

PROJECT REPORT

DAB 402

AIRBNB PRICE PREDICTION IN CANADA

BY

JYOTI BALA	0782019
RAIHAANAH ABUBAKKAR SIDIQ	0785786
HARSH KUMAR	0791250
JOYAL PATEL	0792200
EZEKIEL AYENI	0778486

AGENDA

- Introduction
- Why Airbnb
- Motivation
- Problem Identification
- Challenges
- Target Audience
- Analytical Problem and Uses
- Hypotheses
- Datasets and Sources
- Data Assessment and Pre-Processing
- Exploratory Data Analysis
- Prediction Model and Accuracy
- Prediction Observation
- Tableau Analysis
- Price Dashboard (Effect of Price on Airbnb in Canada).
- Reviews Dashboard (Effect of Reviews on Airbnb in Canada).
- Conclusion and future work

INTRODUCTION

Airbnb, Inc. is an American based company operating an online marketplace for short- term homestays and experiences. The company acts as a broker and charges a commission from each booking. The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com. The company is credited with revolutionizing the tourism industry.

Since its inception, Airbnb has hosted more than 60 million guests in 34,000 locations worldwide, and it has experienced rapid growth. For those who would otherwise have empty space as well as visitors looking for convenient and cheap housing options, Airbnb offers a convenient source of income.

The Canadian economy could gain \$85 million from customer taxes and fees generated by the Airbnb industry in Canada.

WHY AIRBNB

Through the popular platform, Airbnb, local hosts who are renting out their homes or other properties may interact with tourists. Following are a few explanations for why people utilise Airbnb:

Economical: When travelling with a large group or for a longer period of time, Airbnb lodging is frequently less expensive than hotels.

Given that you frequently stay in a residential neighbourhood rather than a popular tourist destination, staying in an Airbnb might provide you with a more genuine sense of the local culture.

Amenities: Compared to regular hotels, Airbnb properties frequently provide more amenities including kitchens, washers, and outside areas.

In general, travellers looking for affordable, distinctive, and flexible lodging may find Airbnb to be a wonderful alternative.

MOTIVATION

Listed below are the motive behind the project while also acknowledging that many things can motivate people to utilise Airbnb, including:

- **Authentic travel encounters:** By booking accommodations in homes and apartments owned by locals, Airbnb gives visitors the chance to get a more genuine sense of the local neighbourhoods and cultures.
- **Savings:** Compared to regular hotels, Airbnb can be more economical, especially for extended stays or large groups.
- **Flexibility:** Airbnb allows for more personalization of the stay and offers more flexibility about check-in and check-out times.
- **Unusual accommodation:** Unlike regular hotels, Airbnb offers a wide range of uncommon lodging options, including treehouses, yurts, and houseboats.
- **Native recommendations:** Many Airbnb hosts offer suggestions for nearby restaurants, events, and views, which can help visitors experience a place like a native.

PROBLEM IDENTIFICATION

Airbnb has become an increasingly popular platform for both hosts and guests to find short-term accommodation. Our goal is to provide local, authentic, diverse, inclusive, and sustainable healthy travel that enables individuals to feel at home everywhere. With Airbnb, you can explore the world in a whole new way and find the perfect place to stay for any occasion. Whether you want a charming cabin in the woods or a villa on the coast, we have accommodations for everyone. In order to optimize pricing and offer budget estimates, it is more crucial to predict the prices of Airbnb listings as they multiply throughout various cities.

Accurate prices of Airbnb properties for users and hosts: We would use data-driven techniques and machine learning algorithms. We would start by gathering and examining the data to look for patterns and connections between the variables. Additionally, we will categorize cities into three groups based on their populations (mega, mid, and small).

Customer reviews based on the stay: Many customers are facing issues regarding the stay and the main issue is their policy of fixed stay on some properties.

To solve the Airbnb price prediction problem, we would use data-driven techniques and machine learning algorithms.

We would start by gathering and examining the data to look for patterns and connections between the variables. Additionally, we will categorize cities into three groups based on their populations (mega, mid, and small).

CHALLENGES

- Selecting suitable dataset from the Airbnb website because it contains data from several cities around the world and grouping the data into different segments.
- Describing how machine learning and predictive modelling algorithms generate forecasts.
- The approach might be applied, for instance, to unfairly discriminate against certain renter groups, such as low-income households or renters who possess particular traits.
- Additionally, it could be challenging to explain how predictive models and machine learning algorithms arrive at specific forecasts, which could result in a lack of transparency. Therefore, before applying any solution, it is crucial to think about any potential ethical ramifications.

TARGET AUDIENCE

Airbnb price prediction will be highly instrumental for traveler's and individuals or group of individuals embarking on a vacation or visits to cities in Canada.

The project will also make it possible for Airbnb hosts and owners to gain a reasonable insight into the prevailing cost of Airbnb within the cities.

ANALYTICAL PROBLEM AND USES

- Directly load various datasets to python then we merged and cleaned it.
- Correct prediction of price leading to easy selection of Airbnb across cities in Canada.
- Customer Satisfaction and reviews would be maintained by solving their issues.

HYPOTHESIS

We have decided to test two hypotheses. Firstly, we want to see if City type has effect on price of Airbnb. It is important for travellers to get an insight into the type of city they intend to select for their vacation and see at a glance how those prices are distributed in different room types, neighbourhoods, cities, and City types across Canada.

Secondly, customer satisfaction is highly imperative in the selection of suitable accommodation. Hence, we have decided to check if reviews and feedbacks from previous users of an accommodation will influence the prices of Airbnb in Canada.

Listed below are the two hypotheses:

- **Hypothesis 1**

Null hypothesis: City type has no effect over Airbnb price.

Alternate hypothesis: City type has an effect over Airbnb price.

- **Hypothesis 2**

Null hypothesis: Reviews has no influence over Airbnb price.

Alternate hypothesis: Reviews has an influence over Airbnb price.

DATASETS AND ITS SOURCES

Airbnb is based in San Francisco, California, the platform is accessible via website and mobile app. The dataset that we would be analyzing was gotten from the inside Airbnb website: <http://insideairbnb.com/get-the-data.html>. We have merged the data and categorized the cities as mid, small, and mega across different provinces in Canada.

