

## Research Article

# A Client-Centric Evaluation System to Evaluate Guest's Satisfaction on Airbnb Using Machine Learning and NLP

**Mohamed Chiny<sup>1</sup>,<sup>2</sup> Omar Bencharef,<sup>2</sup> Moulay Youssef Hadi,<sup>1</sup> and Younes Chihab<sup>1</sup>**

<sup>1</sup>Laboratory of Computer Sciences, Ibn Tofail University, Kenitra, Morocco

<sup>2</sup>Department of Computer Sciences, Cadi Ayyad University, Marrakesh, Morocco

Correspondence should be addressed to Mohamed Chiny; chiny@uca.ma

Received 1 November 2020; Revised 13 January 2021; Accepted 8 February 2021; Published 20 February 2021

Academic Editor: Jun He

Copyright © 2021 Mohamed Chiny et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Understanding the determinants of satisfaction in P2P hosting is crucial, especially with the emergence of platforms such as Airbnb, which has become the largest platform for short-term rental accommodation. Although many studies have been carried out in this direction, there are still gaps to be filled, particularly with regard to the apprehension of customers taking into account their category. In this study, we took a machine learning-based approach to examine 100,000 customer reviews left on the Airbnb platform to identify different dimensions that shape customer satisfaction according to each category studied (individuals, couples, and families). However, the data collected do not give any information on the category to which the customer belongs to. So, we applied natural language processing (NLP) algorithms to the reviews in order to find clues that could help us segment them, and then we trained two regression models, multiple linear regression and support vector regression, in order to calculate the coefficients acting on each of the 6 elementary scores (precision, cleanliness, check-in, communication, location, and value) noted on Airbnb, taking into account the category of customers who evaluated the performance of their accommodation. The results suggest that customers are not equally interested in satisfaction metrics. In addition, disparities were noted for the same indicator depending on the category to which the client belongs to. In light of these results, we suggest that improvements be made to the rating system adopted by Airbnb to make it suitable for each category to which the client belongs to.

## 1. Introduction

In 2015, PricewaterhouseCoopers [1] pointed out that five sharing sectors (travel, car sharing, finance, recruiting, and streaming music and videos) have the potential to increase global revenue from \$15 billion (estimated in 2015) to around \$ 335 billion in 2025. By focusing on the travel sector, the appearance of new peer-to-peer (P2P) models such as Airbnb has the effect of disrupting the classic reservation system [2] by offering an experience different to consumers.

Airbnb is currently the largest P2P hosting platform. It contained around 4 million ads in 2017 and was valued at \$25 billion in 2015 [3].

Taking into account the popularity of Airbnb, customers and hosts are predisposed to view rental sharing as cheaper as the platform allows each of them to rent the

property with global visibility instead of using traditional intermediaries [4] while offering an unprecedented experience given the particularity of certain properties on offer (such as igloos or castles) compared to traditional accommodation, in particular that offered by hotels [5]. In addition, this different experience stimulates the hedonic value of the customers [6].

However, since the customer experience is a determining factor that can have an impact on the recommendations of the latter and taking into account that products related to hosting cannot be tried before purchase [7], consumer interest in listings displayed on P2P platforms (of which Airbnb is one) is influenced by online user reviews [8]. It even happens that the latter have an impact on the purchasing decision since they constitute a major source of information [9] and thus contribute to electronic word of

mouth (eWOM) which has radically restructured the relationship between the business and the consumer [10].

Customer satisfaction is seen as an absolutely high priority in the hospitality industry. It establishes strong links with customer loyalty and improved the financial performance [11].

In the context of understanding customer satisfaction with accommodation exhibited on P2P hosting platforms, it turns out that, in the existing literature, there remain factors influencing rating scores (especially on Airbnb) which have not been empirically examined [3], all the more so if we approach customers by category.

Current work attempts to understand the relationship between Airbnb elementary scores (accuracy, cleanliness, check-in, communication, location, and value) and the overall score by the category of customer (individuals, couples, and families) by examining data regarding properties located in London that have been collected from inside Airbnb [12]. The scores and opinions collected were completed between December 2009 and April 2020.

Since the data provided are not classified by the customer category, we have to examine the opinions left by them by proceeding with text mining. Indeed, after the cleaning and filtering of the collected data which gave rise to 100,000 data ready to be analyzed, the segmentation operation is then carried out in order to deduce the category of customers from their opinions using natural language processing (NLP) algorithms. Due to the large amount of data to be processed as well as their semantic richness, we have adopted an approach based on machine learning. Data analysis is then performed using the multiple linear regression and support vector machine for regression algorithms. The training of these two algorithms allowed us to calculate the acting weights on each of the six elementary scores, which will allow us to better encompass the determining factors of customer satisfaction according to their category.

To validate our two models, we trained an artificial neural network to predict overall scores from elementary scores from a sample of data that is the subject of the study.

The results of this study will be important in helping P2P hosting platforms (in this case, Airbnb) as well as the hosts who offer their products there to better understand the needs of their customers in order to better develop their customer experience during their stays. On the contrary, they will be useful for Airbnb to set up a personalized scoring system that takes into account the category of customers. In other words, for future customers looking for accommodation on the Airbnb platform, it is the scores and reviews that relate to their category that will be exposed. This will better help them make a decision with respect to the offers listed. The results obtained will also provide the existing literature with additional information on the dimensions of customer satisfaction with P2P hosting platforms.

## 2. Literature Review

**2.1. Sharing Economy and P2P Platforms.** The sharing/collaborative consumption economy has made great strides in recent years. It is a form of consumption where people share

goods or services online. It represents collaborative activities to benefit, provide, or share access to goods or services, coordinated by online services that are based on a community of users [13]. Interactions between users of this form of consumption are often provided by P2P sites or platforms that facilitate contact and coordinate the exchange [14]. Moreover, the platforms of this new type of economy have acquired a significant market share in several segments such as transport (Uber and Cabify) and accommodation (Airbnb and CouchSurfing) [13]. Although the sharing economy has long existed in tightly knit communities, its shift to a much larger scale was the result of certain conditions including the rapid adoption of new technologies and low entry requirements for startups [15].

Collaborative consumption allows people to perceive the benefits of ownership while at a reduced cost, ensuring it becomes an alternative to traditional home ownership [16]. Among the services that are part of collaborative consumption is P2P tourism.

P2P tourism brings together activities carried out by tourists who interact with the attributes of the destination (including gastronomy, entertainment, and visits to natural and cultural heritage) made available by peers [17]. P2P tourism ensures a direct relationship between the host and the customer, which has the effect of promoting the authenticity perceived by the latter vis-à-vis their tourist experience [15]. Additionally, hosts can offer local experiences to their guests who, by engaging with the community, discover how the city of their stay lives [18]. Residents themselves can, through P2P tourism, contribute to tourism-related business activities. Indeed, Hamari et al. [14] identified four main factors that arouse the willingness to participate in the activities of the sharing economy as a service provider, namely, sustainability, pleasure, reputation, and economic benefits, especially with the development of political reforms in favor of the collaborative economy by certain large cities including San Francisco, Paris, London, and Singapore [15].

Since 2007, the first P2P tourism platforms have appeared, except they were not very popular. However, some of them have had great success over time, in this case, Airbnb whose activity is linked to P2P hosting which, in March 2018, had more than 150 million users and 640,000 hosts [17].

**2.2. Airbnb Phenomenon and Its Impact on the Accommodation Sector.** Airbnb is a leading platform for short-term accommodation and a pioneer in the sharing economy. It is a service that connects people who have a space to share with people who are looking for accommodation. Airbnb describes itself as a trusted community marketplace for people to list, discover, and book unique accommodations around the world [19].

Since its launch in 2008, Airbnb has grown very rapidly with more than 2 million ownership worldwide and over 50 million customers who used their services in 2015 [20].

As with the rest of the collaborative consumption platforms, technological innovations have simplified the process of entering the market and allowed it to facilitate the

list of searchable for consumers and reduce transaction costs. Airbnb provides better reach by reducing consumer search costs as electronic marketplaces reduce inefficiencies caused by buyer search costs [21]. This significant advantage has placed it at the forefront of competition with traditional providers of accommodation services (such as hotels and guesthouses). Indeed, certain stays with Airbnb can replace certain hotel stays, which affects the turnover of the latter. This impact can be differentiated by geographic area, by hotel market segment, or by season [19]. For example, Credit Suisse analysts estimated that Airbnb led to an 18.6% drop in revenue per room in January 2015 in New York [22].

