Proceedings of the 3rd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/2024.25024

Dynamic Pricing in Short-Term Rentals: An Empirical Examination of Airbnb Listings

Ran Ji[1*], Xuting Huang[2], Yaxin Qian[3]

[1] *University of California, Irvine, USA*
[2] *University of Washington, Washington, USA*
[3] *New York University, New York, USA*
**Corresponding Author. Email: rji6@uci.edu*

Abstract. With the rapid growth of Airbnb, the nature of the short-term rental market has changed. Therefore, getting an in-depth knowledge of those drivers that influence pricing strategies is essential. On a dataset of about 75,000 Airbnb listings, this paper analyzes variables related to room type, cancellation policy, and quality of amenities. This paper uses both linear regression and decision tree models to quantify the direct effects of these variables on rental prices. The results of this study underline strong influences from property type, accommodation capacity, and policy settings as critical factors on pricing, thus helping Airbnb hosts optimize their pricing strategies. This paper has further implications for the sharing accommodation market and provides valuable information to many policy decision-makers and market analysts. The results promote more informed decision-making for hosts and enhance customer satisfaction, benefiting both property owners and guests.

Keywords: Airbnb, pricing strategy, linear regression, decision tree, short-term rental market

1. Introduction

The sharing economy model, also known as collaborative consumption or peer-to-peer sharing, is an economic model that leverages individuals' ability (and perhaps preference) to rent or borrow goods and services rather than buy or own them [1]. With the booming development of the sharing economy, Airbnb, as one of the leading players, has fundamentally changed how people travel and live. Airbnb exemplifies this model by allowing property owners to rent unused rooms or homes to travelers seeking short-term accommodations. This approach helps homeowners earn extra income and provides travelers with a more economical and locally immersive lodging option compared to traditional hotels [1]. As the world's largest online short-term rental platform, it provides multivariant housing from small apartments to luxurious villas and allows the opportunity to profit from millions of hosts worldwide. According to recent studies, the existence of Airbnb has dramatically increased the real estate demand in tourist destinations and indirectly increased the cost of local accommodations [2]. Therefore, in an increasingly competitive and dynamic market, understanding what drives the pricing in Airbnb is crucial to hosts as a way of making the most earnings and to consumers as a means of securing fair and reasonable accommodations.

The extant literature employs a detailed dataset containing approximately 75,000 listings to derive the multi-faceted dynamics driving varied short-term rental market pricing strategies.This data spans multiple dimensions that significantly affect rental prices, such as room type, cancellation policy, cleaning fee, the number of bathrooms, and the availability of instant booking. These factors are further compounded by evolving market conditions and rising consumer expectations. Critical balance, meeting consumer needs and ensuring profitability, is a dual imperative for both property hosts and platform operators; from that perspective, this study is poised to address the same

This study employs a robust analytical framework using linear regression and decision tree models to dissect and understand how these diverse factors collectively influence Airbnb pricing. The multiple linear regression model facilitates a clear quantification of the direct impacts of individual variables on rental prices, offering precise estimations and insights. Complementarily, decision tree models are utilized to unravel the complex interactions and non-linear dependencies between these variables, providing a holistic view of the factors at play [3]. In both cases, we check the accuracy and robustness of the models via MAE, MSE, RMSE, and adjusted R-squared.

The Airbnb market implications of this study are that it provides the scientific base for pricing determination by Airbnb hosts. Previous studies have established that research in this regard provides strategic and informed decisions that are critical for techniques in the sharing economy model [1]. The analysis results obtained valuable insights for policymakers, business analysts, and market researchers, in the way different factors interplay to make effective pricing strategies, as investigated in a broader context on home sharing's impact on housing markets. This knowledge is important to optimize practices of property management, increase customer satisfaction, and ensure sustainable development within a sharing economy[2]. Furthermore, such statistical learning methods include decision trees that conduce to a more just and equal market environment and benefit the hosts and guests of these sharing platforms by giving them a clear framework for price setting based on empirical evidence.

