**Airbnb Booking Analysis**

1. **Ms.Madhuri Bonela**
2. **Mr.Pratik Ghodke**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

Hotel enterprise is a very unstable enterprise and the bookings rely on range of elements such as kind of hotels, seasonality, days of week and many more. This makes inspecting the patterns accessible in the previous records greater vital to assist the motels layout better. Using the historic data, motels can function a variety of campaigns to enhance the business. We can use the patterns to predict the future bookings the use of time collection or selection trees.We will be tackling this problem statement in three stages:

we can analyse some key metrics for hotel bookings like :

1.The number of cancellations

2.Number of bookings on weekday vs weekends

3.Type of hotels booked by customers like City hotels or resort hotels

4.Type of platform most of the customers used for booking a hotel

5. Checking if the booking was from repeated guest or not

6. Type of booking acquired by most of the customers

7. Checking the arrival count for every month and analysing the patterns

Therefore by analysing every data and applying them in future analysis makes profitable business and satisfies every customer from the service provided .

**1. Problem Statement**

These millions of listings generate a lot of dataProblem Statement

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!

1.The type of hotel preferred by most customers.

2.The number of cancelled bookings.

3.The average daily rate (ADR) and it's calculated by dividing the sum of all lodging transactions by the total number of staying nights.

4. Checking the arrival date by months.

5.Market segment.

6.year of arrival date.

7. Customer Type .

8.Distribution Channel.

9.The total number of repeated guests.

10.which are the months of highest and least occupation.

11.which is the most reserved room type .

We are provided with a hotel bookings dataset.

Our main objective is perform EDA on the given dataset and draw useful conclusions about general trends in hotel bookings and how factors governing hotel bookings interact with each other.

**2. Introduction**

Hotel booking analysis is an tourist agent which offers the facilities for reserving hotels, Events, Travels, houses, vacation apartments and different accommodations for customers.

They operate their enterprise in KSA . Any clients want to e bookhotel, or rental they want to go to their workplace which for checking the availability as properly as negotiation.

Recently they have decided to exchange their reserving sample from guide to web primarily based system. It will assist them to manipulate the clients.

**3. What is Exploratory Data Analysis (EDA)?**

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

It is a good practice to understand the data first and try to gather as many insights from it. EDA is all about making sense of data in hand, before getting them dirty with it.

.**4. Exploratory Data Analysis (EDA)**

The following step-by-step example shows how to perform exploratory data analysis in Excel.

Step 1: Create the Dataset.

Step 2: Summarize the Data.

Step 3: Identify Missing Values.

Step 4: Visualize the Data.

Importing Library

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

**4.1.1 import numpy as np**

NumPy can be used **to perform a wide variety of mathematical operations on arrays**. It adds powerful data structures to Python that guarantee efficient calculations with arrays and matrices and it supplies an enormous library of high-level mathematical functions that operate on these arrays and matrices.

**4.1.2 import pandas as pd**

Import = “Bring this functionality or library to my python script”

* Pandas = The library you want to import, in this case, it's pandas.
* As = The python nomenclature for creating as alias. ...
* pd = The standard short name for referencing pandas.

Pandas stands for “**Python Data Analysis Library** ”. It is used for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

**4.1.3 matplotlib.pyplot**

pyplot is **a plotting library used for 2D graphics in python programming language**. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits.

**Following steps were followed:**

1. Define the x-axis and corresponding y-axis values as lists.
2. Plot them on canvas using . plot() function.
3. Give a name to x-axis and y-axis using . xlabel() and . ylabel() functions.
4. Give a title to your plot using . title() function.
5. Finally, to view your plot, we use . show() function.

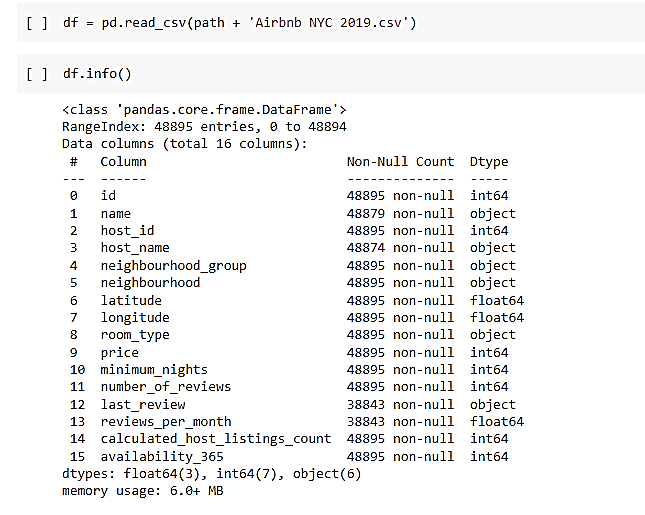
#### **4.2 Mounting Google Drive**

from google.colab import drive

drive.mount('/content/drive/')