Faced with this situation, the managers of hotel chains sometimes make contemptuous statements concerning competitors such as Airbnb, arguing their remarks by the fact that these platforms are a niche market, or target market segments complementary to those targeted by hotels. In fact, Airbnb for its part announced that 70% of the properties offered on its platform are located outside hotel zones [19].

What is certain is that P2P hosting platforms (in this case, Airbnb) have changed customers' perceptions of their trepidation. Many of the latter are looking for low-cost housing and direct interaction with the local community. This interaction was preceded by direct interaction with the host through the P2P platform. This has helped transform the market and attracted mainstream consumers by giving them the opportunity to rent properties as tourist residences [23].

**2.3. Online Reviews/Scoring and Customer Satisfaction.** Electronic word of mouth (eWOM) is a specific form of nonformal communication, which involves the exchange of know-how and customer feedback on a product or service. It helps shape the perception of a product's value and the likelihood of recommending it to others [24]. Electronic word-of-mouth communication, such as online hotel reviews, is gaining more and more attention from tourism marketing players, especially the hotel industry [25]. Indeed, tourism and hotel services are among the most expensive services, which imply a considerable level of risk and uncertainty vis-à-vis travel-related decision-making [26], which has the effect of customers reviewing the reviews and opinions available online in order to minimize the perceived risk to their purchasing decision [25], which makes the level of reliance on online reviews of the very high consumer share, which has a profound impact on the sale of rooms between different hotel segments [27]. Consequently, a number of companies have set up average scores which summarize the information concerning the evaluations of tourist establishments in order to make them more easily accessible to customers, in this case, the star rating system [28] whose robustness is criticized because of the large differences in assigned scores and descriptive words across different systems [10] or because of the distribution of review scores in the same system [29]. Psychic and geographic distance even happens to impact rating scores. Indeed,

travelers with a greater psychic and geographical distance give higher rating scores than travelers with less distance [30]. In addition, when they express their opinion, customers are not only influenced by their tourist experience but also by their experience using the online platform where they booked and wrote their review [4]. In any case, most online hotel reviews show very positive average ratings [31], especially in the area of accommodation which is part of the sharing economy where it is common to have a direct contact between the host and the client, which can affect the latter's decision to express negative opinions [13].

Taking advantage of the proliferation of Web 2.0 and the abundance of travel-related customer review-sharing platforms, recent studies have shed light on the factors that determine the satisfaction of customers who have decided to share their review in cyberspace. Alrawadieh [32], in his study, which looked at Tripadvisor's rating system, found that the quality and size of rooms, as well as the quality of service from staff, are the main factors influencing guest satisfaction. The same study found that relatively young European travelers are the most likely to share their experiences in cyberspace.

Zhu et al. [3] sought to identify the key determinants of rating scores on Airbnb. The results showed that communication and a large space, as well as an accurate description of the accommodation, have a positive impact on guest satisfaction. Regarding the design of the room, it turns out that the demographic characteristics of the guests have an influence on the preferences of the latter in relation to this criterion. Indeed, Bogicevic et al. [33] indicated that age and gender moderate the relationship between room style and guest satisfaction. The study noted that younger guests prefer a contemporary bedroom style, while older guests are indifferent to this criterion. Men prefer a masculine decor style, while women show equal satisfaction with a masculine and feminine decorating style. Yang et al. [34] found that generations X and Y give priority to the design of the room and the quality of offered service. The study noted that Gen Xers also value convenience and food, and Gen Yers value safety. Rather, similar results were found in Brochado et al. [35] where the study tries to identify the factors influencing scores awarded by guests who have stayed in youth hostels, where atmosphere and staff were considered important. Regarding the origin of customers, it turns out that this, linked to the cultural dimension of the destination, has an influence on the evaluation score [36].

The location of the tourist establishment also happens to have an impact on guest satisfaction. In [37], Yang et al. sought to identify whether the location of city hotels is a determining factor in guest satisfaction. Their study suggests that accessibility to points of interest (such as attractions, airport, public transport, or green spaces) is an important determinant. In addition, the effect of location satisfaction appears to be different depending on travelers' experiences and the type of trip described in reviews.

Including the monetary factor, the attributes to which customers are financially sensitive are comfort, staff, and

service [38]. As for low-budget hotels, especially capsule hotels which attract backpackers and young tourists, it has been found that price and convenience of service encourage customers to repurchase the product and share a positive eWOM [39]. However, the price factor is not influenced by the reputation of the host conveyed by online reviews, although the advantages of reputations are very important in P2P tourism experiences because services are closely related to personal hosts' skills [40].

**2.4. Big Data and Machine Learning at the Service of Customer Satisfaction.** Over the past decades, much research has been devoted to the study of factors affecting customer satisfaction using data collected through questionnaires and traditional interviews, implying a paucity of empirical data forcing researchers to focus on a handful of relevant information [11].

Moro et al. [41] pointed out that the advent of social media and opinion-sharing platforms has resulted in a thriving world of big data where consumers have become producers of content who share their opinions on products or services already tested. Indeed, big data has become a popular area of research that can add considerable value to businesses and to the society, in general, whether in the field of air pollution monitoring, assistance in life, disaster management systems, intelligent transport, etc [42]. These data typically require acquisition, cleansing, aggregation, modeling, and interpretation, which raise new challenges to derive meaningful results from these large-scale data, in this case, in terms of computing power in the face of the quantity, heterogeneity, and speed of data that characterize this area [43].

With this colossal mass of data, it is inconceivable to collect them manually. In this regard, some social networking platforms such as Facebook or Twitter offer APIs that facilitate data extraction. On the contrary, in the case of platforms which do not offer such solutions, scrapping seems to be the solution to recover a large mass of data.

This abundance of data is gradually changing research methods in the hospitality industry as it has already done in many other fields of research, leading to the adoption of new analytical tools [11]. Zheng et al. [44] investigated the usefulness of data analysis to better understand the relationship between hotel guest experience and satisfaction and found that several dimensions of guest experience have new and meaningful semantic compositions. In addition, data mining in this colossal mass of megadata can generate new information on variables that have been extensively studied in the existing hotel literature.

Combined with other disciplines that shape data sciences such as machine learning and text mining, big data can inform very specific features allowing to predict certain aspects relating to the tourism sector such as the discovery of the determinants that influence customer satisfaction [4] or the determination of the factors inducing the cancellation of reservations [45].

Finally, the implementation of the big data approach in the tourism sector, in particular the short-term accommodation sector which constitutes Airbnb's action segment, makes it possible to highlight the most relevant attributes relating to the quality of service offered, which can suggest appropriate strategies to prioritize the actions and decisions of stakeholders in this segment to improve the quality of service and make the customer experience even better. For example, Ranjbari et al. [46] were able to map Airbnb's quality of service by carrying out a two-phase survey, following, among other things, a qualitative approach where they resorted to data-mining procedures applied to a big data set of Airbnb customer reviews or in [47] where the study looked at the attributes that influence the Airbnb customer experience by analyzing tens of thousands of online reviews by applying hierarchical clustering algorithms to them, thus helping to illustrate how big data can be used to discover the attributes that facilitate engagement in the sharing economy.

Finally, machine learning techniques are widely used to predict magnitudes related to the tourism sector. For example, Kalehbasti et al. [48] attempted to come up with the best-performing model for predicting Airbnb prices based on a limited set of features using, among others, linear regression and support vector machine for regression. The study produced a promising result in terms of precision given the heterogeneity of the dataset.

### 3. Methodology

Our methodology consists of collecting opinions and scores left by customers on Airbnb. These data are then cleaned and filtered so that only those where the reviews are written in English are kept. We then proceeded to segment them in order to deduce the category to which the customer belongs to. Finally, we trained two regression models in order to calculate the scores acting on the indicators of customer satisfaction before proceeding to the prediction of the scores using a neural network alongside the two algorithms previously used.

