International Journal for Quality Research 14(1) 271–290
ISSN 1800-6450

Burçin Güçlü[1]
David Roche
Frederic Marimon

CITY CHARACTERISTICS THAT ATTRACT AIRBNB TRAVELLERS: EVIDENCE FROM EUROPE

Article info:
Received 13.07.2019
Accepted 12.11.2019

UDC – 366.544
DOI – 10.24874/IJQR14.01-17

Abstract: *This paper investigates the reviews posted by Airbnb customers in order to assess the customer satisfaction, and to understand the criteria of Airbnb customers in short-term accommodation rentals. The objective is to determine the characteristics that customers find important, and to provide the typologies of cities from the customer perspective.*

The analysis holds under scope the five most touristic cities in Europe. The opinion of Airbnb customers were retrieved from the Airbnb website. First, the satisfaction was assessed using sentiment analysis. Second, main characteristics that describe the tourism experience were defined through a set of Exploratory Factor Analysis. Next, segmentation of cities according to these features was settled.

Results indicate that Airbnb customers are satisfied with the service they use, and that their choice of short-term accommodation on Airbnb is a multi-criterion decision process. The results of our research are of utmost importance for governmental institutions, tourism agencies, and Airbnb.

Keywords: *Sharing economy platforms; Service quality assessment; Text mining; Sentiment analysis; Airbnb.*

1. Introduction

Sharing economy is taking momentum and becoming a more common model for transactions (Belk, 2014), especially in the hospitality and tourism sector (Guttentag, 2015; Zervas et al., 2017). Many platforms matching peer accommodation suppliers and potential tourists have emerged in the last decade. One of such revolutionary platforms is Airbnb, founded in 2008, operating in the entire globe and offering all kinds of host services. Every day, Airbnb customers hosted in 33,000 cities across 192 countries post millions of reviews assessing the quality of and rating the services. More and more data of this kind are publicly and privately available, but little is known about how these data can be explored to assess the customer satisfaction, and to understand the main criteria of Airbnb customers in short-term accommodation rentals.

In this paper, we investigate the reviews posted by Airbnb customers in order to assess the customer satisfaction, and to understand the criteria of Airbnb customers in short-term accommodation rentals. We also take into account the robustness of the analyses for different cities. Particularly, we hold under scope the five most touristic cities in Europe. First, we quantify the level of satisfaction in Airbnb feedbacks using sentiment analysis. Second, we define the main factors that describe the tourism experience for each city, and consequently match these city attractiveness factors to the customer

[1] Corresponding author: Burçin Güçlü
Email: burcin.guclu@salle.url.edu

satisfaction analysis.

Although previous research has begun focusing on issues in the sharing economies, the customer satisfaction and short-term accommodation criteria in the sharing economy have been still overlooked. At one hand, the very fact that the transaction is online allows storing digital information for future analyses. The customers are expected to share their opinion in the website of the platform company because of the i) self-regulatory nature of digital platforms (M. Cohen & Sundararjan, 2015), and ii) need to improve the reputation of the peer provider, which is an essential factor in the digital platforms (Ert et al., 2016; Farajallah et al., 2016; L. Liu, Cheung, & Lee, 2016). On the other hand, the use of more comprehensive data and digital decision-making tools becomes part of strategic decision-making (Baralou & Tsoukas, 2015), allowing the analysis of customer opinion available in the website.

Our results indicate that Airbnb customers are generally satisfied with the service they use, and that their choice of short-term accommodation on Airbnb is a multi-criterion decision process based on city- and accommodation- related factors. The results of our research are of utmost importance for governmental institutions, tourism agencies, and Airbnb. The tourism is becoming an important industry and some governments have their own ministry or official departments, which need to take into account this new business model when they issue and revise existing policies. Similarly, there are cities competing for tourists at the city-level, and their tourism offices need to acknowledge this new phenomenon. The tourism department of the main cities are interested in the analysis of their visitors in order to prepare an interesting service for their tourist and to get a high tourist satisfaction, which translates in higher loyalty, higher recommendation, and longer stay in the city.

This paper is first, to our knowledge, to address the service quality assessment in the context of sharing economy platforms in tourism industry, and to conduct complete empirical research using text-mining techniques. Since both the framework of sharing economy and research methodologies based in big data are quite recent, the paper contributes on grounding both: (i) the conceptualization of city and accommodation features of hosting service in sharing economy and (ii) establishing a research methodology based in real opinions of thousands of customers.

The structure of the article is as follows: First, we devote the next section to the literature review on sharing economy platforms and service quality assessment in peer-to-peer evaluation. In the section after, we describe the methodology chosen to study empirically the motive of our paper. We present the results in the proceeding section, and carry out the discussion to fine-tune our findings and insight for future research, and contribution to the literature before we conclude with final remarks.

2. Literature review

2.1. Sharing economy

By sharing economy, we refer to systems of transactions where sharing access to certain goods and services are moderated via online platforms (Hamari et al., 2016). Sharing economy platforms are capable of i) bringing more options and efficient prices for the buyer, ii) reach large segments of the population through user communities, and iii) make use of otherwise under-utilized privately-owned goods via a fee-based online service (Zervas et al., 2017). In doing so, they are disrupting the incumbent traditional business models in their industries. Sharing economy platforms work as intermediaries between the consumer and the producer (Puschmann & Alt, 2016), easing a type of trusted transaction between these two actors secured by the platform (Calo & Rosenblat, 2017).

Sharing economy as a phenomenon has received wide attention from different perspectives (Habibi et al., 2017; Hamari et al., 2016). It is a new model of transaction that provides a new way of consumption. Academics have been conceptualizing it, but here is not yet a consensus on this new concept. Many different labels have been coined in order to differentiate close models (Belk, 2014; Benoit et al., 2017). On one hand, it is an interesting attempt that illuminates the new phenomenon, but on the other hand, it puts more doubts and confusion. Benoit and colleagues (2017) shed light on the distinction of models, providing a good differentiation among some scenarios that are certainly close, but have special characteristics that makes each one unique.

