Summary

Name: Raja Chowdhury
Email: rajachowdhury2468@gmail.com
Github Link:- https://github.com/RajaChowdhury/Airbnb-Bookings-Analysis-EDA-Capstone-Project.git

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?

Our Approach:

Our first step was to understand the dataset of Airbnb NYC (New York City) bookings and then we started exploring and analyzing each column. And we found out that we need to clean our dataset before performing the exploratory data analysis in order to get more accurate outcomes. Hence took the following steps:

- We treated null values of different columns accordingly.
- We removed rows where the price was equal to zero.
- We handled the outliers from columns.
- We looked for data dependencies in our dataset.

After cleaning our dataset, we started with our data visualization on the basis of categorical columns and problem statements.

Key Findings:

1. Manhattan and Brooklyn neighborhood groups hold around 85% of the total booking while Staten Island has the least number of bookings.
2. Manhattan neighborhood hosts have a higher number of properties listed in Airbnb.
3. Entire home/apt is the most preferred room type followed by private room and they hold around 97% of the total booking.
4. Queens has significantly less host listings than Manhattan. So, we should take enough steps to encourage host listings in Queens.
5. Seven out of ten top neighborhood belongs from Manhattan neighborhood group with Entire home/apartment room type.

Conclusion:

- ❖ We can conclude from the analysis that Manhattan is the top neighborhood group with the highest number of bookings and host listings. Seven out of ten top hosts are from Manhattan. One of the probable reasons for being the most preferred Neighborhood Group is that Manhattan is world-famous for its museums, stores, parks, music and cultures. It has a high number of tourist places where Entire Home/Apartment are more preferred as stay options and also for longer periods. These factors increased the demand and led to higher prices compared to any other neighborhood groups.
- ❖ Brooklyn has a significant number of bookings because Brooklyn also has some famous bridges, parks, museums, islands and other tourist places but with more affordable prices as compared to Manhattan. It also received the maximum number of reviews followed by Manhattan.
- ❖ Rest 3 neighborhood groups namely Queens, Bronx and Staten Island are observing very less number of booking and hosts, especially on Staten Island. Considering that these are less popular areas, it is possible that some guests choose these locations to save up money or for official purposes who want to stay for limited time periods. And these neighborhood groups have higher room availability than Brooklyn and Manhattan.