**3.1. Data Collection.** The data for this study were collected from Inside Airbnb [12] which is an investigation site that reports scrapped data regarding the rental of the property on Airbnb. These data, which convey fairly rich information, relate to housing located in London. In addition to the accommodations and their detailed characteristics (such as type, district, and host name), the data also contain the calendar of reservations and the scores and opinions of customers who booked these properties between December 2009 and April 2020. The dataset of the data collected contains 86,357 housing units spread over 32 districts and 1,513,966 reviews left by customers. Figure 1 illustrates an example of information retrieved after cross-referencing and aggregation including the opinion, the overall score (rating), and the elementary scores noted by the customer.

	date	reviewer_name	comments	rating	accuracy	cleanliness	checkin	communication	location	value
0	2012-09-27	LaValle	The Bright Chelsea Apartment was just that. Th...	96.0	10.0	10.0	10.0	10.0	10.0	9.0
1	2010-06-17	Mohd Rafie	Grace's house is the first Airbnb experience f...	96.0	10.0	10.0	10.0	10.0	9.0	10.0
2	2010-07-25	Angela	Grace was a very gracious host; her apartment ...	96.0	10.0	10.0	10.0	10.0	9.0	10.0
3	2010-09-28	Eric	My wife and I had a great time with Grace! We...	96.0	10.0	10.0	10.0	10.0	9.0	10.0
4	2017-08-09	Kate	We had a wonderful stay in Kew Gardens while i...	97.0	10.0	10.0	10.0	10.0	10.0	9.0
5	2010-10-05	Jean Miguel & Jamie	Wonderful experience, great B&B to stay at. Ag...	97.0	10.0	10.0	10.0	10.0	10.0	10.0
6	2010-12-28	Ania	Agri and Roger are a wonderful couple! They w...	Gras 97.0	10.0	10.0	10.0	10.0	10.0	10.0

FIGURE 1: Extraction of the dataset collected after crossing and aggregation.

**3.2. Data Cleaning and Filtering.** To ensure an acceptable level of relevance of the data collected, only the opinions of the customers which do not present missing or inconsistent fields are kept. Then, we filtered the reviews written in English. To do this, we could have proceeded to the language identification based on the recognition of words using a dictionary; however, this method requires a large dictionary, and in addition, it requires a considerable processing time. We therefore used the approach of calculating the probability of the language based on the pronunciation characteristics using the naive Bayes algorithm with character N-gram. This method has proven its performance with an accuracy of 99.8% and a satisfactory detection speed [49].

Despite the fact that the method used to detect the language is considerably fast, but given the number of opinions to be analyzed (1,513,966 opinions), the calculation time required to perform this processing is still significant. We therefore retained 100,000 reviews ready to be used (Figure 2).

**3.3. Segmentation of Reviews according to the Customer Category.** One of the major limitations we were faced with is the fact that the data collected from Inside Airbnb did not convey the personal information of customers nor even the number of people who have booked the accommodation in order to be able to deduce the category (individual, couple, or family) to which they belong. Indeed, the only personal data that the dataset contains are the first name of the reviewer. On the contrary, the information concerning the accommodation is remarkably rich and complete. Taking this into account, we decided to analyze the reviews using natural language processing (NLP) algorithms to find clues that can help us deduce the most plausible category to which the customer belongs to.

First, we isolated the opinions of the customers who were supposed to be part of the “individual” category by using the syntactic dependency analysis [50, 51] which makes it possible to highlight the syntactic structure of a sentence and the relationships that its elements maintain. We sought to identify subjects alluding to the first-person singular in the dependency tree as shown in the example in Figure 3.

At the end, we were able to segment 38,543 reviews of alleged customers belonging to the category “individual.”

We would like to note that we could also treat other syntactical aspects such as complements, but with the method described above, we estimated that the number of filtered reviews is sufficient for our study.

For the remaining two categories (couples and families), we did not find NLP methods or algorithms which allow us to proceed with the expected segmentation, so we proceeded to the processing of strings via regular expressions using keywords such as “my family,” my wife, and *i*. and we could isolate 5495 reviews of couples and 10,874 reviews of families.

For the three categories, we have performed manual testing by analyzing a random sample of opinion to assess the performance of our approach.

The remainder of the reviews of the 100,000 initially selected (which constitute about half) do not contain enough occurrences to determine their membership (Figure 2).

The next step is to train two machine learning models on the elementary scores relating to the reviews selected in order to calculate the weights of the latter using regression algorithms.

## 4. Method Specification for Calculating the Coefficients

The regression models which are based on the weights will allow us, during our study, to calculate the weighting coefficients of each satisfaction indicator noted on Airbnb.

**4.1. Regression Algorithms.** Machine learning, which is a part of artificial intelligence that simulates human cognitive processes, is a field of learning that enables machines to automate intelligent tasks and extract knowledge from a large set of raw data. Machine learning is currently used in many fields such as economics, medicine, or meteorology. Depending on the continuity or discretion of the model output, regression or classification algorithms can be used. In our case, the output defines a score on a scale of 100 where

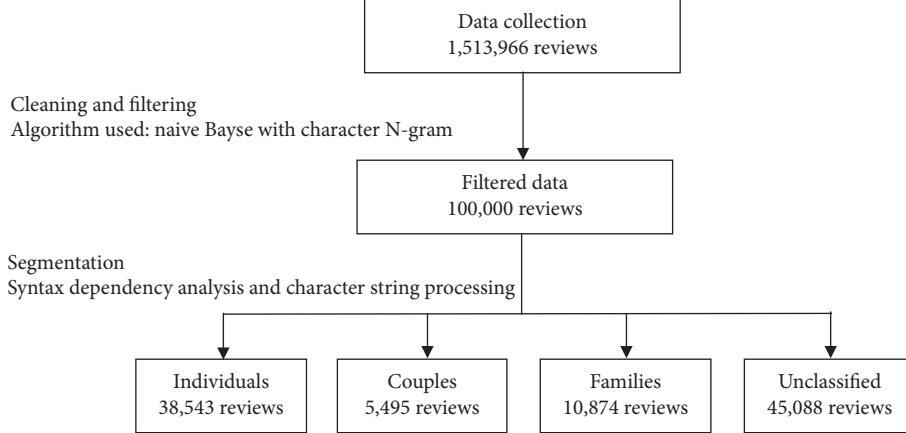


FIGURE 2: Process cleaning, filtering, and segmentation of data collected from Inside Airbnb.

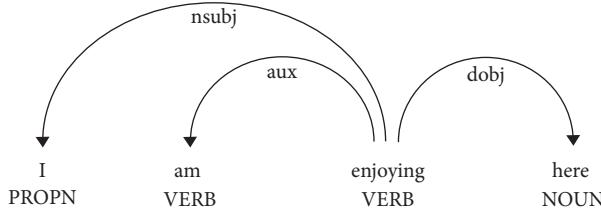


FIGURE 3: The subject of the sentence “I” referring to the first-person singular, which most likely implies that the customer belongs to the category individual.

all the intermediate values are possible. We therefore adopted for regression algorithms.

Regression is a method of modeling a variable (called target) according to independent predictors (called features), where the algorithm involved tries to find the cause and effect relationships between the variables. Among the regression algorithms, we find the multiple linear regression (MLR) and the support vector machine for regression (SVR).

**4.1.1. Multiple Linear Regression (MLR).** Multiple linear regression or simply multiple regression is a statistical technique that uses a set of variables to predict the outcome. Its objective is to model a linear relationship between the explanatory variables which are independent and the response variable which is dependent. Essentially, MLR is an extension of least squares regression that requires more than one explanatory variable [52]. Recent works such as [53] have taken advantage of the power of MLR to establish reliable and conclusive models, particularly in the tourism sector.

**4.1.2. Support Vector Machine for Regression (SVR).** Support vector machine was introduced by Vapnik and colleagues in 1992 initially to resolve classification issues. It is applied in many fields of business, science, and industry to classify and recognize patterns [54]. Its principle has been extended to regression in order to establish predictions. It

relies on the kernel trick where input data are plotted into a new hyperspace whose purpose is to find the most optimal for data fit and prediction [52].