The sharing economy businesses show differences according to the sectors of activity, which range from tourism, transport, education, retail, music, logistics, restauration, to many others. However, there is a couple of sectors that have been intensively analyzed (hospitality and transport), and particularly two companies: Airbnb (Cusumano, 2014; Ert et al., 2016; Guttentag, 2015; Matzler et al., 2014; Möhlmann, 2015; Priporas et al., 2017; Zervas et al., 2017) and Uber (Benoit et al., 2017; B. Cohen & Kietzmann, 2014; Cusumano, 2014; Leighton, 2016). Airbnb and Uber, two Silicon Valley startups that were founded in 2008 and 2009 respectively, today rank among the largest sharing economy firms in terms of market evaluation and highest expansion worldwide (Uzunca et al., 2018).

2.2. Sharing Economy in Lodging

Tourism is an activity where collaborative consumption can easily use for their transactions. Barnes and Mattsson (2016) have documented a list of drivers for this model, and most of these drivers are present in tourism. The real fact is that Airbnb is increasing its business volume each year.

On the other hand, literature also analyzes extensively the tourism sector and its evolution (Belk, 2014; Zervas et al., 2017). It also investigates the main factors at macro level, and the extant literature is full of case studies, which focus in particular issues, as the impact on employment (Berbegal-Mirabent et al., 2016; Fang et al., 2016; Tussyadiah, 2016) and consumer welfare (Koopman et al., 2015). Needless to say that there is a plethora of very specific analyses, attending typologies of tourisms (e.g., Assaf & Josiassen, 2014; Getz & Page, 2014).

This new business model is competing with the traditional hotels based on a differentiated strategy. The collaborative consumption is geared among three agents: the website that makes the match between supply and demand; the particular peer who owns an apartment or lodge and offers the host service; and the consumer, who requires a room or apartment (Benoit et al., 2017; Ertz et al., 2016; Möhlmann, 2015). In this triode, agents have a particular role and each of them looks for particular benefit.

Other agents are affected by the emerging of this new hospitality consumption. The traditional hotels stress out the competence of some of their customers that now are migrating to this new service, which provides something different that the hotel does not provide, the direct contact with the peer. Among the reasons that explain the change of consumption model (from traditional hotel to a platform like Airbnb), it can be found the economic, but this is not the only one. Collaborative consumers also value the interaction with the peer who offers the room. Other tourism activity players are also affected. Travel agencies, which traditionally were matching supply and demand, are competing now with these platforms. Other institutions at governmental level that are influencing the tourism sector are also affected.

2.3. Service Quality Assessment in Sharing Economy

One of the main streams of research is related to the quality of these services. Some papers propose measurement scales, composed by different dimensions, which makes is easier to understand that this is a multifactorial construct (Bardhi & Eckhardt, 2012; Benoit et al., 2017; B. Cohen & Kietzmann, 2014; Del Mar Alonso-Almeida et al., 2014; Schaefers et al., 2016). Marimon et al. (2019), based on this previous literature and on an extensive empirical work propose an instrument valid for any platform operating in the sharing economy, regardless the specific activity sector. They find some quality dimensions related to the platform, and others to the peer server. Similar analysis investigates the impact of perceived quality on other constructs such as loyalty or satisfaction (Cheng et al., 2018).

The interest of the sector analysis at national level is of paramount importance. Thus, the extant literature devoted to find out the motivations of particular destinations, or devoted to analyze the profile of tourists is important. Each country, and each city deploy a tourism strategy in order to attract the segment interesting for their purposes. One determinant issue in all this is the analysis of their current tourists: their demands, their capacity to consume, their preferences, etc... (Farmaki et al., 2015; Nunkoo, 2015).

These collaborative consumers are uploading their experiences, and are assessing and rating the services consumed. More and more data are available and ready for the analysis. In this vain, some articles are publishing results based on big data analysis (e.g., Batista e Silva et al., 2018; Mariani et al., 2018); others on meta-analysis based on previous studies. However, there is still lack of papers that using this data are providing the main characteristics of market segments tourists of a particular city.

3. Methodology

Data collection and most of the analysis were conducted by R, which is a very flexible and trending program due to the availability of packages for specific purposes of data gathering and mining. The opinion of 3,689,879 Airbnb customers were retrieved.

3.1. Sample

Data for this study come from a data portal called Inside Airbnb (Inside Airbnb, 2017). Inside Airbnb is an independent, non-commercial portal of analytics and datasets that allows exploring how Airbnb short-term accommodation services are adapted and diffused worldwide. By analyzing publicly available information about a city’s Airbnb's listings, Inside Airbnb provides filters and key metrics that bring to daylight how Airbnb competes with the residential housing market and lodging industry.

Inside Airbnb portal shares compiled information from Airbnb web-site including the listings, availability calendar for 365 days in the future, neighborhood lists, and the reviews for each listing. Data are verified, cleansed, analyzed and aggregated by Inside Airbnb at city level. The portal uses public information disclosed on Airbnb website, and aims producing a non-commercial derivation to allow public analysis, discussion and community benefit. However, this site is not associated with or endorsed by Airbnb or any of Airbnb's competitors. Moreover, accuracy of the information compiled from the Airbnb site is not the responsibility of Inside Airbnb although due care has been taken with any processing and analysis.

The data presented here are a snapshot of listings available at a particular time. Other snapshots of data from previous dates are available for analysis by request, and a future activity may be to analyze the characteristics of deleted Airbnb listings.

We choose five cities out of the list of most touristic cities in Europe according to Lonely Planet (Lonely Planet, 2017). Table 1 below presents the reviews from several Airbnb accommodations in these cities in detail.

Table 1. Airbnb accommodations and reviews from five cities in Europe

City	Date posted	Number of accommodations	Number of reviews
Barcelona	July 10th 2018	14,748	553,309
Istanbul	July 30th 2018	5,964	87,105
London	August 8th 2018	53,428	1,061,389
Paris	July 8th 2018	44,129	1,074,759
Rome	July 11th 2018	22,604	913,317
Total		**140,873**	**3,689,879**

3.2. Data Analysis

The purpose of our research is to understand the feedback from customers, and to reveal the criteria of the Airbnb customers when choosing a short-term accommodation. For this reason, we apply content analysis (Krippendorff, 2013) and latent semantic analysis. The content analysis comprises of i) sentiment analysis to assess satisfaction feedback, and ii) text mining to understand main concepts arising in the feedback, and iii) principle-component analysis to assess the predictive power of each concept and the dimensions of the preferences.