DATA ASSESSMENT

We have chosen 6 different Airbnb datasets from different cities in Canada. We have also grouped the data based on the population of the cities into 3 segments namely: Mega (Toronto and Montreal), Mid (Vancouver and Winnipeg), and Small (Quebec City and Victoria)

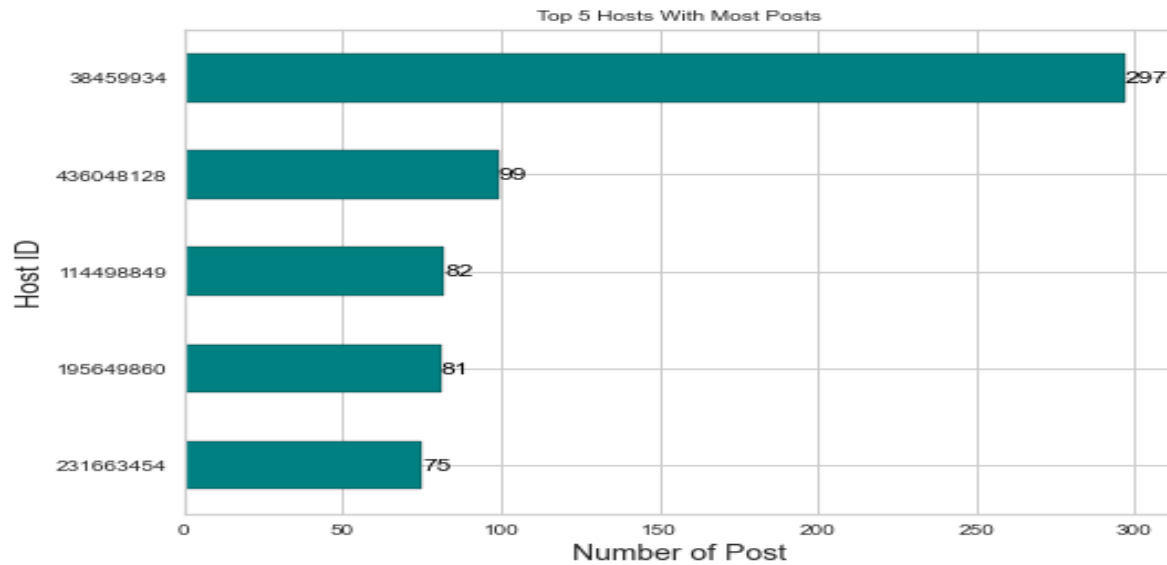
- We segmented cities into Mega, Mid, and Small with respect to population.
- There are 18 variables and 34652 observations in the merged dataset.
- There are no duplicate values in the merged dataset.

PRE-PROCESSING

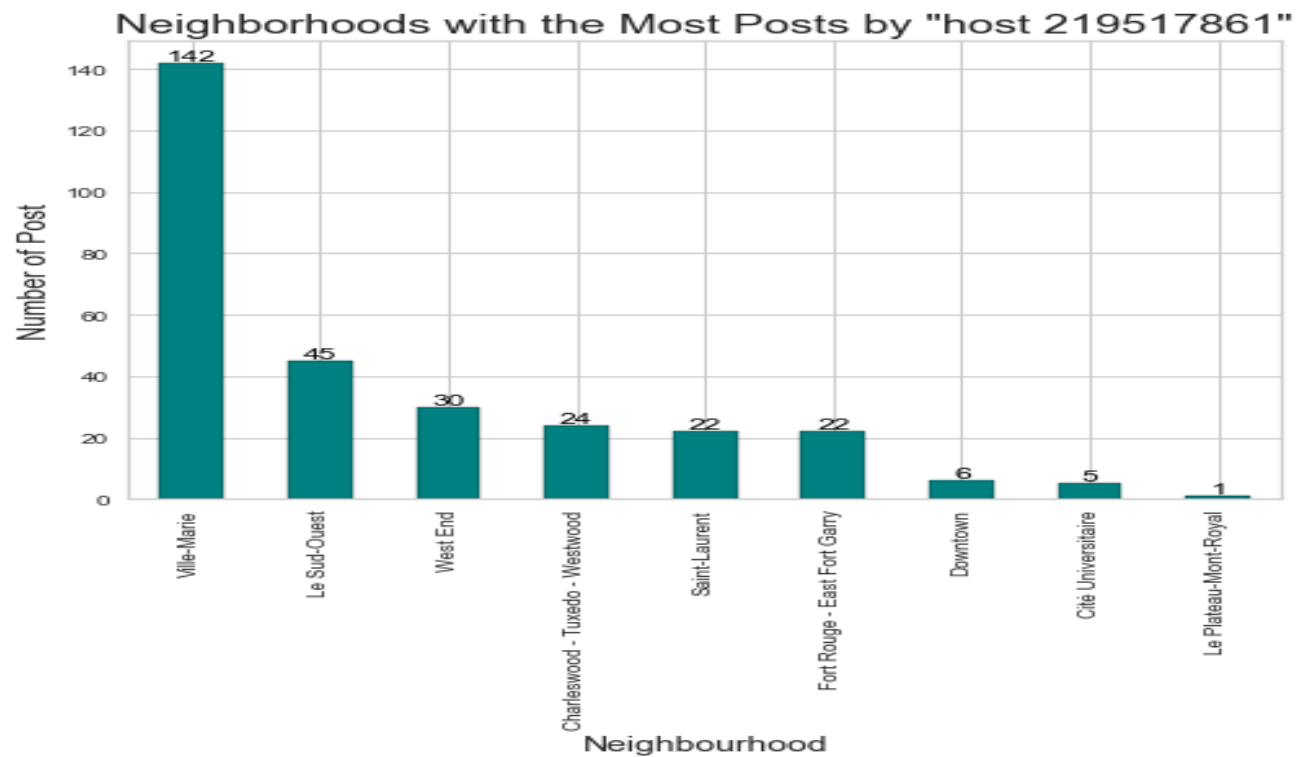
The merged dataset was duly examined and saved by the team. This check comprised of all data cleaning that would be required in terms of incorrect format, inconsistencies, missing values, etc.

EXPLORATORY DATA ANALYSIS (EDA)

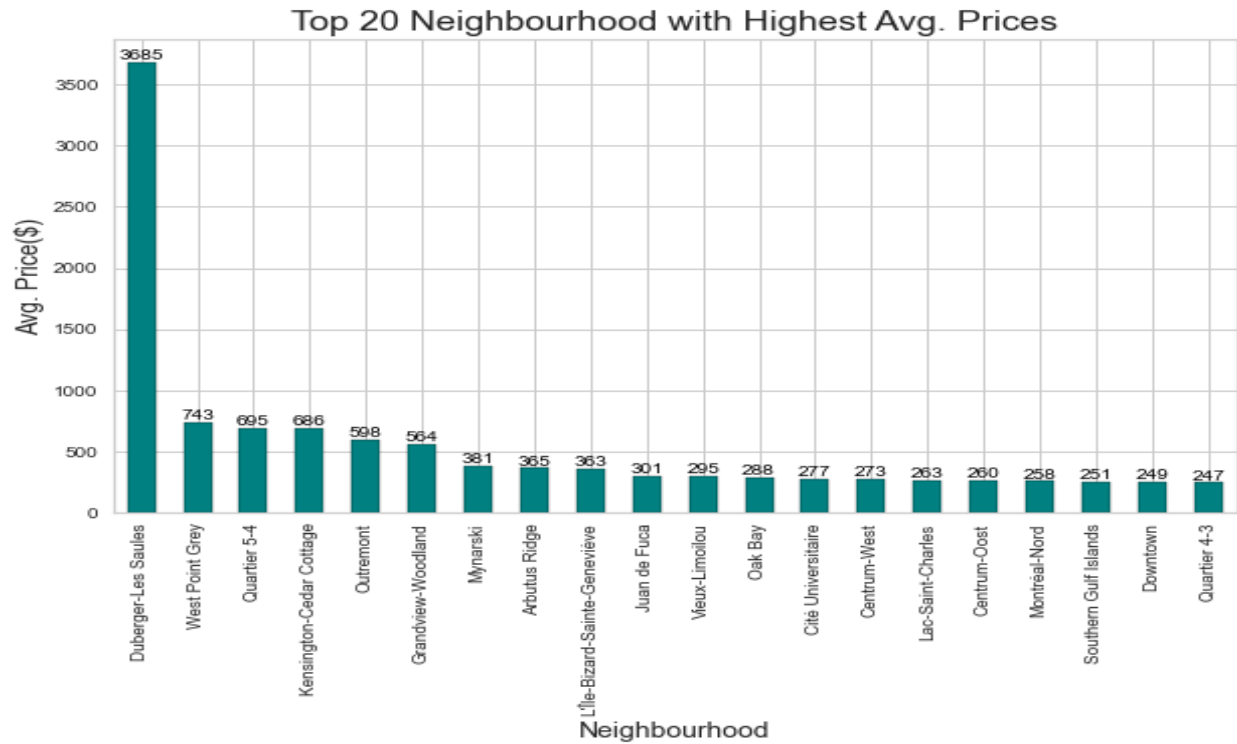
- Using "host_id" column, we can easily analyze the hosts who share the most posts:



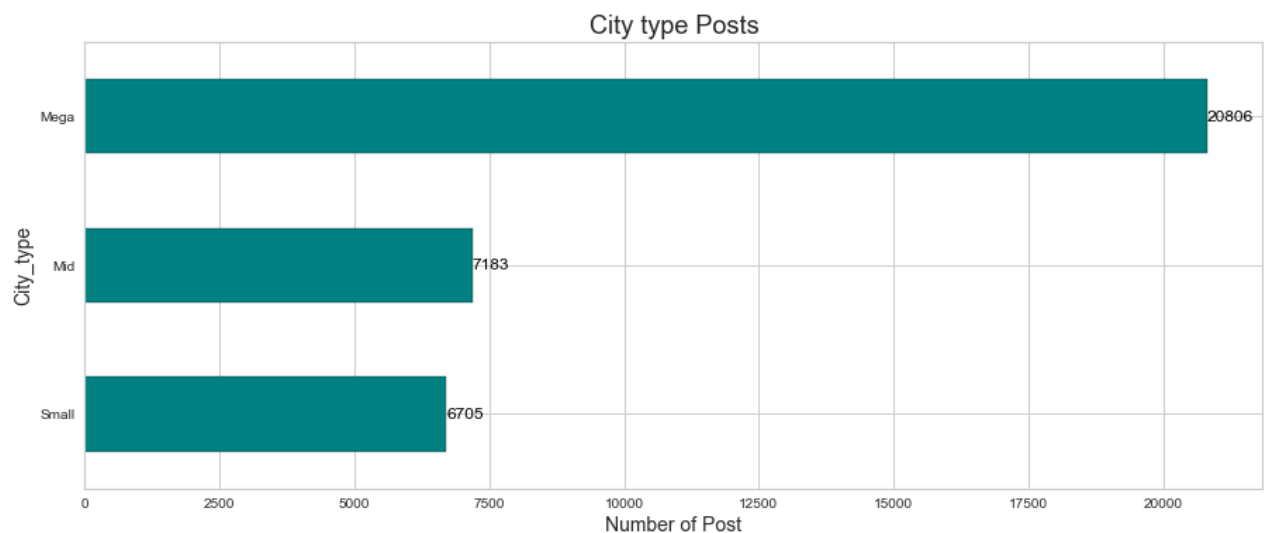
- Neighborhoods with the Most Posts by "host 219517861":



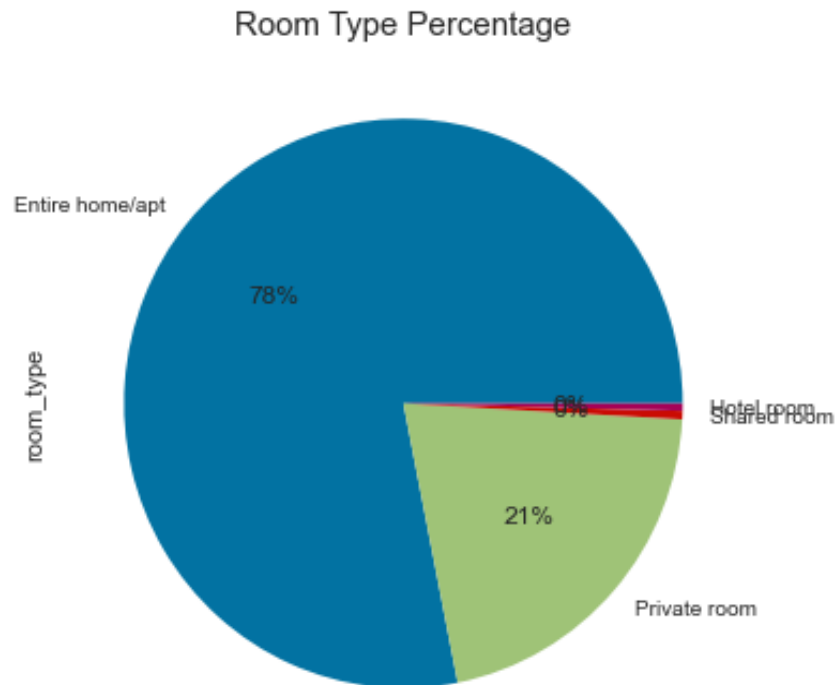
- Top 20 Neighbourhood with Highest Avg. Prices:



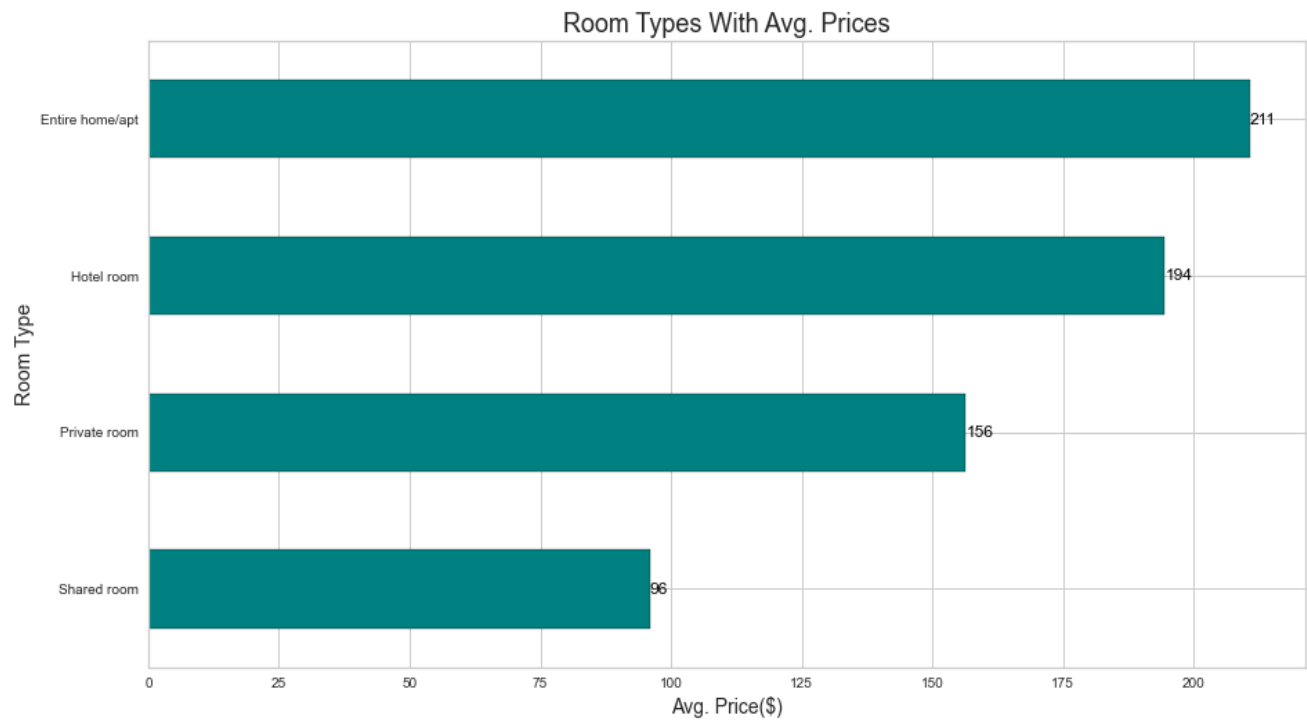
- Using "City type" column, we can easily analyze the City type with the most posts:



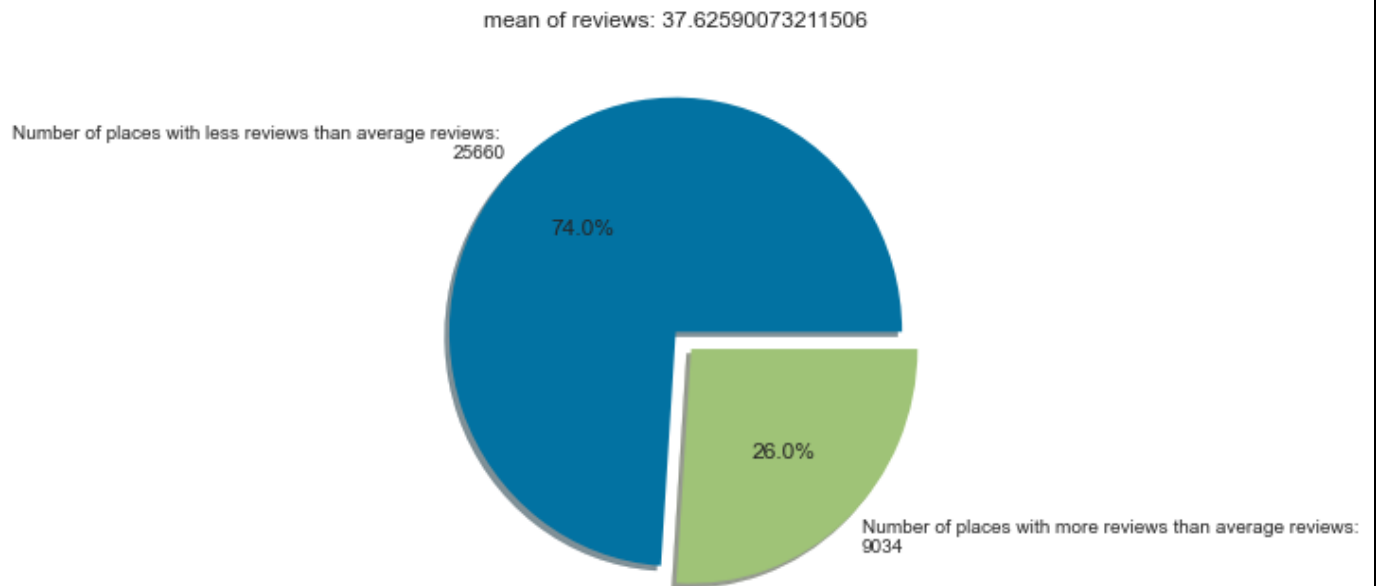
- Room Type Percentage:



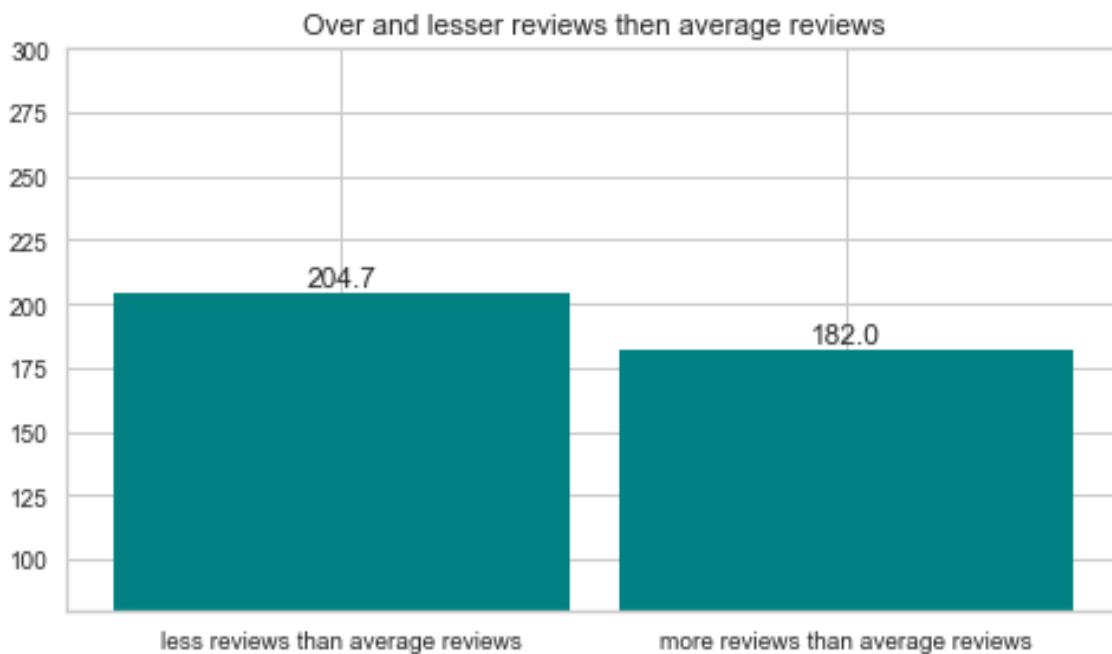
- Room types with average prices:



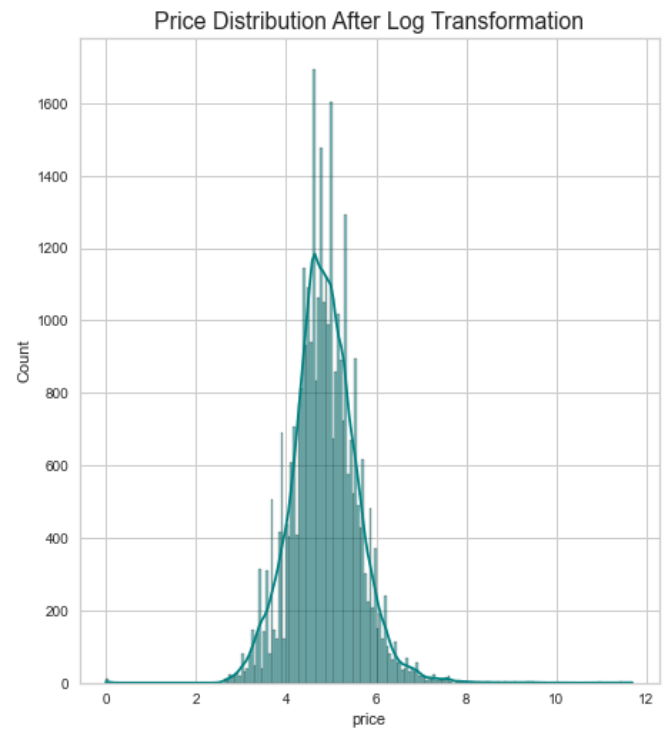
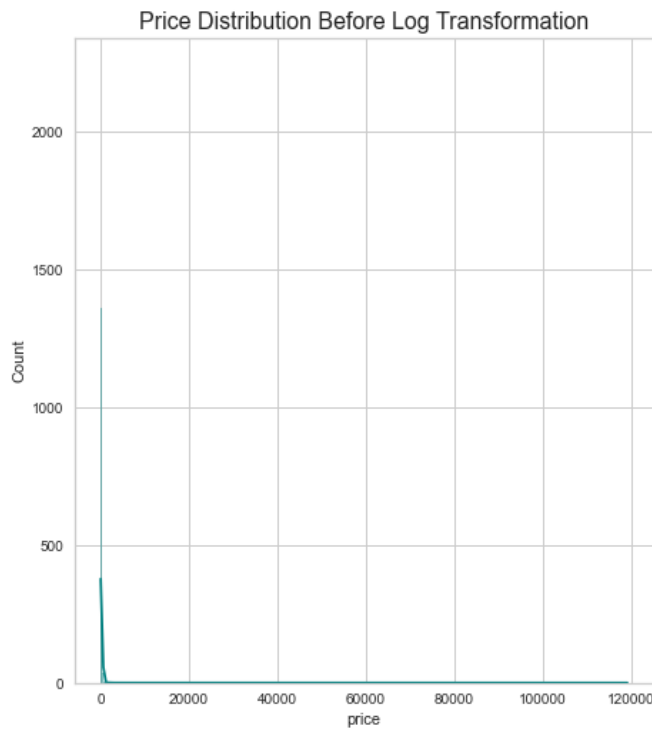
- Here we find the mean of reviews:



- which has more reviews than the average reviews, has significantly less reviews than the average reviews:



- Price Distribution Before/After Log Transformation:



- Checking and removing outliers:

columns = ["price", "minimum_nights", "number_of_reviews",
"reviews_per_month", "calculated_host_listings_count", "availability_365"]

```
: q1_price = df["price"].quantile(0.25)
   q3_price = df["price"].quantile(0.75)
   iqr_price = q3_price - q1_price
   lower_limit_price = q1_price - 1.5 * iqr_price
   upper_limit_price = q3_price + 1.5 * iqr_price

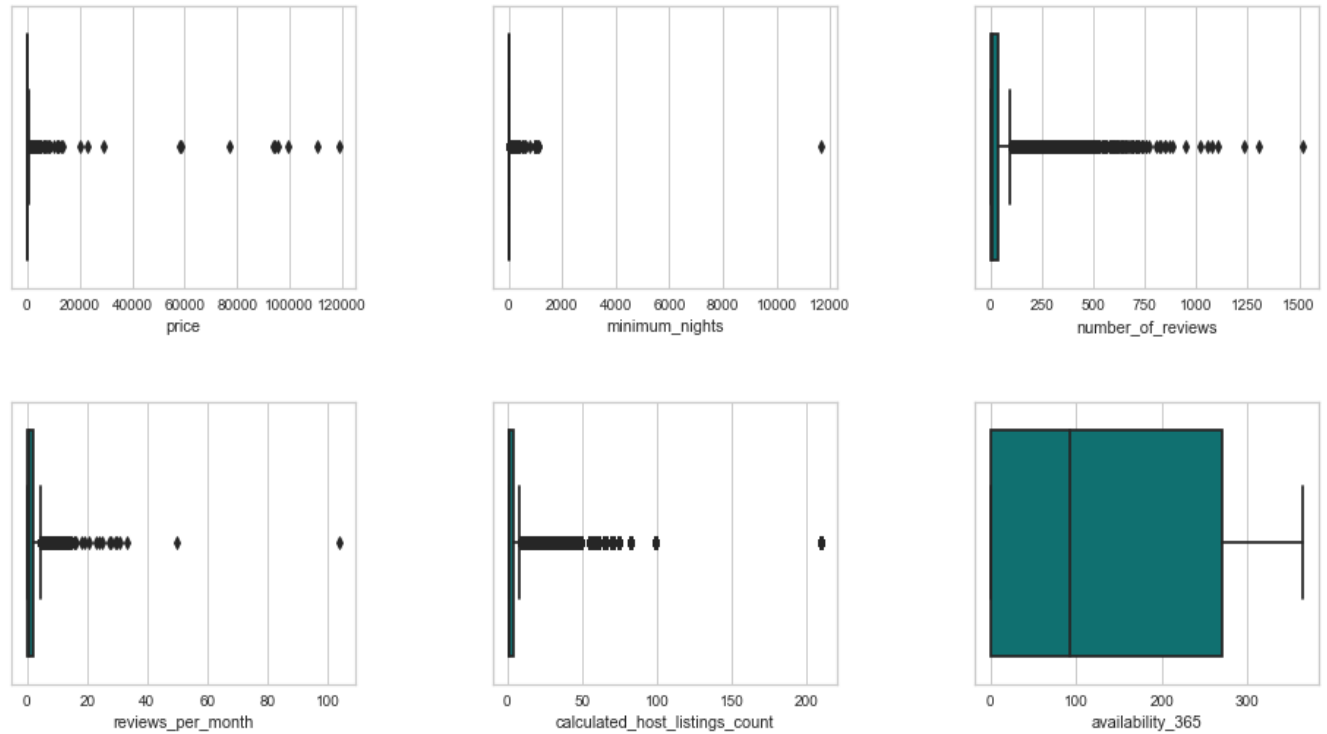
   df_filter_price = df[(df["price"] > lower_limit_price) & (df["price"] < upper_limit_price)]

   df = df_filter_price
   q1_price = df["price"].quantile(0.25)
   q3_price = df["price"].quantile(0.75)
   iqr_price = q3_price - q1_price
   lower_limit_price = q1_price - 1.5 * iqr_price
   upper_limit_price = q3_price + 1.5 * iqr_price

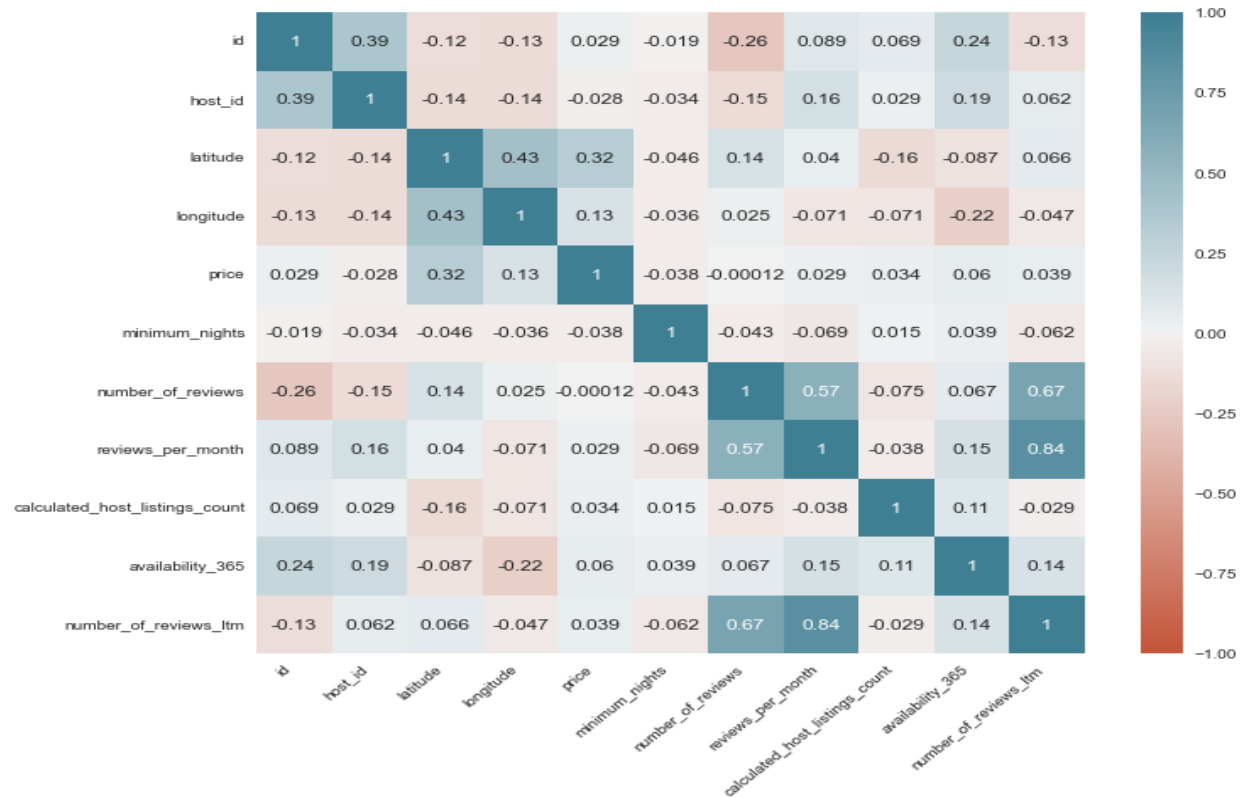
   df_filter_price = df[(df["price"] > lower_limit_price) & (df["price"] < upper_limit_price)]

   df = df_filter_price
```

