

Research on the Influencing Factors of Hotel Booking for Traveling

Wenyi Chen^{1,*}

¹ North China University of Technology, Beijing, 10009, China

*Corresponding author:
22190010312@mail.ncut.edu.cn

Abstract:

This article aims to delve into and identify the key factors that influence consumers' consideration when choosing a hotel. To this end, this study carefully designed a variety of statistical and analytical methods such as linear regression, correlation analysis and random forest model, systematically analyzed a number of potential influencing factors, including hotel rating, rating quantity, price, whether parking service is provided, breakfast service and wireless network, in order to reveal the correlation between them and hotel selection. The results show that there is no significant linear relationship between price, hotel ratings and Wi-Fi availability and hotel selection. At the same time, through the importance test of the random forest model, this paper found that the number of ratings, the provision of parking services and the quality of breakfast services did not reach a significant level. Overall, these findings provide a new perspective on the understanding of the volatility in consumer demand for hotels.

Keywords: Influencing factors; hotel; traveling.

1. Introduction

In recent decades, the global tourism market has witnessed continuous expansion and a profound diversification, solidifying its position as one of the fastest-growing economic sectors worldwide. The number of international tourist arrivals has steadily climbed from a mere 25 million in 1950 to an astonishing total of 1.186 billion in 2015. This upward trend is anticipated to persist, with projections from the World Tourism Organization (UNWTO) indicating that the number of international tourist arrivals will surge to 1.8 billion by the year 2030 [1]. As a form of human mobility, tourism plays a significant role in contributing to the ongoing process of making the planet feel increasingly smaller and more inter-

connected [2].

In the process of travel, choose a suitable hotel, will enhance the whole journey experience, is the travel more relaxed and happy. When choosing a suitable hotel, many factors should be considered, including location, facilities, price and so on. The rapid development of tourism has led to the prosperity of the hotel industry, and various types of hotels have emerged in the market to meet the needs of different tourists. From economy hotels to high-end luxury hotels, from chain hotels to feature lodgings, the hotel market is becoming more and more diversified. In order to explore the difference of different types of travelers in hotel selection, Wang et al put forward a method of image fuzzy TODIM to analyze the results

of hotel selection [3].

Reviews are one of the criteria for choosing a hotel. Technology, leading to access to external information and communication tools is now the network. Consumers' trust in and reliance upon online comments continue to escalate, making online comments a pivotal channel for consumers to acquire comprehensive product information. This channel plays an indispensable role in shaping consumers' attitudes towards products and influencing their purchasing decisions [4]. Prior to purchasing a product or service, customers typically hold specific performance expectations for what they intend to acquire. Subsequently, they compare their actual perception of the product or service with their post-purchase expectations. In this comparison, the absence of uncertainty often leads to customer satisfaction, while positive deviations can further enhance satisfaction, whereas negative deviations result in customer dissatisfaction [5]. Consequently, the more favorable and insightful the reviews are, the greater attention they garner from potential consumers.

The majority of individuals aim to visit as numerous attractions as feasible, and meticulously plan their journeys to ensure the most effective and time-optimal routes. The selection of a hotel's location is also a crucial criterion in the decision-making process. Toledo and others have represented this problem visually through a graph, where each node signifies either a hotel or an attraction. Each attraction node carries a score, reflecting the traveler's inclination to visit that particular attraction. Solving this problem entails identifying the optimal path within the graph, one that maximizes the number of attractions visited during the journey's available timeframe, aligning with the traveler's preferences. This particular challenge, termed the Orienteering Problem with Hotel Selection (OPHS), is recognized as an NP-hard problem. The OPHS falls under a broader category of issues known as the Orienteering Problem with Intermediate Facilities [6]. In the context of OPHS, the objective is to choose hotels and vertices that, within a reasonable timeframe, yield the highest total score, while ensuring that the travel route among the selected points adheres to the time constraints [7].