We do the programming of the data collection and most of the analysis by R (R Core Team, 2017), which is a very flexible and trending program due to the availability of packages for specific purposes of data gathering and mining. After downloading and importing the aforementioned reviews to R, we have focused our attention on reviews in English. In order to retrieve the language of a review automatically, we have used cldr package (McCandless et al., 2013) in R which reports the detected language of a text with high accuracy, and whether the retrieval is reliable. For instance, in Barcelona review data, the package scans 553,309 reviews, and reports that 344,121 of these reviews are detected in English and the detection is reliable. In order to achieve highest accuracy in the data and best prediction in the upcoming analyses, we have only retained reviews in English with reliable detection in each database. Given that English is the common language among travelers who express their opinions and share their experiences on online platforms, we have a good representation of the total reviews in the dataset when we restrict our attention to the reviews in English. Table 2 below reports the reviews retained and their relative frequency in the full dataset.

Table 2. Airbnb reviews from five most touristic cities in Europe used in data analyses

City	Number of accommodations	Number of reviews	Number of reviews in English and reliable	%
Barcelona	14,748	553,309	344,121	62.2%
Istanbul	5,964	87,105	66,204	76.0%
London	53,428	1,061,389	866,401	81.6%
Paris	44,129	1,074,759	645,719	60.1%
Rome	22,604	913,317	598,869	65.6%
Total	**140,873**	**3,689,879**	**2,521,314**	**68.3%**

3.2.1. Sentiment Analysis

Our initial analysis comprises of sentiment scoring, which allows us to quantify customer satisfaction. In order to score the sentiments communicated through Airbnb reviews posted on accommodation feedback portal, we employ two dictionaries which consist of positive words and negative words lists respectively (Hu & Liu, 2004; B. Liu et al., 2005). After configuring the lists of positive and negative words, we run an algorithm which assigns 1 if a word is either encountered in positive words dictionary or negative words dictionary, and 0 if otherwise. We finally compare our words to the dictionaries of positive and negative terms and obtain the sum of positive and negative matches respectively. Subtracting the sum of positive matches from negative matches returns the sentiment score, which can be either a positive number, negative number, or zero.

3.2.2. Text mining

In order to find key terms in the Airbnb reviews, we conduct text mining as a next step. We perform all the preliminary steps with the R package tm (Feinerer & Hornik, 2018; Meyer et al., 2008). We build a corpus for each of the 2,521,314 reviews in order to follow the standard procedures for text mining (Weiss et al., 2005). First, we clean punctuation and spaces from the text. Second, we convert the text to lower case in order to enhance identification. Third, we drop stop words based on the list (488 terms) of the SMART information retrieval system (Lewis et al., 2004). Fourth, we apply stemming based on the Snowball stemmer algorithm (Porter, 1980).

Document-term matrices tend to get very big already for normal sized data sets, which make them inefficient. As a result, we remove sparse terms, meaning terms occurring only in very few documents. Normally, this reduces the matrix dramatically without losing significant relations inherent to the matrix. Before choosing the desired level of sparsity, we have done a scenario simulation of sparsity level and terms remaining in the matrix, together with the total variance explained. We have noticed that there is a trade-off between the total variance explained and the quantity of terms remaining in the matrix, as in Table 3 below. To be more specific, the quantity of terms decreases with more variance explained.

Table 3. Trade-off between the total variance explained and the quantity of terms

	Barcelona	Istanbul	London	Paris	Rome
Sparsity=0.95					
Terms	104	111	95	101	121
Cum. Variance	14.5%	14.4%	15.1%	15.5%	13.2%
Sparsity=0.90					
Terms	33	41	34	35	39
Cum. Variance	28.6%	25.5%	27.7%	27.6%	26.1%
Sparsity=0.85					
Terms	20	20	18	19	22
Cum. Variance	40.0%	40.3%	42.8%	42.7%	37.5%
Sparsity=0.80					
Terms	12	12	10	12	11
Cum. Variance	60.0%	59.6%	68.1%	59.6%	63.8%

We have chosen the sparsity level of 0.90 for the Principle Component Analysis. Consequently, we pick only the terms with threshold level of 90 percent sparsity in each corpus in city database. Table 4 below summarizes the threshold value of sparsity chosen and terms retrieved by the sparsity algorithm.

Table 4. Threshold sparsity chosen and quantity of terms retrieved

City	Number of reviews in the data analysis	Threshold sparsity chosen	Terms retrieved by the algorithm	Cumulative variance
Barcelona	344,121	90%	33	28.6%
Istanbul	66,204	90%	41	25.5%
London	866,401	90%	34	27.7%
Paris	645,719	90%	35	27.6%
Rome	598,869	90%	39	26.1%

Next, we build a term-document matrix, with terms retrieved in the columns and documents (reviews) in the rows, in which each entry (I,j) is equal to the number of times the term j occurs in document i. In the second step, the relevant terms selected for each city were counted and used for an Exploratory Factor Analysis (EFA) which was performed to determine the factors suggested. Five EFAs were conducted, one for each city. These EFAs were analyzed using principal component analysis to explore the natural latent dimensions that emerged.

Third, the common factors among cities were analyzed in order to group the cities in terms of similarity according to the preferences of their visitors.

3.3. Results

3.3.1. Sentiment Analysis

We quantify the sentiments inherent in Airbnb reviews by applying the aforementioned algorithm. We find that the sentiments communicated through the reviews are on average positive, around 5, meaning that the customers make five positive assertions on average per comment. The satisfaction level from Airbnb service is highest for Rome, Barcelona, Istanbul and Paris, and lowest for London. The range of sentiment scores is wide, varying from -37 in Paris to 60 in Rome, but half of the sentiment scores lie between 3 and 7 in all five cities. Table 5 below presents the descriptive statistics on sentiment scores.

Table 5. Descriptive statistics on sentiment scores per city

City	N	Mean	Std. Dev.	Min.	Max.	1st Quartile	Median	3rd Quartile
Barcelona	344,121	5.416	3.577	-24	47	3	5	7
Istanbul	66,204	5.340	3.905	-21	45	3	5	7
London	866,401	5.002	3.512	-28	51	3	4	7
Paris	645,719	5.268	3.787	-37	54	3	5	7
Rome	598,869	5.589	3.717	-27	60	3	5	7

Median sentiment score is 5, slightly more than the average sentiment score in each city, signaling right skewness of the sentiment scores. The histograms of sentiments scores for each city can be found in Appendix A.