Group6



- **Plotting correlation:** This shows a weak correlation among the variables.



PREDICTION MODELS

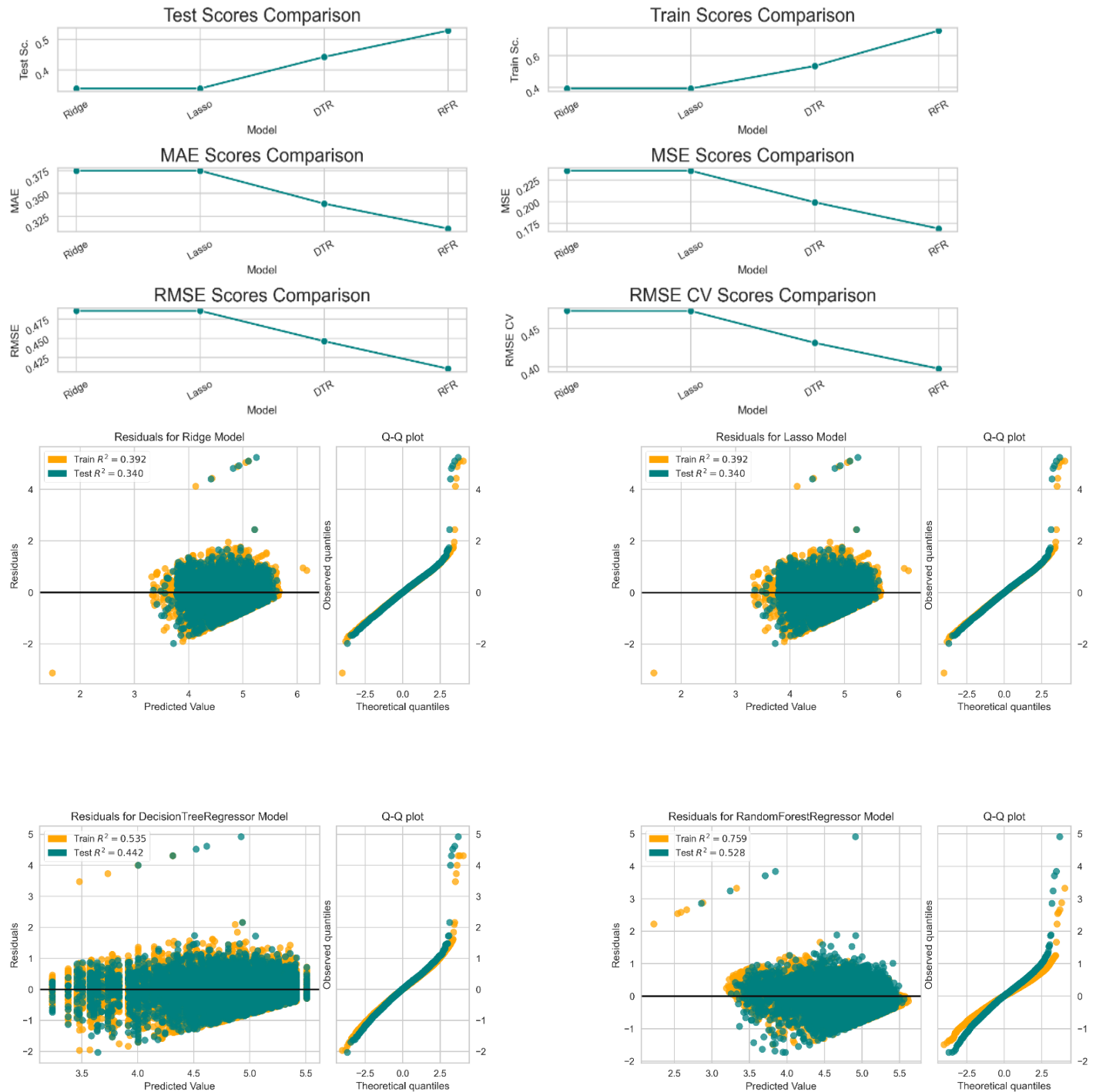
- **Ridge Regression:** Ridge regression is a model-tuning technique that is used to analyze any multicollinear data. L2 regularization is done using this technique. The projected values diverge significantly from the actual values when the multicollinearity problem is present, least-squares are unbiased, and variances are substantial.
- **Lasso:** A regularization method is called lasso regression. For a more accurate forecast, it is preferred to regression techniques. This model makes advantage of shrinkage. When data values shrink toward the mean, this is known as shrinkage. The lasso method promotes simple, sparse models (i.e., models with fewer parameters).
- **Decision Tree Regressor:** A decision tree creates tree-like models for classification or regression. It incrementally develops an associated decision tree while segmenting a dataset into smaller and smaller sections. The outcome is a tree containing leaf nodes and decision nodes.
- **Random Forest Regressor:** A predictor from a random forest. A random forest is a meta estimator that averages many classification decision trees that have been fitted to different dataset subsamples to increase predicted accuracy and reduce overfitting.

```
models(X_train, X_test, y_train, y_test)
```

