Uncovering Key Factors in Airbnb Guest Satisfaction:

A Data-Driven Analysis of Ratings

Yunzhen Wu

1. Executive Summary

The boom in short-term rentals has changed the way travelers choose accommodation. Platforms such as Airbnb dominate the industry. With increasing competition among hosts within the platform, understanding the factors that influence user preferences and ratings is essential to optimize listings and improve customer satisfaction. This study aims to uncover the key attributes that Airbnb users value most when choosing a rental property and how these attributes influence their ratings.

A deep exploratory data analysis (EDA) was conducted using a dataset of Airbnb listings in Melbourne, Australia to explore patterns behind features such as pricing, listing characteristics, and host interactions. Data preparation included handling missing values, resolving outliers, handling highly correlated variables, performing feature engineering, and applying machine learning techniques to extract meaningful insights.

In addition to EDA, machine learning models will be used to analyze how various listing attributes affect customer satisfaction and overall ratings. Using various models, the impact of key factors such as price, location, amenities, and host responsiveness can be quantified. Feature importance analysis further helps explore the most important variables that influence Airbnb ratings.

This study provides Airbnb hosts with actionable recommendations to help them attract more guests and increase revenue by focusing on improving important factors that influence guest satisfaction, such as cleanliness, amenities, location, and communication. The results of this study can also help Airbnb improve its recommendation algorithms and pricing strategies, ultimately enabling both hosts and guests to make informed decisions.

2. Problem Statement

2.1 Problem Statement

The Airbnb marketplace offers a wide variety of listings, but guests often face challenges in choosing the rental that best meets their expectations. At the same time, many hosts struggle to optimize their listings to attract guests and receive high ratings. Lack of clear understanding of factors that influence guest satisfaction can lead to low ratings, decreased occupancy, and lost revenue. While previous studies have compared Airbnb to traditional hotels, there is a need to analyze user preferences specifically in the Airbnb ecosystem to better align listing features with customer expectations.

2.2 Research Question

What are the **most important factors** Airbnb users care about when choosing short-term rentals, and how do these factors affect their ratings?

3. Literature Review

The sharing economy has disrupted every industry (Guttentag, 2015). Many companies are driving the sharing economy in different industries: Airbnb in accommodation, Uber in transportation, and Feastly in restaurants. These companies are attractive to consumers because they offer lower prices, better accessibility, greater flexibility, ease of use, and a “user-centric mission” that includes transparency and interactive communication (Clark, 2014).

Airbnb’s position and performance among these companies are particularly notable. Since its establishment in San Francisco in 2008, it has experienced rapid growth. Evaluating the dimensions of Airbnb’s popularity has become a important research in the industry. Both objective factors, such as cleanliness, and subjective factors, such as service attitude, can reflect consumers’ evaluations.

Previous studies mainly focused on feedback comparisons between Airbnb and the hotel industry. Researchers often question the importance of the host-guest interaction by identifying key attributes of Airbnb (Tussyadiah & Zach, 2016) and examining the differences between Airbnb and traditional hospitality industry (Guttentag 2015). The rapid growth of Airbnb challenges the conventional theories and practices developed in the hotel industry (Bridges and Vásquez, 2016; Cheng, 2016; Zervas et al., 2017).

With Airbnb’s promise to provide a unique human-to-human experience, researchers have embarked on investigations to identify the dimensions that form the basis of the Airbnb experience. Chen et al. (2018) seek to understand the factors affecting the purchase intention of Airbnb users in terms of five key factors: previous ratings, rating volume, reviews, information quality, and media richness. Cheng et. al. (2019) offered an alternative approach and more coherent understanding of the Airbnb experience based on the analysis of online comments.

Many researchers have applied the machine learning techniques to investigate customer experience on hostel platforms. The objective of Joseph et al. (2019) is to present a case of text mining on Airbnb users’ reviews to analyze and understand various aspects that drive customer satisfaction. Turnbull (2019) described the development of a model-based user intent metric, listing view, which combines the signals of various user micro-actions on the listing description page. While these researches illustrated a feasible pathway for this research, they compared the dimensions of the hotel industry with the dimensions of Airbnb, leading to rather general similarities and differences. This study will improve that by directly analyzing the features on Airbnb to gain more specific insights and better explore consumer appeals.

4. Exploratory Data Analysis

4.1 Data Overview

The raw dataset shows Airbnb activities in the city of **Melbourne, Victoria, Australia**. It was collected on **December 7, 2018**, and consists of extensive data on **Airbnb listings**, including information on host characteristics, price, daily & overall availability trends, and guest ratings. Melbourne was ranked **sixth among the top ten global cities for Airbnb users in 2016** and continues to be one of the priority short-term rental markets.

The dataset originally contained **22,895 rows and 84 columns**, including **numerical, categorical, and text-based data**. To enhance efficiency, **irrelevant columns** such as **id, listing_url, and host_name** were removed, reducing the dataset to **53 columns**. A check for **duplicate entries** was conducted, confirming that the dataset contains **no duplicate rows**, ensuring data integrity.

4.2 Missing Value Analysis

The analysis assessed the validity of the dataset by determining the missing values among the variables. The **license** column has the **highest missing rate of 99.9%**, followed by **monthly price (91.74%)** and **weekly price (88.98%)**. Apart from this, the **review scores variables** display around **25% missing rates**, likely due to listings not having enough guest ratings.

4.3 Outlier Detection

Outliers were identified using the **interquartile range (IQR) method**, particularly in financial features like *price*. Outliers can distort data interpretation and model accuracy.

Boxplots visualized numerical variable distributions and highlighted extreme values. The *price*, *weekly_price*, and *monthly_price* variables showed significant variability, suggesting luxury listings or data entry errors.

Notable anomalies include the **boxplot of extra_people (Figure 1)** reveals listings charging over **$400 per extra guest**, far exceeding typical rates, indicating either data entry errors or luxury pricing strategies.

