Group 6: Sinan Wang 620146, Akanksha Saxena 613826, Berivan Teper 618443,
Satyaki Ghosh 613827

1.a- One reason for the missing value is due to respondents seeing different samples of cities, another one is that respondents did not evaluate all the attributes/preference on the questionnaires. After deleting the missing value derived from the first reason, there are still missing values caused by the second reason.

It is suggested that the missing values of each attribute are derived from many different respondents, thus it is impossible to omit the observations according to respondents' number. It will cause unnecessary loss of information. So, adopting the mean value of each attribute to substitute the missing value of the specific attribute can effectively avoid the problem. We also checked the minimum and maximum value of attributes and preference to ensure no outliers. As a result, the dataset contains the evaluation from 266 respondents.

	Berlin	London	Paris	Dublin	Riga	Geneva
the amount of respondents visited the city	236	174	169	43	38	37
the percentage of total 266 respondents	89%	65%	64%	16%	14%	14%

Table 1: The visiting frequencies of cities

1.b- It is suggested from the 266 respondents' data that the average age of respondents is 25.46 years old and 42% of the respondents are male. The most respondents are from Germany (45%), and live in Berlin (47%) currently. The biggest percentage of respondents are Master students, accounting for 50%. 52% of the respondents are in a relationship and 41% of them are single.

As for travel, the most common reason for a weekend trip is to explore a city (89%). The most (83%) respondents go on a weekend trip with friends but the least (6%) with colleagues. 44% of the respondents have travelled 2-3 times in 2015 and 43% of them had a 200-300€ budget for a weekend trip. The Table1 below shows the top three and bottom three cities which the respondents have visited.

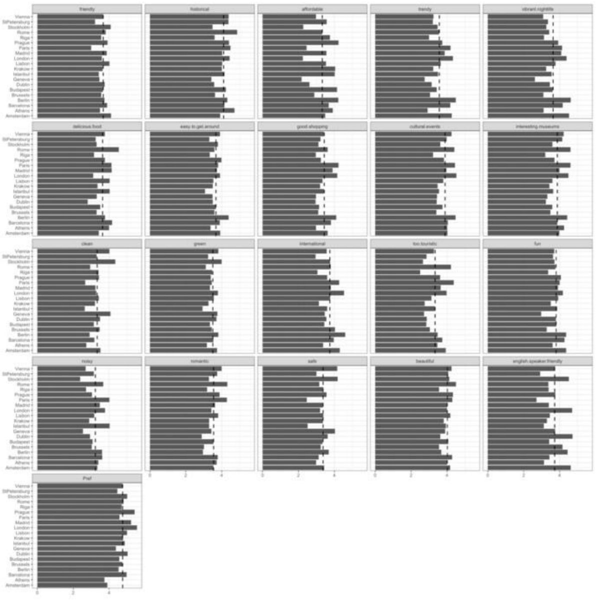


Figure 1: mean attribute and preference across cities

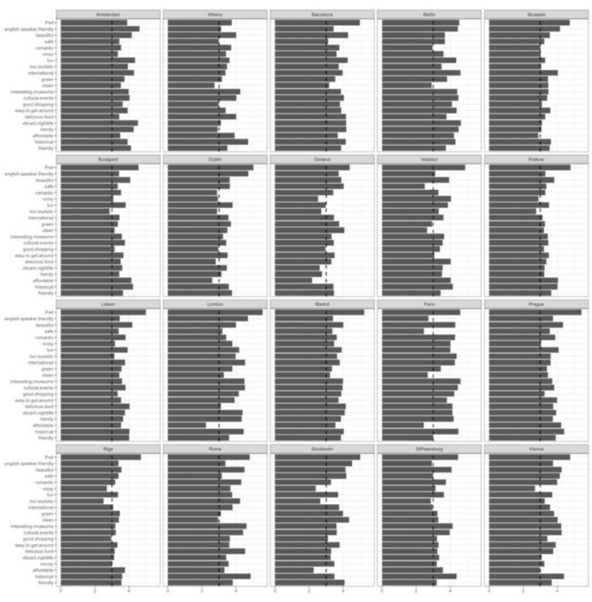


Figure 2: mean attribute and preference for each city

1.c- The Figure 1 and Figure 2 illustrate the mean attribute evaluations and preference across cities and mean attribute evaluations and preference for each city.

