

## Repositório ISCTE-IUL

---

Deposited in *Repositório ISCTE-IUL*:

2019-04-12

Deposited version:

Post-print

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Brochado, A., Rita, P. & Moro, S. (2019). Discovering the patterns of online reviews of hostels in Beijing and Lisbon. *Journal of China Tourism Research*. 15 (2), 172-191

Further information on publisher's website:

10.1080/19388160.2018.1543065

Publisher's copyright statement:

This is the peer reviewed version of the following article: Brochado, A., Rita, P. & Moro, S. (2019). Discovering the patterns of online reviews of hostels in Beijing and Lisbon. *Journal of China Tourism Research*. 15 (2), 172-191, which has been published in final form at <https://dx.doi.org/10.1080/19388160.2018.1543065>. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

---

### Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

---

## **Discovering patterns in online reviews of Beijing and Lisbon hostels**

This study employed a data mining approach to model the quantitative scores given to hostels located in Beijing, China, and Lisbon, Portugal, in guests' online reviews posted on Booking.com. A neural network was built using a total of nine input features (e.g., age, most and least favorite aspects, travel and traveler types, nationality, hostel, and month and weekday of review) to model the score distributions. Each feature's contribution to the scores was then extracted through data-based sensitivity analysis. The most favorite aspect and continent of origin were the two most significant features for hostels in both cities. Lisbon guests were also highly influenced by the hostel itself and traveler type as compared with Beijing travelers. Notably, facilities are the most favorite aspect valued by guests staying in Lisbon, while those that stay in Beijing hostels give more importance to value for money. These findings denote different guest behaviors are associated with each city's particular offer.

Keywords: service quality; hostels; online reviews; data mining; Beijing; Lisbon

### **Introduction**

Hostels are becoming a more popular alternative to traditional hotels for many young travelers aged 18 to 35 years old (Phocuswright, 2016). More specifically, Millennials are driving an increase in hostel-based travel due to low costs, convenient locations, good values, experience-driven accommodations, and opportunities to meet other travelers. According to Phocuswright's (2016) report, this segment, which represents more than 70% of hostel-goers, is particularly impassioned about and determined to take more trips across diverse markets than any other traveler type. Millennials are also more likely to go abroad.

Studies of service quality in hostels have been conducted in Scotland (Nash, Thyne, & Davies, 2006), Canada (Hecht & Martin, 2006), Australia (Chitty, Ward, & Chua, 2007), Malaysia (Musa & Thirumoorthi, 2011), and Portugal (Brochado & Gameiro, 2013; Brochado & Rita, 2018; Brochado, Rita, & Gameiro, 2015). The

present study sought to analyze overall service quality ratings by focusing on hostels located in the capitals of two different countries: Portugal and China. The primary objective was to identify the main variables that explain overall service quality ratings shared online by guests who stayed in hostels located in Lisbon and Beijing.

In a globalized world, travelers can easily shift between destinations across the globe and come in contact with distinct cultures and hostel offers. Comparing two major cities separated by more than 9,000 kilometers is of interest because of one similarity and one divergence between Lisbon and Beijing. First, both are top, renowned cultural city destinations (Hsu & Huang, 2016; Magalhães, 2017). Second, Lisbon has held a special position in Western history even prior to the Middle Ages, while Beijing is considered the cradle of Confucian cultures (Calvo, 2015; Rosker, 2017).

In addition, Lisbon is rather small compared to Beijing, although Lisbon is also considered an alpha-level global city (Globalization and World Cities Research Network, 2016), with lower pollution and traffic levels than Beijing. Lisbon has also been considered in cross-country comparisons in research on the hospitality industry to test convergence regarding customers' experiences (e.g., Brochado, Troilo, & Shah 2017).

Previous studies have also examined service quality, but they have conducted research based mainly on interviews and surveys. An exception to this is Musa and Thirumoorthi's (2011) study, which performed content analyses of text reviews shared by hostel guests. When booking hostels, guests often rely on feedback and ratings provided on hostel websites by other guests (Brochado & Gameiro, 2013; Brochado & Rita, 2018). Thus, the present study opted to gather data from a hostel booking website in order to study the main drivers of service quality ratings. Unlike Musa and Thirumoorthi (2011), the current research model focused on quantitative ratings.

Statistical methods used in previous related studies have included Hecht and Martin (2006) and Nash et al.'s (2006) descriptive statistics and Chitty et al.'s (2007) structural equation modeling. In addition, Brochado and Rita (2018), Brochado et al. (2015), and Musa and Thirumoorthi (2011) employed factor analysis, regression analysis, confirmatory factor analysis, and latent regression analysis. In contrast, the present study applied data mining to analyze the influence of nine relevant features on the quantitative scores given by hostel guests in online reviews (Silva, Moro, Rita, & Cortez, 2018).

The textual content of these reviews has been extensively analyzed using text mining techniques, as reported in the literature on hospitality and tourism (e.g., Calheiros, Moro, & Rita, 2017; Guerreiro & Moro, 2017). Text mining has even been employed to develop decision-support systems for decision making (Nave, Rita & Guerreiro, 2018). However, less attention has been paid to the quantitative features that can also be extracted from online reviews. Recent studies have provided evidence that these features can be effectively used to extract insights into guests' perspectives (e.g., Moro, Rita, & Coelho, 2017).

The rest of this paper is structured as follows. The literature review describes the conceptualization of hostels and discusses the role of Web reviews in hostel bookings, the use of data mining to study online reviews, and the concept of service quality in the hostel sector. Next, the methodology section presents the research context, data collection procedure, and data analysis based on data mining techniques. The results are then analyzed and discussed. The paper ends with theoretical and managerial implications, limitations, and avenues for future research.

## **Theoretical background**

### ***Development of hostels worldwide***

Hostels are a hybrid product that combines accommodation services with an informal, friendly atmosphere (Brochado et al., 2015). Hostelling International—a hostel federation founded in 1932 that has over 4,000 affiliated hostels around the world—defines hostels as a “good quality budget accommodation that offers a comfortable night’s sleep in [a] friendly atmosphere at an affordable price” (Hostelling International, 2017). Hostels provide types of accommodations not offered by hotels, such as shared rooms (i.e., either single gender or mixed) or private rooms. Whereas hotel customers can book a single or double room, they may book an individual bed in a shared room in hostels.

In addition, hostels generally offer more and better opportunities for guests to socialize (Rita, Brochado, & Marques, 2016) and meet new people from different cultures, based on common areas and dormitories. Some hostels are also moving upscale and offering extras. These may include en suite bathrooms, safe storage facilities, bar offers, restaurant and/or dining areas, private rooms and washrooms, “funky” communal areas, and quirky design feature (Brochado et al., 2015).

### ***Online reviews in hostel booking***

Currently, a large number of travelers use the Internet to seek accommodation information (Litvin, Goldsmith, & Pan, 2008). Hostel booking websites or the hostels’ own websites are the most popular sources of information to help travelers choose hostels (Brochado & Gameiro, 2013). Informal and word-of-mouth communication is of utmost importance to hostel guests (Brochado & Rita, 2018; Moshin & Ryan, 2003;

Nash et al., 2006), who consider customers' reviews and ratings reliable sources of information (Brochado & Gameiro, 2013).

If reviews are well reasoned, logical, and persuasive, they can positively influence readers' likelihood of purchase (Park, Lee, & Han, 2007). Travelers read reviews and check hostels' ratings before booking a room. These tourists usually tend to look for hostels with the largest number of reviews, which suggests that more people have stayed there and that evaluations are thus more precise (Brochado & Gameiro, 2013). Hostel guests trust the feedback provided by other guests and write their own reviews, usually giving ratings and reviews of the hostels in which they stay regardless of whether their opinion is positive or negative (Brochado & Rita, 2018).

### ***Mining online reviews in hospitality***

According to Wong, Chaisorn, and Kankanhalli (2014), "we are living in a world of [big] [d]ata." While a few decades ago, most data were gathered and harnessed by corporations, currently the opposite is true. The advent of social media has led to a profusion of user-generated data, which now constitutes one of the largest data sources contributing to the growth of big data (Amado, Cortez, Rita, & Moro, 2018).