2. Method

2.1. Data collection

The chosen dataset derives real-time listing information directly from the official Airbnb website, encompassing a diverse array of listings from various global locations. This extensive dataset introduces realistic variabilities and comprehensive characteristics to our analysis, featuring approximately 75,000 listings. Such a large sample size is conducive to robust statistical analysis. Each listing includes detailed information about room type, available accommodations, number of bathrooms, cancellation policy, cleaning fees, instant booking status, review scores, number of bedrooms, number of beds, and log-transformed prices. The attributes considered are those that directly influence room prices and are highly valued by customers when selecting an Airbnb. The advantage of having a log-transformed dependent variable is that it effectively exponentiates the coefficient, which means that each time the independent variable increases by one unit, this provides the multiplicative factor. The coefficient, for instance, is 0.198, and $\exp(0.198) = 1.218962$. Our dependent variable grows by around 1.22, or 22%, for every unit increase in the independent variable. This increases the robustness and interpretability of our model [4].

DOI: 10.54254/2754-1169/2024.25024

2.2. Data processing

To maintain data integrity, records with incomplete data were omitted to address missing values, leaving the analysis around 60,000 usable rows. Since regression and machine learning algorithms can only interpret numeric input, it is crucial that all categorical data are converted into a series of binary columns. Using the dummy-coding method, each row from a categorical variable is converted into its column. The presence of the value in the original column is represented by a 1 in the new column, and the absence is represented by a 0. Figure 1 demonstrates the distribution by room type after the variable "room type" was transformed into a factor, making it readily interpretable [5].

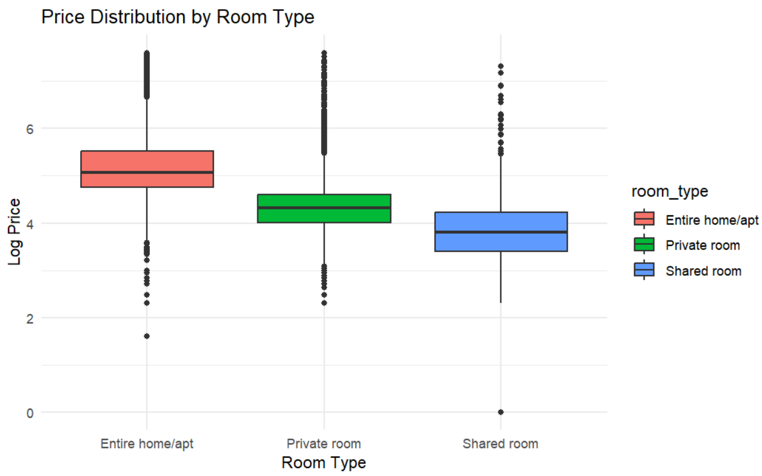

Figure 1: Price distribution by room type

*Notes: The price distribution for different room types in various accommodation options indicates that "Entire homes" generally have the highest price, followed by "Shared room" and then "Lowest price"

Model evaluation of the training and testing sets was done using an 80-20 split of the dataset as is. From here, the log-transformed price of Airbnb listings is the dependent variable. Variables that vary are room type, accommodation (night and day), bathrooms & showers/cancellations policy, cleaning fee, instant bookable status, review scores rating, bedrooms & reviews. Various aspects of the listings that may affect pricing are encompassed by these variables.

2.3. Model building and evaluation

The linear regression model was used as a starting point for the analysis, which was then refined through stepwise regression, as demonstrated in the image below. Patterns of non-linearity were investigated by examining residuals using a Q-Q plot and consolidated with a plot of residuals against the fitted values. To evaluate the performance of the linear regression model, we employed a k-fold cross-validation. The k-fold cross-validation procedure begins with the data being divided into K groups at random, after which the following procedures are carried out for each group:

1. Choose a subset of the training data to use for the experiment.
2. A training set is constructed from the remaining K − 1 groupings.
3. Train the model on the data you've chosen, then put it to the test on the evaluation data.

