forecasting

MDPI

Article

Identifying Key Attributes Associated with Short-Term Rental Occupancy Rates: Case of Airbnb

Sanelisiwe Amanda Nkomo, Ehsan Ahmadi * and Reza Maihami

Stetson-Hatcher School of Business, Mercer University, Atlanta, GA 30341, USA; sanelisiwe.amanda.nkomo@live.mercer.edu (S.A.N.); maihami_r@mercer.edu (R.M.)
* Correspondence: ahmadi_e@mercer.edu

Highlights

What are the main findings?

- Price, location (city), listing capacity, amenities, and host responsiveness were among the variables most strongly associated with occupancy rate variation.
- Moderately priced listings with smaller accommodation capacity exhibited the highest average occupancy rates (approximately 57%), and this pattern was supported by both the Random Forest and cluster analyses.

What are the implications of the main findings?

- Listing and host characteristics should be considered together when evaluating short-term rental performance.
- Future studies may benefit from incorporating seasonality, neighborhood characteristics, local events, and other market conditions to better explain occupancy rate variation.

Abstract

Short-term rental housing plays an important role in the housing market by increasing property utilization and generating income opportunities for property owners. This study investigates the key attributes associated with Airbnb occupancy rates using listing data from five major U.S. cities. Data mining and machine learning techniques, including Random Forest, XGBoost, Deep Neural Networks (DNN), and hierarchical cluster analysis, were applied to identify factors associated with occupancy rate variation. Random Forest achieved the best performance (R^2 = 23.92%). Feature importance analysis identified price, location (city), listing capacity, amenities, and host response rate as the variables most strongly associated with occupancy rates. Cluster analysis supported these findings by identifying a dominant group of moderately priced listings with smaller accommodation capacity, more amenities, higher host responsiveness, and an average occupancy rate of 57%. These findings provide data-driven insights that may help property owners optimize listing performance and improve occupancy rates.

Keywords: Airbnb; short-term rental housing; occupancy rate; property management; machine learning; cluster analysis

Academic Editor: Alessia Paccagnini

Received: 27 April 2026
Revised: 26 June 2026
Accepted: 14 July 2026
Published: 18 July 2026

1. Introduction

Short-term rental housing refers to property leases that range from one night to several weeks but not more than a month [1]. The short-term rental market is expected to

 https://doi.org/10.3390/forecast8040060

grow between 2025 and 2030 at an average annual rate of 11.4%. Due to evolving travel patterns and customer preferences, the global short-term rental market was valued at USD 134.51 billion in 2024 and is expected to grow to USD 256.31 billion by 2030, with the North American market holding a share of 34.6% of the global revenue in 2024 [2]. Airbnb has a notable presence in the United States, with listings available in major cities across the country. The U.S. locations with the highest numbers of Airbnb listings per 1000 inhabitants include Hawaii, New Orleans, Austin, Nashville, Washington DC, Seattle, San Diego, San Francisco, Denver, Portland, Las Vegas, Boston, Los Angeles, and New York [3]. Some cities in the United States have passed laws that have severe restrictions on Airbnb listings, while most cities with these restrictions remain the top locations for Airbnb listings in the country [4]. The popularity of short-term rentals has grown, attracting business professionals, travelers, and individuals in need of temporary accommodation [5]. One of the main advantages of short-term rental housing is its flexibility: guests can choose a property that suits their needs and preferences, while owners retain control over pricing. Short-term rentals also create new income opportunities for homeowners and stimulate growth in service industries [6].

For short-term rental property owners, occupancy rate is one of the most important indicators of business performance because it directly influences revenue generation and asset utilization. While hosts can control factors such as pricing, amenities, property features, and responsiveness to guests, it remains unclear which of these attributes are most strongly associated with occupancy outcomes. Understanding these relationships is important for property owners seeking to improve listing performance and for platform managers interested in supporting efficient marketplace operations. As the short-term rental market continues to expand, identifying the key factors associated with occupancy rates has become both a practical and research-relevant challenge.

After COVID-19, Airbnb experienced significant growth as travel and tourism rebounded, with a notable increase in domestic and international bookings [7,8]. By saving on accommodation, travelers may allocate more of their budget to spending in local neighborhoods. The rise of short-term rental platforms like Airbnb has significantly impacted local economies by increasing tourism revenue and diversifying tourism offerings. It has been particularly valuable in areas with limited traditional hotel options, helping drive economic growth by attracting and sustaining tourism. However, while short-term rentals offer economic advantages, they also raise concerns about housing affordability and market balance [9]. As more housing units are converted to short-term rentals, the reduced availability of long-term rental properties contributes to rising rents and house prices [10].

However, some studies argue that this effect is not driven by Airbnb itself, since its listings can expand the supply of short-term rental units without necessarily reducing the long-term rental stock. This study aims to identify the key attributes associated with short-term rental occupancy rates and to better understand the factors that contribute to improved listing performance and guest experience. This study adopts a data-driven approach using data mining and machine learning techniques, including Random Forest, XGBoost, and cluster analysis, to examine Airbnb listing data. The methodology involves collecting and preprocessing Airbnb listing data and applying machine learning and cluster analysis techniques to identify patterns and examine the attributes associated with occupancy rate variation. Although the study faced limitations due to incomplete data and restricted access to proprietary Airbnb datasets, the findings offer a framework to understand short-term rental trends in order to improve occupancy rates. Ultimately, the study aims to provide valuable strategies for property owners to optimize their listings, which contributes to a more sustainable and thriving short-term rental market.

https://doi.org/10.3390/forecast8040060

2. Literature Review

2.1. Demand and Occupancy Rates

The rapid growth of short-term rental homes has greatly boosted the supply of accommodation, influenced by broader macroeconomic conditions, making the overall lodging industry more competitive [11]. Estimating demand for short-term rental homes involves analyzing historical booking data, market surveys, and broader market trends. The key determinants that influence demand include local attractions, seasonality, events, economic conditions, awareness of sharing economy platforms, and competitive pricing. Host pricing strategies influence guest satisfaction, which in turn affects long-term demand and revenue [12]. External factors like weather, travel restrictions, and consumer preferences also play a role in shaping demand. Guest evaluations, location, accommodation features, management and host service quality impact Airbnb satisfaction, which affects demand [13]. Demand is difficult to measure objectively because it encompasses all inquiries and potential bookings, whether or not they result in actual reservations [14].

Constrained demand is a subset of total demand that reflects the actual booked reservations, excluding inquiries that did not result in bookings. Constrained demand can be objectively measured because it represents confirmed bookings. However, few studies have explicitly quantified this measure. Demand and occupancy are influenced by factors like price, location, cleanliness, facilities, and security. While market demographics also play a role in determining occupancy rates [15], short-term rental occupancy is typically measured by the number of days a property is booked over a given period of time, i.e., within 30 days [16]. In this study, occupancy rate is estimated using the *availability_30* variable, which represents the number of days a listing is available for booking during the next 30 days. Occupancy rate reflects the proportion of available nights that are booked.

2.2. Host Efforts and Property Features

Past studies have focused on Airbnb's affordability, with the price attribute often regarded as a significant factor influencing customers' choice of a listed property [17,18]. Short-term rentals, particularly those listed on Airbnb, highlight the significant role that location, amenities, and reviews play in whether a listing is booked. Proximity to tourist attractions, restaurants, bars, and transportation hubs are highlighted as major influencing factors for potential guests [19]. The provision of amenities is often considered a standard expectation for guests and Airbnb has guidelines on their website for what amenities are considered essential or basic. Inclusion of unique or additional features can significantly differentiate a listing [20]. Hosts are therefore encouraged to invest in thoughtful amenities that go beyond the basics, creating a richer experience for their guests. The property capacity or the number of guests a property can accommodate, and the quality of the sleeping arrangements play a pivotal role in guest satisfaction. Properties with more bedrooms or beds are more likely to yield higher returns when rented as short-term rentals [21].

