Summer 6-26-2019

How to Price Your House: Exploring Price Determinants of Online Accommodation Rental

Yating Zhang
Schoool of Economics and Management, Tongji University, Shanghai, 200000, China, zhangyt@tongji.edu.cn

Yu Zhang
Schoool of Economics and Management, Tongji University, Shanghai, 200000, China, 1650113@tongji.edu.cn

Shan Qian
Schoool of Economics and Management, Tongji University, Shanghai, 200000, China, qianshan98@126.com

How to Price Your House: Exploring Price Determinants of Online Accommodation Rental

Yating Zhang[*], *Yu Zhang, Shan Qian*
[1]Schoool of Economics and Management, Tongji University, Shanghai, 200000, China
[2]Schoool of Economics and Management, Tongji University, Shanghai, 200000, China
[3]Schoool of Economics and Management, Tongji University, Shanghai, 200000, China

Abstract: Tourism and hospitality have emerged as one of the pioneering sectors of sharing economy. However, homeowners who lack knowledge background are confused with pricing. Traditional hotel pricing and rental pricing methods may be not suitable for online accommodation rental. Therefore, CRISP-DM, the data analysis framework, is used to solve this problem. The price prediction model is established via the house data from the Airbnb.com. Finally, 33 determinants closely related to the price are found, and the most important 10 determinants are sorted. The study also finds several interesting rules: (1) the basic situation of housing is an important determinant, (2) online rental houses with more convenient transaction conditions have higher price, (3) providing more facilities and services can increase the price, (4) some determinants in traditional hotel pricing are not efficient in sharing houses. These findings can help the homeowners to understand customers and improve their own house and pricing.

Keywords: sharing economy, price determinants, data analysis framework, house, prediction model

1. INTRODUCTION

In recent years the sharing economy (SE) is seen as a disruptive innovation, made possible by new technologies - mainly the Internet, which is transforming economies and the way business is done. The SE is broadly characterized by peer-to-peer exchanges for renting goods or services utilizing Internet platforms. The SE platforms focus on peer-to-peer economic transactions by facilitating the sharing or renting of space, assets, and labor in real time. Industry practitioners speculatively estimate that sharing economy will potentially increase to 335 billion by 2025 compared with 15 billion in 2015[1]. With its rapid development, many scholars have studied SE from different perspectives. The SE literature includes three broad areas in general: (1) SE's business models and its impacts, (2) nature of SE, and (3) SE's sustainability development as well as two areas of foci in tourism and hospitality specifically: (1) SE's impacts on destinations and tourism services (2) SE's impacts on tourists .We can see that tourism and hospitality are important developing directions for SE.

Actually, since the start of SE, tourism and hospitality have emerged as one of the pioneering sectors for its growth as SE allows for tourists and residents to share diverse contents — their homes, cars, four course meals, and expert local knowledge (e.g. locals being tour guides), and Airbnb becomes a popular example of SE. Merely providing online accommodation rental information, Airbnb is a typical Commission-Based Platform. The Commission-Based Platforms are dominated by (at least) triadic relationships amongst providers, intermediaries and consumers with a utility-bound revenue stream. These business models enable their customers to switch between provider and consumer roles by creating and delivering the value proposition. The platform mainly takes commissions for successful matching and executing trade [2] and let providers and consumers themselves to decide the rest.

[*] Corresponding author. E-mail addresses: [a] zhangyt@tongji.edu.cn, [b] 1650113@tongji.edu.cn, [c] qianshan98@126.com

However, these online accommodation rental platforms also bring some difficult problems such as pricing the house to providers while giving great freedom to providers and consumers. As it is easy to start a tourism business at a relatively low start-up cost, more and more providers with no related knowledge background will participate in house-sharing and face this problem. For traditional accommodation industry, pricing is widely acknowledged to be one of the most critical factors determining the long-term success of the accommodation industry. Many studies have been conducted on pricing strategies in the hospitality industry from both the demand side [3, 4, 5] and the supply side [6, 7]. However, only a few researchers have investigated the factors determining the price of online accommodation rental.

We decide to use CRISP-DM, a data analysis framework, to find meaningful determinants of online accommodation rental price and interesting mode in the house sharing. Through preprocessing the data, building prediction model, we get the determinants of price and sort them by the importance. Not from the demand side neither the supply side, we choose real data from Airbnb.com and let data tell us the pricing rules under such a new sharing age.

