SUPPLEMENTARY MATERIAL FOR PERSISTENT HOMOLOGY IN TOURISM: UNLOCKING THE POSSIBILITIES

A Logistic Regressions

In the main paper results sections we make reference to the results of logistic regressions as an alternative way of identifying candidates for being in the top 10%, and bottom 10% of holiday makers ranked by expenditure. Here we present a series of logistic regression models for predicting whether a particular individual will be in the top 10% (bottom 10%) of spenders amongst inbound UK tourists. We provide inference on the results directing contrast with the main paper.

Following the implementation in R (R Core Team, 2018) logistic regression is specified according to equation (1) as:

$$Log\left(\frac{p_i}{1-p_i}\right) = \beta_j staycat_j + \beta_k Age_k + \beta_m Male + \beta_n Air + \beta_p Persons_p + \beta_q Nation_q + \epsilon_i \tag{1}$$

where $i = \{T, L\}$ denotes the top and lowest 10% of expenditure respectively. p_i is then the probability of being in group i. In the full model we use the six characteristic variable sets and have a parameter matrix β for each. Our first set of data concerns the length of stay, which is categorised into $staycat_j$, $j \in [1, 6]$ according to Table 1 of the main paper, and the categories set out in Table A1 . Age is grouped into eight categories, again as described in the results tables that follow (Table A2), with associated β_k parameters. Male is a dummy which takes the value one if the respondent is male, and hence β_m is a single coefficient value. Likewise β_n is a single value for the impact of using air transportation when leaving the UK. We have six group sizes captured in $Persons_p$ with associated coefficients on each β_p. Finally for the nations which provide larger numbers of tourists we have a vector of coefficients β_q applying to the countries and regions in the vector $Nation_q$[1]. ϵ_i is a white noise error process.

Across the following subsections we work sequentially through the six major sets of control variables used in the Persistent Homology (PH). In each case we report coefficients from the full estimation of model (1). This allows direct comparisons of the effects and a clarity of story to emerge.

A.1 Length of Stay

Our first consideration is the length of time that the respondent stays within the UK. Amongst these β_j coefficients we can see clearly that short stayers are no more, or less,

[1] The precise number of categories in the top 10% and bottom 10% are different as we restrict the Persistent Homology (PH) to either the nationalities of more than 40 respondents in the sample and a regional group that captures the remaining visitors from each of 10 regions. The specific nations are listed in Table A6.

Table A1: Length of stay and expenditure prediction

	Stay Only		Full Model	
	Top 10%	Bottom 10%	Top 10%	Bottom 10%
(Intercept)	-4.773^{***}	-0.575	-5.408^{***}	0.602
	(0.294)	(0.467)	(0.356)	(0.500)
0-2 Days	0.295	-0.235	0.147	-0.109
	(0.238)	(0.466)	(0.269)	(0.479)
3-5 Days	1.554^{***}	-2.183^{***}	1.317^{***}	-1.946^{***}
	(0.301)	(0.470)	(0.329)	(0.484)
6-8 Days	1.505^{***}	-0.644	0.962^{***}	-0.513
	(0.255)	(0.479)	(0.286)	(0.493)
9-12 Days	1.130^{***}	-0.464	0.772^{**}	-0.221
	(0.255)	(0.503)	(0.288)	(0.517)
13-18 Days	0.756^{**}	-0.198	0.916^{**}	-0.252
	(0.256)	(0.520)	(0.289)	(0.533)
19-27 Days	0.310	0.395	-0.023	0.508
	(0.274)	(0.548)	(0.311)	(0.562)
28-30 Days	1.087^{*}	-0.165	0.909	0.001
	(0.429)	(0.855)	(0.484)	(0.877)
AIC	7913.727	8479.104	6666.065	7882.130
BIC	7974.602	8539.979	7000.875	8178.893
Log Likelihood	-3948.863	-4231.552	-3289.032	-3902.065
Deviance	7897.727	8463.104	6578.065	7804.130
Num. obs.	14903	14903	14903	14903

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Notes: Logisitic regressions predicting the probability of an individual with the given stay duration being in the top 10% of spenders amongst tourists visiting the UK. Coefficients are reported for their effects on the odds ratio with the longest stayers (one month or more) as the omitted category. All data from ONS (2017). Regressions performed using glm in R (R Core Team, 2018). Signficance given by $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

likely to be in either the top 10% or bottom 10%. For those staying a few days (3-5 days) they are less likely to be in the bottom 10% and more likely to be in the top decile relative to the longest stayers. A similar observation is made for the next three categories (6-8 days, 9-12 days and 13-18 days), though the negative β_j's in the bottom 10% regressions are not significant. Outwith, this those staying almost one month are more likely to be amongst the highest spenders than those staying more than a month. Given that we are considering total expenditure, many of these observations may seem at odds with what would normally be expected. However the positive association between stay duration and expenditure is often weak.

Table A2: Age categories and expenditure prediction

	Age only		Full model	
	Top 10%	Bottom 10%	Top 10%	Bottom 10%
(Intercept)	-2.046^{***}	-1.833^{***}	-5.408^{***}	0.602
	(0.076)	(0.070)	(0.356)	(0.500)
Under 16 Individual	-2.177^{***}	0.472^{***}	-1.290^{**}	-0.364^{*}
	(0.364)	(0.127)	(0.401)	(0.155)
17-24 Individual	-0.564^{***}	-0.091	-0.701^{***}	-0.108
	(0.121)	(0.100)	(0.147)	(0.115)
17-24 Party	-0.891^{**}	0.162	-0.663^{*}	-0.461^{*}
	(0.285)	(0.179)	(0.333)	(0.209)
25-34	-0.241^{*}	-0.497^{***}	-0.290^{*}	-0.539^{***}
	(0.098)	(0.094)	(0.120)	(0.108)
35-44	-0.175	-0.493^{***}	-0.061	-0.644^{***}
	(0.098)	(0.096)	(0.119)	(0.107)
45-54	-0.007	-0.521^{***}	0.137	-0.682^{***}
	(0.096)	(0.096)	(0.114)	(0.108)
55-64	0.315^{**}	-0.637^{***}	0.320^{**}	-0.604^{***}
	(0.102)	(0.115)	(0.120)	(0.127)
AIC	9564.892	9773.082	6666.065	7882.130
BIC	9625.767	9833.956	7000.875	8178.893
Log Likelihood	-4774.446	-4878.541	-3289.032	-3902.065
Deviance	9548.892	9757.082	6578.065	7804.130
Num. obs.	14903	14903	14903	14903

Notes: Logisitic regressions predicting the probability of an individual witth the stated age being in the top 10% of spenders amongst tourists visiting the UK. Coefficients are reported for their effects on the odds ratio with the youngest group od respondents as the omitted category. All data from ONS (2017). Regressions performed using glm in R (R Core Team, 2018). Signficance given by $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

A.2 Age Categories

For age we use the oldest age category, 65 years and older, as the reference category. Both the youngest category and the youth groups unsurprisingly suggest significant lower probabilities of being in the top 10%. For the bottom 10% dependency on the model specification is noted, whether the full set of characteristics are included or not. Older age groups: 25 to 34, 45 to 44 and 45 to 54, are less likely to be amongst the lowest spenders. Respondents aged between 55 and 64 are more likely to be in the top 10% and less likely to spend in the bottom decile. Again this result is unsurprising since 55 to 64 is typically premium earning age with reduced dependency from their children who are likely to be of working age themselves.

Table A3: Gender and expenditure prediction

	Gender Only		Full Model	
	Top 10%	Bottom 10%	Top 10%	Bottom 10%
(Intercept)	-2.290^{***}	-2.140^{***}	-5.408^{***}	0.602
	(0.040)	(0.037)	(0.356)	(0.500)
Male	0.187^{***}	-0.051	0.188^{**}	-0.095
	(0.055)	(0.054)	(0.065)	(0.060)
AIC	9693.619	9883.148	6666.065	7882.130
BIC	9708.838	9898.367	7000.875	8178.893
Log Likelihood	-4844.810	-4939.574	-3289.032	-3902.065
Deviance	9689.619	9879.148	6578.065	7804.130
Num. obs.	14903	14903	14903	14903

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

A.3 Gender

This single variable model considers the effect of the respondent being Male on their presence in the two categories of expenditure being considered in this paper. As Table A3 attests, males are more likely to be in the top 10%. For the lowest spenders the effect of gender is insignificant. These conclusions apply with near identical strength irrespective of the number of controls included.

A.4 Departure transportation mode

In another analysis of a dummy variable from the dataset we see that the probability of departing by air and being in the top 10% of spenders is greater than if leaving by sea. By contrast sea departures have a higher probability of being in the lowest 10%. Table A4 demonstrates these results clearly. In both cases the magnitude of the effects is reduced when all of the other variables are included as controls.

A.5 Group size

We use lone travellers as the reference category for group size. Of all the other group sizes few are statistically more likely to be in either the bottom, or top, deciles of expenditure; only groups of six or more people are more likely to be in the bottom 10% than lone travellers. Extending to the full set of controls, groups of four are also more likely to be in that lowest decile. For the highest 10% groups of 2,3 or 4 respondents are all less likely to be amongst that highest spending group compared to solo travellers. Noting that expenditure is not calculated on a per-person basis informs that the lower spending behaviour of larger parties comes from group behaviour whilst travelling; PH as employed in the main paper helps break this down further.

