Digitalisation, innovation, and industrial organisation: the pivotal case of the auto industry. *International journal of electronic business*, *2*(2), 193-204.

Boubiche, S., Boubiche, D. E., Bilami, A., & Toral-Cruz, H. (2018). Big data challenges and data aggregation strategies in wireless sensor networks. *IEEE access*, *6*, 20558-20571.

Boudreau, M.-C., Gefen, D., & Straub, D. W. (2001). Validation in information systems research: A state-of-the-art assessment. *MIS quarterly*, 1-16.

Boyle, G. J., Borg, M. G., Falzon, J. M., & Baglioni Jr, A. J. (1995). A structural model of the dimensions of teacher stress. *British Journal of Educational Psychology*, *65*(1), 49-67.

Brace, I. (2018). *Questionnaire design: How to plan, structure and write survey material for effective market research*. Kogan Page Publishers.

Bradley, N. (2007). *Marketing research: tools & techniques*. Oxford University Press, USA.

Bradley, N. (2013). *Marketing research : tools & techniques / Nigel Bradley* (Third edition. ed.). Oxford : Oxford University Press.

Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, *70*, 328-337.

Broothaerts, W., Cordeiro, F., Corbisier, P., Robouch, P., & Emons, H. (2020). Log transformation of proficiency testing data on the content of genetically modified organisms in food and feed samples: is it justified? *Analytical and bioanalytical chemistry*, *412*(5), 1129-1136.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. *McKinsey Quarterly*, *4*(1), 24-35.

Brown, T. (2006). Confirmatory Factor Analysis for Applied Research; Kenny, DA, Ed. In: The Guilford Press: New York, NY, USA.

Brownlow, J., Zaki, M., Neely, A., & Urmetzer, F. (2015). Data and analytics-data-driven business models: A blueprint for innovation. *Cambridge Service Alliance*.

Bryman, A. (2016). *Social research methods*. Oxford university press.

Bryman, A., & Bell, E. (2007). Business Research Methods second edition Oxford University Press UK.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Bucur, C. (2015). *Using big data for intelligent businesses* Proceedings of the Scientific Conference AFASES, Document : French.

Bughin, J., Chui, M., & Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. *McKinsey Quarterly*, *56*(1), 75-86.

Buhl, H. U., Röglinger, M., Moser, F., & Heidemann, J. (2013). Big data. In: Springer.

Burrell, G., & Morgan, G. (2017). *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life*. Routledge.

Byrne, B. M. (2010). Structural equation modeling with AMOS: basic concepts, applications, and programming (multivariate applications series). *New York: Taylor & Francis Group*, *396*, 7384.

Byrne, B. M. (2013). *Structural equation modeling with EQS: Basic concepts, applications, and programming*. Routledge.

Cameron, K. S., Kim, M. U., & Whetten, D. A. (1987). Organizational effects of decline and turbulence. *Administrative Science Quarterly*, 222-240.

Candi, M., van den Ende, J., & Gemser, G. (2013). Organizing innovation projects under technological turbulence. *Technovation*, *33*(4-5), 133-141.

Cao, G., & Duan, Y. (2014). *Gaining Competitive Advantage from Analytics through the Mediation of Decision-Making Effectiveness: an Empirical Study of UK manufacturing Companies* 18th Pacific Asia Conference on Information Systems (PACIS). Chengdu, China. http://eli.johogo.com/AIS-JJAW-PACIS2014/21.pdf

Caputo, F., Mazzoleni, A., Pellicelli, A. C., & Muller, J. (2020). Over the mask of innovation management in the world of Big Data. *Journal of Business Research*, *119*, 330-338.

Cascio, R. P. (2011). Marketing innovation and firm performance research model, research hypotheses, and managerial implications.

Castellan, C. M. (2010). Quantitative and qualitative research: A view for clarity. *International journal of education*, *2*(2), 1.

Chang, S.-J., Van Witteloostuijn, A., & Eden, L. (2010). From the editors: Common method variance in international business research. In: Springer.

Chang, V. (2018). A proposed social network analysis platform for big data analytics. *Technological Forecasting and Social Change*, *130*, 57-68. Chae, H. C., Koh, C. E., & Prybutok, V. R. (2014). Information technology capability and firm performance: contradictory findings and their possible causes. *MIS quarterly*, *38*(1), 305-326.

Charles, V., & Gherman, T. (2013). Achieving competitive advantage through big data. Strategic implications. *Middle-East Journal of Scientific Research*, *16*(8), 1069-1074.

Chau, P. Y. (1997). Reexamining a model for evaluating information center success using a structural equation modeling approach. *Decision Sciences*, *28*(2), 309-334.

Chaudhari, P., & Patel, B. (2017). Future of Big Data. *International Research Journal of Engineering and Technology*, *4*(1), 595-597.

Chen, Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of management information systems*, *32*(4), 4-39.

Chen, C. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, *275*, 314-347.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, *36*(4).

Chen, H.-M., Schütz, R., Kazman, R., & Matthes, F. (2016). *Amazon in the air: Innovating with big data at Lufthansa* Proceedings of the 49th Hawaii international conference on system sciences (HICSS), Koloa, HI, USA. https://doi.org/10.1109/HICSS.2016.631

Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: a data management perspective. *Frontiers of Computer Science*, *7*(2), 157-164.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications*, *19*(2), 171-209.

Chen, M.-J. (1996). Competitor analysis and interfirm rivalry: Toward a theoretical integration. *Academy of management review*, *21*(1), 100-134.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research*, *295*(2), 295-336.

Chin, W. W., & Todd, P. A. (1995). On the use, usefulness, and ease of use of structural equation modeling in MIS research: A note of caution. *MIS quarterly*, 237-246.

Choi, T. M., Wallace, S. W., & Wang, Y. (2018). Big data analytics in operations management. *Production and operations management*, *27*(10), 1868-1883.

Chong, J. K. (2004). Six steps to better crisis management. *Journal of Business Strategy*, *25*(2), 43-46.

Chopra, A., & Madan, S. (2015). Big data: a trouble or a real solution? *International Journal of Computer Science Issues (IJCSI)*, *12*(2), 221.

Chow, W. S., & Chan, L. S. (2008). Social network, social trust and shared goals in organizational knowledge sharing. *Information & Management*, *45*(7), 458-465.

Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016). *Data Cleaning* International conference on management of data, San Francisco, Calif. https://doi.org/10.1145/2882903.2912574

Churchill, G. A., & Iacobucci, D. (2006). *Marketing research: methodological foundations*. Dryden Press New York.

Cochran, W. G. (2007). *Sampling techniques*. John Wiley & Sons.

Cohen, J., & Cohen, P. (1983). Applied multivariate regression/'correlaticot analysis for the behavioral sciences. In: New York: John Wiley.

Colace, F., Lombardi, M., Pascale, F., Santaniello, D., Tucker, A., & Villani, P. (2018). *A multilevel graph representation for big data interpretation* International conference on data science and syste, Exeter, UK. https://doi.org/10.1109/HPCC/SmartCity/DSS.2018.00233

Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X., & Reis, M. S. (2016). How can SMEs benefit from big data? Challenges and a path forward. *Quality and Reliability Engineering International*, *32*(6), 2151-2164.

Collis, J. (2003). *Business research : a practical guide for undergraduate and postgraduate students / Jill Collis and Roger Hussey*. Basingstoke, Hampshire : Palgrave Macmillan.

Collis, J., & Hussey, R. (2009). A practical guide for undergraduate and postgraduate students. In: Basingstoke: Palgrave Macmillan.

Collis, J., & Hussey, R. (2013). *Business research: A practical guide for undergraduate and postgraduate students*. Macmillan International Higher Education.

Commission, E. (2013). *Women and Men in Leadership Positions in the European Union 2013: A Review of the Situation and Recent Progress*. Publications Office of the European Union.

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44-57.

Cooper, D. R., Schindler, P. S., & Sun, J. (2006). *Business research methods* (Vol. 9). Mcgraw-hill New York.

Cooper, R. G. (1990). Stage-gate systems: a new tool for managing new products. *Business Horizons*, *33*(3), 44-54.

Cooper, R. G., Easingwood, C. J., Edgett, S., Kleinschmidt, E. J., & Storey, C. (1994). What distinguishes the top performing new products in financial services. *Journal of Product Innovation Management: an international publication of the product development & management association*, *11*(4), 281-299.

Côrte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of Big Data Analytics in European firms. *Journal of Business Research*, *70*, 379-390.

Côrte-Real, N., Ruivo, P., Oliveira, T., & Popovič, A. (2019). Unlocking the drivers of big data analytics value in firms. *Journal of Business Research*, *97*, 160-173.

Cosic, R., Shanks, G., & Maynard, S. (2012). Towards a business analytics capability maturity model.

Costa, P., Donnelly, A., Rowstron, A., & O'Shea, G. (2012). *Camdoop: Exploiting In-network Aggregation for Big Data Applications* 9th USENIX Symposium on networked systems design and implementation, San Jose, California. https://www.usenix.org/conference/nsdi12/technical-sessions/presentation/costa.

Couldry, N., & Powell, A. (2014). Big data from the bottom up. *Big Data & Society*, *1*(2), 2053951714539277.

Creswell, J. (2003). Research design: Qualitative, quantitative and mixed method approaches", Thousand Oaks; California: Sage Publications.

Creswell, J. W. (2009). Mapping the field of mixed methods research. In (Vol. 3, pp. 95-108): SAGE publications Sage CA: Los Angeles, CA.

Creswell, J. W. (2014). *Qualitative, quantitative and mixed methods approaches*. Sage.

Daas, P. J., Puts, M. J., Buelens, B., & van den Hurk, P. A. (2015). Big data as a source for official statistics. *Journal of Official Statistics*, *31*(2), 249-262.

Dahiya, R., Le, S., Ring, J. K., & Watson, K. (2021). Big data analytics and competitive advantage: the strategic role of firm-specific knowledge. *Journal of Strategy and Management*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JSMA-08-2020-0203

Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of management journal*, *34*(3), 555-590.

Danaee Fard, H. (2012). Research paradigms in public administration. *The International Journal of Humanities*, *19*(4), 55-108.

Das, T., & Kumar, P. M. (2013). Big data analytics: A framework for unstructured data analysis. *International Journal of Engineering Science & Technology*, *5*(1), 153.

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, *6*(1), 1-25.

Data, T. I. (2005). Oslo manual. *Paris and Luxembourg: OECD/Euro-stat*.

Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review*, *84*(1), 98.

Davenport, T. H., Barth, P., & Bean, R. (2012). *How'big data'is different*. MIT Sloan Management Review.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions, better results*. Harvard Business Press.

Davenport, T. H., & Patil, D. (2012). Data scientist. *Harvard Business Review*, *90*(5), 70-76.

Day, G. S. (2011). Closing the marketing capabilities gap. *Journal of marketing*, *75*(4), 183-195.

De Luca, L. M., & Atuahene-Gima, K. (2007). Market knowledge dimensions and cross-functional collaboration: Examining the different routes to product innovation performance. *Journal of marketing*, *71*(1), 95-112.

Debortoli, S., Müller, O., & vom Brocke, J. (2014). Comparing business intelligence and big data skills. *Business & Information Systems Engineering*, *6*(5), 289-300.

