Airbnb Optimization of Pricing Strategy: A Study on the Static Pricing Model of Housing Resources Based on the Characteristic Price Theory

Hangxi Liu[1][*][†], Ruolin Zhang[2][†], Peiyuan Song[3][†], Zixuan Bu[4][†]

[1]*School of Transportation, Fujian University of Technology, Fuzhou, China*
[2]*Dietrich School of Arts and Sciences, University of Pittsburgh, Pittsburgh, USA*
[3]*Rutgers Business School, Rutgers University, New Brunswick, USA*
[4]*BASIS International School, Hangzhou, China*
*Corresponding Author. Email: 2695968603@qq.com
[†]*These authors contributed equally to this work and should be considered as co-first author.*

Abstract. With the surge of tourism and the rapid development of the global homestay market, Airbnb has achieved a significant leap forward in development, among which pricing strategy has become the focus of research. Based on the characteristic price theory, this paper constructs an accurate static pricing model through the linear model algorithm and makes an objective evaluation. The study found that the characteristics of house specifications, cancellation policies, whether to provide cleaning fees and so on have an important impact on the Airbnb housing price. At the same time, the evaluation scores and booking situation also have a certain influence on pricing by integrating a large number of characteristic data, establishing the housing static pricing model, providing a powerful tool for homestay facility pricing, for existing rental records through the evaluation data for pricing maintenance of special research, provides the reference for people host optimization pricing strategy.

Keywords: static pricing, multi-feature, linear model algorithm

1. Introduction

1.1. Research background

The tourism market is booming, and the consumer demand for accommodation has changed significantly. Modern travelers prefer to travel briefly to experience the local culture, and B & B is a new favorite with its unique design and personalized service. The landlord's aesthetic and personality differences give the b & B characteristics, bringing a unique living experience to the passengers. Airbnb With the concept of a sharing economy, it has achieved remarkable development in the global and Chinese markets. In recent years, Airbnb has focused on data-driven product pricing, predicting housing prices by model-fitting eigenvalues, and optimizing the model to help landlords develop efficient rents.

1.2. Questions are raised

In recent years, the market acceptance of Airbnb and homestay products has improved, but product value recognition is facing challenges. In the context of Internet development, competition in the same industry has an impact on Airbnb and the whole industry. Extensive management needs to shift to refinement to meet the economic needs of enterprises, users and landlords. Homestay pricing has become a research hotspot, and the research and application level should be improved.

This paper intends to use the Airbnb housing source data to explore the relationship between characteristics and prices, integrate the data, and use the linear model algorithm to determine the reasonable rent for the housing source. Management benefits and economic value are evaluated after optimization.

1.3. Study significance

The pricing problem of homestay is complex and multi-dimensional. The housing supply on the Airbnb platform has its own characteristics, including geographical location, decoration, facilities, services and landlord personality, which all affect the pricing. Different from traditional pricing strategies, Airbnb pricing needs to accurately reflect the value of each house set. At present, data analysis and computer simulation attract much attention in complex commodity pricing. The price of homestay is affected by regularity (such as the number of residents, number of bedrooms, and cancellation policy) and irregularities (such as emergencies), and the pricing problem is complex. It is of great value to predict the multi-characteristic products through the linear model algorithm to understand the determinants of homestay rental price. Traditional pricing methods, such as cost-oriented, demand-oriented and competitive-oriented, rely on experience or artificial analysis, which are not applicable to homestay. Empirical pricing and human analysis are subjective, and it is difficult to make a completely rational judgment. This study is applied to housing rental pricing, which is expected to improve the economy and rationality of the pricing strategy, meet the needs of all parties, maximize the benefits, and have practical application value.

1.4. Study content

This research takes the pricing problem of Airbnb homestay as the research object.

In order to study the pricing method of homestay, this study integrates multi-feature data, uses the data simulation algorithm to build a static pricing model, evaluates the prediction accuracy and other indicators, and analyzes the management benefit and economic value. The linear model algorithm was used for modeling training to compare the prediction results.

The first is to collect and learn the relevant research literature on the influencing factors of housing rent price, understand the current research results and the latest research trends at home and abroad, on this basis to determine the research direction of this paper, and formulate the experimental method.