In addition, the more recent rise of the Internet of Things has contributed to generating even larger volumes of data (Canito, Ramos, Moro, & Rita, 2018), making extracting useful knowledge from big data a challenging task. By combining artificial intelligence techniques with traditional statistical methods, data mining has become a cornerstone of research seeking to find hidden patterns of knowledge in raw data (Moro, Cortez, & Rita, 2015). Text mining is a variant of data mining that is well suited to dealing with unstructured data such as text comments.

As Moro and Rita (2016) point out, forecasting tourism demand is a key challenge that every organization in the hospitality and tourism industries needs to

embrace. However, the cited authors suggest that most methods used are still based on traditional time series. These techniques are usually outperformed by the most advanced data mining techniques such as neural networks and support vector machines.

In tourism online reviews, two types of data typically occur: text comments and various quantitative features, depending on the platform. For example, TripAdvisor—one of the most well-known hospitality and tourism platforms—displays information about guests (e.g., number of reviews posted). This website also provides specific review information, such as the numeric ratings selected and number of “helpful” votes (Lee, Law, & Murphy, 2011). Booking.com has similarly evolved from an online purchasing tourism platform to a site offering sophisticated customer feedback services through online reviews in which travelers share their experiences (Moro, Rita, & Oliveira, 2018).

While various studies in the literature have focused on unveiling hidden knowledge in text reviews (e.g., Sparks & Bradley, 2014), few researchers have considering quantitative features when modeling the factors impacting tourists’ ratings (e.g., Ye, Li, Wang, & Law, 2014). Moreover, no previous study has specifically applied an advanced data mining approach in order to build a single coherent research model, thereby contributing to a fuller understand of what drives travelers to rate hostels highly or not. The present study is thus the first attempt to adopt this approach.

### ***Service quality in hostels***

Given the specificities of the hostel business, studies have been conducted to assess the service quality provided by hostels to their guests. Nash et al. (2006) concluded that backpackers consider the cleanliness of rooms, value for money, location, and self-catering facilities to be the most important elements of hostels’ service quality. Hecht and Martin (2006) found that the five most important service quality aspects are

cleanliness, location, personal service, security, and hostel services (e.g., Internet and laundry facilities). Chitty et al. (2007) established that brand image and the functional dimension (i.e., staff behavior) have a positive impact on backpackers' satisfaction and that brand loyalty is also directly influenced by brand image.

Musa and Thirumoorthi's (2011) research, in turn, revealed that the most important tangible elements in hostels' service quality are equipment, atmosphere, cleanliness, facilities, central location, and a friendly, welcoming, and home-like atmosphere. Staff excellence elements include courtesy, willingness to help, relevant knowledge, and individualized attention. Brochado and Gameiro's (2013) study further confirmed that the most important item affecting overall satisfaction during stays in hostels is the quality of staff, followed by location, facilities, Internet facilities, atmosphere, cleanliness, and opportunities to meet other travelers. Bar service, security issues, and prices are considered to be less important.

Brochado et al.'s (2015) findings indicate that hostel service quality includes six dimensions: staff, cleanliness, security, facilities, social atmosphere, location, and city connection. The cited authors also identified social atmosphere as a core service dimension crucial to creating hostel guests' overall perception of quality, noting that this dimension is specific to the hostel sector. Brochado and Rita (2018) also found that four core dimensions of service quality in hostels—quality of staff, social atmosphere, hostel tangibles, and city connection—are significant aspects explaining levels of satisfaction, recommendations, and revisiting intentions.

In addition, Santos's (2016) results show that the most important dimensions of service quality that explain premium prices in hostels worldwide are cleanliness, location, and facilities. Cró and Martins (2017) further concluded that hostel guests are willing to pay a premium price for hostels that offer the highest levels of perceived



security in countries with the highest crime indices. In parallel with academic research, Hostelling International (2017b) recognizes the hostel market's need to develop a reliable battery of customer service "assured standards." These should include, among others, a warm welcome, comfort (e.g., a good night's sleep and adequate washing and/or shower facilities), cleanliness, security for backpackers and their possessions, and privacy in showers, washing areas, and toilets. These customer service standards have been assessed by using "mystery shopper" evaluations (Hostelling International, 2017b).

## **Materials and methods**

### ***Data***

The present research focused on highly rated hostels in two culturally contrasting and geographically remote cities. Beijing is one of the largest megacities in the world and the capital of China, which is thriving in terms of economic growth (He, Chen, Mao, & Zhou, 2016). Lisbon is a European city and the capital of Portugal, which World Travel Awards (2018) recently named "Europe's Leading City Destination 2018."

Each city's six top hostels—according to Hostelworld.com's ratings—were selected (i.e., 12 in total), and a hundred reviews were gathered for each hostel from Booking.com. All the hostels selected have an average score of above 8.5 and a good location, including being near public transportation, in the heart of the city, or in an area extremely close or within short walking distance to the main attractions and hot spots.

Booking.com is an online platform that allows travelers to provide quantitative feedback (i.e., from 1 to 10) on 7 individual features: value for money, location, security, atmosphere, facilities, cleanliness, and staff. In the present study, both the favorite and least favorite aspects of hostel experiences were included in the data. Table

1 shows the full list of features collected. Despite the rigorous criteria defined for data collection, five rows were found to have missing values. These records were discarded since they accounted for only 0.4% of the total data, leaving a total of 1,195 reviews for analysis.

**Table 1 - Features collected.**

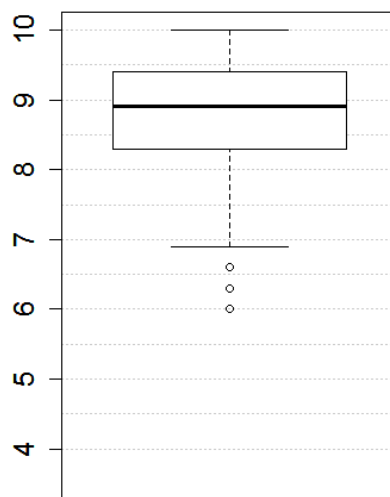
<b>Feature</b>	<b>Category</b>	<b>Possible values</b>
City	Hostel	Beijing; Lisbon
Name		Name of the hostel
Nationality	Traveler	70 different countries
Age		{18-24; 25-30; 31-40; 41+}
Traveling aim		{Business; Leisure; Undefined}
Travel type		{All female group; All male group; Alone; Couple; Mixed group; Undefined}
Traveler type		{Avid traveler; Globetrotter; Novice nomad}
Date	Review	Date of the review (from Sep/2013 to Oct/2016)
Rating (quantitative)		[1; 10]
Review (qualitative)		{Bad; Good; Very good; Fabulous; Superb}
Text Review		Comments written by users
Favorite aspect of the experience		One from the following: value for money, location, security, atmosphere, facilities cleanliness, and staff
Least favorite aspect of the experience		

The selected features were classified into three main categories: hostel features, review characteristics, and traveler profiles (see the “Category” column in Table 1 above). The data regarding the 12 hostels included their name and city. The data on each review listed the date the review was posted, the rating given (i.e., quantitative score), the review’s categorical value computed by Booking.com based on the rating assigned, and the text of the review (i.e., qualitative information). In addition, the dataset encompassed both the favorite and least favorite aspects of each guest’s experience from the following seven features: value for money, location, security, atmosphere, facilities, cleanliness, and staff. Finally, each reviewer’s profile was

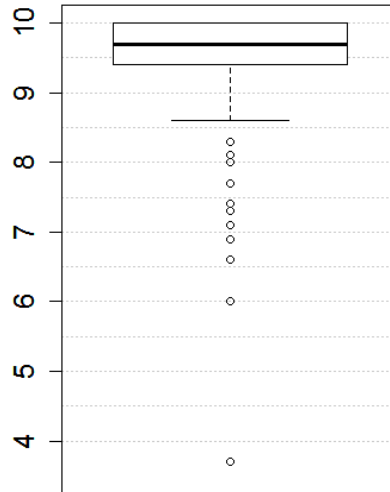
completed in terms of nationality, age, travel type, reasons for traveling, and kind of traveler.