**4.2. Model Specification and Data Study.** Taking into account the six elementary indicators that are rated by the customer (accuracy, cleanliness, check-in, communication, location, and value) and which are provided with the data collected, we plan to identify the degree of their influence on the overall score, depending on the category of customers, by calculating their respective weights. For our study, we summarize the overall score using the following model:

$$GS = \beta_1 PRS + \beta_2 CLN + \beta_3 CHK + \beta_4 COM + \beta_5 PLC + \beta_6 VAL, \quad (1)$$

where PRS, CLN, CHK, COM, PLC, and VAL denote, respectively, precision, cleanliness, check-in, communication, location, and value and  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ , and  $\beta_6$  the respective coefficients which indicate the weight of each indicator vis-à-vis the overall score.

Figure 4 illustrates the representation of the overall score (on a scale of 100) as a function of the elementary scores (on a scale of 10) according to the category of customers.

We have trained two regression algorithms, multiple linear regression (MLR) and support vector regression (SVR), in order to calculate the coefficients acting on each indicator

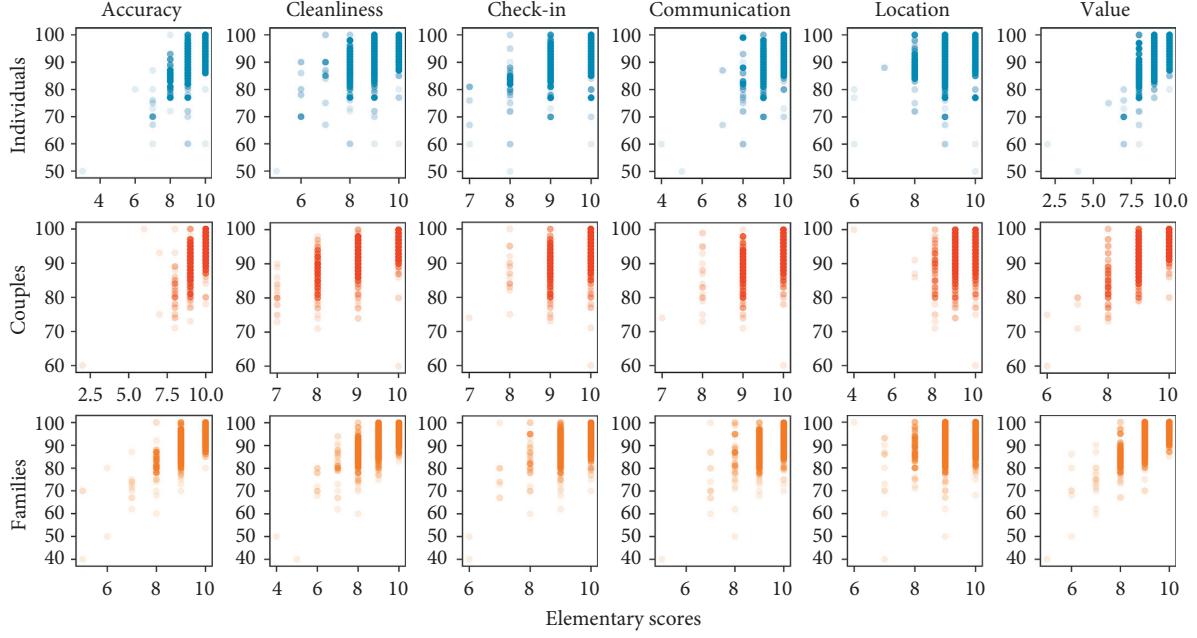


FIGURE 4: Scatter plot illustrating the overall score as a function of Elementary scores for the three categories of customers.

and thus determine the degree of influence these on the overall score. For each of these two algorithms, training is carried out over 3 phases in order to define the appropriate model for each category of customers. The training data are those used in the section devoted to segmentation, namely, 38,543 reviews for individuals, 5,495 reviews for couples, and 10,874 reviews for families. The inputs to our models are the 6 elementary scores noted by the customer (accuracy, cleanliness, check-in, communication, location, and value), and the output is the overall score.

## 5. Results

The dataset was divided into two batches, training data and test data, with percentages of 80% and 20%, respectively. The selection of data for training and testing was completely random.

After training our two models, we were able to obtain the accuracies in Table 1 using the  $R^2$  score. The two algorithms used illustrate fairly close accuracies for each of the three studied categories.

Table 2 illustrates the coefficients calculated using the MLR and SVR algorithms according to the specified customer categories. The average of these coefficients was also calculated because we estimate that they will be useful to us during subsequent interpretations.

Figures 5 and 6, respectively, represent the distribution of the coefficients according to the six elementary scores using the MLR and SVR algorithms according to the 3 categories of customers studied.

Overall, the two algorithms give fairly consistent results with respect to the six indicators studied. In the following, we present these results by treating each indicator

TABLE 1: Accuracy of MLR and SVR algorithms.

	MLR	SVR
Individual	0.660	0.657
Couple	0.684	0.676
Family	0.712	0.711

independently; they have been ranked in the order of importance (from most important to least important) on the basis of the calculated weights.

**5.1. Accuracy.** Based on the results, accuracy appears to be the most important determinant of customer satisfaction. The coefficients calculated with respect to this indicator are the highest among the 6 indicators that constitute the subject of the study. The precision score indicates how well the customer finds the description of the accommodation as well as other related aspects such as neighborhood or convenience to be true to reality and to provide as much information as possible. In addition to the description written by the host, there are also photos which must be up to date and illustrate the current situation of the accommodation. The coefficient calculated for the precision using the MLR and SVR algorithms has a respective mean of 2.298 and 2.334, and it is noted that it is the families which give it the most importance with the coefficients 2.634 and 2.6, respectively, to the MLR and SVR.

The importance of accuracy in the eyes of clients has been made explicit in a recent study which found that providing information about housing and its environment has a positive impact on client satisfaction [55]. According to Guttentag et al. [56], customers who search for

TABLE 2: Coefficients acting on the 6 satisfaction indicators using the algorithms MLR and SVR.

	MLR algorithm				SVR algorithm			
	Individual	Couple	Family	Mean	Individual	Couple	Family	Mean
Accuracy	2.155	2.106	2.634	2.298	2.402	2.000	2.600	2.334
Cleanliness	1.705	2.048	1.781	1.845	1.799	2.000	1.800	1.866
Check-in	0.852	0.610	0.905	0.789	0.602	1.000	1.000	0.867
Communication	1.626	1.609	1.084	1.440	1.799	1.200	1.199	1.399
Location	0.380	0.467	0.649	0.499	0.398	0.800	0.600	0.599
Value	1.670	1.959	1.802	1.810	1.598	2.000	1.800	1.799

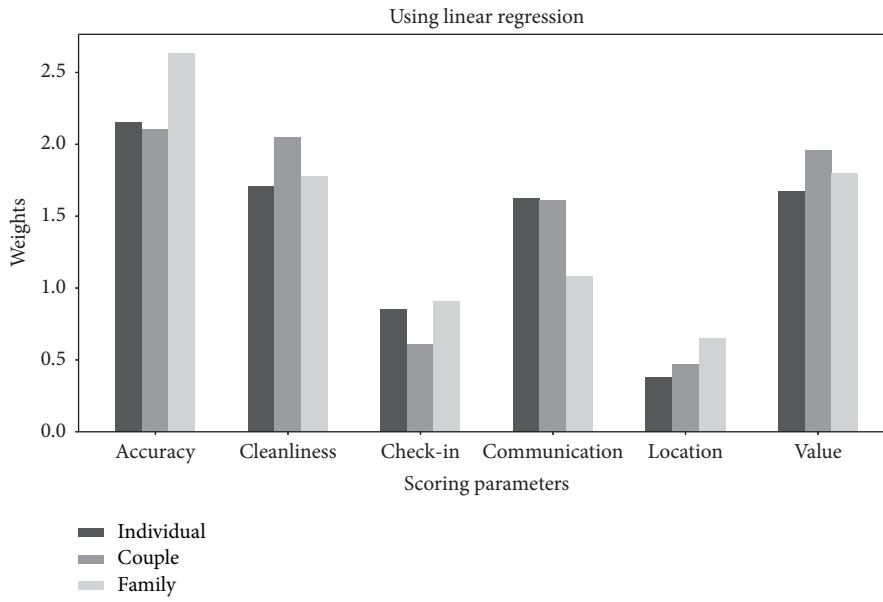


FIGURE 5: Distribution of coefficients calculated using the MLR algorithm.