Table A4: Departure transportation and expenditure prediction

	Stay Only		Full Model	
	Top 10%	Bottom 10%	Top 10%	Bottom 10%
(Intercept)	-3.331^{***}	-1.256^{***}	-5.408^{***}	0.602
	(0.090)	(0.040)	(0.356)	(0.500)
Air departure	1.356^{***}	-1.434^{***}	0.624^{***}	-1.196^{***}
	(0.094)	(0.055)	(0.109)	(0.075)
AIC	9418.602	9227.965	6666.065	7882.130
BIC	9433.821	9243.184	7000.875	8178.893
Log Likelihood	-4707.301	-4611.983	-3289.032	-3902.065
Deviance	9414.602	9223.965	6578.065	7804.130
Num. obs.	14903	14903	14903	14903

Notes: Logisitic regressions predicting the probability of an individual departing the UK by air being in the top 10% of spenders amongst tourists visiting the UK. Coefficients are reported for their effects on the odds ratio. All data from ONS (2017). Regressions performed using glm in R (R Core Team, 2018). Signficance given by $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A5: Group size and expenditure prediction

	Group size only		Full model	
	Top 10%	Bottom 10%	Top 10%	Bottom 10%
(Intercept)	-1.981^{***}	-2.006^{***}	-5.408^{***}	0.602
	(0.043)	(0.044)	(0.356)	(0.500)
2 People	-0.194^{**}	-0.412^{***}	-0.514^{***}	-0.219^{**}
	(0.061)	(0.065)	(0.076)	(0.076)
3 People	-0.552^{***}	-0.265^{**}	-1.134^{***}	0.107
	(0.107)	(0.098)	(0.128)	(0.112)
4 People	-0.877^{***}	0.029	-1.329^{***}	0.238^{*}
	(0.116)	(0.087)	(0.137)	(0.103)
5 People	-0.236	0.060	-1.108^{***}	0.295
	(0.165)	(0.149)	(0.212)	(0.173)
6 or more people	-0.105	0.347^{*}	-1.305^{***}	0.589^{**}
	(0.177)	(0.153)	(0.251)	(0.182)
AIC	9630.029	9828.695	6666.065	7882.130
BIC	9675.685	9874.351	7000.875	8178.893
Log Likelihood	-4809.015	-4908.347	-3289.032	-3902.065
Deviance	9618.029	9816.695	6578.065	7804.130
Num. obs.	14903	14903	14903	14903

Notes: Logisitic regressions predicting the probability of an individual travelling in the specified group size being in the top 10% of spenders amongst tourists visiting the UK. Coefficients are reported for their effects on the odds ratio. All data from ONS (2017). Regressions performed using glm in R (R Core Team, 2018). Signficance given by $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

A.6 Nationality

For this section the creation of a reference region is harder, because we include some nationalities individually based on the numbers of respondents therefrom. As reference therefore we refer to the European Union excluding the individual nations that are listed in Table A6. For brevity the reference region will be referred to as "Other EU ". In the top 10% this means the EU excluding Germany, Italy, The Netherlands and Sweden. Germany, Italy, The Netherlands, Ireland, Belgium, France and Spain are excluded from the EU when considering the bottom 10%. Likewise, the region "Middle East "is smaller in the upper 10% because of the inclusion of Kuwait, the United Arab Emirates and Saudi Arabia as individual countries within the model. Also in the top 10% India is excluded from the "Indian Subcontinent "region, China from the "Asia "region, Australia from the "Australasia "region and the "North America "region is broken into its two constituent nationalities[2]. For the lower 10% there are no non-EU countries included individually.

Compared to the Other EU nationals, British nationals returning to the United Kingdom to holiday are more likely to be in the extremes of the distribution of expenditure at both ends. No significant differential in probability is expected for the North American region being in the lowest 10%, but for both the USA and Canada there is a higher likelihood of being in the upper decile of expenditure compared to the relevant Other EU category. Central and South American respondents are both predicted to be more likely in the upper decile, and also more likely to appear at the other extreme in the lowest 10% when only nationality is considered. Non-EU Europeans are more likely to be in the top 10% and less so in the bottom decile compared to their EU counterparts. For the Middle East, after adjustment for the individual nations the expectation of being in the top 10% remains higher than the EU reference category and a significant negative odds ratio is reported for the lower decile. Africans are significantly more likely to be in the top 10% of spenders, but no significant difference is found for the lower decile. Indian sub-continent nationals, after adjustment for India's inclusion in the top 10% model remain more likely to be in the upper decile. We also found that nationals of the Indian subcontinent are more likely to appear in the lowest decile of expenditure with, and without, the full set of controls than the other EU nationals, but the strength of the effect is much less than the strength of coefficient for being in the top 10%. Asians are, like many long-distance travellers, found more in the top 10% and less in the bottom 10%; Australasians likewise.

Germans, Italians and the Dutch appear in both top and bottom deciles reflecting partially the number of travellers who come from these nations. German nationals have a greater likelihood of being in the top 10% compared to the Other EU reference category, with Dutch reported as more likely to be in the bottom 10%. A negative significance is found for the Germans being in the lowest decile when the full set of controls is used. Swiss nationals are perhaps unsurprisingly more likely to be in the upper decile given the small nations high GDP per capita (OECD, 2018). Likewise we see other leading economies, China and the USA appearing in the top 10% with greater probability than the Other EU group. From the Middle East, Kuwait, United Arab Emirates and Saudi Arabia all show greater expenditure, as do Commonwealth nations Canada, India and Australia.

Four EU members appear separately in the lowest 10% regressions. When only consid-

[2]Because of economic similarities Mexico is treated as Central America.

Table A6: Length of stay and expenditure prediction

	Stay Only Top 10%	Bottom 10%	Full Model Top 10%	Bottom 10%
(Intercept)	-3.750^{***}	-2.455^{***}	-5.408^{***}	0.602
	(0.095)	(0.078)	(0.356)	(0.500)
British Nationals	1.535^{***}	0.999^{***}	0.510^{**}	0.983^{***}
	(0.162)	(0.127)	(0.174)	(0.139)
North America		-0.165		-0.173
		(0.114)		(0.122)
Central America	1.361^{***}	0.794^{**}	1.254^{***}	0.193
	(0.344)	(0.262)	(0.375)	(0.293)
South America	1.513^{***}	-0.026	1.212^{***}	-0.461
	(0.214)	(0.226)	(0.229)	(0.248)
Europe: Non-EU	1.787^{***}	-0.658^{**}	1.121^{***}	-0.698^{**}
	(0.253)	(0.242)	(0.272)	(0.0251)
Middle East	3.225^{***}	-2.193^{***}	2.683^{***}	-1.810^{***}
	(0.167)	(0.456)	(0.185)	(0.465)
Africa	2.910^{***}	-0.430	1.709^{***}	0.015
	(0.185)	(0.334)	(0.205)	(0.352)
Indian Subcontinent	1.853^{***}	0.748^{***}	0.617	0.482^{*}
	(0.448)	(0.187)	(0.487)	(0.221)
Asia	2.097^{***}	-0.384^{*}	1.500^{***}	-0.413^{*}
	(0.139)	(0.164)	(0.149)	(0.177)
Australasia	2.706^{***}	-0.009	1.609^{***}	-0.050
	(0.251)	(0.168)	(0.279)	(0.183)
Germany	0.361^{*}	0.043	0.089	-0.362^{**}
	(0.170)	(0.120)	(0.177)	(0.131)
Italy	-0.260	-0.168	-0.532	-0.088
	(0.277)	(0.158)	(0.284)	(0.166)
Netherlands	0.005	0.938^{***}	0.035	-0.167
	(0.251)	(0.120)	(0.263)	(0.140)
Sweden	0.487		0.426	
	(0.288)		(0.298)	
Switzerland	1.276^{***}		0.919^{***}	
	(0.237)		(0.250)	
Australia	0.012		0.156	
	(0.253)		(0.284)	
Kuwait	4.185^{***}		3.598^{***}	
	(0.216)		(0.238)	
United Arab Emirates	3.850^{***}		3.572^{***}	
	(0.243)		(0.273)	
Saudi Arabia	4.673^{***}		4.073^{***}	
	(0.233)		(0.259)	
India	2.358^{***}		1.675^{***}	
	(0.193)		(0.215)	
China	3.080^{***}		2.163^{***}	
	(0.179)		(0.197)	
Canada	2.195^{***}		1.234^{***}	
	(0.164)		(0.179)	
USA	2.227^{***}		1.567^{***}	
	(0.112)		(0.121)	
Ireland		0.975^{***}		0.003
		(0.131)		(0.143)
Belgium		1.139^{***}		-0.524^{**}
		(0.159)		(0.183)
France		1.155^{***}		-0.038
		(0.102)		(0.123)
Spain		-0.253		-0.171
		(0.174)		(0.182)
AIC	7926.729	9371.816	6660.384	7887.020
BIC	8101.743	9508.783	6995.194	8183.783
Log Likelihood	-3940.364	-4667.908	-3286.192	-3904.510
Deviance	7880.729	9335.816	6572.384	7809.020
Num. obs.	14903	14903	14903	14903

Notes: Logisitic regressions predicting the probability of an individual of the stated nationality being in the top 10% of spenders amongst tourists visiting the UK. Coefficients are reported for their effects on the odds ratio with the European Union nationals who are not from the countries included in the respective lists for each decile as the omitted category. All data from ONS (2017). Regressions performed using glm in R (R Core Team, 2018). Signficance given by $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

ering nationality Irish, Belgians and French are more likely to be found in the lowest decile than the residual Other EU category. When controlling for other characteristics these results disappear leaving Belgians as actually less likely to appear in the low-spending group. When considering the relative distances travelled, and the cost thereof, it is unsurprising that longer distance travellers are more likely to have money for use when in the UK, whilst these neighbouring countries nationals require less disposable income to make the trip.

A.7 Summary

Through this appendix we have evidenced how logisitic regression can be informative to marketeers looking to identify characteristics likely to be associated with being a high or low spending inbound tourist. These relationships necessarily imply a linear function that is often disproved by the PH of the main paper. We thus see that by forcing relationships onto the data a wealth of information can be lost. Be it in the assumption that richer nation citizens spend more, or the prediction that there is no need to focus on small parties to promote expenditure, there are many significant coefficients in the tables above that should not be taken as definitive in promotion planning.

B Cluster Analysis

The main paper posits that Persistent Homology (PH) provides a better way of targeting promotional material than alternative clustering techniques. A major premise of this argument is that through PH more focused groups of respondents are highlighted that make aligning advertising messages a simpler task. In this appendix we show how traditional clustering techniques employed in the literature provide a less effective means of achieving clusters. A primary factor in such is the inclusion of all observations within the clustering of established methods such as k-means, whilst PH only identifies those associated with "holes"and does not include all data. Because of the potential influence of outliers we briefly present an alternative approach using the trimmed clustering approach of Fritz et al. (2012).