### ***Mining procedure***

Prior to modeling, the dataset needed to be subjected to a typical data preparation process as in any data mining research (Han, Pei, & Kamber, 2011). Given that the goal was to construct a model based on quantitative scores, both the qualitative review scores and textual comments were removed from the dataset since usually both types of scores (i.e., qualitative and quantitative) are related (Chung & Tseng, 2012). In addition, the average score for the entire dataset was extremely high, above 8.5, as stated in the previous subsection. Figure 1 shows that the first quartile for Beijing starts at a score around 6.9, with the second and third quartiles encompassing scores from 8.3 to 9.4. The variations are even smaller for Lisbon (see Figure 2), with the first quartile starting at 8.6. As a result, the model had to be able to deal with these small variations in the ratings.



**Figure 1** – Boxplot for the ratings in Beijing.



**Figure 2** - Boxplot for the ratings in Lisbon.

The date was used to compute the aggregate day of the week and month of the reviews, as the dates themselves were meaningless in terms of patterns (Moro et al., 2017). The countries of origin were converted into continents as too many different countries appeared in the dataset. Tourists from 34 European countries were guests at the hostels and posted online reviews related to their experiences. Asia accounted for 14 countries, with Latin America contributing 9. Only 5 African countries were generating markets, followed by North America and Oceania, with 2 countries from each continent. Finally, the reason for traveling was discarded because only one guest wrote a review of a business trip and the remaining travelers were on leisure trips.

For modeling purposes, a total of nine features (see Table 2) were included in the input: the hostels' name, four traveler-related features, and four review-related characteristics. Since the model's output was numerical (i.e., quantitative scores), the data collected were subjected to regression analysis (Moro et al., 2017).

**Table 2 - Features selected.**

<b>Feature</b>	<b>Category</b>	<b>Possible values</b>
Hostel	Hostel	Name of the hostel
Continents	Traveler	6 continents related to traveler's country of origin (Africa, Asia, Europe, North America, Oceania, South America)
Age		{ 18-24; 25-30; 31-40; 41+ }
Travel type		{ All female group; All male group; Alone; Couple; Mixed group; Undefined }
Traveler type		{ Avid traveler; Globetrotter; Novice nomad }
Weekday review	Review	Weekday (Sunday, Monday, ... , Saturday)
Month review		Month (January, February, ..., December)
Favorite aspect of the experience		One from the following: value for money, location, security, atmosphere, facilities cleanliness, and staff
Least favorite aspect of the experience		

Given that Beijing and Lisbon are intrinsically divergent because of the geographical and cultural nature of both cities, two models were built—one for each city. The two final data subsets for each city included 598 reviews for Beijing and 597 for Lisbon. We expected that each model would shed some light on the features influencing most ratings granted. To gain a broader perspective, the results from both models could be compared to highlight the main differences of how guests assess hostels in the two cities.

In addition, the modeling procedure included building a multilayer perceptron, which is a type of artificial neural network that attempts to mimic the human brain by building a network of neurons. The multilayer perceptron developed was configured with one hidden layer of  $H$  hidden nodes (i.e., neurons) and the output node. The number of hidden nodes  $H$  is a hyperparameter that sets the learning model's complexity. Thus, a network with  $H = 0$  is equivalent to a simple logistic regression,

while setting  $H$  to a large number enables the network to apprehend better the complex, inherent, and nonlinear relationships between the input features.

In this study, the state of node  $i$  for input  $x_k$  was computed by using Formula (1):

$$f(w_{i,0} + \sum_{j \in P_i} w_{i,j} \times s_j) \quad (1)$$

in which  $f$  is the sigmoid function,  $w_{i,j}$  is the weight of the connections between nodes  $j$  and  $i$ , and  $P_i$  is the set of nodes reaching node  $i$ . In addition, given that the final output is dependent on the choice of the initial weights, an ensemble of different trained networks was adopted, with the output coming from the average of individual predictions. The multilayer perceptron was then utilized to compute an outcome based on patterns hidden in the data used to create the network (Russell, Norvig, Canny, Malik, & Edwards, 2003). This approach has been successfully adopted in various studies that have compared its results favorably with that of other techniques (e.g., Moro, Cortez, & Rita, 2014).

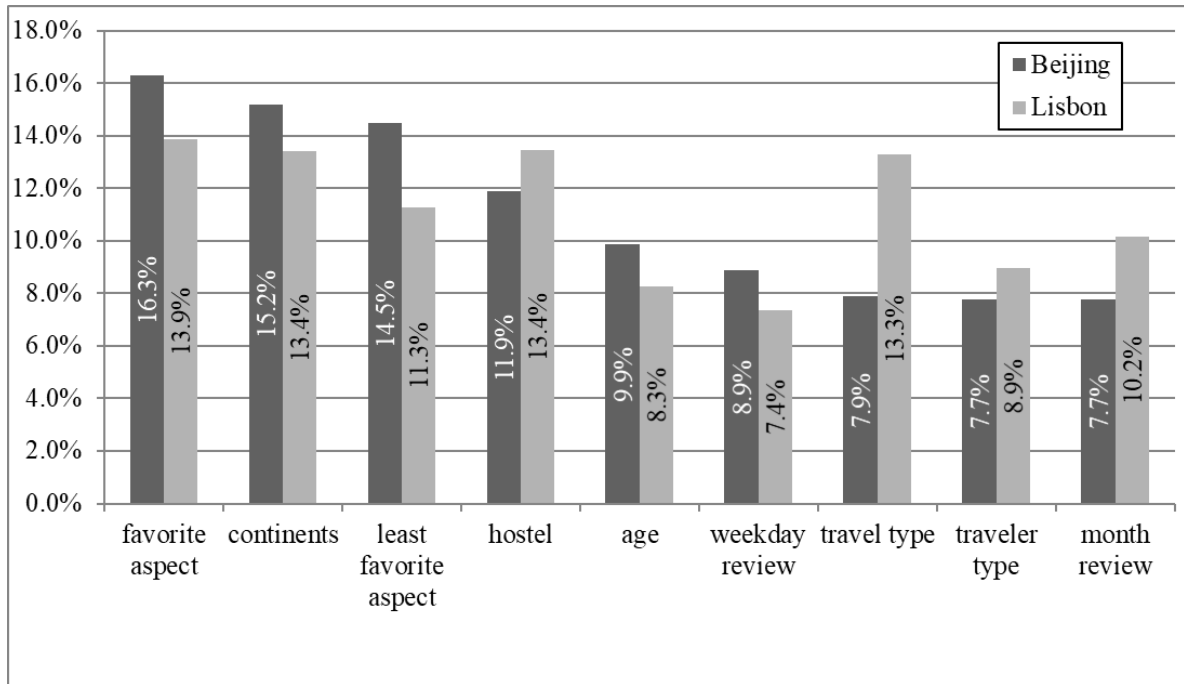
The present model's goal was not to predict future hostel ratings but instead to provide explanations based on insightful knowledge about the features that most strongly influence the ratings for this set of top-rated hostels. Thus, the models' accuracy was first assessed by measuring both the mean absolute error (MAE) and the mean absolute percentage error (MAPE). The former is the real difference between the real score and the outcome derived from the model for the same input values, while the MAPE is the relationship of the MAE with the real score, given as a percentage (Moro et al., 2017). Hence, the lowest possible values are considered more desirable for both metrics. The Beijing model achieved a MAE of 0.31 and a MAPE of 3.6%, while the Lisbon model resulted in a MAE of 0.16 and a MAPE of 1.8%. The values show that both models fit the data used, producing reasonably low errors.

Several techniques can be applied to extract knowledge from black box models, such as rules extraction and sensitivity analysis (Moro et al., 2014). In the present study, the latter option was chosen. More specifically, data-based sensitivity analysis (DSA) was conducted to determine the significance of each feature's contribution to the models developed. DSA uses a sample from the training dataset to assess how the outcomes are affected by varying several of the inputs' features simultaneously (Cortez & Embrechts, 2013). Thus, DSA was performed for both the Beijing and Lisbon models, and graphic representations were generated to facilitate comparisons of the results.

All the procedures described above were executed using R statistical tools since these are available from an open source platform with a large community of enthusiasts, offering a vast number of packages for data analysis (see <https://cran.r-project.org/>). More specifically, the rminer package was selected given that it offers a set of simple, coherent functions for data mining, including DSA (Cortez, 2010).