Delen D, Demirkan H. Data, information and analytics as services. Elsevier; 2013 [accessed 2017 Feb 2]. http://www.sciencedir ect.com/science/article/pii/S0167923612001558.

Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). *Addressing big data issues in scientific data infrastructure* 2013 International conference on collaboration technologies and systems (CTS), San Diego, California, USA. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6567203

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, *55*(1), 412-421.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22.

DeSarbo, W. S., Di Benedetto, C. A., & Song, M. (2007). A heterogeneous resource based view for exploring relationships between firm performance and capabilities. *Journal of modelling in management*, *2*(2), 103-130. https://doi.org/10.1108/17465660710763407

Deshpandé, R., Farley, J. U., & Webster Jr, F. E. (1993). Corporate culture, customer orientation, and innovativeness in Japanese firms: a quadrad analysis. *Journal of marketing*, *57*(1), 23-37.

Desouza, K. C., Dombrowski, C., Awazu, Y., Baloh, P., Papagari, S., Jha, S., & Kim, J. Y. (2009). Crafting organizational innovation processes. *Innovation*, *11*(1), 6-33.

Deutskens, E., De Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing letters*, *15*(1), 21-36.

Dey, N., Das, H., Naik, B., & Behera, H. S. (2019). *Big data analytics for intelligent healthcare management*. Academic Press.

Dezdar, S., & Ainin, S. (2011). The influence of organizational factors on successful ERP implementation. *Management decision*, *49*(6), 911-926.

Di Martino, B., Aversa, R., Cretella, G., Esposito, A., & Kołodziej, J. (2014). Big data (lost) in the cloud. *International Journal of Big Data Intelligence*, *1*(1-2), 3-17.

Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management science*, *35*(12), 1504-1511.

Dillman, D. (2000). Constructing the questionnaire. Mail and internet surveys. In: New York: John Wiley & Sons.

Distanont, A., & Khongmalai, O. (2020). The role of innovation in creating a competitive advantage. *Kasetsart Journal of Social Sciences*, *41*(1), 15-21.

Dong, J. Q., & Yang, C.-H. (2020). Business value of big data analytics: A systems-theoretic approach and empirical test. *Information & Management*, *57*(1), 103124.

Dreyer, B., & Grønhaug, K. (2004). Uncertainty, flexibility, and sustained competitive advantage. *Journal of Business Research*, *57*(5), 484-494.

Du Preez, N. D., & Louw, L. (2008). *A framework for managing the innovation process* Portland international conference on management of engineering & technology, Cape Town, South Africa. https://doi.org/10.1109/PICMET.2008.4599663

Duan, Y., Cao, G., & Edwards, J. S. (2020). Understanding the impact of business analytics on innovation. *European Journal of Operational Research*, *281*(3), 673-686.

Dubey, R., Gunasekaran, A., Childe, S. J., Blome, C., & Papadopoulos, T. (2019). Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture. *British Journal of Management*, *30*(2), 341-361.

Dutta, D., & Bose, I. (2015). Managing a big data project: the case of ramco cements limited. *International Journal of Production Economics*, *165*, 293-306.

Easterby-Smith, M., & Prieto, I. M. (2008). Dynamic capabilities and knowledge management: an integrative role for learning? *British Journal of Management*, *19*(3), 235-249.

Eckerson, W. W. (2002). Data quality and the bottom line. *TDWI Report, The Data Warehouse Institute*, 1-32.

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? *Strategic management journal*, *21*(10-11), 1105-1121.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897-904.

Estrin, S. (2002). Competition and corporate governance in transition. *Journal of Economic Perspectives*, *16*(1), 101-124.

Etikan, I., & Bala, K. (2017). Sampling and sampling methods. *Biometrics & Biostatistics International Journal*, *5*(6), 00149.

Etzion, D., & Aragon-Correa, J. A. (2016). Big data, management, and sustainability: Strategic opportunities ahead. In: SAGE Publications Sage CA: Los Angeles, CA.

Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet research*.

Evans, J. R., & Lindner, C. H. (2012). Business Analytics: The Next Frontier for Decision Sciences. College of Business, University of Cincinnati. *Decision Science Institute. http://www. cbpp. uaa. alaska. edu/afef/business_analytics. htm. Zugegriffen am*, *21*(12), 2018

Eveleens, C. (2010). Innovation management; a literature review of innovation process models and their implications. *Science*, *800*(2010), 900.

Everitt, B. S. (2006). *An R and S-PLUS® companion to multivariate analysis*. Springer Science & Business Media.

Fagerberg, J. (2004). *Innovation: A guide to the literature* http://faculty.bus.olemiss.edu/dhawley/mba622/Fagerberg.pdf

Fan, W., & Bifet, A. (2013). Mining big data: current status, and forecast to the future. *ACM sIGKDD Explorations Newsletter*, *14*(2), 1-5.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 56-83.

Fatorachian, H., & Kazemi, H. (2018). A critical investigation of Industry 4.0 in manufacturing: theoretical operationalisation framework. *Production Planning & Control*, *29*(8), 633-644.

Fernández, P., Del Río, M. L., Varela, J., & Bande, B. (2010). Relationships among functional units and new product performance: The moderating effect of technological turbulence. *Technovation*, *30*(5-6), 310-321.

Ferranti, J. M., Langman, M. K., Tanaka, D., McCall, J., & Ahmad, A. (2010). Bridging the gap: leveraging business intelligence tools in support of patient safety and financial effectiveness. *Journal of the American Medical Informatics Association*, *17*(2), 136-143.

Ferraris, A., Mazzoleni, A., Devalle, A., & Couturier, J. (2019). Big data analytics capabilities and knowledge management: impact on firm performance. *Management decision*.

Ferreira, J., & Fernandes, C. (2017). Resources and capabilities' effects on firm performance: what are they? *Journal of Knowledge Management*.

Fichman, R. G. (2001). The role of aggregation in the measurement of IT-related organizational innovation. *MIS quarterly*, 427-455.

Fihn, S. D., Francis, J., Clancy, C., Nielson, C., Nelson, K., Rumsfeld, J., Cullen, T., Bates, J., & Graham, G. L. (2014). Insights from advanced analytics at the Veterans Health Administration. *Health affairs*, *33*(7), 1203-1211.

Financial Analysis Made Easy. (2022). Bureau van Dijk Electronic Publishing Ltd and VISTRA. https://fame4.bvdinfo.com/version-20211216/fame/1/Companies

Fowler Jr, F. J., & Cosenza, C. (2009). Design and evaluation of survey questions. *The SAGE handbook of applied social research methods*, 375-412.

Frenkel, A. (2003). Barriers and Limitations in the Development of Industrial Innovation in the Region. *European Planning Studies*, *11*(2), 115-137.

Frizzo-Barker, J., Chow-White, P. A., Mozafari, M., & Ha, D. (2016). An empirical study of the rise of big data in business scholarship. *International journal of information management*, *36*(3), 403-413.

Galia, F., & Legros, D. (2004). Complementarities between obstacles to innovation: evidence from France. *Research Policy*, *33*(8), 1185-1199.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144.

Ganesan, S. (1994). Determinants of long-term orientation in buyer-seller relationships. *Journal of marketing*, *58*(2), 1-19.

Garapati, S. L., & Garapati, S. (2018). Application of big data analytics: an innovation in health care. *Comput Intell*, *14*(1), 15-27.

Gärtner, B., & Hiebl, M. R. (2017). Issues with big data. In *The Routledge companion to accounting information systems* (pp. 161-172). Routledge.

Geczy, P. (2014). Big data characteristics. *The Macrotheme Review*, *3*(6), 94-104.

Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the association for information systems*, *4*(1), 7.

Geissdoerfer, M., & Weerdmeester, R. (2019). Managing business model innovation for relocalization in the process and manufacturing industry.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. In: Academy of Management Briarcliff Manor, NY.

George, G., & Lin, Y. (2017). Analytics, innovation, and organizational adaptation. *Innovation*, *19*(1), 16-22.

Gërguri-Rashiti, S., Ramadani, V., Abazi-Alili, H., Dana, L. P., & Ratten, V. (2017). ICT, innovation and firm performance: the transition economies context. *Thunderbird International Business Review*, *59*(1), 93-102.

Germann, F., Lilien, G. L., Fiedler, L., & Kraus, M. (2014). Do retailers benefit from deploying customer analytics? *Journal of retailing*, *90*(4), 587-593.

Ghasemaghaei, M., & Calic, G. (2019). Does big data enhance firm innovation competency? The mediating role of data-driven insights. *Journal of Business Research*, *104*, 69-84.

Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. *Journal of Business Research*, *108*, 147-162.

Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *The Journal of Strategic Information Systems*, *27*(1), 101-113.

Ghasemaghaei, M., Hassanein, K., & Turel, O. (2017). Increasing firm agility through the use of data analytics: The role of fit. *Decision Support Systems*, *101*, 95-105.

Ghauri, P., Gronhaug, K., & Kristinslund, I. (2010). Research methods in business studies. 4. painos. *Essex: Pearson Education Limited*, 977-979.

Grant, R. M. (1991). The resource-based theory of competitive advantage: implications for strategy formulation. *California management review*, *33*(3), 114-135.

Grant, R. M. (1996). Prospering in dynamically-competitive environments: Organizational capability as knowledge integration. *Organization science*, *7*(4), 375-387.

Grover, V., Chiang, R. H., Liang, T.-P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of management information systems*, *35*(2), 388-423.

Groves, P., Kayyali, B., Knott, D., & Kuiken, S. V. (2016). The'big data'revolution in healthcare: Accelerating value and innovation.

Groves, R. M. (2004). *Survey errors and survey costs* (Vol. 536). John Wiley & Sons.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, *2*(163-194), 105.

Gudivada, V. N., Rao, D., & Raghavan, V. V. (2015). Big data driven natural language processing research and applications. In *Handbook of Statistics* (Vol. 33, pp. 203-238). Elsevier.

Gunday, G., Ulusoy, G., Kilic, K., & Alpkan, L. (2011). Effects of innovation types on firm performance. *International Journal of Production Economics*, *133*(2), 662-676.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191-209.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049-1064.

Gupta, S., Drave, V. A., Dwivedi, Y. K., Baabdullah, A. M., & Ismagilova, E. (2020). Achieving superior organizational performance via big data predictive analytics: A dynamic capability view. *Industrial Marketing Management*, *90*, 581-592.

Guyon, I., Matic, N., & Vapnik, V. (1996). Discovering Informative Patterns and Data Cleaning. In.

Hadjimanolis, A. (1999). Barriers to innovation for SMEs in a small less developed country (Cyprus). *Technovation*, *19*(9), 561-570.

Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., & Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly*, *21*(4), 431-454.

Hair, J. F. (2009). Multivariate data analysis.

Hair, J. F., Anderson, R. E., Babin, B. J., & Black, W. C. (2010). Multivariate data analysis: A global perspective (Vol. 7). In: Upper Saddle River, NJ: Pearson.

Hair, J. F., Money, A. H., Samouel, P., & Page, M. (2007). Research methods for business. *Education+ Training*.