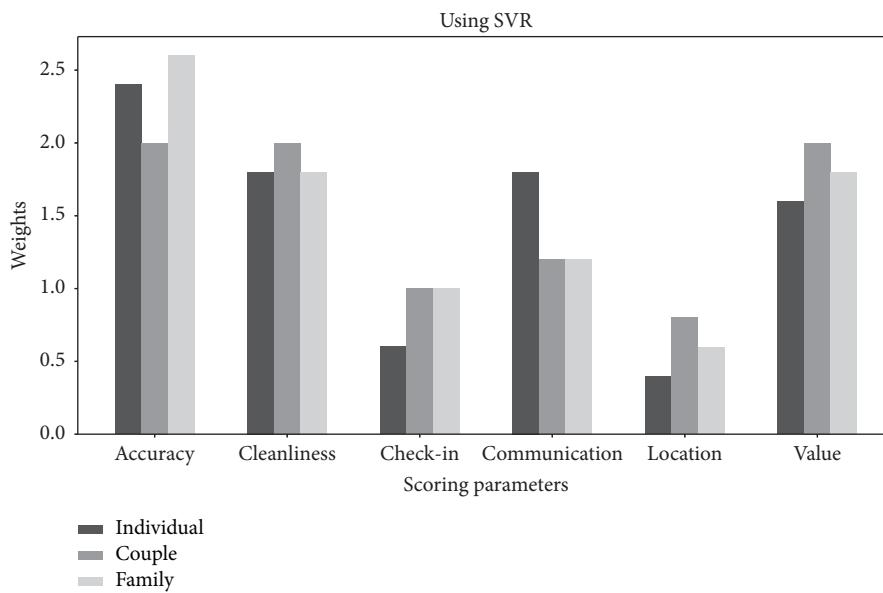


FIGURE 6: Distribution of coefficients calculated using the SVR algorithm.

accommodation on Airbnb appreciate convenience and pay close attention to the hosting environment, and they are likely to make better decisions if this information is provided in a reliable and detailed manner.

**5.2. Cleanliness.** The cleanliness score indicates the degree of customer satisfaction with the cleanliness and good tidiness of the accommodation. For some customers, cleanliness is almost literally a hygiene factor because it is generally associated with the condition of the sheets, mattresses, pillows, the floor, the bathroom, etc., and it is one of the most reported determinants of customer satisfaction in the hotel literature [57], although Lockyer [58] pointed out that customer expectations with regard to cleanliness often exceed the performance offered by accommodation establishments.

Overall, the results of our study align with the fact that cleanliness is an important determinant of customer satisfaction. In fact, cleanliness is the second indicator that is important with regard to the overall score with an average of coefficients of 1.845 and 1.866, respectively, for the MLR and SVR. For these two algorithms, it turns out that it is the couples that are most influenced by cleanliness with the values of coefficients 2.048 and 2, respectively, for the MLR and SVR algorithms.

**5.3. Value.** The value score reflects how satisfied the customer is vis-à-vis the price-quality of the home. Indeed, a key dimension of the hospitality industry is pricing. This is a determining factor in the long-term success of the accommodation industry [59]. However, price indicators used in the conventional hotel industry, such as star rating or membership in a branded hotel chain, are not applicable to P2P accommodation offers [23]. It is for this reason that studies have focused on identifying the price determinants of P2P hosting offers in the digital market, in this case on Airbnb [59].

In our study, we found that the average weight calculated using the MLR and SVR algorithms for the value is 1.81 and 1.799, respectively. These values are quite close to the weights calculated for cleanliness. Moreover, the distribution of coefficients according to the customer category for these two indicators (value and cleanliness) is almost identical. Indeed, it is still the couples which are the most influenced by the value with the coefficients 1.959 and 2 calculated, respectively, using the MLR and SVR algorithms.

It would be worth mentioning that hosts and guests probably perceive rental sharing in advance as cheaper [5], which perhaps justifies why the value in our study is not ranked higher among the indicators studied. Moreover, Wang and Nicolau [59] mentioned that, on the basis of the results they found, they raised the complexity of the price-determining relationship in P2P hosting by mentioning that the determining factors of the price of P2P hosting are different from those that determine hotel prices.

**5.4. Communication.** The communication score indicates how much guests enjoy having their host interact quickly and frequently with their questions reliably and accurately before and during their stay. With regard to this indicator, we recorded an average of 1.44 and 1.4, respectively, for the two algorithms MLR and SVR, and we see that it is the individuals who give it the most importance with the coefficients 1.626 and 1.8, respectively, for the MLR and SVR.

According to Zhu et al. and Madalyn [55, 60], this criterion is often considered decisive in terms of shaping and maintaining the relationship between the host and the guest and constitutes a form of hospitality perceived by the latter. Communication is also seen as a way to foster trust between the host and customers in P2P housing [23]. However, the study of Santos et al. [13] mentioned that communication (or personal contact between the client and the host, in general) during the stay often puts clients in an uncomfortable position to carry out a negative evaluation of the services, which can skew the value of the service evaluation, in particular, with regard to communication.

**5.5. Check-In.** The check-in score indicates whether the user is satisfied with the check-in process upon arrival. According to Ranjbari et al. and Sun et al. [46, 61], this indicator is considered among the key points of customer satisfaction with the quality of service offered by Airbnb.

In our study, the two algorithms MLR and SVR generated respective mean coefficients of 0.789 and 0.867. Different results were recorded with respect to the coefficients relating to individuals and couples according to the two algorithms. However, the latter agree on the importance of the coefficient relating to families which are equal to 0.905 and 1, and it turns out that it is the families that are most influenced by this indicator. However, overall, this indicator comes in the 5th position, which indicates that it is not as significant as those seen previously. Indeed, by referring to [62], this can be explained by the fact that in terms of registration, more than one-third of Airbnb ads allow customers to register by themselves; this means that more than one-third of customers do not face any check-in issues, including waiting time. In addition, according to Cheng and Jin [63], amenities such as self-checking provide a sense of privacy and more flexible options, which lessen the weight of the check-in indicator compared to other indicators already discussed.

**5.6. Location.** The location score indicates whether or not the customer is satisfied with the location of the accommodation. This score can be influenced by the proximity of and access to transport, shopping centers, city centers, etc. It can also take into account particular aspects such as safety or noise. In addition, this indicator may depend on the accuracy of the description provided by the host [20].

Hotel location is an important consideration in hotel selection and is an important factor influencing guest

satisfaction and recall of travel experiences [37]. In [64], the authors concluded that the expected points of a hotel location given by travelers (FIT guests) are security, ease of access to transport, and proximity to attractions.

In the P2P hosting industry, location is a serious consideration for customers who pay a lot of attention to the hosting environment [55].

However, in our study, the mean coefficients calculated, respectively, using the MLR and SVR algorithms are 0.499 and 0.599, which put the location at the bottom of the ranking of the indicators that influence customer satisfaction. The two algorithms used give different results for couples and families. For individuals, on the contrary, the weights of 0.38 and 0.398 found, respectively, using MLR and SVR show that they are the least influenced by the location of the dwelling.

These results can be explained by the fact that clients are already aware of the situation of the location and environment of the accommodation, and they have a clear idea about the convenience and proximity to places and spaces of interest. Indeed, according to Guttentag et al. [56], customers are generally satisfied if more environmental information has been provided, and they are able to make better decisions based on this information. This leads us to consider the effect of a good description on the location, which is supported by Airbnb itself which suggests that the score on the location may depend on the score on the accuracy of the description [20].

