Airbnb Price Prediction

Team 11B – Moiz Nawab, Sharada Sarangan, Jerome Timmons, Anru Wang, Yashu Yang

BANA 273 – Machine Learning

Instructor: Professor Mingdi Xin

02 Dec 2022

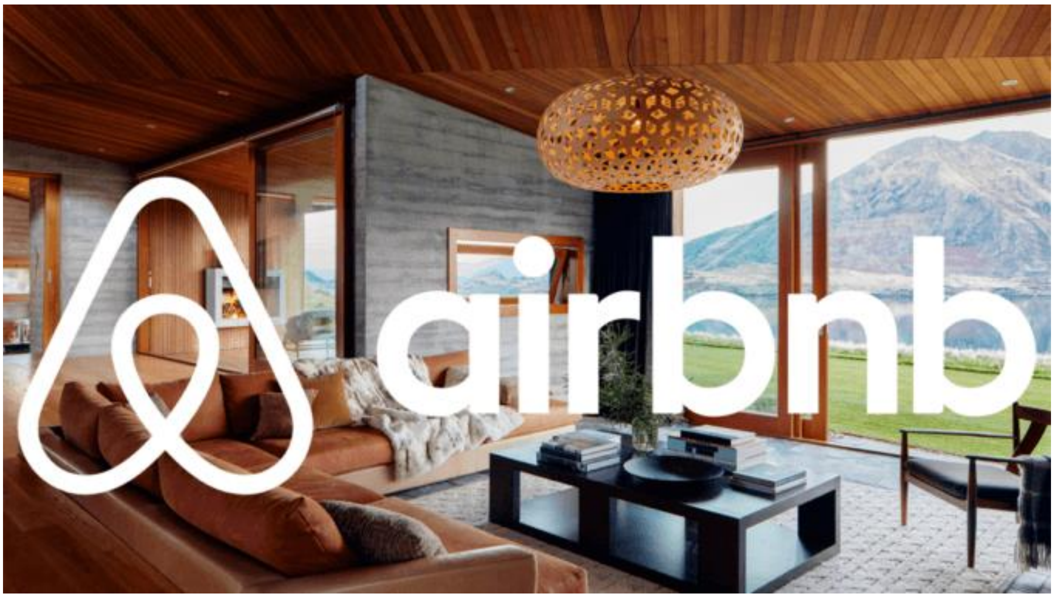


Table of Contents

1. Introduction

1.1. Background

Since its launch in 2008, Airbnb has provided an online marketplace for short-term stays and experiences. The average stay period is around 6 days, and the stays can be booked in homes, apartments, condos and tipis. The company caused a disruption in the hotel industry by providing a more cost-efficient stay, also while giving travelers a "residential feel". Today Airbnb has more than 6 million active listings worldwide, can be found in more than 100,000 cities and towns, and over 220 countries worldwide[1].

Due to its popularity, there is a rise in interest for individuals that want to partake in listing their spaces on Airbnb. Travelers choose Airbnb for various reasons as they offer various benefits such as low-cost stays, household stays and location convenience. Furthermore, it allows travelers to rent entire homes by themselves instead of sharing accommodations with a host. Airbnb is considered by many travelers a more effective substitute for mid-range hotels.

In addition to the viability for using Airbnb, the attractiveness and reputation of the location is crucial to its popularity in a particular location such as Los Angeles[2]. The methodology behind the provision of pricing guidelines for potential Airbnb hosts in Los Angeles can be inferred through the use of datasets.

The methodology behind the provision of pricing guidelines for potential Airbnb hosts in Los Angeles can be inferred using datasets. Due to its increasing popularity, Airbnb's data pool has been leveraged on a wide number of platforms. The data pool has been converted into various datasets by data analysts who have used it to assess the performance of the company through various metrics.

Furthermore, by leveraging and analyzing datasets, Airbnb themselves can execute a more effective decision-making process in attracting a wider number of potential hosts. This includes focusing on variables such as beds, bathrooms, amenities, instant-bookable and cancellation policy. The ratings of beds and bathrooms are important as these are some of the main services which factor into the attractiveness of Airbnb providers. The offering of amenities which hotels lack also adds to the marketability of the vacation rental company. In addition, the variables of instant-bookable and cancellation policies are another set of crucial metrics as flexibility is a very important factor for customers as it saves them time and money.

For the project report, the goal is to provide pricing guidelines for potential Airbnb hosts in Los Angeles. By providing pricing guidelines, potential Airbnb hosts in Los Angeles are

[1] Airbnb. (2022). *About Us*. Airbnb Newsroom. https://news.airbnb.com/about-us/

[2] Toronto Metropolitan University. (2016, October 6). *Why tourists choose Airbnb over hotels*. Toronto Metropolitan University. https://www.torontomu.ca/news-events/news/2016/10/why-tourists-choose-airbnb-over-hotels/

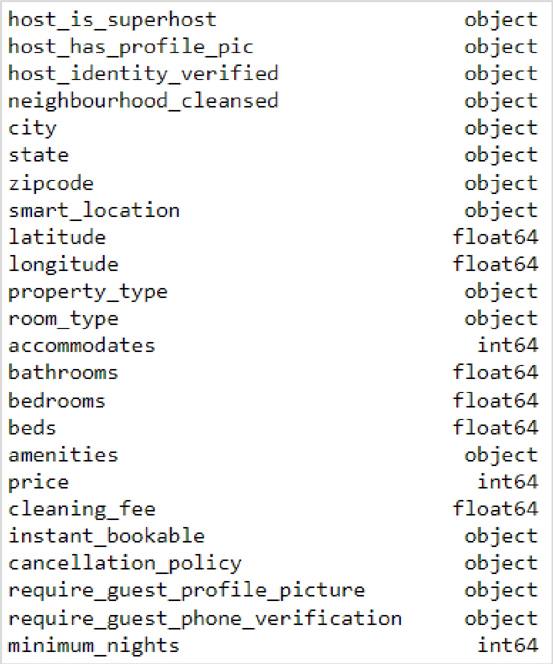
provided with models which enable more effective decision-making in determining the optimal pricing for their rentals. Therefore, enabling interested potential hosts in ensuring profitability, affordability and attractiveness.

1.2. Data Set

The dataset for Airbnb was pulled from Kaggle[3]. This dataset contains Airbnb listings in Los Angeles, CA comprised of 95 columns and 31,253 observations. We first investigated the columns to see what would be relevant for our analysis. The data contained several columns pertaining to location such as street, neighborhood, city, zip code and latitude and longitude. The remaining columns covered information about the host, descriptions about the property and location, URL for the listing, and review scores.