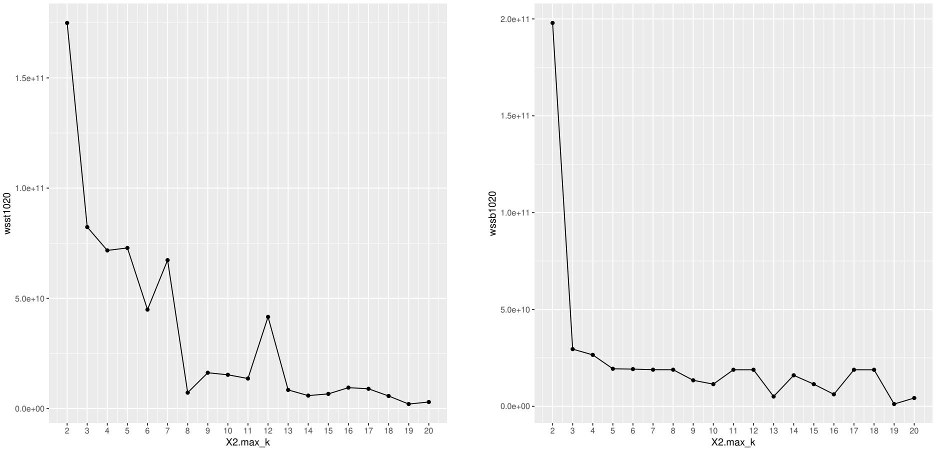
B.1 K-means Clustering

When using traditional clustering methods the alogrithm will seek to allocate all observations to a cluster, splitting the observations repetitively until the full clustering has occurred. Such an approach risks very large clusters in which it is more difficult to extract sufficient targeting information. Many studies in tourism use very low numbers, often giving little explanation as to how the numbers are derived (Dolnicar and Grün, 2008)

Recognising the critique of Dolnicar (2003); Dolnicar and Grün (2008) and others we first obtain the optimal number of clusters using three approaches common in the wider non-tourism clustering literature. The techniques employed are a k-means elbow plot, the distortion function of Pham et al. (2005) and the collection of algorithms contained within Charrad et al. (2014).

Our first approach to establishing the optimum number of clusters involves the creation of Elbow plots of the within cluster sum of squares. Figure A1 shows us that for the top

Figure A1: Elbow plots of k-means clusters within sum of squares



Top 10% Bottom 10%

10% the optimum number of clusters is 8, whilst for the bottom 10% 4 are seen as optimal. These numbers are arrived at as the elbow point of the graph where the line starts to flatten and the additional gain from an extra cluster becomes small. Inevitably such clusters have a large diversity of individuals within them offering limited value over the whole dataset for marketing focus.

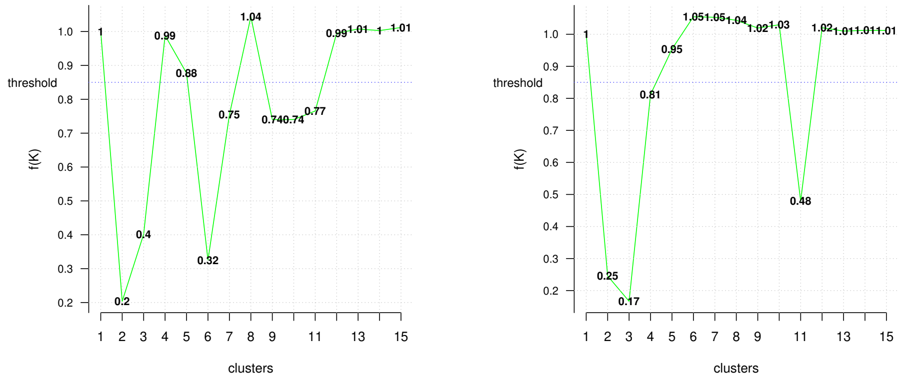
A second approach invokes the Pham et al. (2005) approach implemented in R using Mouselimis (2018). For a given number of clusters K a distortion function $f(K)$ is evaluated, the values from which may then be plotted to identify values below a user specified threshold. In the illustration we limit the maximum number of clusters to 15 and maintain the threshold at 0.85. A disadvantage of this approach is that it only recommends possible K. Running the function for maximum cluster numbers of 150 we find that there is also a possible optimum with 56 clusters for the top 10% and 83 for the bottom 10%. However, because these values are surrounded by values indistinguishable from 1, it is possible that these are peculiarities of the data and therefore they do not represent optimal choices. We provide the $f(K)$ plots over the reduced 15 cluster range as Figure A2; from these plots 11 is suggested for the top 10% and 4 for the bottom 10%[3].

Our third approach is to use Bayesian Inference Criterion implemented in the R package *mclust* (Charrad et al., 2014), which fits 30 different indices for optimal cluster numbers and provides guidance therefrom. Full details of the metrics considered are available in Charrad et al. (2014). For our datasets, the results are summarised in Figure A3 and reveal that for the top 10% just 2 clusters should be chosen and for the bottom 10% 3 are optimal. In the top 10% there are patterns which suggest higher numbers at the level selected by the other methodologies.

The lack of consensus in the methodologies is one of the main challenges of employing a particular method for cluster number selection. Hence we return to the suggestions of the elbow function method as these were also selected by many of the approaches in Charrad

[3]We discount the $K = 11$ solution as it is on its own amongst higher values like the higher 83 suggestion.

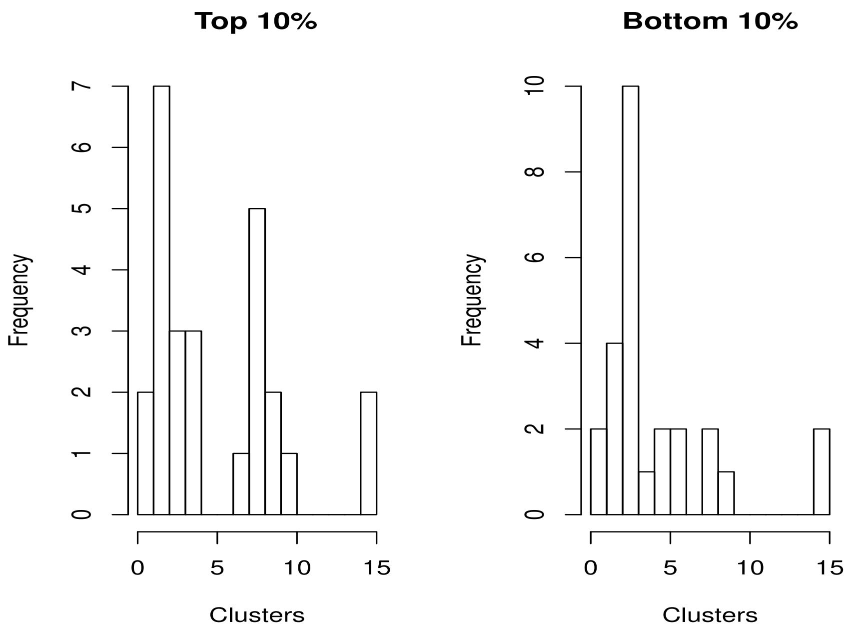
Figure A2: Elbow plots of k-means clusters within sum of squares



Top 10% Bottom 10%

Notes: $f(K)$ is a function of the given number of clusters K. Threshold is a level below which K may be considered to be an optimal cluster number.

Figure A3: Optimal cluster numbers from Charrad et al. (2014)



Notes: Histograms plot the number of methods selecting a given number of clusters as optimal for the top 10% and bottom 10% of expenditure samples. Full details of the algorithms implemented are available in Charrad et al. (2014).

Table A7: Summary statistics for k-means clusters

Sample	Cluster	Stay duration	Age	Male	Flow	Group size	Nationalities	Size
Top 10%	1	3.145	5.491	0.473	0.982	1.855	2	110
		(1.477)	(1.652)	(0.502)	(0.134)	(1.099)		
	2	3.704	6.861	0.591	0.757	1.696	2	115
		(1.835)	(1.583)	(0.494)	(0.431)	(0.948)		
	3	3.338	5.973	0.466	0.953	1.953	3	148
		(1.656)	(1.888)	(0.501)	(0.213)	(1.157)		
	4	3.017	6.981	0.504	0.923	1.83	3	417
		(1.531)	(1.667)	(0.501)	(0.266)	(0.913)		
	5	3.532	5.969	0.587	0.951	2.446	4	327
		(1.726)	(1.947)	(0.493)	(0.216)	(1.524)		
	6	2.132	6.711	0.474	0.579	1.974	1	38
		(1.862)	(1.859)	(0.506)	(0.5)	(0.885)		
	7	2.849	6.005	0.546	0.962	2.178	3	185
		(1.687)	(1.6)	(0.499)	(0.191)	(1.465)		
	8	2.739	6.072	0.569	0.869	1.595	2	153
		(1.888)	(1.821)	(0.497)	(0.338)	(0.869)		
Bottom 10%	1	0.364	4.636	0.636	0.909	1.955	1	22
		(0.727)	(2.216)	(0.492)	(0.294)	(1.463)		
	2	0.981	4.938	0.469	0.719	2.094	4	160
		(1.241)	(2.124)	(0.501)	(0.451)	(1.292)		
	3	0.652	5.254	0.482	0.414	2.314	12	1196
		(1.209)	(2.333)	(0.5)	(0.493)	(1.47)		
	4	1.071	5.929	0.462	0.526	1.923	4	156
		(1.742)	(2.013)	(0.5)	(0.501)	(1.093)		

Notes: Summary statistics calculated on clusters generated by k-means. Figures in parentheses represent standard deviations. Nationalities provides the number of different nationalities within the cluster. Size gives the number of respondents within the cluster.

et al. (2014). In the case of the bottom 10% of expenditure sample, four clusters was not eliminated by the Pham et al. (2005) approach, but the top 10% value of $f(K)$ is above 1. Hence we proceed to estimate k-means clustering using 8 and 4 clusters respectively.

We generate our clusters and provide summary statistics of the type created within the main paper. These are provided in Table A7. Immediately the difference between the two samples on number of nationalities is apparent, where the top 10% are focused on low numbers of nationalities the bottom 10% all have 17 or more nationalities within them. Each cluster is also noticeably larger than the persistent homology values, the smallest containing 38. In PH we found many clusters based solely on gender or travel mode, but here there is no cluster where either the Male dummy or Flow variable take the value 0 or 1 as an average. This lack of focus is one of the reasons why PH is favourable to other clustering methods. For the stay duration, clusters in the top 10% do have much higher average values, meaning there is separation on that dimension.

Recalling the aim of this paper is to show how we can identify focused clusters, to whom marketeers may direct their attentions in the promotion of destinations and expenditure therein, the value of PH against alternatives evidenced in this appendix is clear. The PH approach is thus commended.