In this study, k is fixed as 10, which is an empirical number established after lengthy experimental trials. The typical value of k in a small sample dataset for this study is 10, which was

DOI: 10.54254/2754-1169/2024.25024

arrived at empirically after several experimental attempts. As the validation set, a new testing fold was chosen each time, ranging from D1 to D10. These ten datasets were then consecutively red into the machine learning models. Using 10-fold cross-validation, the inaccurate model assessment brought on by the unintentional partition of the sample datasets may be eliminated entirely. The model was then evaluated using the following metrics: R-square (R2), which measures the proportion of data variance explained by the model, as well as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), both of which measure the average magnitude of errors between the predicted and actual values [6].

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum\left(y_i - \underline{y}\right)^2}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}|$$

*Notes: $\hat{y}$ is the predicted value of y; $\underline{y}$ is the mean value of y.

A decision tree model was employed to identify the factors that have an impact on determining listing prices, following the linear regression model. The decision tree model is a powerful tool for analyzing Airbnb data, providing a structured approach to decision-making by breaking down various factors into a series of nodes and branches. This model effectively combines multiple data points, such as the number of bedrooms, type of room, and review scores, to predict outcomes like pricing and occupancy rates. Airbnb data exhibit non-linear relationships, contain both categorical and numerical variables, and often have complex interactions between features. Decision trees are well-suited for capturing these characteristics, making them an appropriate choice for this type of data. By highlighting key features and their importance, the decision tree model allows Airbnb hosts to understand which factors most significantly impact their listing's performance, enabling them to make informed decisions to optimize their listings and enhance customer satisfaction [7]. Comparable gauges were then employed to assess the model's performance, and feature significance was assessed to determine the relative importance of each characteristic in relation to the model's predictions.

3. Results

3.1. Linear regression model

Our final model achieved an adjusted R-squared value of 0.564, indicating that the predictors explain approximately 56.4% of the variance in log prices. The regression coefficients, their standard errors, t-values, and p-values are summarized in Table 1.

Table 1: Lm model result

Variable	Estimate	Standard Error	T Value	Pr(>|t|)
(Intercept)	3.9482333	0.0263025	150.109	<2e-16 ***
Room Type _ Private room	-0.6283696	0.0049340	-127.354	<2e-16 ***
Room Type _ Shared room	-1.1202108	0.0137827	-81.276	<2e-16 ***
Accommodates	0.0700860	0.0019757	35.475	<2e-16 ***
Bathrooms	0.1334774	0.0046061	28.979	<2e-16 ***
Cancellation Policy _ Moderate	0.0179605	0.0058404	3.075	0.0021 **
Cancellation Policy _ Strict	0.0658267	0.0053897	12.213	<2e-16 ***
Instant Bookable	-0.0611303	0.0047129	-12.971	<2e-16 ***
Review Scores Rating	0.0058645	0.0002674	21.927	<2e-16 ***
Bedrooms	0.1603562	0.0039295	40.809	<2e-16 ***
Beds	-0.0480865	0.0030559	-15.735	<2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 0.442 on 45673 degrees of freedom				
Multiple R-squared: 0.5642		Adjusted R-squared: 0.5641		
F-statistic: 5913 on 10 and 45673 DF		P-value: < 2.2e-16		

3.1.1. Coefficient interpretation

In our final multiple linear regression model, the effect of room type on price is significant (also agree with the result shown on the heat map). Private and shared rooms have significantly lower logarithmic prices compared to the entire home or apartment, with the log price decreasing by 0.628 for private rooms and 1.12 for shared rooms. Besides, the number of people that can be accommodated is also significant (p-value = < 2e-16), with the log price increasing by 0.07 for each additional guest. Similarly, the number of bathrooms shows a significant effect, with the log price increasing by 0.133 for each additional bathroom. Cancellation policy also shows a significant impact on price. Compared to flexible cancellation policies, moderate and strict cancellation policies increase the log price by 0.018 and 0.064, respectively. In contrast, listings with instant bookable options have lower log prices, decreasing by 0.066. Review scores show a positive effect; for each additional rating unit, the log price increases by 0.006. Similarly, the number of bedrooms exhibits a significant positive effect, with each additional bedroom increasing the log price by 0.16. However, the number of beds has a negative impact on price, with each additional bed reducing the log price by 0.048.