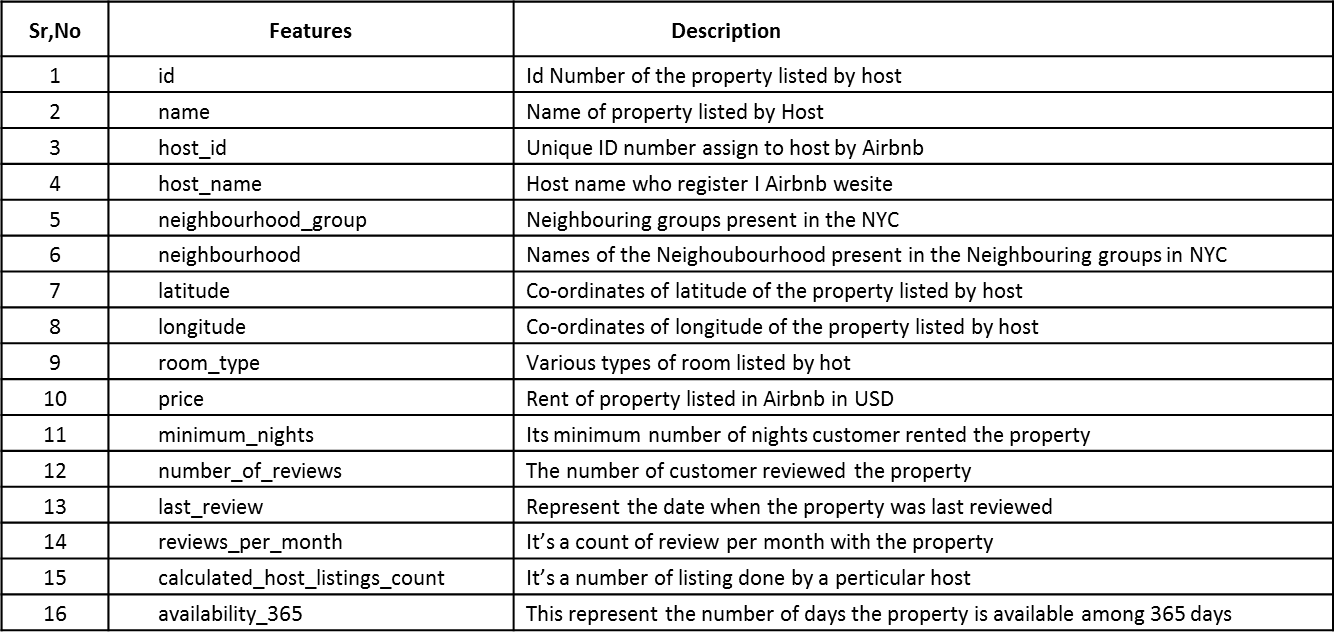
#### **4.3 Importing Dataset From Drive**

df = pd.read\_csv(path + 'Airbnb NYC 2019.csv')



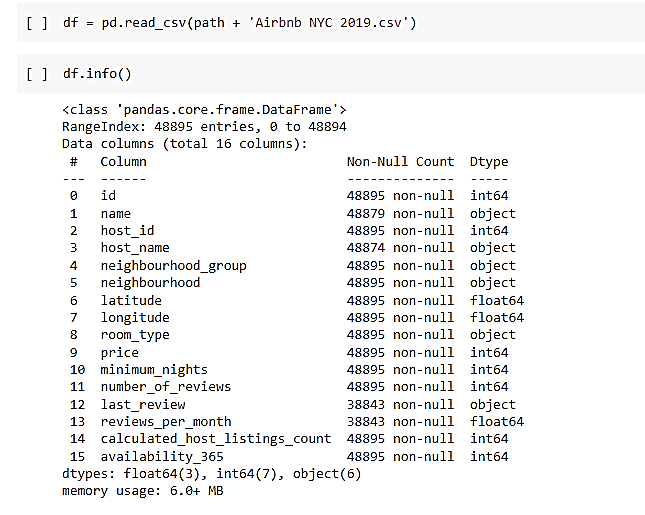
We can observed the, there is non null values in data set and