It is suggested from Figure 1 that London and Prague are the cities with highest preference while Athens with the lowest preference. It is clear from Figure 2 that London has high values in attributes like English speaker friendly, international and cultural events but low values in affordable. Athens has high values in attributes like historical and interesting museum but low values in clean and trendy.

2.a- We analyzed the similarity between 20 cities based on the 20 attributes using Dissimilarity on mean attribute evaluations using Euclidean distances for ease of interpretation. We did not choose the mean similarity measure across respondents because no city was rated by all respondents. We can calculate similarity on mean attribute evaluations as follows:

$$\frac{1}{N}\left[\sum_{k=1}^{K}\left|\sum_{n=1}^{N}(x_{ikn} - x_{jkn})\right|^{2}\right]^{\frac{1}{2}} \quad K = number\ of\ attributes, N = number\ of\ responsondents$$

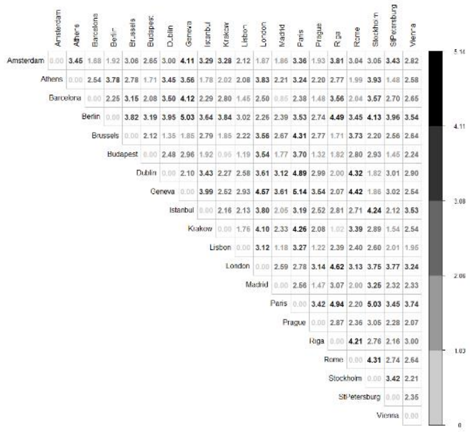


Figure 3: Dissimilarity on Mean Attribute Evaluations

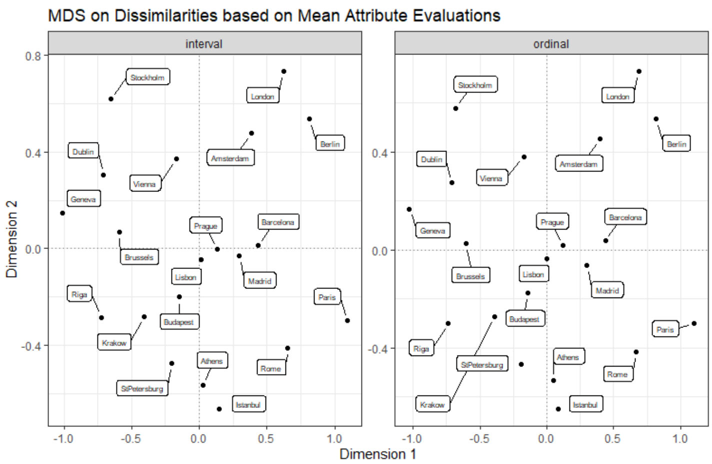


Figure 4: MDS on dissimilarity based on Mean Value Attributions

The data consists of 20 cities evaluated on 20 attributes and a preference scale. The 20 attributes are measured on a 5 point scale and the preference to visit a city is measured on a 7 point scale. Since the features on which cities are not evaluated on the same scale, we chose to scale the data before applying multidimensional scaling (MDS) to the data. We thus scaled the data for the ease of comparison and interpretation of preference with attributes on the perceptual map. We performed metric MDS with interval transformation and non-metric MDS (Figure 4) along with the dissimilarity map (Figure 3), on mean attribute evaluations. In the dissimilarity plot the higher values show higher dissimilarity between cities. From Figure 4 we can see that both metric and non-metric MDS have the similar perceptual maps. The dissimilarity map in Figure 3 closely aligns with both the MDS models.

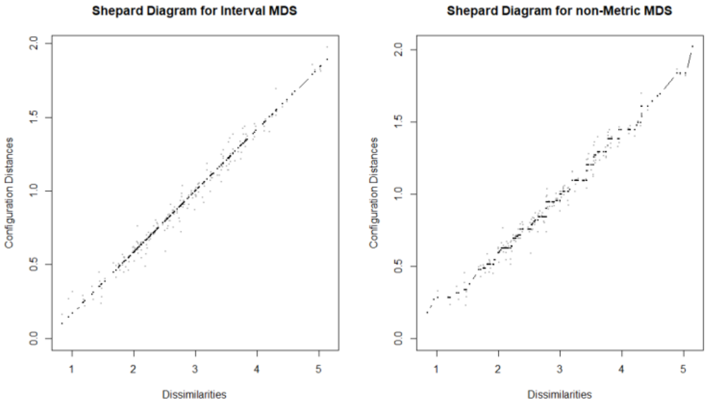


Figure 5: Shepard Diagram

	Interval		Ordinal	
Dim	Stress-1	r^2	Stress-1	r^2
1	0.29	0.73	0.25	0.73
2	0.11	0.93	0.09	0.93
3	0.06	0.98	0.04	0.97
4	0.03	0.99	0.02	0.99
5	0.02	1.00	0.01	0.99

