Behind Customer Satisfaction Metrics: Exploring User Perceptions of Net Promoter Score (NPS) as a Measure of Satisfaction

Jade Logan, Daniele Doneddu, Kevin McLafferty, Muneeb I. Ahmad, and Nicholas Micallef

Swansea University, Swansea, UK
{850852, d.doneddu, kevin.mclafferty, m.i.ahmad, nicholas.micallef}@swansea.ac.uk

Abstract. Customer satisfaction is a critical component of user experience, shaping brand perception and influencing long-term customer relationships. Net Promoter Score (NPS) is a widely used metric to measure satisfaction, but its rigid classification thresholds and simplistic design often fail to capture the complexity of customer perceptions. While critics advocate for alternative models, NPS remains deeply embedded in industry practices. Rather than seeking to replace it, this study focuses on enhancing its interpretability by examining how customers understand and express satisfaction through NPS.

Adopting an approach that focuses on the users point-of-view, this research explores customer perception of NPS scores within a UK banking context, investigating how individuals classify themselves within the metric and how the tone of their feedback reflects their true sentiment. Findings reveal that while many UK banking customers correctly identify their NPS category, misclassification is common, particularly among those who feel positive about their experience but provide scores that place them in lower categories.

These results could highlight the importance of designing customer satisfaction measures that align with user expectations and communication styles. Furthermore, analysis of feedback tone suggests that sentiment-based models could complement existing metrics, offering a more nuanced understanding of customer experiences. This research underscores the need for a more nuanced interpretation of customer satisfaction, ensuring that feedback mechanisms capture the complexity of user sentiment and drive meaningful service improvements. More broadly, these findings may have implications for the design of customer feedback systems, underscoring the need for interpretive approaches that can capture the complexity of user sentiment and drive meaningful service improvements.

Keywords: Customer Satisfaction · Customer Experience · User Study · Net Promoter Score · User Perception

1 Introduction

Understanding customer behaviour is essential for measuring and improving user experience, particularly in service industries where perceptions and emotions shape decision-making [29] such as retail banking. Customer satisfaction is a key determinant of long-term engagement, trust, and advocacy, making it critical to develop accurate and meaningful ways to assess how individuals feel about a company. One widely adopted metric for capturing customer sentiment is the Net Promoter Score (NPS). NPS, a simple yet influential measure, classifies customers based on their likelihood of recommending a company [50]. Despite its prevalence, there is growing concern over whether NPS genuinely reflects the complexity of customer sentiment, particularly given the limitations in how users interpret and classify their own experiences [48, 5, 28].

Academic literature has largely focused on evaluating the accuracy of NPS rather than understanding how customers interact with the metric itself. A key critique is that the categorical boundaries within the NPS scale are rigid and may misrepresent the intent behind a customer's score [33, 17]. Customers unfamiliar with the metric could struggle to classify themselves correctly, leading to misalignment between their actual experience and the way their responses are interpreted by businesses. This raises fundamental questions about whether NPS, in its current form, provides a reliable reflection of customer perceptions. To explore this issue, this study investigates. *RQ1: How accurately do individuals perceive the classification criteria for their NPS response?*

Beyond individual perception, broader behavioural shifts also challenge the conventional NPS framework. Since its introduction in 2003, NPS has assumed that recommendations are shared primarily through personal networks [50]. However, customer behaviour has evolved significantly, with digital platforms and online reviews increasingly shaping purchasing decisions [27]. Traditional word-of-mouth has given way to public, visible, and often more influential digital endorsements or criticisms, particularly in sensitive industries such as finance. Customers now seek validation through online rating systems [7], which often use a Likert scale or five-star ratings rather than NPS-style classifications. This shift raises the question of whether NPS remains a valid method for capturing customer sentiment in the digital age. Hence, our second research question. *RQ2: What is the relationship between traditional online rating systems and NPS?*

A deeper issue arises when considering the way customers express their emotions and experiences. Understanding the language customers use in feedback is critical for interpreting their sentiment accurately [22]. While some research has attempted to link NPS ratings to emotion [44, 34], few studies have explored whether NPS reflects the natural ways in which customers describe their experiences. Traditional survey methodologies often force structured responses, potentially leading to skewed or incomplete insights. In contrast, online reviews—where customers self-initiate feedback—may offer a more authentic and emotionally-driven representation of sentiment [47]. If there is a mismatch between how customers rate their experiences using NPS and how they naturally describe them in reviews, it may indicate a fundamental gap between structured satisfaction metrics and real-world consumer behaviour, which leads to our third research question. *RQ3: How do the language and tone of customer reviews relate to their corresponding NPS ratings?*

To answer these questions, a mixed-methods approach was used, with the aim to focus on how individuals interact with and interpret NPS rather than evaluating its validity as a business metric. The study was conducted within a retail banking context because its high-frequency customer interactions provided a rich dataset for applying NPS, which is further supported by the availability of publicly accessible industry data. It is important to note that NPS research is inherently context-dependent. Each industry and cultural setting introduces distinct factors that influence how customers interpret and respond to NPS questions [2, 25]. As such, findings from this study are not intended to be universally generalisable, but rather specific to the nuances of the UK retail banking sector. While this research focuses specifically on the UK banking sector, studies from other industries also highlight limitations in NPS as a measure of true customer sentiment [13, 3]. Although these works do not directly examine misclassification rates, their findings suggest that NPS may fail to capture the nuances of customer experience—supporting the case for a broader, hypothetical generalisation of these issues across sectors.

The study collected 300 customer responses via a Prolific survey, incorporating both structured NPS questions and free-text feedback to examine how individuals classified their own experiences. Additionally, Trustpilot[1] review data and CMA (Competition and Markets Authority)[2]-published NPS scores were analysed to assess the relationship between traditional NPS classifications and naturally occurring customer sentiment in online environments.

Our findings reveal that while many customers align with their NPS categories, a significant number misclassified themselves, often placing their score in a higher category than assigned. Customers rating their experience as a six were most prone to misclassification, suggesting ambiguity in the lower and neutral categories. While NPS and Likert-scale ratings showed a strong correlation, no significant relationship was found between NPS and Trustpilot sentiment, raising concerns about whether NPS effectively captures customer perceptions.

This study makes three key contributions to HCI literature in customer experience (CX) measurement:

- **User Focused Analysis of NPS:** This study addresses a gap in the literature by comparing customer satisfaction perspectives with their NPS scores, aiming to assess the metric's effectiveness in capturing true customer sentiment rather than merely critiquing its limitations.
- **Customer Perception of NPS Categories:** Many customers misunderstand how their responses are categorised, highlighting the need for greater transparency in NPS interpretation.

