K.J. SOMAIYA COLLEGE OF ENGINEERING
(AUTONOMOUS COLLEGE AFFILIATED TO UNIVERSITY OF MUMBAI)

Mini Project Report for

Python for Data Science Audit Course

TOPIC-

Airbnb listings and metrics in NYC, NY, U.S.A (2019)

(WE HAVE CHANGED OUR DATASET FROM THE ONE WRITTEN IN EXCEL SHEET TO THE LINK GIVEN HERE. NEW LINK FOR OUR DATASET-
https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data)

MADE BY-

ISHIKA BHATT	**Roll No: 1914073**	**SY-IT**
PRANAV KALAMBE	**Roll No: 1913023**	**SY-EXTC**

INTRODUCTION :

Airbnb is an online marketplace for arranging or offering lodging, primarily homestays, or tourism experiences.It is a paid community platform for renting and booking private accommodation founded in 2008 and since then, guests,hosts and travellers have used it to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. It allows individuals to rent all or part of their own home as extra accommodation and offers a search and booking platform between the person offering their accommodation and the vacationer who wishes to rent it.Today,it is being used and recognized by the whole world.Data analysis on millions of listings provided through Airbnb is a crucial factor for the company hence we have decided to perform the analysis and visualizations of the Airbnb data collected for New York City, U.S.A for the year 2019.

ABOUT THE DATASET :

The dataset contains information to find out more about hosts, neighbourhood of the available rooms, total number of rooms and their types, geographical availability and other necessary metrics to make predictions and draw conclusions about the listings in New York City.This dataset has 48,895 observations in it with 16 columns and it is a mix between categorical and numeric values,hence providing us with diverse data type values,which we can clean and alter.

ABSTRACT ANALYSIS :

- We downloaded a csv file of our dataset, named it as "AB_NYC_2019.csv" and loaded it in our notebook.
- The size of the dataset,that is the initial number of rows and columns ,the datatype of each value along with the number of missing values are found.
- After the dataset is cleaned,it's assigned to a new variable.
- Then some basic analysis of each column is performed and their individual features are examined.
- We started with the interpretation of the number of rooms and their respective neighbourhood groups and also classifying the particular type of room like private,shared or entire home/apt,etc.All this is also depicted using visualization through various bar and pie charts.
- Next the hotel rooms were sorted on the basis of maximum and minimum price and the highest priced rooms in a specific neighbourhood were shown.
  The most populated neighbourhood area can also be observed through a bar graph

- Now the maximum and an average of the nights spent by users is calculated and finally the most popular listings from the dataset are categorised based on their ten best reviews.
- With the help of latitude and longitude of rooms different places are marked on the New York city map for better visualization.

EXCEPTIONAL WORK/FUNCTIONALITY :

- Maximum features of the dataset and their distribution are thoroughly picturised and analyzed with 9 different Histograms.
- We have plotted the latitudes and longitudes of the rooms at the 5 main neighbourhood groups using a grid figure and loading the map of New York as an image file.
- Similarly, the different room types in New York City and location of all these rooms,both are portrayed using two separate Scatter graphs.
- The relation between neighbouring groups and availability of days is also visualized with the help of a Boxplot graph.

CONCLUSION :

This Airbnb ('AB_NYC_2019') dataset for the 2019 year appeared to be a very rich dataset with a variety of columns that allowed us to do deep data exploration on each significant column presented. Overall, we discovered a very good number of interesting relationships between its features and explained each step of the process. This data analytics is very much mimicked on a higher level on Airbnb Data team for better business decisions, control over the platform, marketing initiatives, implementation of new features and much more.Therefore this project helped us explore Data Science and get familiar with its basics and visualization using various graphs efficiently.