Secondly, the research in this paper will use the empirical research method, according to the existing research theory and practical needs, using the actual housing data, through the big data processing method, purposefully follow the steps, record and analyze the correlation between the housing characteristics and the housing price.

Finally, the data simulation algorithm will be used to predict the rental price of housing through experiments, and test the effect of the prediction. According to the test situation, adjust the use of

features and data, conduct linear model training, compare the pricing effect of each experimental model, and analyze and evaluate the feasibility of the pricing method.

2. Literature review

2.1. Airbnb research on the influencing factors of housing prices

In the hotel industry, Airbnb, as a sharing economy platform, is both similar and unique to traditional hotels. In terms of price mechanism, the adjustment of Airbnb not only affects consumers' choice, but also concerns the income of landlords and platform profits. Compared with traditional hotels, Airbnb prices are affected by a variety of factors, including [1] social evaluation. The study points out that highly rated housing prices increased more, and reputational capital turned into rent increases by [2]. Dan Wang and Juan L. Nicolau Based on 33 city data, five core categories affect the price of Airbnb, covering 24 variables including landlord, property, facilities, rental rules and online reviews, and the degree of influence changes with price [3]. Property characteristics such as geographical location, property type, and the number of bedrooms and bathrooms directly affect the price. When facilities provide convenience, the rent will increase, and the average evaluation positively affects the price, but the excessive number of comments may negatively affect [4]. Based on the data from 36 cities, the Chinese study found that trust and social degree significantly affected the Airbnb price, the second-hand housing price in cities had a positive impact, and the hotel supply was negatively related [5]. This paper comprehensively considers the research results, finds the significant influence characteristics through the data analysis algorithm, verifies and expands the conclusions of the literature, and builds a foundation for the research on housing pricing.

2.2. Pricing research based on data analysis

With the rapid development of mobile Internet and the fierce competition in the network market, the real-time pricing strategy helps enterprises maintain their competitiveness and improve their profits. The research mainly focuses on static and dynamic pricing. Dynamic pricing flexibly responds to market fluctuations, while static pricing is important in specific scenarios such as product launches. Omar Besbes et al. studied the static pricing strategy of reusable resources and verified its generality and performance [6]. However, the results showed that its efficiency was at least 78.9% and better than the optimal dynamic pricing strategy. Paschalidis and Liu further confirm the asymptotic optimality of static pricing in a smaller-scale user environment [7]. The pricing of homestay is affected by multiple factors. Based on data analysis, this paper discusses the influence of multiple characteristics of B & B on pricing. After optimizing the key features, the static pricing of Airbnb housing is realized, and when the social attributes and market behavior are considered, the pricing model is constructed again to confirm the economic benefits of the pricing maintenance method.

3. Study design and related theory

3.1. Definition of the research questions

This study discusses the relationship between Airbnb housing characteristics and price, constructs the pricing model through a data simulation algorithm, and optimizes the performance of the algorithm. The experimental design focuses on the new housing pricing and the maintenance of existing rental housing prices. The goal is to build models responding to the complexity of pricing. Key features, particularly the evaluation data, will be introduced to comprehensively analyze their

impact on the pricing strategy. The core lies in the analysis of the role and influence of Airbnb housing characteristics in the pricing decision. Special attention is paid to how to deal with the diversified characteristics of housing resources, especially the location and evaluation characteristics.

3.2. Characteristic price theory

This study is based on the characteristic price theory [8], which is a key method for analyzing the composition of house prices in economics. The core point is that goods (especially real estate) are composed of multiple characteristics, and the overall price originates from the utility that characteristics bring to consumers. By refining the characteristic value and using market data estimation, the commodity price can be accurately calculated. Feature price theory includes Lancaster's consumer theory and Rosen's supply and demand equilibrium model. The former emphasizes the difference between commodity characteristics and the nature of demand, while the latter takes the utility and profit maximization in the context of the fully competitive market, providing the basis for theoretical modeling. This study provides an in-depth analysis based on this theoretical framework.

3.3. Study and implementation method design

A linear regression algorithm solves the pricing problem, but simple models face challenges. The integrated learning algorithm improves the performance by integrating the underlying model. This study explores the association of Airbnb characteristics and rental prices based on the characteristic price theory. Through advanced algorithms, we accurately predict feature prices and consider multi-dimensional, temporal feature data. The implementation process includes feature data processing, model building and static pricing, pricing maintenance research, and prediction effect comparison. After experimental analysis, we concluded the feature importance and the pricing effect of the algorithm model, and found that the evaluation data had a significant impact on the pricing, providing theoretical support and practical guidance for Airbnb pricing.