The **boxplot of price (Figure 2)** shows properties priced above **$10,000 per night**, suggesting either ultra-luxury listings or erroneous entries.

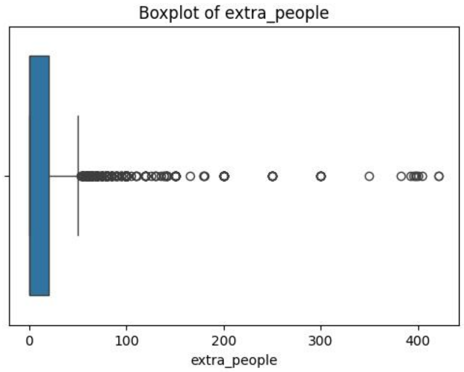


Figure 1: Boxplot of Extra People

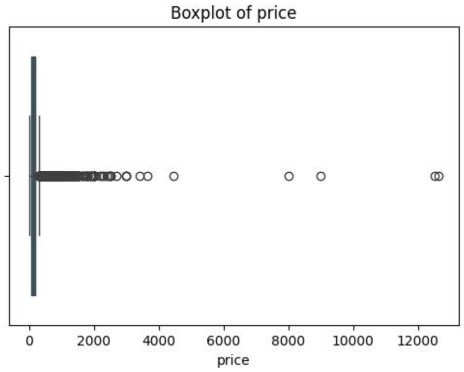


Figure 2: Boxplot of Price

4.4 Correlation Analysis

A **correlation heatmap** identified strong relationships between variables, highlighting redundancy. The **first group**—*accommodates*, *bedrooms*, *bathrooms*, and *beds*—showed correlations between **0.53 and 0.86**, confirming that larger properties accommodate more guests. The **second group**—*price*, *weekly_price*, and *monthly_price*—displayed strong correlations, indicating a proportional pricing structure. The **third group**—*availability_30*, *availability_60* and *availability_90*—had correlations from **0.85 to 0.96**, suggesting that shorter availability often predicts longer availability, making some variables redundant.

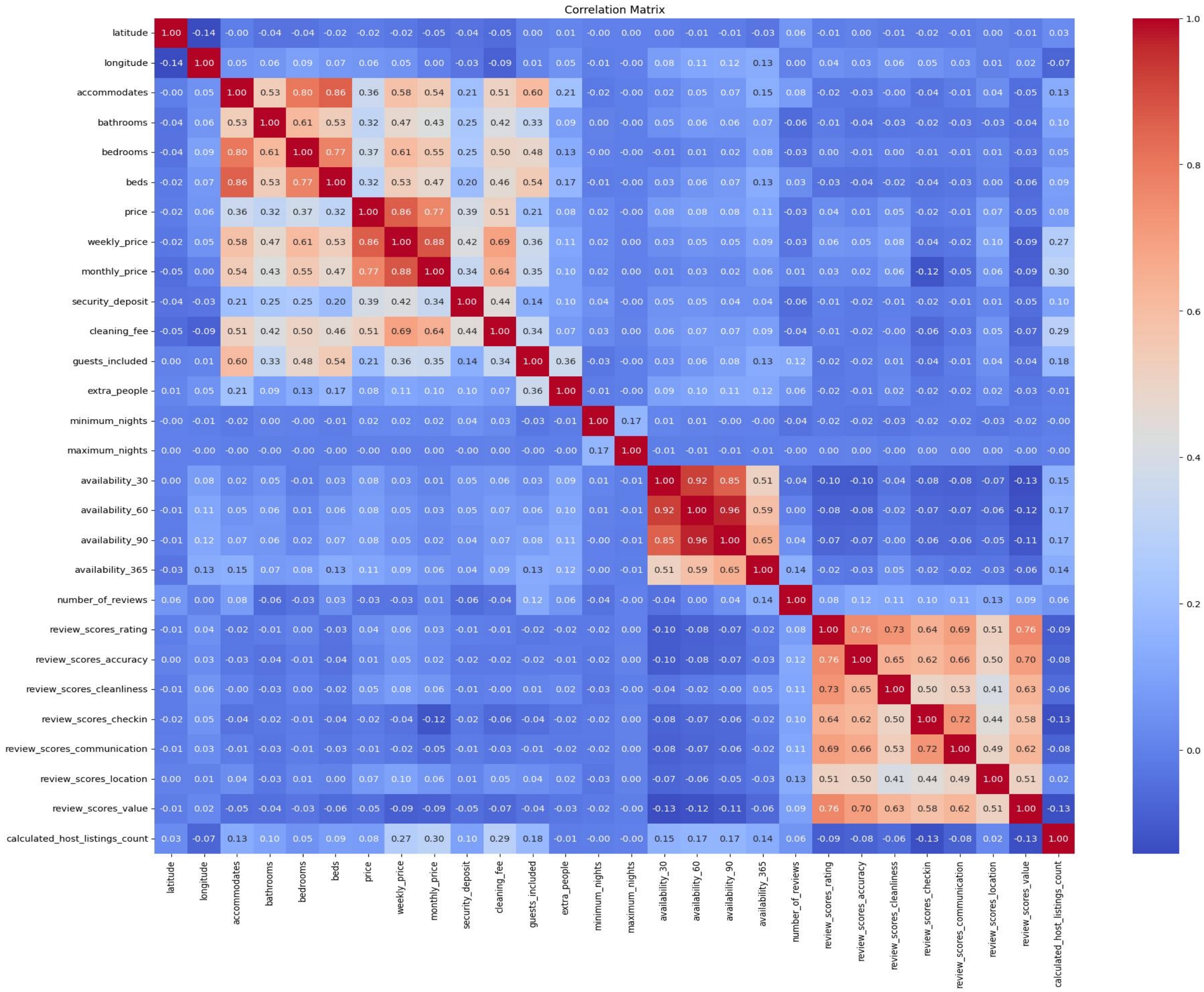
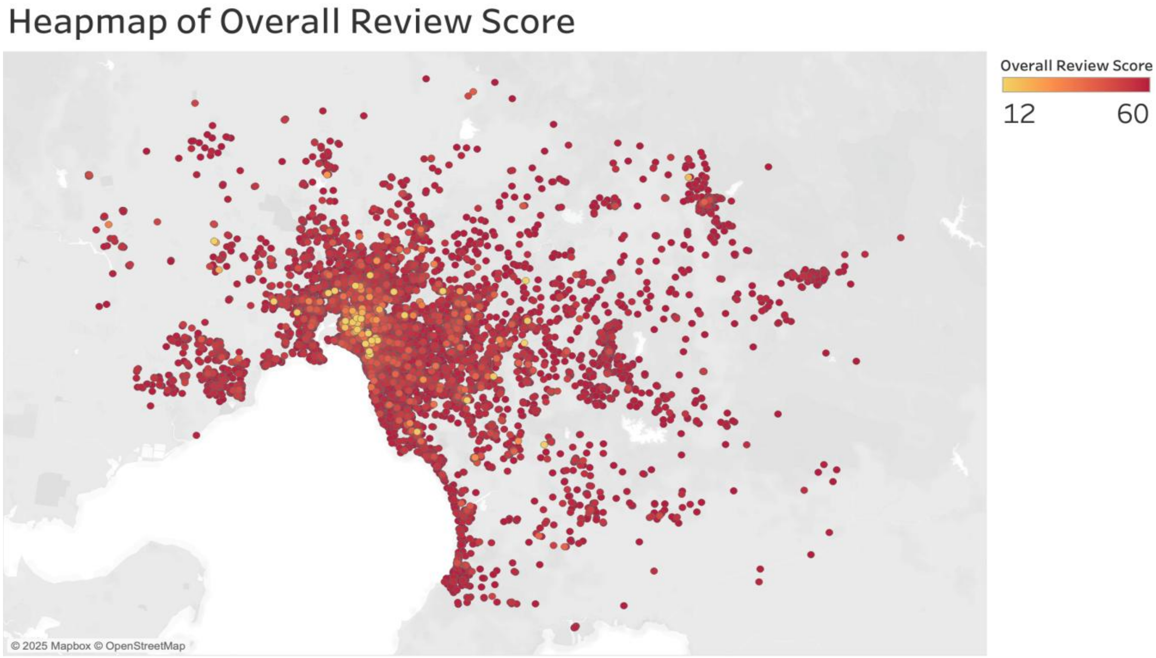


