Airbnb Analysis

Business Problem

In this project we're going to study Airbnb business in New York City of different hotels. We'll explore the data and fetch insights and recommend business decisions based on those insights. We'll computer this for different hotels and compare if one is doing good what's the reason behind it or if other is doing bad what's the reason behind that as well. We'll try to figure out the patterns between different features and how each feature is impacting the target variable (price). From this analysis clients are going to benefit from. If we provide them all the information about areas of hotels, type of services they provide and prices of multiple hotels in that area then client will be able to easily and confidently select the required hotel based on its services and client can get better environment according to his or her needs.

Inspiration:

- What can we learn about different hosts and areas?
- What can we learn from predictions? (ex: locations, prices, reviews, etc)
- Which hosts are the busiest and why?
- Is there any noticeable difference in traffic among different areas and what could be the reason for it?
- What are the features impacting price?
- Is price of the hotels have anything to do with areas of hotel?
- Which hotels have more listings and why?
- Which areas have more hotels and why?
- How do the price and listing vary in hotels?

Dataset

Source: https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data

Original Source: http://insideairbnb.com/new-york-city

The data is related to Airbnb. It's open-source public data available on Kaggle. But the original data is coming from the above-mentioned site. This dataset includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

This dataset is limited to New York city only. it consists of 16 features; we'll use many of them is selected for this project.

- ID of the listing
- name of the listening.
- ID of the host
- host name.
- Location
- Area
- Latitude
- Longitude
- Room types
- Price.
- Review

Models

We are going to perform predictive modeling on our dataset. The target variable will be the price. We'll predict the price based on different machine learning Regression Algorithms. Like

- Lasso Regression
- Ridge Regression,
- Decision Tree Regression
- Random Forest Regression

And then at the end we'll compare the accuracy of the models. We'll select the best performing model based on accuracy and other metrics.

Evaluation Plan

The project will involve three major phases.

1. Preprocess Data

Data Cleaning will involve following steps

- Deleting redundant columns.
- Dropping duplicates.
- Cleaning individual columns.
- Remove the NaN values from the dataset
- Some Transformations

2. Exploratory Analysis

The exploratory analysis will involve visualization of the exploration. It'll be in tabular form and also graphical form.

- Univariate analysis
- Bivariate analysis
- Multivariate analysis
- Comparison analysis (competitor analysis)
- Correlation analysis

3. Predictive Analysis

Predictive Analysis will involve different machine learning techniques to predict the target variable. Different machine learning algorithms will be used to learn the pattern from the data and make predictions based on those patterns.

- Linear Regression
- Decision Tree Regression
- Random Forest Regression
- Deep neural Network

Similar / Past Work

There is a work on Exploratory Analysis before, but very few individuals have worked on predictive analysis. And the accuracy is bad. It is below 25%, and the error rate is high. We'll try to improve accuracy and take the exploratory analysis to the next level.