B.2 Trimmed Clustering

A feature of the k-means clustering technique is that it includes all of the observations within the stated data matrix, hence outliers influence the overall clustering. In this subsection of the appendix we demonstrate quickly how a trimmed clustering approach can first remove outliers before computing any of the cluster allocations. A full exposition of the methodology employed and the benefits thereof are provided within Fritz et al. (2012). By removing the largest and smallest values the subsequent application of the k-means clustering algorithm

Table A8: Summary statistics for trimmed clusters

Sample	Cluster	Stay duration	Age	Male	Flow	Group size	Nationalities	Size
Top 10%	1	3.243	6.466	0.536	0.907	2.023	12	1007
		(1.673)	(1.843)	(0.499)	(0.291)	(1.203)		
	2	3.089	6.005	0.509	0.986	2.201	4	214
		(1.475)	(1.458)	(0.501)	(0.118)	(1.418)		
	3	2.645	6.091	0.589	0.873	1.599	3	197
		(1.859)	(1.77)	(0.493)	(0.334)	(0.861)		
Bottom 10%	1	0.629	5.125	0.492	0.393	2.331	11	1074
		(1.214)	(2.323)	(0.5)	(0.489)	(1.496)		
	2	0.778	6.125	0.431	0.646	2.132	2	144
		(1.106)	(2.215)	(0.497)	(0.48)	(1.253)		
	3	1.028	5.014	0.444	0.711	2.056	3	142
		(1.271)	(2.11)	(0.499)	(0.455)	(1.276)		
	4	0.897	5.866	0.526	0.392	1.99	2	97
		(1.571)	(1.858)	(0.502)	(0.491)	(1.168)		

Notes: Summary statistics calculated on clusters generated by the *tclust* package of (Fritz et al., 2012). Figures in parentheses represent standard deviations. Nationalities provides the number of different nationalities within the cluster. Size gives the number of respondents within the cluster.

Table A9: Summary statistics for trimmed clusters removing 25% outliers

Sample	Cluster	Stay duration	Age	Male	Flow	Group size	Nationalities	Size
Top 10%	1	3.229	6.421	0.537	0.921	2.045	11	912
		(1.611)	(1.848)	(0.499)	(0.27)	(1.221)		
	2	2.789	6.13	0.553	0.854	1.602	1	123
		(1.964)	(1.882)	(0.499)	(0.355)	(0.903)		
	3	2.964	5.679	0.524	1	2.512	1	84
		(1.617)	(1.554)	(0.502)	(0)	(1.624)		
Bottom 10%	1	0.47	5.123	0.492	0.308	2.338	7	909
		(0.948)	(2.339)	(0.5)	(0.462)	(1.503)		
	2	0.84	6.24	0.424	0.704	2.128	1	125
		(1.16)	(2.201)	(0.496)	(0.458)	(1.211)		
	3	1.129	4.948	0.44	0.784	2.078	1	116
		(1.342)	(2.076)	(0.498)	(0.413)	(1.339)		

Notes: Summary statistics calculated on clusters generated by the *tclust* package of (Fritz et al., 2012). 25% of observations are removed where the algorithm classifies them as outliers. Figures in parentheses represent standard deviations. Nationalities provides the number of different nationalities within the cluster. Size gives the number of respondents within the cluster.

will split the remaining mass of data points into groups optimised for that set. This has obvious benefits over the alternative use of the full sample where numbers are dictated by an optimisation biased from the long distances between outliers and potential cluster centre points.

We run the trimmed clustering function from the *tclust* (Fritz et al., 2012) package in R setting the number of clusters equal to those used in the previous subsection. In this case the algorithm informs when there are empty clusters and suggests reducing the number modelled. After several iterations we are led to the conclusion that the optimal number of clusters for the top 10% is 3 and for the lower 10% the best choice of cluster numbers is 4. Table A8 provides summary statistics showing how large the resulting clusters are.

Given the aim to obtain focused clusters for marketing, the trimmed method as implemented is producing clusters too large. When removing more of the "outliers" we still obtain a large first cluster with 2 and 3 smaller clusters in the top and bottom 10% groups respectively. Table A9 offers summary statistics from this case. An interesting feature of the trimmed clustering is that it removes the nationalities with only a few observations; whether this is beneficial is open to interpretation.

Table A10: Summary statistics for hierarchical clusters

Sample	Cluster	Stay duration	Age	Male	Flow	Group size	Nationalities	Size
Top 10%	1	3.238	6.302	0.524	0.968	2.175	2	63
		(1.757)	(1.691)	(0.503)	(0.177)	(1.476)		
	2	3.704	6.861	0.591	0.757	1.696	2	115
		(1.835)	(1.583)	(0.494)	(0.431)	(0.948)		
	3	2.739	6.072	0.569	0.869	1.595	2	153
		(1.888)	(1.821)	(0.497)	(0.338)	(0.869)		
	4	3.231	6.402	0.532	0.941	2.069	9	854
		(1.618)	(1.869)	(0.499)	(0.235)	(1.24)		
	5	3.282	5.986	0.451	0.951	1.923	2	142
		(1.63)	(1.868)	(0.499)	(0.217)	(1.137)		
	6	2.964	5.679	0.524	1	2.512	1	84
		(1.617)	(1.554)	(0.502)	(0)	(1.624)		
	7	2.318	6.159	0.659	0.886	1.614	1	44
		(1.736)	(1.599)	(0.479)	(0.321)	(0.841)		
	8	2.132	6.711	0.474	0.579	1.974	1	38
		(1.862)	(1.859)	(0.506)	(0.5)	(0.885)		
Bottom 10%	”1	0.647	5.243	0.484	0.423	2.308	13	1218
		(1.202)	(2.332)	(0.5)	(0.494)	(1.47)		
	2	0.981	4.968	0.458	0.716	2.103	3	155
		(1.24)	(2.103)	(0.5)	(0.452)	(1.295)		
	3	1.067	5.695	0.429	0.61	1.829	2	105
		(1.666)	(2.015)	(0.497)	(0.49)	(1.096)		
	4	1.071	6.196	0.554	0.393	2.089	3	56
		(1.857)	(2.118)	(0.502)	(0.493)	(1.083)		

Notes: Summary statistics calculated on clusters generated by the *tclust* package of (Fritz et al., 2012). 25% of observations are removed where the algorithm classifies them as outliers. Figures in parentheses represent standard deviations. Nationalities provides the number of different nationalities within the cluster. Size gives the number of respondents within the cluster.

B.3 Hierarchical Clustering

Alternative to the centroid approaches discussed in the first two examples are the tree based methodologies which cut across the tree in order to generate clusters. Most common of the tree techniques is hierarchical clustering and it is such which forms the example here. At the top of the tree all data is joined in one cluster, but as the similarity requirement for clustering strengthens so the tree becomes split and more groups form. When drawing the tree this similarity parameter defines the vertical dimension, with the horizontal drawn such that each observation is ultimately a node at the branch end of the tree.

Hierarchical clustering is performed using the optimal number of clusters suggested by k-means. There is less motivation for using 8 clusters in the top 10% than there is in k-means, but the choice allows more direct comparison of the methodologies. K-means splits the top 10% into more equal sized clusters relative to the hierarchical approach. Consequently, it is seen that one of the hiearchical clusters, number 4, contains 9 different nationalities where the highest in k-means was 4. All others are 1 or 2, however, where k-means provides some clusters with 3 different nationalities contained within them. Both k-means and hierarchical clustering produce groups that have variation within for all characteristics. A single exception to this is noted for the top 10%, cluster 6 containing only respondents who leave the UK by air. In the bottom 10% size disparities are equally pronounced, with the largest group containing 1218 respondents compared to just 155 in the second largest. Comparative figures for the k-means approach are 1196 and 160. While k-means identified one cluster with just 22 respondents of a single nationality, the smaller groups in the hierarchical clustering contain 2 or 3 different nationalities. Overall there are small differences between the clusters generated by the two approaches, and these come through in the summary statistics. However, there is much commonality between k-means and hierarchical clustering for the dataset considered.

Table A11: Information criteria for model based clustering

Sample	Model	log-likelihood	n	df	BIC	ICL
Top 10%	VEV	-25894.47	1493	50	-52154.37	-52186.06
Bottom 10%	EEE	-25794.03	1534	83	-52196.92	-52475.39

Notes: Model summary statistics calculated for the optimal choices of ellipsoid volume, shape of the contour density and the orientation of the ellipsoid. E implies even variance, whilst V implies varied variance. In this case the column Model reports the relevant selection. n is the number of observations within the sample. *df* provides the degrees of freedom within the model. BIC is the Bayesian Information Criteria and the ICL is the integrated complete-data likelihood criteria.

B.4 Model Based Clustering

A large literature follows Fraley (1998) and Fraley and Raftery (2002) in implementing model based clustering to classify datasets. Much of this uses the R package *mclust* which was first introduced in (Fraley and Raftery, 1999). Subsequent growth of the approach recommends model based clustering for analysis here. For this clustering and comparisons are generated using the latest version of *mclust* (Scrucca et al., 2016).

In general terms for a dataset with n observations, $\mathbf{x} = \{\mathbf{x_1}, \mathbf{x_1}, ..., \mathbf{x_1}, ..., \mathbf{x_1}\}$ a set of G groups will have the probability density function $f(\mathbf{x_i}, \mathbf{\Psi})$. This is noted by Scrucca et al. (2016) to have the form:

$$f(\mathbf{x_i}, \mathbf{\Psi}) = \sum_{k=1}^{G} \pi_k f_k(\mathbf{x_i}; \boldsymbol{\theta}_\mathbf{k}) \tag{2}$$

in which $\mathbf{\Psi} = \{\boldsymbol{\pi}_\mathbf{1}, ..., \boldsymbol{\pi}_\mathbf{G-1}, \boldsymbol{\theta}_\mathbf{1}, ..., \boldsymbol{\theta}_\mathbf{G-1}\}$ are the parameters of the mixture model estimated by the algorithm. $f_k(\mathbf{x_i}; \boldsymbol{\theta}_\mathbf{k})$ is the kth component desnity for observation $\mathbf{x_i}$ with parameter vector $\boldsymbol{\theta}_\mathbf{k}$, $\boldsymbol{\pi}_\mathbf{1}$, ..., $\boldsymbol{\pi}_\mathbf{G-1}$ are the mixing probabilities and therefore sum to 1. It is also imposed that $\boldsymbol{\pi}_\mathbf{k} > 0$. For equation (2) we can compute the corresponding likelihood function $l()$ using:

$$l(\mathbf{\Psi}; \mathbf{x_1}, ...\mathbf{x_n}) = \sum_{i=1}^{n} log(f(\mathbf{x_i}; \boldsymbol{\theta}_\mathbf{k}))$$

This is then optimised using the Dempster et al. (1977) algorithm. As implemented in *mclust* all components have a Gaussian distribution of the form $f(\mathbf{x}; \boldsymbol{\theta}_\mathbf{k})$ $N(\boldsymbol{\mu}_\mathbf{k}, \mathbf{\Sigma_k})$. Clusters are thus ellipsoidal centered on $\boldsymbol{\mu}_\mathbf{k}$ and with $\mathbf{\Sigma_k}$ determining the other geometric qualities of the ellipsoids. Models are thus defined in terms of being of equal variance, E, and varied variance, V, in their volume of the ellipsoid, shape of the density contours and the orientation of the ellipsoid.