Figure 3: Correlation Heatmap

4.5 Spatial Distribution of Review Scores

To further analyze the relationship between location and guest satisfaction, **Tableau** was used to generate a rating heat map, plotting Airbnb listings based on their latitude, longitude, and ratings. The visualization shows a clear spatial pattern of guest ratings, and it is interesting that low-rated listings are concentrated in the central city. In contrast, high-rated listings tend to be scattered in suburbs and surrounding areas.


Heapmap of Overall Review Score
Overall Review Score
12
60
© 2025 Mapbox © OpenStreetMap

Map based on Longitude and Latitude. Color shows details about Overall Review Score. Details are shown for Smart Location.

Figure 4: Heatmap of Overall Review Score

5. Data Preparation

5.1 Handling Missing Values

(1) Dropping Entire Columns

Some columns were excluded because they were **not informative enough** or had t**oo many missing values,** such as the columns *"license", "monthly_price"*, and *"weekly_price"* were removed as they did not have a bearing on the property's quality which helped in eliminating data redundancy. Also, *host_neighborhood*, *access*, *space*, and *host_verifications* had **an extraordinarily high percentage of missing values** which made their imputation expensive and problematic. For this reason, they were eliminated entirely.

(2) Dropping Rows with Missing Values

Property reviews are essential in determining the quality of a property and the satisfaction of the guests. In order to maintain the quality of the dataset, listings without review scores were deleted. The main variables related to the review are:

- *review_scores_cleanliness* - the cleanliness of the property.
- *review_scores_communication* - communication with the host.
- *review_scores_checkin* - check-in experience.
- *review_scores_accuracy* - accuracy of the listings' description.
- *review_scores_location* - location score.
- *review_scores_value* - value for money.

(3) Dropping Rows with Few Missing Values

While other columns **do not have a large proportion missing,** they still provide important descriptive details about the property, for example *"beds"* (number of beds)*, "bathrooms"* (number of bathrooms)*, "host_is_superhost"* (whether host is superhost**).**

(4) Imputing Missing Values for Text Variables

To ensure the dataset is useful, some textual variables are **manually filled with some values,** preparing for the following feature engineering, for example*"transit"* (how accessible is public transport), *'neighborhood_overview'* (how is the neigborhood), and *"house_rules"* (what are the specific house rules/set out). This information is missing for a listing may imply that the host **has not supplied enough information** making the data incomplete. For the sake of maintaining the usability and applicability of the dataset, the gaps across *transit, notes, neighborhood_overview,* and *house_rules* columns were not removed. Instead, these missing data were filled by 0 for the following feature engineering.

Also, missing values in ***"summary"*** and ***"description"*** fields were filled with empty strings ***("")***. Such an approach guarantees that all calculations based on text length will not lose value during later feature engineering.

5.2 Handling Outliers

A couple of numerical variables contain unrealistic minimum values, thus these components were clipped to make sure they are able to remain within a valid range:

- ***"Price"*:** To avoid situations where prices are set at free or close to no rentals, a minimum clip set at 10 was placed.
- ***"Accommodates"*:** This variable which marks maximum capacity by guests was clipped to a minimum of 1 in order to ensure that there are no listings that claim to accommodate 0 guests.
- ***"Bathrooms"*, *"bedrooms"*, *"beds"*, *"security_deposit"*, *"cleaning_fee"*, *"extra_people"*, *"number_of_reviews"*:** These columns were clipped to 0, restricting the negative values that were placed into the columns.
- The ***"maximum_night"*** feature also removes an extreme value, 99,999,999, which will seriously affect a row of data in the subsequent model.