Hajli, N., Tajvidi, M., Gbadamosi, A., & Nadeem, W. (2020). Understanding market agility for new product success with big data analytics. *Industrial Marketing Management*, *86*, 135-143.

Hameed, M. A., Counsell, S., & Swift, S. (2012). A conceptual model for the process of IT innovation adoption in organizations. *Journal of Engineering and Technology Management*, *29*(3), 358-390.

Han, Q., Liang, S., & Zhang, H. (2015). Mobile cloud sensing, big data, and 5G networks make an intelligent and smart world. *IEEE Network*, *29*(2), 40-45.

Hans, N., Mahajan, S., & Omkar, S. (2015). Big data clustering using genetic algorithm on hadoop mapreduce. *Int. J. Sci. Technol. Res*, *4*, 58-62.

Hao, S., Zhang, H., & Song, M. (2019). Big data, big data analytics capability, and sustainable innovation performance. *Sustainability*, *11*(24), 7145.

Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data*, *6*(1), 1-16.

Harrington, D. (2009). *Confirmatory factor analysis*. Oxford university press.

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2014). Big data for big business? A taxonomy of data-driven business models used by start-up firms. *A taxonomy of data-driven business models used by start-up firms*.

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., & Chiroma, H. (2016). The role of big data in smart city. *International journal of information management*, *36*(5), 748-758.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information systems*, *47*, 98-115.

Hashi, I., & Stojčić, N. (2013). The impact of innovation activities on firm performance using a multi-stage model: Evidence from the Community Innovation Survey 4. *Research Policy*, *42*(2), 353-366.

Hauser, J., Tellis, G., & Griffin, A. (2005). Research on Innovation: A Review and Agenda for Marketing Science. Special Report 05-200. *Marketing Science Institute*.

Hekkert, M. P., Suurs, R. A., Negro, S. O., Kuhlmann, S., & Smits, R. E. (2007). Functions of innovation systems: A new approach for analysing technological change. *Technological Forecasting and Social Change*, *74*(4), 413-432.

Helfat, C. E., & Peteraf, M. A. (2009). Understanding dynamic capabilities: progress along a developmental path. In: Sage publications Sage UK: London, England.

Henao-García, E., Arias-Pérez, J., & Lozada, N. (2021). Fostering big data analytics capability through process innovation: Is management innovation the missing link? *Business Information Review*, *38*(1), 28-39.

Hess, A. M., & Rothaermel, F. T. (2011). When are assets complementary? Star scientists, strategic alliances, and innovation in the pharmaceutical industry. *Strategic management journal*, *32*(8), 895-909.

Higón, D. A. (2012). The impact of ICT on innovation activities: Evidence for UK SMEs. *International Small Business Journal*, *30*(6), 684-699.

Hirschheim, R. (1985). Information systems epistemology: An historical perspective. *Research methods in information systems*, *9*, 13-35.

Hobday, M. (2005). Firm-level innovation models: perspectives on research in developed and developing countries. *Technology analysis & strategic management*, *17*(2), 121-146.

Hong, K.-K., & Kim, Y.-G. (2002). The critical success factors for ERP implementation: an organizational fit perspective. *Information & Management*, *40*(1), 25-40.

Hong, S., Hyoung Kim, S., Kim, Y., & Park, J. (2019). Big Data and government: Evidence of the role of Big Data for smart cities. *Big Data & Society*, *6*(1), 2053951719842543.

Hooi, T. K., Abu, N. H. B., & Rahim, M. (2018). Relationship of big data analytics capability and product innovation performance using smartPLS 3.2. 6: Hierarchical component modelling in PLS-SEM. *Int. J. Supply Chain Manag*, *7*, 51.

Horng, J.-S., Liu, C.-H., Chou, S.-F., Yu, T.-Y., & Hu, D.-C. (2022). Role of big data capabilities in enhancing competitive advantage and performance in the hospitality sector: Knowledge-based dynamic capabilities view. *Journal of Hospitality and Tourism Management*, *51*, 22-38.

Hou, R., Kong, Y., Cai, B., & Liu, H. (2020). Unstructured big data analysis algorithm and simulation of Internet of Things based on machine learning. *Neural Computing and Applications*, *32*(10), 5399-5407.

Hu, L. t., & Bentler, P. M. (1995). Evaluating model fit.

Hu, L. t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, *3*(4), 424.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Hu, M., Li, Y., & Wang, J. (2018). Efficient ignorance: Information heterogeneity in a queue. *Management science*, *64*(6), 2650-2671.

Huang, T., Lan, L., Fang, X., An, P., Min, J., & Wang, F. (2015). Promises and challenges of big data computing in health sciences. *Big Data Research*, *2*(1), 2-11.

Hudson, L. A., & Ozanne, J. L. (1988). Alternative ways of seeking knowledge in consumer research. *Journal of consumer research*, *14*(4), 508-521.

Hughes, J. A., & Sharrock, W. W. (2016). *The philosophy of social research*. Routledge.

Hull, C. E., & Rothenberg, S. (2008). Firm performance: The interactions of corporate social performance with innovation and industry differentiation. *Strategic management journal*, *29*(7), 781-789.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic management journal*, *20*(2), 195-204.

Hunt, S. D. (1997). Resource-advantage theory: an evolutionary theory of competitive firm behavior? *Journal of Economic Issues, 31*(1), 59-78.

Hurwitz, J. S., Nugent, A., Halper, F., & Kaufman, M. (2013). *Big data for dummies*. John Wiley & Sons.

Iacobucci, D. (2009). Everything you always wanted to know about SEM (structural equations modeling) but were afraid to ask. *Journal of consumer psychology*, *19*(4), 673-680.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of consumer psychology*, *20*(1), 90-98.

Ianni, M., Masciari, E., Mazzeo, G. M., Mezzanzanica, M., & Zaniolo, C. (2020). Fast and effective Big Data exploration by clustering. *Future Generation Computer Systems*, *102*, 84-94.

Ikeda, K., & Marshall, A. (2016). How successful organizations drive innovation. *Strategy & Leadership*.

Ilieva, J., Baron, S., & Healey, N. M. (2002). Online surveys in marketing research. *International Journal of Market Research, 44*(3), 1-14.

Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, *57*(7), 86-94.

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, *70*, 338-345.

Jaseena, K., & David, J. M. (2014). Issues, challenges, and solutions: big data mining. *CS & IT-CSCP*, *4*(13), 131-140.

Jaworski, B. J., & Kohli, A. K. (1993). Market orientation: antecedents and consequences. *Journal of marketing*, *57*(3), 53-70.

Jha, A. K., Agi, M. A., & Ngai, E. W. (2020). A note on big data analytics capability development in supply chain. *Decision Support Systems*, *138*, 113382.

Jiménez-Jiménez, D., & Sanz-Valle, R. (2011). Innovation, organizational learning, and performance. *Journal of Business Research*, *64*(4), 408-417.

Johnson, J. E. (2012). Big data+ big analytics= big opportunity: big data is dominating the strategy discussion for many financial executives. As these market dynamics continue to evolve, expectations will continue to shift about what should be disclosed, when and to whom. *Financial Executive*, *28*(6), 50-54.

Johnson, J. S., Friend, S. B., & Lee, H. S. (2017). Big data facilitation, utilization, and monetization: Exploring the 3Vs in a new product development process. *Journal of product innovation management*, *34*(5), 640-658.

Johnson, J. W., & Rapp, A. (2010). A more comprehensive understanding and measure of customer helping behavior. *Journal of Business Research*, *63*(8), 787-792.

Jokonya, O. (2015). *Towards a conceptual framework for big data adoption in organizations* 2015 International Conference on Cloud Computing and Big Data (CCBD), Shanghai, China. https://doi.org/10.1109/CCBD.2015.59

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*(4), 443-477.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. Scientific Software.

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *IT Professional*, *15*(6), 43-48.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and psychological measurement*, *34*(1), 111-117.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). *Big data: Issues and challenges moving forward* 2013 46th Hawaii International Conference on System Sciences (HICSS), Wailea, Maui, HI USA. https://doi.org/10.1109/HICSS.2013.645

Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, *74*(7), 2561-2573.

Kamioka, T., & Tapanainen, T. (2014). Organizational Use of Big Data and Competitive Advantage-Exploration of Antecedents. *PACIS, 2014*, 18th.

Katal, A., Wazid, M., & Goudar, R. (2013, 8-10 Aug). *Big data: issues, challenges, tools and good practices* 2013 Sixth international conference on contemporary computing (IC3), Noida, India. https://doi.org/10.1109/IC3.2013.6612229

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Ali, M., Kamaleldin, W., Alam, M., Shiraz, M., & Gani, A. (2014). Big data: survey, technologies, opportunities, and challenges. *The Scientific World Journal*, *2014*.

Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78-85.

Kim, M., Song, J., & Triche, J. (2015). Toward an integrated framework for innovation in service: A resource-based view and dynamic capabilities approach. *Information Systems Frontiers*, *17*(3), 533-546.

Kim, M.-K., & Park, J.-H. (2017). Identifying and prioritizing critical factors for promoting the implementation and usage of big data in healthcare. *Information Development*, *33*(3), 257-269.

Kiron, D., Prentice, P. K., & Ferguson, R. B. (2012). Innovating with analytics. *MIT sloan management review*, *54*(1), 47.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography*, *3*(3), 262-267.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Klein, K. J., & Sorra, J. S. (1996). The challenge of innovation implementation. *Academy of management review*, *21*(4), 1055-1080.

Kline. (2011). Convergence of structural equation modeling and multilevel modeling. In: na.

Kline. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kline, R. (2011). Principles and Practice of Structural Equation Modeling (DA Kenny & TD Little Eds. In: New York, London: The Guilford Press: A Division of Guilford Publications, Inc.

Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*. Sage.

Knight, K. E. (1967). A descriptive model of the intra-firm innovation process. *The journal of business*, *40*(4), 478-496.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization science*, *3*(3), 383-397.

Koman, G., Kubina, M., Holubčík, M., & Soviar, J. (2018). *Possibilities of Application a Big Data in the Company Innovation Process* International Conference on Knowledge Management in Organizations, Žilina, Slovakia. https://doi.org/10.1007/978-3-319-95204-8_54

Koronios, A., Gao, J., & Selle, S. (2014). *Big data project success-A meta analysis* PACIS, Chengdu, China. https://aisel.aisnet.org/pacis2014/376/

Kozlenkova, I. V., Samaha, S. A., & Palmatier, R. W. (2014). Resource-based theory in marketing. *Journal of the Academy of Marketing Science*, *42*(1), 1-21.

Kraaijenbrink, J., Spender, J.-C., & Groen, A. J. (2010). The resource-based view: A review and assessment of its critiques. *Journal of management*, *36*(1), 349-372.

Kraska, T. (2013). Finding the needle in the big data systems haystack. *IEEE internet Computing*, *17*(1), 84-86.

Krishnan, S., Wang, J., Franklin, M. J., Goldberg, K., Kraska, T., Milo, T., & Wu, E. (2015). SampleClean: Fast and Reliable Analytics on Dirty Data. *IEEE Data Eng. Bull.*, *38*(3), 59-75.

Kung, L., Kung, H.-J., Jones-Farmer, A., & Wang, Y. (2015). Managing big data for firm performance: a configurational approach.