The quality of hotel facilities is an important index for hotel selection. For example, smart hotel. The essence of an intelligent hotel lies in the seamless integration of communication, computer, and control technologies, all aimed at delivering an unparalleled service experience while

reducing labour and energy costs. These next-generation hotels prioritize safety, energy efficiency, and comfort [8]. These facilities serve as the bedrock for the hotel's intelligent services, encompassing features such as key-less entry, self-service kiosks for check-in and check-out, advanced navigation systems, biometrics, facial recognition technology for access control, and contactless payment options [9]. These facilities improve hotels security, improved customer satisfaction, but also to improve the hotel's visibility. The quality of a hotel has a direct impact on customer satisfaction with the hotel, and determines the competitiveness and sustainable development of the hotel [10].

This paper chooses Beijing Hotel as the sample for analysis. The author will examine the impact of 6 factors, including Hotel rating, number of ratings, hotel price, whether parking is available, whether breakfast is available, whether Wi-Fi is available.

2. Methods

2.1 Data Source and Description

The data of this paper comes from Ctrip website. This paper collected data on all the hotels in Beijing. There are more than 14,000 hotels in Beijing, which is a huge amount of data. So, this paper chose 500 hotels as my sample data.

2.2 Indicator Selection and Description

The dataset consisted of 14,000 sets, which was too much data, so this paper ended up choosing 500 of them. Hotel rates vary greatly depending on peak season, holidays or special events. At the same time, the promotion of culture and tourism will also affect the number of tourists, and if the competition in the region where the hotel is located is tense, the hotel will also affect the change of hotel prices in order to attract more customers. Due to holidays, special activities and other factors, the price of the hotel changes too much, so this paper chooses to delete these variables, and only selects the prices of various hotels without any special activities on ordinary days. Finally, the data variables this paper selected included hotel rating, number of ratings, hotel price, whether parking is available, whether breakfast is available, and whether Wi-Fi is available (table 1).

Table 1. List of variables

Variable	Meaning	Value range
Mark	Consider the quality of all aspects of the hotel service	4.0-5.0
Comment number	The number of customer reviews reflects the popularity of the hotel	0-9,377
Price	Hotel rates for queen bed rooms	¥108-¥1,880
Parking lot	Parking is available at the hotel	Not have
breakfast	Does the hotel have breakfast	Not have
Wireless network	Wireless network is available at the hotel	Not have

2.3 Method Introduction

Correlation analysis is a statistical method used to measure the strength and direction of a linear relationship between two or more variables. In hotel selection, this paper collects data on multiple characteristics of the hotel, such as rating, number of ratings, price, whether there is parking, whether there is breakfast, whether there is wireless Internet, and calculate the correlation between these characteristics and the target variable.

Linear regression model method is a statistical method used to describe the linear relationship between independent variables and dependent variables. The method selects a suitable model form, and uses optimization algorithms such as least square method to estimate model parameters, So as to minimize the error between the two value. Linear regression model is easy to understand and easy to calculate. It is widely used in the fields of prediction analysis, trend prediction, feature selection and so on. At the same time, the feature scaling and regularization techniques can be used to expand and optimize the linear regression model to improve the performance and generalization ability of the model.

The Random Forest method is an ensemble learning algorithm. It builds multiple models using decision trees. It combines the results of these models to improve prediction accuracy and stability. This method creates training subsets from the original data by self-sampling. Each subset trains one decision tree. For regression problems, it combines the results by averaging.

3. Results and Discussion

3.1 Descriptive Analysis

The data set contains basic information about the hotel, including ratings, number of reviews, price, whether park-

ing is available, whether breakfast is served, and whether Wi-Fi is available. These variables may influence consumers' choice behavior to a certain extent, so understanding the interrelationship between these variables can help better analyze and predict the attractiveness or profitability of a hotel (Table 2).