We also look at the sentiment polarity, meaning how positive, negative and neutral sentiments are dispersed. In order to carry out this analysis, we count positive, negative and neutral sentiments, and table each of these sentiments separately. We find that the dominant sentiment is positive for all cities, most common in Rome, Barcelona, London, Istanbul, and least common in Paris with respect to the relative frequencies reported in percentages. Negative reviews, on the other hand, do not occupy more than 1.50 percent in any of the city data: We find that the negative comments are most common in Istanbul, Barcelona, London, Paris, and least common in Rome with respect to the relative frequencies reported in percentages. Table 6 below presents the results from sentiment polarity.

Table 6. Sentiment polarity per city

City	Negative	%	Neutral	%	Positive	%	Total
Barcelona	4334	1.26	12,978	3.77	326,809	94.97	344,121
Istanbul	957	1.45	3325	5.02	61,922	93.53	66,204
London	9911	1.14	37,620	4.34	818,870	94.51	866,401
Paris	7296	1.13	38,299	5.93	600,124	92.94	645,719
Rome	5491	0.92	17,265	2.88	576,113	96.20	598,869
Total	**27,989**	**1.11**	**109,487**	**4.34**	**2,383,838**	**94.55**	**2,521,314**

3.3.2. Text Mining

In order to retrieve the predictive power of each term on city basis, we conduct five independent Principle-Component Analyses using R package irlba (Baglama et al., 2018), which provides a fast way to compute partial singular value decompositions (SVD) and principal components analysis (PCA) of large sparse or dense matrices. For instance, we initially retrieve the following 39 terms from Barcelona reviews: also, apartment, Barcelona, can, city, clean, close, comfortable, definitely, easy, everything, flat, friendly, good, great, helpful, host, just, located, location, lovely, metro, nice, perfect, place, really, recommend, restaurants, room, stay, time, walk, and well. Lists of terms retrieved for each city can be found in Appendix B.

We retain the terms according to the following criteria: First, a term needs to gradually load on a factor by 0.20 or more by magnitude. If a factor loads by 0.20 by magnitude or more on more than one factor, that term is not useful for the analysis. Second, neutral terms, i.e. terms whose meanings are ambiguous are discarded from the analysis. For example, we retain the following 10 terms from Barcelona reviews: Clean, comfortable, flat, friendly, great, helpful, host, nice, restaurants, and time. Lists of terms retrieved for all cities can be found in Appendix C along with the Principle – Component Analysis conducted.

After the analyses of these factors (attending to the words included), some factors from different cities were identified with similar content. Thereafter, the same label was assigned to different factors of different cities that appeared to be close in terms of content.

Consequently, we enlist 6 dimensions named after the remaining terms:

- F1 – City's image
- F2 – City's entertainment
- F3 – Apartment's location
- F4 – Tangibles & host characteristics
- F5 – Rate of the apartment
- F6 – Rate of stay.

The factors listed above can be found in Appendix C. The labels of factors in Appendix C refer to each city, and six factors have a second label (F-i) to identify those similar factors that emerge in different cities. Table 7 below presents these six factors that are common across some cities.

Table 7. Touristic profile of each city

	Factors	Barcelona	Istanbul	London	Paris	Rome
City	F1. City's Image			X	X	X
	F2. City's Entertainment	X	X			
Housing	F3. Apartment's location		X	X		X
	F4. Tangibles & host characteristics	X	X	X	X	
	F5. Rate of the apartment	X	X	X	X	
	F6. Rate of the stay					X

Paris and Rome are quite similar in their attractiveness for Airbnb. To be more specific, they are both attractive in one city dimension and in two apartment dimensions. Istanbul and London also show similarity in a sense that they are both attractive in one city dimension and in three apartment dimensions. Barcelona, on the other hand, is a particular case with one city dimension and two apartment dimensions. It is not far from Istanbul because they share the same three dimensions, but it is far from London because it has a different profile.

3.3.3. Robustness Checks

In order to validate the robustness of our results, we replicated the aforementioned PCA on the aggregate database (Table 8). Lists of terms retrieved for each city can be found in Appendix D along with the Principle – Component Analysis conducted. Total variance in the document-term matrix was 9.404, meaning that 34 terms retrieved accounted for 27.65 percent of the total variance.

Table 8. Touristic profile of each city

	Barcelona	Istanbul	London	Paris	Rome	Variance	Variance explained
PC1	-0.0293	0.0384	0.2515	-0.1225	-0.2192	2.8791	0.3062
PC2	0.0445	0.0621	0.0762	-0.0662	-0.0713	1.4657	0.1559
PC3	0.0297	-0.2032	-0.3028	0.2929	0.1277	1.4016	0.1490
PC4	-0.0359	-0.1004	-0.1490	0.1796	0.0537	1.2999	0.1382
PC5	0.0868	0.0063	-0.0384	-0.0433	0.0516	1.2066	0.1283
PC6	0.0982	0.1441	-0.0478	-0.0409	0.0408	1.1508	0.1224

The sum of the proportion of variance explained by the first three components accounted for 61.11 percent of the variance, indicating the adequacy of a 3-D plot of the cities for visualizing. Figure 1 below depicts the clusters of cities.

The results confirm our previous findings: Paris and Rome are quite similar in their attractiveness for Airbnb as they are plotted very close to each other. Istanbul and London are apart from the cluster of Paris and Rome, but they show closeness among each other. Barcelona is plotted on its own, and it is not far from Istanbul because they share the same three dimensions, but it is far from London because it has a different profile.

4. Discussion

The sentiment analysis findings are satisfactory. We find that the sentiments communicated through the reviews are on average positive, around 5, meaning that the customers make five positive assertions on average per comment. We find that the dominant sentiment is positive for all cities, meaning that Airbnb service is satisfactory for customers in the European market.

Text mining and Principle – Component Analysis provide exploratory findings. The main conclusions drawn are three. First, experiences of Airbnb consumers are multifactorial. In all cases (in all cities), there is a factor related to the city and other factors to the accommodation facilities. Both groups of factors are important.

Second, the main factors related with to the city are: (i) City's Brand and (ii) Entertainment in the city. Both are exclusive. Each city has only one of them. Customers are attracted by the city brand or by the entertainment. This is particularly significant in order to describe what are the common characteristics of the Airbnb tourists and what they are looking for in the focal city. The tourism factors related to city brand consist of looking for museums, art galleries, theaters, special monuments and buildings that are only found in this city; whereas the tourism factors related to entertainment consist of looking for restaurants, beaches, spectacular activities, and gastronomy.