## **Results and discussion**

The importance of each individual features to the models of Beijing and Lisbon hostels' satisfaction ratings is shown in Figure 3. In general, the findings are consistent with the globalization effect resulting from communication and technology improvements, which has had a standardization effect on managerial education (Stromquist & Monkman, 2014). This can be observed in the large number of significant features emerging for both Beijing and Lisbon hostels. Nevertheless, some differences could reflect contrasting cultural patterns among Chinese and Portuguese hostel managers.



**Figure 3** - Features' relevance (in %) for Beijing and Lisbon.

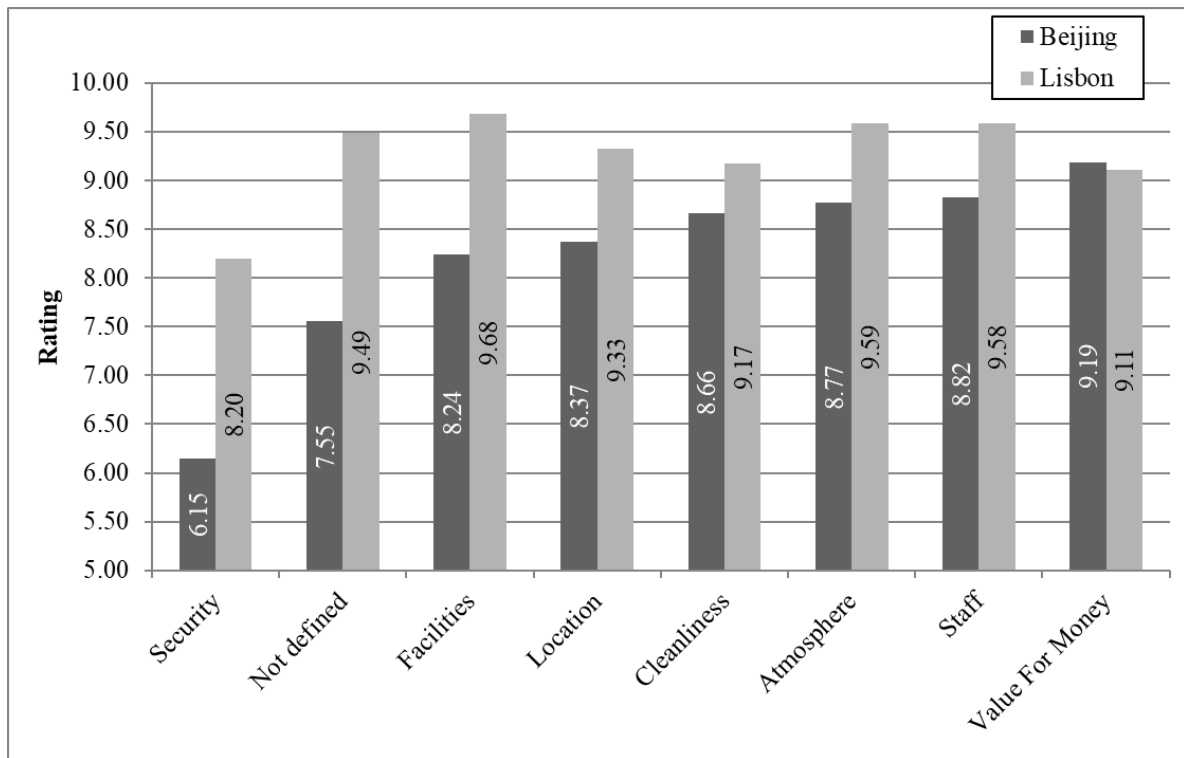
The most important features that explain the overall ratings for both Beijing and Lisbon hostels are favorite aspect of experiences (16.3% and 13.9%, respectively) and continent of nationality (15.2% and 13.4%). However, in Beijing, the third most significant feature is the least favorite aspect (14.5%), while, in Lisbon, hostel (13.4%) is a close third, which is immediately followed by travel group (13.3%). Interestingly the latter feature came only seventh for Beijing (7.9%) since hostel (11.9%), age (9.9%), and day of the week of review (8.9%) came ahead in fourth, fifth, and sixth place, respectively. Another important insight gained is that both traveler type (8.9% for Lisbon vs. 7.7% for Beijing) and month of review (10.2% vs. 7.7%) are more significant for Lisbon hostels than they are for Beijing hostels' ratings.

These results highlight that the overall ratings are explained by service quality dimensions (e.g., staff), hostel characteristics (i.e., location and name of the hostel), guest profiles (i.e., age, traveler type, and origin), and occasions (i.e., month and day of the review). Thus, the findings support the conclusion that hostel guests are a



heterogeneous market, leaving room for segmentation of service quality ratings based on demographics (Brochado & Gameiro, 2013; Brochado & Rita, 2018).

Since favorite aspect of experiences was found to be the most important feature for both Beijing and Lisbon hostels, specific aspects were analyzed in terms of their influence on ratings (see Figure 4). The results show that, for Beijing, value for money is the number one aspect (9.19), followed by staff (8.82), atmosphere (8.77), and cleanliness (8.66). In contrast, for Lisbon, the most influential aspects are facilities (9.68), atmosphere (9.59), and staff (9.58).



**Figure 4** - Influence of favorite aspect on rating.

The results also highlight similarities between hostels located in Lisbon and Beijing in terms of the most important attributes explaining service quality—staff and atmosphere—and differences in the relative importance of value for money, facilities, and cleanliness. Regarding value for money, the results reveal that this is the only attribute that is more important for Beijing hostels than for Lisbon hostels. A further

interesting result is that security is more important for hostels located in Lisbon than in Beijing.

The findings for the staff factor are in accordance with previous studies (Brochado & Gameiro, 2013; Brochado et al., 2015; Musa & Thirumoorthi, 2011). Human resource management has proved to be an important component in high quality service for hostel guests (Musa & Thirumoorthi, 2011). This may even be the most important dimension that explains guests' likelihood to return to the same hostel and recommend it (Brochado et al., 2015).

Brochado et al.'s (2015) study revealed that social atmosphere is a core service dimension needed to develop hostel guests' perceptions of overall quality, which differentiates hostels from other types of accommodation. Therefore, the importance of this variable for hostels located in both Portugal and China is not surprising. Guests may consider hostels a lifestyle choice, as they target young guests who seek to explore other cultures, expand their knowledge, and meet other travelers with common interests (Hecht & Martin, 2006).

In relation to the importance of value for money, Brochado and Gameiro (2013) found that, for hostels located in Lisbon, price is one of the least important correlates of overall service quality ratings. Nonetheless, males, younger clients, and guests from North America tend to place more importance on price issues. Brochado et al. (2015) and Wu and Ko (2013) also concluded that price is not a relevant determinant of service quality, but Nash et al.'s (2006) research showed value for money is one of the most important drivers of hostel service quality in Scotland hostels. Although hostels are known for offering budget accommodations, hostel guests in China expect to receive an adequate level of service value for money, as hostels are moving upscale and offering different types of services to guests (Brochado & Rita, 2018).

In addition, cleanliness proved to be an important driver of service quality in hostels located in Scotland (Nash et al., 2006), Canada (Hetch & Martin, 2006), Malaysia (Musa & Thirumoorthi, 2011), and Portugal (Brochado et al., 2015). Brochado et al. (2015) confirmed that the dimension of cleanliness is one of the more important aspects of service quality, comprising the cleaning of all hostel areas, such as rooms, dorms, bathrooms, kitchens, and social areas.

