**Airbnb Booking Analysis**

**Contributors: Beesa Sri Sai Meghana Goud**

**Neha Pasi**

**Alma Better Capstone Project**

**Abstract:**

Airbnb is an online marketplace connecting travelers with local hosts. On one side, the platform enables people to list their available space and earn extra income in the form of rent. On the other, Airbnb enablestravelers to book unique homestays from local hosts, saving them money and giving them a chance to interact with locals. Catering to the on-demand travel industry, Airbnb is present in over 190 countries across the world.

**Problem Statement:**

The aim is to identify the parameters that influence property listings on the platform. I will use basic exploratory data analysis (EDA) techniques to explore and visualize the dataset from Airbnb in NYC. I will find the distribution of every Airbnb listing based on its location, price, room type and other relevant factors. I examined this data from various perspectives and came up with some interesting findings. This can help Airbnb's marketing, finance, and technical teams in making strategic data-driven decisions.

**Introduction:**

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.

**Data Summary:**

Based upon the initial assessment we found that the data was pretty much clean except for missing values in some columns. Upon using the info() method, we draw out the following key insights about the data:-

1. The dataset has a shape of (48895,16) which means that it contains approximately 49000 rows and 16 columns.
2. Our dataset has 3 columns with float64 dtype, 7 columns with int64 dtype and 6 columns with object dtype.
3. In our dataset, we observed null values in the following columns:

* 16 null values in name column
* 10052 null values in last\_review column
* 10052 null values in reviews\_per\_month column

We have the following column provided to us in the dataset:

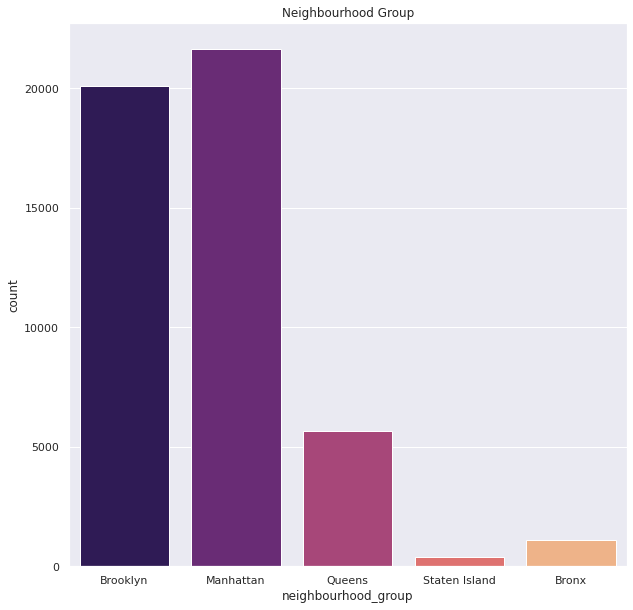
* id: Column id is a unique column in the dataset
* name: This column contains the name of the listing
* host\_id: This column contains the host IDs of the various hosts. Each host has a unique host ID
* host\_name: This column contains the name of the hosts for a listing
* neighbourhood\_group: It is a categorical column containg different neighbourhood groups
* neighbourhood: It is a categorical column containg the various neighbourhoods of a listing
* lattitude: It is a numerical column containg the latitude of the geographical location of the listing
* longitude: It is a numerical column containg the longitude of the geographical location of the listing
* room\_type: It is an categorical column containg different room types
* price: This column contains the price of the listings
* minimum\_nights: It contains the minimum number of nights spend by tourists in a listing
* number\_of\_reviews: This column shows how many reviews are there for a particular listing
* last\_review: This column contains the last date when the listing was reviewed
* reviews\_per\_month: This column contains the number of reviews for a particular listing in a month
* calculated\_host\_listings\_count: This column shows number of listings of a particular host
* avalaibility\_365: This column shows the avalaibilty of a listing on yearly basis.