[1] https://www.trustpilot.com/

[2] a UK non-ministerial government department dedicated to preventing anti-competitive practices and protecting consumer interests in competitive markets

- **Emotional Expression in Customer Feedback:** By comparing NPS responses with online reviews, this research underscores the importance of behavioural and context-sensitive satisfaction evaluation methods.

The following sections provide a literature review on customer satisfaction measurement, digital recommendation behaviours, and the psychology of sentiment expression. The methodology details the strategies for data collection and analysis, followed by the research findings, a discussion, and conclusions that provide recommendations for enhancing satisfaction metrics so that they more accurately capture user sentiment.

2 Background

NPS, introduced by Fred Reichheld in a 2003 Harvard company Review article, was developed following two years of research focused on predicting company growth [50]. Reichheld's research concluded that a single question, centred on the likelihood of customers recommending a company, served as a key indicator of customer loyalty - *"How likely are you to recommend us to a friend or family?"*. His claim was that the propensity to recommend a brand was the ultimate measure of loyalty. External research supports this claim, as 92% of consumers reportedly trust recommendations from friends and family [46]. Reichheld's hypothesis that companies with higher NPS grow faster, due to promoters driving repeat company referrals, is therefore grounded in this broader understanding of consumer behaviour.

Promoters (those answering 9 or 10) are considered highly valuable to the company due to their enthusiasm and likelihood to promote the brand, passives (7 or 8) exhibit a lack of enthusiasm and are more vulnerable to competitive influence and detractors (0 to 6) pose a risk through negative word-of-mouth communication [50]. Despite Reichheld pre-emptively addressing concerns by highlighting the need to avoid grade inflation, where mildly satisfied customers are often counted as loyal [50], there is significant debate in the academic community regarding the robustness of these categories [14, 29, 33].

The broad categorisation of users into promoters, passives, and detractors does not fully capture the nuances of customer sentiment. Kristensen and Eskildsen's analysis suggests that a more accurate grouping would place detractors between zero and four, passives between five and seven, and promoters between eight and ten [28]. This misclassification limits NPS's ability to accurately measure customer loyalty and satisfaction, as the metric often oversimplifies context and relies heavily on industry-specific factors despite being industry ambiguous, explaining why NPS is better suited for certain industries than others [28]. Furthermore, the grouping of customers is not fully supported by empirical analysis. They propose alternative groupings based on statistical significance, highlighting the need for more sophisticated tools to measure customer satisfaction.

3 Related Work

Despite criticisms, NPS remains widely used due to its simplicity and efficiency in tracking customer loyalty [2]. However, its broad classification system and rigid scoring thresholds limit its ability to capture nuanced customer sentiment; this oversimplification misrepresents user experiences and can lead to misaligned business decisions. Additionally, NPS lacks a neutral or unsure option, forcing users into predefined categories that may not accurately reflect their opinions [29]. Studies indicate that when users can provide neutral responses, the distribution of scores shifts, reducing the number of detractors and improving reliability [28]. The absence of such flexibility weakens NPS's ability to serve as a fully representative satisfaction metric.

To address these limitations, various technological solutions have been proposed. Sentiment analysis and natural language processing (NLP) enrich NPS data with qualitative insights from customer feedback [55], allowing businesses to identify common pain points beyond numerical scores. However, requiring written responses increases cognitive effort and reduces survey participation, undermining NPS's key advantage—its speed and ease of use [28].

Another suggested improvement is adaptive survey methodologies, which adjust follow-up questions based on initial responses to clarify intent and sentiment in real time [42]. While this aligns with HCI principles by reducing the cognitive burden of rigid surveys, excessive questioning may deter participation or introduce bias if users feel overwhelmed [9].

Hybrid models combining NPS with alternative satisfaction metrics, such as Customer Effort Score (CES) or the American Customer Satisfaction Index (ACSI), provide a more comprehensive measure of CX [39]. However, while these models enhance insight depth, they risk complicating NPS's appeal as a simple, industry-benchmarked tool [2]. Balancing usability with accuracy remains a key challenge to ensure customer feedback mechanisms remain both actionable and representative. Building on the widespread integration of NPS within banking culture, this study aims to enhance its effectiveness by examining it from the customer's perspective. It seeks to not only explore how customers interpret NPS but also investigate the underlying factors driving their responses, particularly in relation to demographic characteristics. The following sections review existing literature that similarly endeavours to understand these dynamics.

3.1 Users perceptions of customers satisfaction metrics

Customer perceptions of satisfaction metrics are shaped by usability, clarity, and alignment with real-world experiences. Research suggests customers favour intuitive feedback systems that require minimal effort and accurately reflect their experiences [41]. However, many satisfaction metrics, including NPS, often fail to meet these expectations, leading to frustration and disengagement [17]. Customers may struggle to understand how their responses contribute to business decisions, particularly when surveys use ambiguous rating scales or lack meaningful follow-up actions.

The one-dimensional nature of some metrics, such as NPS, can prevent customers from expressing the complexity of their experiences, resulting in disengagement or inaccurate responses [45]. HCI research highlights that feedback systems perform best when they allow users to clarify or contextualise their ratings while maintaining ease of use [8]. Without this balance, satisfaction metrics risk being perceived as corporate formalities rather than meaningful tools for improving CX, ultimately reducing their credibility and effectiveness. While the simplicity of NPS makes it an appealing and near-ideal metric for engaging customers, our research seeks to determine the extent to which its ease of use may compromise the accurate capture of nuanced customer sentiment.

3.2 Online Reviews: Their Role in Customer Perception and Decision-Making

The influence of online reviews has grown exponentially, shaping consumer perceptions and decision-making [27]. Studies indicate that over 70% of consumers trust online reviews as much as personal recommendations [1], making them a key factor in purchasing decisions. For industries like banking, where customer trust is critical, satisfaction and loyalty function as both business objectives and powerful marketing tools. The rapid spread of customer opinions via online platforms amplifies word-of-mouth marketing globally [6].

Consumer behaviour research shows that purchase decisions are closely tied to the information available [24]. Online reviews provide this information, helping customers evaluate businesses. Given that 93% of consumers consult online reviews before making decisions [7], their influence on brand perception is significant. Unlike traditional surveys, online reviews offer public, candid feedback at scale, making them an attractive alternative for understanding customer sentiment [12]. Companies increasingly use sentiment analysis to mine these insights for CX improvements [56].

As online reviews gain prominence, companies adjust their strategies to manage public perception. Research suggests that while NPS scores correlate with Likert-style ratings on these platforms [25], their functions differ: NPS is an internal metric for tracking long-term loyalty, whereas online ratings are public and shaped by social proof. Given that nearly half of all consumers avoid businesses rated below four stars [19], online reputation now plays a crucial role in customer engagement. As reliance on online reviews grows, the potential integration of these metrics with NPS presents an opportunity for a more comprehensive CX measure. This research will investigate the relationship between NPS and online reviews to study whether this approach represents a viable method for more comprehensive understanding of customer perceptions.