2. Data Cleaning

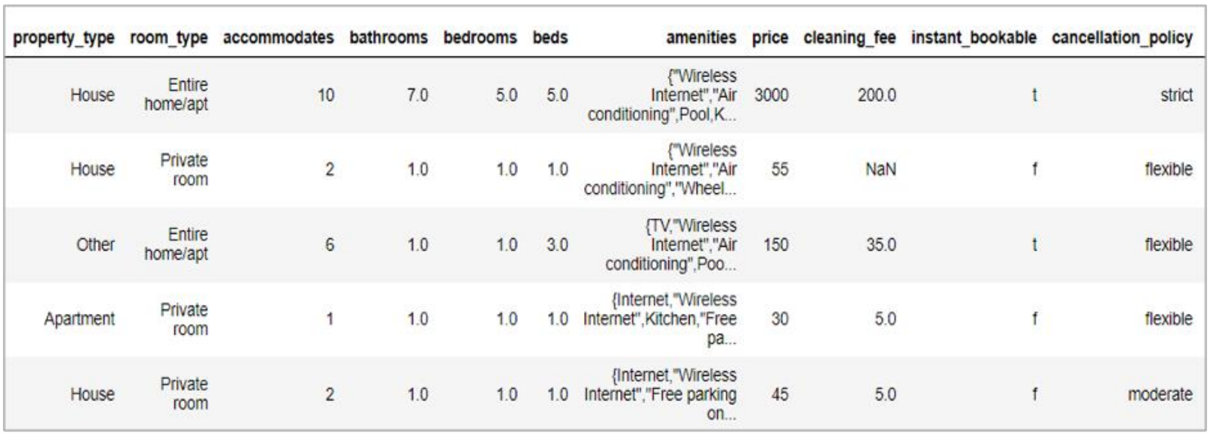
In order to properly analyze the data, we removed several irrelevant columns. These columns contained data referring to website URLs. For the location we kept columns such as neighborhood, city, zip code, latitude and longitude and removed the remaining columns pertaining to the location. There were columns such as "is this location exact", which provided True/False results, but the column was 98% True so we decided to remove the column. Another significant portion of columns we removed were the columns pertaining to reviews of the Airbnb as we are not considering those for the price analysis.
After sifting through all the columns, we settled on 21 columns that would be relevant for our analysis:



```
host_is_superhost                     object
host_has_profile_pic                  object
host_identity_verified                object
neighbourhood_cleansed                object
city                                  object
state                                 object
zipcode                               object
smart_location                        object
latitude                             float64
longitude                            float64
property_type                         object
room_type                             object
accommodates                           int64
bathrooms                            float64
bedrooms                             float64
beds                                 float64
amenities                             object
price                                  int64
cleaning_fee                         float64
instant_bookable                      object
cancellation_policy                   object
require_guest_profile_picture         object
require_guest_phone_verification      object
minimum_nights                         int64
```

Figure 1: Columns selected for the analysis

[3] Sen, O. (2018). *LA Airbnb Listings*. Www.kaggle.com. https://www.kaggle.com/datasets/oindrilasen/la-airbnb-listings



property_type	room_type	accommodates	bathrooms	bedrooms	beds	amenities	price	cleaning_fee	instant_bookable	cancellation_policy
House	Entire home/apt	10	7.0	5.0	5.0	{"Wireless Internet","Air conditioning",Pool,K...	3000	200.0	t	strict
House	Private room	2	1.0	1.0	1.0	{"Wireless Internet","Air conditioning","Wheel...	55	NaN	f	flexible
Other	Entire home/apt	6	1.0	1.0	3.0	{TV,"Wireless Internet","Air conditioning",Poo...	150	35.0	t	flexible
Apartment	Private room	1	1.0	1.0	1.0	{Internet,"Wireless Internet",Kitchen,"Free pa...	30	5.0	f	flexible
House	Private room	2	1.0	1.0	1.0	{Internet,"Wireless Internet","Free parking on...	45	5.0	f	moderate

Figure 2: Sample Data

The host is superhost column identifies whether the Airbnb host has been coined a "superhost" by the platform. A host becomes a superhost when they have a significant number of reviews from customers, they are highly rated, and very responsive to customers. The other two columns relating to host describe if the host has a profile picture on Airbnb and if their identify has been verified. We decided to select these columns as we wanted to consider if the hosts profile had any effect on the price of the Airbnb. Following the host are the columns pertaining to the location and then the columns containing the details of the location. To describe the location, we selected the columns with the number of bedrooms, bathrooms, and the number of beds in the Airbnb. We also selected amenities which contained features such as internet, TV, washer, dryer, and gym. Price and cleaning fee were also selected, and whether the Airbnb was instantly bookable along with the cancellation policy for the Airbnb. The cancellation policy was categorized into three categories: moderate, flexible, or strict. Next, we selected the columns pertaining to whether the Airbnb required the guest to have a profile picture and also verify they have a valid phone number. Last but not least in order to narrow the dataset down we only selected the minimum nights required to stay in the Airbnb as "1". If it required more than 1 night then we did not consider for this analysis.

3. Analysis

3.1. Analysis - Part 1

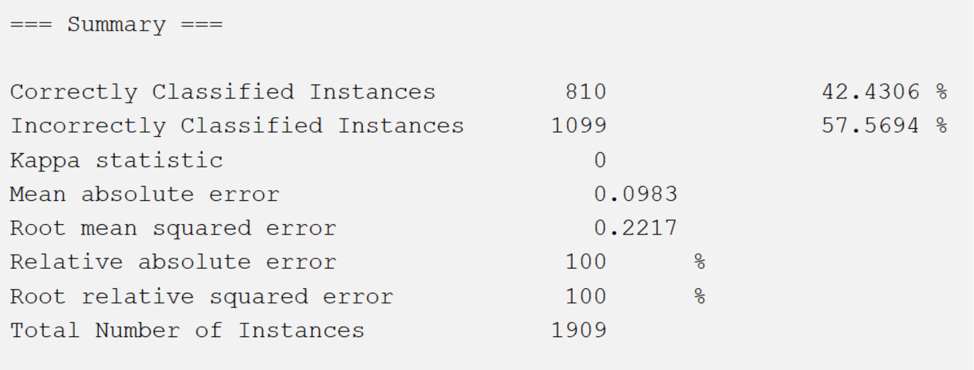
For our Airbnb analysis there were several models taken into consideration. Gaussian Naïve Bayes was one of the considerations. Naïve Bayes classifier is a probabilistic machine learning model that's used for classification tasks, which means that it predicts on the basis of the probability of an object. The classifier is based on Bayes Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Using Bayes Theorem, we can find the probability of A happening, give that B has occurred. B is the evidence and A is the hypothesis; there is an assumption being that the predictors/features are independent[4]; this assumption can also be a disadvantage in that variables are not always independent. Naïve Bayes only uses categorical variables and cannot process numerical or continuous variables, therefore the "price" column in our dataset was binned into ranges of $50 up to $750. We also removed the zip code column.

Benchmark Accuracy

We decided to use WEKA in order to analyze our data and create our model. WEKA is very simple and a powerful tool for machine learning. In order to confirm the efficiency of our model we had to establish a benchmark accuracy. Our method to obtain a benchmark accuracy was using the ZeroR Model. The ZeroR Model is a simple classifier that predicts the majority category (class)[5]. There were no further columns removed to obtain the benchmark accuracy. Using the Zero R Model we were able to obtain a Benchmark Accuracy of 42.43%.



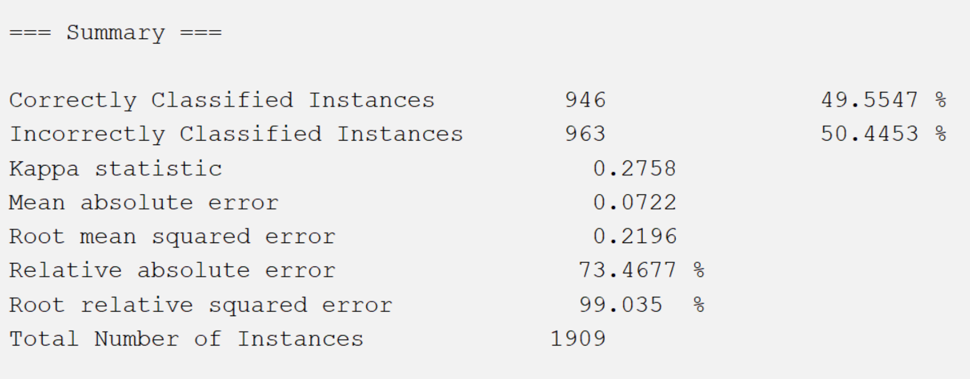
```
=== Summary ===

Correctly Classified Instances         810               42.4306 %
Incorrectly Classified Instances      1099               57.5694 %
Kappa statistic                          0
Mean absolute error                      0.0983
Root mean squared error                  0.2217
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances             1909
```

Figure 3: ZeroR Model Benchmark Accuracy

Once we established the Benchmark Accuracy for the model we ran Naïve Bayes using an 80/20 split; that is 20 percent of the data used for training the model and 80 percent used for testing the model. The initial accuracy for Naïve Bayes was at 49.55%.