Instead of the log price, we further consider how the price changes accordingly. By exponentiating the regression coefficients, we can obtain the multiplicative effect of each variable on the original price. For instance, if the coefficient of a predictor variable is β, then for each unit increase in that variable, the log price increases by β units, corresponding to a multiplicative effect on the original price:

$$\Delta Price = e^{\beta}$$

DOI: 10.54254/2754-1169/2024.25024

Specifically, the log prices for private and shared rooms are substantially less than those of complete homes or apartments. As the room types change from the entire home or apartment to the private rooms, the log price decreases by 0.628, indicating an approximately 53.4% reduction in the original price. Similarly, the log price for shared rooms drops by 1.12, which translates to about 32.6% savings in the initial cost. Moreover, the price increases by approximately 7.3% for each extra guest that can be accommodated, 14.2% for every extra bathroom, and 17.4% for every extra bedroom. However, the number of beds is negatively correlated with the price, as the number of beds increases by 1 unit, the price will decrease by approximately 4.7%.

3.1.2. Model diagnostics

After analyzing the coefficients, we examined the normality, homoscedasticity, and independence of the residuals, to ensure the robustness of our model.

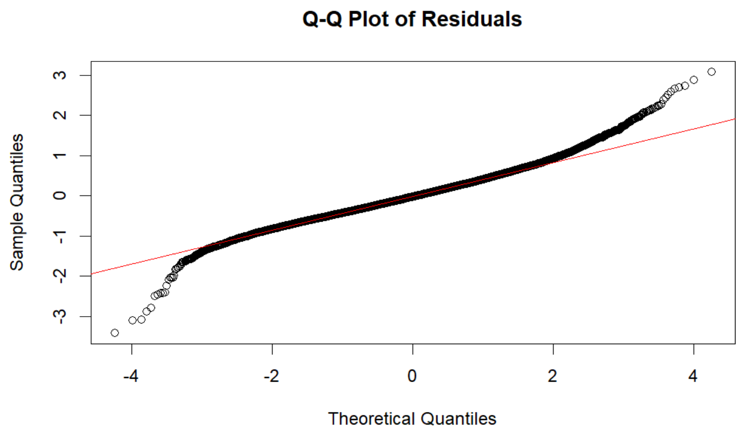

Figure 2: Q-Q plot

*Notes: The Red Line represents the expected quantiles if the residuals were perfectly normally distributed.

The Points represent the actual quantiles of the residuals. If the points closely follow the red line, it indicates that the residuals are approximately normally distributed.

As shown in the above Q-Q plot of the residuals (Figure 2), the points mostly follow the red line, indicating that the residuals are roughly normally distributed. The deviations from the line at the ends (tails) indicate potential deviations from normality, which could be due to outliers or heavy tails. This suggests that the normality assumption is fairly satisfied.

DOI: 10.54254/2754-1169/2024.25024

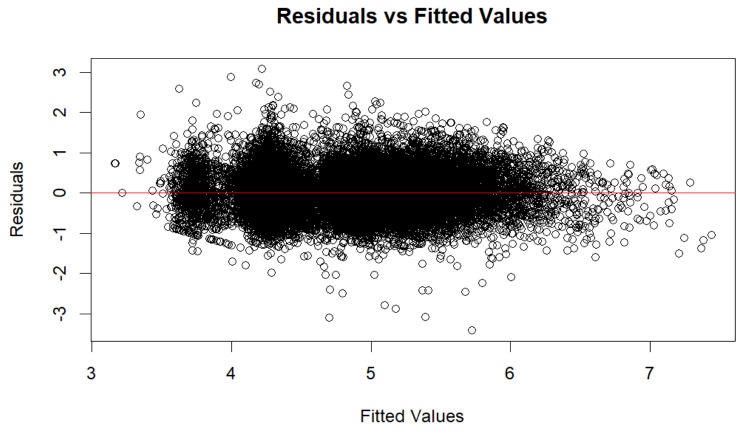

Figure 3: Residuals vs fitted values

*Notes: The red line represents where the residuals would be if they had a mean of zero.

The residuals should be randomly scattered around the red line.

As shown in the above graph (Figure 3), the residuals appear to be randomly distributed around the red line, and there is no clear pattern, suggesting that the linearity assumption is reasonably met. Also, the spread of the residuals are relatively constant, although there is a slight funnel shape indicating possible heteroscedasticity (non-constant variance).