Hosts that prioritize comfort and offer diverse accommodation options are more likely to be appealing to a wider range of guests. A secure verification process is a critical factor that ensures trust and security in the short-term rental platforms and systems [22]. The host's profile photo also influences whether a listing is booked. Negative expressions or absence of a facial image reduce customer interest, while positive or neutral expressions help to enhance it [23]. The verification process is done using identity checks via photo, ID, phone number, email, and address, which is designed to foster a secure environment and build trust between users [24]. Hosts who complete their profiles with descriptions, have positive reviews, and achieve Superhost status further reinforce their credibility [25,26].

https://doi.org/10.3390/forecast8040060

2.3. Occupancy Drivers: Attribute Association Insights

The rise of data analytics and machine learning has transformed how businesses make decisions. In this data-driven environment, companies increasingly use sophisticated algorithms to convert raw data into actionable insights that shape business strategy and support better decision-making. Dhillon et al. [27] used multiple machine learning techniques for Airbnb price predictions and analyzed the models with higher accuracy. Studies on Airbnb often focus on predicting pricing and understanding the factors that influence the price of a listing. These models forecast price responsiveness by considering variables such as location, seasonality, attractions, and property features [28]. Hosts frequently adjust prices based on seasonal demand, location, events, competitor pricing and booking trends. This makes the price a moving variable and not a static feature. Airbnb datasets are sourced from open-source data repositories such as Inside Airbnb [29], which often contain static data that does not capture time series or seasonal pricing variations.

Although prior literature concentrates on price prediction, more accurate results could be obtained by incorporating the factors that drive prices, including location and seasonality. Amenities offered in a listing can significantly influence a guest's decision to book. Additionally, the purpose of the trip, whether for tourism or business, affects which amenities guests consider most important [21]. Customer reviews and host descriptions come in the form of text data. The number of reviews by a customer on a listing is an indicator of trust and reliability for potential guests. A high volume of positive reviews may significantly boost a listing's attractiveness. The listing description written by hosts helps shape a guest's perception of the property. Engaging language can increase the appeal of a listing and lead to higher occupancy rates [30]. Sentimental analysis of customer reviews and host descriptions has helped both hosts and Airbnb optimize user experience and improve business outcomes. Analyzing review data, ratings, and host descriptions helps to gain insights into guest preferences.

2.4. Research Gaps

Previous research has focused on price prediction and sentiment analysis by understanding customer experience through analyzing Airbnb review data and host descriptions. In contrast, relatively little attention has been paid to the key factors associated with changes in short-term rental occupancy, leaving a gap in understanding their influence. This study aims to identify the key attributes that influence occupancy rates by quantifying property performance and identifying attributes that affect a listing's occupancy. Table 1 is a list of the literature obtained from previous research on the various attributes that measure or contribute to the success of Airbnb. The literature has explored each attribute, such as price, amenities, location, and sentiment, and how each impacts Airbnb listings, but there remains a lack of integrative studies that directly examine how these attributes collectively influence Airbnb occupancy.

Table 1. Attributes contributing to Airbnb performance and related literature.

Attribute	Literature
Price	[17,18,28,31–33]
Amenities	[20,21,34]
Location	[13,19,35]
Sentiment Analysis	[30,36–38]
Host Practices & Verification	[22–26]
Reviews & Trust	[25,26,30]
Demand & Occupancy Determinants	[11,12,14,15]

Most research focuses on pricing strategies or consumer behavior, whereas relatively few studies have examined the factors associated with occupancy performance, particularly using real-world booking or availability data. This study utilizes two ensemble methods (Random Forest and XGBoost) and unsupervised machine learning (Cluster Analysis) to analyze the Airbnb listing data. The Airbnb dataset is characterized by high dimensionality, heterogeneous features, and complex nonlinear relationships, which can be difficult to model using traditional statistical approaches. Random Forest and XGBoost are ensemble machine learning methods that combine the outputs of multiple decision trees to capture complex relationships within the data. To provide a comparative assessment, a Deep Neural Network (DNN) was also evaluated alongside Random Forest and XGBoost. The DNN was included to examine whether a deep learning approach could capture additional nonlinear relationships and interactions within the Airbnb dataset. In addition, cluster analysis is used to identify underlying patterns and groupings within the dataset without relying on predefined labels. In this study, cluster analysis is employed to uncover relationships among listing attributes and determine which characteristics are most strongly associated with higher occupancy rates.

3. Data

Airbnb does not officially provide open-access datasets for general use. Therefore, this study used publicly available data obtained from the Inside Airbnb project, which provides Airbnb listing data for cities worldwide. The datasets used in this study were downloaded from the Inside Airbnb repository (https://insideairbnb.com/get-the-data/ (accessed on 10 November 2025)). The analysis focuses on U.S. cities with the highest Airbnb listing densities [3]. Table 2 presents the cities with the greatest number of Airbnb listings per 1000 inhabitants. These data provide insights into the distribution of Airbnb properties relative to local population size, offering a comparative perspective on short-term rental density across major urban and tourist destinations in the United States.

Table 2. Airbnb listings per 1000 inhabitants in selected U.S. cities [29].

City	State	Total Listings
–	Hawaii	22.4
New Orleans	Louisiana	18.8
Austin	Texas	15.3
Nashville	Tennessee	13.0
Washington, D.C.	District of Columbia	9.5
Seattle	Washington	9.0
San Diego	California	8.8
San Francisco	California	8.5
Denver	Colorado	7.5
Portland	Oregon	7.3
Las Vegas	Nevada	6.9
Boston	Massachusetts	5.9
Los Angeles	California	5.8
New York City	New York	4.9

The selected cities represent a diverse mix of tourism, business, and administrative destinations within the United States. The analysis focuses on five U.S. cities: Austin, New Orleans, New York City, Denver, and San Francisco. Each city dataset contains 75 variables, while the number of observations varies across locations. Austin includes 14,861 listings, New Orleans 56,783 listings, New York City 39,453 listings, Denver 5388 listings, and San Francisco 7418 listings. The dataset contains information on host characteristics, listing details, pricing, guest engagement metrics (e.g., reviews), and property features. It is heterogeneous in nature, comprising numerical, categorical, and text-based variables. Amenities are recorded in a JSON (JavaScript Object Notation) format, allowing detailed information on listing features to be incorporated into the analysis.

3.1. Data Cleaning and Dimension Reduction

The data-cleaning process involved merging the datasets from the five cities into a single dataset, making it a total of 73,903 observations and 75 variables. Variables with over 70% missing values, URLs (Uniform Resource Locators), and redundant and irrelevant data were removed. Listings with a minimum stay of over 10 days (mid-term and long-term stays) were excluded from this study. Special characters such as "%" and "$" were removed from numeric variables (e.g., host response rate and price) to facilitate analysis. Boolean variables were converted into binary indicator variables, while categorical variables with multiple levels were transformed into factors. To reduce the influence of extreme values, outliers in the price variable were removed by excluding listings with prices below $50, above $750, or more than three standard deviations from the mean. Figure 1 illustrates a right-skewed distribution of listing prices, indicating that most properties are concentrated in the lower price range, while a smaller number of high-priced listings create a long right tail. As a result, the mean price exceeds the median, as shown in the descriptive statistics presented in Table 3. A logarithmic transformation was applied to the price variable, and Figure 2 shows how the transformation makes the distribution more symmetric, which brings the mean closer to the median. This approach was chosen to stabilize variance and mitigate the influence of extreme outliers, ensuring the price distribution was more suitable for model performance.

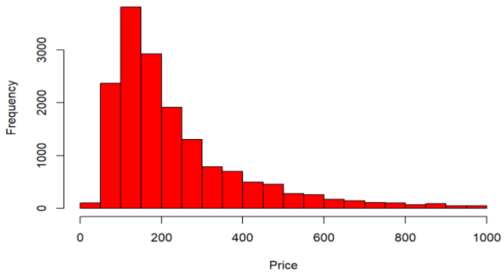

Figure 1. Price distribution.