Table 2: Comparison of MDS models

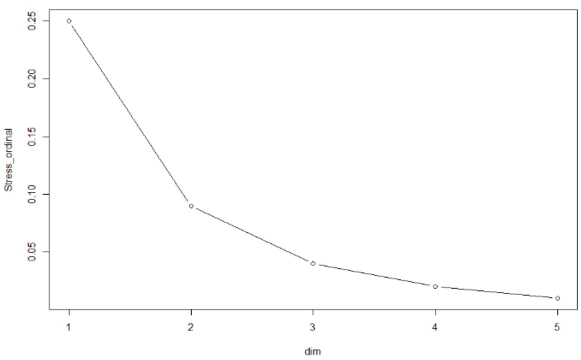


Figure 6: Elbow criterion

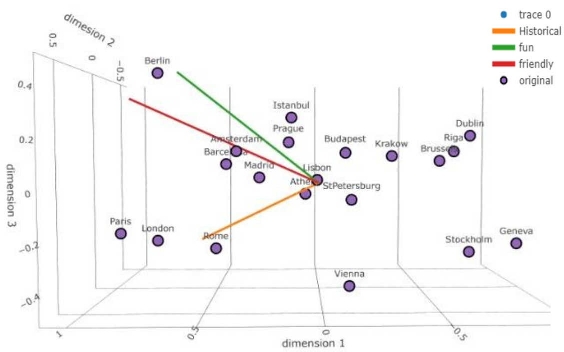


Figure 7: Metric MDS on dissimilarity based on Mean value attributions in 3 dimensions

Although on further analysis using the Shepard Diagram in Figure 5, we see that ordinal MDS has less variance as compared to the Metric Interval transformation. Also, from table 2, it is quite evident that ordinal MDS performs better than Metric MDS on both stress and r^2 values in all the dimensions. Furthermore, the data from the questionnaire is a likert scale, based on rating, which are perceptions of people. It is an ordinal scale as the difference between values of rating scale may not necessarily be the same. The ratings are perceptions of people and not real integers. We select non-Metric MDS on ordinal transformation for further analysis, which is least restrictive as it has least assumptions and has better performance in terms of less variance in data and less stress.

The mathematical formulation of ordinal transformation is given by –

$f(\cdot)$ is a monotone function such in non-metric MDS that if $\delta_{ij} < \delta_{kl}$ then $\hat{d_{ij}} \leq \hat{d_{kl}}$. i, j, k, and l represent different objects, δ_{ij} is the proximities which are a function of distance metric $d_{ij}(X)$,example Euclidean distance or city block distance, and X are the points of configuration of objects in m dimensional space. $\hat{d_{ij}}$ are the disparities which are the estimated distances from dissimilarity measures. Non-metric MDS says that if objects i and j are more similar than k and l then the estimated distance on the perceptual map of i and j is less than equal to k and l. Unlike, metric MDS, non-metric MDS does not try to fit true distances but to find the true rank order.

Coming to the selection of dimensions we can see from the elbow criteria of non-metric MDS in Figure 6, that there is a significant difference between stress for dimensions 1,2 and 3. The performance of dimension 3 on stress outperforms that of dimension 2. Although from Figure 7, we can see that the interpretability will be compromised in dimension 3. Unlike Factor Analysis which has only one vector for preference, MDS will have 21 vectors for attributes and preference which will be hard to represent in 3-dimensional space. In dimension 3, there are multiple rotations of the plot which makes it difficult to interpret the similarities between 20 cities on 20 attributes. Additionally, the non-MDS model in 2 dimensions has a r^2 = 0.93 which is almost the same as that of 3 dimensions (0.97), meaning it covers most of the data.

2.b-Property fitting is used to map the original attributes onto the derived MDS space. In this context, the value of attributes represents the degree of agreement and the rating of preference can be interpreted as ‘the higher the better’. But an ideal-point corresponds to an optimal value in the Ideal-Point Model. Therefore, the Ideal-Point Model is not applicable for preference. A joint mapping of perceptions and preferences can also be used in property fitting, which can show the preference of each respondent. It is clearer for joint mapping of perceptions and preferences to segment individuals and plot the preference vectors of the customer segment. In a vector model, objects are represented as points and attributes/respondents as vectors in the joint space. The direction of the vector indicates the direction of the increase in values of a particular attribute/respondent. Therefore the vector model is applicable for attributes and preference in this case.