3.1.3. Significance assessment of interaction terms

In order to test the possible effect of the potential interaction term on the model, we performed the ANOVA test. The results are summarized in Table 2 below. The result shows that the interaction terms between room type and cancellation policy ($F = 86.639$, $p < 2.2e-16$), room type and instant booking status ($F = 13.929$, $p < 8.96e-07$), and cancellation policy and cleaning fee ($F = 3.536$, $p < 0.007$) are significant. Specifically, room types performed differently across different cancellation policies and whether or not instant booking was available. Also, the effect of cleaning fees varies across combinations, e.g., a strict cancellation policy may have a greater impact on price due to a high cleaning fee in some room types, while it may have less of an impact in other room types.

DOI: 10.54254/2754-1169/2024.25024

Table 2: ANOVA test result

Variable	Sum Sq	Df	F Value	Pr(>F)
Room Type	7445.6	2	14525.3856	<2e-16 ***
Cancellation Policy	133.9	2	261.2193	<2e-16 ***
Cleaning Fee	12.0	1	46.8638	7.705e-12***
Instant Bookable	25.0	1	97.6720	<2e-16 ***
Room Type: Cancellation Policy	88.8	4	86.6386	<2e-16 ***
Room Type: Cleaning Fee	1.9	2	3.6454	0.026118*
Cancellation Policy: Cleaning Fee	0.4	2	0.8638	0.421563
Room Type: Instant Bookable	7.1	2	13.9292	8.964e-07 ***
Cancellation Policy: Instant Bookable	0.1	2	0.1508	0.860020
Cleaning Fee: Instant Bookable	0.8	1	3.2111	0.073146
Room Type: Cancellation Policy: Cleaning Fee	3.6	4	3.5355	0.006861 **
Room Type: Cancellation Policy: Instant Bookable	6.9	4	6.7368	2.044e-05 ***
Room Type: Cleaning Fee: Instant Bookable	0.3	2	0.6351	0.529883
Cancellation Policy: Cleaning Fee: Instant Bookable	0.1	2	0.2042	0.815290
Room Type: Cancellation Policy: Cleaning Fee: Instant Bookable	2.5	4	2.3996	0.047783 *
Residuals	11699.5	45648		
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

3.1.4. Cross-validation

To fully test the performance of our linear model, we continued with a 10-fold cross-validation test. The test results show that the R-squared (coefficient of determination) is 0.5640624, implying that the model explains about 56.4% of the data variance. Besides, the error indicators RMSE (Root Mean Square Error) is 0.442006 and MAE (Mean Absolute Error) is 0.3436586. These indicate that the model has some error in predicting the listing price but has relatively good prediction accuracy in most cases. While showing the limitation of our current model, the CV result also leads to some possible guidance for our next step, such as performing nonlinear transformations or using more complex models (e.g., Random Forest or Gradient Boosting Tree) to improve the prediction accuracy.

3.2. Decision tree model

The decision tree model was employed to analyze various factors affecting Airbnb listing prices. The dataset included features such as room type, number of bedrooms, bathrooms, accommodation, cleaning fees, review scores, and cancellation policies. The summary of the results obtained from the decision tree model is shown in Table 3.

Table 3: Decision tree model result

Metric	Value
Mean Absolute Error (MAE)	0.4059
Mean Squared Error (MSE)	0.2930
Root Mean Squared Error (RMSE)	0.5413
R Squared Score	0.4297
Adjusted Squared Score	0.4293

Instead of using p-value in the linear regression model, the decision tree model uses feature importance. P-values are used in linear regression to test the significance of individual predictors. These models assume a linear relationship between the predictors and the outcome variable. Since the decision tree models do not assume a linear relationship, instead, they recursively partition the data into subsets based on the most significant splits at each node. The decision tree algorithm selects features and thresholds that maximize the separation of the data at each node, but it does not generate coefficients for predictors that would be tested for significance [3]. The importance of features in decision trees is measured by how often a feature is used in the splitting process and the improvement in the splitting criterion it brings. The results of feature importance are shown in Table 4.