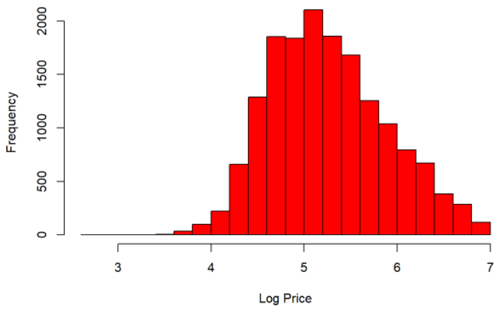

Figure 2. Log (Price).

The amenities variable has 6464 unique items stored in a JSON format. Text-mining techniques were applied to extract and process amenity information, resulting in 229 binary variables representing the presence or absence of individual amenities within each listing. This transformation enabled a more effective analysis of amenity-related features. To improve model efficiency, redundant variables were removed, while key listing characteristics and host-related attributes were retained for further analysis.

Following data cleaning and preprocessing, the dataset contained 16,182 observations. Additional feature selection and dimensionality reduction procedures were then applied to prepare the data for model development.

Table 3. Summary statistics of Airbnb dataset.

Attribute	N	Mean	Std	Min	Pctl. 25	Pctl. 50	Pctl. 75	Max
accommodates	16,182	5	3	1	3	4	6	16
bathrooms	16,182	2	1	0	1	1	2	9
bedrooms	16,182	2	1	1	1	2	3	13
beds	16,182	3	2	1	2	2	3	26
price	16,182	237	170	15	120	181	293	999
minimum_nights	16,182	2	1	1	1	2	2	9
has_availability	16,182	1	0	0	1	1	1	1
availability_30	16,182	14	10	0	5	13	22	30
number_of_reviews	16,182	71	104	1	11	34	85	1834
reviews_per_month	16,182	2	2	0	1	2	3	80
review_scores_rating	16,182	5	0	1	5	5	5	5
review_scores_accuracy	16,182	5	0	1	5	5	5	5
review_scores_cleanliness	16,182	5	0	1	5	5	5	5
review_scores_check-in	16,182	5	0	1	5	5	5	5
review_scores_communication	16,182	5	0	1	5	5	5	5
review_scores_location	16,182	5	0	1	5	5	5	5
review_scores_value	16,182	5	0	1	5	5	5	5
instant_bookable	16,182	0	0	0	0	0	1	1
host_response_rate	16,182	1	0	0	1	1	1	1
host_acceptance_rate	16,182	1	0	0	1	1	1	1
host_is_super_host	16,182	1	0	0	0	1	1	1
host_listings_count	16,182	32	162	0	1	2	10	2766
host_total_listings_count	16,182	57	374	1	2	4	14	9176
host_has_profile_pic	16,182	1	0	0	1	1	1	1
host_identity_verified	16,182	1	0	0	1	1	1	1

3.2. Variable Selection

Variable selection was performed to reduce model complexity, improve interpretability, and minimize redundancy among predictors [39]. The process involved feature engineering, correlation analysis, and stepwise selection procedures to identify the variables most relevant to occupancy rate analysis.

To support the modeling objectives, two derived variables were created: *occupancy rate* and *amenities count*. *Occupancy rate* was selected as the target variable and was calculated from the *availability_30* variable, which represents the number of days a listing is available for booking during the next 30 days in the Inside Airbnb dataset. *Occupancy rate* was then estimated as the percentage of days booked during that 30-day period:

$$\text{Occupancy Rate} = \left(\frac{30 - \text{availability}_{30}}{30}\right) \times 100$$

Amenities count represents the total number of amenities available in a listing. This variable was derived from the text mining analysis of the amenities field, which generated 229 binary variables indicating the presence or absence of individual amenities. The *amenities count* was calculated by summing these binary variables for each listing.

Further dimension reduction was performed through correlation analysis by removing highly correlated variables in guest engagement attributes (reviews) and property features. Figure 3 presents the correlation matrix for guest-rating variables, including cleanliness, location, communication, check-in, and overall review score rating. The strong correlations among these variables suggest the presence of multicollinearity, indicating that they capture similar aspects of guest satisfaction and may provide overlapping information. Figure 4 presents the correlation matrix for property-related features. Strong correlations were observed among the variables *accommodates*, *beds*, and *bedrooms*, indicating that these variables capture overlapping information related to property capacity. To reduce multicollinearity and redundancy, *accommodates* was retained as the representative measure of property capacity. Similarly, *review_scores_rating* was retained as a summary measure of overall guest satisfaction. The remaining highly correlated variables were excluded from further analysis to improve model interpretability and performance. Stepwise variable selection was applied as a dimensionality reduction technique to identify the most relevant features based on statistical significance. This process helped simplify the model by removing less informative variables while retaining those that contributed most to explaining variation in the target variable [40]. The final dataset contained 42 variables and 16,182 observations.

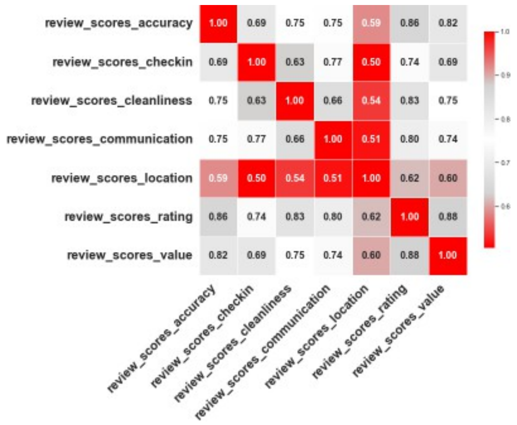

Figure 3. Correlation matrix of guest engagement attribute.

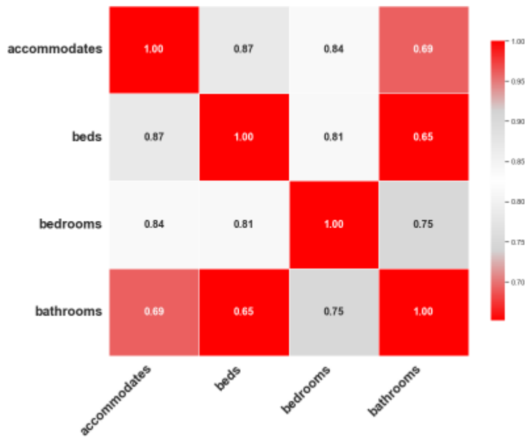

Figure 4. Correlation matrix of property features.

The Airbnb listing dataset was then integrated with the amenities dataset for model development. Based on the feature engineering, correlation analysis, stepwise selection process, and findings from prior studies, the final set of variables included city, essential amenities, price, log(price), host response time, host response rate, host acceptance rate, host is superhost, host listings count, host profile photo, host identity verification status, room type, accommodates, minimum nights, review scores value, instant bookable status, amenities count, occupancy rate, and the top 22 amenities identified from the text-mining analysis.

To facilitate data reproducibility, the complete preprocessing workflow consisted of downloading the raw Inside Airbnb datasets, merging the five city datasets, applying the data cleaning and preprocessing procedures described above, engineering the occupancy rate and amenity count variables, and performing feature selection to generate the final analytical dataset.

4. Methodology

This section outlines the methodology used for the analysis. The Airbnb dataset comprises heterogeneous data with numerical and categorical variables and complex feature interactions. Two tree-based ensemble methods, Random Forest and XGBoost, were selected as the primary analytical models to examine factors associated with occupancy rates. Deep Neural Network (DNN) was used as a benchmark model to compare the core models with more advanced machine learning algorithms and evaluate predictive performance. Agglomerative hierarchical clustering, a bottom-up clustering approach, was employed to identify natural groupings within the data and explore attributes associated with occupancy rates.

Figure 5 shows the modeling framework used for Random Forest and XGBoost. The cleaned dataset was partitioned into training (70%) and testing (30%) subsets. To ensure reproducibility of the results, fixed random seeds were used during data partitioning and model training. A random seed of 42 was used for the Random Forest and Deep Neural Network analyses, while a seed of 123 was used for the XGBoost analysis. Models were developed using the training dataset and evaluated using the testing dataset based on Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). The best-performing model was subsequently optimized through hyperparameter tuning and selected for further analysis.

