**Summary**

**Capstone Project (1) – Airbnb dataset Exploratory Data Analysis**

Airbnb, is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app. Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking.

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 providers' (hosts) behaviour and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

Firstly, we have imported the raw dataset and started to get it in a desired state to draw any information out of it. We have selected the features that are required, added additional features, replaced or eliminated the null values with a sense so as there is not a significant loss of information. Then we used ‘describe’ method to get the idea of the numerical columns present in the dataset and their value range. We have also identified categorical columns and the total categories that are present in each.

Then we started analysing our dataset firstly by plotting a correlation heatmap to get the overview of which feature is likely affecting the other feature(s). After that we started with univariate analysis like finding the top ten host on the basis of their listing count, finding the borough that has maximum listings present in it, the room types that are mostly offered by the listings.

We have filtered our dataset on the basis of price<=500, to eliminate the outliers that are present to get the unbiased conclusion wrt price. Then we moved forward by doing bivariate analysis using features like ‘neighbourhood\_groups’, ‘price’, ‘minimum\_nights’, ‘number\_of\_reviews’, ‘reviews\_per\_month’ and ‘room\_type’. We have also determined the top five busiest host according to the maximum average occupancy and determined the area which has maximum traffic.

We are able to draw out various inferences from our analysis and able to figure out, guests mostly prefer popular places with a lot of tourist attractions and corporate sectors, for short stay private rooms are preferred. Shared rooms are least in number and are least preferred by the guests. Most of the people like a booking that has lesser minimum nights.

Contributor Lists:

1. Bishakha Kumari

Feature selection

Null values handling

Outliers detection

Analysis of data through different techniques

Visualizing the data

1. Aashish

Feature selection

Null values handling

Outliers detection

Analysis of data through different techniques

Visualizing the data

1. Mrityunjay Dubey

Feature selection

Null values handling

Outliers detection

Analysis of data through different techniques

Visualizing the data

Github Link: <https://github.com/bishakha1995/Eda-Airbnb>

Drive Link: <https://drive.google.com/drive/folders/1UJ668NLPKpItCEJa2-VB6yXEhJ-BQ6_Z?usp=sharing>