2. LITERATURE REVIEW

In different product markets, there are various methods to fix a price. Though, there is little theory to help us fix online rental house's price. Online accommodation rental has dual attributes of hotel and housing. Therefore, we can get help from the existing theories on hotel room pricing and housing pricing. These factors can be divided into three aspects: external factors, the property of the house itself and customers' preferences.

External factors refer to uncontrollable objective factors such as economy development condition and tourism seasonality. Hotel price is closely related to the seasonality of tourism [8]. Hotel accessibility also affects the price level. Low market accessibility (high flight cost) leads to lower hotel prices. The effect of sea view to room rates exhibited a significant spatial variability, indicating that local natural and/or tourism resources may have a substantial role in aesthetic values [9].

Housing attributes refer to the size, grade, service types and some other attributes that determine the quality of housing. These factors are the most basic information to consider when pricing a house. Some scholars have revealed how the properties of housing sites, such as swimming pool, spa and free breakfast, affect house prices, and found the spatial relationship between house prices and hotel attributes, such as size, age, grade and service quality.

The third kind of factor comes from the demand side, customers, whose special preferences also affect pricing. Customers also appreciate some types of hotels, such as boutique, quaint or trendy hotels, but view others negatively, such as family-friendly or business hotels [10]. High-star hotels and chain hotels are also favored by consumers.

Some factors related to online platforms and trading parties emerge to influence prices and transaction rates. Under the mode of shared economy, people can be both providers and consumers, and play an increasingly important role in transactions. The degree of discrimination in the rental market has been significantly reduced, but some scholars have found that Airbnb's current design choices help promote discrimination [11]. Some scholars also found that customers will consider the appearance of the owner when selecting a house online. We can see from the photos that the more trustworthy the owner is, the higher the listing price is, the greater the possibility of being selected. The reputation of the owner, conveyed through his online comment score, has no effect on the listed price or the possibility of consumer booking [12]. But it's also found that there is a significant relationship between online consumer reviews and hotel performance [13].

Remove customers' preferences which are dynamic and intersect with other factors, and we can summarize the factors affecting the price of online accommodation rental into three categories: external uncontrollable factors, self-attributes and platform related factors. The three classes of factors come from the summary of previous literatures. These studies were mainly developed from one city’s data and with limited variables and no one have made a comprehensive survey of these factors. It’s necessary to use a larger dataset with abundant variables to explore the price determinants of online accommodation rental. So, we started our research with a large

3. DATA ANALYSIS

3.1 Variables and data

Airbnb, an online market for accommodation, has reached a large global scale. We selected the housing data provided by the Kaggle data platform. The dataset includes detailed house information of six cities in the United States and has a total of 74.1k records, each record has 29 variables. The introduce of 29 variables is shown in Table 1.

Table 1. The variable list

Variable Name	Data type	Definition
Id	Numeric	Serial number of house.
Log_price	Numeric	Listed price per night on Airbnb.com.
Property_type	String	The house's type: apartment, condominium, townhouse and so on.
Room_type	String	The room's type: entire room, private room and shared room.
Facilities and Services	String	Combined provided facilities and services.
Accommodates	Numeric	The number of people the house can hold.
Bathrooms	Numeric	The number of bathrooms the house has.
Bed_type	String	The type of the bed: real bed, futon, pull-out sofa and so on.
Cancellation_policy	String	How easy to cancel an order.
Cleaning_fee	Boolean	Whether to charge a cleaning fee or not.
City	String	The city where the house is located.
Description	String	Description of the house.
First_review	String	The date of the first review.
Host_has_profile_pic	String	Whether to provide house pictures or not.
Host_identity_verified	String	Whether the house's identity is verified.
Host_response_rate	String	The rate of the households response.
Host_since	Date	The data of the house was put on the platform.
Instant_bookable	String	Whether can book the house instantly.
Last_review	String	The date of the lase review.
Latitude	Numeric	The house's latitude.
Longitude	Numeric	The house's longitude.
Name	String	The house's name.
Neighbourhood	String	The house's neighbourhood.
Number_of_reviews	Numeric	The number of the reviews.
Rating	String	The house's rating given by customers.
Zipcode	Numeric	The house's zipcode.
Bedrooms	Numeric	The number of the bedrooms the house has.
Beds	Numeric	The number of the beds the house has.