Estimation enables all combinations of E and V, with the model format chosen to optimise either the Bayesian Information Criteria (BIC) or the integrated complete-data likelihood (ICL) criteria. In the case of the IPS data they both inform on the same models.

An immediate observation from Table A12 is that the model based approach has only generated two clusters for the top 10% and has produced eight for the bottom 10%. Both

Table A12: Summary statistics for model based clustering

Sample	Cluster	Stay duration	Age	Male	Flow	Group size	Nationalities	Size
Top 10%	1	3.323	6.248	0.231	0.859	2.218	20	905
		(1.711)	(1.956)	(0.422)	(0.349)	(1.398)		
	2	2.889	6.43	1	1	1.643	17	588
		(1.665)	(1.606)	(0)	(0)	(0.728)		
Bottom 10%	1	0.904	5.367	1	1	2.154	17	311
		(1.276)	(2.179)	(0)	(0)	(1.342)		
	2	4.828	4.621	0	1	2.345	9	29
		(1.197)	(2.896)	(0)	(0)	(1.471)		
	3	0.79	5.231	0	1	2.172	17	290
		(0.794)	(2.068)	(0)	(0)	(1.272)		
	4	1.071	5.929	0.462	0.526	1.923	4	156
		(1.742)	(2.013)	(0.5)	(0.501)	(1.093)		
	5	0.615	5.179	0	0	5.692	9	39
		(1.227)	(2.187)	(0)	(0)	(0.893)		
	6	0.323	6.99	0	0	2.404	12	198
		(0.71)	(1.258)	(0)	(0)	(1.046)		
	7	0.498	6.695	1	0	3.108	14	203
		(1.145)	(1.773)	(0)	(0)	(1.56)		
	8	0.268	2.771	0	0	1.452	15	157
		(0.763)	(1.25)	(0)	(0)	(0.937)		

Notes: Summary statistics calculated on clusters generated by the *mclust* package of (Scrucca et al., 2016). Top 10% calculated with a varied variance for the ellipsoid volume, even variance for the shape of the contour density of the ellipsoid and a varied variance for the orientation of the ellipsoid (VEV). Bottom 10% calculated with an even variance for the ellipsoid volume, contour density of the ellipsoid, and ellipsoid orientation (EEE). Figures in parentheses represent standard deviations. Nationalities provides the number of different nationalities within the cluster. Size gives the number of respondents within the cluster.

numbers of clusters differ greatly from those selected by k-means elbow plots. It is also apparent that the focus has moved away from nationality with most clusters containing more than ten. The lowest number of nationalities in a group is 4, in cluster number 4 of the bottom 10%; this has been considered a high number in the other approaches. Both dummy variables are often the focus of clusters, with zero standard deviation observed in the majority of cases. For the top 10% one cluster is all males who leave by air, whilst the other is dominated by females. The high proportion of fliers in the dataset means that flying also dominates that cluster. Variation in other variables is in keeping with that identified in the other approaches reviewed here.

B.5 Summary of Clustering

Most clustering techniques used within the literature involve the allocation of all observations to a cluster, creating large groupings not suitable for marketing. By removing outliers we reduce the number of clusters generated rather than the size, failing to provide the kind of narrow groupings that would be suitable for target promotion. Contained within this appendix are many useful start points for the advancement of clustering analysis within the tourism literature. However, for our purpose we continue to advocate the benefit of data topography, commending the narrower focus of the PH clusters in the main paper.

A further challenge from the clustering techniques exposited here is the tendency to replicate the expected relationships from the existing literature, particularly the correspondence between using aeroplanes and being higher spending. Stay duration is longer in the top 10% and age is slightly higher. Both of these were associated with being in the top 10% in the logit modelling. Consequentially when considering clustering approaches thought should also be given to the value it brings over the regression. On this front we demonstrate how PH can produce some very counter-intuitive clustering.

C Full Cluster Set

In this appendix we provide the full list of clusters identified by the homology with a filter level of $\epsilon = 1.5$. A full set of summary statistics is included within the main manuscript alongside a discussion of the patterns identified there from. We provide this full set of clusters for reference.

C.1 Top 10%

For the highest spending 10% we note that the majority of identified respondents are from the Middle East, India and the United States of America. This should not be seen as surprising but there are more interesting patterns within other variables. For example we see a large number of high spending individual travellers. Other features are more commonly associated with high spending, such as being in the 45 to 64 year old band, arriving by air and staying more than one week. Tables A13 to A15 provide the full set of details.

C.2 Bottom 10%

Tables A16 to A18 provide the full set of clusters in the bottom 10% sorted by the cluster number assigned within the homology. Summaries for these groupings are provided within the main paper. Broadly we can see a dominance of shorter stays, less air arrivals and a larger number of European nationalities than were evident in the top 10%. As noted we do see some surprises within this bottom set, such as the number from USA or Australasia in this bottom 10% where most methodologies have associated these nationalities with high expenditure.

References

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., and Charrad, M. M. (2014). Package 'nbclust'. *Journal of Statistical Software*, 61:1–36.

Dempster, A. P., Laird, N. M., and 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.

Dolnicar, S. (2003). Using cluster analysis for market segmentation-typical misconceptions, established methodological weaknesses and some recommendations for improvement. *Australasian Journal of Market Research*, 11(2):5–12.

Dolnicar, S. and Grün, B. (2008). Challenging "factor–cluster segmentation". *Journal of Travel Research*, 47(1):63–71.

Fraley, C. (1998). Algorithms for model-based gaussian hierarchical clustering. *SIAM Journal on Scientific Computing*, 20(1):270–281.

Table A13: Full Top 10% Cluster Limit

Cluster	Stay	Age	Male	Air Departure	People	Nationality	Min ϵ
1	Between 6 and 8 days	17 to 24 years	0	1	1	Kuwait	1.245
	Between 6 and 8 days	25 to 34 years	0	1	1	Kuwait	
	Between 6 and 8 days	25 to 34 years	0	1	2	Kuwait	
	Between 9 and 12 days	17 to 24 years	0	1	2	Kuwait	
	Between 9 and 12 days	17 to 24 years	0	1	3	Kuwait	
	Between 9 and 12 days	35 to 44 years	0	1	2	United Arab Emirates	
	Between 9 and 12 days	25 to 34 years	0	1	2	United Arab Emirates	
	Between 13 and 18 days	Under 16	0	1	3	United Arab Emirates	
	Between 9 and 12 days	35 to 44 years	0	1	3	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	0	1	4	Saudi Arabia	
	Between 9 and 12 days	17 to 24 years	0	1	4	Saudi Arabia	
	Between 13 and 18 days	Under 16	0	1	4	Saudi Arabia	
	Between 9 and 12 days	25 to 34 years	0	1	4	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	0	1	2	Saudi Arabia	
2	Between 6 and 8 days	25 to 34 years	1	1	1	Kuwait	1.380
	Between 9 and 12 days	25 to 34 years	1	1	1	Kuwait	
	Between 9 and 12 days	35 to 44 years	1	1	1	Kuwait	
	Between 13 and 18 days	45 to 54 years	1	1	1	Kuwait	
	Between 6 and 8 days	25 to 34 years	1	1	1	United Arab Emirates	
	Between 6 and 8 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 3 and 5 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 3 and 5 days	35 to 44 years	1	1	1	Saudi Arabia	
	Between 19 and 26 days	45 to 54 years	1	1	1	Saudi Arabia	
	Between 3 and 5 days	35 to 44 years	1	1	1	India	
	Between 6 and 8 days	25 to 34 years	1	1	1	India	
	Between 13 and 18 days	35 to 44 years	1	1	1	India	
	Between 9 and 12 days	35 to 44 years	1	1	1	India	
	Between 19 and 26 days	45 to 54 years	1	1	1	India	
	Between 9 and 12 days	25 to 34 years	1	1	1	India	
3	Between 19 and 26 days	45 to 54 years	1	1	1	Saudi Arabia	1.485
	Between 27 and 29 days	45 to 54 years	1	1	2	Saudi Arabia	
	More than 30 days	55 to 64 years	1	1	2	Saudi Arabia	
	More than 30 days	45 to 54 years	1	1	2	India	
	More than 30 days	35 to 44 years	1	1	1	India	
	Between 19 and 26 days	45 to 54 years	1	1	1	India	
4	More than 30 days	35 to 44 years	0	1	6	Kuwait	1.365
	Between 27 and 29 days	45 to 54 years	0	1	5	Kuwait	
	Between 27 and 29 days	25 to 34 years	0	1	4	Kuwait	
	Between 9 and 12 days	35 to 44 years	0	1	4	Saudi Arabia	
	Between 13 and 18 days	25 to 34 years	0	1	4	Saudi Arabia	
	Between 6 and 8 days	35 to 44 years	0	1	5	Saudi Arabia	
	Between 13 and 18 days	35 to 44 years	0	1	7	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	0	1	6	Saudi Arabia	
	Between 6 and 8 days	35 to 44 years	0	1	6	Saudi Arabia	
	Between 19 and 26 days	25 to 34 years	0	1	7	Saudi Arabia	
	Between 9 and 12 days	25 to 34 years	0	1	4	Saudi Arabia	
	More than 30 days	25 to 34 years	0	1	7	Saudi Arabia	
	Between 19 and 26 days	35 to 44 years	0	1	4	Saudi Arabia	
5	Between 9 and 12 days	25 to 34 years	0	1	1	EU (Other)	1.170
	Between 19 and 26 days	17 to 24 years	0	1	1	EU (Other)	
	Between 13 and 18 days	17 to 24 years	0	1	1	EU (Other)	
	Between 13 and 18 days	35 to 44 years	0	1	1	EU (Other)	
	Between 27 and 29 days	17 to 24 years	0	1	1	EU (Other)	
	Between 13 and 18 days	17 to 24 years	0	1	1	EU (Other)	
	Between 27 and 29 days	35 to 44 years	0	1	1	EU (Other)	
	More than 30 days	25 to 34 years	0	1	1	EU (Other)	
6	Between 9 and 12 days	35 to 44 years	1	1	1	Kuwait	1.380
	Between 9 and 12 days	25 to 34 years	1	1	1	Kuwait	
	Between 6 and 8 days	25 to 34 years	1	1	1	Kuwait	
	Between 13 and 18 days	45 to 54 years	1	1	1	Kuwait	
	Between 6 and 8 days	25 to 34 years	1	1	1	United Arab Emirates	
	Between 9 and 12 days	25 to 34 years	1	1	2	Saudi Arabia	
	Between 6 and 8 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 6 and 8 days	25 to 34 years	1	1	2	Saudi Arabia	
	Between 19 and 26 days	45 to 54 years	1	1	1	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	1	1	2	Saudi Arabia	
	Between 3 and 5 days	45 to 54 years	1	1	2	India	
	Between 13 and 18 days	35 to 44 years	1	1	1	India	
	Between 3 and 5 days	45 to 54 years	1	1	1	India	
	Between 9 and 12 days	35 to 44 years	1	1	2	India	
	Between 6 and 8 days	45 to 54 years	1	1	2	India	
	Between 6 and 8 days	25 to 34 years	1	1	1	India	
	Between 3 and 5 days	35 to 44 years	1	1	1	India	
	Between 9 and 12 days	35 to 44 years	1	1	1	India	
	Between 19 and 26 days	45 to 54 years	1	1	1	India	
	Between 9 and 12 days	25 to 34 years	1	1	1	India	
	Between 9 and 12 days	45 to 54 years	1	1	2	India	