## 6. Synthesis

After exploring these results, it is clear that customer expectations for P2P hosting, in this case, at Airbnb, are quite different. An overview of Figures 4 and 5 illustrates these disparities, especially between the 3 categories of customers studied. Basically, the general tendency of the determinants of satisfaction is almost the same for these 3 categories because although there are differences between the results found using the MLR and SVR algorithms, we still see that all the indicators are listed in the following order: precision, cleanliness, value, communication, check-in, and location, despite the fact that, by using the SVR algorithm, we have found that individuals attach more importance to communication than the value, inversely to the result found using the MLR algorithm. Yet, some indicators are clear, as is the case for accuracy which is the most important determinant for families, followed by cleanliness and value for couples, and then communication for individuals.

## 7. Prediction of Scores

In the light of the results found and which consisted in the calculation of the weights which act on the elementary scores, it would be interesting to proceed to the prediction of

the overall scores of certain collected listings. We therefore proceeded to the selection of a sample of 24 accommodations already reserved and belonging to the 4 listing categories offered in London (private room, entire home/apt, hotel room, and shared room), and we calculated the overall score which would have been given by the client based on the actual elementary scores that were assigned (Table 3). This score will mainly depend on the category to which the customer belongs to. We therefore proceeded to predict the overall scores using the two algorithms used in our study, namely, MLR and SVR. In order to consolidate the results, we also calculated the score using an artificial neural network (NN). The results of the predictions are listed in Table 4.

Our neural network has three hidden layers with 30 nodes each. It is this configuration that allowed us to have the best possible precision (Table 5).

The results of the prediction confirm the conclusion drawn from the calculation of the weights acting on the satisfaction indicators set by Airbnb for the rating. Indeed, despite the difference in scores calculated using the three algorithms (MLR, SVR, and NN), it is clear that the overall scores calculated on the basis of the same elementary scores change depending on the category to which the client belongs to.

However, if we pay attention to the scores of certain listings, for example, L24, who obtained 10 out of the 6 elementary scores, we would probably think that the score of the overall score whose value we predicted would be equal to 100, but this is not the case. Indeed, although the overall scores that were predicted are significantly higher than those of the other listings, which had lower elementary scores, the maximum score obtained is 98.194. This is the score which the families would have given based on the predictions of the NN algorithm. This may mean that the determinants of customer satisfaction exceed the six metrics rated on Airbnb. Indeed, age can be decisive in terms of shaping satisfaction. Moreover, this indicator has been the subject of study in several works such as [34]. The sex [33], cultural dimensions [36, 65], and geographic and psychological dimensions of clients [30] can also influence the score.

Some studies have gone even further by studying the effect of gamification integrated into review sites, in this case, on TripAdvisor [66], such as the rating system or badges, and it turned out that such features may affect the behavior of the traveler when writing the review.

It even happens that the design of the room [33] and the features available to customers such as WiFi [67] have a significant impact on their satisfaction.

To sum up, it would be too unfair to claim that the overall score is shaped purely based on the elementary scores noted on Airbnb. Indeed, to have a broader idea of the determinants of customer satisfaction, we must take into account all these indicators, and many others, to elucidate the factors that affect this complex feeling of customers which is satisfaction.

TABLE 3: Real scores by indicators attributed to a sample of 24 selected listings.

Listing ID	Room type	Accuracy	Cleanliness	Check-in	Communication	Location	Value
L1	Private room	10.0	10.0	10.0	10.0	9.0	9.0
L2	Private room	10.0	10.0	10.0	10.0	9.0	10.0
L3	Private room	9.0	10.0	10.0	10.0	10.0	8.0
L4	Private room	10.0	10.0	10.0	10.0	10.0	10.0
L5	Private room	10.0	10.0	10.0	10.0	9.0	10.0
L6	Private room	10.0	10.0	10.0	10.0	9.0	10.0
L7	Entire home/apt	10.0	9.0	9.0	9.0	10.0	9.0
L8	Entire home/apt	10.0	10.0	10.0	10.0	10.0	9.0
L9	Entire home/apt	10.0	10.0	10.0	10.0	10.0	10.0
L10	Entire home/apt	10.0	9.0	10.0	10.0	9.0	9.0
L11	Entire home/apt	9.0	9.0	10.0	9.0	9.0	9.0
L12	Entire home/apt	10.0	10.0	10.0	10.0	9.0	9.0
L13	Hotel room	9.0	10.0	10.0	10.0	9.0	10.0
L14	Hotel room	9.0	9.0	10.0	9.0	9.0	9.0
L15	Hotel room	10.0	10.0	10.0	10.0	9.0	10.0
L16	Hotel room	10.0	10.0	10.0	10.0	10.0	9.0
L17	Hotel room	10.0	10.0	10.0	10.0	10.0	9.0
L18	Hotel room	10.0	10.0	10.0	10.0	10.0	10.0
L19	Shared room	10.0	9.0	10.0	10.0	10.0	10.0
L20	Shared room	10.0	10.0	10.0	10.0	9.0	10.0
L21	Shared room	10.0	10.0	10.0	10.0	10.0	10.0
L22	Shared room	8.0	9.0	9.0	8.0	8.0	8.0
L23	Shared room	8.0	8.0	9.0	9.0	10.0	8.0
L24	Shared room	10.0	10.0	10.0	10.0	10.0	10.0

TABLE 4: Overall scores predicted on the basis of real elementary scores on a sample of 24 listings according to the category of customers using MLR, SVR, and NN algorithms.

Listing ID	Room type	MLR			SVR			NN		
		Individual	Couple	Family	Individual	Couple	Family	Individual	Couple	Family
L1	Private room	95.038	95.091	95.337	95.302	95.100	95.500	95.167	94.982	95.645
L2	Private room	96.708	97.050	97.139	96.899	97.100	97.300	96.453	96.993	97.411
L3	Private room	91.592	91.494	91.551	91.700	91.900	91.700	91.565	92.041	92.221
L4	Private room	97.088	97.518	97.788	97.297	97.900	97.900	97.176	97.952	98.194
L5	Private room	96.708	97.050	97.139	96.899	97.100	97.300	96.453	96.993	97.411
L6	Private room	96.708	97.050	97.139	96.899	97.100	97.300	96.453	96.993	97.411
L7	Entire home/apt	91.235	91.291	92.216	91.499	91.701	92.100	90.620	91.269	91.884
L8	Entire home/apt	95.418	95.559	95.986	95.700	95.900	96.100	95.321	95.941	96.435
L9	Entire home/apt	97.088	97.518	97.788	97.297	97.900	97.900	97.176	97.952	98.194
L10	Entire home/apt	93.333	93.043	93.556	93.502	93.100	93.700	93.860	92.922	93.791
L11	Entire home/apt	89.552	89.329	89.838	89.301	89.900	89.900	88.962	89.166	89.626
L12	Entire home/apt	95.038	95.091	95.337	95.302	95.100	95.500	95.167	94.982	95.645
L13	Hotel room	94.553	94.945	94.505	94.497	95.100	94.700	94.650	95.105	94.991
L14	Hotel room	89.552	89.329	89.838	89.301	89.900	89.900	88.962	89.166	89.626
L15	Hotel room	96.708	97.050	97.139	96.899	97.100	97.300	96.453	96.993	97.411
L16	Hotel room	95.418	95.559	95.986	95.700	95.900	96.100	95.321	95.941	96.435
L17	Hotel room	95.418	95.559	95.986	95.700	95.900	96.100	95.321	95.941	96.435
L18	Hotel room	97.088	97.518	97.788	97.297	97.900	97.900	97.176	97.952	98.194
L19	Shared room	95.383	95.469	96.007	95.498	95.900	96.100	95.209	95.893	96.360
L20	Shared room	96.708	97.050	97.139	96.899	97.100	97.300	96.453	96.993	97.411
L21	Shared room	97.088	97.518	97.788	97.297	97.900	97.900	97.176	97.952	98.194
L22	Shared room	82.868	82.578	82.764	82.502	82.901	82.701	81.890	81.697	81.958
L23	Shared room	83.549	83.073	83.366	83.298	83.701	83.300	83.097	83.421	83.334
L24	Shared room	97.088	97.518	97.788	97.297	97.900	97.900	97.176	97.952	98.194

TABLE 5: Accuracy of the NN algorithms after training the model on the data retained from the dataset.

	Neural network
Individual	0.668
Couple	0.680
Family	0.710

## 8. Conclusion

Given that Airbnb is currently the largest platform for short-term rental accommodation and which has considerably disrupted the traditional reservation system, many studies have been interested in demystifying the dimensions of customer satisfaction with towards the P2P hosting domain over the past decade.