Third, the factors related to the apartment are: (i) Apartment's location, (ii) Tangible & host encounter, (iii) Rate of the apartment, and (iv) Rate of the stay. These factors encompass both the tangible dimensions of the accommodation facilities and the host encounter experience.

Fourth, we can establish a category of city according to the factors related to the city. Barcelona and Istanbul are cities whose Airbnb tourists are looking for entertainment, whereas London, Paris and Rome are cities whose attractiveness rests in the city brand.

All in all, the dimensions “Tangible & host characteristics” alongside with “Rate of the apartment” are the most demanded. Both are related to the housing and accommodation, and as well as to the city. Additionally, in all these five cases, there are at least two dimensions related to the housing. We can conclude that in any case, the housing has the highest priority over other issues related to the city.

On the other hand, only one city factor is stressed out for each city. In two cities, the city factor is entertainment, whereas in other three is image. The reason that travelers have in mind to choose destination is “image” or “entertainment”. They chose destination according to that, and the second decision is the housing, and the criteria are different and independent from the first decision (destination).

5. Future research

This research can be extended both in theoretical and empirical dimensions. We consider the following extensions of the theoretical research: Despite the increasing popularity of sharing platforms, very little is known about why customers choose these platforms and how satisfied they are. Given the lack of government regulation and control, feedback received from customers plays an essential role in assessing customer preferences and service quality. This research particularly focuses on the former, while the latter can be addressed in a future research.

We also consider the following extensions of the empirical research: First, we can carry out a latent class analysis to identify latent concepts. To be more specific, the latent concepts can be combinations of terms with different loadings that would allow the profiling of such latent dimensions. We plan to use the R package lsa (Wild, 2015), which follows the ideas of Landauer and colleagues (1998). Second, we can apply a cluster analysis based on the spherical *k*-means algorithm (Dillon & Modha, 2001). We plan to use the R package spherical k-means (Hornik et al., 2012). Third, we also plan to explore different normalizations of the term frequencies to get the maximum profit from the analysis. The last but not the least, we plan to conduct a cross-classification of topics versus clusters and concepts.

6. Conclusion

The main contribution of this study is to assess holistically customer preference for choosing Airbnb. To our knowledge, this is the first attempt to propose the criteria for choosing Airbnb. Our study is valuable for governmental institutions, city councils, tourism agencies and Airbnb itself that want to have a complete and reliable assessment of its quality to their customers. It also can assist for benchmarking purposes, due to the fact the criteria found in our study are applicable to any company operating in tourism. One limitation of our study is that the empirical application uses a sample from a particular continent; consequently, results cannot be generalized across worldwide. However, a recommendation for further studies is to relate our findings to cross-country comparisons.

This comprehensive study aims to be an essential reference source for the use of text mining in the context of service quality assessment, building on the available literature in the field of sharing economy while providing further research opportunities in big data and tourism and services.

Acknowledgments: This article was written as part of a research project titled 'improvement of quality in collaborative consumption companies: model, scale and loyalty (CC-Qual)' (ref: rti 2018-096279-b-i00) financed by the ministry of science, innovation and universities of Spain within the aid program for R&D "Retos de Investigación" project.

References:

Assaf, A. G., & Josiassen, A. (2014). Frontier Analysis: A State-of-the-Art Review and Meta-Analysis. *Journal of Travel Research*, *55*(5), 612-627.

Baglama, J., Rachel, L., & Lewis, B. W. (2018). Irlba: Fast Truncated Singular Value Decomposition and Principal Components Analysis for Large Dense and Sparse Matrices. R package version 2.3.2. Retrieved from https://CRAN.R-project.org/package=irlba

Baralou, E., & Tsoukas, H. (2015). How is new organizational knowledge created in a virtual context? An ethnographic study. *Organization Studies*, *36*(5), 593-620.

Bardhi, F., & Eckhardt, G. M. (2012). Access-Based Consumption: The Case of Car Sharing. Journal of Consumer Research, *39*(4), 881-898.

Barnes, S. J., & Mattsson, J. (2016). Understanding current and future issues in collaborative consumption: A four-stage Delphi study. *Technological Forecasting and Social Change*, *104*, 200-211.

Batista e Silva, F., Marín Herrera, M. A., Rosina, K., Ribeiro Barranco, R., Freire, S., & Schiavina, M. (2018). Analyzing spatiotemporal patterns of tourism in Europe at high resolution with conventional and big data sources. *Tourism Management*, *68*, 101-115.

Belk, R. (2014). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research*, *67*(8), 1595-1600.

Benoit, S., Baker, T. L., Bolton, R. N., Gruber, T., & Kandampully, J. (2017). A triadic framework for collaborative consumption (CC): Motives, activities and resources & capabilities of actors. *Journal of Business Research*, *79*, 219-227.

Berbegal-Mirabent, J., Mas-Machuca, M., & Marimon, F. (2016). Antecedents of online purchasing behaviour in the tourism sector. *Industrial Management & Data Systems*, *116*(1), 87-102.

Calo, R., Rosenblat, A. (2017). The taking economy: Uber, information, and power. *Columbia Law Review*, *117,* 1623-1690.

Cheng, X., Fu, S., & de Vreede, G. J. (2018). A mixed method investigation of sharing economy driven car-hailing services: Online and offline perspectives. *International Journal of Information Management*, *41*, 57-64.

Cohen, B., & Kietzmann, J. (2014). Ride On! Mobility Business Models for the Sharing Economy. *Organization and Environmen*t, *27*(3), 279-296.

Cohen, M., & Sundararajan, A., (2015). Self-regulation and innovation in the peer-to-peer sharing economy. *University of Chicago Law Review Online*, *82*(1), 116-133.

Cusumano, M. A. (2014). How traditional firms must compete in the sharing economy. *Communications of the ACM*. https://doi.org/10.1145/2688487

Del Mar Alonso-Almeida, M., Bernardo, M., Llach, J., & Marimon, F. (2014). Building loyalty through functional and hedonic quality. *Industrial Management & Data Systems*, *114*(3), 387-404.

Dillon, I. S., & Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. *Machine Learning*, *42*, 143-175.

Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, *55*, 62-73.

Ertz, M., Durif, F., & Arcand, M. (2016). Collaborative consumption: conceptual snapshot at a buzzword. *Journal of Entrepreneurship Education*, *19*(2), 1-23.