4.4 **Feature descriptions**



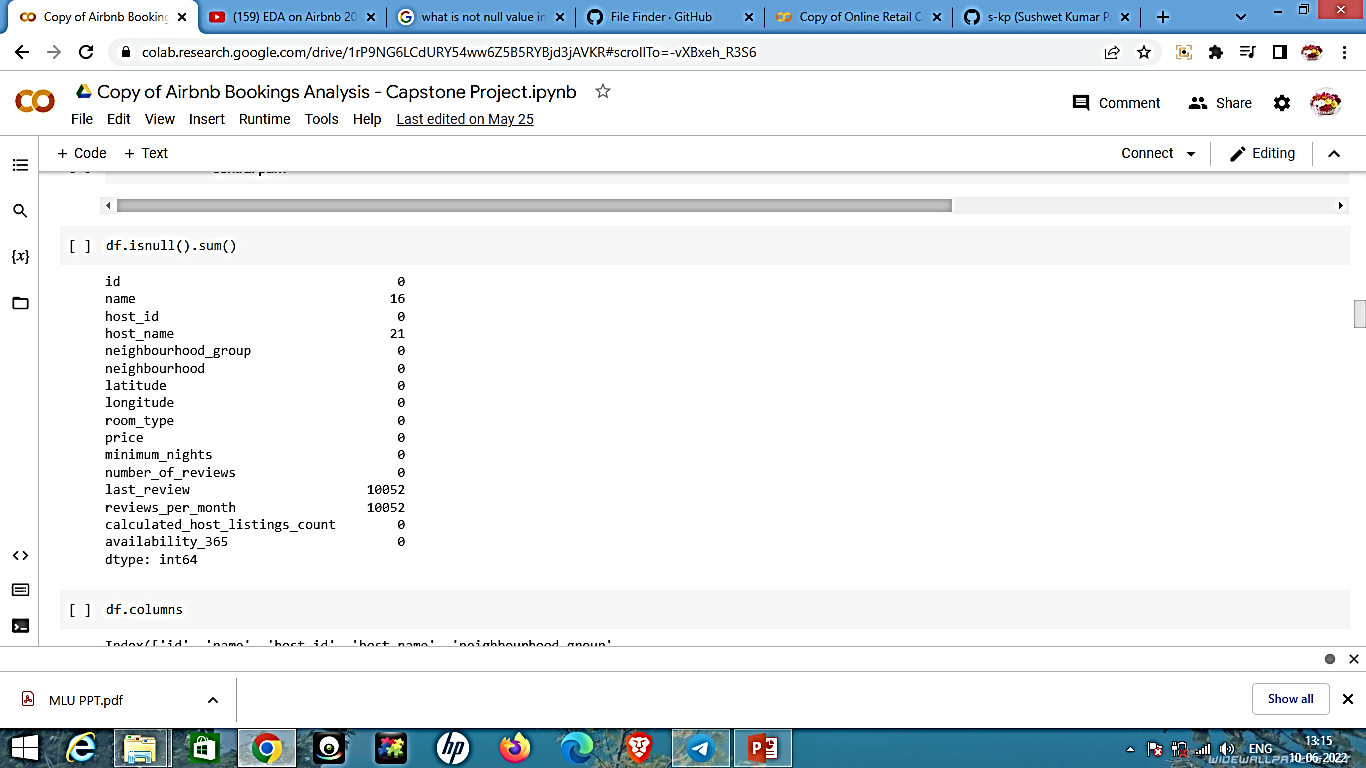
**4.5 Data Exploration**

Data exploration is **the first step in data analysis involving the use of data** visualization tools and statistical techniques to uncover data set characteristics and initial patterns.

