Capstone Project Submission

Team Member's Name, Email and Contribution:

1) Priya Debrani
E-mail:priyadebrani3005@gmail.com

- Data visualization.
- Approach towards plan.
- Bar plot and Heat map.
- Pi-plot and pair plot
- Technical documentation.
- Project summery template.
- Frame work of project.
- Debug all Errors
- Data sorting.
- Pie plot and Histogram plot
- PPT presentation
- Data analysis.

Problem definition:

❑ The dataset has around 49000 observations and 16 columns in it which is a mix of categorical and numerical variables. The main objective here is to explore and analyse the data to find out insights such as:

- What can we learn about different hosts and areas?
- How are rentals distributed among the five boroughs?
- What's the price distribution and what's the range of fair prices available?
- Which hosts are the busiest?
- Which are the top 10 hosts on the basis of reviews?
- Which are the top 10 hosts on the basis of count of listings?

❑ EDA DATASET ANALYSIS:-

The different features the dataset contains:

- **id**: It is an unique id given to the property listed in airbnb NYC which is a numerical variable.

- **name**: It represents the name of the airbnb listed property which is a categorical variable.

- **host_id**: This is an unique id given to the host of the property which is a numerical variable.

- **host_name**: The name of the host of the property listed which is a categorical variable.

- **neighbourhood_group**: This represents a big neighborhood inside which there are many mini neighborhoods which is a categorical variable. There are 5 neighborhood groups in the data:

  i)   Manhattan
  ii) Brooklyn
  iii) Staten Island
  iv) Queens
  v)   Bronx

Conclusion:

❑ We have reached the end of our analysis of Airbnb listings in NYC. We started from looking out for duplicate values, then missing value treatment and finally used EDA to discover many insights from the dataset. To summarize few of the important insights we gathered:

   ● Host Maya is the busiest host in NYC and there are multiple reasons in favor of it like price, minimum nights, availability and number of reviews. She has a total of 5 properties listed in the same neighbourhood.
   ● Manhattan and Brooklyn are the most expensive neighborhoods and they receive the most traffic as well. Due to many tourist attractions and the number of properties available, people tend to visit these two areas comparatively more than other ones.
   ● Entire Home/Apt is the costliest room type available but still the most preferred ones for the customers. Entire Home/Apt and Private Rooms receive way more traffic than Shared Rooms and as a result Shared Rooms stay available for most of the time out of the 365 days.

❑ These insights generated can definitely help everyone make better decisions in future to enhance their experience of staying in an Airbnb in NYC.

Github Link:- PriyaDebrani/Airbnb-NYC-2019-Analysis-: EDA_PROJECT (github.com)

Summary of Capstone project

❑ 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 analysed and used for security, business decisions, understanding of customers' and hosts' behaviour and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

❑ The main objective of this exploratory data analysis project is to understand customer and Hosts behaviour and Performance