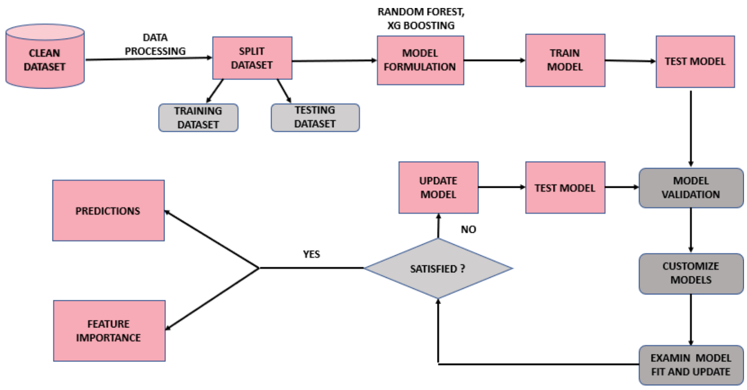

Figure 5. Proposed methodology: ensemble methods.

To complement the machine learning analysis, hierarchical clustering was applied to identify groups of listings with similar characteristics. The clustering methodology is shown in Figure 6. A dendrogram was used to determine the optimal number of clusters, and cluster profiles were examined by comparing group means to identify attributes associated with occupancy rates.

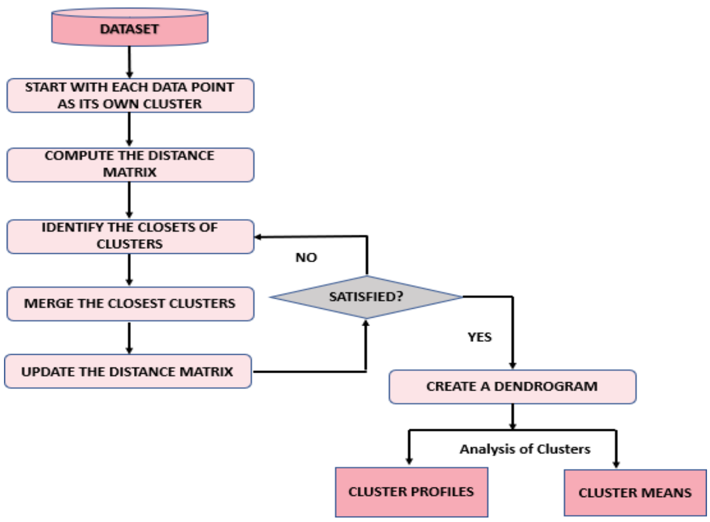

Figure 6. The proposed methodology: cluster analysis.

4.1. Model Assumption

Modeling assumptions were established to simplify the analysis and provide a consistent framework for evaluating occupancy rates. These assumptions help focus the analysis on key variables while improving model interpretability. The assumptions used in this study are as follows:

- Prices were analyzed as recorded in the dataset and were assumed to remain unchanged during the observed 30-day availability period.
- Listings with no availability during the 30-day period were assumed to be fully occupied.
- The analysis was limited to short-term rentals with a minimum stay requirement of 10 days or less.

4.2. Random Forest

Random Forest is a tree ensemble machine learning algorithm that is used for classification and regression tasks. It works by creating a forest of decision trees and merging their output to improve prediction accuracy [41,42]. Random Forest creates multiple decision trees by bootstrapping data and selecting random subsets of features at each split to promote diversity. This study uses a regression task for the continuous target variable occupancy rate; the prediction is the average of all tree outputs.

$$\hat{y} = \frac{1}{M}\sum_{m=1}^{M} f_m(x)$$

where:

- $\hat{y}$ is the predicted occupancy rate.
- M is the number of trees in the forest.
- $f_m(x)$ is the prediction from the m^{th} tree of the features.

To optimize the performance of the Random Forest model, hyperparameter tuning was conducted using Grid Search with five-fold cross-validation. The hyperparameter search included the number of trees (*n_estimators*) with values of 100, 300, and 500; the maximum tree depth (*max_depth*) with values of 10, 20, and 30; the minimum number of samples required to split an internal node *(min_samples_split*) with values of 5, 10, and 15; and the minimum number of samples required at a leaf node (*min_samples_leaf*) with values of 3, 4, and 5. The optimal Random Forest model selected by Grid Search used 500 trees, a maximum depth of 20, a minimum sample split of 10, and a minimum leaf size of 3. These hyperparameters were subsequently used for model evaluation and feature importance analysis.

4.3. XG Boosting

XGBoost (Extreme Gradient Boosting) is a machine learning algorithm that uses an ensemble of decision trees to improve prediction accuracy. It is based on gradient boosting, where models are trained sequentially, with each tree focusing on correcting errors made by the previous tree [43,44].

$$\hat{y} = \sum_{m=1}^{M} f_m(x)$$

where:

- $\hat{y}$ is the predicted occupancy rate.
- M is the number of trees in the model.
- $f_m(x)$ is the prediction from the m^{th} tree of the features.

The XGBoost hyperparameters were optimized using Grid Search with five-fold cross-validation. The hyperparameter search included the number of trees (*n_estimators*) with values of 100, 200, and 300; the maximum depth of trees (*max_depth*) with values of 3, 5, and 7; the learning rate (*learning_rate*) with values of 0.01, 0.1, and 0.2; the subsample ratio (*subsample*) with values of 0.7, 0.8, and 1.0; and the column sampling ratio (*colsample_bytree*) with values of 0.7, 0.8, and 1.0. The optimal XGBoost model selected by Grid Search used 300 trees, a maximum depth of 7, a learning rate of 0.01, a subsample ratio of 0.8, and a column sampling ratio of 1.0. The optimal parameter combination is selected based on the highest cross-validated R^2 score, resulting in improved predictive performance and generalization of the model.

4.4. Cluster Analysis

Hierarchical clustering is an unsupervised machine learning technique used to group similar objects, data points, or observations into clusters based on certain characteristics or features. In hierarchical clustering, each data point starts as its own cluster and pairs of clusters are progressively merged based on similarity. The process continues until all points are grouped into distinct clusters [45]. The methodology used for this study is Ward's method, a variance-minimizing approach to form well-separated clusters. The dataset was normalized using z-score standardization to re-scale data so that it has a mean of 0 and a standard deviation of 1. Euclidean distance matrix was used to compute the distance or similarity between the data points. A dendrogram was used to visualize how the data points merge into clusters at various similarity levels. The optimal clusters were identified by cutting the tree at a height where large gaps indicate natural divisions in the data. The formula for Z-score standardization is $z_i = \frac{x_i - \mu}{\sigma}$, where x_i is the original value, μ is the mean of the dataset, σ is the standard deviation of the dataset, and z_i is the standardized value (z-score).

The Euclidean distance between two points $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots,$ $y_n)$ in n-dimensional space is given by Distance $= \sqrt{\sum_{i=1}^{n} (x_\mathrm{i} - y_\mathrm{i})^2}$. Hierarchical clustering provides a clear visualization of how different attributes group together, making it easier to identify patterns in the data and to explore attributes that are associated with high occupancy rates. The methodology does not require pre-defining the number of clusters, offering flexibility in discovering natural groupings.