[4] Gandhi, R. (2018, May 5). *Naive Bayes Classifier*. Towards Data Science; Towards Data Science. https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c

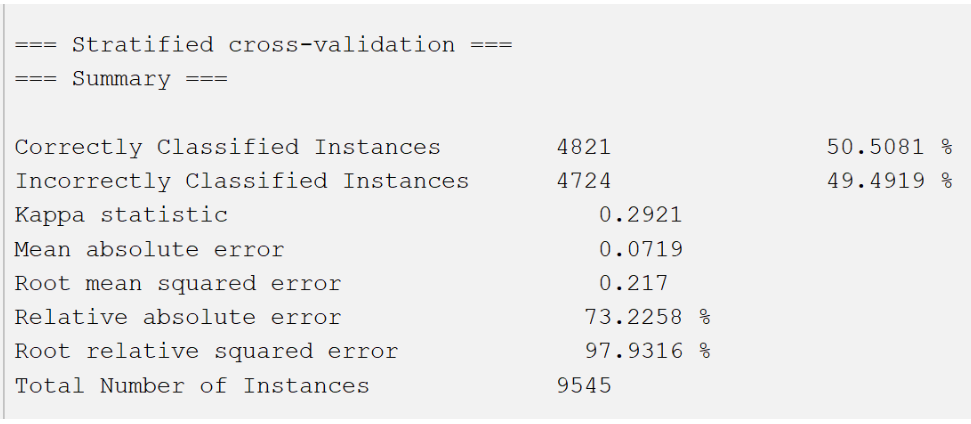
[5] Sayad, S. (n.d.). *ZeroR*. Www.saedsayad.com. https://www.saedsayad.com/zeror.htm#:~:text=ZeroR%20is%20the%20simplest%20classification



```
=== Summary ===

Correctly Classified Instances         946               49.5547 %
Incorrectly Classified Instances       963               50.4453 %
Kappa statistic                          0.2758
Mean absolute error                      0.0722
Root mean squared error                  0.2196
Relative absolute error                 73.4677 %
Root relative squared error             99.035  %
Total Number of Instances             1909
```

Figure 4: Naive Bayes initial run

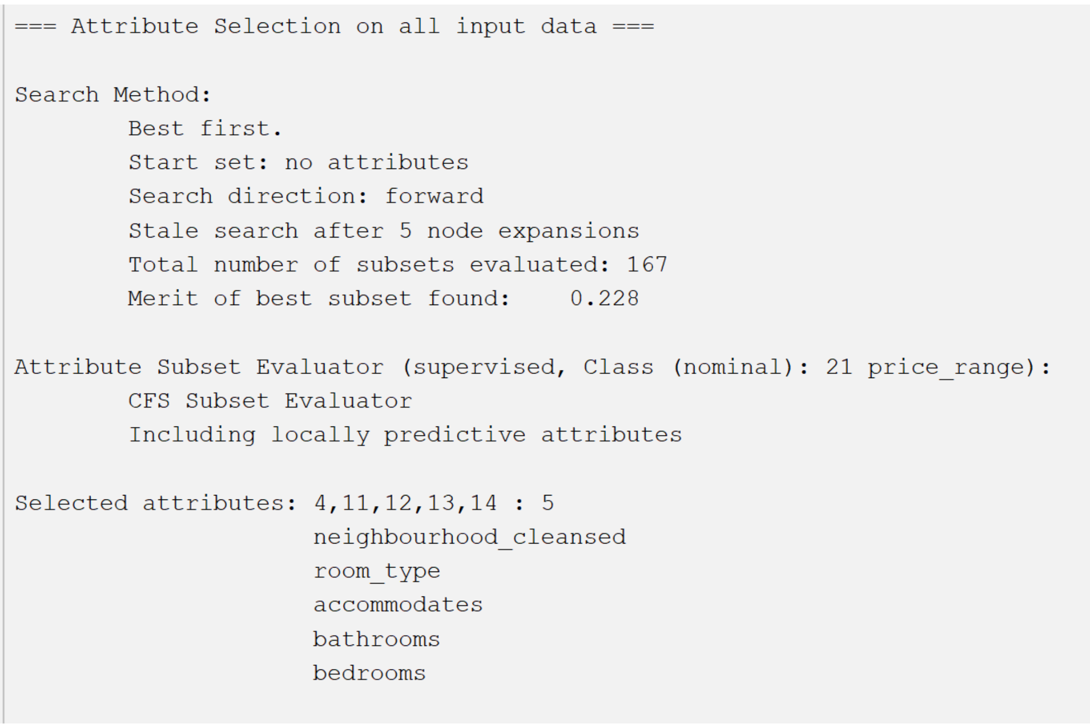
In order to strengthen the accuracy, we used ten-fold cross validation. Ten-fold cross validation is a method where the model takes 10% of the data to validate 90% of the remaining data, and does the process ten times; using ten-fold cross validation we only seen a minute improvement in the accuracy, increasing to 50.51%.



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4821               50.5081 %
Incorrectly Classified Instances      4724               49.4919 %
Kappa statistic                          0.2921
Mean absolute error                      0.0719
Root mean squared error                  0.217
Relative absolute error                 73.2258 %
Root relative squared error             97.9316 %
Total Number of Instances             9545
```

Figure 5: Ten-fold cross validation result

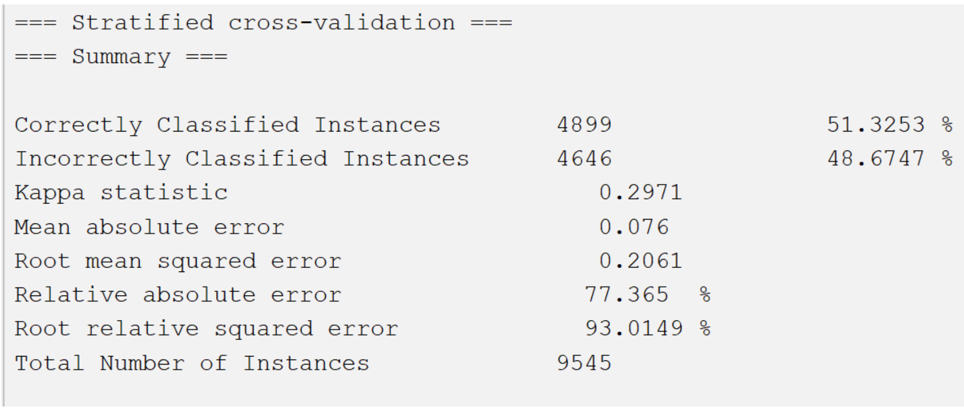
We wanted to see if we can improve the model further and ran the Subset Evaluator in Weka to determine if there was any redundancy in our initial variable selection. The subset evaluator determined that the best variables were neighborhood cleansed, room type, accommodates, bathrooms, and bedrooms.