4. Data description

4.1. Data acquisition and preprocessing

4.1.1. Acquisition of data

The data base of this paper is derived from the housing source data publicly released by the Airbnb platform. Airbnb Housing data is been available around the world, where China is available for Beijing and Hong Kong. The publicly available data set in the Airbnb area covers three categories: housing source data, order data and comment data. The research work mainly focuses on the use of housing data pricing experiments. The data used a total of 74111, involving 10 key features, including but not limited to housing summary, housing facilities, rental rules, housing type, pricing, the maximum number, bathroom, bedroom number of beds and cleaning costs, etc (see Figure 1).
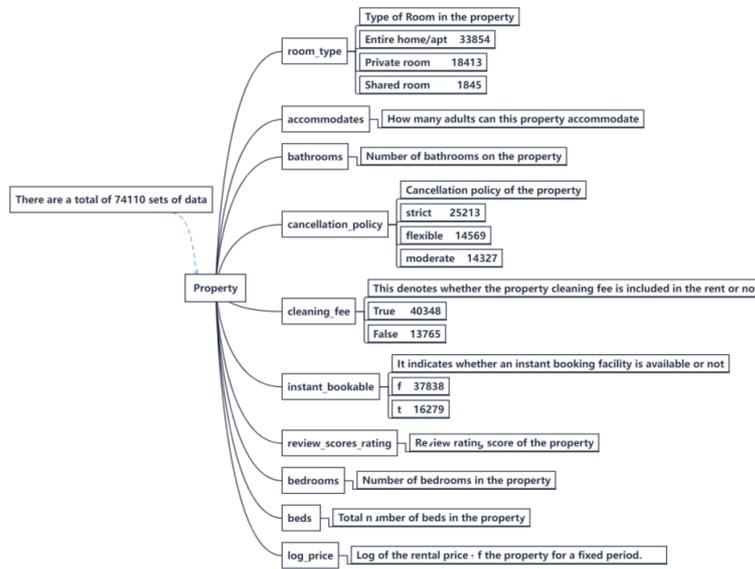
Figure 1: Data overview and feature value display

Given the many characteristics involved in the study subjects, the study first needs to screen out the characteristic data suitable for this study. Previous studies have shown that housing characteristics, room facilities, rental rules, trust and social factors all have a significant impact on Airbnb housing prices [9].

4.1.2. Data preprocessing

After completing the preliminary feature selection, the preprocessing of the above feature data is required. Since the data generated in the actual business inevitably has missing data and wrong outliers, in order to make the modeling effect not affected by this part of the data, the data preprocessing needs to be carried out before the subsequent steps.

The first is the missing value treatment. After statistical analysis of the above characteristics and rental price data, missing values in the following characteristics:

Table 1: Missing values of the primary characteristics

order number	Feature name	data type	Number of missing values
1	room_type	object	1
2	accommodates	float64	0
3	bathrooms	float64	27
4	cancellation_policy	object	2
5	cleaning_fee	object	2
6	instant_bookable	object	0
7	review_scores_rating	float64	16722
8	bedrooms	float64	11
9	beds	float64	25
10	log_price	float64	2

As shown in Table 1, except for the review_scores_rating feature, the amount of missing data of all features accounts for less than 0.1 ‰ of the total data, while the main reason for the missing data of review_scores_rating is the user habit, which accounts for about 22% of the total proportion, and has limited impact on the overall data, so there is no need to avoid the missing data of review_scores_rating.

In data processing, we identify outliers and decided to remove them, while populating the remaining data with missing value processing. It is found that, except for the rent price, other numerical data have similar values, so it is not normalized in this analysis.

4.2. Exploratory analysis of characteristic data

The following will conduct exploratory data analysis on the characteristic data representing the housing characteristics and the landlord characteristics in the primary characteristics obtained after data preprocessing, and present the relationship between each feature and the rent price through visualization.

4.2.1. Numfeature description

Data for numerical variables include the maximum number of people available in the price, number of bathrooms, number of bedrooms, scores, number of beds, and price.