Generational differences further complicate this evolving relationship. Age significantly impacts NPS scoring patterns [54] and influences how individuals seek recommendations online. Generation Z (Gen Z), for example, conducts more online research before purchasing than previous generations [16], increasing the importance of online reputation management. As Gen Z enters adulthood and makes financial decisions, they become a key demographic for banks and other industries. With 40% of the British population remaining loyal to one bank for life [15], financial institutions must adapt strategies to attract and retain younger consumers. Understanding generational differences in feedback interpretation will be essential in refining NPS strategies and customer engagement efforts. It is for this reason that this research will examine the impact of demographic factors on how customers articulate their opinions, thereby providing deeper insights into the true perceptions underlying all forms of qualitative feedback—not just online reviews.

3.3 Sentiment Analysis and Customer Feedback

Sentiment analysis provides valuable insights into customer emotions as expressed in online reviews. Computational models like VADER classify textual feedback as positive, neutral, or negative, offering a more detailed understanding of satisfaction and loyalty trends [22]. Compared to structured survey responses, sentiment analysis enhances NPS by identifying underlying factors influencing promoters, passives, and detractors [55].

However, discrepancies often arise between NPS classifications and underlying customer sentiment. Customers with limited brand exposure may assign NPS scores based on superficial impressions rather than well-formed opinions, raising concerns about the long-term validity of early-stage NPS scores [14]. This misalignment underscores the importance of incorporating sentiment analysis into NPS evaluations to capture customer sentiment more accurately over time.

The rigid NPS structure further complicates sentiment interpretation. The absence of a "no answer" option forces an indifferent user to select a score between zero and five, classifying them as detractors despite lacking dissatisfaction [28]. Similarly, customers who do not typically engage in brand recommendations may provide scores that fail to reflect their actual satisfaction levels [52]. Without sentiment analysis, companies risk misinterpreting these classifications, leading to skewed insights and flawed decision-making. This research investigates whether sentiment analysis can strengthen NPS's predictive value and provide a more holistic view of customer satisfaction.

4 Methodology

This study employs a mixed-methods approach to provide a comprehensive examination of customer perceptions related to the NPS in the UK banking landscape. A combination of qualitative and quantitative methods was essential to effectively address the research questions, ensuring both statistical validity and contextual depth in the analysis. By integrating survey data with sentiment analysis and online review metrics, this approach enables a multi-faceted investigation into how accurately customers understand NPS classifications *(RQ1)*, the relationship between NPS and traditional online rating systems *(RQ2)*, and the linguistic and tonal characteristics of customer reviews in relation to their NPS scores *(RQ3)*. The mixed-methods approach allows for triangulation across multiple data sources, enhancing the reliability of findings while providing richer insights into user perceptions. Given the complexity of customer interactions with feedback mechanisms, this approach is particularly well-suited to HCI research, where quantitative patterns must be contextualised within user experiences and perceptions. All study procedures were reviewed and approved by Swansea University's Institutional Review Board (IRB).

4.1 Primary Data: Survey

Survey Design and Pilot Study: The survey was distributed through Prolific to collect a representative sample of UK adult population. Considering only 2% of UK adults do not have a bank account,[3] this was fairly representative of the UK banking population, further exclusions included those who have had experiences with 16 of the major UK banks. These banks were selected as they are recognised by the CMA and have publicly available NPS results. The banks included in the study were HSBC, Barclays, Virgin Money, NatWest, Lloyds, Santander, Monzo, Co-Operative Banking, Starling Bank, First Direct, Metro Bank, Halifax, Bank of Scotland, TSB, The Royal Bank of Scotland, and Nationwide.

Before the full survey launch, a pilot study was conducted with 20 participants to ensure the clarity of the survey questions and to confirm the estimated completion time. The pilot highlighted several areas that required improvement. A number of users provided brief responses to the open-ended questions, and many encountered difficulties understanding how to classify their NPS ratings. In response to these findings, we revised the questions. Furthermore, the NPS classification question was revised and divided into two distinct questions. The quantitative question remained but was reworded to emphasise the key instructions with bold font, while an additional question specifically asked participants to classify themselves as promoters, detractors, or passives, each accompanied by clear definitions.

Survey Distribution and Sample: After revisions based on the pilot study, the survey was distributed to 300 participants. The target sample was carefully constructed to be representative of the UK population, which in turn is representative of the UK banking population with strict inclusion criteria ensuring that participants were aged 18 or older, residing in the UK, and customers of one of the 16 aforementioned banks. These criteria ensured that all responses were directly relevant to our research focus. The representative sample consisted of 152 females and 148 males. In terms of ethnicity, the majority identified as White (256), followed by Asian or Asian British (23), Mixed or Multiple Ethnic Groups (10), and Black (8). Employment status varied across the sample, with 118 participants in full-time employment, 54 in part-time roles, 46 retired, 26 self-employed, and 28 unemployed. The remaining participants categorised their employment status or ethnicity as 'Other'.

[3] https://www.fca.org.uk/publications/financial-lives/financial-lives-2023

The resulting sample was mostly representative of the UK population. After comparing our Prolific-collected sample with national demographic data from the Office for National Statistics (ONS),[4] Chi-square goodness-of-fit tests indicated that there were no significant differences in gender, income or ethnicity distributions. However, significant deviations were observed in age and employment status. Specifically, the dataset skews younger and includes fewer full-time employed individuals, with a slight over representation of retired participants. These differences are likely due to the nature of Prolific's participant pool, which is known to attract younger users and individuals who may seek supplemental income—characteristics less common among those in full-time employment.

To minimise social desirability bias, demographic questions were placed after the primary content of the survey [11]. These demographic variables included gender, ethnicity, employment status, annual income, education level, and geographic location until the age of 18, as these factors were hypothesised to influence participants' understanding of NPS. Participants were offered £1.50 for the completion of a 10 minute survey. The monetary incentive provided through Prolific raised concerns about potential response bias, where participants might rush through the survey to obtain the reward. However, Prolific's built-in quality control mechanisms mitigated this risk. Additionally, an attention check question was strategically placed within the demographic section to identify participants who were not responding attentively, further ensuring the integrity and reliability of the collected data.

4.2 Secondary Data: Trustpilot Reviews

A Python script was developed to scrape publicly available reviews from each of the banks Trustpilot pages using Selenium, which is a widely used automated browser interaction tool [37, 49]. Trustpilot was selected as the primary data source for this research due to its extensive collection of publicly available customer reviews and its established role as a key platform for consumer feedback. Its large-scale dataset ensures sufficient volume for robust sentiment analysis while maintaining accessibility. Moreover, Trustpilot's structured rating system and textual review format allow for a direct comparison between structured numerical feedback and unstructured sentiment analysis, making it particularly well-suited for addressing *RQ2* and *RQ3*. Additionally, Trustpilot has been widely used in prior sentiment analysis research [38, 40], enhancing the validity and comparability of this study. Given its influence on consumer decision-making and cross-industry applications, Trustpilot provides an ideal dataset for investigating the relationship between NPS, online ratings, and customer sentiment in retail banking.