```
=== Attribute Selection on all input data ===

Search Method:
        Best first.
        Start set: no attributes
        Search direction: forward
        Stale search after 5 node expansions
        Total number of subsets evaluated: 167
        Merit of best subset found:    0.228

Attribute Subset Evaluator (supervised, Class (nominal): 21 price_range):
        CFS Subset Evaluator
        Including locally predictive attributes

Selected attributes: 4,11,12,13,14 : 5
                     neighbourhood_cleansed
                     room_type
                     accommodates
                     bathrooms
                     bedrooms
```

Figure 6: Subset Evaluator - Variable Selection

After selecting the variables according to the Subset Evaluator, we re-ran the ten-fold cross validation and was able to achieve a slight increase in the accuracy with a final accuracy for Naïve Bayes of 51.32%.



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4899               51.3253 %
Incorrectly Classified Instances      4646               48.6747 %
Kappa statistic                          0.2971
Mean absolute error                      0.076
Root mean squared error                  0.2061
Relative absolute error                 77.365  %
Root relative squared error             93.0149 %
Total Number of Instances             9545
```

Figure 7: Final result for Naive Bayes

Naïve Bayes was only one of three methods used in analyzing our model. We will continue with Decision Tree analysis next in our process.

3.2. Analysis - Part 2

Since our dependent variable is continuous, we decided to leverage the regressor models of Decision Tree and K Nearest Neighbor for the next part of our project. The dependent variable – **Price** would be used as a continuous variable and would not be binned into a categorical variable.

The following pre-processing steps were performed on the data before proceeding with the modelling:

1. **Outlier Treatment**
   It is observed from Figure 1 that the 'price' variable is highly skewed and has a lot of outliers. Outliers are data points that are far away from other data points and they are likely to degrade the quality of statistical procedures. Therefore, it is essential to remove the outliers. After analyzing the quantile distribution from Table 1, we removed the data points from the last two percentiles.

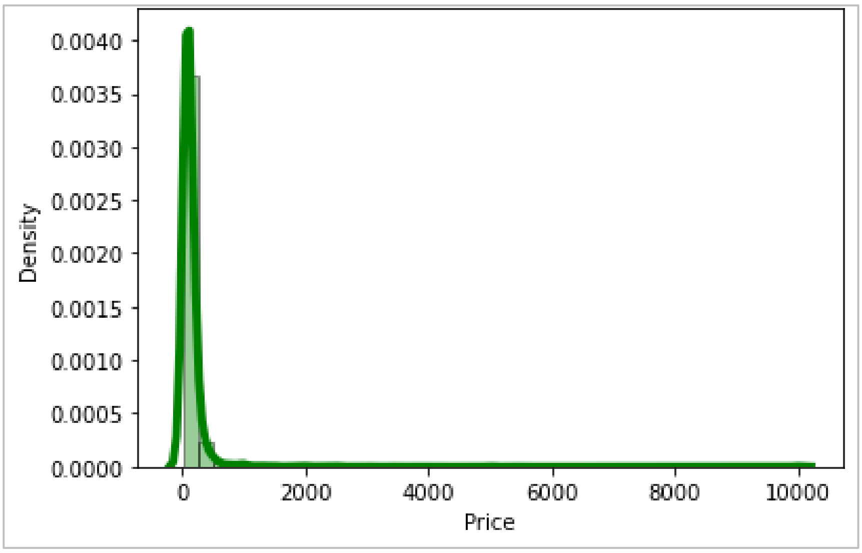


Figure 8: Frequency distribution of 'Price' before outlier treatment

Quantile	Price Value
0.00	$ 10
0.10	$ 45
0.25	$ 65
0.30	$ 70
0.50	$ 95
0.60	$ 109
0.75	$ 149

0.80	$ 165
0.90	$ 247
0.95	$ 350
0.98	$ 750
0.99	$ 1,350
1.00	$ 10,000

Table 1: Value of ‘Price’ variable for different quantiles

The frequency distribution in Figure 8 represents how ‘price’ is distributed after the removal of outliers

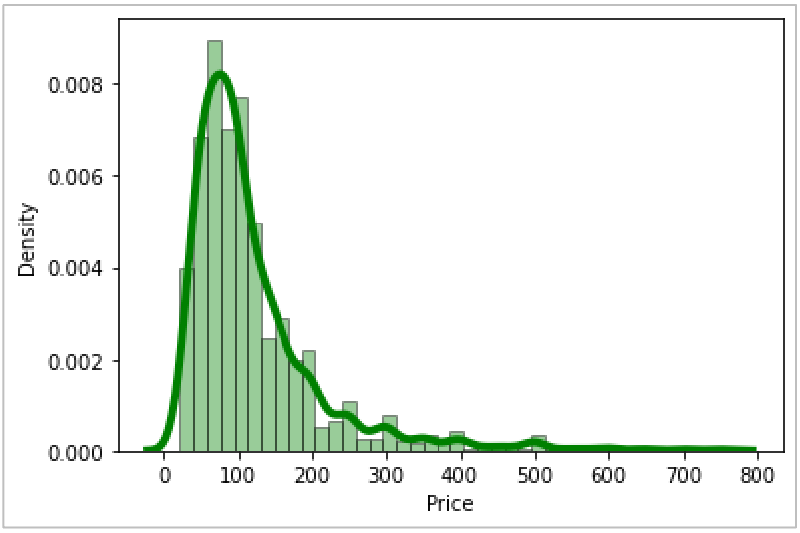
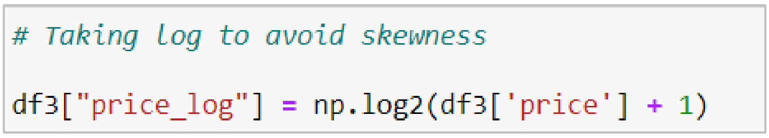


Figure 9: Frequency distribution of ‘Price’ after outlier treatment

2. Log Transformation

It is observed (Figure 9) that even after the removal of outliers, the data is right-skewed. Skewed data impairs the performance of the model, particularly regression-based models. If the data is skewed, the model has to deal with rare instances of extreme values. For example, a model with right-skewed data will predict data points with lower values better as compared to data points with higher values. Hence, it is crucial to eliminate any skewness in the data. For our project, we used logarithmic transformation on the required variable to attain normally distributed data.



```
# Taking log to avoid skewness

df3["price_log"] = np.log2(df3['price'] + 1)
```

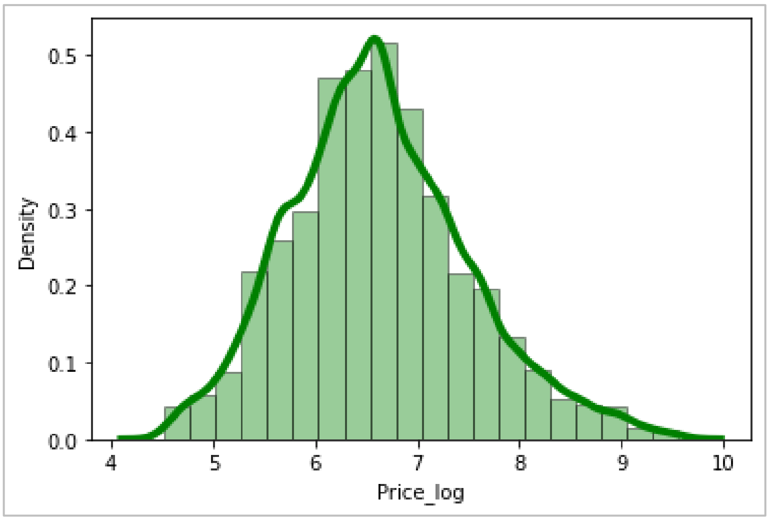
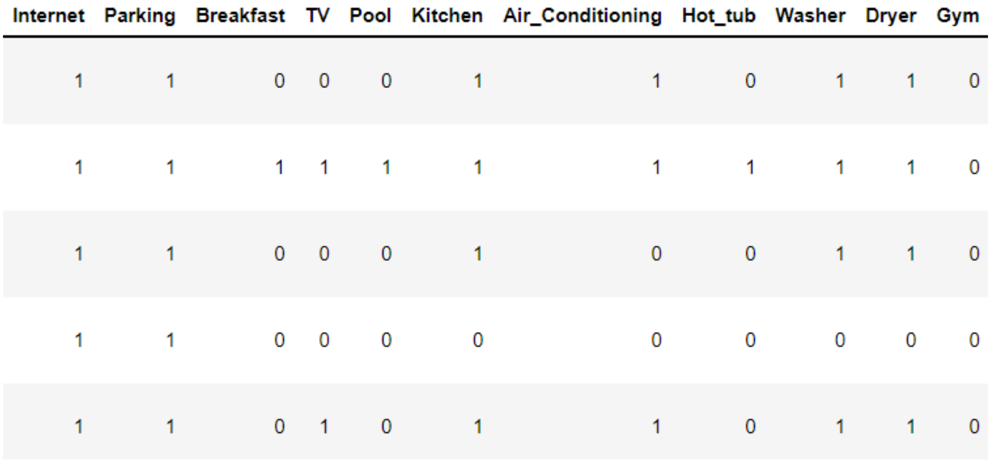