Based on practical experience and basic logical analysis, it can be clearly said that the main direct driving force of homestay pricing comes from the number of bedrooms and the number of beds. These two not only directly reflect the size of the housing accommodation, but also directly related to the number of passengers that the house can accommodate. Therefore, the key feature of the maximum number of residents and its specific value will undoubtedly have a significant and direct impact on the rent pricing. This point is also strongly verified by the data visualization results.
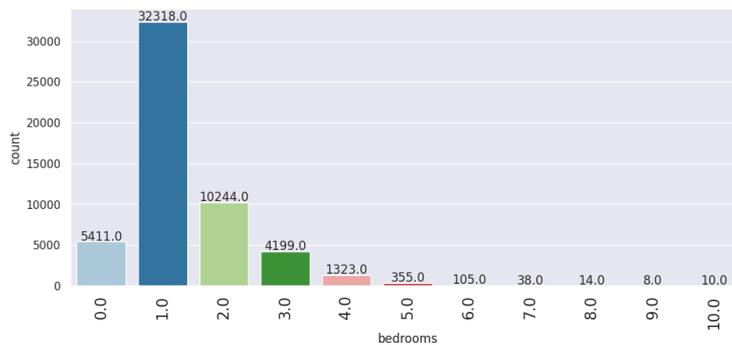

Figure 2: Distribution diagram of bedroom number characteristics

Figure 2 shows that in the statistics of the number of bedrooms, 43% of the houses have one bedroom, while most of the rest are distributed between two and three bedrooms, while the other number of bedrooms account for a relatively small proportion.

Figure 3: Distribution diagram of bathroom number characteristics

Figure 3, in the distribution of data values of the number of people available, 24% accommodate 2,13% accommodate 4,8% accommodate 3 or 1, while the number of the rest is relatively small.
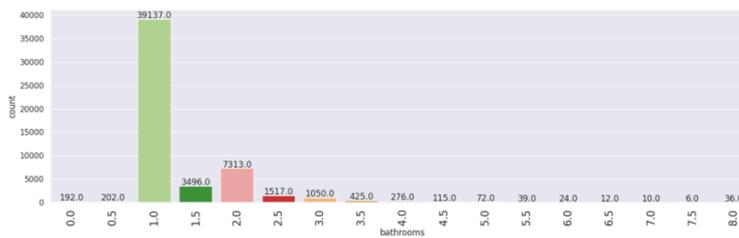

Figure 4: Distribution map of the number of people available

As shown from Figure 4, 52% of the data value of the bathroom number is 1, while the rest are mostly 2 or 1.5. The total amount of other cases is not much.
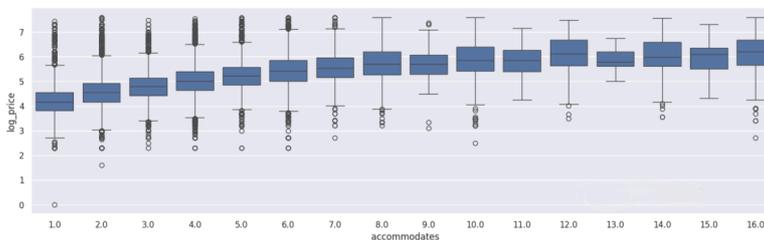