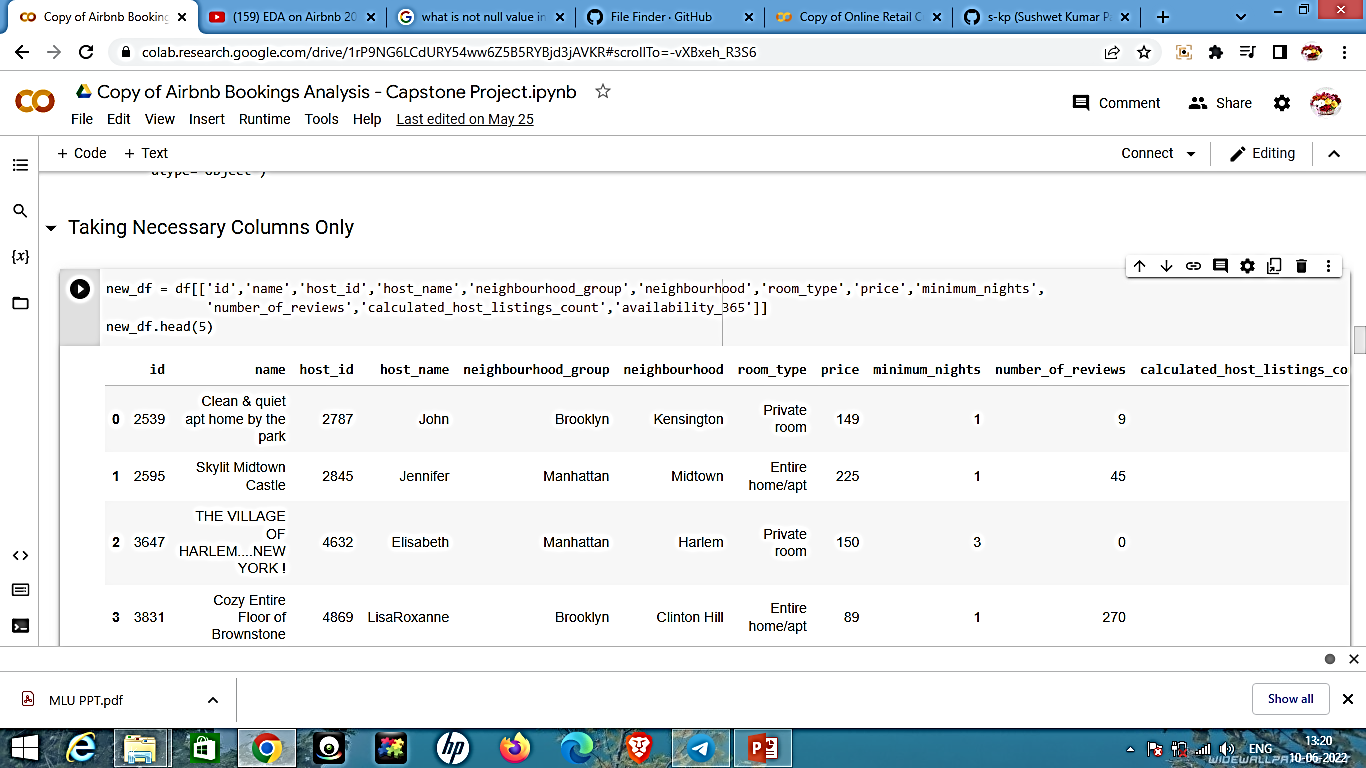


* The data set consist of 16 columns/features
* Total 48895 Entries in data set
* We can observed the, there is non null values in data set and

**4.6 Identify and remove null values**



* The table shows us a null values are present in data set
* There is 4 category is present whose null value is Gretter and its not required
* We can skip those column's and keep necessary columns only



* **Null values Treatment**

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project inorder to get a better result.

* **Encoding of categorical columns**

We can used One Hot Encoding to produce binary integers of 0 and 1 to encode our categorical features because categorical features that are in string format cannot be understood by the machine and needs to be converted to numerical format.

### **4.7 Taking Necessary Columns Only**

new\_df = df[['id','name','host\_id','host\_name','neighbourhood\_group','neighbourhood','room\_type','price','minimum\_nights','number\_of\_reviews','calculated\_host\_listings\_count','availability\_365']]

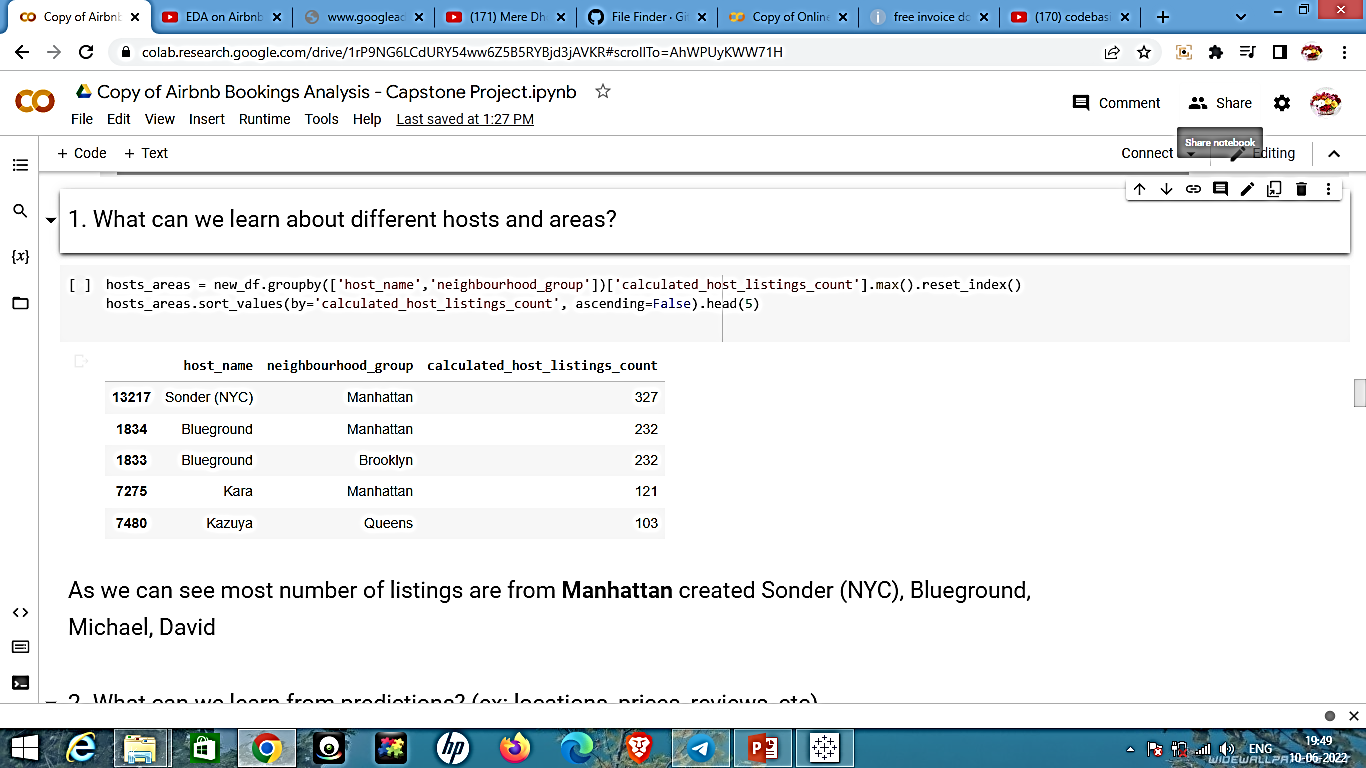
The new data frame has created by dropping unnecessary columns.