4.5. Deep Neural Networks Model

Artificial neural network (ANN) is a computational model inspired by the structure and functioning of the human brain, consisting of interconnected nodes (neurons) organized in layers. A Deep Neural Network (DNN) is a type of ANN that contains multiple hidden layers between the input and output, enabling it to learn complex patterns and hierarchical representations from data. Each layer transforms the input using weighted connections and nonlinear activation functions, improving the model's ability to capture intricate relationships [46]. In this study, DNN hyperparameters were optimized using a Grid Search procedure that evaluated alternative network architectures, dropout rates, learning rates, and batch sizes. The hidden-layer structures evaluated included (64, 32) and (128, 64, 32). Dropout rates of 0.2 and 0.3, learning rates of 0.001 and 0.0005, and batch sizes of 32 and 64 were also evaluated. The optimal model consisted of an input layer followed by three hidden layers containing 128, 64, and 32 neurons, respectively. Rectified Linear Unit (ReLU) activation functions were used in all hidden layers, while a linear activation function was used in the output layer for regression. Batch normalization layers were incorporated after each hidden layer, and a dropout rate of 0.3 was applied for regularization. The network was trained using the Adam optimizer with a learning rate of 0.0005 and a batch size of 32. Early stopping with a patience of 10 epochs was implemented to reduce overfitting, and the final model converged after 29 training epochs. Figure 7 illustrates the Deep Neural Network model process.

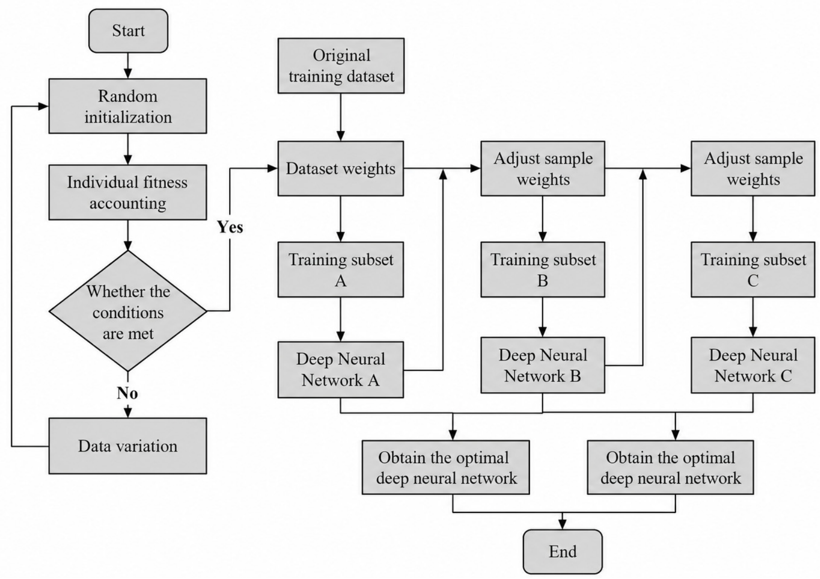

Figure 7. The proposed methodology: Deep Neural Network.

5. Results and Discussions

5.1. Model Performance Analysis

Table 4 is a summary of the model performance metrics. The performance of the Random Forest, DNN, and XGBoost models was evaluated using statistical metrics, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). These metrics offer a comprehensive view of the model's prediction accuracy for the continuous target variable. The DNN model was used as a challenger model to benchmark and test its performance against the traditional ensemble methodologies.

Table 4. Model performance comparison.

Model	MAE	MSE	RMSE	R^2
Random Forest	**22.99**	**783.07**	**27.98**	**23.92%**
XGBoost	23.44	794.14	28.18	21.37%
DNN	24.13	845.38	29.08	16.30%

The algorithm with the best performance is shown in bold.

The Random Forest model outperformed both the XGBoost and DNN models across all metrics, with the lowest MAE (22.99), MSE (783.07), and RMSE (27.98), and the highest R^2 (23.92%). Based on the complexity of the data, a lower MAE and RMSE indicate that the Random Forest predictions are closer to the actual values. R^2 of 23.92% means that the model explains 23.92% of the variation in the occupancy rates. Therefore, the Random Forest model is the best choice for this analysis due to its superior predictive accuracy and is used for further analysis of feature importance.

Figure 8 presents the residual distribution for the Random Forest model. The residuals are approximately symmetric and centered around zero, suggesting that the model does not exhibit substantial systematic bias and provides a reasonable fit to the data, although some unexplained variation remains.

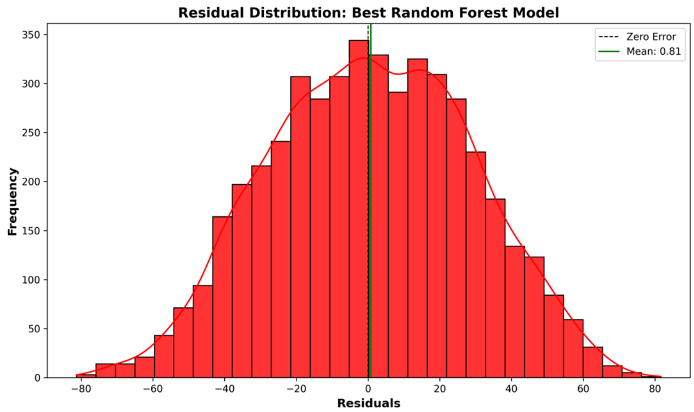

Figure 8. Residuals plot: Random Forest.

The feature importance plot in Figure 9 for the Random Forest model highlights the most influential variables in the prediction of occupancy rates. *Price* is the most significant feature, suggesting that pricing strategy plays a critical role in determining booking performance. This is followed by *review_scores_value* and *host_listings_count*, highlighting the importance of listing quality and host experience. Notably, location-based variables such as *city_New Orleans* and *city_New York* also rank highly, confirming that geographic factors and travel purpose significantly influence demand patterns. Variables such as *amenities_count*, *host_acceptance_rate*, *minimum_nights* and *accommodates* also contributed to occupancy rate variation, reflecting both listing attractiveness and booking flexibility. Lower-ranked features, including specific amenities, *host_is_superhost*, *host_identity_verified* and *room_type* have relatively small individual contributions, indicating that while they add value, they are less critical compared to core pricing, reputation, and location factors.

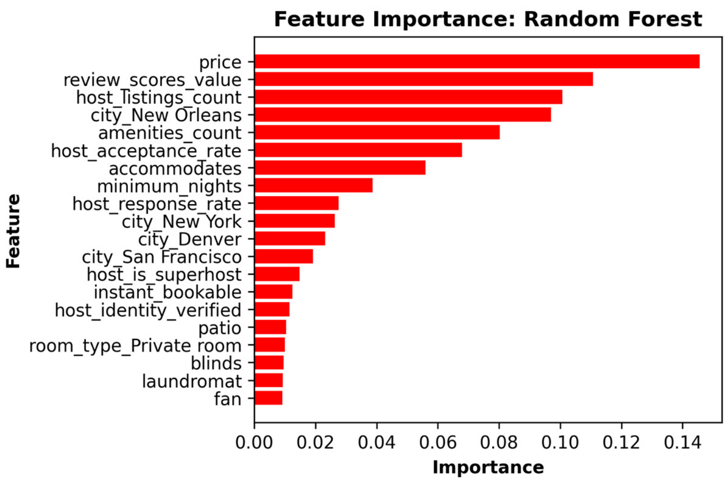

Figure 9. Feature importance: Random Forest.

5.2. Association Insights

This section presents the findings from the cluster analysis. The purpose of the analysis is to identify attributes in the listing dataset that are associated with high occupancy rates. Figure 10 shows a dendrogram which was used to visualize how the data points merge into clusters at various similarity levels. The optimal number of clusters was identified by cutting the tree at a height where large gaps indicate natural divisions in the data. The longest vertical lines that do not intersect with other lines represent the largest distances at which clusters are merged. In the case of the Airbnb listing data, the optimal number of clusters identified is three clusters.

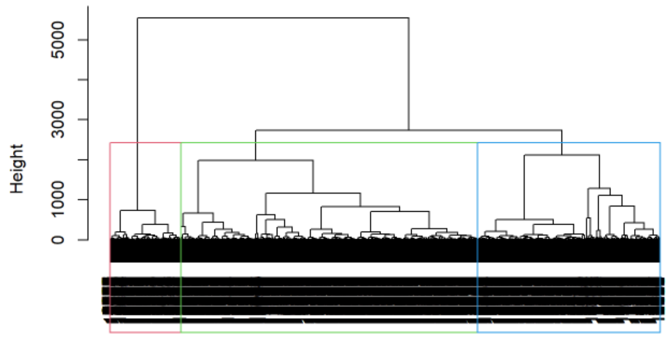

Figure 10. Dendrogram.