Kuo, S.-Y., Lin, P.-C., & Lu, C.-S. (2017). The effects of dynamic capabilities, service capabilities, competitive advantage, and organizational performance in container shipping. *Transportation Research Part A: Policy and Practice*, *95*, 356-371.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International journal of information management*, *34*(3), 387-394.

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, *5*(12), 2032-2033.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review*, *52*(2), 21-32.

Lee, J., Kao, H.-A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia Cirp*, *16*, 3-8.

Lee, R., Lee, J.-H., & Garrett, T. C. (2019). Synergy effects of innovation on firm performance. *Journal of Business Research*, *99*, 507-515.

Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural equation modeling*, *12*(1), 1-27.

Lei, P. W., & Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: issues and practice*, *26*(3), 33-43.

Lepkowski, J. M., Landis, J. R., & Stehouwer, S. A. (1987). Strategies for the analysis of imputed data from a sample survey: The national medical care utilization and expenditure survey. *Medical Care*, 705-716.

Levin, M. (1991). The reification-realism-positivism controversy in macromarketing: A philosopher's view. *Journal of Macromarketing*, *11*(1), 57-65.

Li, J., Tao, F., Cheng, Y., & Zhao, L. (2015). Big data in product lifecycle management. *The International Journal of Advanced Manufacturing Technology*, *81*(1-4), 667-684.

Li, S., Peng, G. C., & Xing, F. (2019). Barriers of embedding big data solutions in smart factories: insights from SAP consultants. *Industrial Management & Data Systems*.

Lichtenthaler, U. (2009). Outbound open innovation and its effect on firm performance: examining environmental influences. *R&d Management*, *39*(4), 317-330.

Lichtenthaler, U. (2017). Shared value innovation: linking competitiveness and societal goals in the context of digital transformation. *International Journal of Innovation and Technology Management*, *14*(04), 1750018.

Liebowitz, J. (2013). *Big data and business analytics*. CRC press.

Lin, H.-F. (2014). Understanding the determinants of electronic supply chain management system adoption: Using the technology–organization–environment framework. *Technological Forecasting and Social Change*, *86*, 80-92.

Liu, H., Ke, W., Wei, K. K., & Hua, Z. (2013). The impact of IT capabilities on firm performance: The mediating roles of absorptive capacity and supply chain agility. *Decision Support Systems*, *54*(3), 1452-1462.

Liu, Y. (2013). Sustainable competitive advantage in turbulent business environments. *International Journal of Production Research*, *51*(10), 2821-2841.

Liyanage, S., Greenfield, P. F., & Don, R. (1999). Towards a fourth generation R&D management model-research networks in knowledge management. *International journal of technology management*, *18*(3-4), 372-393.

Lu, J. W., & Beamish, P. W. (2001). The internationalization and performance of SMEs. *Strategic management journal*, *22*(6-7), 565-586.

Lu, Y., & Ramamurthy, K. (2011). Understanding the link between information technology capability and organizational agility: An empirical examination. *MIS quarterly*, 931-954.

Ma, H. (2000). Competitive advantage and firm performance. *Competitiveness Review: An International Business Journal*.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, *1*(2), 130.

Madrid-Guijarro, A., Garcia, D., & Van Auken, H. (2009). Barriers to innovation among Spanish manufacturing SMEs. *Journal of Small Business Management*, *47*(4), 465-488.

Maheshwari, S., Gautam, P., & Jaggi, C. K. (2021). Role of big data analytics in supply chain management: current trends and future perspectives. *International Journal of Production Research*, *59*(6), 1875-1900.

Majumdar, S. K. (1997). The impact of size and age on firm-level performance: some evidence from India. *Review of industrial organization*, *12*(2), 231-241.

Makadok, R. (1999). Interfirm differences in scale economies and the evolution of market shares. *Strategic management journal*, *20*(10), 935-952.

Malaka, I., & Brown, I. (2015, September 28 - 30). *Challenges to the organisational adoption of big data analytics: A case study in the South African telecommunications industry* Proceedings of the 2015 annual research conference on South African institute of computer scientists and information technologists, Stellenbosch, South Africa. https://dl.acm.org/doi/pdf/10.1145/2815782.2815793

Maleki, M., & Anand, D. (2008). The Critical Success Factors in Customer Relationship Management (CRM)(ERP) Implementation. *Journal of Marketing & Communication*, *4*(2).

Malhotra, N. K., Kim, S. S., & Patil, A. (2006). Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. *Management science*, *52*(12), 1865-1883.

Mandal, S. (2018). An examination of the importance of big data analytics in supply chain agility development. *Management Research Review*.

Manogaran, G., Shakeel, P. M., Baskar, S., Hsu, C.-H., Kadry, S. N., Sundarasekar, R., Kumar, P. M., & Muthu, B. A. (2020). FDM: Fuzzy-optimized data management technique for improving big data analytics. *IEEE Transactions on Fuzzy Systems*, *29*(1), 177-185.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Mariani, M. M., & Wamba, S. F. (2020). Exploring how consumer goods companies innovate in the digital age: The role of big data analytics companies. *Journal of Business Research*, *121*, 338-352.

Maroufkhani, P., Wagner, R., Wan Ismail, W. K., Baroto, M. B., & Nourani, M. (2019). Big data analytics and firm performance: A systematic review. *Information*, *10*(7), 226.

Marr, B. (2015). *Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance*. John Wiley & Sons.

Marshall, A., Mueck, S., & Shockley, R. (2015). How leading organizations use big data and analytics to innovate. *Strategy & Leadership*.

Martinez, M. G., & Briz, J. (2000). Innovation in the Spanish food & drink industry. *The International Food and Agribusiness Management Review*, *3*(2), 155-176.

Mata, F. J., Fuerst, W. L., & Barney, J. B. (1995). Information technology and sustained competitive advantage: A resource-based analysis. *MIS quarterly*, 487-505.

Matsunaga, M. (2010). How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *International journal of psychological research*, *3*(1), 97-110.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60-68.

McCosker, A., & Wilken, R. (2014). Rethinking 'big data'as visual knowledge: the sublime and the diagrammatic in data visualisation. *Visual Studies*, *29*(2), 155-164.

McDermott, C. M., & O'connor, G. C. (2002). Managing radical innovation: an overview of emergent strategy issues. *Journal of Product Innovation Management: an international publication of the product development & management association*, *19*(6), 424-438.

McGregor, A., Hall, M., Lorier, P., & Brunskill, J. (2004). *Flow clustering using machine learning techniques* 5th International workshop on passive and active network measurement, Antibes Juan-les-Pins, France. https://link.springer.com/content/pdf/10.1007%2Fb96961.pdf

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. Guilford Press.

McLafferty, S. L. (2003). Conducting questionnaire surveys. *Key methods in geography*, 87-100.

McMahon, S. R. (2019). Lean data: How small insights drive big data innovation. *Organizational Dynamics*, *48*(1), 56-62.

Mellinger, C. D., & Hanson, T. A. (2016). *Quantitative research methods in translation and interpreting studies*. Taylor & Francis.

Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2016). A data quality in use model for big data. *Future Generation Computer Systems*, *63*, 123-130.

Merlo, O., Bell, S. J., Mengüç, B., & Whitwell, G. J. (2006). Social capital, customer service orientation and creativity in retail stores. *Journal of Business Research*, *59*(12), 1214-1221.

Meyers, L. S., Gamst, G., & Guarino, A. (2006). Applied multivariate research: Design and interpretation Sage Publications. *Inc. Thousand Oaks, CA*.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019a). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, *98*, 261-276.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019b). Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment. *British Journal of Management*, *30*(2), 272-298.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2020). The role of information governance in big data analytics driven innovation. *Information & Management*, *57*(7), 103361.

Mikalef, P., Framnes, V. A., Danielsen, F., Krogstie, J., & Olsen, D. (2017). Big data analytics capability: antecedents and business value.

Mikalef, P., & Krogstie, J. (2018). *Big Data Analytics as an Enabler of Process Innovation Capabilities: A Configurational Approach* In Business Process Management 16th International Conference, BPM, Sydney, NSW, Australia. https://link.springer.com/content/pdf/10.1007%2F978-3-319-98648-7.pdf

Mikalef, P., & Krogstie, J. (2020). Examining the interplay between big data analytics and contextual factors in driving process innovation capabilities. *European Journal of Information Systems*, 1-28.

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management*, *57*(2), 103169.

Mithas, S., Ramasubbu, N., & Sambamurthy, V. (2011). How information management capability influences firm performance. *MIS quarterly*, 237-256.

Montash, M., Dwivedi, A., & Vidgen, R. (2014). *Knowledge sharing and professional online communities acceptance: An integrated model* Proceedings of the 58th Annual Meeting of the ISSS-2014., Washington DC, USA. https://journals.isss.org/index.php/proceedings58th/article/view/2314/784

Morgan, N. A., Slotegraaf, R. J., & Vorhies, D. W. (2009). Linking marketing capabilities with profit growth. *International Journal of Research in Marketing*, *26*(4), 284-293.

Mortensen, P. S., & Bloch, C. W. (2005). Oslo Manual-Guidelines for collecting and interpreting innovation data. *Organisation for Economic Cooperation and Development, OECD*.

Muhtaroğlu, F. C. P., Demir, S., Obalı, M., & Girgin, C. (2013). *Business model canvas perspective on big data applications* 2013 IEEE International Conference on Big Data, Silicon Valley, CA, USA. https://doi.org/10.1109/BigData.2013.6691684

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological bulletin*, *105*(3), 430.

Müller, O., Fay, M., & Vom Brocke, J. (2018). The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of management information systems*, *35*(2), 488-509.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Jama*, *309*(13), 1351-1352.

Naganathan, V. (2018). Comparative Analysis of Big Data, Big Data Analytics: Challenges and Trends. *Int. Res. J. Eng. Technol*, *5*, 1948-1964.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(1), 1.

Nambiar, R., Bhardwaj, R., Sethi, A., & Vargheese, R. (2013). *A look at challenges and opportunities of big data analytics in healthcare* 2013 IEEE international conference on Big Data, Silicon Valley, CA, USA. https://doi.org/10.1109/BigData.2013.6691753

Narasimhan, R., & Bhuvaneshwari, T. (2014). Big data—a brief study. *Int. J. Sci. Eng. Res*, *5*(9), 350-353.

Narver, J. C., Slater, S. F., & MacLachlan, D. L. (2004). Responsive and proactive market orientation and new‐product success. *Journal of product innovation management*, *21*(5), 334-347.

Neirotti, P., & Pesce, D. (2018). ICT-based innovation and its competitive outcome: the role of information intensity. *European Journal of Innovation Management*.

Neuman, W. L. (2006). *Workbook for Neumann Social research methods: qualitative and quantitative approaches*. Allyn & Bacon.

Neuman, W. L. (2007). *Student workbook with data disc for" Basics of social research": qualitative and quantitative approaches*. Pearson.

Newton-Smith, W. H., & Newton-Smith, W. (2000). *A companion to the philosophy of science*. Blackwell Oxford.