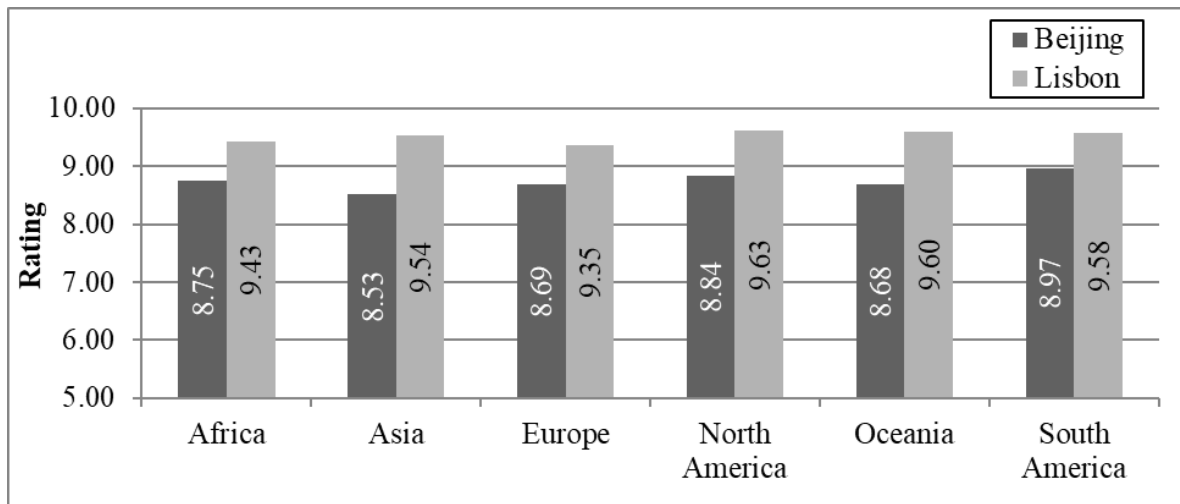
Facilities were further identified as a dimension of service quality in Lisbon hostels by Brochado et al. (2015), including a comfortable ambiance, a well-equipped kitchen, and an appealing decorative design of the hostels overall and of their rooms. Musa and Thirumoorthi (2011) also concluded that facilities are one of the most important drivers of the success of the Red Palm hostel in Kuala Lumpur, Malaysia, which received an award for being the best backpackers' hostel in Asia.

Security can be linked to the safety of both guests and their possessions. Hetch and Martin (2006) identified hostel safety and/or security as an important component that enhances hostel experiences. According to Brochado et al. (2015), the security dimension includes items such as location in a safe neighborhood, existence of a 24-hour front desk service, and guests' perception of safeness. Santos (2016), in turn, concluded that security is one of the most important variables explaining hostels asking premium prices worldwide.

Musa and Thirumoorthi (2011) argue that a central location is a key driver of success for the best hostel (i.e., Red Palm) in Malaysia. Location is related to hostels' convenient placement and proximity to city attractions, bars and restaurants, and public transportation (Brochado et al., 2015).

The present study also sought to assess how strongly travelers' nationality (i.e., continent of origin) influences their hostel ratings (see Figure 5). Whereas, in Beijing,

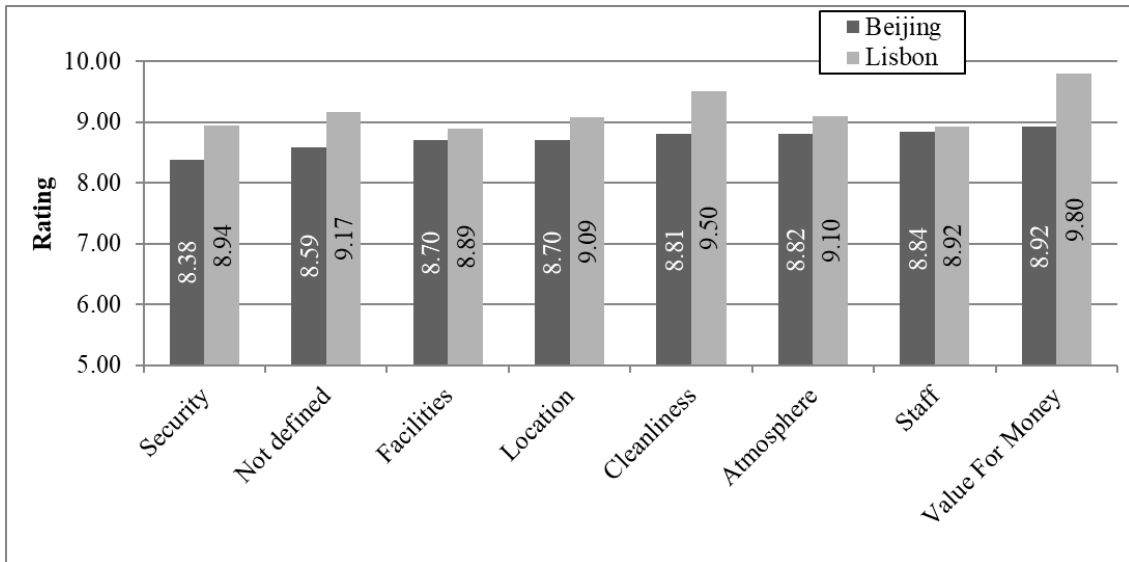
the Americas contribute the most to the highest ratings (i.e., South America [8.97] and North America [8.84]), in Lisbon, North America (9.63) is closely followed by Oceania (9.60). The lowest ratings of Lisbon hostels come from European guests, whereas the lowest ratings for Beijing hostels are generated by Asian guests.



**Figure 5** - Influence of nationality (continent) on rating.

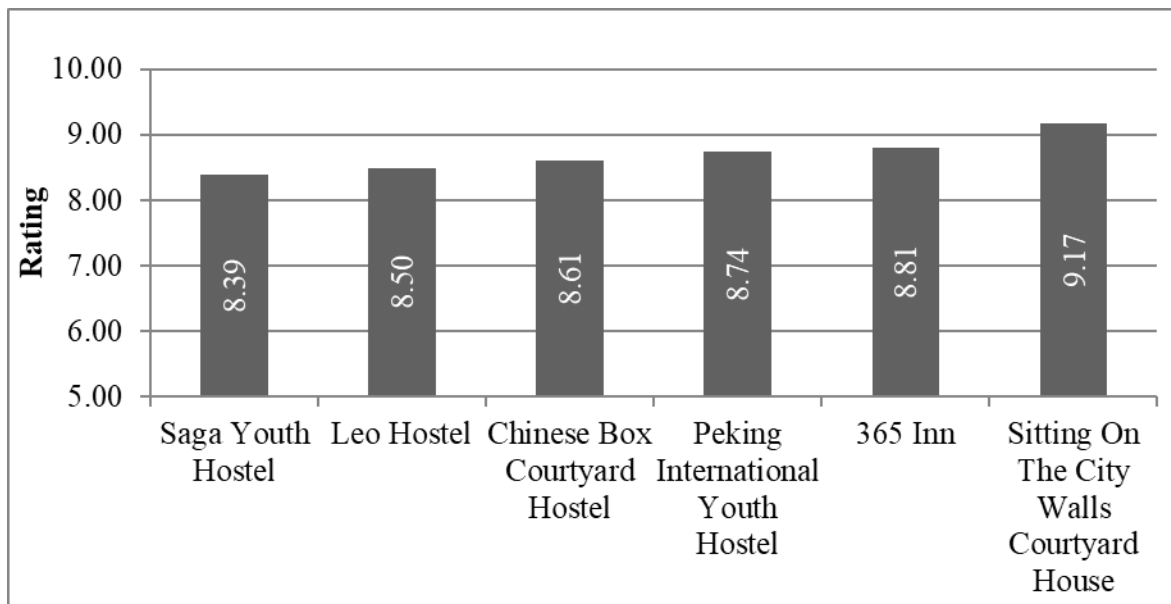
These differences in service quality ratings reinforce that hostel managers need to pay attention to guests' countries of origin (Brochado & Gameiro, 2013). According to Brochado and Rita (2018), the dimensions of service quality with the strongest impact on overall service quality vary according to hostel guests' nationality. For instance, city connections have the strongest impact on North Americans' satisfaction, while hostel tangibles stand out for Latin Americans and the quality of staff for Europeans and Australians.

After evaluating the influence of the least favorite aspect of hostel experiences on ratings (see Figure 6), the present results show that guests in both cities put value for money in first place. However, Beijing hostel clients rate staff and atmosphere as next in importance, while Lisbon guests emphasize cleanliness and location as their least favorite aspect after value for money. Security has the smallest impact on ratings among hostel clients' least favorite aspects in both cities.