As depicted in the correlation heat map, there exists a notable correlation (0.42) between hotel ratings and the availability of wireless Internet (Figure 1). This suggests that hotels offering wireless Internet services cater more effectively to the demands of contemporary consumers, potentially enhancing their overall satisfaction. In today's digital era, where staying connected is paramount, the provision of wireless Internet has become a crucial aspect of the hotel experience.

Furthermore, a positive correlation (0.22) is observed between ratings and the availability of breakfast services. This could be attributed to the fact that breakfast services are highly appealing to guests, and hotels that offer comprehensive breakfast options are more likely to receive favorable ratings. Breakfast plays a significant role in shaping the overall guest experience and can significantly impact a hotel's reputation.

Additionally, a slight positive correlation (0.15) is noted between the number of reviews and the ratings received. Although this correlation is relatively weak, hotels with a higher number of reviews tend to enjoy greater visibility and foot traffic, which may indirectly influence consumer ratings. The number of reviews serves as an indicator of a hotel's popularity and can play a role in shaping consumer perceptions.

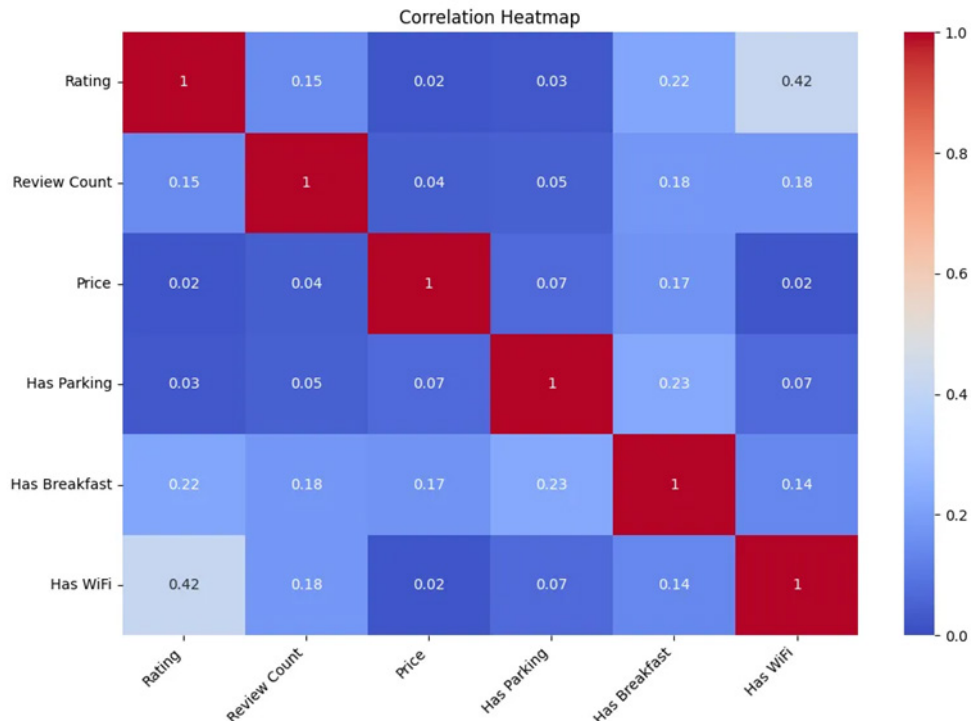


Fig. 1 Correlation heart map

Table 2. Correlation coefficient between major variables

Variable	Mark	Comment number	Price	Parking lot	Breakfast	Wireless network
Mark	1	0.15	0.02	0.03	0.22	0.42
Comments number	0.15	1	0.04	0.05	0.18	0.18
Price	0.02	0.04	1	0.07	0.17	0.02
Parking lot	0.03	0.05	0.07	1	0.23	0.07
Breakfast	0.22	0.18	0.17	0.23	1	0.14
Wireless network	0.42	0.18	0.02	0.07	0.14	1

3.2 Model Results

The model uses the following formula: Multiple linear regression model equation: $y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_n \cdot x_n + \epsilon$. Random forest

model prediction formula: $F(X) = \frac{1}{M} \cdot \sum_{m=1}^M n_m(x)$.