**Data Analysis**

Basic Data Analysis of categorical column

Objective: - To understand the count and types of listings across different neighborhood and neighborhood group

1. Count of Airbnb listings across different neighborhood groups

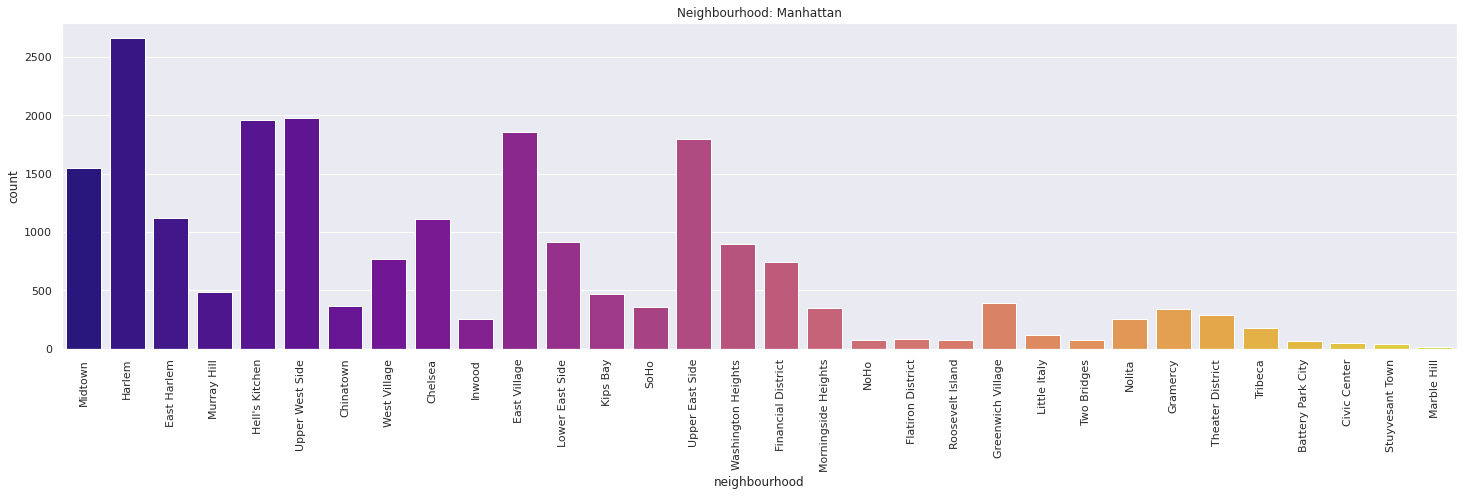


1. Count of Airbnb listings across neighborhoods in different neighborhood groups:-
2. Brooklyn

Chart, waterfall chart

Description automatically generated

1. Manhattan



1. Queens

Chart

Description automatically generated

1. Staten Island

Chart, waterfall chart

Description automatically generated

1. Bronx

Chart, bar chart

Description automatically generated

1. Count of different types of room types in NYC

Chart, pie chart

Description automatically generated

I answered following questions to get better understanding of the data and conclusions

Question 1 What is the average price based on the location and room type.

Answer : The average price according to different neighborhood are:-

A screenshot of a computer

Description automatically generated

Visualization:-

Chart, bar chart

Description automatically generated

Question 2 Minimum and maximum prices in different neighborhood according to room type.