**Figure 6** - Influence of the least favorite aspect on rating.

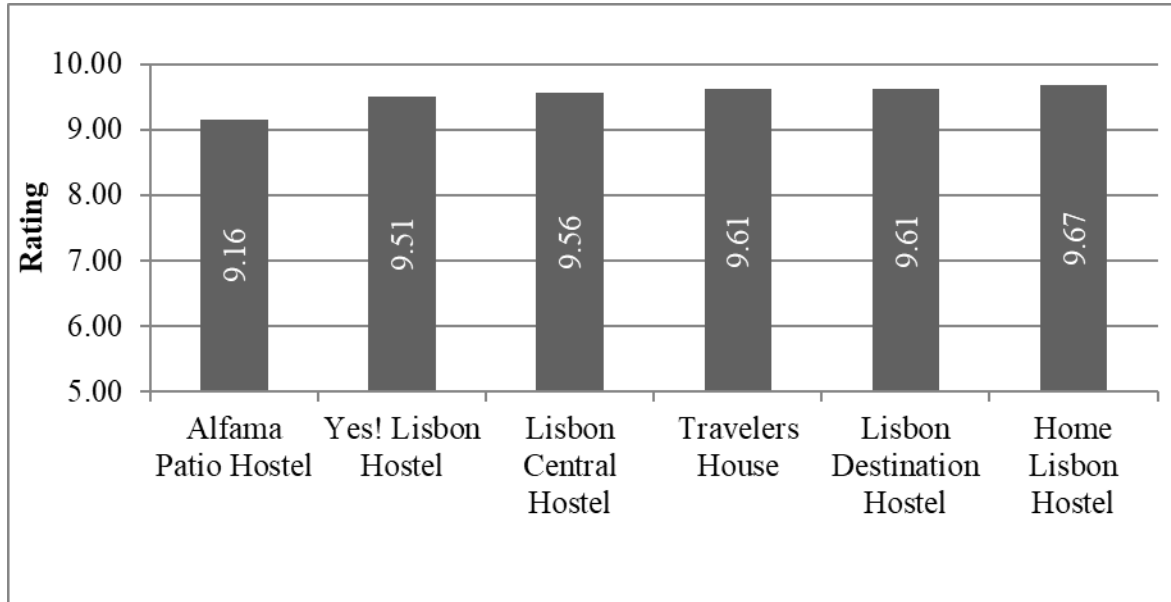
Regarding specific hostels, Sitting on the City Walls Courtyard House is the leading Beijing hostel (9.17), with a positive influence on overall ratings for this city's hostels (see Figure 7). The next two with the highest ratings are 365 Inn (8.81) and Peking International Youth Hostel (8.74).



**Figure 7** - Influence of the hostel on rating for Beijing.

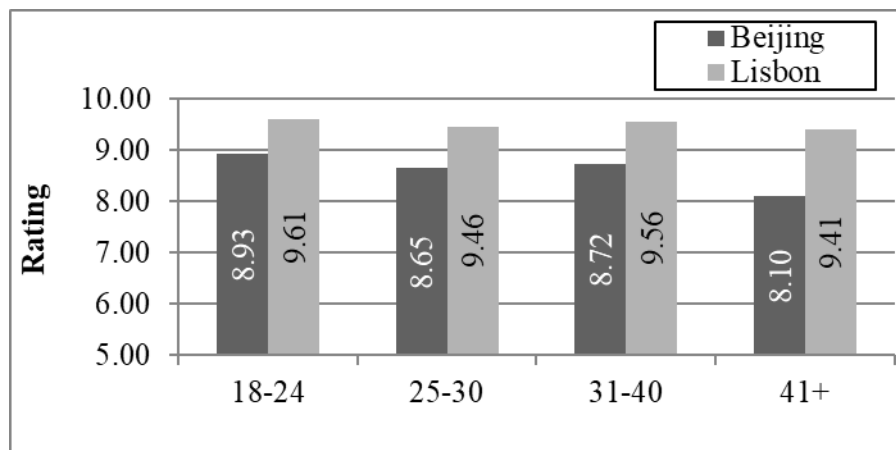
The Lisbon hostel with the strongest influence on ratings for this city's hostels (see Figure 8) is the Home Lisbon Hostel (9.67), followed by Lisbon Destination Hostel and Travelers House (both with 9.61). Although small differences were found

associated with the specific hostel variable, all hostels' contributions to Lisbon and Beijing hostels' overall ratings are extremely significant because the sample was selected from the list of hostels in each city with the highest ratings.



**Figure 8** - Influence of the hostel on rating for Lisbon.

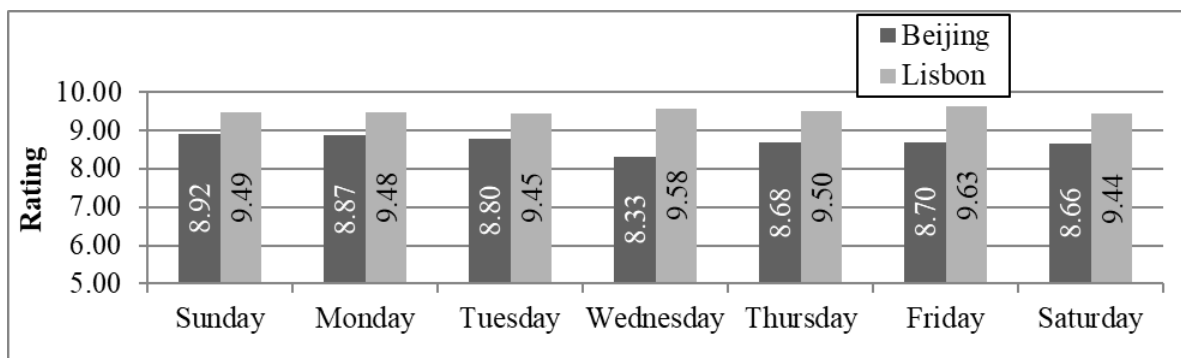
For both Beijing and Lisbon, the age group with the strongest influence on ratings (see Figure 9) is the 18–24-years-old segment (9.61 for Lisbon and 8.93 for Beijing). Notably, the 31–40-years-old group appears to have slightly more influence (9.56 and 8.72, respectively) than the 25–30-years-old segment (9.46 and 8.65) does. In both cities, the oldest age group makes the smallest contribution to the overall rating (8.10 in Beijing and 9.41 in Lisbon).



**Figure 9** - Influence of age on rating.

The existing literature on hostel guests has also confirmed heterogeneity based on age groups (Brochado & Gameiro, 2013; Brochado & Rita, 2018; Hecht & Martin, 2006). Clients' willingness to pay more for greater privacy and comfort and better facilities is positively correlated with age (Hecht & Martin, 2006). The importance of security issues is also positively correlated with age, while meeting other travelers is of utmost importance to the youngest age groups (Brochado & Gameiro, 2013). Finally, Brochado and Rita (2018) concluded that the quality of staff is the most important dimension explaining overall service quality for guests older than 30 years old.

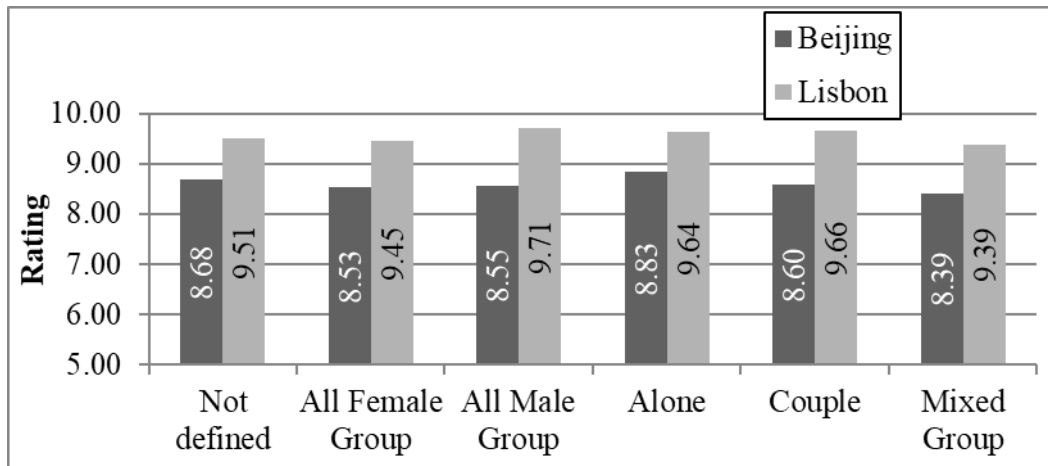
Whereas, for Beijing, Sunday (8.92) is the day of the week of reviews with the strongest influence on ratings (see Figure 10), for Lisbon, Friday (9.63) is the most important. The second most significant day in Beijing is Monday (8.87), but, for Lisbon, Wednesday (9.58) is next in importance.