3.2 Data preprocessing

- Correcting the rating

The two variable "rating" and "number of commentators" are closely related. True and reliable ratings should not only be the average of multiple ratings, but also be adjusted according to the number of commentators. The more commentators there are, the more reliable the ratings will be. The ranking method of some ranking websites can be used to correct the rating. The method of Bayesian star rating is to constantly revise the prior probability by introducing the latest observation results, so as to obtain the final posterior probability. According to the calculation method of ranking websites such as Douban Movie Score, assuming that the initial rating of a house is the average score of all houses at present, the following revised formula can be deduced.

$$\text{Corrected rating} = \frac{\text{Average rating} \times \text{ total reviews number} + \text{this house}'\text{s rating} \times \text{ this house}'\text{s reviews number}}{\text{number of total reviews} + \text{number of this house}'\text{s reviews}} \quad (1)$$

- Split variables

The variable "Facilities and Service" is a description set of 77 types of facilities or services, such as {Air Conditioning, Tableware, Dryer, Cable TV}. In order to find out which types of services or facilities have a significant impact on price, this variable needs to be split up to 77 facilities and service variables in1 or 0.

- Data cleaning

Several variables have some missing variables and solutions are given in Table 2. Additionally, we find three records have abnormal value and delete them directly for data accuracy.

Table 2. Missing variables and resolution

Variable Name	Missing rate	Resolution
Neighbourhood	9.273%	The variable city already contains the information of neighbourbood. Don't use this variable.
Zipcode	1.304%	
Bathrooms	0.270%	
Host_has_profile_pic	0.254%	
Host_identity_verified	0.254%	The missing rate is acceptable. Delete the record that has missing value.
Host_since	0.254%	
Bedrooms	0.177%	
Beds	0.123%	

3.3 Data exploration

After data cleaning, there are 99 variables and 55584 records in current data set. Variables can be classified as Table 3.

Table 3. Variable classification

Variable classification	Count	Detailed information
Self-attibutes	94	17 variables related to house's basic information and 77 variables of facilities and services.
Platform related factors	5	Variables related to the platform and transaction.
Total	99	

- Dependent variable

The descriptive statistics of log_price are shown in Table 4, and the frequency histogram is shown in Figure 1. It can be seen that the price distribution has a strong normality and is suitable for the prediction model.

Table 4. Descriptive statistics of log_price

	Mean	Variance	Minimum	Lower quartile	Median	Upper quartile	Maximum
Log_price	4.7535	0.6670	0	4.3175	4.7005	5.1648	7.6004

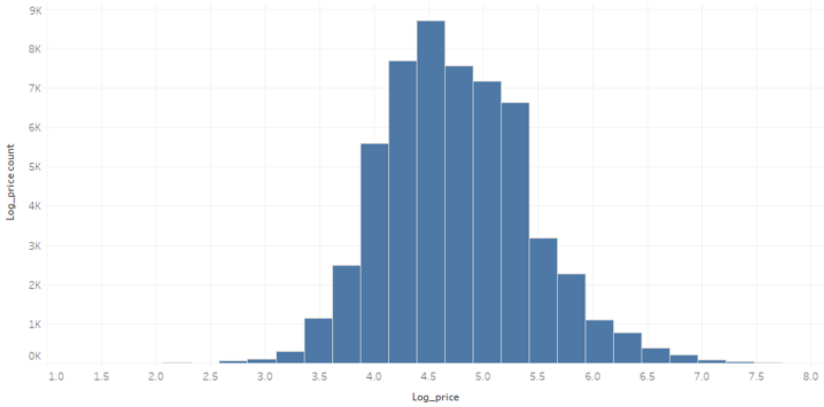

Figure 1. Frequency histogram of log_price

- Independent variable

Firstly, from descriptive statistics, we can see that some variables have strong correlation. When modeling, we can only choose one variable from them or integrate them into a variable. For example, the number of accommodation, the number of bedrooms and the number of beds all reflect the number of residents. Most houses only provide one bedroom and one bed, which can accommodate two people, perhaps because the customers in these big cities are mainly for business travel rather than traveling with friends and family.

Second, some fields are mutually exclusive and can only use one variable to model. For example, "Cable TV" and "Radio TV" both reflect whether the house provides the service of TV. We can combine the two variables into one variable.

In addition, through descriptive statistics, we find that some variables are not divergent enough and include little information, so we can consider eliminating them when using prediction model. For example, the number of bathrooms and the bed type are relatively concentrated, most houses only have one bathroom, the type of bed is ordinary bed. Among the 77 facility service variables, there are also values that are not divergent enough and are concentrated in 1 or 0. Such variables are directly excluded and not used.

We calculate the spearman coefficients of the correlation between all variables and housing price. After sorting, we select the variables whose absolute value of the spearman coefficients is above 0.1 for further analysis.