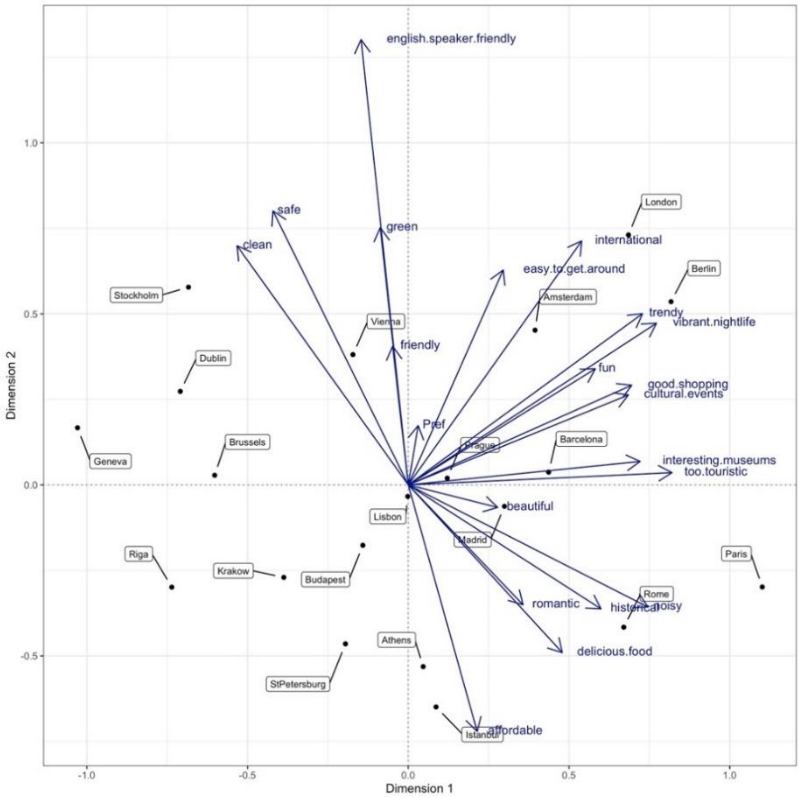


Figure 8: Property fitting of attributes and preferences: Vector Model

2.c- In the vector model, cities are represented as points and attributes/preference as vectors in the joint space. The direction of arrows indicates the increasing evaluation value of attributes respectively. The projections of the cities to a specific vector represents the order of cities by specific attribute.

It is suggested from Figure 8, that respondents have strongest preference on London and Berlin, but least preference on Riga. Respondents perceive Berlin, London and Amsterdam as the most international cities while evaluating Geneva and Riga as the least international ones, which corresponds to the description of the data in Figure 1. Although there is some noise in the ranking of cities on some attributes, the graph is generally interpretable and understandable.

3.a- We conducted Exploratory Factor Analysis (EFA) on this dataset to group the attributes and find out the underlying latent variables. We did not conduct Confirmatory Factor Analysis as we do not have a proper model or theory to combine the attributes. There are a total of 20 attributes and 1590 responses.

For performing factor analysis, we first checked the correlation between attributes. We found clusters of medium to high correlation between the attributes green, clean and safe; fun, trendy, good shopping and vibrant nightlife; historical, cultural events and interesting museums.