In our study, we sought to understand which indicators determine customer satisfaction, taking into account their category with regard to the accommodation offered by Airbnb. We have therefore collected a large body of reviews from users who have booked properties in London between December 2009 and April 2020 from Inside Airbnb. After cleaning, filtering, and segmenting these opinions, using mainly natural language processing (NLP) algorithms, we used regression algorithms, in this case, multiple linear regression (MLR) and support vector regression (SVR), in order to calculate the coefficients which act on the elementary scores noted by the customers and which influence the overall score. Then, we simulated global scores by applying the artificial neural networks, as well as the two algorithms MLR and SVR, to the real elementary scores assigned to a listing sample according to one of the three categories of clients studied.

The results suggest that these indicators are viewed by customers in the following order: accuracy, cleanliness, value, communication, check-in, and location. In addition, the understanding of these indicators changes from one category to another, for example, families attach importance to precision more than other categories, couples are interested, in addition to precision, by cleanliness and value, and communication is an indicator taken seriously by individuals.

However, it is clear that the six scores rated by customers are not the only ones influencing the overall score. Indeed, dimensions that are not taken into account by this study such as age, gender, and cultural and geographic dimensions of customers can have an impact on customer satisfaction, and therefore on score attributed to housing. It would be useful to adopt the approach of this study by taking into account other indicators in order to better encompass the determinants of customer satisfaction according to their category.

However, we recognize that there is a limitation in the way we have segmented customer reviews in order to derive categories. Indeed, we have searched in the existing literature for algorithms that would allow us to perform such an operation, but unfortunately, we have not found anything. Although the tests on the samples that we took manually to confirm category membership were satisfying, it would be

fair to mention that a scientifically approved method would give more relevant results.

In the end, we believe that this work can help contribute in the field of P2P hosting, in particular, through Airbnb, to better understand the expectations of the customers in order to apprehend them by taking into account the category to which they belong in the aim to guarantee them the best possible experience. In addition, Airbnb may consider implementing a new scoring system that takes into account the category of customers, something that will most likely help future customers to better make decisions about the offers listed. On the contrary, current work provides the literature with additional answers on the determinants of customer satisfaction, especially in the area of P2P hosting.

## Data Availability

The raw data used in this study were downloaded from <http://insideairbnb.com/>. Then, these data have undergone a lot of processing through Jupyter Notebook. If necessary, the codes and their execution results can be made available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] <https://www.pwc.com/us/en/technology/publications/assets/pwc-consumer-intelligence-series-the-sharing-economy.pdf>, 2015.
- [2] G. Zhang, R. Cui, M. Cheng, Q. Zhang, and Z. Li, "A comparison of key attributes between peer-to-peer accommodations and hotels using online reviews," *Current Issues in Tourism*, vol. 23, no. 5, pp. 530–537, 2019.
- [3] L. Zhu, M. Cheng, and A. Wong, "Determinants of peer-to-peer rental rating scores: the case of Airbnb," *International Journal of Contemporary Hospitality Management*, vol. 31, no. 9, 2019.
- [4] S.. Moro, P. Rita, J. Esmerado, and C. Oliveira, "Unfolding the drivers for sentiments generated by Airbnb Experiences," *International Journal of Culture, Tourism and Hospitality Research*, vol. 13, no. 4, 2019.
- [5] M. Chattopadhyay and S. K. Mitra, "Do airbnb host listing attributes influence room pricing homogenously?" *International Journal of Hospitality Management*, vol. 81, pp. 54–64, 2019.
- [6] S. Rosengren, "Experience value as a function of hedonic and utilitarian dominant services," *International Journal of Contemporary Hospitality Management*, vol. 28, no. 1, 2014.
- [7] G. Cetin and A. Walls, "Understanding the customer experiences from the perspective of guests and hotel managers: empirical findings from luxury hotels in istanbul, Turkey," in *Proceedings of the 17th Annual Graduate Student Research Conference in Hospitality and Tourism*, Washington, DC, 2012.
- [8] D. E. Boyd, T. B. Clarke, and R. E. Spekman, "The emergence and impact of consumer brand empowerment in online social networks: a proposed ontology," *Journal of Brand Management*, vol. 21, no. 6, pp. 516–531, 2014.
- [9] E. Martin-Fuentes, C. Mateu, and C. Fernandez, "The more the merrier? number of reviews versus score on TripAdvisor