Ng, R. T., & Han, J. (1994). *E cient and E ective clustering methods for spatial data mining* VLDB '94: Proceedings of the 20th International Conference on Very Large Data Bases, San FranciscoCAUnited States. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.27.49&rep=rep1&type=pdf

Niebel, T., Rasel, F., & Viete, S. (2019). BIG data–BIG gains? Understanding the link between big data analytics and innovation. *Economics of Innovation and New Technology*, *28*(3), 296-316.

Norris, M., & Lecavalier, L. (2010). Evaluating the use of exploratory factor analysis in developmental disability psychological research. *Journal of autism and developmental disorders*, *40*(1), 8-20.

Nunnally, J. C. (1994). *Psychometric theory 3E*. Tata McGraw-hill education.

O'Leary, D. E. (2013). Artificial intelligence and big data. *IEEE Intelligent Systems, 28*(2), 96-99.

Obal, M. (2017). What drives post-adoption usage? Investigating the negative and positive antecedents of disruptive technology continuous adoption intentions. *Industrial Marketing Management*, *63*, 42-52.

Oliver, C. (1997). Sustainable competitive advantage: combining institutional and resource-based views. *Strategic management journal*, *18*(9), 697-713.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information systems research*, *2*(1), 1-28.

Otero, C. E., & Peter, A. (2014). Research directions for engineering big data analytics software. *IEEE Intelligent Systems*, *30*(1), 13-19.

Ottum, B. D., & Moore, W. L. (1997). The role of market information in new product success/failure. *Journal of Product Innovation Management: an international publication of the product development & management association*, *14*(4), 258-273.

Oussous, A., Benjelloun, F.-Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, *30*(4), 431-448.

Özköse, H., Arı, E. S., & Gencer, C. (2015). Yesterday, today and tomorrow of big data. *Procedia-Social and Behavioral Sciences*, *195*, 1042-1050.

Pallant, J. (2003). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS*. Berkshire: McGraw-Hill Education.

Pallant, J. (2010). SPSS survaival manual: a step by step guide to data analysis using SPSS. In: McGraw-Hill Education.

Panger, G. (2016). Reassessing the Facebook experiment: critical thinking about the validity of Big Data research. *Information, Communication & Society*, *19*(8), 1108-1126.

Park, J.-H., Kim, M.-K., & Paik, J.-H. (2015). The factors of technology, organization and environment influencing the adoption and usage of big data in Korean firms.

Pavlou, P. A., & El Sawy, O. A. (2006). From IT leveraging competence to competitive advantage in turbulent environments: The case of new product development. *Information systems research*, *17*(3), 198-227.

Pavlou, P. A., & El Sawy, O. A. (2010). The "third hand": IT-enabled competitive advantage in turbulence through improvisational capabilities. *Information systems research*, *21*(3), 443-471.

Perera, C., Ranjan, R., Wang, L., Khan, S. U., & Zomaya, A. Y. (2015). Big data privacy in the internet of things era. *IT Professional*, *17*(3), 32-39.

Perneger, T. V., Courvoisier, D. S., Hudelson, P. M., & Gayet-Ageron, A. (2015). Sample size for pre-tests of questionnaires. *Quality of Life Research*, *24*(1), 147-151.

Peteraf, M. A. (1993). The cornerstones of competitive advantage: a resource-based view. *Strategic management journal*, *14*(3), 179-191.

Peteraf, M. A., & Barney, J. B. (2003). Unraveling the resource-based tangle. *Managerial and decision economics*, *24*(4), 309-323.

Peterson, R. A. (2000). *Constructing effective questionnaires* (Vol. 1). Sage publications Thousand Oaks, CA.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. sage.

Pisano, G. P. (2015). You need an innovation strategy. *Harvard Business Review, 93*(6), 44-54.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology, 88*(5), 879.

Polyakova, A., Loginov, M., Strelnikov, E., & Usova, N. (2019). Managerial decision support algorithm based on network analysis and big data. *International Journal of Civil Engineering and Technology, 10*(2), 291-300.

Ponta, L., Puliga, G., Oneto, L., & Manzini, R. (2020). Identifying the determinants of innovation capability with machine learning and patents. *IEEE transactions on engineering management*.

Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2014). How information-sharing values influence the use of information systems: An investigation in the business intelligence systems context. *The Journal of Strategic Information Systems*, *23*(4), 270-283.

Popovič, A., Hackney, R., Tassabehji, R., & Castelli, M. (2018). The impact of big data analytics on firms' high value business performance. *Information Systems Frontiers*, *20*(2), 209-222.

Porter, M. E., & Millar, V. E. (1985). How information gives you competitive advantage. In: Harvard Business Review Reprint Service.

Powell, T. C., & Dent-Micallef, A. (1997). Information technology as competitive advantage: The role of human, business, and technology resources. *Strategic management journal*, *18*(5), 375-405.

Power, D. J. (2014). Using 'Big Data'for analytics and decision support. *Journal of Decision Systems*, *23*(2), 222-228.

Prajogo, D. I. (2006). The relationship between innovation and business performance—a comparative study between manufacturing and service firms. *Knowledge and process management*, *13*(3), 218-225.

Prescott, M. E. (2014). Big data and competitive advantage at Nielsen. *Management decision*.

Priem, R. L., & Butler, J. E. (2001). Is the resource-based "view" a useful perspective for strategic management research? *Academy of management review*, *26*(1), 22-40.

Punch, K. F. (2013). *Introduction to social research: Quantitative and qualitative approaches*. sage.

Qi, C.-c. (2020). Big data management in the mining industry. *International Journal of Minerals, Metallurgy and Materials*, *27*(2), 131-139.

Rachinger, M., Rauter, R., Müller, C., Vorraber, W., & Schirgi, E. (2019). Digitalization and its influence on business model innovation. *Journal of Manufacturing Technology Management*.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems*, *2*(1), 3.

Raguseo, E., & Vitari, C. (2018). Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects. *International Journal of Production Research*, *56*(15), 5206-5221.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.*, *23*(4), 3-13.

Rahman, M. J., & Zhao, N. (2020). Information technology capability and firm performance in the era of Big Data analytics: Evidence from the United States. *Journal of Accounting and Management Information Systems*.

Rajagopalan, R., & Varshney, P. K. (2006). Data aggregation techniques in sensor networks: A survey.

Rajaguru, R., & Matanda, M. J. (2013). Effects of inter-organizational compatibility on supply chain capabilities: Exploring the mediating role of inter-organizational information systems (IOIS) integration. *Industrial Marketing Management*, *42*(4), 620-632.

Rakesh, A., Gehrke, J., Dimitrios, G., & Prabhakar, R. (1998). *Automatic subspace clustering of high dimensional data for data mining applications.* In Proceedings of the 1998 ACM SIGMOD international conference on Management of data,

Ramadan, M., Shuqqo, H., Qtaishat, L., Asmar, H., & Salah, B. (2020). Sustainable competitive advantage driven by big data analytics and innovation. *Applied Sciences*, *10*(19), 6784.

Ransbotham, S., & Kiron, D. (2017). Analytics as a source of business innovation. *MIT sloan management review*, *58*(3).

Rapp, A., Beitelspacher, L. S., Grewal, D., & Hughes, D. E. (2013). Understanding social media effects across seller, retailer, and consumer interactions. *Journal of the Academy of Marketing Science*, *41*(5), 547-566.

Rasmussen, J. L. (1988). Evaluating outlier identification tests: Mahalanobis D squared and Comrey Dk. *Multivariate behavioral research*, *23*(2), 189-202.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, *41*(2), 79-82.

Rehm, S.-V., & Goel, L. (2015). The emergence of boundary clusters in inter-organizational innovation. *Information and Organization*, *25*(1), 27-51.

Ren, J.-f., Fosso Wamba, S., Akter, S., Dubey, R., & Childe, S. J. (2017a). Modelling quality dynamics on business value and firm performance in big data analytics environment.

Ren, J.-f., Fosso Wamba, S., Akter, S., Dubey, R., & Childe, S. J. (2017b). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, *55*(17), 5011-5026.

Retrialisca, F., & Chotijah, U. (2020). The Maturity Measurement of Big Data Adoption in Manufacturing Companies Using the TDWI Maturity Model. *Journal of Information Systems Engineering and Business Intelligence, 6*(1), 70-78.

Ribarsky, W., Wang, D. X., & Dou, W. (2014). Social media analytics for competitive advantage. *Computers & Graphics*, *38*, 328-331.

Rice, R. E., & Rogers, E. M. (1980). Reinvention in the innovation process. *Knowledge*, *1*(4), 499-514.

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational research methods*, *12*(4), 762-800.

Rindfleisch, A., Malter, A. J., Ganesan, S., & Moorman, C. (2008). Cross-sectional versus longitudinal survey research: Concepts, findings, and guidelines. *Journal of marketing research*, *45*(3), 261-279.

Rios-Morales, R., & Brennan, L. (2009). Ireland's innovative governmental policies promoting internationalisation. *Research in International Business and Finance*, *23*(2), 157-168.

Roberts, J. K. (1999). Basic Concepts of Confirmatory Factor Analysis.

Robertson, T. S. (1967). The process of innovation and the diffusion of innovation. *Journal of marketing*, *31*(1), 14-19.

Rogers, D. M. A. (1996). The challenge of fifth generation R&D. *Research-technology management*, *39*(4), 33-41.

Rogers, E. M. (2003). Diffusion of innovations. Free Press. *New York, 551*.

Rothwell, R. (1994). Towards the fifth-generation innovation process. *International marketing review*, *11*(1), 7-31.

Roussel, P., Saad, K., & Erickson, T. (1991). Third generation R&D management. *Harvard Business School Press, Boston*.

Rubera, G., & Kirca, A. H. (2012). Firm innovativeness and its performance outcomes: A meta-analytic review and theoretical integration. *Journal of marketing*, *76*(3), 130-147.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and social psychology bulletin*, *28*(12), 1629-1646.

Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, *19*(4), 1-34.

Russom, P. (2013). Managing big data. *TDWI Best Practices Report*, *TDWI Research*, 1-40.

Rust, R. T., Ambler, T., Carpenter, G. S., Kumar, V., & Srivastava, R. K. (2004). Measuring marketing productivity: Current knowledge and future directions. *Journal of marketing*, *68*(4), 76-89.

Saboo, A. R., Kumar, V., & Park, I. (2016). Using Big Data to Model Time-Varying Effects for Marketing Resource (Re) Allocation. *MIS quarterly*, *40*(4).

Sagiroglu, S., & Sinanc, D. (2013). *Big data: A review* 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA. https://doi.org/10.1109/CTS.2013.6567202

Saha, B., & Srivastava, D. (2014). *Data quality: The other face of big data* 2014 IEEE 30th International conference on data engineering, Chicago, IL, USA. https://doi.org/10.1109/ICDE.2014.6816764

Salavou, H. E., & Sergaki, P. (2013). Generic business strategies in Greece: Private food firms versus agricultural cooperatives. *Journal of rural cooperation*, *41*(886-2016-64701), 44-59.

Saleem, H., Li, Y., Ali, Z., Mehreen, A., & Mansoor, M. S. (2020). An empirical investigation on how big data analytics influence China SMEs performance: do product and process innovation matter? *Asia Pacific Business Review*, 1-26.