Figure 10: Frequency distribution of log(Price +1)

3. **Conversion of Data Types**
   The columns 'host_is_superhost', 'room_type', 'instant_bookable', 'cancellation_policy' were converted from categorical data type to numeric data type using LabelEncoder.

4. **Introducing new columns**
   In the original data set, all the amenities information was consolidated in a single column. The amenities information was parsed and converted into new columns with values 0 and 1 depending on whether the amenity is available or not. The following columns were added to the data set - Internet, Parking, Breakfast, TV, Pool, Kitchen, Air Conditioning, Hot tub, Washer, Dryer, and Gym.



Internet	Parking	Breakfast	TV	Pool	Kitchen	Air_Conditioning	Hot_tub	Washer	Dryer	Gym
1	1	0	0	0	1	1	0	1	1	0
1	1	1	1	1	1	1	1	1	1	0
1	1	0	0	0	1	0	0	1	1	0
1	1	0	0	0	0	0	0	0	0	0
1	1	0	1	0	1	1	0	1	1	0

Figure 11: Newly added columns to include amenities information

3.2.1. Decision Tree Regressor

A decision tree builds models in the form of a tree structure. It can handle both categorical as well as continuous dependent variable. For our project, we would be using the decision tree regressor. It breaks down the data set into smaller subsets while the decision tree is developed at each step incrementally. Finally, a tree with decision nodes and leaf nodes is constructed. A decision node has two or more branches while the leaf node represents a decision on the dependent variable. The root node is the topmost decision node in a tree and corresponds to the best predictor.

In order to identify the optimal number of leaf nodes, a graph is plotted between MAE (Mean absolute error) and number of leaf nodes. We have considered the node with the least mean absolute error as the optimal node. It is observed (Figure 12) that **n = 45** has the least MAE of 35.55. The attribute 'room type' was the root node for the model.

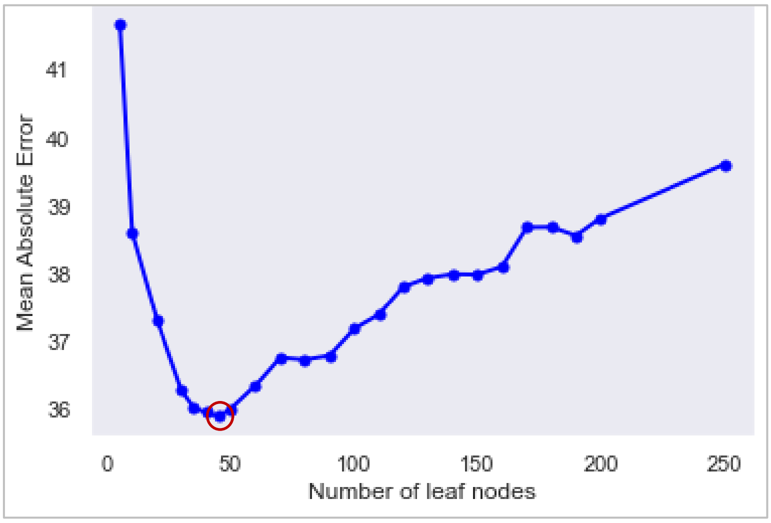


Figure 12: MAE vs Number of leaf nodes

The following measures are analysed in order to under the performance of the model:

1. R Squared – explains the amount of variation in dependent variable due to the variation in the feature variables
2. Mean Absolute Error – The absolute value of the difference between the predicted values and actual values
3. Mean Absolute Percentage Error – measures the accuracy of the model and represents the average of the absolute percentage errors of each instance in the dataset.

R Squared	51.15%
Mean Absolute Error (MAE)	35.55
Mean Absolute Percentage Error (MAPE)	29.98%

Table 2: Decision Tree Regressor performance measures

3.2.2 K Nearest Neighbor (KNN)

KNN is a supervised learning, non-parametric method, which uses proximity to make classifications or predictions about the grouping of an instance. In KNN regression, the output is the average of the values of *k* nearest neighbors.

In order to pin down the ideal count of nearest neighbors, a graph is plotted between mean absolute error and nearest neighbors. It is observed (Figure 13) for n = 17, the mean absolute error is the least. To evaluate the performance of the model, R2, MAE and MAPE is calculated.