Fang, B., Ye, Q., & Law, R. (2016). Effect of sharing economy on tourism industry employment. *Annals of Tourism Research*, *57*(January 2013), 264-267.

Farajallah, M., Hammond, R. G., & Pénard, T. (2016). *What Drives Pricing Behavior in Peer-to Peer Markets? Evidence from the Carsharing Platform BlaBlaCar.* Paper presented at the 65th Annual meeting of the French Economic Association.

Farmaki, A., Altinay, L., Botterill, D., & Hilke, S. (2015). Politics and sustainable tourism: The case of Cyprus. *Tourism Management*, *47*, 178-190.

Feinerer, I., & Hornik, K. (2018). Tm: Text Mining Package. R Package version 0.7-5. Retrieved from https://CRAN.R-project.org/package=tm

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, *25*(5), 1-54.

Getz, D., & Page, S. J. (2014). Progress and prospects for event tourism research. *Tourism Management*, *52*, 593-631.

Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, *18*(12), 1192-1217.

Habibi, M. R., Davidson, A., & Laroche, M. (2017). What managers should know about the sharing economy? *Business Horizons*, *60*(1), 113-121.

Hamari, J., Sjöklint, M., & Ukkonen, A. (2016). The sharing economy: Why people participate in collaborative consumption. *Journal of the Association for Information Science and Technology*, *67*(9), 2047-2059.

Hornik, K., Feinerer, I., Kober, M., & Buchta, C. (2012). Spherical k-means clustering. *Journal of Statistical Software*, *50*(10), 1-22.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD 2004), Aug 22-25. Seattle, Washington, USA.

Koopman, C., Mitchell, M., & Thierer, A, (2015). The sharing economy and consumer protection regulation: The case for policy change. *Journal of Business, Entrepreneurship & the Law*, *8*, 529.

Krippendorff, K. (2013). *Content Analysis* (3rd Ed). Thousand Oaks, CA: Sage Publications.

Landauer, T. K., Foltz, P. W. & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284.

Lewis, D. D., Yang, Y., Rose, T. G., & Li, F. (2004). RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, *5*, 361-397.

Liu, B., Hu, M., & Cheng, J. 2005. Opinion observer: Analyzing and comparing opinions on the web. *Proceedings of the 14th International World Wide Web conference* (WWW2005), May 10-14. Chiba, Japan.

Lonely Planet (2017). Las ciudades de Europa más visitadas. Retrieved from https://www.lonelyplanet.es/blog/las-ciudades-de-europa-mas-visitadas

Leighton, P. (2016). Professional self-employment, new power and the sharing economy: Some cautionary tales from Uber. *Journal of Management and Organization*, *22*(6), 859-874.

Liu, L., Cheung, C. M. K., & Lee, M. K. O. (2016). An empirical investigation of information sharing behavior on social commerce sites. *International Journal of Information Management*, *36*(5), 686-699.

Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, 30(12), 3514-3554.

Marimon, F., Llach, J., Alonso-Almeida, M., & Mas-Machuca, M. (2019). CC-Qual: A holistic scale to assess customer perceptions of service quality of collaborative consumption services. *International Journal of Information Management*, *49*, 130-141.

Matzler, K., Veider, V., & Kathan, W. (2014). Adapting to the Sharing Economy. *MIT Sloan Management Review*, *56*(2), 71.

McCandless, M., Sanford, M., & Fırat, A. (2013). Cldr: Language Identifier based on CLD library. R package version 1.0.1. Retrieved from https://www2.uaem.mx/r-mirror/web/packages/cldr/cldr.pdf

Möhlmann, M. (2015). Collaborative consumption: Determinants of satisfaction and the likelihood of using a sharing economy option again. *Journal of Consumer Behaviour*, *14*(3), 193-207.

Nunkoo, R. (2015). Tourism development and trust in local government. *Tourism Management*, *46*, 623-634.

Priporas, C. V., Stylos, N., Rahimi, R., & Vedanthachari, L. N. (2017). Unraveling the diverse nature of service quality in a sharing economy: A social exchange theory perspective of Airbnb accommodation. *International Journal of Contemporary Hospitality Management*, *29*(9), 2279-2301.

Puschmann, T., & Alt, R., (2016). Sharing economy. *Business & Information Systems Engineering*, *58*(1), 93-99.

Schaefers, T., Wittkowski, K., Benoit (née Moeller), S., & Ferraro, R. (2016). Contagious Effects of Customer Misbehavior in Access-Based Services. *Journal of Service Research*, *19*(1), 3-21. https://doi.org/10.1177/1094670515595047

Tussyadiah, I. P. (2016). Factors of satisfaction and intention to use peer-to-peer accommodation. *International Journal of Hospitality Management*, *55*, 70-80.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, *14*(3), 130-137.

R Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation, Vienna, Austria. Retrieved from https:// www.R-project.org.

Uzunca, B., Rigtering, J. P., & Ozcan, P. (2019). Sharing and shaping: A cross-country comparison of how sharing economy firms shape their institutional environment to gain legitimacy. *Academy of Management Discoveries*, *4*(3), 248-272.

Weiss, S. M., Indurkhya, N., Zhang, T., & Damerau, F. J. (2005). *Text Mining*. New York, NY: Springer Verlag.

Wild, F. (2015). lsa: Latent Semantic Analysis. R package version 0.73.1. Retrieved from https://CRAN.R-project.org/package=lsa

Zervas, G., Proserpio, D., & Byers, J. W. (2017). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research*, *54*(5), 687-705.