Table 4: Feature importance values

Feature	Importance
Room Type_Private Room	41.60%
Bedrooms	18.58%
Bathrooms	9.15%
Room Type_Shared Room	8.82%
Review Scores Rating	6.55%
Accommodates	6.18%
Beds	3.57%
Cleaning Fee	1.60%
Instant Bookable	1.50%
Cancellation Policy _ Strict	1.31%
Cancellation Policy _ Moderate	1.14%

3.3. Model analysis

Firstly, we analyze the model performance metrics, which include five parts. The mean absolute error (MAE) is 0.04059. This metric indicates the average absolute difference between the predicted and actual log prices is approximately 0.4059, which means the model's predictions are reasonably close to the actual values on average. The Linear Regression model has a lower MAE (0.34) which indicates that the absolute errors of the Linear Regression model's predictions are smaller on average.

The main squared error measures the average of the squared differences between the predicted and actual values. An MSE of 0.2930 for the decision tree model suggests that, on average, the squared differences between the predicted and actual log prices are 0.2930. While the MSE value of the linear regression model is 0.1936. The linear regression model excels in prediction accuracy and interpretability, making it ideal for simpler, linear relationships. The decision tree model, despite its higher MSE, offers valuable insights into feature importance and can capture more complex relationships, making it suitable for more nuanced analyses.

The RMSE is 0.5413, corresponding to a moderate level of error. The linear regression model has a lower RMSE at 0.44 and is more sensitive to major errors, so it has less number of large errors in the predictions. An R^2 of 0.4297 means that a decision tree model fosters about 42.97% of the variability of log prices. Though the above model captured much of the variance, there is still much more variability in log prices that the model is not able to explain. This indicates that there may be some factors left out of the model or that there might be an element of inherent randomness in the data. This is a considerable improvement over the decision tree model, where the R^2 is 0.4297, and with a Linear Regression model $R^2 = 0.565$. A higher R^2 indicates that this linear regression model fits the data better than this Decision Tree model.

A value of 0.4293 for the adjusted R^2 is taken to mean that after controlling for too many predictors, the decision tree model can explain up to about 42.93% of the variance in log prices. The model has a slight reduction in the R^2 score, which reflects pessimism, while it is good not to severely suffer from overfitting because the value has reduced slightly due to too many predictors. By contrast, the R^2 score for the linear regression model, appropriately adjusted, computes the same performance at a worse adjusted R^2 of 0.5649. Moderate RMSE and an R^2 score suggest that the decision tree model is being somewhat accurate but still it is leaving grey area towards opportunity of improvement. From the result of the random forests, it is clear that about 57% of the variance in log prices is unexplained. Therefore, there is a possibility for an increase in the model's effectiveness by adding more relevant features or using contemporary methods such as ensemble methods.

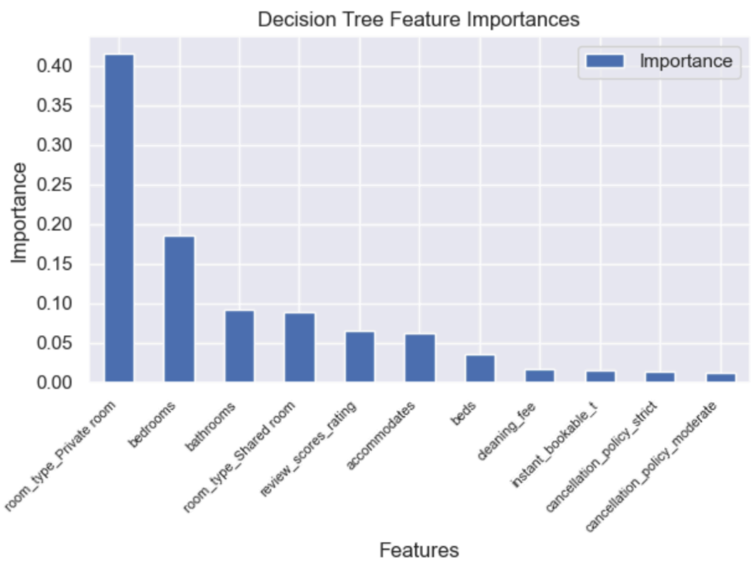

Figure 4: Decision tree feature importance

The decision tree model provides insights into which features are most influential in determining Airbnb listing prices included in Figure 4. For example, room type (private room), number of bedrooms, and bathrooms are identified as the most important features, highlighting their significant impact on pricing. Therefore, for Airbnb hosts, understanding the importance of these features can