3.4 Feature extracting

We exclude variables which are not divergent enough and have low spearman coefficients. Finally, 33 variables are extracted for modeling, including 29 self-attributes combined with 20 facilities and services variables and 9 basic self-attributes and 4 platform related factors. The variables are bedroom door lock, pets, TV, child-friendly, breakfast, air conditioning, smoke alarm, necessities, hangers, carbon monoxide detectors, elevators, shampoo, independent access, personal computer use space, self-help check-in, irons, hair dryers, washing machines, dryers, cable TV, city, registration time, postal code, attribute type, source type, bathroom number, bedroom number, immediate booking, whether there are house pictures, type of bed, cancellation policy, number of beds and accommodation number.

4. MODELING AND RESULTS

4.1 Model selection

Before modeling, the data set is divided into test set and training set according to the ratio of 3:7. Second, the root mean square error (RMSE) is selected as the regression evaluation index. Third, we use one-hot method to code the value of all variables for a more accurate model. We select nine commonly used regression algorithms, adopt default parameters, and use 5-fold cross-validation to get the RMSE value to evaluate each model.

Three models with RMSE less than 0.36 were selected for analysis: random forest regression, gradient upgrade regression and extreme random forest regression. Random forest regression can give the importance of each variable, and the top ten variable with high importance were obtained, as shown in Table 6.

Table 6. Top 10 variable with high mportance

Variable Name	Importance
Bed_type	0.4066
Property_type	0.1809
Bedrooms	0.1243
Instant_bookable	0.0720
Beds	0.0390
Air conditioner	0.0321
Room_type	0.0148
Host_has_profile_pic	0.0115
Breakfast	0.0092
Carbon dioxide detector	0.0068

4.2 Feature construction

After preliminary modeling, we use gradient lifting regression model (the result is 0.3392021064382331) to test the constructed features. We integrate some variables which have strong correlation and increase high-correlation variables' proportion in the model via squaring their value. However, the accuracy of the model doesn't be increased and the error reduced. It show that the variables already reflect the changes of housing price.

4.3 Model fusion

After Optimizing parameters including iteration times, maximum depth of decision tree and number of samples required for internal node re-partitioning, we fused the three models and adjust the specific proportion for a better result in Table 7. In summary, the final result of our prediction is 0.3298.

Table 7. Model fusion result

Random Forest Regression	Gradient Boosting Regression	Extra-Trees Regression	Result
0.7	0.15	0.15	0.3447
0.6	0.3	0.1	0.3407
0.26	0.64	0.1	0.3306
0.15	0.65	0.2	0.3303
0.2	0.7	0.1	0.3303
0.3	0.6	0.1	0.3302
0.25	0.65	0.1	0.3299
0.225	0.65	0.125	0.3298

5. DISCUSSION

5.1 Result

This study explores the determinants of online accommodation rental prices and sorts them by importance. Based on 99 variables, the online accommodation rental price is analyzed. Through preprocessing the data, modeling prediction model, 33 variables which can predict the price well are obtained: 29 self-attributes combined with 20 facilities and services variables and 9 basic self-attributes and 4 platform related factors. Random forest model is used to rank the importance of variables, and 10 variables that have significant importance on price prediction are obtained.

We find that the basic situation of housing is an important price determinant. Different from hotels' standardized facilities and services, online rental houses has higher richness and diversity. However, only rental houses with comfortable beds, more bedrooms and more beds tend to have higher prices, indicating that customers always put comfort and sleeping environment in the first place when selecting a rental house. No one likes sleeping on a couch or a futon for a long time. Therefore, the type of bed has the most significant impact on the price.

Secondly, this study show that online rental houses with more convenient transaction conditions have higher prices. Immediate booking and providing-pictures can increase the price. Immediate booking keeps with people's fast-paced lifestyle and providing real-time booking service can shorten the distance between online and offline. Displaying real and reliable housing pictures on the website can provide more information for consumers to make decisions.

Thirdly, we find that providing some facilities and services can increase the price of houses. Short rentals with air conditioning, breakfast and carbon dioxide detectors are more expensive. More facilities and services will undoubtedly be popular with consumers. In addition, it also shows that safety is a new factor that consumers attach increasing importance to. For example, carbon dioxide detectors can guarantee consumers' safety.

Another finding is that the price determinants of online accommodation rental are different from those of traditional hotels, but similar to the price determinants of sale housing. Hotel star has significant impact on room price of traditional hotels, but rating, the similar variable in online accommodation rental, has little relevance with house price. It maybe because these two variables have different sources: the hotel star is assessed from many aspects and online rating only comes from the customers.