Burçin Güçlü
BES La Salle, Universitat Ramon Llull
Barcelona, Spain
burcin.guclu@salle.url.edu

David Roche
Universitat Internacional de Catalunya,
Barcelona, Spain
droche@uic.es

Frederic Marimon
Universitat Internacional de Catalunya,
Barcelona, Spain
fmarimon@uic.es

Appendix A. Histograms of sentiment scores for each city

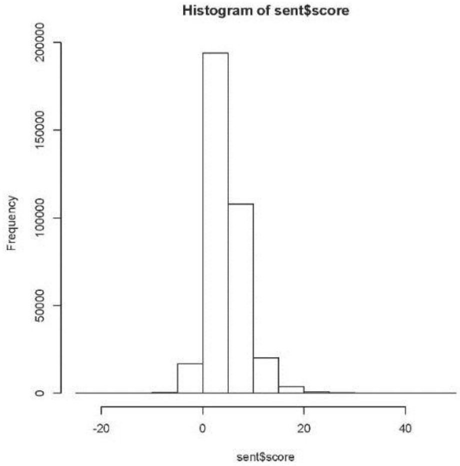

Figure A.1 Barcelona

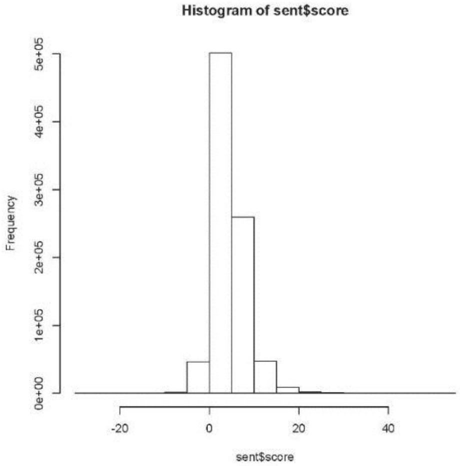

Figure A.2 Istanbul

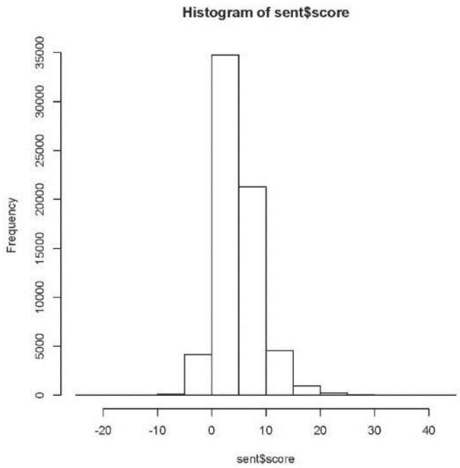

Figure A.3 London

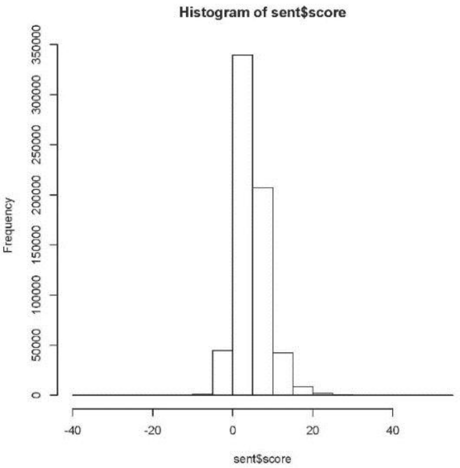

Figure A.4 Paris

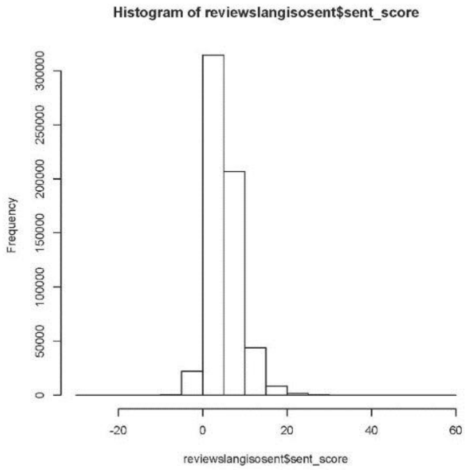

Figure A.5 Rome

Appendix B. List of terms retrieved for each city

Barcelona
[1] "also" "apartment" "barcelona" "can" "city" "clean" "close" "comfortable" "definitely" "easy" "everything" "flat"
[13] "friendly" "good" "great" "helpful" "host" "just" "located" "location" "lovely" "metro" "nice" "perfect"
[25] "place" "really" "recommend" "restaurants" "room" "stay" "time" "walk" "well"

Istanbul

[1] "also" "amazing" "apartment" "area" "can" "city" "clean" "close" "comfortable" "definitely" "easy" "even"
[13] "everything" "flat" "friendly" "get" "good" "great" "helpful" "host" "istanbul" "just" "like" "located"
[25] "location" "nice" "one" "perfect" "place" "really" "recommend" "restaurants" "room" "stay" "street" "taksim"
[37] "time" "view" "walk" "well" "will"

London

[1] "also" "apartment" "area" "clean" "close" "comfortable" "definitely" "easy" "everything" "flat" "friendly" "good"
[13] "great" "helpful" "home" "host" "house" "just" "location" "london" "lovely" "nice" "perfect" "place"
[25] "quiet" "really" "recommend" "room" "station" "stay" "time" "tube" "walk" "well"

Paris

[1] "also" "apartment" "area" "clean" "close" "comfortable" "definitely" "easy" "everything" "flat" "get" "good"
[13] "great" "helpful" "host" "just" "located" "location" "lovely" "metro" "nice" "paris" "perfect" "place"
[25] "quiet" "really" "recommend" "restaurants" "room" "small" "stay" "time" "walk" "well" "wonderful"

Rome
[1] "also" "apartment" "area" "around" "can" "city" "clean" "close" "comfortable" "definitely" "easy" "even"
[13] "everything" "get" "good" "great" "helpful" "highly" "host" "just" "located" "location" "lovely" "metro"
[25] "nice" "perfect" "place" "really" "recommend" "restaurants" "rome" "room" "station" "stay" "time" "walk"
[37] "walking" "well" "wonderful"

APPENDIX C. Principle – Component Analysis using Terms on City Basis

Barcelona

Terms	B1 (F2)	B2 (F5)	B3	B4	B5 (F4)	B6
clean	0.17	0.14	0.05	0.07	-0.28	0.01
comfortable	0.16	0.01	0.06	-0.02	-0.27	0.11
flat	0.10	0.21	0.04	0.10	0.14	0.08
friendly	0.07	0.11	0.08	0.17	-0.45	-0.02
great	0.14	-0.38	-0.07	0.12	-0.06	-0.07
helpful	0.10	-0.01	0.11	0.08	-0.44	-0.03
host	0.06	-0.03	0.09	0.14	-0.20	0.13
nice	0.11	0.37	-0.10	0.12	0.06	-0.13
restaurants	0.20	-0.14	-0.18	-0.19	-0.03	-0.10
time	0.19	0.04	0.08	0.12	0.16	0.22