**5.0 Exploration Data and Visualizing**

Data exploration is the first step of data analysis used to explore and visualize data to uncover insights from the start or identify areas or patterns to dig into more. Using interactive dashboards and point-and-click data exploration, users can better understand the bigger picture and get to insights faster.

## **5.1 What can we learn about different hosts and areas?**

Host that have maximum number of hosting



**5.2 What can we learn from predictions?**

area = areas\_reviews['neighbourhood\_group']

review = areas\_reviews['number\_of\_reviews']

fig = plt.figure(figsize = (10, 5))

# creating the bar plot

plt.bar(area, review, color ='maroon',

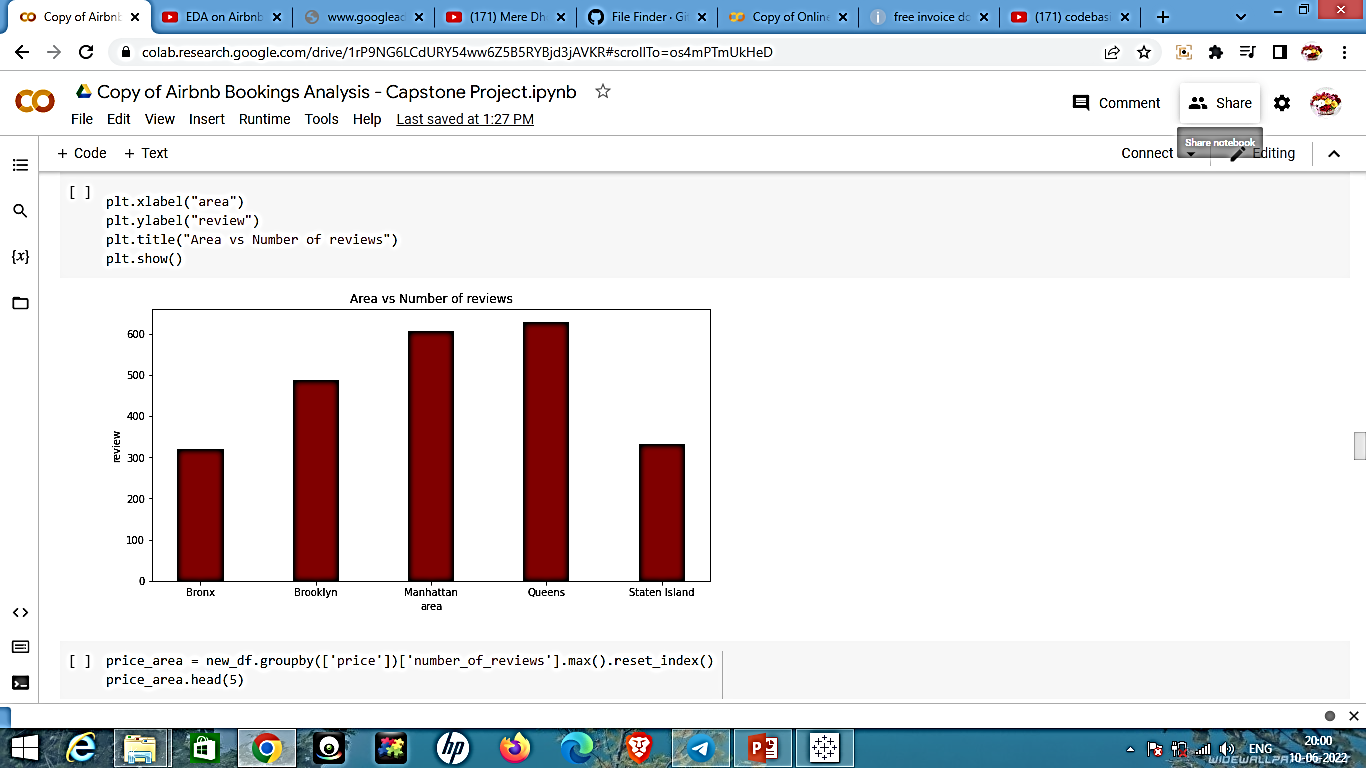
        width = 0.4)

plt.xlabel("area")

plt.ylabel("review")

plt.title("Area vs Number of reviews")

plt.show()



area = price\_area['price']

price = price\_area['number\_of\_reviews']

fig = plt.figure(figsize = (10, 5))