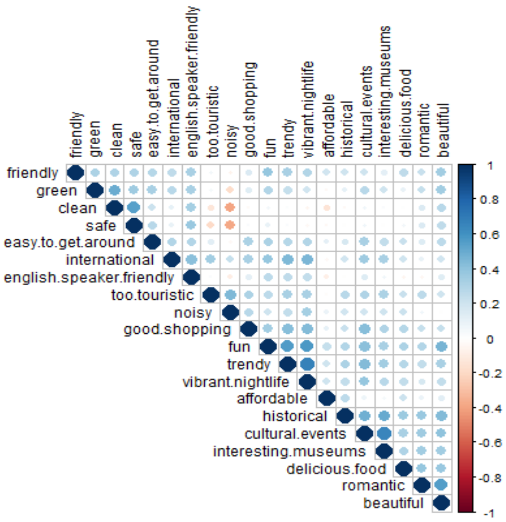


Figure 9: Correlation plot between the attributes

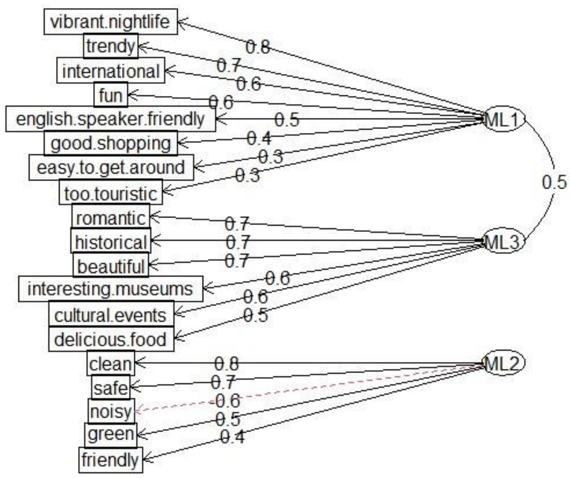


Figure 10: Path diagram for chosen model with three factors and Oblique Rotation

The evaluation of the attributes and the preferences were conducted on different scales. Therefore, we chose to standardize the data for comparability. After scaling the data, we assessed its fit by using Kaiser-Meyer-Olkin test (KMO) and Bartlett test. The overall Measure of Sampling Adequacy (MSA) was 0.88 with each items having MSA ≥ 0.7. The KMO scale varies from 0 to 1 and MSA>0.5 is considered adequate for Factor Analysis. Bartlett test of sphericity was also significant with $p<0.001$. This signifies that the data is suitable for EFA.

Determining the number of factors

To determine the number of factors, we generated a scree plot. From Figure 11, we can see that there are 4 factors which have eigenvalues greater than 1. As it is normally suggested that factors with eigenvalues greater than 1 should be taken into consideration, we decided to perform Factor analysis with at most 4 factors.

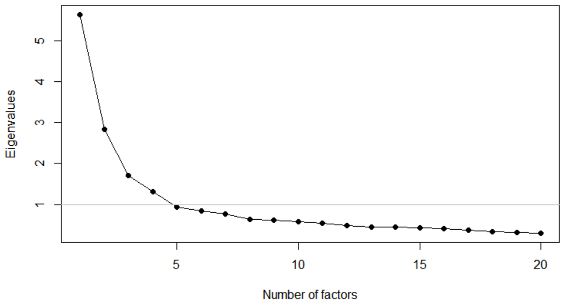


Figure 11: Scree plot considering all 20 attributes.

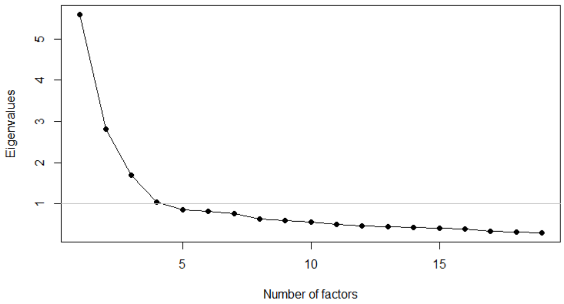


Figure 12: Scree plot after dropping attribute affordable.

Fitting the model

We performed EFA with both unrotated and rotated models. For unrotated model, the 3 factor and 4 factor models were really poor. In 3 factor model, the attribute 'affordable' was not assigned to any factors and also one of the factors was assigned only 1 attribute, 'English speaker friendly'. In 4 factors model there was no variables allotted to one factor.

For rotated models, these discrepancies were mostly resolved. We got a well distributed factor analysis with high factor loadings for 4 factor model, but the proportional variance of the 4th factor was very less, with a value of 0.07. (Table 3)

	ML1	ML2	ML3	ML4
SS loadings	2.99	2.52	2.44	1.46
Proportional Var	0.15	0.13	0.12	0.07
Cumulative Var	0.15	0.28	0.40	0.47

Table 3: Variance loadings vs Factors for 4 Factor model by taking all 20 attributes and applying Orthogonal Rotation

	ML1	ML2	ML3
SS loadings	3.22	2.92	2.39
Proportional Var	0.17	0.15	0.13
Cumulative Var	0.17	0.32	0.45

Table 4: Variance loadings vs Factors for 3 Factor model after dropping the attribute "affordable" and applying Oblique Rotation.