5.3 Handling Highly Correlated Variables

The removal of highly correlated variables was done to **eliminate multicollinearity and redundancy**. According to the previous **heatmap**, *availability_30, availability_60*, and *availability_365* were dropped because of their high correlation with *availability_90*. Likewise, *weekly_price* and *monthly_price* were removed as they had a strong correlation with price. Such actions guarantee a better understanding of the dataset during modeling.

6. Feature Engineering

6.1 Creating New Features

To improve feature representation, raw geographical coordinates (*latitude, longitude*) and categorical location descriptors (*smart_location*) were replaced with a single numerical feature: ***distance_to_city_center***. Using the *Haversine formula*, the great-circle distance was calculated between each property and **Melbourne city center (-37.8136, 144.9631)**, providing a more meaningful measure of location impact on review scores. This transformation reduces feature redundancy, enhances interpretability, and allows the model to better capture potential non-linear relationships between location and ratings.

Another new feature was created, ***overall_review_score***, to aggregate the extensive guest satisfaction measures obtained from all the review score variables, which include aspects of accuracy, cleanliness, check-in, communication, location, and value. As a result, this transformation gives an overall guest satisfaction score, improving on the complexity of guest satisfaction scores.

To quantify a host's experience, the *host_since* was transformed feature into a numerical variable, ***host_duration_days***, representing the total number of days since the host's registration.

This feature captures the potential influence of host experience on review scores, as more experienced hosts may provide better services, leading to higher ratings.

Moreover, two additional numerical features were created that track the overall richness of a property listing. These features depict the **length of the *"summary"* and *"description"*** fields respectively.

6.2 Binary Encoding for Text Variables

Several text-based columns (*transit, notes, neighborhood_overview, and house_rules*) contain additional descriptive information about the property. These columns were converted into **binary variables (0/1)** to make them compatible with machine learning models. The ones with content are marked as 1, and the ones without content are marked as 0 (which has been marked when dealing with missing values before). Then, these columns were binned together to a new feature, ***"descriptive_features"***, which shows the total impact of these textual variables.

6.3 One-Hot Encoding for Categorical Variables

All categorical characteristics (e.g. bed type, immediate availability, cancellation policy *room_type"* and *"neighbourhood_group*). These features, **one-hot encoding** was performed to handle them. This was done with *pd.get_dummies(),* the categorical variables are preserved for analysis while eliminating potential biases into the models. Moreover, multicollinearity was avoided by setting **drop_first=True,** thus enhancing the overall robustness of the model.

6.4 Transformation of Target Variable

To mitigate the impact of outliers and improve the distribution of the target variable, *overall_review_score*, a **winsorization** approach was applied by capping extreme values at the 5th and 95th percentiles. This adjustment prevents highly skewed ratings from disproportionately influencing the model. Following this, a **log transformation** was performed to further normalize the distribution and reduce right skewness.

7. Methodology

7.1 Linear Regression

Linear Regression was the first model used as **baseline** model to explore the relationship between Airbnb guest satisfaction and various features. This model assumes a linear relationship between the target variable (*overall review score*) and the input features. Linear regression is a simple and interpretable method that works well when the relationship between the features and target is approximately linear.

Before fitting the model, feature scaling was applied using **StandardScaler** to standardize the input features. This step ensures that each feature contributes equally to the model and prevents any single feature with larger numerical values from disproportionately influencing the outcome. The dataset was split into an 80% training set and a 20% test set to assess the model's generalization ability.

While Linear Regression is relatively simple and computationally efficient, it may struggle with capturing complex, non-linear relationships in the data compared to more sophisticated models like ElasticNet, XGBoost, and Random Forest. Nonetheless, Linear Regression provides a good baseline for comparison with other models in this study.

7.2 ElasticNet

ElasticNet Regression was used to determine the determinants of **Airbnb guest satisfaction**. The model, by means of combining **Lasso (L1)** and **Ridge (L2) regularization**, offers meaningful selection of features and simultaneously prevents serious **multicollinearity**. This **hybrid method** is designed to filter the data exclusively with the essential variables and to avoid **overfitting**.

In order to ensure that the predictive model is consistent in all features, non-target variables are standardized using **StandardScaler**. The dataset was split into an **80% training set and a 20%test set** to evaluate the effectiveness of the model. **Hyperparameter tuning** achieves an optimum L1 ratio and alpha for the ElasticNet model using the **ElasticNetCV method**; the process also employs **five-fold cross-validation**. **Cross-validation** shows that the best choice of the parameters, **alpha = 0.0002976** and **L1 ratio = 0.8**, makes a compromise between the **coefficient freeze and decline**.

7.3 XGBoost

In the beginning, a **parameter search space** was set up in the variable named "param_grid." This grid specified ranges for essential **XGBoost** hyperparameters, including **n_estimators**, **learning_rate**, **max_depth**, **subsample**, **colsample_bytree**, **gamma**, and **min_child_weight**. Each of these parameters plays a role in controlling the model's complexity. For instance, **n_estimators** governs the number of trees, while **max_depth** limits the depth of each tree, and **learning_rate** determines how quickly the model adapts during training. This approach balance between underfitting and overfitting effectively.

To explore the parameter space without testing every combination exhaustively, **RandomizedSearchCV** was employed. This technique randomly samples parameter combinations, making it computationally efficient while still providing a good chance of finding near-optimal settings. A **5-fold cross-validation** was also coupled. Cross-validation shows that the best choice of the parameters, **n_estimators = 800**, **min_child_weight = 5**, **max_depth = 4** and **learning rate = 0.01**.

To refine the model further, a smaller parameter grid (grid_param) was used around the best **learning_rate** to discover during the randomized search. By applying **GridSearchCV** with this narrowed focus, the learning rate was fine-tuned to achieve even better model performance.