The heat map in Figure 11 illustrates the cluster profiles of the various features, representing differences in attributes across three distinct clusters. The clusters are color-coded to show the normalized feature values, ranging from −1 to 1. The warmer colors (red and beige) indicate higher values, where the variable values are relatively high. The cooler colors (blue and white) indicate lower values, where the variable values are relatively low. Table 5 is a summary of the cluster profiles highlighting unique characteristics across three distinct groups.

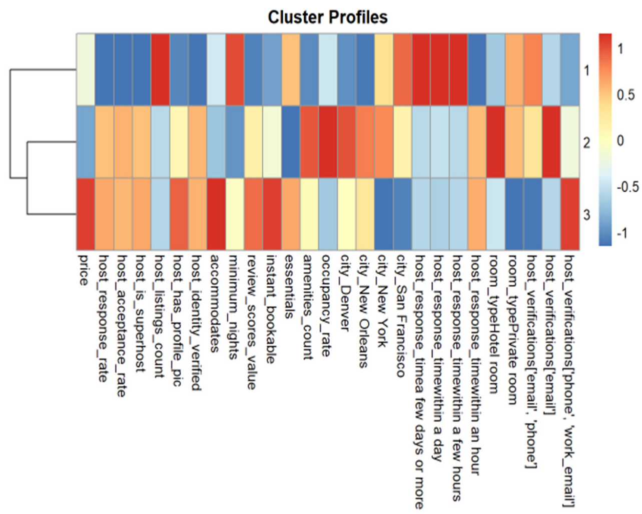

Figure 11. Heatmap showing the profiles of the identified clusters.

Cluster 1 exhibits casual and selective hosts who prioritize accepting only select guests and maintaining a less automated hosting approach. The listings require longer stays (high minimum nights) and are not instantly bookable, with lower acceptance and response rates. Cluster 1 is characterized by relatively low occupancy rates. Listings in this cluster are more commonly associated with private rooms, stricter minimum-night requirements, and a limited set of essential amenities. Occupancy rates and prices remain moderate, indicating that these characteristics are associated with lower utilization compared with the other clusters.

Cluster 2 is characterized by high occupancy rates, moderate pricing, short minimum-night requirements, and smaller accommodation capacity. The host prioritizes faster turnover with faster verifications done over the phone or by prompt email responses. Cluster 2 is the largest group in terms of the number of observations, suggesting that the hosts who dominate the Airbnb landscape aim for high-volume turnovers with moderate prices, also providing more amenities rather than just Airbnb essentials. Listings in this cluster are distributed across all major cities included in the study, including New York City, San Francisco, New Orleans, and Denver, suggesting that this listing profile is prevalent across multiple markets.

Table 5. Cluster profiles.

Cluster	Observations	High	Low
Cluster 1: Casual & Selective Hosts	755	*Email and Phone Verifications, Minimum Nights, Room type—Private Room Host Listing Counts, Response Time (Day, Few Hours, Few Days), Essentials, city—New York and city—San Francisco*	*Host Acceptance Rates, Host Response Rate, Instant Bookable, Host is Super Host, Occupancy Rates, Amenities, Prices*
Cluster 2: Moderately Priced, Frequently Booked Listings	12,917	*Phone and email Verifications, Private, Room & Hotel Room, Response within an hour* (moderate), *city—New York and city—San Francisco, city—New Orleans and city—Denver, Amenities, Occupancy Rates, Host Identity Verified, Response Rates, Host is Super host, Review Scores Value*	*Minimum Nights, Prices, Capacity (Accommodates), Price* (low to moderate)
Cluster 3: Premium & Professional Hosts	2510	*Prices, Capacity (Accommodates), Host Verifications (Work Email and Phone), Host Identity Verified, Response Rates, Host has profile picture, Review Score Value, Instant Bookable, Essentials, Host Identity Verified, Response Rates, Host is Super host, city—New Orleans* and *city—Denver*	*Occupancy Rates, Host Verifications (Email and Phone), city—New York* and *city—San Francisco*

Cluster 3 is characterized by hosts with verified identities, high response rates, and more complete profiles, including profile pictures and verification credentials. Listings in this cluster tend to accommodate more guests and have higher prices, representing the highest-priced segment among the identified clusters. These listings are also more likely to offer instant booking and receive higher review scores. Cluster 3 is most prevalent in New Orleans and Denver, indicating that these markets contain a greater concentration of listings with these characteristics.

Table 6 presents the cluster means, which summarizes the average values of the variables within each cluster. Cluster 2 contains 12,917 observations and represents the majority of listings in the dataset. This cluster is characterized by moderately priced accommodations, shorter minimum-night requirements, a broader range of amenities, and the highest average occupancy rate (57%) among the three clusters. Cluster 1 is characterized by moderate prices, stricter minimum-night requirements, and a more limited set of amenities. In contrast, Cluster 3 represents the highest-priced listings, with an average price of $465 per night. Listings in this cluster tend to accommodate more guests and are more likely to have verified host profiles, higher response rates, and instant-booking availability.

Table 6. Cluster means.

Attribute		Mean	Lower (95%CI)	Upper (95%CI)
Occupancy Rate	Cluster 1	47	45	50
	Cluster 2	57	56	57
	Cluster 3	46	45	47
Accommodates (capacity)	Cluster 1	5	5	5
	Cluster 2	4	4	4
	Cluster 3	9	9	10
Minimum nights	Cluster 1	2	2	2
	Cluster 2	2	2	2
	Cluster 3	2	2	2
Price	Cluster 1	286	273	299
	Cluster 2	190	188	192
	Cluster 3	465	456	474
Amenities	Cluster 1	38	37	39
	Cluster 2	39	39	39
	Cluster 3	39	38	39

6. Discussion

This study examined factors associated with Airbnb occupancy rates using data-driven analytical methods. As shown in Table 4, Random Forest achieved the best overall performance among the models evaluated, with the lowest MAE (22.99), MSE (783.07), and RMSE (27.98), and the highest R^2 value (23.92%). These results indicate that the Random Forest model explained a greater proportion of the variation in occupancy rates than the Neural Network (16.30%) and XGBoost (21.37%). However, the relatively low R^2 values suggest that a substantial portion of the variation in occupancy rates remains unexplained.

The relatively modest explanatory power of the models suggests that additional factors not included in the dataset may contribute to variation in occupancy rates. Variables such as seasonality, local events, competition, and other market conditions may provide further insight into booking patterns. The feature importance analysis identified *price*, *city*, *listing capacity*, *amenities*, and *host response rate* as the strongest predictors of occupancy rate variation. In contrast, host characteristics such as *superhost status* and *identity verification* were less influential in the model. These findings suggest that listing attributes and location-related factors may play a larger role in explaining occupancy rate variation than certain host characteristics.

The cluster analysis complements these findings by identifying listing profiles that share characteristics associated with higher occupancy rates. As shown in Tables 5 and 6, Cluster 2 represents the largest group in the dataset (12,917 observations) and exhibits the highest average occupancy rate (57%), exceeding the national Airbnb and Vrbo average occupancy rate of 54.3% reported by AirDNA [47]. Listings in this cluster are generally moderately priced, accommodate fewer guests, offer a broader range of amenities, and are associated with higher host response rates. These characteristics closely align with the variables identified as important in the Random Forest analysis, providing additional support for their association with occupancy rates. Together, the machine learning and clustering results suggest that pricing, listing capacity, amenities, location, and host responsiveness are consistently associated with occupancy rate variation across the analyzed dataset.

7. Conclusions

This study used data mining and machine learning techniques to identify the key attributes associated with short-term rental occupancy rates. Through detailed analysis and interpretation, the key insights highlighted that *price*, *location (city)*, *property capacity (accommodates)* and *host response time* were among the variables most strongly associated with occupancy rate variation. These findings were further supported by the cluster analysis, which identified higher-occupancy listing profiles characterized by moderate pricing, smaller accommodation capacity, a broader range of amenities, and higher host responsiveness.