Sambamurthy, V., Bharadwaj, A., & Grover, V. (2003). Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *MIS quarterly*, 237-263.

Sandryhaila, A., & Moura, J. M. (2014). Big data analysis with signal processing on graphs. *IEEE Signal Processing Magazine*, *31*(5), 80-90.

Santos, M. Y., e Sá, J. O., Andrade, C., Lima, F. V., Costa, E., Costa, C., Martinho, B., & Galvão, J. (2017). A big data system supporting bosch braga industry 4.0 strategy. *International journal of information management*, *37*(6), 750-760.

Santoso, L. W. (2017). Data warehouse with big data technology for higher education. *Procedia Computer Science*, *124*, 93-99.

Sarantakos, S. (2005). Social Research. Palgrave Macmillan. In: China.

Sarantakos, S. (2012). *Social research*. Macmillan International Higher Education.

Saris, W. E., & Gallhofer, I. (2007). Estimation of the effects of measurement characteristics on the quality of survey questions. Survey research methods,

Saunders, M. (2016). *Research methods for business students / Mark Saunders, Philip Lewis and Adrian Thornhill* (Seventh edition. ed.). New York : Pearson Education.

Saunders, M., Lewis, P., & Thornhill, A. (2007). Research methods. *Business Students 4th edition Pearson Education Limited, England*.

Saunders, M., Lewis, P., and Thornhill, A. (2007). Research methods. *Business Students 4th edition Pearson Education Limited, England*.

Saunders, M., Lewis, P., and Thornhill, A. (2009). *Research methods for business students*. Pearson education.

Saunders, M., Lewis, P., and Thornhill, A. (2015a). *Research methods. Business*

Students Pearson 7th edition

Saunders, M., Lewis, P., and Thornhill, A. (2015b). *Research methods. Business*

Students Pearson 7th edition

Schilke, O. (2014). On the contingent value of dynamic capabilities for competitive advantage: The nonlinear moderating effect of environmental dynamism. *Strategic management journal*, *35*(2), 179-203.

Schumacker, R., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. psychology press.

Schumacker, R., & Lomax, R. G. (2010). A beginner’s guide to structural equation modeling (third). In: Mahwah, NJ: Lawrence Erlbaum Associates.

Schumaker, R., & Lomax, R. (2004). A beginner’s guide to statistical equation modeling. In: Mahwah, NJ: Erlbaum.

Schwager, S. J., & Margolin, B. H. (1982). Detection of multivariate normal outliers. *The annals of statistics*, *10*(3), 943-954.

Segars, A. H., & Grover, V. (1993). Re-examining perceived ease of use and usefulness: A confirmatory factor analysis. *MIS quarterly*, 517-525.

Sejahtera, F., Wang, W., Indulska, M., & Sadiq, S. W. (2018). *Enablers And Inhibitors Of Effective Use Of Big Data: Insights From A Case Study* PACIS, Yokohama, Japan. https://core.ac.uk/download/pdf/301375956.pdf

Sekaran, U., & Bougie, R. (2002). Research Methods For Business. New York. In: John Wiley & Sons, Inc.

Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach*. John Wiley & Sons.

Senawi, A., Wei, H.-L., & Billings, S. A. (2017). A new maximum relevance-minimum multicollinearity (MRmMC) method for feature selection and ranking. *Pattern Recognition*, *67*, 47-61.

Sergeeva, N., & Trifilova, A. (2018). The role of storytelling in the innovation process. *Creativity and Innovation Management*, *27*(4), 489-498.

Serrato, M., & Ramirez, J. (2017). The strategic business value of big data. In *Big Data Management* (pp. 47-70). Springer.

Setia, P., & Patel, P. C. (2013). How information systems help create OM capabilities: Consequents and antecedents of operational absorptive capacity. *Journal of Operations Management*, *31*(6), 409-431.

Shah, R., & Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management, 24*(2), 148-169.

Shamim, S., Zeng, J., Shariq, S. M., & Khan, Z. (2019). Role of big data management in enhancing big data decision-making capability and quality among Chinese firms: A dynamic capabilities view. *Information & Management*, *56*(6), 103135.

Shan, S., Luo, Y., Zhou, Y., & Wei, Y. (2019). Big data analysis adaptation and enterprises' competitive advantages: the perspective of dynamic capability and resource-based theories. *Technology analysis & strategic management*, *31*(4), 406-420.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research*, *58*(7), 935-943.

Sharma, V., Pandey, B., & Kumar, V. (2016). Importance of big data in financial fraud detection. *International Journal of Automation and Logistics*, *2*(4), 332-348.

Shirazi, F., Tseng, H.-T., Adegbite, O., Hajli, N., & Rouhani, S. (2021). New product success through big data analytics: an empirical evidence from Iran. *Information Technology & People*.

Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. *Journal of Network and Computer Applications*, *71*, 151-166.

Siddiqa, A., Karim, A., & Gani, A. (2017). Big data storage technologies: a survey. *Frontiers of Information Technology & Electronic Engineering*, *18*(8), 1040-1070.

Simon, P. (2013). *Too big to ignore: the business case for big data* (Vol. 72). John Wiley & Sons.

Singh, S., Singh, P., Garg, R., & Mishra, P. (2015). Big Data: Technologies, Trends and Applications. *Sudhakar Singh, Pankaj Singh, Rakhi Garg, PK Mishra,"Big Data: Technologies, Trends and Applications", International Journal of Computer Science and Information Technologies (IJCSIT), ISSN*, 0975-9646.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263-286.

Slater, S. F., & Narver, J. C. (2000). Intelligence generation and superior customer value. *Journal of the Academy of Marketing Science*, *28*(1), 120-127.

Smys, S., & Raj, J. S. (2019). Internet of things and big data analytics for health care with cloud computing. *Journal of Information Technology*, *1*(01), 9-18.

Sobolevsky, S., Sitko, I., Des Combes, R. T., Hawelka, B., Arias, J. M., & Ratti, C. (2014). *Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. the case of residents and foreign visitors in spain* 2014 IEEE international congress on big data, Anchorage, AK, USA. https://doi.org/10.1109/BigData.Congress.2014.28

Sousa, M. J., & Rocha, Á. (2019). Digital learning: Developing skills for digital transformation of organizations. *Future Generation Computer Systems*, *91*, 327-334.

Sowmya, R., & Suneetha, K. (2017). *Data mining with big data* 11th International conference on intelligent systems and control (ISCO), Coimbatore, India. https://doi.org/10.1109/ISCO.2017.7855990

Srinivasan, S., Pauwels, K., Silva-Risso, J., & Hanssens, D. M. (2009). Product innovations, advertising, and stock returns. *Journal of marketing*, *73*(1), 24-43.

Stieneker, J. (2018). Digitization and data analytics will impact upstream operations favorably. *World Oil*, 86-90.

Story, V., O'Malley, L., & Hart, S. (2011). Roles, role performance, and radical innovation competences. *Industrial Marketing Management*, *40*(6), 952-966.

Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the association for information systems*, *13*(1), 24.

Strohbach, M., Daubert, J., Ravkin, H., & Lischka, M. (2016). Big data storage. In *New horizons for a data-driven economy* (pp. 119-141). Springer, Cham.

Struijs, P., Braaksma, B., & Daas, P. J. (2014). Official statistics and big data. *Big Data & Society*, *1*(1), 2053951714538417.

Subramaniyaswamy, V., Vijayakumar, V., Logesh, R., & Indragandhi, V. (2015). Unstructured data analysis on big data using map reduce. *Procedia Computer Science*, *50*, 456-465.

Suikki, R., Tromstedt, R., & Haapasalo, H. (2006). Project management competence development framework in turbulent business environment. *Technovation*, *26*(5-6), 723-738.

Sun, S., Cegielski, C. G., Jia, L., & Hall, D. J. (2018). Understanding the factors affecting the organizational adoption of big data. *Journal of Computer Information Systems*, *58*(3), 193-203.

Sun, S., Hall, D. J., & Cegielski, C. G. (2020). Organizational intention to adopt big data in the B2B context: An integrated view. *Industrial Marketing Management*, *86*, 109-121.

Sundmaeker, H., Guillemin, P., Friess, P., & Woelfflé, S. (2010). Vision and challenges for realising the Internet of Things. *Cluster of European Research Projects on the Internet of Things, European Commision*, *3*(3), 34-36.

Suoniemi, S., Meyer-Waarden, L., Munzel, A., Zablah, A. R., & Straub, D. (2020). Big data and firm performance: The roles of market-directed capabilities and business strategy. *Information & Management*, *57*(7), 103365.

Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. *Information & Management*, *57*(1), 103146.

Sutapa, S., Mulyana, M., & Wasitowati, W. (2017). The role of market orientation, creativity and innovation in creating competitive advantages and creative industry performance. *JDM (Jurnal Dinamika Manajemen)*, *8*(2), 152-166.

Süzen, H. N., & Mamur, N. (2014). Reflection of "Philosophy" on Art and Philosophy of Art. *Procedia-Social and Behavioral Sciences*, *122*, 261-265.

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics: International edition. *Pearson2012*.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Pearson Boston, MA.

Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *How to Test the Validation of a Questionnaire/Survey in a Research (August 10, 2016)*.

Tambe, P. (2014). Big data investment, skills, and firm value. *Management science*, *60*(6), 1452-1469.

Tang, N. (2014). *Big data cleaning* In Web Technologies and Applications 16th Asia-Pacific Web Conference, APWeb, https://link.springer.com/content/pdf/10.1007%2F978-3-319-11116-2.pdf

Tashakkori, A., Teddlie, C., & Teddlie, C. B. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (Vol. 46). sage.

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic management journal*, *28*(13), 1319-1350.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, *18*(7), 509-533.

Tellis, G. J., Prabhu, J. C., & Chandy, R. K. (2009). Radical innovation across nations: The preeminence of corporate culture. *Journal of marketing*, *73*(1), 3-23.

Teras, M., & Raghunathan, S. (2016). Big Data Visualisation in Immersive Virtual Reality Environments: Embodied Phenomenological Perspectives to Interaction. *ICTACT Journal on Soft Computing*, *5*(4), 1009-1015.

Tharenou, P., Donohue, R., & Cooper, B. (2007). *Management research methods*. Cambridge University Press.

Thode, H. C. (2002). *Testing for normality* (Vol. 164). CRC press.

Tidd, J., & Bessant, J. R. (2018). *Managing innovation: integrating technological, market and organizational change*. John Wiley & Sons.

Tidd, J., & Bessant, J. R. (2020). *Managing innovation: integrating technological, market and organizational change*. Wiley.

Tinsley, H. E., & Brown, S. D. (2000). *Handbook of applied multivariate statistics and mathematical modeling*. Academic press.

Tippins, M. J., & Sohi, R. S. (2003). IT competency and firm performance: is organizational learning a missing link? *Strategic management journal*, *24*(8), 745-761.

To, M. L., & Ngai, E. W. (2006). Predicting the organisational adoption of B2C e-commerce: an empirical study. *Industrial Management & Data Systems*.

To, W.-M., & Lai, L. S. (2015). Data analytics in China: trends, issues, and challenges. *IT Professional*, *17*(4), 49-55.