7.4 Random Forest

The Random Forest Regressor was optimized using **GridSearchCV** and evaluate its predictive performance. The objective is to identify the optimal combination of hyperparameters that minimizes prediction errors. To facilitate this process, a parameter grid was defined, **param_grid**, which specifies a range of hyperparameter values to be explored during the grid search. The hyperparameters considered include **n_estimators**, representing the number of trees in the forest; **max_depth**, which sets the maximum depth of individual decision trees; **min_samples_split**, defining the minimum number of samples required to split an internal node; and **min_samples_leaf**, which determines the minimum number of samples that must be present at a leaf node. By systematically iterating through all possible combinations of these hyperparameters, the grid search allows us to identify the configuration that yields the best model performance. To ensure robust model evaluation, a 5-fold cross-validation was implemented.

7.5 Model Evaluation

Predicted **review scores** of four models are evaluated through basic metrics, R2 and MSE. In addition, **plots of True vs. Predicted values** were applied to show the models' performance**. Distribution of residuals** were also plotted to determine the level of **accuracy** of the model. These measures would show how well the model ensures guest satisfaction and find the **predilection of discrepancy between models to predictions**.

The focus of this study is mainly on **feature importance**. Particularly, a feature importance analysis was conducted in each model to show what features contribute most to **guest satisfaction** to enhance corresponding strategies for **hosting and optimizing the platform**.

8. Results and Findings

8.1 Model Performance Metrics

As the baseline model, **Linear Regression** model has **very few errors** in the model's prediction, which was evidenced by the **low MSE** (0.001399 for training set, and 0.001405 for testing set). However, the **R² scores are not outstanding**, which were 0.1609 for training data, and 0.1649 for testing data.

The performance of ElasticNet Regression model is similar to baseline**.** The **MSE** of the **ElasticNet Regression model** for the **training data** stands at **0.001403**, with **0.001407** for the **test set data**. However, the **R² scores** are only **0.159 and 0.163**, and the model accounts for only **around 16% variance** for **review scores**. Because the **R² statistics** are relatively low, **following models**, using **XGBoost** and **Random Forest**, will be considered for improved prediction and complex relationship analysis of this study.

XGBoost model **performs better** than previous two linear models, which increased the R² score of testing set to 0.232, but the improvement was not significant despite of the hypertuning and cross-validation conducted in the implementation. **MSE remained very low**, 0.00112 for train set and 0.00129 for test set.

For Random Forest, the test MSE is 0.00129, and the R² score is 0.233, indicating that the model **still performs excellent in keeping low prediction errors**, and still explains only 23.2% of the variance in review scores on unseen data. Given the situation of low R² in four models, it's recommended to **combine the results of four models together** to get comprehensive insights from this study.

8.2 Model Evaluation Visualizations

(1) True vs. Predicted Review Score Scatter Plot

The **True vs. Predicted Review Score Scatter Plot** (Figure 5-8) specifically shows the variance between the score given by the model and the **actual review scores**. The **red dashed line** shows the **ideal one-to-one prediction relationship**; the model approaches this line, but some **variance** appears, particularly where **higher review numbers** are found. This figure speaks to the **accuracy of the predictive nature of the model**.

All four models show a tendency to predict the review scores fairly **closely to the true scores**, with most of the points clustering near the red dashed line. This suggests that the models are generally able to predict the review scores with a reasonable level of accuracy.

A key observation is that for **higher review scores**, especially those close to the maximum (4.1), there appears to be **greater variance** between the true and predicted scores.

This indicates that while the models can predict low to mid-range scores well, they tend to be less precise when it comes to predicting very high ratings.