Figure 5: The relation diagram between score and price characteristics

Figure 5 shows that the price change showed an obvious linear growth trend; that is, with the increase of the number of people accommodating, the price also increased accordingly. However, it is worth noting that there is no obvious direct correlation between the lowest and highest prices set at the different stages and the actual number of people.
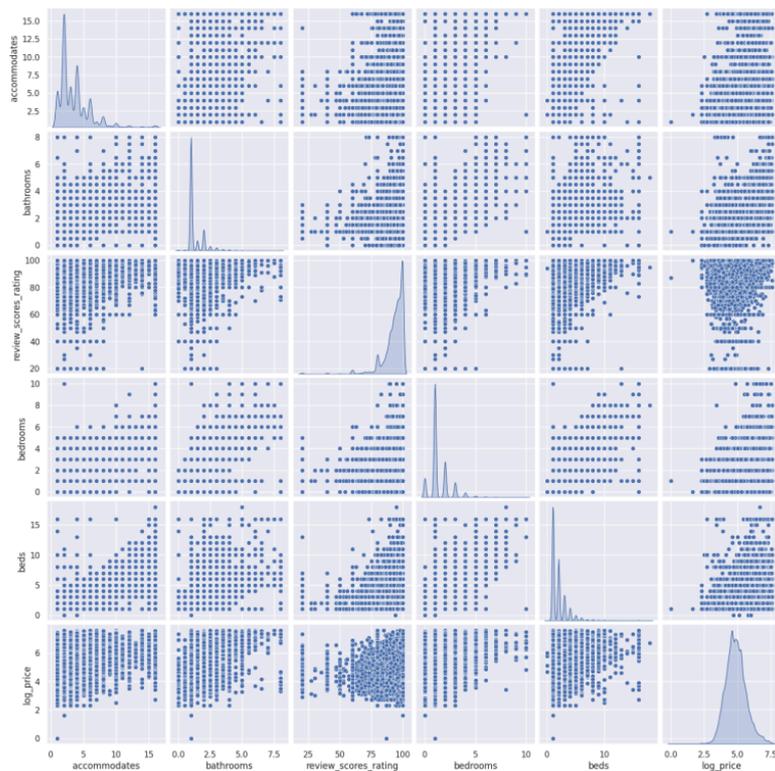
Figure 6: Correlation heat map

The latest analysis of the correlation heat map further confirmed the significant correlation between the maximum number of residents, scores, number of bathrooms, number of bedrooms and number of beds and the rental price. This finding once again highlights the important impact of these key factors on rent pricing (see Figure 6).

4.2.2. Description of classification characteristics

Characteristics of the data as categorical variables included house type, cancellation policy, and whether reservations are supported.

Figure 7: Number of housing types

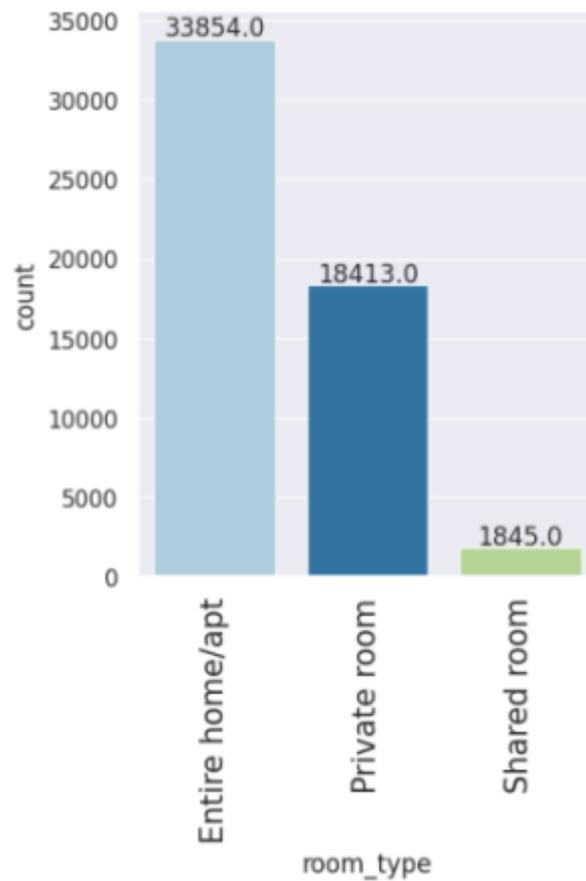Specifically, the housing source types are mainly divided into three categories, among which the whole rental houses or apartments account for more than 60%. As can be clearly seen from the data in Figure 7, the rent pricing range, median price and mean of the whole rental house / apartment are at the highest, followed by private rooms, while the rent of shared rooms is relatively low.
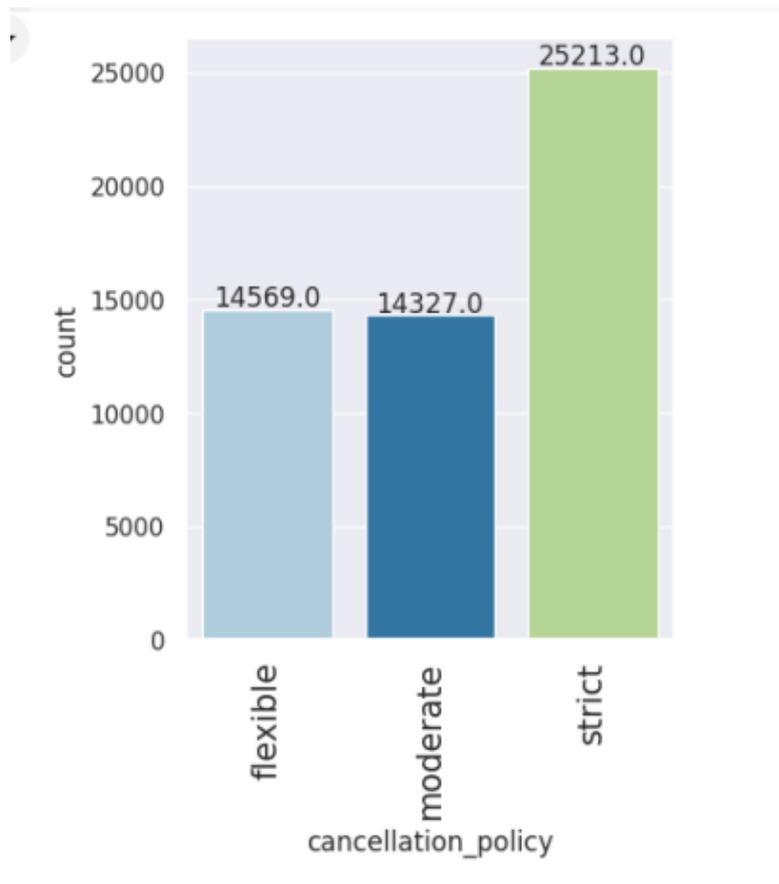
Figure 8: Number of cancellation policies