Istanbul

Terms	I1	I2 (F2)	I3 (F5)	I4	I5 (F4)	I6 (F3)
amazing	-0.09	-0.33	-0.09	0.05	0.18	-0.19
clean	-0.15	0.13	-0.02	-0.08	-0.32	0.08
close	-0.12	0.15	0.11	-0.05	-0.15	-0.28
everything	-0.13	-0.07	-0.05	0.01	-0.03	-0.22
flat	-0.11	0.09	-0.11	-0.05	0.06	-0.43
friendly	-0.06	0.06	-0.09	-0.02	-0.31	0.13
helpful	-0.08	0.00	0.02	-0.11	-0.37	0.04
like	-0.16	0.03	-0.20	0.16	0.06	0.08
located	-0.12	0.19	-0.06	-0.50	0.18	0.00
place	-0.15	-0.11	-0.28	0.01	-0.14	-0.03
taksim	-0.15	0.20	0.16	-0.03	-0.06	-0.07

London

Terms	L1 (F1)	L2 (F5)	L3	L4	L5 (F4)	L6 (F3)
everything	0.15	0.00	-0.13	-0.08	-0.14	-0.26
flat	0.17	0.18	-0.04	0.10	0.04	-0.25
host	0.03	-0.04	-0.13	-0.08	0.36	0.07
house	0.13	-0.34	0.14	0.14	0.08	-0.07
London	0.29	0.04	-0.13	0.02	-0.11	-0.11
lovely	0.10	-0.18	-0.16	0.37	0.10	-0.10
perfect	0.11	0.09	-0.18	0.05	0.01	-0.32
time	0.18	-0.09	-0.07	-0.11	-0.01	-0.25

Paris

Terms	P1 (F1)	P2	P3	P4	P5 (F4)	P6 (F5)
clean	-0.15	-0.09	0.09	0.10	-0.04	0.36
comfortable	-0.17	-0.02	0.08	-0.03	-0.19	0.21
easy	-0.16	0.05	-0.12	-0.14	0.12	-0.20
flat	-0.09	-0.09	0.07	0.15	0.15	-0.47
host	-0.01	0.07	0.08	0.06	-0.21	-0.12
lovely	-0.10	0.10	0.14	-0.07	-0.06	-0.23
Paris	-0.28	0.14	0.17	0.11	-0.04	-0.15
perfect	-0.12	0.26	-0.02	0.12	-0.09	-0.05
walk	-0.23	0.00	-0.17	-0.15	-0.04	-0.15

Rome

Terms	R1 (F1)	R2 (F6)	R3 (F5)	R4 (F3)	R5	R6
area	0.16	-0.02	0.07	0.02	0.21	0.10
around	0.20	-0.05	0.15	0.02	-0.01	-0.12
clean	0.14	-0.09	-0.21	-0.04	-0.04	0.16
close	0.12	-0.14	-0.01	0.25	-0.02	-0.02
easy	0.14	-0.02	0.11	0.16	-0.03	-0.22
great	0.12	0.19	0.26	0.08	-0.18	0.04
perfect	0.10	0.21	0.13	-0.12	-0.10	0.05
Rome	0.27	0.16	-0.01	-0.17	0.02	-0.18
time	0.17	0.02	0.04	-0.26	0.00	-0.06
wonderful	0.09	0.20	-0.01	-0.06	0.11	-0.19

Appendix D. Principle – Component Analysis using Terms on Aggregate Data

	PC1	PC2	PC3	PC4	A5	A6
easy	-0.1508	-0.0290	0.1030	0.0223	-0.0230	-0.0671
flat	-0.1114	0.0671	-0.1318	-0.0526	-0.1686	-0.0503
metro	-0.2332	0.1163	0.2777	0.1881	0.2540	-0.0977
nice	-0.1214	0.3184	-0.1525	-0.1072	0.1282	-0.0658
clean	-0.1762	0.0837	-0.2052	-0.0610	0.0673	0.1121
great	-0.1429	-0.2710	0.2171	-0.2260	-0.0362	0.0546
location	-0.0995	-0.2528	0.3437	-0.3828	0.1067	0.2025
recommend	-0.1987	-0.4131	-0.3205	0.1775	0.2947	-0.0279
apartment	-0.2739	-0.1150	0.2339	0.1870	-0.1319	0.2532
good	-0.1068	0.2608	0.0324	-0.1814	0.1511	0.1706
well	-0.2006	0.1136	-0.0790	0.2469	-0.2392	0.1179
stay	-0.2267	-0.1347	-0.1159	-0.1319	-0.2415	-0.2488
time	-0.1695	0.0260	-0.0503	-0.1653	-0.1319	-0.0042
also	-0.2720	0.0956	0.0066	-0.0980	0.0396	0.0620
definitely	-0.1472	-0.2124	-0.1298	-0.0995	-0.1698	-0.2825
helpful	-0.1026	-0.0421	-0.1366	-0.0582	0.0123	0.2742
host	-0.0383	-0.0946	-0.0765	-0.0770	0.0205	0.3035
quiet	-0.1495	0.0800	-0.0255	0.0792	-0.1314	0.0833
friendly	-0.0645	0.0514	-0.2235	-0.1226	0.0925	0.3231
perfect	-0.1134	-0.1897	0.0999	-0.1446	-0.1407	-0.1034
room	-0.1182	0.2150	-0.2450	-0.2682	0.1453	0.1417
comfortable	-0.1740	0.0094	-0.1275	-0.0262	-0.1180	0.1559
area	-0.1916	0.0710	0.0013	0.0852	-0.1903	0.0973
everything	-0.1639	-0.0728	0.0088	-0.0618	-0.1823	-0.1379
highly	-0.1159	-0.4069	-0.2362	0.2599	0.3569	0.1503
place	-0.1559	-0.0579	-0.1954	-0.1019	0.1721	-0.4612
really	-0.1657	0.1551	-0.1762	-0.2057	0.0049	-0.0924
restaurants	-0.2311	-0.0312	0.2527	0.1225	-0.0547	0.0715
located	-0.1373	0.1505	-0.1283	0.4591	-0.2298	0.0196
station	-0.2050	0.2161	0.1064	0.0847	0.3311	-0.1429
close	-0.1723	0.0951	0.1872	0.1343	0.2125	-0.1821
just	-0.2189	0.0413	0.0615	-0.1288	-0.0390	-0.0597
lovely	-0.0900	-0.0874	-0.1134	-0.0236	-0.2263	0.0145
walk	-0.2610	0.0583	0.1922	0.0197	0.1067	-0.0124