4.3 Data Analysis

Qualitative Analysis: For the qualitative analysis of the survey data thematic analysis was used. NVivo software was employed to code and categorise the open-

[4] https://www.ons.gov.uk/peoplepopulationandcommunity

ended responses and was conducted by the first author, having been reviewed by the other authors which agreed with all the generated codes. NVivo enabled the identification of key themes related to customer satisfaction. The identified themes were analysed across different NPS scores, allowing for the comparison of patterns and trends among specific subgroups, particularly when interesting trends were identified within the data.

Tone Analysis: Sentiment analysis was conducted using VADER's SentimentIntensityAnalyzer [22], a widely used tool for sentiment classification [53]. In this study, it was adapted to classify Trustpilot reviews as positive, neutral, or negative, with the compound sentiment score serving as the primary metric. This approach aligned with other averaged variables, allowing for trend identification without separately analysing individual sentiment dimensions. 30 randomly selected reviews were manually reviewed and classified as either positive, neutral or negative and compared to the classifications completed by VADER, resulting in an 80% accuracy. This level of agreement reflects reasonable reliability for identifying sentiment trends in naturally expressed feedback.

Politeness analysis was performed using ConvoKit [10], which assigns binary scores to text indicators: polite (1), neutral (0), and negative (-1). These were aggregated to generate a net politeness score per review. Toxicity was assessed using Detoxify [18], quantifying offensive or aggressive language indicative of negative experiences, including insult levels.

A challenge arose from the presence of emojis in customer reviews, as sentiment analysis tools struggled to interpret them accurately. Since emojis contribute to tone and sentiment, simply removing them would risk losing critical context. To address this, we developed the replace_emojis() function, converting emojis into textual descriptions for improved sentiment interpretation [32, 51].

Both qualitative survey responses and Trustpilot reviews underwent sentiment, politeness, and toxicity analysis using these Python tools. The Trustpilot data was aggregated into six-month periods and correlated with NPS scores for the corresponding intervals. Statistical correlation tests were performed to assess relationships between review sentiment, tone, and NPS scores. Prior research has explored the impact of politeness strategies on customer satisfaction [21] and examined how toxic content can affect overall product sentiment [43]. Building on these studies, our analysis aims to determine if similar patterns are observable within our dataset.

5 Findings

The findings are structured according to the research questions, with each section presenting the most relevant findings. The analysis follows a consistent format: first, results from primary data collection (the survey) are reported, followed by demographic influences, and finally, secondary data findings from online sources. As *RQ1* pertains solely to survey responses, no secondary data is included in that section.

5.1 RQ1: Customer Understanding of NPS Categories

This section reports the findings from the survey before analysing the demographical results which were found to be related to the research question.

Study Results: The collected data revealed that 110/300 (37%) individuals misclassified themselves, believing their NPS score placed them in a different category than it actually did. A Cohen's Kappa value of 0.388 was calculated, despite indicating a fair level of agreement between participants' self-classifications and the actual categories, this is still a high number of participants who failed to fully comprehend how their opinions are being measured by the NPS metric.

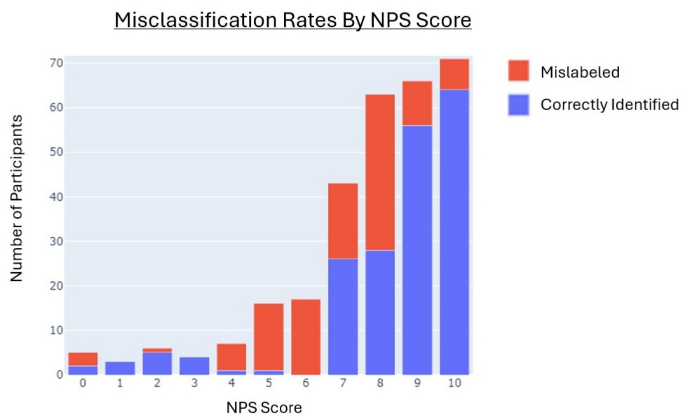

Fig. 1. This graph demonstrates the misclassification rates found split by their corresponding NPS scores.

Given that the Kappa value is closer to zero than to one, implying a greater proximity to random chance than to certainty, it is crucial to consider the factors influencing participants' misjudgements. Among the participants who misclassified their NPS category, 92/110 (82%) believed they belonged to a higher category than their score indicated. Further analysis (see Figure 1), revealed that the NPS score of 8 was the most frequently misinterpreted, with users often believing that this score categorised them as promoters of the brand. Qualitative analysis of these responses uncovered overwhelmingly positive feedback, with frequent references to positive experiences with the bank's app and customer service. A recurring theme was a strong intention to recommend the bank, as 25 of the 34 responses associated with a score of 8 included the word "recommend".

More notably, misclassification was highest within the upper range of the detractor category. None of the participants who provided a score of six were able to accurately identify their corresponding category, and only a small proportion of those who responded with a score of four or five correctly recognised their classification. Furthermore, all users who selected a score of six perceived themselves to belong to a higher NPS category than was actually the case.

An in-depth analysis of these scores revealed that the most frequently cited factor was the willingness to recommend if prompted, combined with a general sense of satisfaction. However, participants didn't perceive the banks as performing exceptionally well. These findings suggest that NPS scores may be artificially deflated and may fail to accurately capture customers' true sentiments, potentially due to a lack of understanding of the metric on the part of the users.

Additionally, it is noteworthy that all of those who misclassified themselves when responding 0 on the NPS scale, which is typically associated with being a detractor, classified themselves as passives. These participants selected 0 because they would not recommend the bank under any circumstances, reflecting a fundamentally different interpretation of the NPS rating system, highlighting a disconnect between their intentions and the NPS metric itself.

Demographics: To investigate whether demographic characteristics influenced the likelihood of NPS misclassification, we conducted a series of statistical tests, including chi-squared tests for categorical variables and ANOVA for the continuous variable of age. Across all variables examined—gender, income, education level, employment status, ethnic background, area lived in until age 18, and age—no statistically significant associations were found (all p-values $> .05$). This suggests that misclassification is not systematically related to demographic factors and may instead reflect a broader misunderstanding of NPS categories across the sample. These findings imply that improved clarity and design of NPS feedback systems may be more impactful than targeting demographic-specific interventions.