The findings of this study contribute to the growing body of literature on factors associated with short-term rental occupancy rates. Consistent with previous research, the results suggest that pricing, location, listing capacity, amenities, and host responsiveness are important characteristics associated with occupancy rate variation. The cluster analysis further indicates that higher-occupancy listings tend to be moderately priced, accommodate fewer guests, offer a broader range of amenities, and have higher host response rates. In contrast, larger and higher-priced properties were not consistently associated with higher occupancy rates. These findings highlight the importance of considering multiple listing and host characteristics when evaluating short-term rental performance.

However, the relatively low explanatory power of the models suggests that additional factors, such as seasonality, local events, competition, and market conditions, may also play an important role in occupancy rate variation. For example, New Orleans, one of the cities included in this study, experiences strong tourism demand associated with cultural attractions, festivals, and seasonal events. According to available market statistics, listings in New Orleans achieve an average occupancy rate of approximately 58%, with seasonal demand peaks occurring during major tourism periods [48]. Such location-specific and temporal factors were not included in the present analysis and may partially explain the unexplained variation in occupancy rates. Future studies could improve model performance by incorporating neighborhood-level characteristics, seasonal demand indicators, and local market conditions.

Guest-related variables, including *host response rate*, *host verification status*, and *review score value*, were associated with occupancy rate variation in the analyzed dataset. These variables may provide useful information about listing quality, host engagement, and guest experiences. However, review-based metrics should be interpreted with caution, as they may reflect past occupancy patterns rather than independently explain occupancy rates. Figure 12 illustrates the relationship between *review score values* and *occupancy rates*. Most listings in the dataset received high review scores regardless of their occupancy rates, indicating limited variation in review ratings across listings. This concentration of positive ratings may reduce the ability of review scores to distinguish between higher- and lower-performing listings. Furthermore, *review score values* summarize overall guest evaluations but may not fully capture the context contained in written reviews. As a result, important information about guest experiences may be overlooked when relying solely on numerical review ratings [49].

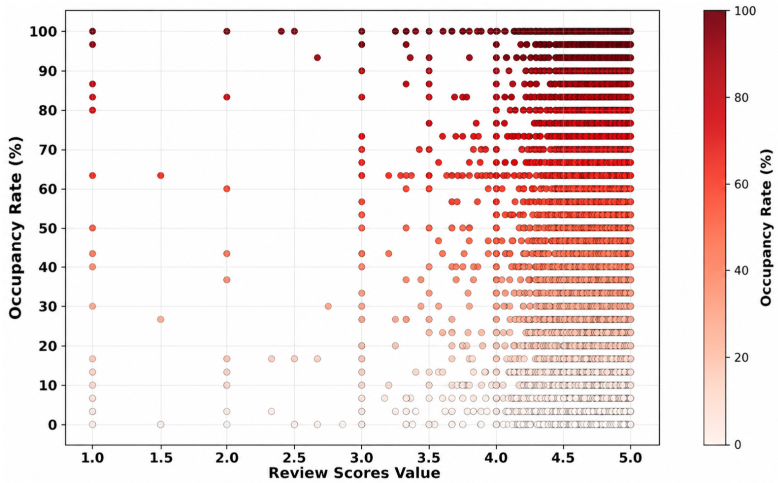

Figure 12. Review score rating analysis.

Future research could examine the relationship between written reviews and numerical review ratings to better understand how guest feedback is represented on short-term rental platforms. Natural Language Processing (NLP) techniques, such as sentiment analysis, could be used to assign sentiment scores to written reviews and compare them with the corresponding numerical ratings. This approach may help identify inconsistencies between review text and assigned ratings, providing additional insight into guest experiences and the reliability of review-based measures. Advanced language models, such as Bidirectional Encoder Representations from Transformers (BERT), may be particularly useful for capturing the context and sentiment expressed in detailed reviews.

Future studies could also investigate temporal aspects of guest feedback. While listings accumulate reviews over time, the total number of reviews may not accurately reflect current guest experiences. Incorporating time-sensitive measures, such as recent review activity or reviews per month, may provide a more current representation of guest perceptions and allow for fairer comparisons across listings with different ages and review histories.

Author Contributions: Conceptualization, S.A.N., E.A. and R.M.; methodology, S.A.N. and E.A.; software, S.A.N.; validation, E.A. and R.M.; formal analysis, S.A.N. and E.A.; investigation, S.A.N.; resources, S.A.N., E.A. and R.M.; data curation, S.A.N. and E.A.; writing—original draft preparation, S.A.N.; writing—review and editing, E.A. and R.M.; visualization, S.A.N.; supervision, E.A. and R.M.; project administration, E.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The Airbnb datasets analyzed in this study are publicly available from Inside Airbnb at https://insideairbnb.com/get-the-data/. The Python 3.11 scripts used for data cleaning, preprocessing, model development, hyperparameter tuning, evaluation, and visualization are publicly available at https://github.com/muresearchlab/Airbnb_Project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. De Jaureguizar Cervera, D.; Pérez-Bustamante Yábar, D.C.; de Esteban Curiel, J. Factors Affecting Short-Term Rental First Price: A Revenue Management Model. *Front. Psychol.* **2022**, *13*, 994910. [CrossRef] [PubMed]
2. Short-Term Vacation Rental Market. Available online: https://www.grandviewresearch.com/industry-analysis/short-term-vacation-rental-market-report (accessed on 11 September 2025).
3. Zandt, F. Which U.S. Cities Have High Airbnb Densities? Available online: https://www.statista.com/chart/30761/airbnb-listings-per-1000-inhabitants-in-the-us/ (accessed on 11 September 2025).
4. Hospitable Airbnb Restrictions: Cities with Strict Regulations | Hospitable. Available online: https://hospitable.com/airbnb-restrictions/ (accessed on 11 September 2025).
5. Furukawa, N.; Onuki, M. The Design and Effects of Short-Term Rental Regulation. *Curr. Issues Tour.* **2022**, *25*, 3245–3260.
6. Bekkerman, R.; Cohen, M.C.; Kung, E.; Maiden, J.; Proserpio, D. The Effect of Short-Term Rentals on Residential Investment. *Mark. Sci.* **2023**, *42*, 819–834. [CrossRef]
7. Türk, U.; Sap, S. The Effect of the COVID-19 on Sharing Economy: Survival Analysis of Airbnb Listings. *Bus. Manag. Stud. Int. J.* **2021**, *9*, 215–226. [CrossRef]
8. Armstrong, M. Airbnb Bookings Already Past Pre-Pandemic High. 2023. Available online: https://www.statista.com/chart/23520/number-of-nights-and-experiences-booked-on-airbnb/ (accessed on 11 September 2025).
9. Granicus Are Short-Term Vacation Rentals Contributing to the Housing Crisis? Available online: https://granicus.com/blog/are-short-term-vacation-rentals-contributing-to-the-housing-crisis/ (accessed on 11 September 2025).
10. Barron, K.; Kung, E.; Proserpio, D. The Sharing Economy and Housing Affordability: Evidence from Airbnb. *EC* **2018**, *5*, 1–62. [CrossRef]
11. Suess, C.; Mody, M.; Bulut, U.; Sirakaya-Turk, E. What Caused the Rise of Airbnb? An Examination of Key Macroeconomic Factors. *Tour. Manag.* **2020**, *81*, 104134. [CrossRef]
12. Carnehl, C.; Schaefer, M.; Stenzel, A.; Tran, K.D. Value for Money and Selection: How Pricing Affects Airbnb Ratings. *Innocenzo Gasparini Inst. Econ. Res. Work. Pap. Ser.* **2022**. [CrossRef]
13. Sutherland, I.; Kiatkawsin, K. Determinants of Guest Experience in Airbnb: A Topic Modeling Approach Using LDA. *Sustainability* **2020**, *12*, 3402. [CrossRef]
14. Lee, S. Study of Demand Models and Price Optimization Performance. Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA, USA, 2011.
15. Vasić, V.; Hristov-Stančić, B.; Zečević, B. Empirical Modeling of Hotel Occupancy Rate with Dynamic Panel Data. *Ekon. Preduz.* **2022**, *70*, 240–253. [CrossRef]
16. BNBFacts Airbnb Occupancy Rate: The Most Important Factors Ranked—BnB Facts. 2023. Available online: https://bnbfacts.com/airbnb-occupancy-rate-the-most-important-factors-ranked/ (accessed on 11 September 2025).
17. Zhang, Z.; Chen, R.J.C.; Han, L.D.; Yang, L. Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach. *Sustainability* **2017**, *9*, 1635. [CrossRef]
18. Lin, W.; Yang, F. The Price of Short-Term Housing: A Study of Airbnb on 26 Regions in the United States. *J. Hous. Econ.* **2024**, *65*, 102005. [CrossRef]