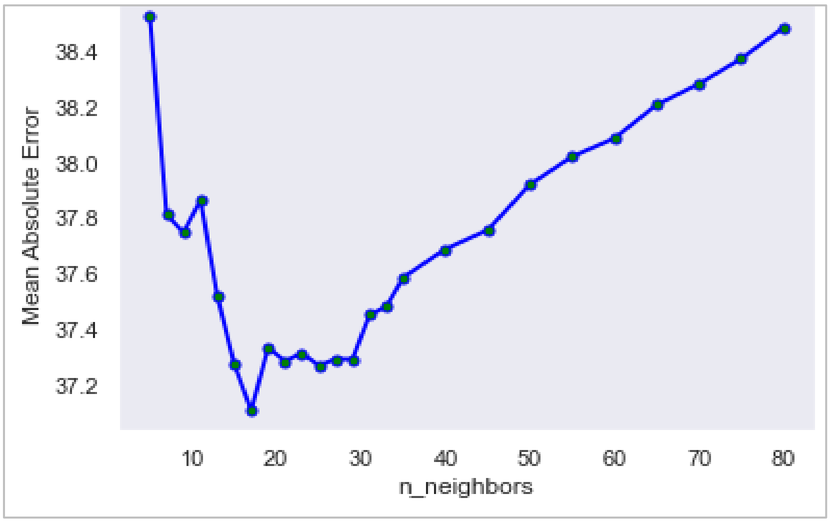


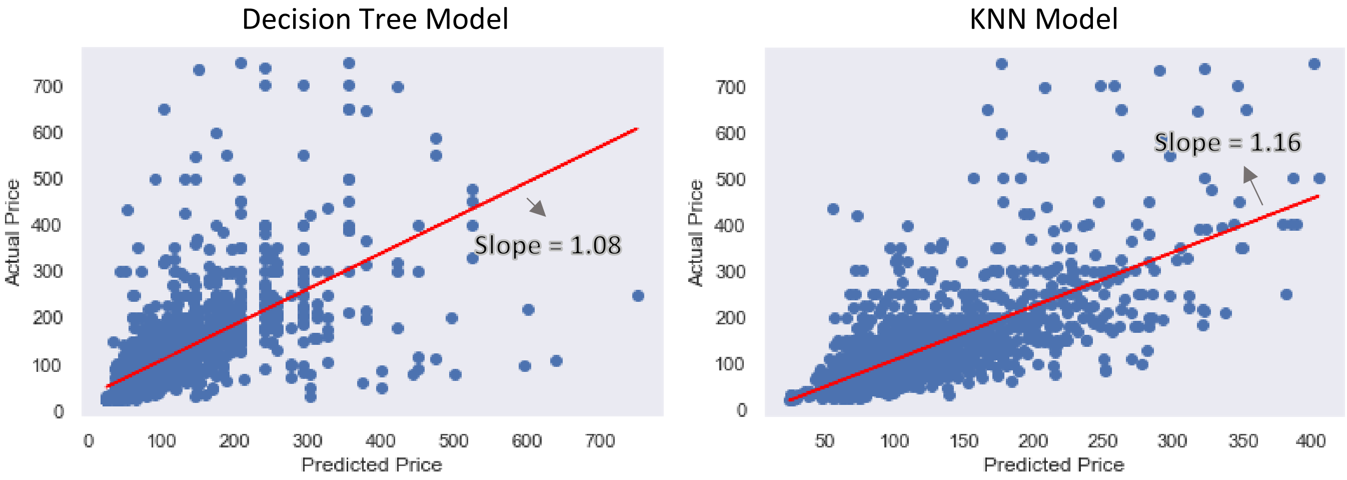
Figure 13: Mean Absolute Error vs n-neighbors

R Squared	50.65%
Mean Absolute Error (MAE)	37.10
Mean Absolute Percentage Error (MAPE)	32.51%

Table 3: KNN Model Performance measures

4. Conclusion

In conclusion we have carefully evaluated our analysis on Airbnb price prediction using the three models. Due to the limitations with categorical variables and assumption of independent variables, Naïve Bayes' performance in this analysis did not compare to that of the Decision Tree and KNN model. It is observed that MAE and MAPE is lesser for Decision Tree as compared to KNN by 4.2% and 7.8% respectively. The R squared for Decision Tree is higher than KNN model by 1%. Additionally, when we analyse the scatter plot between actual price and predicted price, we observe that slope of the fitted line for Decision Tree model is closer to 1 as compared to slope of the fitted line for KNN model



Decision Tree Model

KNN Model

Figure 14: Scatter Plot between Predicted Price and Actual Price

After examining the performance of all the models, we can conclude that Decision Tree Regressor does a better job at predicting the nightly prices.

5. Data Limitations & Key Takeaways

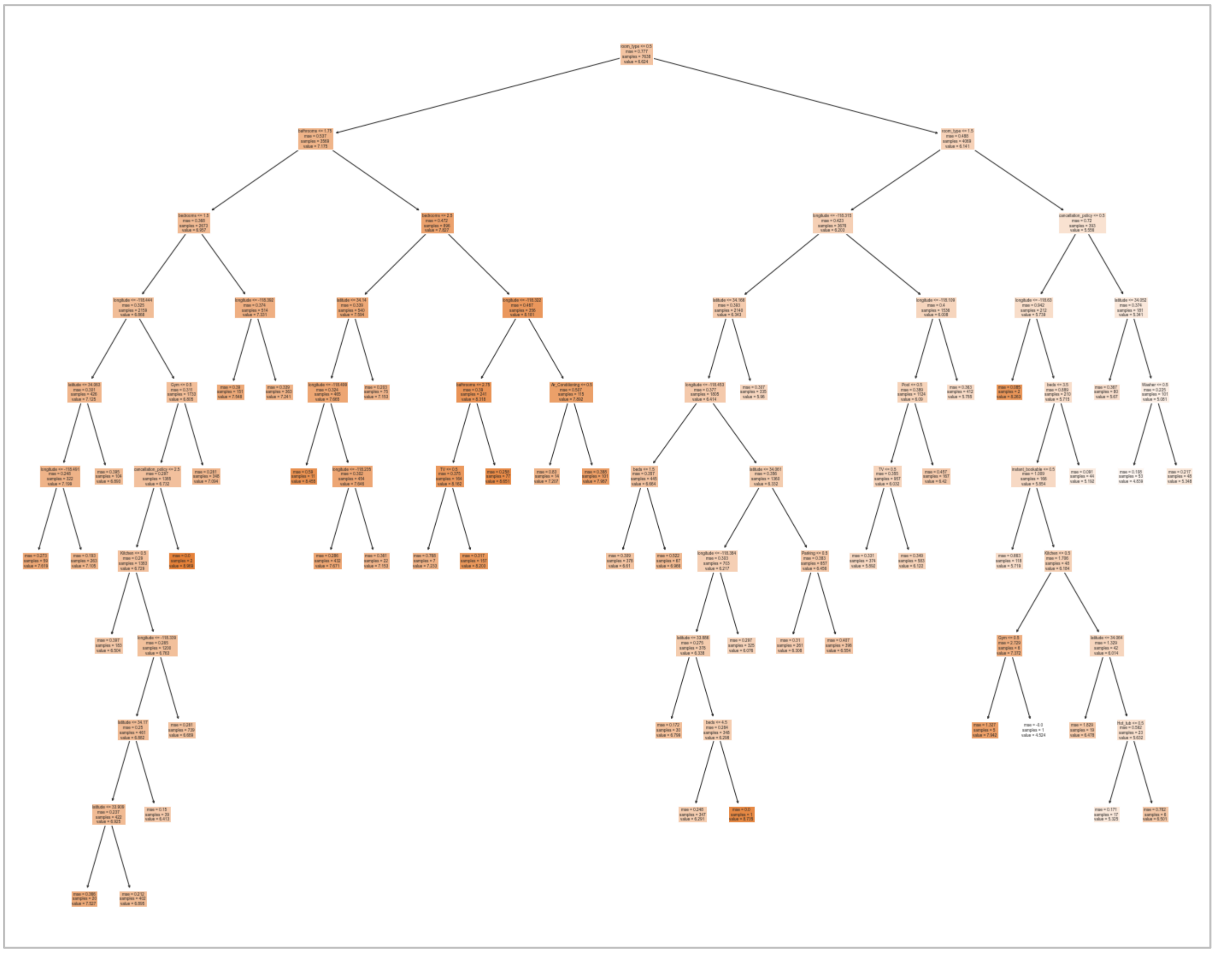
- The ‘cleaning_fee’ column of our dataset predominantly had null values and needed to be dropped before modelling. However, cleaning fee is one of the most important feature that hosts need to consider before pricing their rental space. Having this data point in the future would help in predicting nightly prices better.
- Understood the impact of skewed data on model performance
- Understood the relevance of exploratory data analysis and modelling on real life applications
- For Guassian Naïve Bayes, the main limitation is the assumption of independent predictor features, while in real life, it is almost very unlikely to get a set of features that are mutually independent
- Decision Tree outputs are easy to interpret and visualize. However, the cost of building a decision tree is high for large datasets as each node required field sorting
- For KNN model, the main limitation is that it does not work well with large datasets. As it is a distance based algorithm, the cost of computing distance between a new point and existing point is very hugh which in turn impairs the performance of the model

6. Future Scope

- We have incorporated three models as part of this project. More models such can be developed to understand and predit the data better
- A weighted KNN model can be implemented and tested
- Different types of KNN models can be developed and analyed by changing the distance measure - Euclidean distance, cosine similarity measure, Minkowsky, correlation, and Chi square
- Feature selection in KNN models can be implemented. This will allow in reducing the number of features and complexity and thereby improving the performance of the classifier