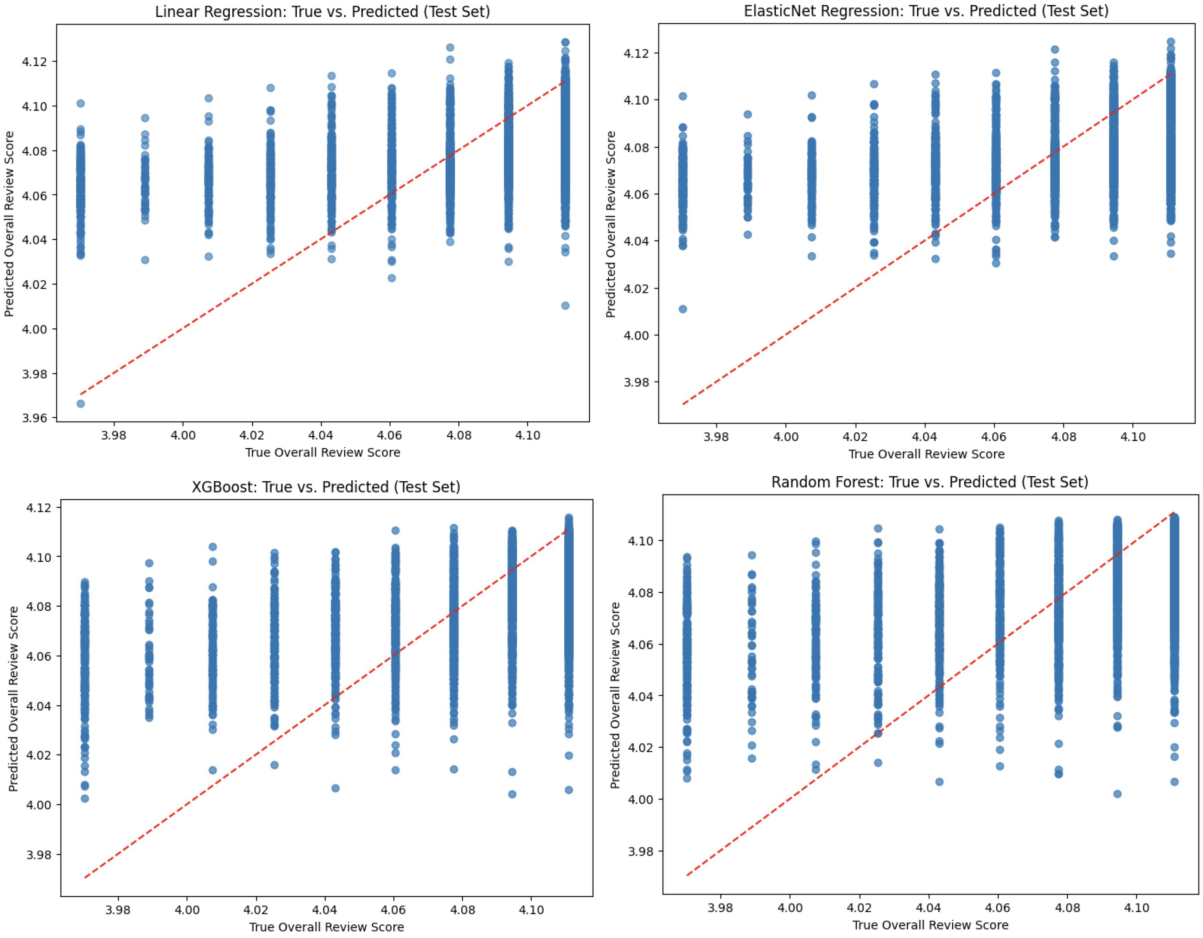
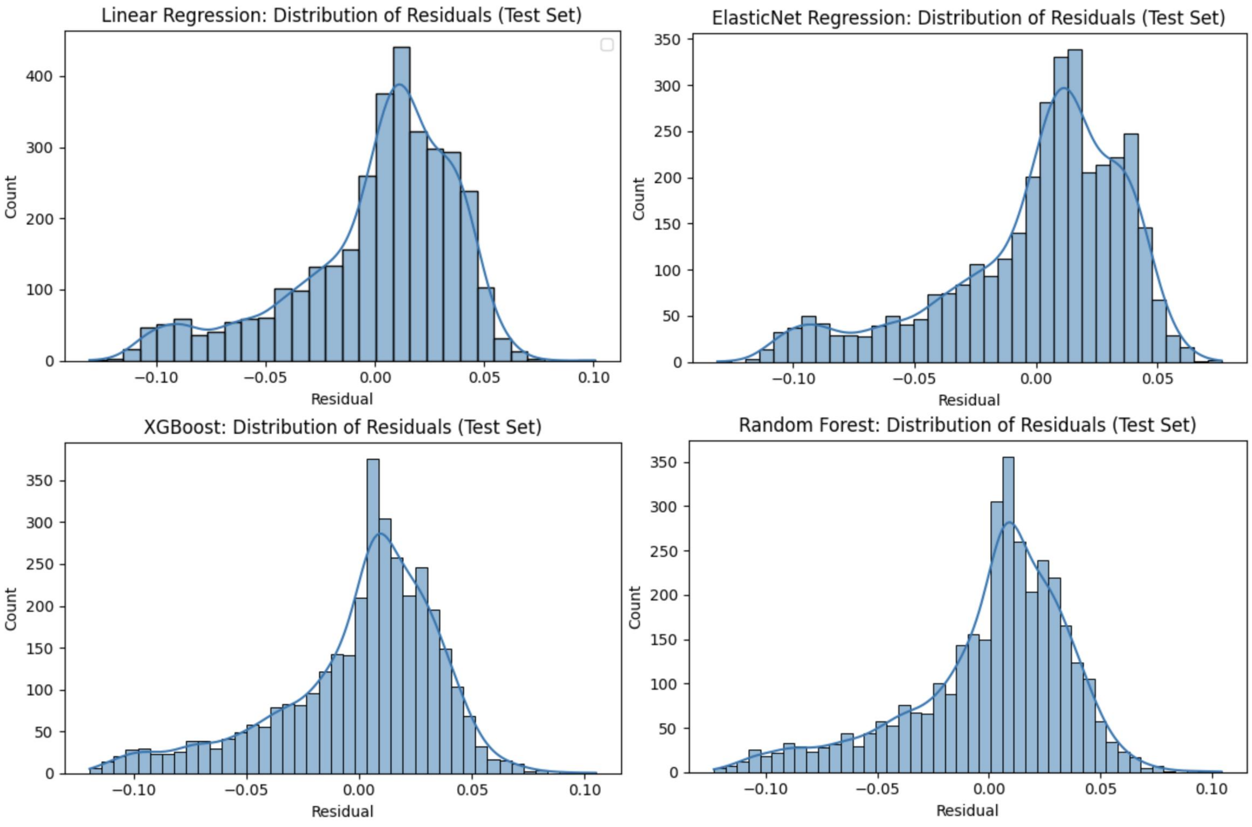


Figure 5-8: True vs. Predicted Overall Review Score – Test Set of the four models

(2) Distribution of Residuals

In figures 9-12, a simple **residual analysis** is performed in order to identify the **error in the model predictions**. The histogram of the residuals shows **errors predominantly around zero**, which **implies the accuracy of prediction**; however, some **discrepancies** can still be detected. The extreme errors demonstrate that there are certain factors beyond the attributes of the highlighted items, but **they are very personal (job location preference, expectations, and guest reviews)**.

The residuals formed a roughly symmetrical bell shape centered near zero. This distribution indicates that the errors are mostly small and spread evenly, suggesting that the model does not favor overprediction or underprediction. The fact that most residuals cluster close to zero reinforces the idea that the models captures the key relationships within the data without introducing systematic bias. Generally, **the two tree-based models performed better** than linear models.


Linear Regression: Distribution of Residuals (Test Set)

ElasticNet Regression: Distribution of Residuals (Test Set)

XGBoost: Distribution of Residuals (Test Set)

Random Forest: Distribution of Residuals (Test Set)

Figure 9-12: Distribution of Residuals – Test Set of the four models

8.3 Feature Importance Analysis

From **figures 13-16**, for all the four models, it is obvious that **superhost status** is the strongest predictor of **guest satisfaction**. The **superhost designation** is a sure indicator of a **seasoned host**, typically correlating with **high ratings** because this designation depicts **hosting qualities** such as **complete stays, reliability, responsiveness, and positive past guest experiences**. The superhost badge stands for a **trust signal** and features **high-quality services**. Visitors have most of the time **positive feedback**, which is an important factor of their **satisfaction**.