Trieu, V.-H. (2017). Getting value from Business Intelligence systems: A review and research agenda. *Decision Support Systems, 93*, 111-124.

Trkman, P., McCormack, K., De Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, *49*(3), 318-327.

Tsai, K.-H., & Yang, S.-Y. (2014). The contingent value of firm innovativeness for business performance under environmental turbulence. *International Entrepreneurship and Management Journal*, *10*(2), 343-366.

Tsai, M.-S., Tsai, M.-C., & Chang, C.-C. (2013). The direct and indirect factors on affecting organizational sustainability. *J. Mgmt. & Sustainability*, *3*, 67.

Tsikriktsis, N. (2005). A review of techniques for treating missing data in OM survey research. *Journal of Operations Management*, *24*(1), 53-62.

Uddin, M. F., & Gupta, N. (2014). *Seven V's of Big Data understanding Big Data to extract value* Proceedings of the 2014 zone 1 conference of the American Society for Engineering Education, Bridgeport, CT, USA. https://doi.org/10.1109/ASEEZone1.2014.6820689

Ulaga, W., & Eggert, A. (2001). *Developing a standard scale of relationship value in business markets* 17th Annual IMP Conference Proceedings, Oslo, Norway. https://www.impgroup.org/uploads/papers/271.pdf

Ularu, E. G., Puican, F. C., Apostu, A., & Velicanu, M. (2012). Perspectives on big data and big data analytics. *Database Systems Journal*, *3*(4), 3-14.

Upadhyay, P., & Kumar, A. (2020). The intermediating role of organizational culture and internal analytical knowledge between the capability of big data analytics and a firm's performance. *International journal of information management*, *52*, 102100.

Van de Vrande, V., De Jong, J. P., Vanhaverbeke, W., & De Rochemont, M. (2009). Open innovation in SMEs: Trends, motives and management challenges. *Technovation*, *29*(6-7), 423-437.

Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Med*, *2*(10), e267.

Van Rijmenam, M. (2014). *Think bigger: Developing a successful big data strategy for your business*. Amacom.

Vargas-Solar, G., Zechinelli-Martini, J.-L., & Espinosa-Oviedo, J.-A. (2017). Big data management: what to keep from the past to face future challenges? *Data Science and Engineering*, *2*(4), 328-345.

Varian, H. R. (2010). Computer mediated transactions. *American Economic Review*, *100*(2), 1-10.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS quarterly*, 21-54.

Verhoef, P. C., Kooge, E., & Walk, N. (2016). *Creating value with big data analytics: Making smarter marketing decisions*. Routledge.

Vermesan, O., Friess, P., Guillemin, P., Gusmeroli, S., Sundmaeker, H., Bassi, A., Jubert, I. S., Mazura, M., Harrison, M., & Eisenhauer, M. (2011). Internet of things strategic research roadmap. *Internet of things-global technological and societal trends, 1*(2011), 9-52.

Wahyuni, D. (2012). The research design maze: Understanding paradigms, cases, methods and methodologies. *Journal of applied management accounting research*, *10*(1), 69-80.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. In: Wiley Online Library.

Wamba, S. F., Akter, S., & De Bourmont, M. (2018). Quality dominant logic in big data analytics and firm performance. *Business Process Management Journal*.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 165,* 234-246.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356-365.

Wang, Y., & Byrd, T. A. (2017). Business analytics-enabled decision-making effectiveness through knowledge absorptive capacity in health care. *Journal of Knowledge Management*.

Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research, 70,* 287-299.

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3-13.

Wang, Y., Kung, L., Gupta, S., & Ozdemir, S. (2019). Leveraging big data analytics to improve quality of care in healthcare organizations: a configurational perspective. *British Journal of Management*, *30*(2), 362-388.

Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management*, *55*(1), 64-79.

Wang, Z., & Wang, N. (2012). Knowledge sharing, innovation and firm performance. *Expert systems with applications*, *39*(10), 8899-8908.

Waqas, M., Honggang, X., Ahmad, N., Khan, S. A. R., & Iqbal, M. (2021). Big data analytics as a roadmap towards green innovation, competitive advantage and environmental performance. *Journal of cleaner production*, *323*, 128998.

Weber, R. (2004). Editor's comments: the rhetoric of positivism versus interpretivism: a personal view. *MIS quarterly*, iii-xii.

Weersink, A., Fraser, E., Pannell, D., Duncan, E., & Rotz, S. (2018). Opportunities and challenges for big data in agricultural and environmental analysis. *Annual Review of Resource Economics*, *10*, 19-37.

Weinman, J. (2015). *Digital disciplines: Attaining market leadership via the cloud, big data, social, mobile, and the Internet of things*. John Wiley & Sons.

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic commerce research and applications*, *9*(6), 476-487.

Weston, R., & Gore Jr, P. A. (2006). A brief guide to structural equation modeling. *The counseling psychologist*, *34*(5), 719-751.

White, A. (2019). Our top data and analytics predicts for 2019. *Retrieved*, *12*, 2019.

Wickham, H. (2013). Bin-summarise-smooth: a framework for visualising large data. *had. co. nz, Tech. Rep*.

Wiklund, J., & Shepherd, D. (2005). Entrepreneurial orientation and small business performance: a configurational approach. *Journal of business venturing*, *20*(1), 71-91.

Willetts, M., Atkins, A. S., & Stanier, C. (2020). *Barriers to SMEs Adoption of Big Data Analytics for Competitive Advantage* “2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS),” Fez, Morocco. https://doi.org/10.1109/ICDS50568.2020.9268687

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian journal of paramedicine*, *8*(3).

Williamson, O. E. (1999). Strategy research: governance and competence perspectives. *Strategic management journal*, *20*(12), 1087-1108.

Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic management journal*, *24*(10), 991-995.

Wixom, B. H., Yen, B., & Relich, M. (2013). Maximizing value from business analytics. *MIS Quarterly Executive*, *12*(2).

Wortmann, F., & Flüchter, K. (2015). Internet of things. *Business & Information Systems Engineering, 57*(3), 221-224.

Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2013). Data mining with big data. *IEEE transactions on knowledge and data engineering*, *26*(1), 97-107.

Wymbs, C. (2016). Managing the innovation process: Infusing data analytics into the undergraduate business curriculum (lessons learned and next steps). *Journal of Information Systems Education*, *27*(1), 61.

Xu, Z., & Shi, Y. (2015). Exploring big data analysis: fundamental scientific problems. *Annals of Data Science*, *2*(4), 363-372.

Yadegaridehkordi, E., Hourmand, M., Nilashi, M., Shuib, L., Ahani, A., & Ibrahim, O. (2018). Influence of big data adoption on manufacturing companies' performance: An integrated DEMATEL-ANFIS approach. *Technological Forecasting and Social Change*, *137*, 199-210.

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth*, *10*(1), 13-53.

Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. *International journal of information management*, *36*(6), 1231-1247.

Yasmin, M., Tatoglu, E., Kilic, H. S., Zaim, S., & Delen, D. (2020). Big data analytics capabilities and firm performance: An integrated MCDM approach. *Journal of Business Research*, *114*, 1-15.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). sage.

Yin, S., & Kaynak, O. (2015). Big data for modern industry: challenges and trends [point of view]. *Proceedings of the IEEE*, *103*(2), 143-146.

Yu, B., Gong, J., & Xu, C.-Z. (2008). *Catch-up: a data aggregation scheme for vanets* Proceedings of the fifth ACM international workshop on VehiculAr Inter-NETworking, San Francisco California USA. https://dl.acm.org/doi/pdf/10.1145/1410043.1410053

Yu, W., Jacobs, M. A., Salisbury, W. D., & Enns, H. (2013). The effects of supply chain integration on customer satisfaction and financial performance: An organizational learning perspective. *International Journal of Production Economics*, *146*(1), 346-358.

Yu, W., Wong, C. Y., Chavez, R., & Jacobs, M. A. (2021). Integrating big data analytics into supply chain finance: The roles of information processing and data-driven culture. *International Journal of Production Economics*, *236*, 108135.

Yunis, M., Tarhini, A., & Kassar, A. (2018). The role of ICT and innovation in enhancing organizational performance: The catalysing effect of corporate entrepreneurship. *Journal of Business Research*, *88*, 344-356.

Zahra, S. A., & Das, S. R. (1993). Innovation strategy and financial performance in manufacturing companies: An empirical study. *Production and operations management*, *2*(1), 15-37.

Zeng, D., & Lusch, R. (2013). Big data analytics: Perspective shifting from transactions to ecosystems. *IEEE Intelligent Systems*, *28*(2), 2-5.

Zeng, J., & Glaister, K. W. (2018). Value creation from big data: Looking inside the black box. *Strategic Organization*, *16*(2), 105-140.

Zerdoumi, S., Sabri, A. Q. M., Kamsin, A., Hashem, I. A. T., Gani, A., Hakak, S., Al-Garadi, M. A., & Chang, V. (2018). Image pattern recognition in big data: taxonomy and open challenges: survey. *Multimedia Tools and Applications*, *77*(8), 10091-10121.

Zhang, C., & Wang, X. (2019). The influence of ICT-driven innovation: a comparative study on national innovation efficiency between developed and emerging countries. *Behaviour & Information Technology*, *38*(9), 876-886.

Zhang, D., Ge, L., Hardy, R., Yu, W., Zhang, H., & Reschly, R. (2013). *On effective data aggregation techniques in host-based intrusion detection in manet* 2013 IEEE 10th

Consumer Communications and Networking Conference (CCNC), Las Vegas, USA. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6488430

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied artificial intelligence*, *17*(5-6), 375-381.

Zhang, X., & Lam, J. S. L. (2019). A fuzzy Delphi-AHP-TOPSIS framework to identify barriers in big data analytics adoption: case of maritime organizations. *Maritime Policy & Management*, *46*(7), 781-801.

Zhong, J., & Nieminen, M. (2015). Resource-based co-innovation through platform ecosystem: experiences of mobile payment innovation in China. *Journal of Strategy and Management*, *8*(3), 283.

Zhong, R. Y., Newman, S. T., Huang, G. Q., & Lan, S. (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers & Industrial Engineering*, *101*, 572-591.

Zhou, K. Z., Brown, J. R., & Dev, C. S. (2009). Market orientation, competitive advantage, and performance: A demand-based perspective. *Journal of Business Research*, *62*(11), 1063-1070.

Zhu, K., Kraemer, K. L., & Xu, S. (2006). The process of innovation assimilation by firms in different countries: a technology diffusion perspective on e-business. *Management science*, *52*(10), 1557-1576.

Zikmund, W. (2003). Business research methods (7th edn) Mason. *Ohio: Thomson South Western*.

Zikmund, W. G., Carr, J. C., & Griffin, M. (2013). *Business Research Methods (Book Only)*. Cengage Learning.