Cancellation policies can be subdivided into three categories, among which flexible cancellation policies account for 46%, and moderate cancellation policies and strict cancellation policies account for 26%, respectively. On the whole, as shown in Figure 8, the proportion of strict cancellation policies is nearly twice the sum of the flexible and moderate cancellation policies, while the moderate cancellation policies are slightly inferior to the number of flexible cancellation policies.

Based on 74111 housing data, data preprocessing and feature exploration of the Airbnb platform. We handled missing values with outliers and filtered the data to ensure quality. Numerical feature analysis shows that the number of bedrooms and beds is positively correlated with the pricing. Classification characteristic analysis shows that the rent of the whole rental house or apartment is high. The chart intuitively shows the characteristics and price correlation model, such as the number of bedrooms, the number of bathrooms, the number of people and price linear growth. Furthermore, the proportion of different cancellation policies in the data set was studied. To sum up, housing characteristics, rental rules and other key factors significantly affect the Airbnb housing price, providing a reference for the formulation of pricing strategies.

5. Research on static pricing of housing resources

5.1. Model fitting

noted how generalization ability is exhibited. On the training set, the RMSE is 0.513981, while on the test set, it gives 0.51951. This is minute, showing that the model will remain in the same

prediction when exposed to unseen data. The MAE values are 0.398918 for the training set and 0.403697 for the test set; in other words, there is a very slight growth in absolute prediction error on test data. The R-squared value is the proportion of variance explained by the model; it amounts to 0.510795 for the training set and 0.505223 for the test set. Besides, the figures are next to each other: 0.510653 and 0.504887 for training and testing datasets, respectively, as the adjusted R-squared value. This means that the ability of the model between both data sets is close to explaining the variation of the outcome. Also, generalization is great. MAPE increases slightly between the training dataset to 8.437099% and the test dataset, to 8.362512%, indicating that this model shows stability in the prediction accuracy of percentage errors. Generally, small discrepancies between training and test sets would suggest a well-fitted model with new data, containing no significant overfitting and underfitting. The model would be reliable.

5.2. Characteristic significance analysis according to the multiple regression results

This study uses a linear regression model to analyze the impact of multiple variables on log price, as well as the significance of different variables in predicting log price. The formula of the model is $\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$. This formula constructs a regression equation for predicting the dependent variable log-price by linearly combining each independent variable and its corresponding coefficient. The chart "coefficient of linear model" shows the direction and degree of influence of each variable on log price and, more intuitively, shows the relative relationship of the influence of different variables.