19. Guttentag, D.; Smith, S.; Potwarka, L.; Havitz, M. Why Tourists Choose Airbnb: A Motivation-Based Segmentation Study. *J. Travel Res.* **2018**, *57*, 342–359.
20. Sanchez-Franco, M.J.; Aramendia-Muneta, M.E. Why Do Guests Stay at Airbnb versus Hotels? An Empirical Analysis of Necessary and Sufficient Conditions. *J. Innov. Knowl.* **2023**, *8*, 100380. [CrossRef]
21. Tafesse, W.; Tariq, U. Airbnb Amenities: Quantifying Their Contribution to the Guest Experience. *Ann. Tour. Res. Empir. Insights* **2025**, *6*, 100167. [CrossRef]
22. Akhmedova, A.; Vila-Brunet, N.; Mas-Machuca, M. Building Trust in Sharing Economy Platforms: Trust Antecedents and Their Configurations. *Internet Res.* **2021**, *31*, 1463–1490. [CrossRef]
23. Fagerstrøm, A.; Pawar, S.; Sigurdsson, V.; Foxall, G.R.; Yani-de-Soriano, M. That Personal Profile Image Might Jeopardize Your Rental Opportunity! On the Relative Impact of the Seller's Facial Expressions upon Buying Behavior on Airbnb™. *Comput. Human Behav.* **2017**, *72*, 123–131. [CrossRef]
24. Ert, E.; Fleischer, A.; Magen, N. Trust and Reputation in the Sharing Economy: The Role of Personal Photos in Airbnb. *Tour. Manag.* **2016**, *55*, 62–73. [CrossRef]
25. Sengupta, P.; Biswas, B.; Kumar, A.; Shankar, R.; Gupta, S. Examining the Predictors of Successful Airbnb Bookings with Hurdle Models: Evidence from Europe, Australia, USA and Asia-Pacific Cities. *J. Bus. Res.* **2021**, *137*, 538–554. [CrossRef]
26. Liang, S.; Schuckert, M.; Law, R.; Chen, C.-C. Be a "Superhost": The Importance of Badge Systems for Peer-to-Peer Rental Accommodations. *Tour. Manag.* **2017**, *60*, 454–465. [CrossRef]
27. Dhillon, J.; Eluri, N.P.; Kaur, D.; Chhipa, A.; Gadupudi, A.; Eravi, R.C.; Pirouz, M. Analysis of Airbnb Prices Using Machine Learning Techniques. In *Proceedings of the 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 27–30 January 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 297–303.
28. Samwel, M. What Factors Drives the Airbnb Listing's Prices? *Int. Bus. Econ. Stud.* **2022**, *4*, 26. [CrossRef]
29. Inside Airbnb. Available online: https://insideairbnb.com/get-the-data/ (accessed on 11 September 2025).
30. Martinez, R.D.; Carrington, A.; Kuo, T.; Tarhuni, L.; Abdel-Motaal, N.A.Z. The Impact of an AirBnb Host's Listing Description'sentiment'and Length on Occupancy Rates. *arXiv* **2017**, arXiv:1711.09196.
31. Islam, M.D.; Li, B.; Islam, K.S.; Ahasan, R.; Mia, M.R.; Haque, M.E. Airbnb Rental Price Modeling Based on Latent Dirichlet Allocation and MESF-XGBoost Composite Model. *Mach. Learn. Appl.* **2022**, *7*, 100208. [CrossRef]
32. Gyódi, K.; Nawaro, Ł. Determinants of Airbnb Prices in European Cities: A Spatial Econometrics Approach. *Tour. Manag.* **2021**, *86*, 104319. [CrossRef]
33. Kim, M.; Darban, M.; Lim, S.E. The Impact of Airbnb's Competitive Pricing on Traveler Spending at Destinations. *Int. J. Hosp. Manag.* **2025**, *124*, 103994. [CrossRef]
34. Janssens, B.; Bogaert, M.; Van den Poel, D. Evaluating the Influence of Airbnb Listings' Descriptions on Demand. *Int. J. Hosp. Manag.* **2021**, *99*, 103071. [CrossRef]
35. Han, L.; Xu, F.; Xu, F.; Hu, M.; Hou, G. Inside Airbnb Titles: The Location Choices of Listings from Chinese Hosts' Advertising Expressions. *Tour. Manag. Perspect.* **2023**, *48*, 101151. [CrossRef]
36. Sun, X.; Gui, L.; Cai, B. To Be Honest or Positive? The Effect of Airbnb Host Description on Consumer Behavior. *Decis. Support Syst.* **2024**, *181*, 114200. [CrossRef]
37. Gunter, U. What Makes an Airbnb Host a Superhost? Empirical Evidence from San Francisco and the Bay Area. *Tour. Manag.* **2018**, *66*, 26–37. [CrossRef]
38. Zhang, L.; Wei, W.; Line, N.D.; Cheng, Y. When Positive Reviews Backfire: The Effect of Review Dispersion and Expectation Disconfirmation on Airbnb Guests' Experiences. *Int. J. Hosp. Manag.* **2021**, *96*, 102979. [CrossRef]
39. Heinze, G.; Wallisch, C.; Dunkler, D. Variable Selection–a Review and Recommendations for the Practicing Statistician. *Biom. J.* **2018**, *60*, 431–449. [CrossRef] [PubMed]
40. Zhang, Z. Variable Selection with Stepwise and Best Subset Approaches. *Ann. Transl. Med.* **2016**, *4*, 136. [CrossRef] [PubMed]
41. Salman, H.A.; Kalakech, A.; Steiti, A. Random Forest Algorithm Overview. *Babylon. J. Mach. Learn.* **2024**, *2024*, 69–79. [CrossRef]
42. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random Forests. In *Ensemble Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 157–175. [CrossRef]
43. Ali, Z.A.; Abduljabbar, Z.H.; Tahir, H.A.; Sallow, A.B.; Almufti, S.M. EXtreme Gradient Boosting Algorithm with Machine Learning: A Review. *Acad. J. Nawroz Univ.* **2023**, *12*, 320–334. [CrossRef]
44. Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
45. Nazari, Z.; Nazari, M.; Kang, D. A Bottom-up Hierarchical Clustering Algorithm with Intersection Points. *Int. J. Innov. Comput. Inf. Control* **2019**, *15*, 291–304.
46. Islam, M.; Chen, G.; Jin, S. An Overview of Neural Network. *Am. J. Neural Netw. Appl.* **2019**, *5*, 7–11. [CrossRef]
47. Gallagher, B. What's the Average Occupancy Rate on Airbnb? 2025. Available online: https://www.airdna.co/blog/average-occupancy-rate-airbnb (accessed on 11 September 2025).

48. Airbtics New Orleans Tourism Statistics: Insights from Airbnb User. Available online: https://airbtics.com/tourism-statistics-new-orleans-us/ (accessed on 11 September 2025).
49. Abrardi, L.; Raguseo, E.; Rondi, L. Kindness Capital and Rating Bias. A Sentiment Analysis on Airbnb Reviews. *A Sentim. Anal. Airbnb Rev.* **2022**. [CrossRef]