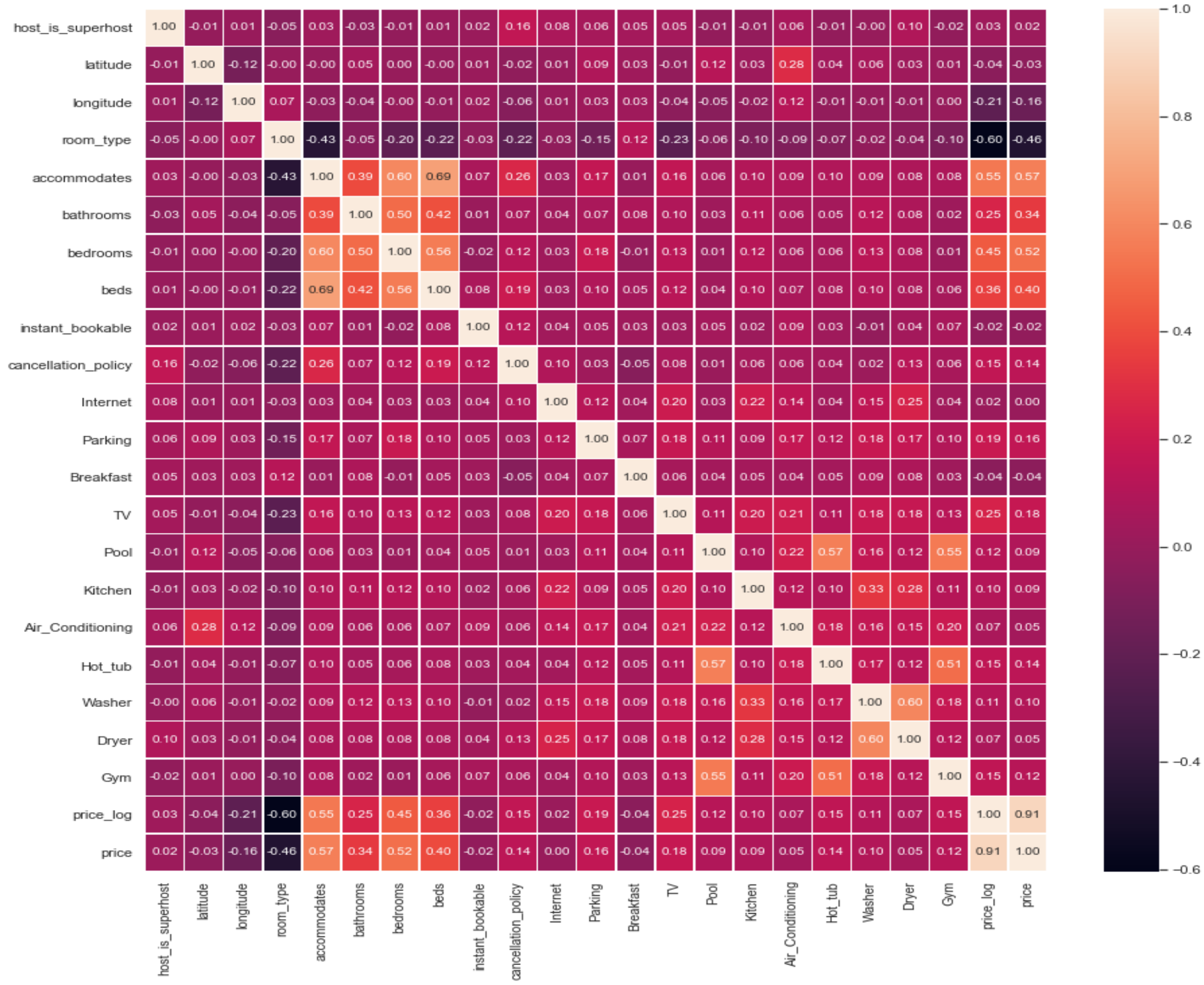
Appendix

1. Decision Tree Diagram for n = 45



room_type <= 0.5 mse = 0.777 samples = 7638 value = 6.624
bathrooms <= 1.75 mse = 0.537 samples = 3569 value = 7.175
room_type <= 1.5 mse = 0.485 samples = 4069 value = 6.141
bedrooms <= 1.5 mse = 0.368 samples = 2673 value = 6.957
bedrooms <= 2.5 mse = 0.472 samples = 896 value = 7.827
longitude <= -118.315 mse = 0.423 samples = 3676 value = 6.203
cancellation_policy <= 0.5 mse = 0.72 samples = 393 value = 5.556
longitude <= -118.444 mse = 0.325 samples = 2159 value = 6.868
longitude <= -118.392 mse = 0.374 samples = 514 value = 7.331
latitude <= 34.14 mse = 0.339 samples = 540 value = 7.594
longitude <= -118.322 mse = 0.467 samples = 356 value = 8.181
latitude <= 34.168 mse = 0.383 samples = 2140 value = 6.343
longitude <= -118.109 mse = 0.4 samples = 1536 value = 6.008
longitude <= -118.63 mse = 0.942 samples = 212 value = 5.739
latitude <= 34.052 mse = 0.374 samples = 181 value = 5.341
latitude <= 34.063 mse = 0.381 samples = 426 value = 7.125
Gym <= 0.5 mse = 0.311 samples = 1733 value = 6.806
mse = 0.39 samples = 151 value = 7.548
mse = 0.339 samples = 363 value = 7.241
longitude <= -118.499 mse = 0.324 samples = 465 value = 7.668
mse = 0.283 samples = 75 value = 7.153
bathrooms <= 2.75 mse = 0.39 samples = 241 value = 8.318
Air_Conditioning <= 0.5 mse = 0.587 samples = 115 value = 7.892
longitude <= -118.453 mse = 0.377 samples = 1806 value = 6.414
mse = 0.387 samples = 335 value = 5.96
Pool <= 0.5 mse = 0.389 samples = 1124 value = 6.09
mse = 0.363 samples = 412 value = 5.765
mse = 0.085 samples = 2 value = 8.263
beds <= 3.5 mse = 0.889 samples = 210 value = 5.715
mse = 0.387 samples = 80 value = 5.67
Washer <= 0.5 mse = 0.225 samples = 101 value = 5.081
longitude <= -118.491 mse = 0.248 samples = 322 value = 7.199
mse = 0.395 samples = 104 value = 6.893
cancellation_policy <= 2.5 mse = 0.297 samples = 1385 value = 6.732
mse = 0.281 samples = 348 value = 7.094
mse = 0.59 samples = 11 value = 8.455
longitude <= -118.235 mse = 0.302 samples = 454 value = 7.646
TV <= 0.5 mse = 0.375 samples = 164 value = 8.162
mse = 0.255 samples = 77 value = 8.651
mse = 0.83 samples = 14 value = 7.207
mse = 0.385 samples = 101 value = 7.987
beds <= 1.5 mse = 0.387 samples = 445 value = 6.664
latitude <= 34.061 mse = 0.356 samples = 1360 value = 6.332
TV <= 0.5 mse = 0.355 samples = 957 value = 6.032
mse = 0.457 samples = 167 value = 6.42
instant_bookable <= 0.5 mse = 1.009 samples = 166 value = 5.854
mse = 0.091 samples = 44 value = 5.192
mse = 0.188 samples = 53 value = 4.839
mse = 0.217 samples = 48 value = 5.348
mse = 0.273 samples = 59 value = 7.619
mse = 0.193 samples = 263 value = 7.105
Kitchen <= 0.5 mse = 0.29 samples = 1383 value = 6.729
mse = 0.0 samples = 2 value = 8.969
mse = 0.286 samples = 432 value = 7.671
mse = 0.381 samples = 22 value = 7.153
mse = 0.788 samples = 7 value = 7.233
mse = 0.317 samples = 157 value = 8.203
mse = 0.389 samples = 378 value = 6.61
mse = 0.522 samples = 67 value = 6.968
longitude <= -118.384 mse = 0.383 samples = 703 value = 6.217
Parking <= 0.5 mse = 0.383 samples = 657 value = 6.458
mse = 0.331 samples = 374 value = 5.892
mse = 0.349 samples = 583 value = 6.122
mse = 0.663 samples = 118 value = 5.719
Kitchen <= 0.5 mse = 1.786 samples = 48 value = 6.184
mse = 0.397 samples = 183 value = 6.504
longitude <= -118.339 mse = 0.285 samples = 1200 value = 6.763
latitude <= 33.888 mse = 0.275 samples = 378 value = 6.338
mse = 0.287 samples = 325 value = 6.078
mse = 0.31 samples = 261 value = 6.308
mse = 0.487 samples = 396 value = 6.554
Gym <= 0.5 mse = 2.729 samples = 6 value = 7.372
latitude <= 34.064 mse = 1.329 samples = 42 value = 6.014
latitude <= 34.17 mse = 0.25 samples = 461 value = 6.882
mse = 0.281 samples = 739 value = 6.689
mse = 0.172 samples = 30 value = 6.799
beds <= 4.5 mse = 0.284 samples = 348 value = 6.298
mse = 1.327 samples = 5 value = 7.942
mse = -0.0 samples = 1 value = 4.524
mse = 1.829 samples = 19 value = 6.478
Hot_tub <= 0.5 mse = 0.582 samples = 23 value = 5.632
latitude <= 33.909 mse = 0.237 samples = 422 value = 6.925
mse = 0.15 samples = 39 value = 6.413
mse = 0.248 samples = 347 value = 6.291
mse = 0.0 samples = 1 value = 8.735
mse = 0.171 samples = 17 value = 5.325
mse = 0.782 samples = 6 value = 6.501
mse = 0.386 samples = 20 value = 7.527
mse = 0.212 samples = 402 value = 6.895