- and booking.com," *International Journal of Hospitality & Tourism Administration*, vol. 21, no. 1, pp. 1–14, 2018.
- [10] J. Mellinas, "Average scores integration in official star rating scheme," *Journal of Hospitality and Tourism Technology*, vol. 10, no. 3, 2019.
- [11] T. Radojevic, N. Stanisic, and N. Stanic, "Inside the rating scores: a multilevel analysis of the factors influencing customer satisfaction in the hotel industry," *Cornell Hospitality Quarterly*, vol. 58, no. 2, pp. 134–164, 2017.
- [12] <http://insideairbnb.com/get-the-data.html>, 2020.
- [13] G. Santos, V. F. S. Mota, F. Benevenuto, and T. H. Silva, "Neutrality may matter: sentiment analysis in reviews of Airbnb, booking, and Couchsurfing in Brazil and USA," *Social Network Analysis and Mining, Issue*, vol. 1, 2020.
- [14] J. Hamari, M. Sjöklint, and A. Ukkonen, "The sharing economy: why people participate in collaborative consumption," *Journal of the Association for Information Science and Technology*, vol. 67, no. 9, 2015.
- [15] Dianne Dredge and S. Gyimóthy, "The collaborative economy and tourism: critical perspectives, questionable claims and silenced voices," *Tourism Recreation Research*, vol. 40, no. 3, 2015.
- [16] R. Botsman and R. Rogers, *Beyond Zipcar: Collaborative Consumption*, Harvard Business Publishing, Boston, MA, USA, 2010.
- [17] J. Batle and B. Joan, "Are locals ready to cross a new frontier in tourism? factors of experiential P2P orientation in tourism," *Current Issues in Tourism*, vol. 23, no. 10, 2020.
- [18] B. Hasan, K. Berezina, and C. Cobanoglu, "Comparing customer perceptions of hotel and peer-to-peer accommodation advantages and disadvantages," *International Journal of Contemporary Hospitality Management*, vol. 30, no. 2, 2018.
- [19] G. Zervas, D. Proserpio, and J. W. Byers, "The rise of the sharing economy: estimating the impact of airbnb on the hotel industry," *Journal of Marketing Research*, vol. 54, no. 5, pp. 687–705, 2017.
- [20] <https://www.Airbnb.com>.
- [21] J. Bakos, "Reducing buyer search costs: implications for electronic marketplaces," *Management Science*, vol. 43, no. 12, 1997.
- [22] *New York City hotel rooms are getting cheaper thanks to Airbnb*, Quartz, New York, NY, USA, 2015.
- [23] D. Guttentag, "Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector," *Current Issues in Tourism*, vol. 18, no. 12, pp. 1192–1217, 2013.
- [24] T. W. Gruen, Osmontbekov, T. Czaplewski, and A. J. Czaplewski, "eWOM: the impact of customer-to-customer online know-how exchange on customer value and loyalty," *Journal of Business Research*, vol. 59, no. 4, pp. 449–456, 2006.
- [25] E. Ellen, M. Anna, and S. Baloglu, "Effects of gender and expertise on consumers' motivation to read online hotel reviews," *Cornell Hospitality Quarterly*, vol. 52, no. 4, pp. 399–406, 2011.
- [26] J. Y. Chung and D. Buhalis, "(2008) Web 2.0: a study of online travel community. In," in *Information and Communication Technologies in Tourism 2008*, W. Connor, W Höpken, and U. Gretzel, Eds., Springer, Vienna, Austria, 2008.
- [27] A. Jin and O. Hengxuan Chi, "Zhe Ouyang categorizing peer-to-peer review site features and examining their impacts on room sales," *Journal of Hospitality Marketing & Management*, vol. 28, no. 7, 2019.
- [28] R. Leung, N. Au, J. Liu, and R. Law, "Do customers share the same perspective? a study on online OTAs ratings versus user ratings of Hong Kong hotels," *Journal of Vacation Marketing*, vol. 24, no. 2, pp. 103–117, 2018.
- [29] M. Marcello, "Effects of the Booking.com rating system: bringing hotel class into the picture," *Tourism Management*, vol. 66, pp. 47–52, 2018.
- [30] P. Phillips, N. Antonio, Ana de Almeida, and L. Nunes, "The influence of geographic and psychic distance on online hotel ratings," *Journal of Travel Research*, 2020.
- [31] E. Smironva, K. Kiatkwasin, S. K. Lee, J. Kim, and C.-H. Lee, "Self-selection and non-response biases in customers' hotel ratings - a comparison of online and offline ratings," *Current Issues in Tourism*, vol. 23, no. 10, pp. 1191–1204, 2020.
- [32] Z. Alrawadieh, "Determinants of hotel guests' satisfaction from the perspective of online hotel reviewers," *International Journal of Culture, Tourism and Hospitality Research*, vol. 13, no. 1, 2019.
- [33] V. Bogicevic, M. Bujisic, C. Cobanoglu, and A. H. Feinstein, "Gender and age preferences of hotel room design," *International Journal of Contemporary Hospitality Management*, vol. 30, no. 2, 2018.
- [34] X. Fiona and V. M. C. Yang, "LuXurY hotel loyalty—a comparison of Chinese Gen X and Y tourists to Macau," *International Journal of Contemporary Hospitality Management*, vol. 27, pp. 1685–1706, 2015.
- [35] A. Brochado, P. Rita, and S. . Moro, "Discovering patterns in online reviews of beijing and lisbon hostels," *Journal of China Tourism Research*, vol. 15, no. 2, 2019.
- [36] S. . Moro, "Guest satisfaction in East and West: evidence from online reviews of the influence of cultural origin in two major gambling cities, Las Vegas and Macau," *Tourism Recreation Research*, 2020.
- [37] Y. Yang, Z. Mao, and J. Tang, "Understanding guest satisfaction with urban hotel location," *Journal of Travel Research*, vol. 57, no. 2, pp. 243–259, 2018.
- [38] J. Nicolau and J. Pedro Mellinas, "Satisfaction measures with monetary and non-monetary components: hotel's overall scores," *International Journal of Hospitality Management*, 2020.
- [39] C.-F. Chiang, "Influences of price, service convenience, and social servescapes on post-purchase process of capsule hotels," *Asia Pacific Journal of Tourism Research*, vol. 23, no. 4, 2018.
- [40] E. Ert, A. Fleischer, and N. Magen, "Trust and reputation in the sharing economy: the role of personal photos in Airbnb," *Tourism Management*, vol. 55, pp. 62–73, 2016.
- [41] S. . Moro, J. Esmeraldo, and P. Ramos, "Evaluating a guest satisfaction model through data mining," *International Journal of Contemporary Hospitality Management*, 2019.
- [42] L.-M. Ang and K. P. Seng, "Big sensor data applications in urban environments," *Big Data Research*, vol. 4, no. C, 2016.
- [43] F. Mehdipour, H. Noori, and B. Javadi, "Energy-efficient big data analytics in datacenters," *Advances in Computers*, pp. 59–101, 2016.
- [44] X. Zheng, Z. Schwartz, H. John, and P. Gerdes, "Muzaffer Uysal what can big data and text analytics tell us about hotel guest experience and satisfaction?" *International Journal of Hospitality Management*, vol. 44, pp. 120–130, 2015.
- [45] A. J. Sanchez-Medina and E. C-Sanchez, "Using machine learning and big data for efficient forecasting of hotel booking cancellations," *International Journal of Hospitality Management*, vol. 89, p. 2020.
- [46] M. Ranjbari, Z. Esfandabadi, and S. Scagnelli, "A big data approach to map the service quality of short-stay

- accommodation sharing,” *International Journal of Contemporary Hospitality Management*, 2020.
- [47] L. Hang, Y. K. Tse, M. Zhang, and J. Ma, “analysing online reviews to investigate customer behaviour in the sharing economy the case of Airbnb,” *Information Technology & People*, 2020.
- [48] P. Kalehbasti, L. Nikolenko, and H. Rezaei, “Airbnb price prediction using machine learning and sentiment analysis,” 2019, <http://arxiv.org/abs/12665>.
- [49] S. Nakatani, “Langage detection librairy for java,” 2010, <http://code.google.com/p/language-detection>.
- [50] M. Honnibal, Y. Goldberg, and M. Johnson, “A non-monotonic arc-eager transition system for dependency parsing,” in *Proceedings of the Conference on Computational Natural Language Learning*, Beijing, China, 2013.
- [51] M. Honnibal and M. Johnson, “An improved non-monotonic transition system for dependency parsing,” in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Honolulu, HI, USA, 2015.
- [52] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, Springer Series in Statistics, Berlin, Germany, 2008.
- [53] S. Deep Arora and S. Mathur, “Hotel pricing at tourist destinations—a comparison across emerging and developed markets,” *Tourism Management Perspectives*, 2020.
- [54] Y. Chihab, Z. Bousbaa, M. Chihab, B. Omar, and S. Ziti, “Algo-trading strategy for intraweek foreign exchange speculation based on random forest and probit regression,” *Applied Computational Intelligence and Soft Computing*, 2019.
- [55] L. Zhu, M. Cheng, and I. A. Wong, “Determinants of peer-to-peer rental ranting scores: the case of Airbnb,” *International Journal of Contemporary Hospitality Management*, 2018.
- [56] D. Guttentag<sup>1</sup>, Stephen Smith<sup>2</sup>, L. Potwarka<sup>3</sup>, and M. Havitz, “Why tourists choose Airbnb: a Motivation-Based segmentation study,” *Journal of Travel Research*, vol. 57, no. 3, 2016.
- [57] H. Gu and C. Ryan, “Chinese clientele at Chinese hotels- Preferences and satisfaction,” *International Journal of Hospitality Management*, vol. 27, no. 3, pp. 337–345, 2008.
- [58] T. Lockyer, “Hotel cleanliness-how do guests view it? let us get specific. a New Zealand study,” *International Journal of Hospitality Management*, vol. 22, no. 3, pp. 297–305, 2003.
- [59] D. Wang and J. L. Nicolau, “Price determinants of sharing economy based accommodation rental: a study of listings from 33 cities on Airbnb.com,” *International Journal of Hospitality Management*, vol. 62, pp. 120–131, 2017.
- [60] A. Madalyn, “Airbnb superhosts’ talk in commercial homes,” *Annals of Tourism Research*, vol. 80, 2020.
- [61] S. Sun, J. Zheng, M. Schuckert, and R. Law, “Exploring the service quality of airbnb,” *Tourism Analysis*, vol. 24, no. 4, pp. 531–534, 2019.
- [62] Yu Meng, M. Cheng, Z. Yu, J. Tan, and Z. Li, “Investigating Airbnb listings’ amenities relative to hotels,” *Current Issues in Tourism*, 2020.
- [63] M. Cheng and X. Jin, “What do Airbnb users care about? An analysis of online review comments,” *International Journal of Hospitality Management*, vol. 76, pp. 58–70, 2018.
- [64] K.-W. Lee, H.-b. Kim, H.-S. Kim, and D.-S. Lee, “The determinants of factors in FIT guests’ perception of hotel location,” *Journal of Hospitality and Tourism Management*, vol. 17, no. 1, pp. 167–174, 2010.
- [65] G. Zhang, R. Wang, and M. Cheng, “Peer-to-peer accommodation experience: a Chinese cultural perspective,” *Tourism Management Perspectives*, vol. 33, 2020.
- [66] S. Moro, P. Ramos, J. Esmeraldo, and S. M. J. Jalali, “Can we trace back hotel online reviews’ characteristics using gamification features?” *International Journal of Information Management*, vol. 44, pp. 88–95, 2019.
- [67] N. Eriksson, “The Relative impact tf Wi-Fi service on young consumers’ hotel booking online,” *Journal of Hospitality & Tourism Research*, vol. 42, no. 7, 2018.