5.2 RQ2: NPS and Likert Rating Relationship

Survey Results: To investigate the potential relationship between Likert scale ratings commonly used on online review platforms and NPS, the survey included two questions: one asked participant to rate their bank on a scale of one to five (Likert scale), while the other followed the traditional NPS format, asking participants to rate their likelihood of recommending the bank on a scale of zero to ten. Pearson and Spearman correlation tests were applied to compare each participant's score across both scales. The results yielded a correlation coefficient of 0.761, with a p-value of <0.001, indicating a significant positive correlation between the two scoring systems (see Figure 2). This suggests that as the Likert score increases, so does the corresponding NPS score.

Demographics: Interestingly, it was only the age demographics that revealed variation. While most age groups showed strong and significant correlations, participants aged 25-29 and 30-34 demonstrated no statistically significant correlations. This contrasts with other age groups, which generally showed significant and strong correlations, particularly those aged 40-44, 45-49, and 50-54. A detailed breakdown of correlations by age is presented in Table 1.

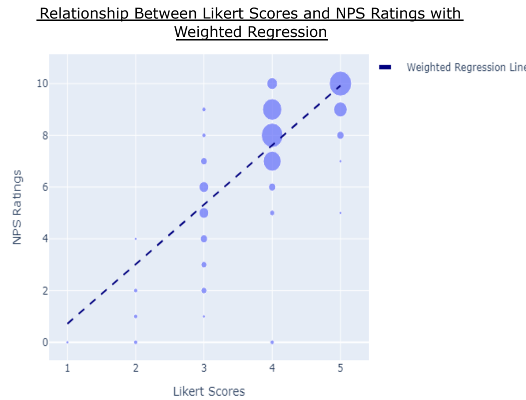

Fig. 2. This graph visualises the relationship between Likert Scores and NPS Ratings, with the size of the bubbles representing the frequency of occurrences. A weighted regression line is included to highlight the trend in the data, indicating a positive correlation between the two variables.

Table 1. Correlation and P-Values by Age Group

Age Group	Correlation (R)	P-Value
18-24	0.552	$<$0.05
25-29	0.614	0.079
30-34	0.186	0.632
35-39	0.514	$<$0.05
40-44	0.957	$<$0.001
45-49	0.940	$<$0.001
50-54	0.874	$<$0.001
55-59	0.800	$<$0.001
60-64	0.757	$<$0.001
65-69	0.866	$<$0.001

These results indicate that while there is generally a strong correlation between Likert and NPS scores across various demographics, certain groups — particularly some age ranges — show weaker or no statistically significant correlations, suggesting that factors beyond age may influence how individuals perceive and report their satisfaction across different scoring systems. However, the strength of the correlation especially in the age groups of 40-44 and 45-49 makes this a particularly interesting finding.

Publicly Available Data: The Pearson correlation test and Spearman Correlation test both returned no statistically significant correlations suggesting that there is no relationship in this instance. A key consideration is the temporal mismatch between the datasets, which may influence comparability. While CMA NPS scores are collected and published biannually, Trustpilot reviews were aggregated over a six-month period to align with this cycle. This approach, while necessary, could blur specific trends and reflect variations due to shifting sentiment, campaign activity, or isolated events within that time frame. As such, public review data should be interpreted with caution when used alongside formal NPS scores, given the differing contexts and collection methods.

5.3 RQ3: NPS Relationship to Language and Tone

Study Results: This question will focuses on analysing the qualitative data that was collected during this study. In the study this was the question which asked the individual to describe their Likert rating and the question which asked them to elaborate on their NPS score. In the secondary data we analysed the reviews left online. Every qualitative piece of data was analysed using the same python packages and for the secondary data the scores were combined to create six month averages which were comparable to the NPS scores used.

Toxicity and Politeness – The Spearman correlation analysis revealed a very weak but significant negative relationship between the toxicity of the Likert Explanation and the Likert Rating, with a Spearman coefficient of -0.169 ($p = 0.003$). Similarly, a slightly stronger but still low negative correlation was observed between the toxicity of the NPS Explanation and the NPS Rating, with a Spearman coefficient of -0.250 ($p < 0.001$). Although these results indicate that as the toxicity in explanations increases, both Likert and NPS ratings tend to decrease, the correlations are low suggesting that toxic language is not a strong indicator of low customer satisfaction on its own. When examining the correlation between politeness in both Likert and NPS explanations and NPS scores, no statistically significant results were found. Therefore similar to toxicity, these findings suggest that politeness does not indicate customer satisfaction.

Sentiment – The correlation between compound sentiment in explanations and NPS scores demonstrated a moderately positive relationship. For NPS explanations, the correlation coefficient was low but present with 0.23 ($p < 0.001$). In contrast, the correlation between Compound Sentiment in Likert explanations and NPS scores was stronger, with a coefficient of 0.44 ($p < 0.001$). These results suggest that more positive sentiment in customer explanations is strongly associated with higher NPS scores, with Likert explanations exhibiting an especially pronounced positive relationship.

Demographics – Income levels were more closely related to differences in tone. Participants in the lowest income category exhibited significantly higher levels of toxicity and insult in their Likert explanations. Conversely, participants in the highest income category showed the lowest toxicity and insult levels across both types of explanations, indicating a more neutral or polite tone. Interestingly, participants in the £70,000-£79,999 income bracket had the highest levels of insult in their Likert explanations, suggesting income alone doesn't predict tone consistently. Employment status also influenced the tone of explanations. Participants who were unemployed or students showed higher levels of toxicity in both NPS and Likert explanations compared to those who were employed full-time. In contrast, retired participants demonstrated lower levels of toxicity, particularly in NPS explanations, indicating that life experience may contribute to more polite or neutral language.

Gender differences were also apparent, with female participants generally showing lower levels of toxicity and insult. Age-related differences in tone were also observed. Participants aged 40-44 had the highest overall insult levels in both NPS and Likert explanations whilst participants aged 50-54 had the lowest insult scores, indicating a more polite tone overall. Participants aged 65-69 demonstrated lower insult levels and higher politeness, suggesting that older individuals may be more likely to use neutral or positive language.

Publicly Available Data: ***Toxicity and Politeness*** – The Spearman correlation analysis of the web-scraped data revealed a statistically significant negative correlation between toxicity levels in customer reviews and NPS, with a medium correlation coefficient of -0.31 ($p = 0.001$). This indicates that as toxicity in reviews increases, NPS scores tend to decrease. In contrast, a positive correlation was observed between politeness in reviews and NPS scores, with a stronger Spearman correlation coefficient of 0.43 ($p < 0.001$). This finding suggests that in the online space, more polite reviews are associated with higher NPS scores and more toxic reviews are linked to lower NPS scores.

Sentiment – The overall correlation between compound sentiment and NPS scores across all banks revealed a moderate positive relationship. A high Spearman correlation coefficient was found at 0.52, ($p < 0.001$). This suggests that as the sentiment of customer reviews becomes more positive, NPS scores tend to increase. These results suggest compound sentiment to be a strong predictor of NPS.