**Figure 10** - Influence of week day of review on rating.

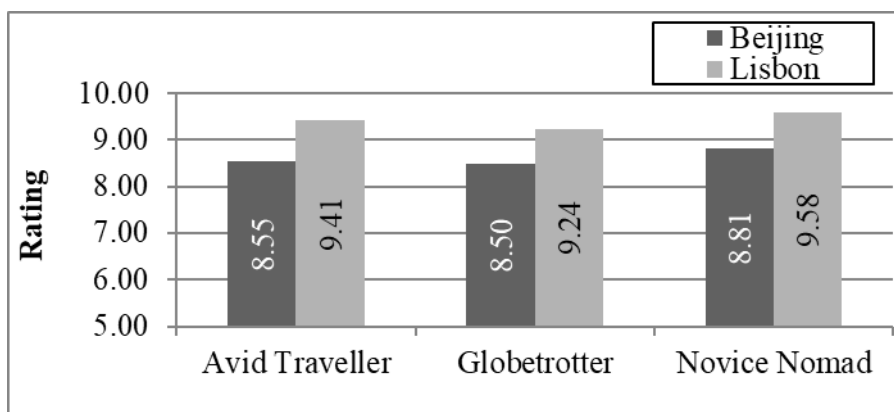
Traveling alone (8.83) exerts more influence on ratings of Beijing hostels (see Figure 11) but the all-male group (9.71) occupies the leading position for Lisbon, followed by couples (9.66) and, in third place, those who travel alone (9.64). The mixed group is the least important in both cities (9.39 in Lisbon and 8.39 in Beijing). These results are innovative as previous studies have only reported on gender differences in service quality ratings (Brochado & Gameiro, 2013; Brochado & Rita, 2018; Hecht & Martin, 2006). In prior research, social atmosphere has the strongest impact on service

quality for the all-male group, whereas tangibles have the strongest effect on the all-female group.



**Figure 11** - Influence of travel type on rating.

Travelers were categorized (i.e., novice nomad, avid traveler, and globetrotter) according to the number of reviews these hostel guests had each submitted and thus their level of experience. Novice nomads have the most influence on ratings compared with other traveler types (see Figure 12) for both Beijing (8.81) and Lisbon (9.58). Avid travelers come second (8.55 and 9.41, respectively) and globetrotters third (8.50 and 9.24). Therefore, in both cities, a negative relationship exists between guests' level of experience and their ratings.

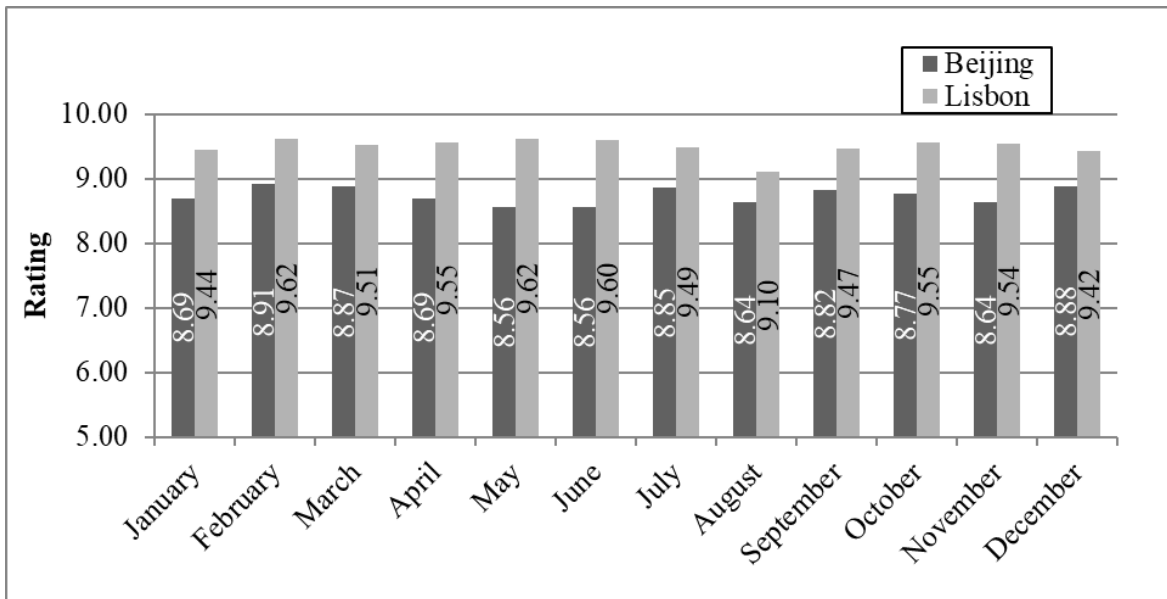


**Figure 12** - Influence of traveler type on rating.

February appears to be the most influential month of reviews in terms of ratings (see Figure 13) for Beijing (8.91) and Lisbon (9.62). However, the second and third most influential months are different. Whereas, in Beijing, these are December (8.88)



and March (8.87), respectively, in Lisbon, they are May (9.62 and so nearly as important as February) and June (9.60).



**Figure 13** - Influence of month of review on rating.

### Theoretical and practical implications

The findings emerging from this study have important implications on a theoretical and practical level. Travelers who are most concerned about security clearly give the lowest scores to hostels while staying in Lisbon or Beijing. This result reveals a common trend among guests in both cities, which has no precedent in the tourism literature. However, this finding raises the question of whether insecure guests are genuinely more demanding or whether they are also timid about giving scores and prefer to be more cautious when choosing ratings. This question will need to be explored in future research.

Notably, the results reveal that the guests' most favorite aspect and continent of origin are the two most significant features for both cities. These factors are the most influential given the evidence found for these features' significant contribution to the overall scores guests assigned to the hostels under study. In contrast, the next most

important factors are different for each city. Lisbon guests are also more strongly influenced by traveler type and the hostels themselves than travelers in Beijing are. Conversely, whereas facilities are the most valued aspect for guests staying in Lisbon hostels, in Beijing, travelers give more importance to value for money.

These findings denote different guest behaviors are associated with each city's specific hostel offer. Lisbon appears to attract diverse travelers regarding mode of travel, while visitors to Beijing are influenced by the least favorite aspect of hostels. Moreover, Beijing hostel managers need to pay special attention to the value for money variable, but Lisbon managers should invest in facilities.

A common tendency in both cities is that older guests are generally more demanding in regard to hostels. This trend was previously verified by Brochado and Gameiro (2013) among hostel guests staying in Lisbon, and the present study confirmed the same pattern for Beijing. As hostel visitors are starting to show a broader range of ages, hostels need to be prepared to host a larger spectrum of travelers, including more diverse and demanding tourists.

## **Conclusion**

This study sought to identify the main variables that explain overall service quality ratings posted online by guests who stayed in top-rated hostels located in two capital cities: Lisbon, Portugal, and Beijing, China. Based on analyses of 1,195 reviews, text mining procedures identified those factors that explain guests' service quality ratings. The correlates of overall ratings include service quality dimensions, traveler characteristics, and occasions.

The results reveal several similarities between Beijing and Lisbon. The favorite aspects of experiences and guests' country of origin (i.e., continent) are the most important variables for both cities. Regarding the favorite aspects, hostels' atmosphere

and staff were found to be significant for both cities. However, value for money has a stronger impact on Beijing versus Lisbon hostel ratings, and facilities are more important in Lisbon. Differences were also detected in security components, revealing a higher contribution to Beijing hostel ratings. In addition, a nonlinear relationship exists between age groups and overall ratings. Finally, a negative relationship was confirmed between the hostels' overall ratings and guests' level of experience with hostels.

The results offer several theoretical contributions. First, in contrast to most previous studies, the present research used user-generated content posted online instead of data collected through interviews and surveys to study service quality in hostels. Second, the current study's findings include a cross-country comparison, examining data on hostels located in Lisbon and Beijing. Although hostels are a global phenomenon, the results reveal similarities and differences between cities. Moreover, compared with past studies, this research considered a large number of attributes that might explain service quality ratings.