The linear regression model was employed to predict the data, yielding a rather disappointing R-squared value of -0.0688. This figure clearly indicates that the model has a poor fitting effect and is unable to effectively capture and explain the variations within the dependent variable. The mean absolute error (MAE) for this model stands at 0.5725, signifying that on average, the predicted values deviate from the actual values by 0.5725 units. This un-

derscores the inadequacy of the linear regression model in handling nonlinear relationships within the dataset.

In stark contrast, the random forest regression model exhibits an even more concerning R-squared value of -0.8980, which further emphasizes its failure to grasp the intricate internal relationships within the data. The MAE for this model, at 0.6747, is slightly elevated compared to the linear regression model, highlighting a larger prediction error. These results collectively suggest that neither model is well-suited for the given dataset, and alternative approaches or models may need to be explored to achieve better predictive accuracy.

4. Conclusion

Random forest method is an ensemble learning algorithm, which builds multiple models based on decision trees and integrates the prediction results of these models to improve the accuracy and stability of prediction. For the regression problem, the integration is carried out by means of average value. The dataset reveals a correlation between hotel ratings and some key services, particularly in the areas of Wi-Fi and breakfast availability. The correlation between the rating and wireless network is 0.42, indicating that wireless network plays a significant role in improving customer satisfaction, and modern consumers may prefer to choose hotels that offer wireless network. The correlation between the score and whether there is breakfast is 0.22, although the correlation is low, it also reflects that the breakfast service affects the customer's score to some extent. The score has a slight correlation with the number of reviews, with a correlation coefficient of 0.15, suggesting that the number of reviews is not a decisive factor. Hotels that attract more reviews might gain higher scores due to their greater popularity or customer flow, thereby drawing more attention.

References

- [1] Chang X U, Li Li W. Influencing Factors of Hotel Internship Satisfaction of Tourism Management Undergraduate. *Journal of Tonghua Normal University*, 2012.
- [2] Alan A Lew. Why travel - travel, tourism, and global consciousness. *An International of Tourism Space, Place and Environment*, 2018, 20.
- [3] Wang L, et al. The differences in hotel selection among various types of travellers: A comparative analysis with a useful bounded rationality behavioural decision support model. *Tourism Manage*, 2020, 76.
- [4] Xudong Lin, Ting Bu, Qiuting He, Yuting Liang. A study on the impact of positive and negative comments on consumers' perceived usefulness. *Internet of Things Technology*, 2019.
- [5] Mingyang Li, Yumei Ma, Pingping Cao. Revealing Customer Satisfaction With Hotel Through Multi-Site Online Reviews: A Method Based on The Evidence Theory. *Australian Journal of Hospitality Management*, 2020.
- [6] Divsalar A, Vansteenwegen P, Cattrysse D. A variable neighborhood search method for the orienteering problem with hotel selection. *Int. J. Prod. Econ*, 2013, 145: 150-160.
- [7] Alan Toledo, Maria -Cristina Riff, Bertrand Neveu. A Hyper-Heuristic for the Orienteering problem With Hotel Selection. *Australian Journal of Hospitality Management*, 2020.
- [8] Guan S Y, Liu X H. Design and application of smart hotel system based on the six-domain model of the internet of things. *Internet of Things Technology*, 2022, 3: 95-99.
- [9] Gurucharan Singh, Sandeep Raheja, Rakesh Sharma. Elevating Hospitality with Smart Hotel Technologies: A Guest-Centric Perspective. *Internet of Things Technology*, 2023
- [10] Huaju Xue, Xiumei Chen. Hotel Quality Evaluation and Promotion Strategy Based on IPA Analysis: An Empirical Case of Sofitel Hotel in Xining City. *Australian Journal of Hospitality Management*, 2022.