Capstone Project

On

Airbnb Booking Analysis

Prepared by: Ankitkumar S Dandiwala

Project Summary

This exploratory data analysis (EDA) project focused on understanding the factors that influence Airbnb booking patterns and trends, The objective was to provide valuable insights to hosts and travellers, enabling them to enhance their experience on the platform. The dataset used in this analysis contained information about Airbnb bookings, including geographical coordinates, room type, availability, prices, neighbourhood details, and review metrics.

The project began with a thorough understanding of the dataset, which comprised 48,884 rows and 16 columns. Manipulations were performed to remove irrelevant columns and handle missing values, resulting in a clean dataset ready for analysis.

Thirteen different visualizations were employed to extract insights and facilitate discussion. These visualizations covered various aspects of the data, ranging from average and median prices based on location and room type to the distribution of numeric features and the correlation between different variables. Each visualization provided specific insight that could potentially impact business strategies positively or negatively.

Some key findings from the visualizations include: 1) The mean price of entire home/apartment room types was higher across all locations, with Manhattan having the highest prices overall. This insight could guide pricing strategies and promotional schemes to attract more customers. 2)Outliers were identified in variables like minimum nights and calculated host listing count, suggesting the need for outlier management to improve booking accuracy and customer experience. 3) The scatter plots of price vs number of reviews indicated an inverse relationship, implying that lower-priced accommodations tend to attract more reviews. Adjusting pricing strategies could potentially drive positive customer reviews and satisfaction. 4) The correlation matrix revealed relationships between variables such as calculated host listing count, number of reviews, and minimum nights. These insights could guide optimization of hosting strategies and resource allocation. 5) The pie charts illustrated the distribution of room types and services across different neighbourhood groups. Understanding these distributions could inform market targeting resource allocation, and expansion opportunities.

Based on these insights, several recommendations can be made to improve business outcomes: 1) Implement attractive promotional schemes and discounts for entire home/apartment room types to mitigate the potential negative impact of higher prices. 2) Address outliers and anomalies in the dataset to enhance booking accuracy and customer experience. 3) Optimize marketing efforts, pricing strategies, and resource allocation based on the preferences and demand for specific room types in each neighbourhood group. 4) Monitor and respond to customer

feedback to enhance customer satisfaction and drive positive growth. 5) Identify potential expansion opportunities in neighbouhrhood groups with lower service offerings to capture a large market share.

In conclusion, this EDA project has successfully delved into the vast trove of Airbnb booking data, unearthing invaluable insights that prove beneficial for both hosts and travellers alike. Leveraging these insightful findings and implementing the recommended strategies enables business to finely tune their pricing models, enhance marketing efforts, and optimize resource allocation, thereby fostering positive growth and an elevated level of customer satisfaction. It is of utmost importance to maintain a diligent and regular monitoring of data trends and actively seek customer feedback in order to remain competitive within the ever-evolving landscape of the dynamics Airbnb market.

GitHub Link

https://github.com/ankitdandiwala/Airbnb-Booking-Analysis

Problem Statement

This exploratory data analysis project aims to uncover valuable insights from the Airbnb booking dataset, enabling informed decisions for the client.

Business Objective

To identify the key factors that influence Airbnb booking patterns and trends in order to provide insights that can be used by the client grow their business.

About dataset

Dataset represents information about the Airbnb booking that includes geographical coordinates of the property, types of room, availability, price, details of neighbourhood alongwith number of reviews received till the last date of review, and rate of review. The shape of the dataset is 48895 rows and 16 columns. Some features of the dataset include null values, however, dataset does not possess duplicate values. Dataset contains categorical data too.

What all manipulations have you done and insights you found?

Here are the key data cleaning steps that have been performed:

No Duplicate Rows: It is verified that no duplicate rows exist in the dataset, ensuring that each entry is unique.

Excluded Columns: To streamline the analysis, certain columns that are not essential for the current objectives. These columns are 'name,' 'host_name,' and 'last_review.'

Excluded Rows with Zero Price: This is important to eliminate erroneous data that might affect the analysis. Therefore, rows having zero price for room services have been excluded from the analysis.

Replaced Null Values: Column 'reviews_per_month' contains approximately 20.56% of the values were null, They were replaced with zero to maintain consistency in the analysis.

The data cleaning resulted in a final dataset with shape of 48884 rows and 13 columns, ensuring reliable analysis.

Conclusion

Based on the analysis of the Airbnb booking data, it is evident that there are several factors that influence booking patterns and trends.

- Manhattan and Brooklyn have a noticeably higher number of services compared to other neighbourhood groups. This insight allows us to pinpoint potential expansion opportunities in neighbourhood groups that currently have fewer services.
- Price and number of reviews, and price and reviews per month are inversely proportional to each other. By monitoring and responding to customer feedback, customer satisfaction can be enhanced and client can drive positive growth for the business.
- The box plot analysis revealed outliers in the variables minimum nights. It seems unrealistic that the minimum night for the booking is 1250. It is important to eliminate the anomalies and outliers.
- The correlation matrix reveals the positive correlation between number of reviews and reviews per month. Any of the variable can be removed before applying the model in order to reduce the computing power and complexity.