Appendices

Appendix 4.1 Survey instruments

Survey Questionnaire of Big Data Analytics in innovation Context

Thank you for agreeing to help us understand the staff skills required for big data analysis, big data impact on the innovation process, competitive advantage, and subsequently on firm performance. I am a PhD student at Swansea University and this research is part of my thesis. This questionnaire should only take 12-15 minutes. Please submit your answers through this Google.doc. Your answers will be treated with complete confidentiality, and unless you choose to provide an e-mail address, will be entirely anonymous. If you have any questions about this questionnaire, please contact me on:

E-mail: [redacted]

Mobile: [redacted]

*Required

1. Instruction:

   Please answer the following questions by putting an "x" in the box for the answer or answers that come closest to your opinion. For example if you were asked to indicate your gender and you are female, you should answer the question as follows:

   Mark only one oval.

   ◯ Female

   ◯ Male

Background information

2. Your Age? *

   Tick all that apply.

   ☐ 18-24

   ☐ 25-44

   ☐ 45-64

   ☐ 65+

7. How many employees are at your company? (select only one.) *

Tick all that apply.

- [ ] < 200
- [ ] 200~500
- [ ] 500~1000
- [ ] 1000~3000
- [ ] 3000~5000

8. When was your company established? *

Tick all that apply.

- [ ] fewer than 5 years
- [ ] 5-10 years
- [ ] 10-15 years
- [ ] 15-20 years
- [ ] more than 20 years

9. What is your company's activity sector? *

Tick all that apply.

- [ ] Agriculture
- [ ] Automotive
- [ ] Energy
- [ ] High Tech and communication
- [ ] Financial and Insurance
- [ ] Health Care
- [ ] Real Estates
- [ ] Services

Other: [ ] ______________________

14. Technological breakthroughs provide big opportunities in this product/service area. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

15. In our kind of business, customers' product/service preferences change a lot over time. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

16. Marketing practices in our product/service area are constantly changing. New product/service introductions are very frequent *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

Effective use of data aggregation tools

Please, rate the effectiveness by which your firm uses the following business data aggregation tools.

21. Predicting product/service patterns in response to customers' needs. *

Mark only one oval.

	1	2	3	4	5	
Poorly developed	◯	◯	◯	◯	◯	Well developed

22. Analyzing data in near-real or real-time that allows responses to unexpected market threats. *

Mark only one oval.

	1	2	3	4	5	
Poorly developed	◯	◯	◯	◯	◯	Well developed

23. Analyzing social media data to understand current trends from a large population. *

Mark only one oval.

	1	2	3	4	5	
Poorly developed	◯	◯	◯	◯	◯	Well developed

Effective use of data interpretation tools

Please, rate the effectiveness by which your firm uses the following business data interpretation tools.

24. Providing systemic and comprehensive reporting to help recognize feasible opportunities for product/service improvement. *

Mark only one oval.

	1	2	3	4	5	
Poorly developed	◯	◯	◯	◯	◯	Well developed

29. Our big data analytics staff has the right skills to accomplish their jobs successfully *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

30. Our big data analytics staff has suitable education to fulfil their jobs

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

31. Our big data analytics staff holds suitable work experience to accomplish their jobs successfully *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

32. Our big data analytics staff is well trained *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

36. Our big data analytics managers are able to anticipate the future business needs of innovation process's team, functional managers, suppliers, and customers. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

37. Our big data analytics managers have a good sense of where and when to apply big data. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

38. Our big data analytics managers are able to understand and evaluate the output extracted from big data *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

Innovation Process

Overall, our innovation process leads to...

39. increasing in our technology competitiveness. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

44. pioneering disposition to introduce new process. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

45. clever response to new process introduced by the companies in the same sector. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

46. reducing the cost of product/service better than our competitors in. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

Section E: Competitive Advantage

Overall, our competitive advantages lead to ....
Statement

47. Continuously investing in producing new abilities that give us an advantage compared to our competitors. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

52. Our advantages are optimised in the firm and not in individuals - no one can copy us by stealing our staffs away from us *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

53. No one can copy our business procedures, methods and culture *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

54. Reduce the cost of our product/service *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

55. Increase delivery dependability *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

60. Overall financial performance *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

61. We have moved to new marketplaces more speedily than our competitors *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

62. We have presented new products/services to the market quicker than our competitors *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

63. Our achievement rate of new products/services has been upper than our competitors. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

64. Our market segment has over weighed that of our competitors. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

65. Please add your comments about this research or questionnaire:

This content is neither created nor endorsed by Google.

Google Forms

Appendix 4.2 Approved ethical form

SCHOOL OF MANAGEMENT, SWANSEA UNIVERSITY

<u>FIRST STAGE</u> ETHICAL REVIEW FORM

To be completed for all research involving human subjects OR datasets of any kind OR the environment

Name of PI or PGR Student	Abdullah Hamadi
Staff Number or Student ID	[redacted]
Supervisors*	Dr Nick Hajli. Dr Mike Williams
Date Submitted	01/08/2019
Title of Project	Big Data Analytics in Innovation Context
Name of Funder / Sponsor*	Self – fund
Finance Code / Reference*	
Duration of Project	3 years

Aim of research project
Study aims to address the gap and answers to the following main questions:
1. How do digital transformation through the big data influence the way work is carried out in the innovation process?
2. To take full advantage of digital transformation via big data, what type of new capabilities/skills do workers involved need?
3. Combination of skilled staff and big data as an innovation process resource to what extent can affect competitive advantage and firm performance positively.
4. Is the innovation process lead to competitive advantages?

* *Complete if appropriate*

***Risk evaluation**: Does the proposed research involve any of the following?*
✓Tick** those boxes for which the answer is **YES
X Cross** those boxes for which the answer is **NO

Participants

☒ Will the study involve recruitment of patients or staff through the NHS or the use of NHS data or premises and/or equipment? If this is the case, the project **<u>must</u>** be reviewed by the NHS. Please see the following NHS online tools for help with this http://www.hra-decisiontools.org.uk/research/ and http://www.hra-decisiontools.org.uk/ethics/

☒ Does the study involve participants aged 16 or over who are unable to give informed consent? (e.g. people with learning disabilities: see Mental Capacity Act 2005. All research that falls under the auspices of the Act **<u>must</u>** be reviewed by the NHS)

☒ Does the research involve other vulnerable groups: children, those with cognitive impairment or in unequal relationships? (e.g. your students). This **<u>may</u>** require NHS review, and will typically require the researcher to get **Disclosure & Barring Service (DBS) clearance** (formerly CRB checks)

☒ Will the research harm or pose any risk to the environment? (e.g. research in environmentally sensitive areas (e.g. SSSIs); permission needed to access field sites; transport of samples between countries (e.g. soil); sampling of rare or hazardous material (e.g. invasive species) that could deplete or endanger)

Please describe the participants involved in your research (if no participants, state 'none'): *max 250 words.* Participants of this study consist of managers, senior staffs, Executive/Vice from different companies and organisations across in the UK and Irish and other countries. All participants are volunteers that means there is not any payments. This data collection will be confidential and for the purpose of the study only.

Recruitment

☒ Will the study require the co-operation of a gatekeeper for initial access to the groups or individuals to be recruited? (e.g. students at school, members of self-help group or residents of nursing home?)

☒ Will it be necessary for participants to take part in the study without their knowledge and consent at the time? (e.g. covert observation of people or use of social media content)

☒ Will the research involve any form of deception? (e.g. misinformation or partial information about the purpose or nature of the research)

☒ Will financial inducements (other than reasonable expenses and compensation for time) be offered to participants?

☒ Does the research involve members of the public in a research capacity? (e.g. participant research; participants as co-producers or data collectors)

Please explain the recruitment of participants involved in your research (if no participants, state 'none'): *max 250 words.* Participants involved in this study include managers, senior staffs, Executive/Vice from different business companies and the participation in this survey is voluntary. The methods that will use for this data collection is the **questionnaire**, which explores the views of senior staff on big data and its role in the innovation process.

Research Design

☒ Will the study discuss sensitive topics or require the collection of sensitive information? (e.g. terrorism and extremism; sexual activity, drug use or criminal activity; collection of security sensitive documents or information)

☒ Could the study induce psychological stress or anxiety or cause harm or negative consequences beyond the risks encountered in normal life?

☒ Is pain or more than mild discomfort likely to result from the study?

☒ Will the study involve prolonged or repetitive testing?

☒ Are drugs, placebos or other substances (e.g. foods or vitamins) to be administered to study participants, or will the study involve invasive, intrusive or potentially harmful procedures of any kind? (If any substance is to be administered, this **may** fall under the auspices of the Medicines for Human Use (Clinical Trials) Regulations 2004, and require review by the NHS)

☒ Will tissue samples (including blood) be obtained from participants? (This would fall under the terms of the Human Tissue Act 2004. All research that falls under the auspices of the Act **must** be reviewed by the NHS)

Please summaries your methodology in detail and provide reflective comments with regards to the design of your research: *max 250 words.* Quantitative deductive research; the data collection will be done through the questionnaire. The questionnaire will be sent to the different companies, especially those who are working in senior management and have enough knowledge in terms of firm's activities. The student will use 'Fame' database that provides information about companies across the UK and Irish to collect data.

Data Storage and anonymity

☒ Will the research involve administrative or secure data that requires permission from the appropriate data controllers and/or individuals before use?

☒ Will the research involve the sharing of data or confidential information beyond the initial consent given?

☒ Will the research involve respondents to the Internet or other visual/vocal methods where respondents may be identified?

Please describe how you will store your research data and for how long, and, if appropriate, how you will ensure anonymity of your data subjects: *max 250 words.* Data will be stored on the university computer system, and it will be used in the study until finishing the PhD. The expected date for completing this degree is 2020.

Safety and Risk

☒ Has a risk assessment been completed?

☒ Is there a possibility that the safety of the researcher may be in question? (e.g. in international research: locally employed researchers)

☒ Will the research take place outside the UK where there may be issues of local practice and political or other sensitivities?

☒ Could the research impact negatively upon the reputation of the University, researcher(s), research participants, other stakeholders or any other party?

☒ Do any of the research team have an actual or potential conflict of interest?

☒ Are you aware of any other significant ethical risks or concerns associated with the research proposal? (If yes, please outline them in the space below)

Please describe the health and safety considerations in relation to both participants and researchers (*250 words max): If there are significant concerns an appropriate risk assessment and management plan must be attached.*

There is no risk for both participants and the researcher.

Other significant ethical issues or concerns: (If None, then please state 'None')

None

If any answer to the questions above is ***YES****, then a* ***Second Stage (Full) Ethical Review MAY*** *be required.*

If the project involves ***none of the above****, complete the* ***Declaration****, send this form and a* ***copy of the proposal*** to **Amy Jones the School of Management Research Support Officer: amy.e.jones@swansea.ac.uk**. Research may only commence once approval has been given.

Declaration: *The project will be conducted in compliance with the University's Research Integrity Framework (P1415-956). This includes securing appropriate consent from participants, minimizing the potential for harm, and compliance with data-protection, safety & other legal obligations. Any significant change in the purpose, design or conduct of the research will be reported to the SOM-REC Chair, and, if appropriate, a new request for ethical approval will be made to the SOM-REC.*

Signature of PI or PGR Student	[redacted]
Signature of first supervisor (if appropriate)	[redacted]
Decision of SOM-REC	Approved
Signature of SOM-REC Chair or SOM-REC deputy Chair	[redacted]
Date	[redacted] 23/09/19
SOM-REC Reference number (office use only)	[redacted]

Appendix 5.1 Structural Model