The results also have managerial implications since they include the most important features influencing perceived service quality of hostels in Beijing and Lisbon. In addition, this study examined the impacts on these hostels' ratings of important determining variables such as guests' nationality and age and the day and month reviews were generated, among other aspects.

The present research's limitations are mainly due to the selection of only two cities, although these were carefully chosen for their overall importance in terms of recipients of international hostel awards. The cities are, at the same time, located in quite different countries and geographical regions, which are likely to generate incoming tourists from disparate source markets.

Future studies could consider not only expanding the number of features examined. Researchers can also conduct further comparisons between major cities that are hostel destinations within the same country or between different countries and geographical regions' hostels. Furthermore, other studies could address segmentation issues among hostel guests, analyzing, for example, Millennials' behavior versus that of other generations.

## References

- Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on big data in marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics*, 24(1), 1–7.
- Brochado, A., & Gameiro, C. (2013). Toward a better understanding of backpackers' motivations. *TÉKHNE—Review of Applied Management Studies*, 11, 92–99.
- Brochado, A., & Rita, P. (2018). Exploring heterogeneity among backpackers in hostels. *Current Issues in Tourism*, 21(13), 1502–1520.
- Brochado, A., Rita, P., & Gameiro, C. (2015). Exploring backpackers' perceptions of the hostel service quality. *International Journal of Contemporary Hospitality Management*, 27(8), 1839–1855.
- Brochado, A., Troilo, M., & Shah, A. (2017). Airbnb customer experience: Evidence of convergence across three countries. *Annals of Tourism Research*, 63, 210–212.
- Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management*, 26(7), 675–693.
- Calvo, D. M. (2015). The meaning of centrality and margin in Lisbon's Rossio: Spatializing urban processes before and after the 1755 earthquake. *Portuguese Journal of Social Science*, 14(2), 123–141.
- Canito, J., Ramos, P., Moro, S., & Rita, P. (2018). Unfolding the relations between companies and technologies under the big data umbrella. *Computers in Industry*, 99, 1–8.

- Chitty, B., Ward, S., & Chua, C. (2007). An application of the ECSI model as a predictor of satisfaction and loyalty of satisfaction and loyalty of backpacker hostel. *Marketing Intelligence & Planning*, 25(6), 563–580.
- Chung, W., & Tseng, T. L. B. (2012). Discovering business intelligence from online product reviews: A rule-induction framework. *Expert Systems With Applications*, 39(15), 11870–11879.
- Cortez, P. (2010). Data mining with neural networks and support vector machines using the R/rminer tool. In P. Perner (Ed.), *ICDM 2010. Advances in data mining : applications and theoretical aspects: Proceedings of the industrial conference on data mining* (pp. 572–583). Berlin: Springer.
- Cortez, P., & Embrechts, M. J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, 225, 1–17.
- Cró, S., & Martins, A. M. (2017). The importance of security for hostel price premiums: European empirical evidence. *Tourism Management*, 60, 159–165.
- Globalization and World Cities Research Network (GaWC). (2016). *The world according to GaWC 2016*. Retrieved from <http://www.lboro.ac.uk/gawc/world2016t.html>
- Guerreiro, J., & Moro, S. (2017). Are Yelp’s tips helpful in building influential consumers? *Tourism Management Perspectives*, 24, 151–154.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. New York: Elsevier.
- He, C., Chen, T., Mao, X., & Zhou, Y. (2016). Economic transition, urbanization and population redistribution in China. *Habitat International*, 51, 39–47.
- Hecht, J.-A., & Martin, D. (2006). Backpacking and hostel-picking: An analysis from Canada. *International Journal of Contemporary Hospitality Management*, 18(1), 69–77.
- Hostelling International. (2017a). *Hostels FAQs* [Web page]. Retrieved from <https://www.hihostels.com/pages/469#two>
- Hostelling International. (2017b). *Safe and clean hostels* [Web page]. Retrieved from <https://www.hihostels.com/pages/471#standards>

- Hsu, C. H., & Huang, S. S. (2016). Reconfiguring Chinese cultural values and their tourism implications. *Tourism Management*, *54*, 230–242.
- Lee, H. A., Law, R., & Murphy, J. (2011). Helpful reviewers in TripAdvisor, an online travel community. *Journal of Travel & Tourism Marketing*, *28*(7), 675–688.
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*, 458–468.
- Magalhães, F. P. O. (2017). The historic center of Lisbon as an ecomuseum: Building bridges between tourists, places and heritage. *Revista Iberoamericana de Turismo (RITUR)*, *7*(3), 114–136.
- Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, *62*, 22–31.
- Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, *42*(3), 1314–1324.
- Moro, S., & Rita, P. (2016). Forecasting tomorrow's tourist. *Worldwide Hospitality and Tourism Themes*, *8*(6), 643–653.
- Moro, S., Rita, P., & Coelho, J. (2017). Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip. *Tourism Management Perspectives*, *23*, 41–52.
- Moro, S., Rita, P., & Oliveira, C. (2018). Factors influencing hotels' online prices. *Journal of Hospitality Marketing & Management*, *27*(4), 443–464.
- Moshin, A., & Ryan, C. (2003). Backpackers in the northern territory of Australia—motives, behaviours and satisfactions. *International Journal of Tourism Research*, *5*(2), 113–131.
- Musa, G., & Thirumoorthi, T. (2011). Red Palm: Exploring service quality and service scape of the best backpacker hostel in Asia. *Current Issues in Tourism*, *14*(2), 103–120.
- Nash, R., Thyne, M., & Davies, S. (2006). An investigation into customer satisfaction levels in the budget accommodation sector in Scotland: A case study of backpacker tourists and the Scottish Youth Hostels Association. *Tourism Management*, *27*, 525–532.

- Nave, M., Rita, P., & Guerreiro, J. (2018). A decision support system framework to track consumer sentiments in social media. *Journal of Hospitality Marketing & Management*, 27(6), 693–710.
- Park, D., Lee, J., & Han, I. (2007). The effect of online consumer reviews on consumer purchase intention: The moderating role of involvement. *International Journal of Electronic Commerce*, 11(4), 125–148.
- Phocuswright. (2016). *The global hostel marketplace: 2014–2018* [Web page]. Retrieved from <https://www.phocuswright.com/Travel-Research/Market-Overview-Sizing/The-Global-Hostel-Marketplace-2014-2018>
- Rita, P., Brochado, A., & Marques, R. (2016). Sociability and the intention to return to a hostel. In K. Rontos, J. Filipe, & P. Tsartas (Eds.), *Modelling and new trends in tourism*. Nova Science Publishers, UK.
- Rosker, J. S. (2017). Between tradition and modernity: Modern Confucianism as a form of East Asian social knowledge. *Asian Studies*, 5(2), 43–62.
- Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., & Edwards, D. D. (2003). *Artificial intelligence: A modern approach* (Vol. 2). Upper Saddle River: Prentice Hall.
- Silva, A. T., Moro, S., Rita, P., & Cortez, P. (2018). Unveiling the features of successful eBay smartphone sellers. *Journal of Retailing and Consumer Services*, 43, 311–324.
- Sparks, B. A., & Bradley, G. L. (2014). A “triple A” typology of responding to negative consumer-generated online reviews. *Journal of Hospitality & Tourism Research*, 41(6), 719–745.
- Stromquist, N. P., & Monkman, K. (2014). Defining globalization and assessing its implications for knowledge and education, revisited. *Globalization and Education: Integration and Contestation Across Cultures*, 1, 1–21.
- Wong, Y., Chaisorn, L., & Kankanhalli, M. S. (2014). Discovering person identity via large-scale observations. In C. Jawahar, & S. Shan (Eds.), *ACCV 2014. Computer vision—ACCV 2014 workshops* (pp. 602–616). Cham, Switzerland: Springer International Publishing.
- World Travel Awards. (2018). *Europe Winners 2018*. Retrieved from <https://www.worldtravelawards.com/winners/2018/europe>

Ye, Q., Li, H., Wang, Z., & Law, R. (2014). The influence of hotel price on perceived service quality and value in e-tourism an empirical investigation based on online traveler reviews. *Journal of Hospitality & Tourism Research*, 38(1), 23–39.