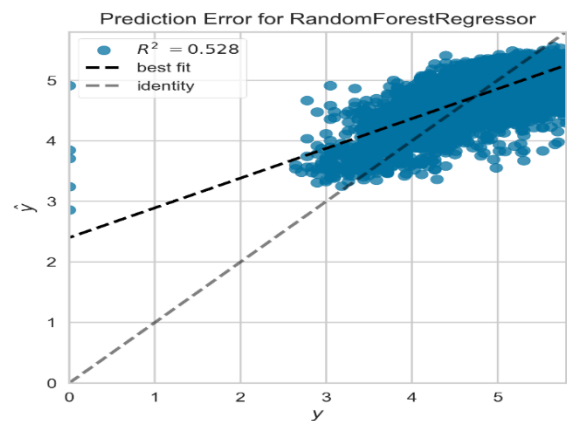
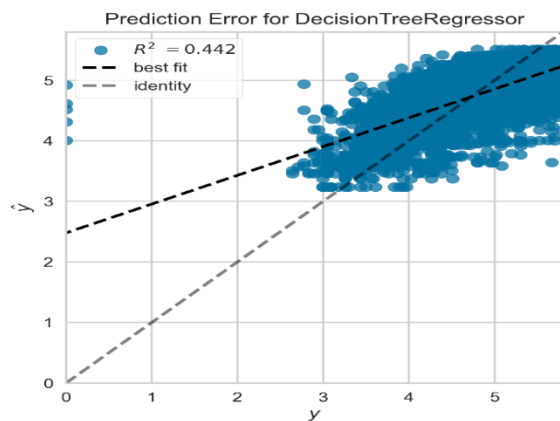
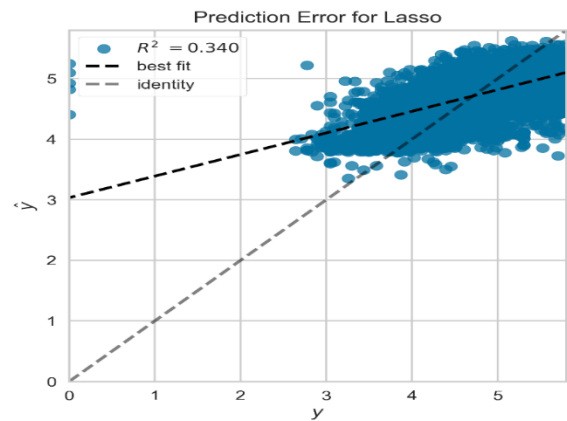
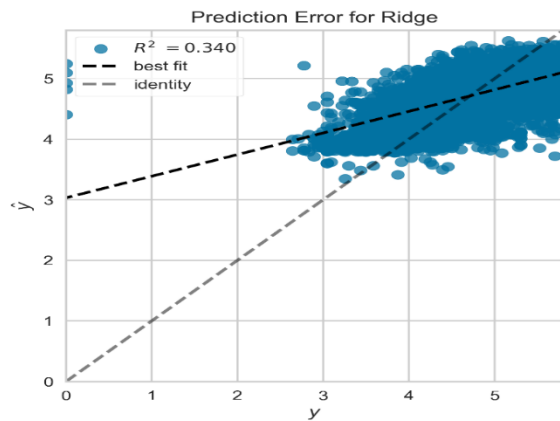
	Model	Test Sc.	Train Sc.	MAE	MSE	RMSE	RMSE CV
0	Ridge	0.339595	0.392320	0.374694	0.235820	0.485613	0.472811
1	Lasso	0.339642	0.392308	0.374724	0.235803	0.485595	0.472660
2	DTR	0.442185	0.535044	0.338806	0.199187	0.446303	0.431209
3	RFR	0.527679	0.758563	0.311523	0.168658	0.410680	0.397709

COMPARISM OF PREDICTION ACCURACY

The evaluation measures used were r-squared for accuracy and mean squared error for loss.