2. Correlation Plot



	host_is_superhost	latitude	longitude	room_type	accommodates	bathrooms	bedrooms	beds	instant_bookable	cancellation_policy	Internet	Parking	Breakfast	TV	Pool	Kitchen	Air_Conditioning	Hot_tub	Washer	Dryer	Gym	price_log	price
host_is_superhost	1.00	-0.01	0.01	-0.05	0.03	-0.03	-0.01	0.01	0.02	0.16	0.08	0.06	0.05	0.05	-0.01	-0.01	0.06	-0.01	-0.00	0.10	-0.02	0.03	0.02
latitude	-0.01	1.00	-0.12	-0.00	-0.00	0.05	0.00	-0.00	0.01	-0.02	0.01	0.09	0.03	-0.01	0.12	0.03	0.28	0.04	0.06	0.03	0.01	-0.04	-0.03
longitude	0.01	-0.12	1.00	0.07	-0.03	-0.04	-0.00	-0.01	0.02	-0.06	0.01	0.03	0.03	-0.04	-0.05	-0.02	0.12	-0.01	-0.01	-0.01	0.00	-0.21	-0.16
room_type	-0.05	-0.00	0.07	1.00	-0.43	-0.05	-0.20	-0.22	-0.03	-0.22	-0.03	-0.15	0.12	-0.23	-0.06	-0.10	-0.09	-0.07	-0.02	-0.04	-0.10	-0.60	-0.46
accommodates	0.03	-0.00	-0.03	-0.43	1.00	0.39	0.60	0.69	0.07	0.26	0.03	0.17	0.01	0.16	0.06	0.10	0.09	0.10	0.09	0.08	0.08	0.55	0.57
bathrooms	-0.03	0.05	-0.04	-0.05	0.39	1.00	0.50	0.42	0.01	0.07	0.04	0.07	0.08	0.10	0.03	0.11	0.06	0.05	0.12	0.08	0.02	0.25	0.34
bedrooms	-0.01	0.00	-0.00	-0.20	0.60	0.50	1.00	0.56	-0.02	0.12	0.03	0.18	-0.01	0.13	0.01	0.12	0.06	0.06	0.13	0.08	0.01	0.45	0.52
beds	0.01	-0.00	-0.01	-0.22	0.69	0.42	0.56	1.00	0.08	0.19	0.03	0.10	0.05	0.12	0.04	0.10	0.07	0.08	0.10	0.08	0.06	0.36	0.40
instant_bookable	0.02	0.01	0.02	-0.03	0.07	0.01	-0.02	0.08	1.00	0.12	0.04	0.05	0.03	0.03	0.05	0.02	0.09	0.03	-0.01	0.04	0.07	-0.02	-0.02
cancellation_policy	0.16	-0.02	-0.06	-0.22	0.26	0.07	0.12	0.19	0.12	1.00	0.10	0.03	-0.05	0.08	0.01	0.06	0.06	0.04	0.02	0.13	0.06	0.15	0.14
Internet	0.08	0.01	0.01	-0.03	0.03	0.04	0.03	0.03	0.04	0.10	1.00	0.12	0.04	0.20	0.03	0.22	0.14	0.04	0.15	0.25	0.04	0.02	0.00
Parking	0.06	0.09	0.03	-0.15	0.17	0.07	0.18	0.10	0.05	0.03	0.12	1.00	0.07	0.18	0.11	0.09	0.17	0.12	0.18	0.17	0.10	0.19	0.16
Breakfast	0.05	0.03	0.03	0.12	0.01	0.08	-0.01	0.05	0.03	-0.05	0.04	0.07	1.00	0.06	0.04	0.05	0.04	0.05	0.09	0.08	0.03	-0.04	-0.04
TV	0.05	-0.01	-0.04	-0.23	0.16	0.10	0.13	0.12	0.03	0.08	0.20	0.18	0.06	1.00	0.11	0.20	0.21	0.11	0.18	0.18	0.13	0.25	0.18
Pool	-0.01	0.12	-0.05	-0.06	0.06	0.03	0.01	0.04	0.05	0.01	0.03	0.11	0.04	0.11	1.00	0.10	0.22	0.57	0.16	0.12	0.55	0.12	0.09
Kitchen	-0.01	0.03	-0.02	-0.10	0.10	0.11	0.12	0.10	0.02	0.06	0.22	0.09	0.05	0.20	0.10	1.00	0.12	0.10	0.33	0.28	0.11	0.10	0.09
Air_Conditioning	0.06	0.28	0.12	-0.09	0.09	0.06	0.06	0.07	0.09	0.06	0.14	0.17	0.04	0.21	0.22	0.12	1.00	0.18	0.16	0.15	0.20	0.07	0.05
Hot_tub	-0.01	0.04	-0.01	-0.07	0.10	0.05	0.06	0.08	0.03	0.04	0.04	0.12	0.05	0.11	0.57	0.10	0.18	1.00	0.17	0.12	0.51	0.15	0.14
Washer	-0.00	0.06	-0.01	-0.02	0.09	0.12	0.13	0.10	-0.01	0.02	0.15	0.18	0.09	0.18	0.16	0.33	0.16	0.17	1.00	0.60	0.18	0.11	0.10
Dryer	0.10	0.03	-0.01	-0.04	0.08	0.08	0.08	0.08	0.04	0.13	0.25	0.17	0.08	0.18	0.12	0.28	0.15	0.12	0.60	1.00	0.12	0.07	0.05
Gym	-0.02	0.01	0.00	-0.10	0.08	0.02	0.01	0.06	0.07	0.06	0.04	0.10	0.03	0.13	0.55	0.11	0.20	0.51	0.18	0.12	1.00	0.15	0.12
price_log	0.03	-0.04	-0.21	-0.60	0.55	0.25	0.45	0.36	-0.02	0.15	0.02	0.19	-0.04	0.25	0.12	0.10	0.07	0.15	0.11	0.07	0.15	1.00	0.91
price	0.02	-0.03	-0.16	-0.46	0.57	0.34	0.52	0.40	-0.02	0.14	0.00	0.16	-0.04	0.18	0.09	0.09	0.05	0.14	0.10	0.05	0.12	0.91	1.00