The variables that made up the 4th factor were 'affordable', 'romantic', 'beautiful' and 'delicious food' and we didn't think that these attributes made up for a meaningful factor. In both 2 factor and 3 factor rotated models the attribute 'affordable' was not assigned to any factors. On checking the correlation plot between the attributes from Figure 9, it was observed that 'affordable' has almost no correlation with any of the other attributes. So, we decided to drop 'affordable' as an attribute and continued with the other 19 attributes for performing EFA.

We made a scree plot again and this time we saw that there are 3 factors with eigenvalues more than 1 (Figure 12). We conducted a model fit using rotated and unrotated EFA for both 2 and 3 factor models. The 3 factors model gave much better insights in terms of grouping of attributes. We took the 3 factors as 'Metropolitan', 'Clean, safe and green' and 'Cultural Heritage'. For the 2-factor model, we were getting all the attributes of Cultural Heritage grouped along with Metropolitan as 1 factor. So, we took the 3 factor model as our main model.

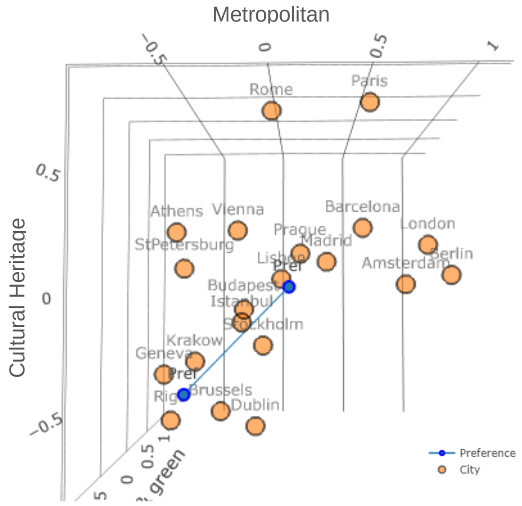
As for rotation, the unrotated model had very less variance (only 0.07) for the 3rd factor. Orthogonal rotation gave proportional variance of 0.13 for the 3rd factor. When we applied oblique rotation, apart from good proportional variance for all factors, the attribute 'English speaker friendly' was assigned with the factor Metropolitan in place of factor 'Clean, safe and green'. We think this attribute is more suitable with the Metropolitan factor rather than the 'Clean, safe and green' one. Oblique rotation also gave a correlation of 0.5 between the factors Metropolitan and Cultural heritage, which we thought to be a valid and useful correlation as Metropolitan cities often have a rich cultural heritage with museums, cultural events and a significant history.

From Figure 10, we can see the 3 factors as ML1, ML2 and ML3. ML1 factor is Metropolitan with the assigned attributes vibrant nightlife, trendy, international, fun, English speaker friendly, good shopping, easy to get around and too touristic. ML3 is Cultural Heritage with the attributes romantic, historical, beautiful, interesting museums, cultural events and delicious food. ML2 is Clean, green and safe with the attributes clean, safe, noisy, green and friendly. We got factor loadings of 0.3 or more for all the attributes, which justifies taking them into account. Only noisy is negatively correlated with the factor. All other attributes are positively correlated. The diagram further suggests correlation between factors ML1 and ML3 with medium loading of 0.5.

3.b- The subset of attributes that were associated to each factor were aggregated by their mean value by city to be used as dimensions for the perceptual map. A perceptual map in this case shows the positioning of cities on the factors. Lasty, the position of the respondents was represented by an augmentation of a preference in a vector model. Since the preference follows a 7-Point-Scale where seven is the most preferred and one is the least preferred a vector is more fit to showcase the preferences appropriately.

The perceptual map shows that most preferred cities are Vienna, London, Barcelona, Prague, Berlin, Amsterdam, and Madrid. Least preferred are Athens, St. Peterburg, Krakow Dublin, Riga and Brussels, Istanbul. One can see that the increasing direction of the vector is towards the positive values of axes "metropolitan" (Figure 13). (The actual figure showing the preferences appropriately was not included due to pages limits).

3.c- Figures 13 and 14 show that the cities that are associated as metropolitan Berlin, London, Amsterdam Paris. They are loading high on attributes such as fun, vibrant nightlife, trendy, international, English-speaker-friendly, too touristic. Those cities are also mapped as having a rich cultural heritage. The positive correlation of 0.5 (Figure 10) between the factors 'metropolitan' and 'cultural heritage' is also derivable in the map. The more metropolitan a city the more likely it is to have a larger cultural heritage. The map demonstrates that cities as Dublin, Brussels, Riga, Geneva, Krakow are scoring low on Cultural heritage. Not metropolitan but culturally rich are St. Peterburg, Athens, Vienna Rome. The safest, cleanest, and greenest city is Stockholm followed by Vienna, Geneva, Dublin. Istanbul and Paris are by far perceived as unclean, unsafe, and not green.