These finding is not only more detailed and specific than previous studies [10, 14], but also has practical significance. It's recommended that households make a price by assessing house self-attributes and platform related factors. Households can improve their house by choosing a comfortable bed and providing more facilities within budget, and attract more customers by setting immediate booking mode and unloading real and beautiful house pictures.

5.2 Summary

Using data sets of Airbnb from six cities in the United States, this study explores the price determinants of online accommodation rental. These findings provide a comprehensive understanding of the determinants of product prices under SE condition, and provide a perspective for stakeholders such as homeowners to improve house conditions and increase profits. In addition, this study also provide suggestions to online short-rent platforms to provide homeowners with pricing services based on current price determinants.

Nevertheless, we acknowledge an important limitation of this study. Firstly, although this study considers the influence of geographical factors, the data used in this study are only rent data of Boston, San Francisco, Chicago, Washington, New York and Los Angeles, which have limitations. Secondly, the price difference caused by social or psychological factors is not considered. Therefore, it is very important to conduct qualitative research to explore the price decision under the influence of social and psychological factors. Finally, the exploration of the interaction between variables is not deep enough, so follow-up studies can explore the interaction between different variables.

REFERENCES

[1] Masiero L, Nicolau J L, Law R. (2015). A demand-driven analysis of tourist accommodation price: a quantile regression of room bookings. International Journal of Hospitality Management, 50:1–8.

[2] Ioanna C, Marton A, Tuunainen V K. (2017). Four Models of Sharing Economy Platforms. MIS Quarterly Executive, 16(4): 231–51.

[3] Becerra M, Santaló J, Silva R. (2013). Being better vs. being different: differentiation, competition: and pricing strategies in the Spanish hotel industry. Tourism Management, 34:71–79.

[4] Chen C F, Rothschild R. (2010). An application of hedonic pricing analysis to the case of hotel rooms in Taipei. Tourism Economics, 16(3):685–694.

[5] Ritter M, Schanz H. (2019). The sharing economy: A comprehensive business model framework. Journal of Cleaner Production, 213:320-331.

[6] Gutt D, Herrmann P. (2015). Sharing Means Caring? Hosts' Price Reaction to Rating Visibility. Association for Information Systems, 5-29.

[7] Lee C G. (2011). The determinants of hotel room rates: another visit with Singapore’s data. International Journal of Hospitality Management, 30 (3):756–758.

[8] Ye Q, Law R, Gu B. (2009). The impact of online user reviews on hotel room sales. International Journal of Hospitality Management, 28:180-182.

[9] Latinopoulos D. (2018). Using a spatial hedonic analysis to evaluate the effect of sea view on hotel prices. Tourism Management, 65:87-99.

[10] Hung W T, Shang J K, Wang F C. (2010). Pricing determinants in the hotel industry: quantile regression analysis. International Journal of Hospitality Management, 29(3):378–384.

[11] Edelman B, Luca M, Svirsky D. (2017) Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. American Economic Journal-Applied Economics, 9(2):1-22.

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

[13] Cheng M M. (2016). Sharing economy: A review and agenda for future research. International Journal of Hospitality Management, 57:60-70.

[14] Wang D, Nicolau J L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. International Journal of Hospitality Management, 62:120-131.

[15] Heo C Y, Hyun S S. (2015). Do luxury room amenities affect guests’ willingness to pay? International Journal of Hospitality Management, 46:161–168.

[16] Kakar V, Franco J, Voelz J, Wu J. (2016). Effects of Host Race Information on Airbnb Listing Prices in San Francisco. Munich Personal RePEc Archive.

[17] Soler I P, Gemar G, Correia M B, Serra F. (2019).Algarve hotel price determinants: A hedonic pricing model. Tourism Management, 70:311-321.

[18] Saló A, Garriga A, Rigall-I-Torrent R, Vila M, Fluvià M. (2014). Do implicit prices for hotels and second homes show differences in tourists’ valuation for public attributes for each type of accommodation facility? International Journal of Hospitality Management, 36:120–129.

[19] Teubner T, Hawlitchek F, Dann D. (2017). Price Determinants on Airbnb: How Reputation Pays off in The Sharing Economy. Journal of Self-Governance and Management Economics, 5(4):53-80.

[20] Sutherland W, Jarrahi M H. (2018). The sharing economy and digital platforms: A review and research agenda. International Journal of Information Management, 43:328-341.

[21] Wang X R, Sun J X, Wen H Z. (2019). Tourism seasonality, online user rating and hotel price: A quantitative approach based on the hedonic price model. International Journal of Hospitality Management, 79:140-147.