PREDICTION ERRORS



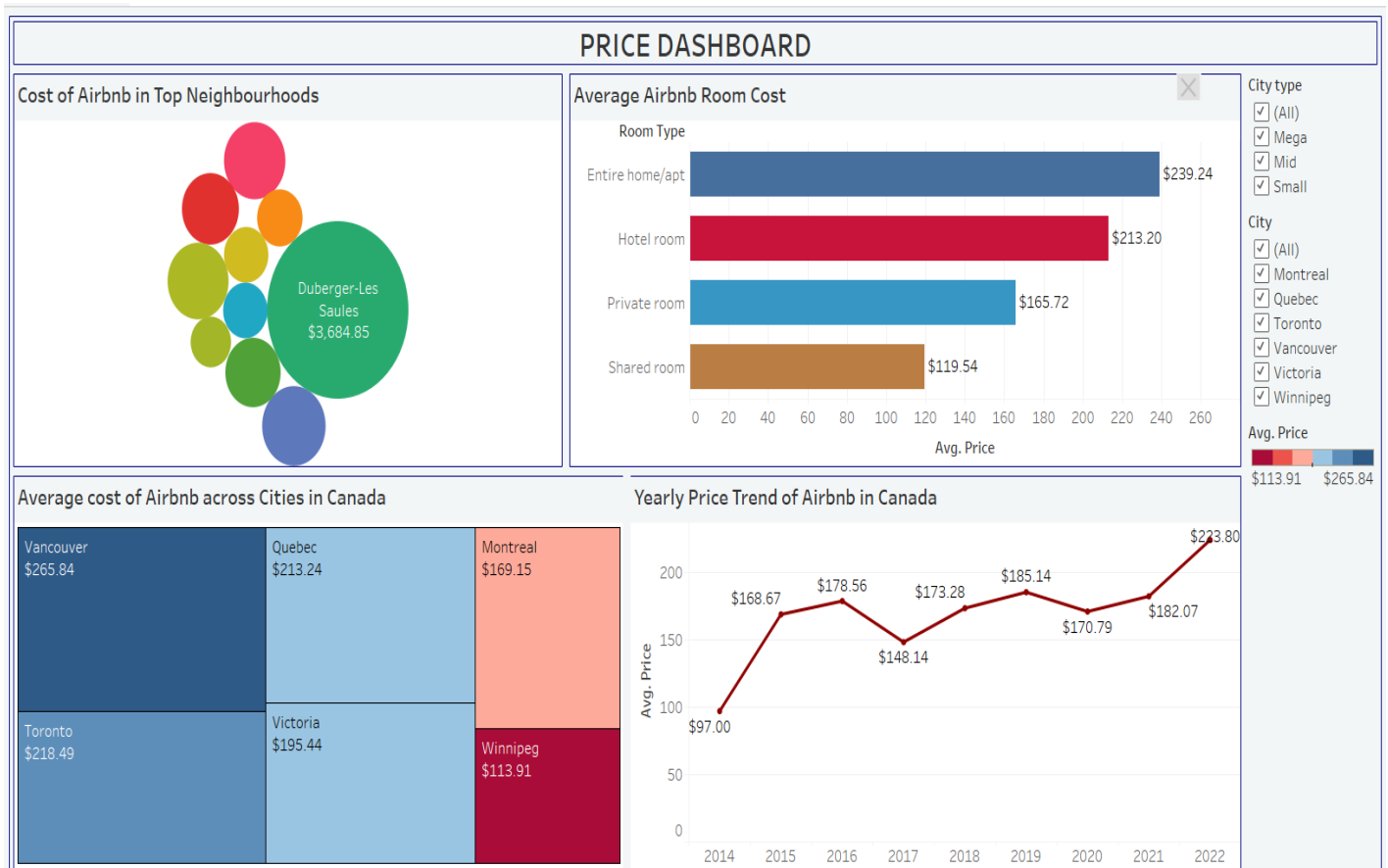
PREDICTION OBSERVATION

The best model to choose would be the Random Forest Model. It performs better than the other three models: Ridge, Lasso, and Decision Tree models. Although it only predicts about 76% of the variation in price. It has the minimum error in predicting the price.

TABLEAU ANALYSIS

From our analysis, it is evident that both city type and review have effect on the price of Airbnb. We created two dashboards based on our hypotheses; one dashboard completely visualizes the effect of price on Airbnb while the second dashboard illustrates customer satisfaction based on reviews received in different Airbnb locations in Canada.

PRICE DASHBOARD

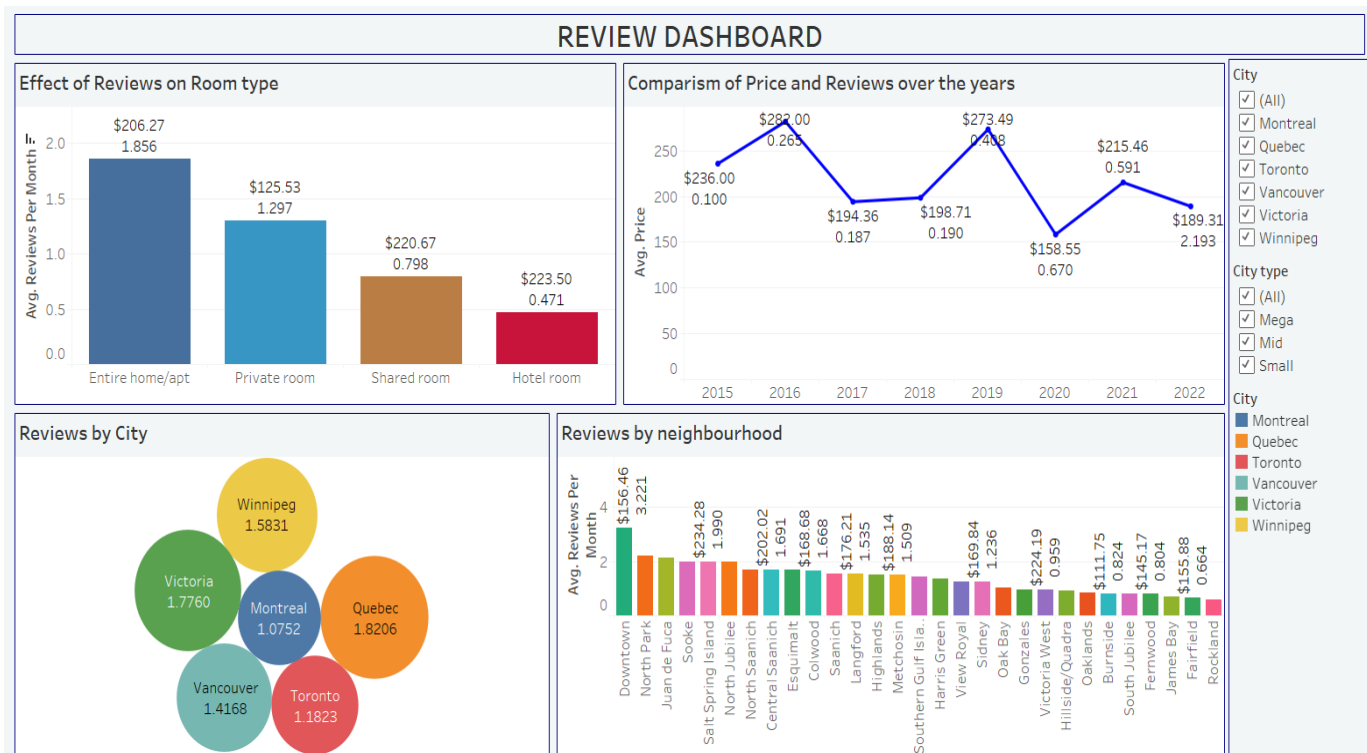


The price dashboard describes the average prices of Airbnb in Canada. The dashboard is made in an interactive way. This way, we can gain full insights into the effect of prices on accommodation in top neighbourhoods, room types, cities, and city types. The dashboard also illustrates how prices change over period of years.

EFFECT OF PRICE ON AIRBNB IN CANADA

- Mid cities of Vancouver and Winnipeg has the highest and lowest average prices respectively.
- The top and most expensive neighborhood is found in a Small city of Quebec City.
- Private room is most expensive generally, while entire home is cheapest in Vancouver.
- Prices of Airbnb is on the increase in Toronto, Quebec City, and Montreal
- Prices in Vancouver, Winnipeg, and Victoria are dropping.
- Most Expensive private room is in the city of Quebec.
- Least price of private room is found in Winnipeg.
- Entire home in Winnipeg is more affordable than any other city.

REVIEWS DASHBOARD:



The reviews dashboards as shown above, which is also an interactive dashboard is designed to visualize the effect of reviews on room type, city, and different neighbourhoods while also comparing price and reviews over the years.

EFFECT OF REVIEWS ON AIRBNB IN CANADA

- The higher the reviews, the more the cost of Airbnb in both mega and small cities
- Quebec City has the highest customer satisfaction based on the review while Montreal is the lowest.
- Surprisingly, the cheapest room has the highest review. This could mean customers prefer lower cost rather than comfortability.

CONCLUSION AND RECOMMENDATION

- Travellers should consider locations in Montreal, city of Quebec City and Toronto where there is increase in reviews which may be due to higher customer satisfaction.
- Travellers should also consider holidays in Vancouver, Winnipeg, and Victoria where Airbnb prices are currently on the downward trend.
- For group of travellers, friends, or families looking for affordable entire home, Winnipeg is the best bet as it is cheaper than any other city.

FUTURE PROJECT

- An improved model might result from additional investigation of the seasonality tendencies. There are undoubtedly encouraging seasonal and weekly trends that affect how much an Airbnb listing should cost.

REFERENCES

Dataset: <http://insideairbnb.com/get-the-data.html>

Airbnb Website: <https://www.airbnb.ca/>

Airbnb Disclaimer: <http://insideairbnb.com/about.html>

<https://en.wikipedia.org/wiki/Airbnb>