6 Discussion and Implications

Below we discuss the salient insight gained from our findings.

6.1 Customer Misclassification in NPS Responses

The overall misclassification rate among participants indicates only a "fair" agreement between perceived and actual NPS categories. While better than random chance, this level of agreement suggests that customers frequently misunderstand the distinctions between promoters, passives, and detractors. Given that NPS scores are highly sensitive to classification errors [28], such misinterpretations can significantly distort the intended measurement of customer sentiment. The most common error involved overestimation, where participants perceived themselves as more loyal or satisfied than their numerical score would indicate. This misalignment was reinforced by qualitative data in which customers who misclassified themselves used positive language and words like "recommend," despite their assigned NPS category suggesting otherwise.

These results highlight the need for a more user-centred approach to NPS data collection, as companies continue to rely on this metric for key performance evaluations [41]. A potential solution is the introduction of visual aids to help users better grasp what each score represents, improving internalisation and reducing misclassification errors [26, 9].

Additionally, users who rated their bank a "0" but self-identified as passives demonstrated a fundamental misunderstanding of the implications of their score. This is particularly relevant in industries like banking, where recommending a provider is not a common behaviour due to the complexity and stakes involved in financial decisions [4]. The assumption that all customers understand or engage with the concept of brand advocacy in the same way seems flawed, reinforcing the argument that NPS should offer a "no response" option [29]. However, this approach may reduce usable data, making an alternative—such as emphasising the hypothetical nature of the recommendation question—a more viable solution.

6.2 Correlation Between NPS and Traditional Online Ratings

A strong correlation between NPS and Likert scale ratings echoes prior research indicating that both reflect overall CX [25]. While Likert measures specific satisfaction aspects and NPS assesses brand advocacy, customers use both in similar ways [50]. However, analysis of Trustpilot data did not show a statistically significant correlation. This is likely due to selection bias being inherent in open review platforms like Trustpilot, where contributors are typically self-selected and disproportionately represent those with highly polarised experiences — often the most satisfied or dissatisfied customers [20]. Also, response motivation differs substantially, NPS responses are usually prompted and structured within a specific feedback window, whereas Trustpilot reviews are voluntary, public-facing, and often emotionally charged.

6.3 Emotional Tone and Customer Sentiment in NPS Responses

Toxic language strongly correlated with lower NPS scores, reinforcing findings that negative emotional expressions—such as frustration or anger—are linked to lower satisfaction ratings [12]. Since toxicity in both Likert and NPS explanations affected ratings similarly, organisations can use sentiment analysis to monitor customer dissatisfaction and intervene proactively. AI-driven tools can detect trends in negative feedback, allowing businesses to enhance service quality and customer retention.

While toxicity indicated dissatisfaction, politeness showed no significant correlation with NPS, suggesting that courteous language does not necessarily reflect satisfaction or loyalty [31]. In some cases, politeness may even mask criticism. Sarcasm, particularly in British culture, further complicates sentiment analysis, as polite phrasing can disguise negative sentiment [30]. This complicates automated sentiment analysis, as sarcasm can distort conventional politeness metrics, making it harder to extract true customer sentiment from text-based feedback. To further examine the impact of such cultural influences, future work should replicate this work across diverse cultural contexts, thereby elucidating the extent to which British sarcasm contributes as a significant factor. This highlights just one example of how cultural nuances — such as British politeness and sarcasm, can shape NPS responses, demonstrating why understanding NPS through emotinoal tone cannot be reliably generalised beyond the specific cultural context in which it is studied.

6.4 Influence of Demographics on Sentiment Expression

Demographics played a key role in how customers expressed sentiment. Lower-income participants exhibited higher toxicity and insult levels, while higher-income users used more neutral or polite language. Socioeconomic background influences communication style, with individuals from lower-income areas more likely to use direct, emotionally charged language [35]. This can lead to misinterpretations in sentiment analysis, as blunt feedback may be mistakenly classified as overly negative.

Generational differences also shaped sentiment expression. Millennials and Gen Z use more unfiltered language, including profanity, without necessarily indicating dissatisfaction [23]. In contrast, older generations view such language as negative [36]. This generational gap suggests that sentiment analysis must be adapted to consider evolving communication styles when assessing feedback.

As AI tools become increasingly prevalent in CX management, there's a risk that critical factors—such as the influence of demographic characteristics on sentiment expression—may be overlooked. Consequently, trends analysed without accounting for these variables may be misinterpreted, leading to poor management decisions that could ultimately prove detrimental to business performance.

6.5 Limitations

We addressed dataset-specific constraints by incorporating both primary and secondary data, which was essential for answering our research questions. While NPS effectiveness varies across industries and our findings are specific to the banking sector, they still provide valuable insights into customer satisfaction within this context. Additionally, computational linguistic tools for sentiment analysis face challenges in detecting nuances such as sarcasm and cultural differences, which can lead to misclassification. To mitigate this limitation, we employed multiple sentiment measures—compound sentiment, toxicity, and politeness—offering a more nuanced perspective on perceptions of customer satisfaction to strengthen the reliability of our conclusions. Finally, using Prolific may have introduced some bias regarding the sample leaning to slightly younger participants; however, it was the most practical and efficient method for data collection within the project's limited time frame.

7 Conclusion

This study contributes to the literature in three key areas. First, it fills an important gap by comparing customers' perceived satisfaction with their actual NPS classifications, revealing that 37% misclassified themselves, often overestimating their status. This misalignment, highlights the need for greater transparency in NPS interpretation and adaptive survey designs that account for user literacy and provide in-the-moment explanation of scoring consequences.

Second, the study challenges the assumption that NPS is universally understood by those interacting with it in the UK banking sector, reinforcing concerns that many UK banking customers misunderstand how their responses are categorised. For HCI practitioners, this underscores the importance of interface transparency and user-centred explanations in feedback tools—particularly in industries where customer trust and clarity are paramount such as banking but with potential to expand to other similar industries.

Finally, by analysing sentiment in both structured NPS surveys and online reviews, we highlight the role of emotional expression in customer feedback. While NPS correlated well with Likert scales, publicly available Trustpilot data showed a weaker correlation with no statistical significance, indicating contextual differences in customer engagement. Additionally, sentiment analysis revealed higher toxicity linked to lower scores and politeness sometimes masking dissatisfaction, reinforcing the need for context-sensitive satisfaction evaluation methods. These insights suggest that sentiment-aware feedback interfaces—capable of detecting linguistic cues such as politeness or emotional intensity—should offer real-time prompts or clarifications to help users accurately express their experiences, ultimately improving the reliability and interpretability of the feedback collected.