The number of listings managed by a host is another important factor for the four models. From the coefficients of the two linear models, the higher the calculated_host_listings_count, the more negative the impact will be. This means that the host may not be able to manage and maintain multiple houses at the same time, and may not be able to reply to customers' messages online in time, thus reducing the user experience.

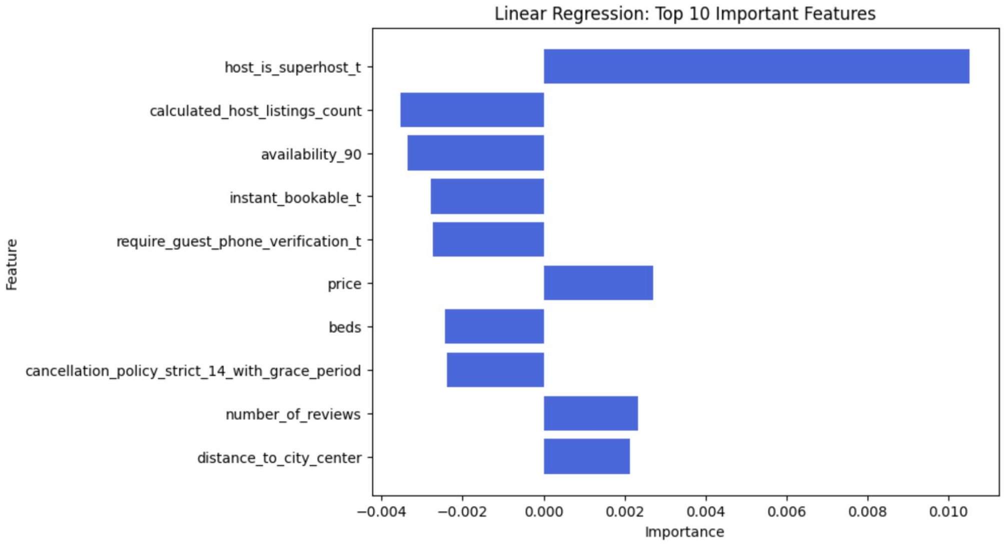
Being **available over the past three months** is evidently correlated with **guest dissatisfaction**. The **high demand** for **listings that are frequently booked** suggests active pricing of the product, the presence of demand, and the occurrence of positive guest experiences. More availability generally means less of an inclination of a property to be replaced by a competitor. And this can usually be attributed to the owners' maintenance properly, good management, and the engagement of the host.

The slight **positive results of price** on **guest ratings** in three models are interpreted as the **property having better accommodation and/or first-class service**. At the same time, guests, who prioritize value for money, appear inclined to pay extra for an exclusive experience, which includes superior amenities, prime locations, and very responsive staff.

In addition, the high negative impact of cancellation_policy_strict_14_with_grace_period on the score shows that users still care about the **flexibility of house booking**. Houses with low flexibility may have problems themselves, so the homeowners do not intend to give users more flexibility, but choose to retain customers.

Number of reviews has a positive effect on ratings among all the four models, indicating that well-reviewed properties tend to maintain guest satisfaction.

Other variables like **instant_bookable_t** and **host_identity_verified_t** appeared among the top drivers in some models, hinting that ease of booking and host verification—factors tied closely to trust and convenience—are also key influences on guest ratings.