# creating the bar plot

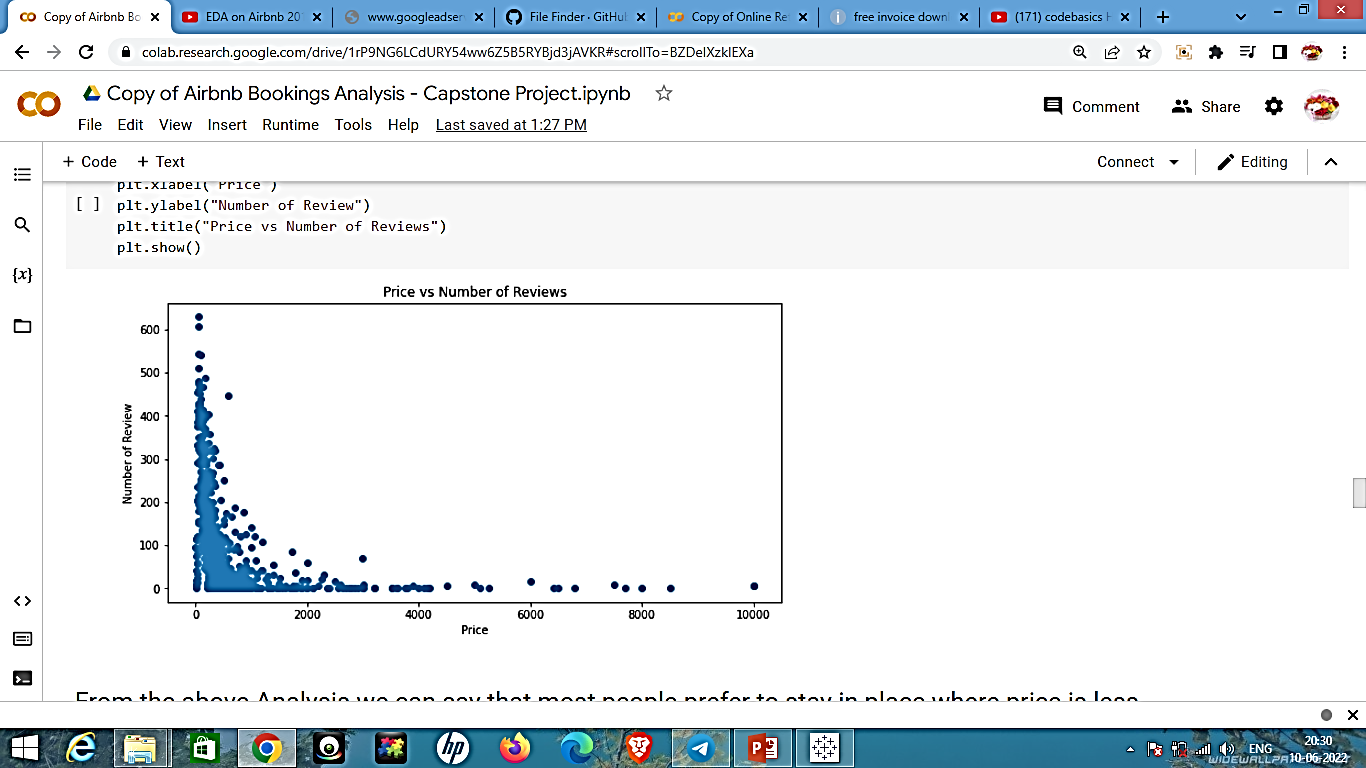
plt.scatter(area, price)