DOI: 10.54254/2754-1169/2024.25024

guide decisions on how to optimize their listings, such as investing in additional bedrooms or improving bathroom facilities might lead to higher listing price

4. Discussion

Through the examination of Airbnb listing data, it becomes clear that there are several factors that have a significant impact on rental prices, providing valuable insights into the short-term listing market.

4.1. Effect of space type

A significant difference in cost between private or shared rooms and entire homes or apartments is a Noteworthy finding. Prices for entire properties are much higher, reflecting strong consumer demand for space and privacy. The statement is in line with Airbnb's distinctive advantage of providing accommodation options that cater to a wide range of budgets and needs. But the higher cost of whole properties understates the value that travelers place on having a complete and personal living space, which is often the most desirable feature of travel – or for families, groups with whom one could feel at home.

Furthermore, the number of bedrooms and bathrooms was also an important factor of price. Properties that have more bedrooms and bathrooms are typically priced higher, indicating that the market is focused on finding lodging that can accommodate larger groups or provide additional comfort and convenience. This indicates that property owners could potentially receive higher returns if they invest in expanding or upgrading these facilities.

4.2. Regulations on cancellation and pricing mechanisms

Strict cancellation policies are associated with higher prices, which is a curious finding in the study. As a result of this correlation, it seems that hosts are charging more in risk for less flexible booking terms. I think the reason this is so, maybe it's because properties that enforce strict cancellation policies appeal to a certain group of travelers who want peace of mind and are willing to pay. In the Airbnb industry, pricing strategies and booking policies are interdependent, as demonstrated by this study.

4.3. Consequences for hosts

Airbnb hosts can now optimize their listings and maximize their returns with these findings: Optimizing Property Types: Hosts should consider offering complete properties where possible, as they are the most expensive. In the absence of offering entire houses, highlighting the unique features of private rooms that offer privacy may be beneficial to justify higher rates. Enhancements to facilities: If possible, adding more bedrooms or bathrooms would increase the value of a property in the Airbnb market. Prices may justify larger upgrades that enhance both comfort and convenience.

It is important for hosts to carefully evaluate their cancellation policies and how they are compatible with their pricing strategy. Hosts must weigh the potential loss of bookings from guests who want more freedom, even if stricter policies lead to higher prices. Host marketing should be targeted: To attract more customers, hosts must highlight the unique features of their properties that make them attractive. These might encompass promoting the use of space, privacy or other specific features catering to their target audience.

DOI: 10.54254/2754-1169/2024.25024

4.4. Consumer behavior in the Airbnb sector

When it comes to the Airbnb sector consumers have an array of choices tailored to their preferences and budgets. For budget travelers or solo adventurers looking to spend time at their accommodations, private or shared rooms are a great option that offers affordability without compromising on essential amenities.

On the side property owners who invest in buildings can provide luxurious experiences for those seeking more space and privacy. These properties cater to travelers in search of stays and families or larger groups requiring room. Cancellation policies also play a role in pricing decisions influencing the balance between cost savings and booking flexibility. Travelers watching their budget may lean towards listings with cancellation terms to save money while those prioritizing flexibility might opt for pricier options offering booking freedom.

Looking ahead, there are trends shaping the term rental market. The significant price variations across room types indicate a market that values diversity, and this trend is expected to persist as Airbnb and hosts continue to introduce categories catering to traveler needs. The preference for space and privacy is also evident, with significant importance assigned to complete properties featuring bedrooms and bathrooms. This reflects a desire for ample space and privacy, which might impact the way properties are developed and upgraded in sought-after Airbnb locations as hosts endeavor to cater to this increasing demand.

