Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Sanjay Yadav neer.ping@gmail.com Complete EDA of given dataset Preparing technical documentation Preparing presentation Preparing video presentation
Please paste the GitHub Repo link.
GitHub Link:- https://github.com/sanjay2097/airbnb-analysis
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one-of-a-kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data - data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.
This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

PROBLEM STATEMENT -
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?**

APPROACH-

- **Adding the data into our dataframe.**
- **Getting information about rows and column and datatypes that we have to work with.**
- **Cleaning data by treating Null values, Outliers duplicates, dropping obsolete columns unnecessary for our analysis.**
- **Making a new dataframe with all the columns we are going to analyze.**
- **Analyzing the data by using Pandas and NumPy libraries.**
- **Plotting graphs using matplotlib and seaborn to establish relationship between various attributes.**

Conclusions-

- **Noticed that Manhattan has nearly 4 times the listings when compared to Queens.**
- **Noticed that the price distribution is right skewed with very few listings more costly than $1000 per night.**
- **The median price of listings in Manhattan is around $150 which is double the median price in Queens.**
- **There are a lot of listings available for a minimum stay of 1, 2 as well as 30 days.**
- **Also, one can notice that, majority of the listings are of Entire home and Private room type and minimum for Shared room.**
- **Most of the highly reviewed listings are present in Staten Island ,Bronx and Queens.**