Figure 9: Coefficients of the linear model

As shown in Figure 9, variables with positive coefficients include bathrooms, bedrooms, accommodations, and review scores rating. This means that as these variables increase, the log-price will increase. Variables with negative Coefficients include room type, shared room, private room, cleaning fee, etc. This means that when this type of variable increases, the log-price will decrease accordingly, so housing prices can also be increased by reducing this type of variable. At the same time, combined with the ODS chart, the significance level of different variables is verified through p-value to evaluate the role of each variable in the model, thereby determining which variables are important predictor variables.

```
                        OLS Regression Results
==============================================================================
Dep. Variable:                     y   R-squared:                     0.506
Model:                           OLS   Adj. R-squared:                0.506
Method:                Least Squares   F-statistic:                    3528.
Date:               Sat, 06 Jul 2024   Prob (F-statistic):             0.00
Time:                       07:17:19   Log-Likelihood:               -28681.
No. Observations:              37872   AIC:                        5.739e+04
Df Residuals:                  37860   BIC:                        5.749e+04
Df Model:                         11
Covariance Type:           nonrobust
==============================================================================
                                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const                          3.2373      0.052     61.938      0.000       3.135       3.340
accommodates                   0.0935      0.003     31.535      0.000       0.088       0.099
bathrooms                      0.1799      0.007     24.763      0.000       0.166       0.194
cleaning_fee                  -0.0801      0.007    -12.260      0.000      -0.093      -0.067
instant_bookable              -0.0578      0.006     -9.921      0.000      -0.069      -0.046
review_scores_rating           0.0137      0.001     25.412      0.000       0.013       0.015
bedrooms                       0.1572      0.005     31.518      0.000       0.147       0.167
beds                          -0.0440      0.005     -8.039      0.000      -0.055      -0.033
room_type_Private room        -0.6042      0.007    -88.917      0.000      -0.617      -0.591
room_type_Shared room         -1.0490      0.015    -68.196      0.000      -1.079      -1.019
cancellation_policy_moderate  -0.0681      0.007     -9.108      0.000      -0.083      -0.053
cancellation_policy_strict    -0.0059      0.007     -0.862      0.388      -0.019       0.007
==============================================================================
Omnibus:                    2396.202   Durbin-Watson:                  2.014
Prob(Omnibus):                 0.000   Jarque-Bera (JB):            3978.924
Skew:                          0.503   Prob(JB):                       0.00
Kurtosis:                      4.229   Cond. No.                    1.87e+03
==============================================================================
```

Figure 10: OLS regression results

As shown in the figure 10, the p-values of most variables are extremely small (shown as 0.000), which indicates that they have a significant impact on the regression model, which suggests that these variables are strong predictors of log-price. However, the p-value of the cancellation policy strict is 0.388, and the p-value is greater than 0.05, indicating that the impact of the cancellation policy strict on log price is not significant.

6. Business insights and suggestions

On the OLS summary output, there is a particularly large p-value (0.388) appearing as we mentioned above. Such an unusual p-value was contributed by the dummy variable we created based on the category of rooms with strict cancellation policy. To address quantitative predictor which shows insignificance in the model, we can drop the variable from the model. However, in this case, the problematic variable represents a sub-category, and thus, we cannot ignore and delete it to address its insignificance, as other sub-categories, such as flexible or moderate cancellation policy, both have a p-value less than 0.05, which reveals their high significance. Then, we can only conclude that the prices of rooms with a strict cancellation policy do not show a noticeable difference from those of cancellation-friendly rooms. Accordingly, there are several plausible assumptions toward the result. Firstly, most travellers tend to book AirBnB rental rooms out of an immediate need. For example, people who go on business trips will be less likely to change their travelling plan, hence paying less attention to whether the rent is refundable. Secondly, the number of rooms for each type of cancellation policy is about the same, so rents will not differ as customers will not immediately differentiate certain types. Therefore, we suggest the business focus more on other factors, for example, paying more attention to improving users' experience on room types, number of bedrooms, or number of bathrooms.