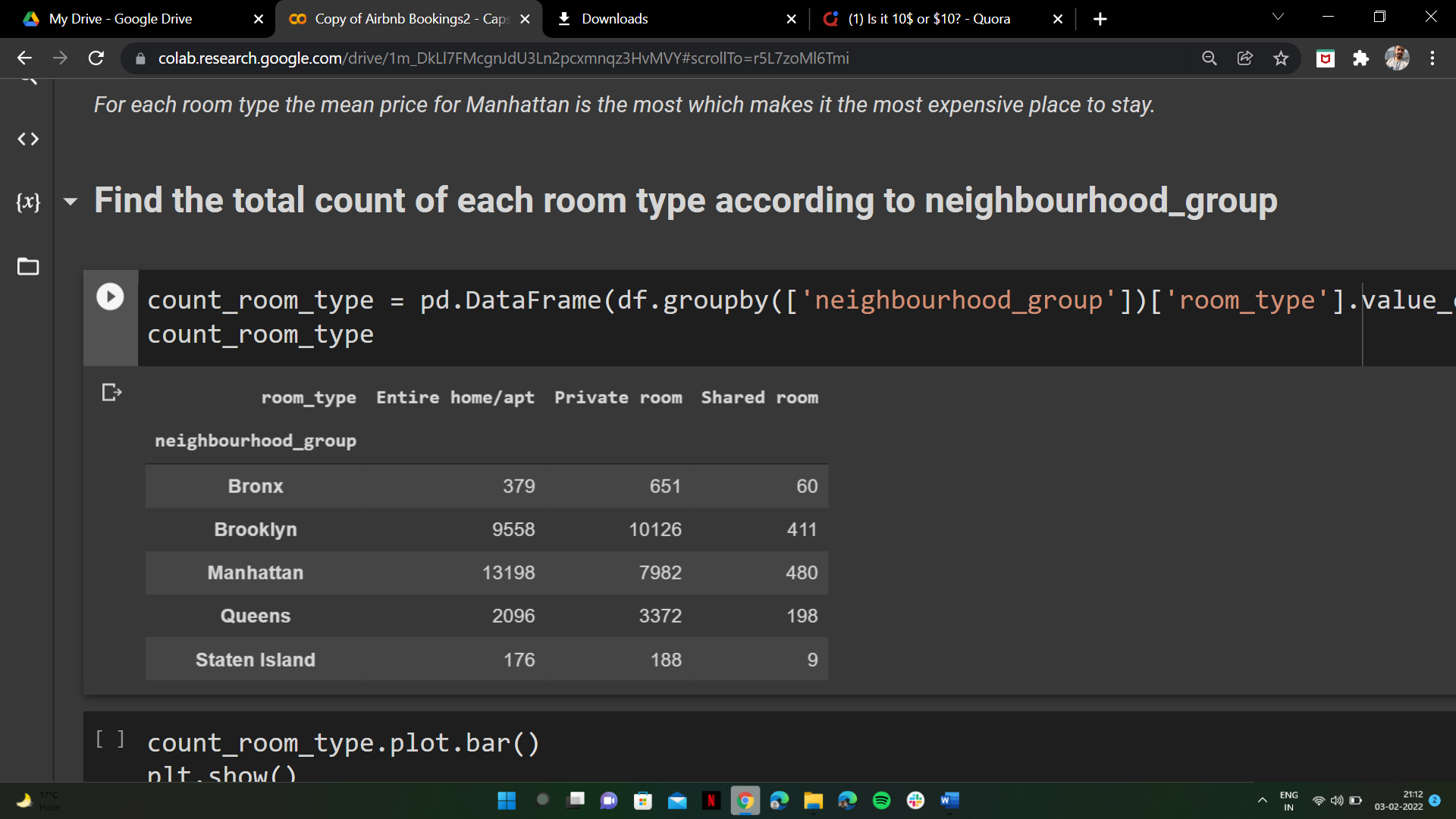
Answer: The minimum and maximum price are as follows:-

A screenshot of a computer

Description automatically generated

Question 3 Find the total count of each room type according to neighborhood group

Answer: The total count of each room type are as follows:-



Visualization:-

Chart, bar chart

Description automatically generated

Question 4 Find the relationship between the neighborhood group and availability of rooms.

Answer: The median of availability of room type in different neighborhood are as follows:-

A screenshot of a computer

Description automatically generated

Visualization:-

Chart, bar chart

Description automatically generated

Question 5 The average number of nights stay in every room type in a neighborhood.

Answer: The average number of nights stay are as follows:-

A screenshot of a computer

Description automatically generated

Question 6 Top 10 hosts according to count of listings

Answer: The top host has 327 number of listings in entire NYC

Chart, bar chart, histogram

Description automatically generated

Question 7 Top 10 highest listing neighborhood

Answer: The top 10 highest listing neighborhood are as follows:-

A screenshot of a computer

Description automatically generated

Question 8 Find the most frequent words used by host in the name of their listings.

Answer:

A picture containing text, newspaper

Description automatically generated

**Conclusion:-**

1. Manhattan is the most emphasized area in New York for hosting.
2. The most popular room type is Entire home / apt with 52 % of listings, and the least popular is shared room with only 2.4 % of listings.
3. The average price for Entire home / apt room type is the most but there is no significant difference between the price of Shared room and Private room in most of the neighborhoods.
4. The prices for Entire home / apt as room type in Manhattan and Brooklyn varies from $10 to $10000.
5. Manhattan is the most busiest place of New York.
6. In every neighborhood, people stay in the Entire home / apt room type for longer period of time.
7. The top host of Airbnb New York has 327 number of listings.
8. Since the availability of Entire home/apt and Private room is relatively low in Manhattan new hosts should invest in these room type.
9. Host used to describe their property in the name of the listing.