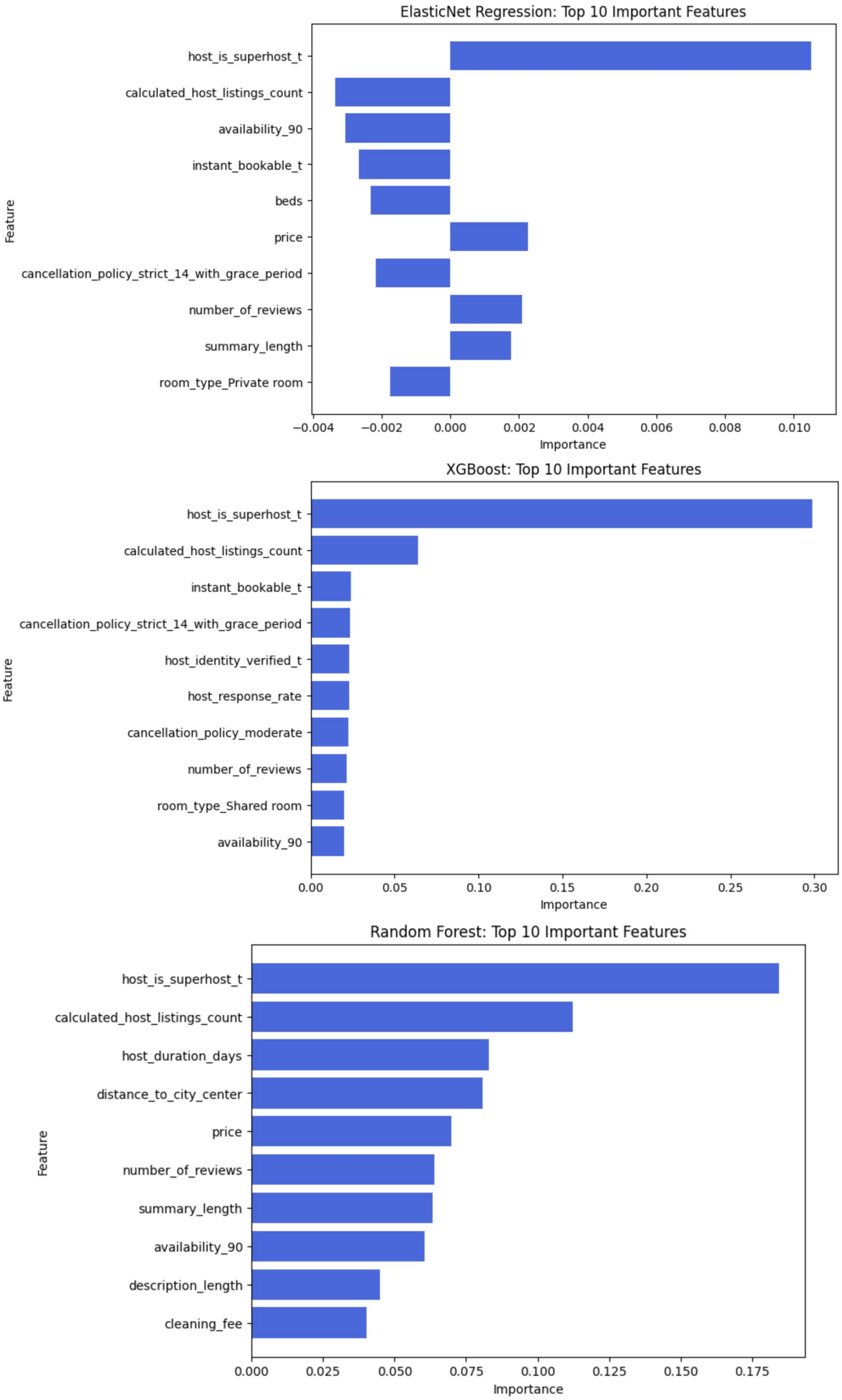


Figure 13-16: Top 10 Most Important Features in the four models

9. Conclusion and Recommendations

9.1 Summary of Findings

This study used machine learning techniques to assess the primary determinants of guest satisfaction in Airbnb. As expected, **Superhost status** was found to be the strongest predictor of high ratings, since well-reviewed hosts provide better guest experiences. The host's number of listings affects ratings, as more experienced hosts generally receive higher ratings.

Moreover, other variables related to the policies and properties facilities also contributed to customers' ratings. In the end, **XGBoost** and **Random Forest** outperformed **Elastic Net** in predictive accuracy, but the harshly low **R^2** values in the end denote that personal guest preferences and specific experiences largely impact ratings.

9.2 Recommendations

For **Airbnb hosts**, To improve guest satisfaction, hosts should aim for **Superhost status** by maintaining high **service quality, prompt communication**, and **positive reviews**. Managing multiple listings effectively and ensuring consistent service can also enhance ratings. Additionally, **offering flexible cancellation policies** can lead to better guest experiences and higher ratings.

For **Airbnb platform,** improving guest satisfaction at Airbnb is possible by shifting the focus of the recommendation algorithm to the aspects, like Superhost criteria, instant booking, and host experience. Revenue optimization for hosts can be accomplished with data-driven pricing analysis while better guest-host communication tools can improve service quality.

9.3 Future Research Directions

To enhance the precision of predictions and capture intricacies in the relationships present in the data, **MLP or text-based transformer models** could be utilized for the analysis of guests' textual reviews. **A Bayesian approach** could as well be used to integrate uncertainty estimation within the predictions which assists in determining the accuracy of different features affecting ratings. Besides, unsupervised learning techniques like **clustering (K-Means, DBSCAN)** can complement the Airbnb listing segmentation motivated by user preferences and reveal data that is not captured through traditional regression analysis.

The addition of these available external datasets or models can aid further research in enhancing guesstimates, exploitable gaps, and trends that might be more relevant to the customers and the hosts of the platform and the site itself. These models can make use of the available data that incorporate **tourism activities, crime rates, and other seasonal trends or special event information.** Incorporating seasonal changes in visitor numbers and crime rates can help analyze how different variables such as neighborhood safety affects guest ratings.

Reference

Bridges, J., and C. Vásquez. 2016. "If Nearly All Airbnb Reviews Are Positive, Does That Make Them Meaningless?" *Current Issues in Tourism* 21 (18): 2057–75. https://doi.org/10.1080/13683500.2016.1267113.

Chen, C., and Y. Chang. 2018. "What Drives Purchase Intention on Airbnb? Perspectives of Consumer Reviews, Information Quality, and Media Richness." *Telematics and Informatics* 35 (5): 1512–23. https://doi.org/10.1016/j.tele.2018.03.019.

Cheng, M. 2016. "Sharing Economy: A Review and Agenda for Future Research." *International Journal of Hospitality Management* 57: 60–70. https://doi.org/10.1016/j.ijhm.2016.06.003.

Cheng, M., and X. Jin. 2019. "What Do Airbnb Users Care About? An Analysis of Online Review Comments." *International Journal of Hospitality Management* 76: 58–70. https://doi.org/10.1016/j.ijhm.2018.04.004.

Clark, J. 2014. "Making Connections via Peer-to-Peer Travel." *USA Today*, January 1, p. 8.

Guttentag, D. 2015. "Airbnb: Disruptive Innovation and the Rise of an Informal Tourism Accommodation Sector." *Current Issues in Tourism* 18 (12): 1192–1217. https://doi.org/10.1080/13683500.2013.827159.

Joseph, G., and V. Varghese. 2019. “Analyzing Airbnb Customer Experience Feedback Using Text Mining: Managerial Approaches, Techniques, and Applications.” In *Big Data and Innovation in Tourism, Travel, and Hospitality*, 147–62.

Turnbull, B. 2019. “Learning Intent to Book Metrics for Airbnb Search.” WWW’19: *The World Wide Web Conference*, 3265–71. https://doi.org/10.1145/3308558.3313648.

Tussyadiah, L., and F. Zach. 2016. “Identifying Salient Attributes of Peer-to-Peer Accommodation Experience.” *Journal of Travel & Tourism Marketing* 34 (5): 1–17. https://doi.org/10.1080/10548408.2016.1209153.

Zervas, G., D. Proserpio, and J.W. Byers. 2017. “The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry.” *Journal of Marketing Research* 54 (5): 687–705. https://doi.org/10.2139/ssrn.2366898.