Then, according to the statistic summary again, the largest coefficients within the model both come from the "room type" category. Setting the "Entire room/apt" as the baseline category and introducing two dummy variables to represent the other two categories, we can see the two variables, which are "Private room" and "Shared room," have a relatively large absolute value

(negative values in this case) of coefficients. Considering the coefficients as well as extremely small p-values (rounded to 0 in the output), it is reasonable to conclude that room type does play a significant role in determining the (log) prices. On the basis of such analysis, we suggest that AirBnB should make special clarification on the room type to customers. For example, the web designer team can apply highlighting effects or special fonts on the information page of room resources that are identified with "Entire room/apt" or "Private room." This action can help customers efficiently recognize the advantage, thus better justifying the high prices for these rooms.

7. Conclusion

In this article, we delve into the complexity and diversity of Airbnb listing pricing, proposing a comprehensive pricing strategy through the comprehensive analysis of multiple influencing factors. Our research not only verifies many influencing factors mentioned in previous literature, such as location, listing type, amenities, and user reviews, but also introduces new perspectives, such as seasonal fluctuations, holiday effects, and market competition, to comprehensively analyze the pricing mechanism.

The core highlight of this research lies in its data-driven methodology. With the help of big data analysis and machine learning technology, we have conducted in-depth mining of massive Airbnb listing information, extracted key pricing factors, and built an accurate prediction model. This method not only improves the accuracy of pricing but also provides a scientific decision-making basis for hosts, helping them optimize their pricing strategies and enhance market competitiveness.

In addition, this study emphasizes the importance of trust-building between hosts and guests in the context of the shared economy. Through transparent pricing mechanisms, hosts can send signals of honesty and integrity to guests, thereby enhancing their trust and loyalty. The establishment of this trust relationship is of great significance for the long-term development of the Airbnb platform.

Looking to the future, with the continuous development of the shared economy, the issue of Airbnb listing pricing will become more complex and variable. Therefore, we need to continuously monitor market trends and optimize pricing models to adapt to new market environments. At the same time, we should explore more innovative pricing strategies, such as personalized pricing and dynamic pricing, to further enhance the flexibility and competitiveness of pricing.

In summary, this study provides new perspectives and solutions for the issue of Airbnb listing pricing, which has important academic value and practical significance. We believe that in future research, we will be able to deepen our understanding of this issue and contribute more wisdom and power to the development of the shared economy.

Acknowledgement

References

[1] Sun Yanqing. Research on Intelligent Pricing of Airbnb Listings Based on Multi-feature Fusion [D]. Shanghai University of Finance and Economics, 2023. DOI: 10.27296/d.cnki.gshcu.2020.002408.
[2] Tapio Ikkala and Airi Lampinen. 2014. Defining the price of hospitality: networked hospitality exchange via Airbnb. In Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW Companion '14). Association for Computing Machinery, New York, NY, USA, 173–176.

[3]   Gutt Dominik, Herrmann Philipp, 2015, "Sharing Means Caring? Hosts'Price Reaction to Rating Visibility", ECIS 2015 Research-in-Progress Papers, Paper 13.

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

[5]   Wu Xiaojun, Qiu Zhuilu, 2019, Research on the Influencing Factors of Airbnb Room Prices - Based on Data from 36 Cities in China [J], Journal of Tourism Studies, 2019.34(4), pages 13-28.

[6]   Omar Besbes, Adam N. Elmachtoub, and Yunjie Sun, 2019, "Static Pricing: Universal Guarantees for Reusable Resources", In ACM EC 19: ACM Conference on Economics and Computation (EC 19), June 24–28, 2019, Phoenix, pp.393-394.

[7]   Ioannis Ch. Paschalidis, Yong Liu, 2002, "Pricing in Multiservice Loss Networks: Static Pricing, Asymptotic Optimality, and Demand Substitution Effects", IEEE/ACM Transactions on Networking, VOL. 10, NO. 3, JUNE 2002, pp.425-438.

[8]   Peng Ye, Julian Qian, Jieying Chen, Chen-hung Wu, Yitong Zhou, Spencer De Mars, Frank Yang, and Li Zhang, 2018, "Customized Regression Model for Airbnb Dynamic Pricing", In KDD 18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom. ACM, New York, NY, USA, pp.932-940.

[9]   Jia Shenghua, Wen Haizhen, 2004, The theoretical development and application of the real estate characteristic price model [J], Foreign Economic and Management, 26th volume, 5th edition, pages 42-44.