Further studies could investigate the implications identified here, particularly the role of visual cues in improving NPS response accuracy. A key question is whether such prompts help users better align their responses with their true sentiments or risk introducing bias by subtly guiding choices. Controlled experiments could explore how users interact with these elements, assessing whether they enhance comprehension or distort feedback, supported by in-depth interviews and follow-ups to more accurately capture participants' intended meanings and interpretations. Additionally, future research could examine the use of real-time sentiment analysis to proactively address dissatisfaction before it impacts NPS scores, and explore how socio-technical characteristics can be meaningfully integrated into CX measurement. To advance inclusivity and accuracy, future work should also consider how personalised feedback systems can accommodate individual differences in communication styles. Crucially, we advocate for a more human-centred approach to the design of customer satisfaction measures. This includes incorporating qualitative interviews and co-design workshops, enabling stakeholders to collaboratively shape feedback mechanisms that reflect natural user expression while balancing conceptual rigour and practical usability.

Acknowledgements

This work was funded by EPSRC grant number EP/S021892/1 and HSBC UK Bank PLC. We thank the study participants for their time and effort. We are grateful to anonymous reviewers who helped us further improve the paper.

References

1. Anderson, E.: Customer satisfaction and word of mouth. Available at: https://journals.sagepub.com/doi/10.1177/109467059800100102 (1998)
2. Baehre, S., O'Dwyer, M., O'Malley, L., Lee, N.: The use of net promoter score (nps) to predict sales growth: Insights from an empirical investigation. Journal of the Academy of Marketing Science **50**(1), 67–84 (2022). https://doi.org/10.1007/s11747-021-00790-2
3. Baquero, S.D., Castañeda, J.A., Gómez, M.: Does net promoter score reflect real customer satisfaction? evidence from the hotel industry. Sustainability **14**(4), 2011 (2022). https://doi.org/10.3390/su14042011
4. Bennett, R., Kottasz, R.: Public attitudes towards the uk banking industry following the global financial crisis. International Journal of Bank Marketing **30**(2), 128–147 (2012). https://doi.org/10.1108/02652321211210877
5. Brandt, D.R.: For good measure. Marketing Management **16**(1), 20–25 (2007)
6. Busalim, A., Hollebeek, L.D., Lynn, T.: The effect of social commerce attributes on customer engagement: An empirical investigation. Internet Research **34**(7), 187–214 (2023). https://doi.org/10.1108/INTR-03-2022-0165
7. Chen, T., Samaranayake, P., Cen, X., Qi, M., Lan, Y.C.: The impact of online reviews on consumers' purchasing decisions: Evidence from an eye-tracking study. Frontiers in Psychology **13** (2022). https://doi.org/10.3389/fpsyg.2022.865702
8. Chevalier, J.A., Mayzlin, D.: The effect of word of mouth on sales: Online book reviews. Journal of Marketing Research **43**(3), 345–354 (2006). https://doi.org/10.1509/jmkr.43.3.345
9. Couper, M.P., Traugott, M.W., Lamias, M.J.: Web survey design and administration. Public Opinion Quarterly **65**(2), 230–253 (2001). https://doi.org/10.1086/322199
10. Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., Potts, C.: A computational approach to politeness with application to social factors. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 250–259 (2013), https://aclanthology.org/P13-1025
11. Dillman, D.A., Smyth, J.D., Christian, L.M.: Internet, Phone, Mail, and Mixed Mode Surveys: The Tailored Design Method, 4th Ed. John Wiley & Sons Inc, Hoboken, NJ, US (2014)
12. Fan, W., Liu, Y., Li, H., Tuunainen, V.K., Lin, Y.: Quantifying the effects of online review content structures on hotel review helpfulness. Internet Research **32**(7), 202–227 (2021). https://doi.org/10.1108/INTR-11-2019-0452
13. Fisher, N.I., Kordupleski, R.E.: Good and bad market research: A critical review of net promoter score. arXiv preprint arXiv:1806.10452 (2018), https://arxiv.org/abs/1806.10452
14. Fisher, N.I., Kordupleski, R.E.: Good and bad market research: A critical review of net promoter score. Applied Stochastic Models in Business and Industry **35**(1), 138–151 (2019). https://doi.org/10.1002/asmb.2417
15. Gordon, J.: 40% of brits stay loyal to one bank for their entire lives (2020), https://www.theinvestmentobserver.co.uk/money/2020/09/25/40-of-brits-stay-loyal-to-one-bank-for-their-entire-lives/
16. Grigoreva, E.A., Garifova, L.F., Polovkina, E.A.: Consumer behavior in the information economy: Generation z. International Journal of Financial Research **12**(2), 164 (2021). https://doi.org/10.5430/ijfr.v12n2p164

17. Grisaffe, D.B.: Questions about the ultimate question: Conceptual considerations in evaluating reichheld's net promoter score (nps). Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior **20**, 36–53 (2016)
18. Hanu, L., Team, U.: Detoxify (2020). https://doi.org/10.5281/zenodo.7925667
19. Howarth, J.: 81 online review statistics (new 2024 data) (2022), https://explodingtopics.com/blog/online-review-stats, accessed: 2024-09-28
20. Hu, N., Pavlou, P.A., Zhang, J.J.: Overcoming the j-shaped distribution of product reviews (2009), https://papers.ssrn.com/abstract=2369332
21. Hu, Y., Huang, D., Kim, H., Koh, R.: Read this, please? the role of politeness in customer service engagement on social media. Proceedings of the 52nd Hawaii International Conference on System Sciences pp. 1–10 (2019)
22. Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 8, pp. 216–225 (2014). https://doi.org/10.1609/icwsm.v8i1.14550
23. Jay, T., Janschewitz, K.: The pragmatics of swearing. Journal of Politeness Research. Language, Behaviour, Culture **4**(2) (2008). https://doi.org/10.1515/JPLR.2008.013
24. Karimi, S., Holland, C.P., Papamichail, K.N.: The impact of consumer archetypes on online purchase decision-making processes and outcomes: A behavioural process perspective. Journal of Business Research **91**, 71–82 (2018). https://doi.org/10.1016/j.jbusres.2018.05.038
25. Keiningham, T.L., Cooil, B., Andreassen, T.W., Aksoy, L.: A longitudinal examination of net promoter and firm revenue growth. Journal of Marketing **71**(3), 39–51 (2007). https://doi.org/10.1509/jmkg.71.3.039
26. Khandelwal, M.: Expert advice on the impact of color coding the nps scale (2023), https://www.surveysensum.com/blog/color-coding-nps-scale, accessed: 2024-09-28
27. Kim, W.G., Park, S.A.: Social media review rating versus traditional customer satisfaction: Which one has more incremental predictive power in explaining hotel performance? International Journal of Contemporary Hospitality Management **29**(2), 784–802 (2017). https://doi.org/10.1108/IJCHM-11-2015-0627
28. Kristensen, K., Eskildsen, J.: Is the net promoter score a reliable performance measure? In: 2011 IEEE International Conference on Quality and Reliability. pp. 249–253 (2011). https://doi.org/10.1109/ICQR.2011.6031719
29. Kristensen, K., Eskildsen, J.: Is the nps a trustworthy performance measure? The TQM Journal **26**(2), 202–214 (2014). https://doi.org/10.1108/TQM-03-2011-0021
30. Kádár, D.Z.: Politeness, Impoliteness and Ritual: Maintaining the Moral Order in Interpersonal Interaction. Cambridge University Press (2017). https://doi.org/10.1017/9781107280465
31. Kádár, D.Z., Haugh, M.: Politeness as social practice. In: Understanding Politeness, pp. 57–80. Cambridge University Press (2013). https://doi.org/10.1017/CBO9781139382717.005
32. LeCompte, T., Chen, J.: Sentiment analysis of tweets including emoji data. In: 2017 International Conference on Computational Science and Computational Intelligence. IEEE (2017). https://doi.org/10.1109/CSCI.2017.137, https://ieeexplore.ieee.org/document/8265116
33. Lewis, C., Mehmet, M.: Does the nps® reflect consumer sentiment? a qualitative examination of the nps using a sentiment analysis approach. International Journal of Market Research **62**(1), 9–17 (2020). https://doi.org/10.1177/1470785319863623