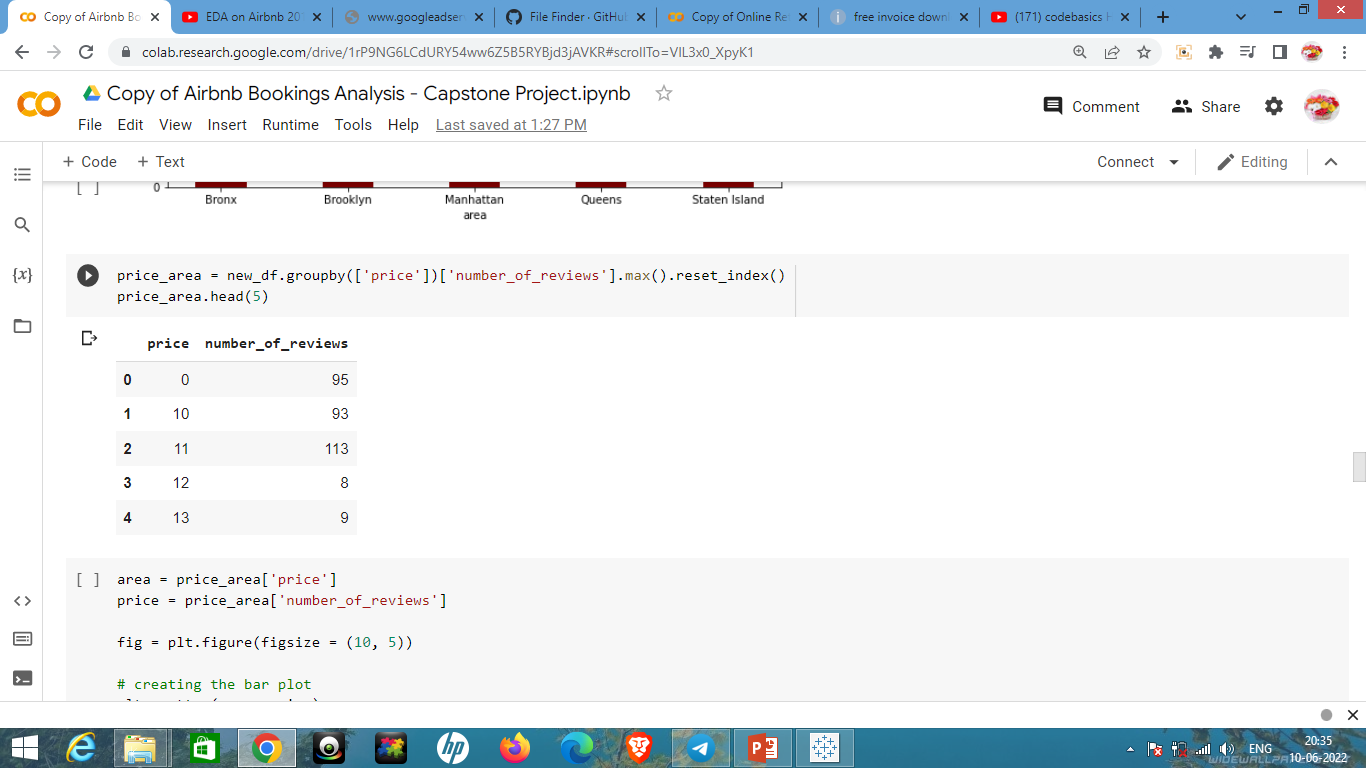
plt.xlabel("Price")

plt.ylabel("Number of Review")

plt.title("Price vs Number of Reviews")

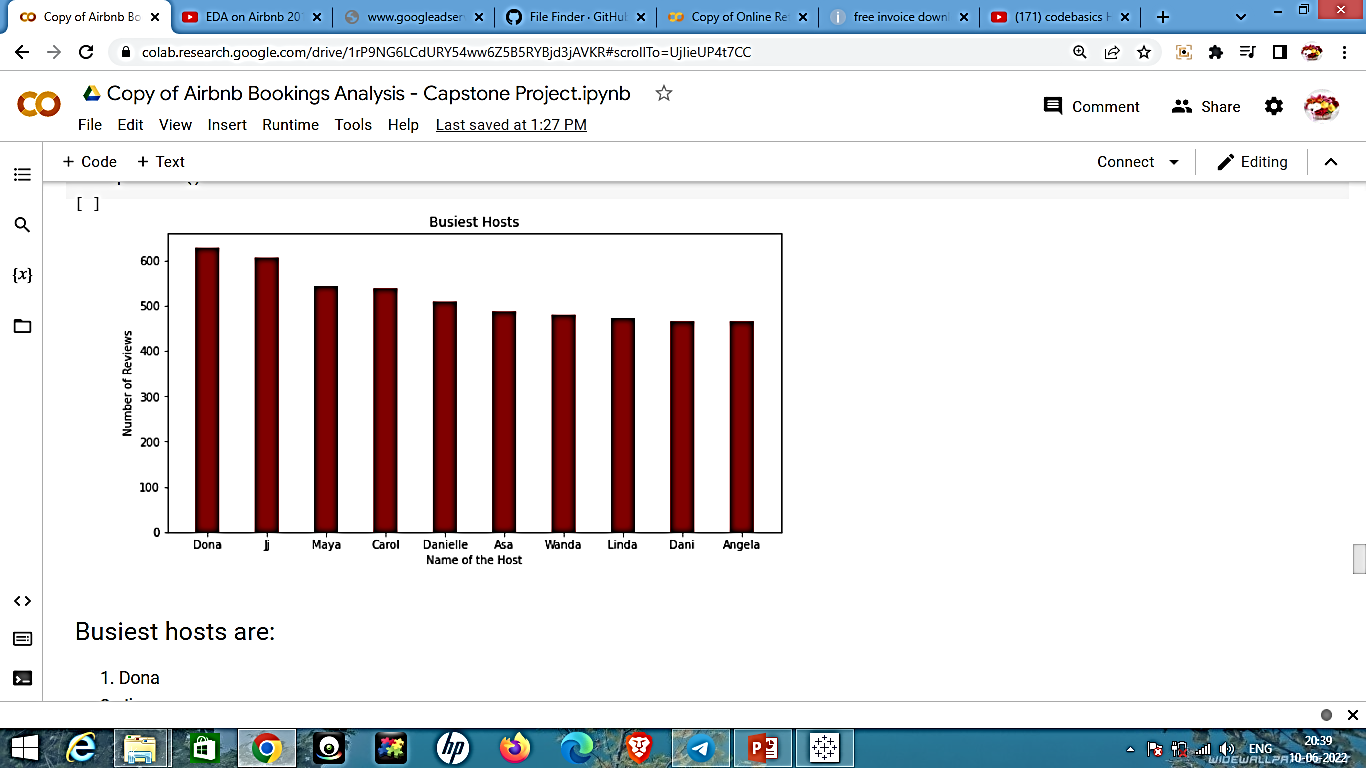
plt.show()





## **From the above Analysis we can say that most people prefer to stay in place where price is less.**

**5.3 Which hosts are the busiest and why?**



* Busiest hosts are:1.Dona 2.Ji 3. Maya 4.Carol 5. Danielle
* Because these hosts listed room type as Entire home and Private room which is preferred by most number of people.