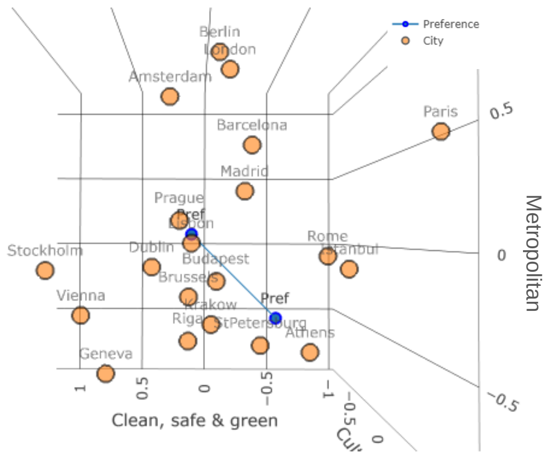


Figure 13: Augmented Perceptual Map (1)

Figure 14: Augmented Perceptual Map (2)

4.a- To make the results comparable and easier to interpret both analyses were conducted on a two dimensionality. The overall mappings we obtained from Multidimensional Scaling in 2 dimensions (Figure 8) and 2-factored Factor Analysis (Figure 15) are very similar. Figures embody mutual clusters as well as some variances. The similarities help us identify some core characteristics of the data while the differences highlight disparities in lost information caused by different methods in data reduction.

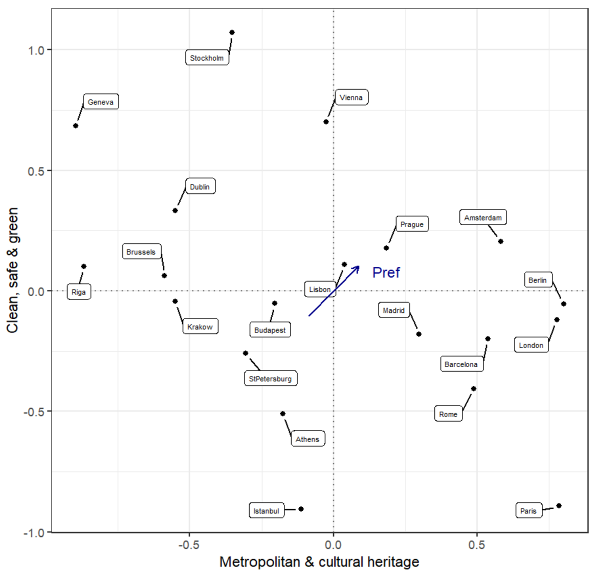


Figure 15: Perceptual Map - Factor analysis 2D

The first observable group in EFA is Stockholm, Vienna and Geneva, Dublin. These cities rank the highest in the factor 'Clean, safe & green', i.e., they were ranked as clean, safe, and green, less noisy, English-speaker-friendly, and friendly. At the same they are positioned in the lower ranking of the second factor 'Metropolitan & cultural heritage'. Stockholm is ranked the most and Dublin is ranked the least in ‘clean, safe & green’. MDS is similar distributed. In EFA Geneva is more similar to Stockholm than in MDS. Different to EFA, MDS groups Brussels into that cluster. Because Dublin, Geneva, Vienna, and Stockholm are correlating more positively with the factor ‘clean, safe & clean’, and Brussels is just breaking the zero point it was more appropriate to group Brussels with Riga and Krakow only in EFA.

The next group contains cities that correlate negatively with the factor ‘Clean, safe & green’ as well as seem less metropolitan and offer less cultural events and nightlife in EFA. Similar attributes can be derived of the perceptual map that was computed from MDS. Although e.g., St. Petersburg is perceived as more metropolitan than Riga, both are scoring low in comparison to e.g., London and Berlin. This cluster is not showing any strong suits in comparison to other clusters. In both EFA analysis Riga and St. Petersburg seem the least similar to each other within that cluster in EFA analysis. In MDS St. Petersburg is nor grouped with Krakow, Riga, and Budapest. The distance measure is smaller to Athens and Istanbul.

