Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Contribution Role : 1. Praful M. Gedam (Prafulgedam12@gmail.com) :- 1) Data Cleaning and Data Preprocessing 2) General and Business Analysis • All neighbourhood groups • Availability of number of days for room when listing is available for booking • Room type • Neighbourhood group based on latitude and longitude • Price exploration • Relation between neighbourhood group and availability of number of days for Room • Property owned by each neighbourhood group • Most successful host • Most expensive neighbourhood in each group
Please paste the GitHub Repo link.
Github Link:- PrafulGedam/airbnb-booking-analysis (github.com)
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Problem Statement:

This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values. Explore and analyze the data to discover key understandings (not limited to these) such as

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 of traffic among different areas and what could be the reason for it?

Approaches:

First thing first, I started by getting some basic information about the data types, null values as well as some descriptive statistics of the features. This allowed us to decide which columns and data to keep while removing the null values. The statistical analysis also allowed us analyze and keep the required data.

After cleaning the data, we simply studied the distribution of bookings with respect to the columns or features. This allowed us to get some important information on different neighborhoods and their property listings and their prices etc.

Some of the distributions were compared based on the existing knowledge of the relationship between the columns. This multivariate analysis helped us make some interesting observations, namely about the price variation, price variation between different neighborhoods, and most expensive neighborhoods and correlation between price and availability etc.

Bar plot showed the mean price of different room types and inferred that Entire home/apt is the expensive one as it has more amenities, followed by a private room and the shared rooms lesser than the former rooms. Pie plot showed the room type which was the most booked/wished for and the percentage was higher for an entire room/apartment followed by a private room and shared rooms.

Conclusion:

80% of the hosts are in Manhattan and Brooklyn, Sonder (NYC) and blue ground has the greatest number of host listings. most of the people likes Queens and Manhattan

Even with moderate number of hosts in Queens, it generated the highest revenue out of all the neighborhood group

Drive Link :-
https://drive.google.com/drive/folders/1sBQfdDXn4PovOokUeQyS1hVRCeULxC1x