5. Conclusion

Overall, this study provides an in-depth analysis of the impact of multiple factors on Airbnb pricing through linear regression and decision tree modeling, revealing the complexity of pricing strategies in the short-term rental market. It provides a possible reference for landlords to optimize their pricing strategies, understand market dynamics, and forecast.

Based on our model, we found that the impact of room type and room capacity on Airbnb pricing is significant, with whole homes or apartments priced much higher than private or shared rooms, reflecting a strong consumer preference for privacy and personal space. Also, for the room capacity (accommodates), as the number of people a room can accommodate increases, the log price (correspondingly, the actual home price) tends to increase. Besides, the rating of the hard furnishings of the Airbnb home, such as the number of bathrooms and bedrooms, is also positively correlated with prices. Higher prices usually accompany strict cancellation policies. If the cancellation policy is not flexible enough, the landlords may compensate for potential booking losses through higher pricing.

In terms of pricing strategies, our research shows that landlords should first consider offering a complete house or apartment to maximize rental income. Additionally, investing in the hard furnishings of the home, such as adding bathrooms or enhanced amenities, can also increase rental prices. At the same time, emphasizing unique features and high-quality services can help garner better reviews, which can further increase the appeal and pricing of a listing. Understanding the interplay between pricing and cancellation policies can help landlords strike a balance between attracting bookings and maximizing revenue.

Specifically from the base model of this paper, we suggest that housing operators can use the key variables identified in the model, such as room type, cancellation policy, and quality of amenities, to optimize pricing and make continuous adjustments and upgrades to the model by going over actual operational data. In detail, during peak travel seasons or surges in demand in specific areas,

DOI: 10.54254/2754-1169/2024.25024

operators can adjust their pricing strategy based on the predictive model to maximize revenue. These data-driven decisions will help landlords maintain an advantageous position in a competitive market.

Regarding the limitations of this study's findings, the dataset may suffer from regional biases that do not adequately reflect the diversity of global markets. Besides, the models lacked sufficient training and debugging due to constraints related to time factors and research access, which may reduce the accuracy and usability of the models. Therefore, this study only provides possible suggestions for Airbnb owners to consider the determinants of house prices and does not involve specific pricing models. Continuing to conduct more detailed analyses for different regions, taking into account the impact of regional economic conditions, tourism seasonality, local competition, and other factors on pricing, as well as collecting and analyzing data over long time horizons and exploring long-term trends and changes in pricing strategies, would be a possible next step for this study.

Acknowledgment

Ran Ji and Xuting Huang contributed equally to this work and should be considered co-first authors.

References

[1] Quattrone, G., Kusek, N., & Capra, L. (2022). A global-scale analysis of the sharing economy model – an Airbnb case study. EPJ Data Science, 11(1). https: //doi.org/10.1140/epjds/s13688-022-00349-3

[2] Barron, K., Kung, E., & Proserpio, D. (n.d.). The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb *. https: //marketing.wharton.upenn.edu/wp-content/uploads/2019/08/09.05.2019-Proserpio-Davide-Paper.pdf

[3] Hastie, T., Tibshirani, R., & Friedman, J. H. (2004). The elements of statistical learning: data mining, inference, and prediction. Springer.

[4] Interpreting log transformations in a linear model. Interpreting Log Transformations in a Linear Model | UVA Library. (n.d.). https: //library.virginia.edu/data/articles/interpreting-log-transformations-in-a-linear-model

[5] Coding System for Categorical Variables in Regression Analysis. UCLA: Statistical Consulting Group. from https: //stats.oarc.ucla.edu/spss/faq/coding-systems-for-categorical-variables-in-regression-analysis/ (accessed August 4, 2024).

[6] Malakouti, S. (2023). Babysitting hyperparameter optimization and 10-fold-cross-validation to enhance the performance of ML methods in predicting wind speed and energy generation. Intelligent Systems with Applications, 19, 200248. https: //doi.org/10.1016/j.iswa.2023.200248

[7] Corbo, A. (2023, January 23). What Is a Decision Tree? (Definition, When to Use). Built In. https: //builtin.com/machine-learning/decision-tree#: ~: text=Decision%20trees%20combin e%20multiple%20data%20points%20and%20weigh