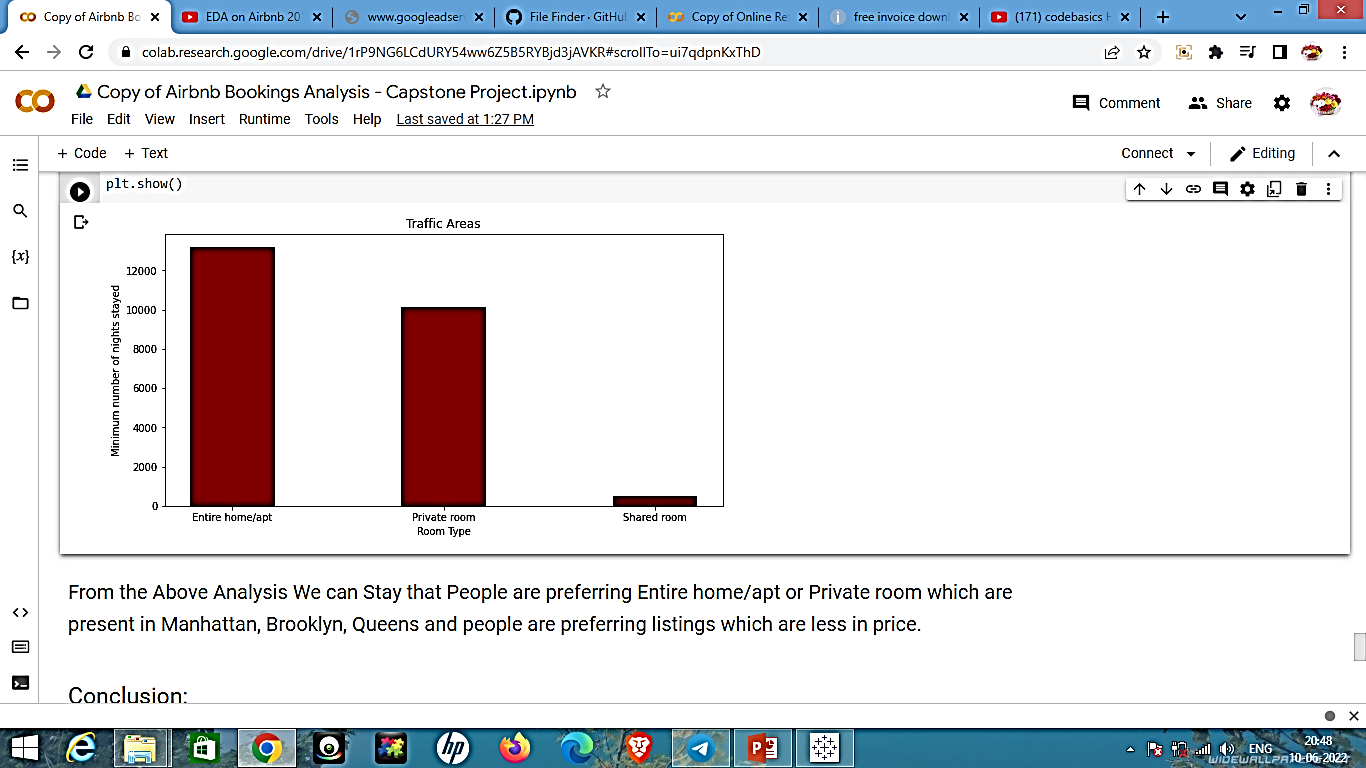
Busiest hosts are:

1. Dona
2. Ji
3. Maya
4. Carol
5. Danielle

Because these hosts listed room type as Entire home and Private room which is preferred by most number of people.

**5.4 Is there any noticeable difference of traffic?**

Among different areas and what could be the reason for it?



* From the Above Analysis We can Say that People are preferring Entire home/apt or Private room which are present in Manhattan, Brooklyn, Queens and people are preferring listings which are less in price.
* Share rooms are not preferred much by the people

**Conclusion**

The people who prefer to stay in Entire home or Apartment they are going to stay bit longer in that particular Neighbourhood only.

2. The people who prefer to stay in Private room they won't stay longer as compared to Home or Apartment.

3. Most people prefer to pay less price.

4. If there are more number of Reviews for particular Neighbourhood group that means that place is a tourist place.

5. If people are not staying more then one night means they are travellers.