34. Liang, X., Yu, Y., Socolinsky, D., Mao, T., Rojas, C., Goryainov, A., Wang, F., Chen, X.: An end-to-end solution for net promoter score estimation and explanation from social media using natural language processing. In: Fuzzy Systems and Data Mining IX, pp. 54–66. IOS Press (2023), https://doi.org/10.3233/FAIA231006
35. Liu, R.W., Lapinski, M.K., Kerr, J.M., Zhao, J., Bum, T., Lu, Z.: Culture and social norms: Development and application of a model for culturally contextualized communication measurement (mc3m). Frontiers in Communication **6** (2022). https://doi.org/10.3389/fcomm.2021.770513
36. Livingstone, S.: igen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy – and completely unprepared for adulthood. Journal of Children and Media **12**(1), 118–123 (2018). https://doi.org/10.1080/17482798.2017.1417091
37. Loke, R.E., Kachaniuk, O.: Sentiment polarity classification of corporate review data with a bidirectional long-short term memory (bilstm) neural network architecture. In: Proceedings of the 9th International Conference on Data Science, Technology and Applications. pp. 310–317. SCITEPRESS – Science and Technology Publications, Lda. (2020). https://doi.org/10.5220/0009892303100317
38. Loke, R.E., Pathak, S.: Decision support system for corporate reputation based social media listening using a cross-source sentiment analysis engine. In: Proceedings of the 12th International Conference on Data Science, Technology and Applications. pp. 559–567. SCITEPRESS – Science and Technology Publications, Lda. (2023). https://doi.org/10.5220/0012136400003541
39. Melnic, E.: The science of customer satisfaction in the retail banking system - a critical comparison between the two international indexes: Net promoter score (nps) and american customer satisfaction index (acsi) (2016), https://www.etimm.ase.ro/?p=109
40. Miracula, V., Picone, A.: Unleashing the power of user reviews: Exploring airline choices at catania airport, italy. arXiv **2306.15541** (2023), https://arxiv.org/abs/2306.15541v1
41. Morgan, N.A., Rego, L.L.: The value of different customer satisfaction and loyalty metrics in predicting business performance. Marketing Science **25**(5), 426–439 (2006). https://doi.org/10.1287/mksc.1050.0180
42. Morgeson, F.V., Hult, G.T.M., Sharma, U., Fornell, C.: The american customer satisfaction index (acsi): A sample dataset and description. Data in Brief **48**, 109123 (2023). https://doi.org/10.1016/j.dib.2023.109123
43. Mukhopadhyay, M., Sahney, S.: Effect of toxic review content on overall product sentiment. Journal of Marketing Theory and Practice (2022). https://doi.org/10.1177/10591478241309439
44. Müller, S., Seiler, R., Völkle, M.: Should net promoter score be supplemented with other customer feedback metrics? an empirical investigation of net promoter score and emotions in the mobile phone industry. International Journal of Market Research **66**(2–3), 303–320 (2024), https://doi.org/10.1177/14707853231219648
45. Parasuraman, A., Berry, L.L., Zeithaml, V.A.: Refinement and reassessment of the servqual scale. Journal of Retailing **67**(4), 420–450 (1991)
46. Patel, N.: The benefits and importance of customer satisfaction, https://widewebadvisor.com/wp-content/uploads/2019/01/The-Benefits-and-Importance-of-Customer-Satisfaction/1851/-1.pdf, accessed: 2024-09-28
47. Pechter, J., Kuusik, A.: Nps from the customer's perspective: The influence of the recent experience. International Journal of Market Research **66**(2–3), 261–277 (2024). https://doi.org/10.1177/14707853231214188

48. Pingitore, G., Morgan, N., Rego, L., Gigliotti, A., Meyers, J.: The single-question trap. Marketing Research **19**(2), 8–13 (2007)
49. Reddy, D., Singh, A., Chopra, R., Patel, R.: Leveraging machine learning algorithms and natural language processing for ai-enhanced social media marketing analytics. Journal of AI and Marketing (2023), https://joaimlr.com/index.php/v1/article/view/28, preprint
50. Reichheld, F.: The one number you need to grow. Harvard Business Review (December 2003), https://hbr.org/2003/12/the-one-number-you-need-to-grow
51. Sakode, H., Surekha, T.L., Kumar, G.B., Sri, M.D.: Sentiment analysis using text and emojis. Proceedings of the International Conference on Inventive Computation Technologies (ICICT 2023) (2023). https://doi.org/10.1109/ICICT57646.2023.10134459
52. Schulman, A., Sargeant, K.: Measuring donor loyalty beyond net promoter score (2013), https://www.nonprofitpro.com/article/measuring-donor-loyalty-beyond-net-promoter-score/
53. Sha, Y., Micallef, N., Wu, Y.: Dissecting the advocacy discourse behind the #StopAsianHate movement on X/Twitter. In: Social Networks Analysis and Mining. pp. 211–229. Springer Nature Switzerland, Cham (2025)
54. Situmorang, S.H.: Gen c and gen y: Experience, net emotional value and net promoter score. In: Proceedings of ICOSOP-16. pp. 259–265 (2016). https://doi.org/10.2991/icosop-16.2017.38
55. Thelwall, M.: Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology **63**(1), 163–173 (2012). https://doi.org/10.1002/asi.21662
56. Walter, V., Kölle, M., Collmar, D.: Measuring the wisdom of the crowd: How many is enough? PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science **90**(3), 269–291 (2022). https://doi.org/10.1007/s41064-022-00202-2