Notes: Nationalities are provided where more than 20 visitors shared a particular nationality. Where applied (Other) informs that the nationality of the respondent is not in the set provided in the summary statistics of the main paper.

Table A14: Full Top 10% Cluster Limit

Cluster	Stay	Age	Male	Air Departure	People	Nationality	Min ϵ
7	Between 9 and 12 days	35 to 44 years	1	1	1	Kuwait	1.485
	Between 6 and 8 days	25 to 34 years	1	1	1	Kuwait	
	Between 13 and 18 days	45 to 54 years	1	1	1	Kuwait	
	Between 9 and 12 days	25 to 34 years	1	1	1	Kuwait	
	Between 6 and 8 days	25 to 34 years	1	1	1	United Arab Emirates	
	Between 6 and 8 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 19 and 26 days	45 to 54 years	1	1	1	Saudi Arabia	
	Between 13 and 18 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 9 and 12 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 13 and 18 days	35 to 44 years	1	1	1	India	
	Between 19 and 26 days	45 to 54 years	1	1	1	India	
8	Between 13 and 18 days	35 to 44 years	0	1	6	Kuwait	1.365
	Between 13 and 18 days	17 to 24 years	0	1	5	Kuwait	
	Between 13 and 18 days	25 to 34 years	0	1	6	Kuwait	
	Between 13 and 18 days	35 to 44 years	0	1	7	United Arab Emirates	
	Between 9 and 12 days	35 to 44 years	0	1	4	Saudi Arabia	
	Between 6 and 8 days	35 to 44 years	0	1	5	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	0	1	6	Saudi Arabia	
	Between 9 and 12 days	25 to 34 years	0	1	4	Saudi Arabia	
	Between 9 and 12 days	17 to 24 years	0	1	4	Saudi Arabia	
	Between 6 and 8 days	35 to 44 years	0	1	6	Saudi Arabia	
	Between 13 and 18 days	35 to 44 years	0	1	7	Saudi Arabia	
9	Between 9 and 12 days	45 to 54 years	1	1	3	East Asia (Other)	1.485
	Between 6 and 8 days	25 to 34 years	1	1	6	East Asia (Other)	
	Between 13 and 18 days	35 to 44 years	1	1	2	East Asia (Other)	
	Between 3 and 5 days	17 to 24 years	1	1	5	East Asia (Other)	
	Between 9 and 12 days	45 to 54 years	1	1	4	East Asia (Other)	
	Between 9 and 12 days	17 to 24 years	1	1	5	East Asia (Other)	
	Between 9 and 12 days	45 to 54 years	1	1	2	East Asia (Other)	
	Between 13 and 18 days	25 to 34 years	1	1	2	East Asia (Other)	
	Between 13 and 18 days	45 to 54 years	1	1	2	East Asia (Other)	
	Between 9 and 12 days	35 to 44 years	1	1	5	East Asia (Other)	
	Between 13 and 18 days	25 to 34 years	1	1	4	East Asia (Other)	
	Between 13 and 18 days	25 to 34 years	1	1	3	East Asia (Other)	
10	Between 6 and 8 days	25 to 34 years	0	1	1	Kuwait	1.380
	Between 6 and 8 days	25 to 34 years	0	1	2	Kuwait	
	Between 6 and 8 days	25 to 34 years	0	1	1	United Arab Emirates	
	Between 9 and 12 days	25 to 34 years	0	1	2	United Arab Emirates	
	Between 9 and 12 days	35 to 44 years	0	1	2	United Arab Emirates	
	Between 9 and 12 days	35 to 44 years	0	1	1	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	0	1	2	Saudi Arabia	
	Between 6 and 8 days	25 to 34 years	0	1	1	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	0	1	2	India	
	Between 9 and 12 days	35 to 44 years	0	1	1	India	
11	1 or 2 days	45 to 54 years	0	1	2	Australia	1.230
	Between 13 and 18 days	55 to 64 years	0	1	2	Australia	
	Between 3 and 5 days	65 years +	0	1	2	Australia	
	Between 13 and 18 days	65 years +	0	1	1	Australia	
	Between 13 and 18 days	65 years +	0	1	2	Australia	
	Between 13 and 18 days	45 to 54 years	0	1	2	Australia	
	Between 6 and 8 days	35 to 44 years	0	1	2	Australia	
	Between 3 and 5 days	65 years +	0	1	1	Australia	
	Between 9 and 12 days	65 years +	0	1	1	Australia	
	Between 3 and 5 days	35 to 44 years	0	1	2	Australia	
	Between 9 and 12 days	35 to 44 years	0	1	2	Australia	
12	Between 9 and 12 days	25 to 34 years	1	1	1	Kuwait	1.380
	Between 6 and 8 days	25 to 34 years	1	1	1	Kuwait	
	Between 9 and 12 days	35 to 44 years	1	1	1	Kuwait	
	Between 13 and 18 days	45 to 54 years	1	1	1	Kuwait	
	Between 6 and 8 days	25 to 34 years	1	1	1	United Arab Emirates	
	Between 9 and 12 days	25 to 34 years	1	1	2	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	1	1	2	Saudi Arabia	
	Between 19 and 26 days	45 to 54 years	1	1	1	Saudi Arabia	
	Between 6 and 8 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 6 and 8 days	25 to 34 years	1	1	2	Saudi Arabia	
	Between 6 and 8 days	45 to 54 years	1	1	2	Saudi Arabia	
	Between 3 and 5 days	35 to 44 years	1	1	1	India	
	Between 3 and 5 days	45 to 54 years	1	1	1	India	
	Between 9 and 12 days	25 to 34 years	1	1	1	India	
	Between 3 and 5 days	45 to 54 years	1	1	2	India	
	Between 9 and 12 days	35 to 44 years	1	1	1	India	
	Between 13 and 18 days	35 to 44 years	1	1	1	India	
	Between 19 and 26 days	45 to 54 years	1	1	1	India	
	Between 6 and 8 days	45 to 54 years	1	1	2	India	
	Between 6 and 8 days	25 to 34 years	1	1	1	India	

Notes: Nationalities are provided where more than 20 visitors shared a articular nationality. Where applied (Other) informs that the nationality of the respondent is not in the set provided in the summary statistics of the main paper.