Another Cluster of cities is Athens and Istanbul in EFA. Similar to the group of cities in the previous paragraph, it is correlation negatively on cleanliness, safety and greenness in EFA. Istanbul ranks lower than Athens on Cleanness, safeness, and greenness. For MDS we can see similar results. A dissimilarity from MDS to EFA, is that St, Petersburg is suggested to be more similar to Athens and Istanbul.

Next Cluster of cities can be seen located in the center of the perceptual map in MDS in Figure 8 consists of Prague, Lisbon, Madrid, and Barcelona. These cities are on average positively associated with trendy, vibrant nightlife, fun, good shopping, cultural events, beautiful, interesting museums, touristic, historic, delicious food, romantic, clean, safe, and green. Although, Prague is the most affordable city as can be seen in Figure 2 and is misrepresented in MDS. This happens due to the loss of information in a two-dimensional space. The findings from MDS for this group of cities are quite similar to that of factor analysis except that Barcelona seems to be dissimilar from the group of Prague, Madrid, and Lisbon. From Figure 15 it’s evident that Prague, Madrid, and Lisbon have high values on both factors. This is in coherence with our findings from the MDS model.

Let us consider the cities Amsterdam, Berlin, and London, which are grouped in both MDS and EFA. They are located near the international, trendy, multicultural, vibrant nightlife, shopping, fun, easy to get around, friendly, shopping, and fun vectors in the Vector Model of MDS. The cities have a very high value on factor metropolitan & cultural heritage which is in parallel to the MDS model. Although, we see that Barcelona has a high value on the x-axis and this makes it similar to London, Berlin, and Amsterdam in the perceptual map of Factor analysis. It can be discernible from Figure 3 that Amsterdam, Berlin, and London are similar to each other.

Lastly, we can see from both MDS and Factor Analysis is Rome and Paris. Rome and Paris are located near delicious food, romantic, historic, romantic, and beautiful vectors in the vector model of MDS in Figure 8. We can also see from the factor analysis that Rome and Paris have a higher value for the factor metropolitan & cultural. Also, from Figure 15 we see that they are low on the factor clean, safe & green, which is similar to the findings of the MDS model. From the dissimilarity matrix in Figure 3 we see that Rome and Paris are quite similar with a distance measure of 2.20. However, Figure 8 shows Istanbul and Athens to be rated more than Paris and Rome on Romantic, beautiful, and historic which is inaccurate. Paris and Rome both appear to be affordable from the perceptual map of MDS but are not affordable as is apparent from Figure 2. Coming to the preference rating of the cities we see in Figure 15 that Stockholm is the most preferred city followed by Vienna, Geneva, Dublin, Amsterdam, and Prague in the order they are mentioned. This is not in coherence with Figure 2, which shows that London is the most preferred city followed by Prague and Madrid. Unlike Factor analysis, MDS correctly depicts London as the most preferred city to travel to, which is precise. But erroneously represents the preference of Prague and Madrid which are ranked quite low on preference.

4.b- There is some loss of data in both MDS and Factor analysis, but they both can capture the essence of the data quite well. Both methodologies are giving similar results in the given scenario. One drawback of MDS worth noting is that dimensions are undefined and labelled based on subjective judgment. Factor analysis requires that the relationships of underlying data are linear and distributed as multivariate normal. In our given data and generally, MDS performs better in terms of readability and interpretability of results in comparison to factor analysis, which tends to extract more factors (dimensions) than MDS. The 2-dimension comparison reveals that: in MDS all attributes are separately plotted through vectors, thus MDS is more insightful than EFA. In EFA we cannot tell which attributes from 'clean, safe & green' make Paris low in the y-axis value, but it is clear in MDS for each attribute. This is where MDS performs better. In this 2D comparison, MDS can be recommended for dimensionality reduction.

5- One can argue that if an airline tries to target consumer segments based on this model, they should definitely target the cities such as London, Berlin and Amsterdam as this is the most preferred group as can be seen in Figure 8 and 15. These cities are more metropolitan high on vibrant nightlife, trendy and fun. Average age of respondents in the survey is 25.46, with the 50% master students. This can be a potential target segment of customers. They could increase the number of flights on the weekend to these destinations. They could also run some promotional campaigns for these cities to attract tourists where they offer discounts to clubs or bars and also list specific events in the cities. The airline company can offer discounted flights to the cities in the same cluster if customers have already visited one in the cluster before. For example, if the customer has visited Prague, they may have high preference on trendy, vibrant nightlife and fun, the possibility to book a visit to Lisbon, Madrid and Barcelona would also be high.