Table A15: Full Top 10% Cluster Limit

Cluster	Stay	Age	Male	Air Departure	People	Nationality	Min ϵ
13	Between 9 and 12 days	35 to 44 years	1	1	1	Kuwait	1.380
	Between 6 and 8 days	25 to 34 years	1	1	1	Kuwait	
	Between 13 and 18 days	45 to 54 years	1	1	1	Kuwait	
	Between 9 and 12 days	25 to 34 years	1	1	1	Kuwait	
	Between 6 and 8 days	25 to 34 years	1	1	1	United Arab Emirates	
	Between 6 and 8 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 19 and 26 days	45 to 54 years	1	1	1	Saudi Arabia	
	Between 9 and 12 days	25 to 34 years	1	1	1	India	
	Between 19 and 26 days	45 to 54 years	1	1	1	India	
	Between 6 and 8 days	25 to 34 years	1	1	1	India	
	Between 13 and 18 days	35 to 44 years	1	1	1	India	
	Between 9 and 12 days	35 to 44 years	1	1	1	India	
14	Between 6 and 8 days	25 to 34 years	1	1	2	Kuwait	1.365
	Between 9 and 12 days	45 to 54 years	1	1	6	Kuwait	
	Between 3 and 5 days	25 to 34 years	1	1	2	United Arab Emirates	
	Between 6 and 8 days	45 to 54 years	1	1	4	United Arab Emirates	
	Between 6 and 8 days	45 to 54 years	1	1	5	United Arab Emirates	
	Between 6 and 8 days	25 to 34 years	1	1	3	United Arab Emirates	
	Between 13 and 18 days	45 to 54 years	1	1	6	United Arab Emirates	
	Between 13 and 18 days	Under 16	1	1	3	United Arab Emirates	
	Between 6 and 8 days	45 to 54 years	1	1	6	United Arab Emirates	
	Between 13 and 18 days	Under 16	1	1	5	United Arab Emirates	
	Between 6 and 8 days	45 to 54 years	1	1	3	United Arab Emirates	
	Between 6 and 8 days	25 to 34 years	1	1	2	United Arab Emirates	
	Between 6 and 8 days	35 to 44 years	1	1	3	United Arab Emirates	
	Between 6 and 8 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 13 and 18 days	17 to 24 years	1	1	7	Saudi Arabia	
	Between 13 and 18 days	Under 16	1	1	4	Saudi Arabia	
	Between 6 and 8 days	25 to 34 years	1	1	2	Saudi Arabia	
	Between 13 and 18 days	Under 16	1	1	6	Saudi Arabia	
	Between 3 and 5 days	25 to 34 years	1	1	2	Saudi Arabia	
	Between 9 and 12 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 13 and 18 days	45 to 54 years	1	1	6	Saudi Arabia	
	Between 13 and 18 days	25 to 34 years	1	1	6	Saudi Arabia	
	Between 9 and 12 days	Under 16	1	1	2	Saudi Arabia	
	Between 9 and 12 days	17 to 24 years	1	1	1	Saudi Arabia	
15	Between 13 and 18 days	55 to 64 years	0	1	1	United Kingdom	1.290
	Between 19 and 26 days	25 to 34 years	0	1	1	United Kingdom	
	Between 13 and 18 days	55 to 64 years	0	1	2	United Kingdom	
	Between 27 and 29 days	55 to 64 years	0	1	2	United Kingdom	
	Between 13 and 18 days	45 to 54 years	0	1	1	United Kingdom	
	More than 30 days	35 to 44 years	0	1	1	United Kingdom	
	More than 30 days	45 to 54 years	0	1	1	United Kingdom	
16	Between 13 and 18 days	45 to 54 years	1	1	1	Kuwait	1.380
	Between 6 and 8 days	25 to 34 years	1	1	1	Kuwait	
	Between 9 and 12 days	25 to 34 years	1	1	1	Kuwait	
	Between 9 and 12 days	35 to 44 years	1	1	1	Kuwait	
	Between 6 and 8 days	25 to 34 years	1	1	1	United Arab Emirates	
	Between 9 and 12 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 6 and 8 days	25 to 34 years	1	1	1	Saudi Arabia	
	Between 19 and 26 days	45 to 54 years	1	1	1	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	1	1	1	India	
	Between 9 and 12 days	25 to 34 years	1	1	1	India	
	Between 19 and 26 days	45 to 54 years	1	1	1	India	
	Between 13 and 18 days	35 to 44 years	1	1	1	India	
17	Between 6 and 8 days	65 years +	0	0	2	USA	1.140
	Between 3 and 5 days	55 to 64 years	0	0	1	USA	
	Between 6 and 8 days	35 to 44 years	0	0	1	USA	
	Between 3 and 5 days	35 to 44 years	0	0	1	USA	
	Between 9 and 12 days	65 years +	0	0	2	USA	
	Between 9 and 12 days	45 to 54 years	0	0	2	USA	
	Between 6 and 8 days	35 to 44 years	0	0	2	USA	
18	Between 6 and 8 days	25 to 34 years	0	1	2	Kuwait	1.485
	Between 6 and 8 days	25 to 34 years	0	1	1	Kuwait	
	Between 6 and 8 days	25 to 34 years	0	1	1	United Arab Emirates	
	Between 9 and 12 days	35 to 44 years	0	1	2	United Arab Emirates	
	Between 9 and 12 days	25 to 34 years	0	1	2	United Arab Emirates	
	Between 6 and 8 days	25 to 34 years	0	1	1	Saudi Arabia	
	Between 13 and 18 days	45 to 54 years	0	1	3	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	0	1	3	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	0	1	1	Saudi Arabia	
	Between 9 and 12 days	35 to 44 years	0	1	2	Saudi Arabia	
	Between 13 and 18 days	45 to 54 years	0	1	2	India	
	Between 9 and 12 days	35 to 44 years	0	1	2	India	
	Between 9 and 12 days	35 to 44 years	0	1	1	India	
	Between 13 and 18 days	35 to 44 years	0	1	3	India	

Notes: Nationalities are provided where more than 20 visitors shared a particular nationality. Where applied (Other) informs that the nationality of the respondent is not in the set provided in the summary statistics of the main paper.

Table A16: Full Bottom 10% Cluster Limit

Cluster	Stay	Age	Male	Air Departure	People	Nationality	Min ϵ
1	Between 3 and 5 days	45 to 54 years	1	0	5	France	1.200
	Between 3 and 5 days	55 to 64 years	1	0	5	France	
	Between 6 and 8 days	65 years +	1	0	4	Germany	
	Between 3 and 5 days	45 to 54 years	1	0	6	Germany	
	Between 9 and 12 days	45 to 54 years	1	0	4	Germany	
	Between 9 and 12 days	45 to 54 years	1	0	5	Germany	
	Between 6 and 8 days	35 to 44 years	1	0	6	Germany	
2	1 or 2 days	45 to 54 years	1	0	5	France	1.395
	1 or 2 days	35 to 44 years	1	0	4	France	
	1 or 2 days	35 to 44 years	1	0	2	France	
	1 or 2 days	25 to 34 years	1	0	2	France	
	1 or 2 days	45 to 54 years	1	0	2	France	
	1 or 2 days	55 to 64 years	1	0	2	France	
	1 or 2 days	35 to 44 years	1	0	5	France	
	1 or 2 days	25 to 34 years	1	0	3	France	
	Between 3 and 5 days	55 to 64 years	1	0	2	France	
	1 or 2 days	35 to 44 years	1	0	3	France	
	Between 3 and 5 days	45 to 54 years	1	0	5	France	
	1 or 2 days	55 to 64 years	1	0	5	France	
	Between 3 and 5 days	55 to 64 years	1	0	5	France	
	Between 3 and 5 days	45 to 54 years	1	0	6	Germany	
	Between 6 and 8 days	35 to 44 years	1	0	6	Germany	
	Between 6 and 8 days	45 to 54 years	1	0	2	Germany	
	Between 9 and 12 days	45 to 54 years	1	0	4	Germany	
	Between 9 and 12 days	45 to 54 years	1	0	2	Germany	
	Between 9 and 12 days	45 to 54 years	1	0	5	Germany	
3	1 or 2 days	17 to 24 years	0	0	1	France	1.290
	1 or 2 days	Under 16	0	0	1	France	
	1 or 2 days	17 to 24 years	0	0	1	Germany	
	1 or 2 days	25 to 34 years	0	0	1	Germany	
	1 or 2 days	17 to 24 years	0	0	1	Germany	
	Between 3 and 5 days	Under 16	0	0	1	Germany	
	1 or 2 days	Under 16	0	0	1	Germany	
	1 or 2 days	25 to 34 years	0	0	1	Italy	
	Between 3 and 5 days	Under 16	0	0	1	Netherlands	
	1 or 2 days	Under 16	0	0	2	Netherlands	
	1 or 2 days	17 to 24 years	0	0	2	Netherlands	
	1 or 2 days	25 to 34 years	0	0	1	Netherlands	
	1 or 2 days	17 to 24 years	0	0	1	Netherlands	
4	1 or 2 days	45 to 54 years	0	1	4	USA	1.125
	1 or 2 days	35 to 44 years	0	1	1	USA	
	1 or 2 days	55 to 64 years	0	1	2	USA	
	1 or 2 days	45 to 54 years	0	1	3	USA	
	Between 9 and 12 days	55 to 64 years	0	1	4	USA	
	Between 3 and 5 days	35 to 44 years	0	1	1	USA	
	1 or 2 days	45 to 54 years	0	1	1	USA	
	Between 6 and 8 days	45 to 54 years	0	1	4	USA	
	Between 3 and 5 days	45 to 54 years	0	1	4	USA	
	Between 6 and 8 days	45 to 54 years	0	1	1	USA	
	Between 9 and 12 days	65 years +	0	1	2	USA	
	Between 13 and 18 days	65 years +	0	1	2	USA	
	Between 9 and 12 days	45 to 54 years	0	1	1	USA	
	Between 13 and 18 days	45 to 54 years	0	1	3	USA	
5	1 or 2 days	17 to 24 years	0	0	1	Australasia (Other)	1.260
	1 or 2 days	25 to 34 years	0	0	1	United Kingdom	
	1 or 2 days	45 to 54 years	0	0	1	Eire	
	1 or 2 days	Under 16	0	0	1	Eire	
	1 or 2 days	25 to 34 years	0	0	1	Eire	
	1 or 2 days	35 to 44 years	0	0	1	Eire	
	1 or 2 days	45 to 54 years	0	0	1	Belgium	
	1 or 2 days	Under 16	0	0	1	Belgium	
	1 or 2 days	25 to 34 years	0	0	1	France	
	1 or 2 days	Under 16	0	0	1	France	
	1 or 2 days	45 to 54 years	0	0	1	France	
	1 or 2 days	35 to 44 years	0	0	1	France	
	1 or 2 days	17 to 24 years	0	0	1	France	
	1 or 2 days	17 to 24 years	0	0	1	Germany	
	1 or 2 days	25 to 34 years	0	0	1	Germany	
	1 or 2 days	17 to 24 years	0	0	1	Germany	

Notes: Nationalities are provided where more than 20 visitors shared a articular nationality. Where applied (Other) informs that the nationality of the respondent is not in the set provided in the summary statistics of the main paper.

Table A17: Full Bottom 10% Cluster List Part 2

Cluster	Stay	Age	Male	Air Departure	People	Nationality	Min ϵ
6	Between 6 and 8 days	25 to 34 years	1	1	5	Belgium	1.200
	Between 3 and 5 days	25 to 34 years	1	1	3	France	
	Between 3 and 5 days	35 to 44 years	1	1	5	France	
	Between 6 and 8 days	17 to 24 years	1	1	5	France	
	Between 3 and 5 days	Under 16	1	1	5	Germany	
	Between 3 and 5 days	25 to 34 years	1	1	4	Germany	
	Between 3 and 5 days	25 to 34 years	1	1	2	Germany	
	Between 3 and 5 days	35 to 44 years	1	1	4	Germany	
	Between 6 and 8 days	Under 16	1	1	4	Germany	
	1 or 2 days	Under 16	1	1	5	Italy	
	1 or 2 days	35 to 44 years	1	1	2	Italy	
	Between 3 and 5 days	25 to 34 years	1	1	2	Italy	
	1 or 2 days	25 to 34 years	1	1	2	Italy	
	1 or 2 days	17 to 24 years	1	1	4	Italy	
	1 or 2 days	25 to 34 years	1	1	4	Italy	
	1 or 2 days	45 to 54 years	1	1	2	Italy	
	Between 3 and 5 days	35 to 44 years	1	1	3	Italy	
	1 or 2 days	35 to 44 years	1	1	4	Italy	
	Between 3 and 5 days	45 to 54 years	1	1	3	Italy	
	Between 3 and 5 days	35 to 44 years	1	1	4	Italy	
	Between 3 and 5 days	45 to 54 years	1	1	2	Italy	
7	1 or 2 days	25 to 34 years	1	1	2	Italy	1.455
	1 or 2 days	45 to 54 years	1	1	2	Italy	
	1 or 2 days	25 to 34 years	1	1	4	Italy	
	1 or 2 days	35 to 44 years	1	1	4	Italy	
	1 or 2 days	35 to 44 years	1	1	2	Italy	
	Between 3 and 5 days	35 to 44 years	1	1	4	Italy	
	Between 3 and 5 days	35 to 44 years	1	1	3	Italy	
	Between 3 and 5 days	45 to 54 years	1	1	3	Italy	
	1 or 2 days	17 to 24 years	1	1	4	Italy	
	Between 3 and 5 days	45 to 54 years	1	1	2	Italy	
	1 or 2 days	17 to 24 years	1	1	2	Netherlands	
8	Between 6 and 8 days	17 to 24 years	0	1	1	Australasia (Other)	1.440
	Between 3 and 5 days	25 to 34 years	0	1	1	Australasia (Other)	
	Between 3 and 5 days	17 to 24 years	0	1	1	United Kingdom	
	Between 3 and 5 days	35 to 44 years	0	1	1	United Kingdom	
	Between 6 and 8 days	35 to 44 years	0	1	1	Eire	
	Between 3 and 5 days	45 to 54 years	0	1	1	Eire	
	Between 9 and 12 days	25 to 34 years	0	1	1	Eire	
9	1 or 2 days	17 to 24 years	0	0	7	Netherlands	1.470
	1 or 2 days	17 to 24 years	0	0	5	Netherlands	
	1 or 2 days	35 to 44 years	0	0	3	Netherlands	
	1 or 2 days	25 to 34 years	0	0	3	Netherlands	
	1 or 2 days	45 to 54 years	0	0	4	Netherlands	
	1 or 2 days	35 to 44 years	0	0	5	Netherlands	
	1 or 2 days	25 to 34 years	0	0	4	Netherlands	
	1 or 2 days	35 to 44 years	0	0	7	Netherlands	
	1 or 2 days	45 to 54 years	0	0	3	Netherlands	
10	1 or 2 days	35 to 44 years	0	0	2	EU (Other)	1.305
	1 or 2 days	17 to 24 years	0	0	2	EU (Other)	
	Between 3 and 5 days	35 to 44 years	0	0	1	EU (Other)	
	Between 3 and 5 days	17 to 24 years	0	0	2	EU (Other)	
	Between 3 and 5 days	17 to 24 years	0	0	1	EU (Other)	
11	Between 3 and 5 days	17 to 24 years	0	1	3	France	1.200
	1 or 2 days	17 to 24 years	0	1	5	France	
	1 or 2 days	25 to 34 years	0	1	1	France	
	1 or 2 days	17 to 24 years	0	1	4	France	
	1 or 2 days	35 to 44 years	0	1	1	France	
	Between 3 and 5 days	17 to 24 years	0	1	2	France	
	Between 3 and 5 days	25 to 34 years	0	1	1	France	
	Between 3 and 5 days	25 to 34 years	0	1	5	France	
	Between 3 and 5 days	17 to 24 years	0	1	1	Germany	
	Between 3 and 5 days	25 to 34 years	0	1	1	Germany	
	1 or 2 days	35 to 44 years	0	1	4	Germany	
	1 or 2 days	45 to 54 years	0	1	2	Germany	
	1 or 2 days	45 to 54 years	0	1	1	Germany	
	1 or 2 days	45 to 54 years	0	1	3	Germany	
	Between 3 and 5 days	17 to 24 years	0	1	2	Germany	
	1 or 2 days	35 to 44 years	0	1	3	Germany	
	Between 3 and 5 days	17 to 24 years	0	1	1	Germany	

Notes: Nationalities are provided where more than 20 visitors shared a articular nationality. Where applied (Other) informs that the nationality of the respondent is not in the set provided in the summary statistics of the main paper.

Table A18: Full Bottom 10% Cluster List Part 3

Cluster	Stay	Age	Male	Air Departure	People	Nationality	Min ϵ
12	1 or 2 days	25 to 34 years	0	0	2	United Kingdom	1.17
	1 or 2 days	25 to 34 years	0	0	1	United Kingdom	
	1 or 2 days	45 to 54 years	0	0	1	Eire	
	1 or 2 days	17 to 24 years	0	0	5	Eire	
	1 or 2 days	35 to 44 years	0	0	1	Eire	
	1 or 2 days	35 to 44 years	0	0	3	Eire	
	1 or 2 days	25 to 34 years	0	0	3	Eire	
	1 or 2 days	Under 16	0	0	5	Eire	
	1 or 2 days	25 to 34 years	0	0	4	Eire	
	1 or 2 days	25 to 34 years	0	0	1	Eire	
	1 or 2 days	35 to 44 years	0	0	2	Eire	
	1 or 2 days	Under 16	0	0	5	Belgium	
	1 or 2 days	45 to 54 years	0	0	1	Belgium	
	1 or 2 days	Under 16	0	0	5	France	
	1 or 2 days	17 to 24 years	0	0	1	France	
	1 or 2 days	17 to 24 years	0	0	2	France	
	1 or 2 days	35 to 44 years	0	0	1	France	
	1 or 2 days	45 to 54 years	0	0	1	France	
	1 or 2 days	17 to 24 years	0	0	4	France	
	1 or 2 days	17 to 24 years	0	0	3	France	
	1 or 2 days	Under 16	0	0	4	France	
	1 or 2 days	25 to 34 years	0	0	1	France	
	1 or 2 days	25 to 34 years	0	0	1	Germany	
	1 or 2 days	17 to 24 years	0	0	1	Germany	
	1 or 2 days	17 to 24 years	0	0	1	Germany	
13	Between 3 and 5 days	25 to 34 years	1	1	3	France	1.425
	Between 6 and 8 days	17 to 24 years	1	1	5	France	
	Between 3 and 5 days	25 to 34 years	1	1	4	Germany	
	Between 3 and 5 days	Under 16	1	1	5	Germany	
	Between 6 and 8 days	Under 16	1	1	4	Germany	
	Between 3 and 5 days	25 to 34 years	1	1	2	Germany	
	1 or 2 days	45 to 54 years	1	1	2	Italy	
	1 or 2 days	35 to 44 years	1	1	2	Italy	
	1 or 2 days	Under 16	1	1	5	Italy	
	Between 3 and 5 days	25 to 34 years	1	1	2	Italy	
	Between 3 and 5 days	35 to 44 years	1	1	4	Italy	
	1 or 2 days	17 to 24 years	1	1	4	Italy	
	Between 3 and 5 days	35 to 44 years	1	1	3	Italy	
	1 or 2 days	25 to 34 years	1	1	4	Italy	
	1 or 2 days	25 to 34 years	1	1	2	Italy	
	1 or 2 days	35 to 44 years	1	1	4	Italy	
	Between 3 and 5 days	45 to 54 years	1	1	2	Italy	
	Between 3 and 5 days	45 to 54 years	1	1	3	Italy	
14	1 or 2 days	25 to 34 years	1	1	2	USA	1.470
	1 or 2 days	45 to 54 years	1	1	5	USA	
	1 or 2 days	45 to 54 years	1	1	3	USA	
	1 or 2 days	35 to 44 years	1	1	2	USA	
	1 or 2 days	45 to 54 years	1	1	2	USA	
	1 or 2 days	17 to 24 years	1	1	5	USA	
	1 or 2 days	45 to 54 years	1	1	4	USA	
	1 or 2 days	17 to 24 years	1	1	3	USA	
15	Between 3 and 5 days	25 to 34 years	0	1	1	Australasia (Other)	1.425
	Between 6 and 8 days	25 to 34 years	0	1	3	Australasia (Other)	
	Between 6 and 8 days	17 to 24 years	0	1	3	Australasia (Other)	
	Between 3 and 5 days	17 to 24 years	0	1	2	Australasia (Other)	
	Between 9 and 12 days	45 to 54 years	0	1	1	United Kingdom	
	Between 3 and 5 days	35 to 44 years	0	1	1	United Kingdom	
	Between 6 and 8 days	55 to 64 years	0	1	1	United Kingdom	
	Between 3 and 5 days	17 to 24 years	0	1	1	United Kingdom	
	Between 3 and 5 days	65 years +	0	1	1	United Kingdom	
	Between 6 and 8 days	65 years +	0	1	2	United Kingdom	
	Between 3 and 5 days	65 years +	0	1	2	United Kingdom	
	Between 9 and 12 days	45 to 54 years	0	1	2	United Kingdom	
	Between 3 and 5 days	55 to 64 years	0	1	1	Eire	
	Between 3 and 5 days	45 to 54 years	0	1	1	Eire	

Notes: Nationalities are provided where more than 20 visitors shared a particular nationality. Where applied (Other) informs that the nationality of the respondent is not in the set provided in the summary statistics of the main paper.

Fraley, C. and Raftery, A. E. (1999). Mclust: Software for model-based cluster analysis. *Journal of classification*, 16(2):297–306.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, 97(458):611–631.

Fritz, H., Garcia-Escudero, L. A., and Mayo-Iscar, A. (2012). tclust: An R package for a trimming approach to cluster analysis. *Journal of Statistical Software*, 47(12):1–26.

Mouselimis, L. (2018). *ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans and K-Medoids Clustering*. R package version 1.1.3.

OECD (2018). Gdp per capita. *https://data.oecd.org/switzerland.htm*.

ONS (2017). International Passenger Survey [data collection] SN: 8016.

Pham, D. T., Dimov, S. S., and Nguyen, C. D. (2005). Selection of k in k-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1):103–119.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Scrucca, L., Fop, M., Murphy, T. B., and Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1):205–233.