**Airbnb Booking Analysis**

**Contributor:**

**1) Khushboo Khanrah**

**2) Rahul Verma**

**3) Vishal Tomer**

**4) Sumit Sharma**

**5) Khushboo Khobragade**

**Alma Better Capstone Project**

**Abstract:**

Airbnb is a service that lets property owners rent out their spaces

to travelers looking for a place to stay. Travelers can rent a space

for multiple people to share, a shared space with private rooms, or the entire property for themselves. On the other hand, airbnb enables travel to book unique homestays from local hosts, saving their money and giving them a chance to interact with locals. Catering is on demand by travel industry, Airbnb is present in over 190 countries across the world.

**Problem statement:**

The main motive of this is to identify the parameters that influence property listings on the platform. We will basic exploratory data analysis (EDA) techniques to explore and visualise the dataset from Airbnb in NYC. This dataset will help us to find the distribution of every Airbnb listing based on its location , price , reviews , room type ,etc. We together examined this data from various perspective and ideas , also came up with some of the other interesting findings. This will help Airbnb’s marketing , finance and technical teams in making strategic data-driven decisions .

**Introduction:**

**Exploratory Data Analysis (EDA)**is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations. Exploratory data analysis is essential for any business.

**Data Summary:**

We found that dataset was pretty much clean except from missing values in some columns, we can specify it based upon initial assessment . By using info() method, we will 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:

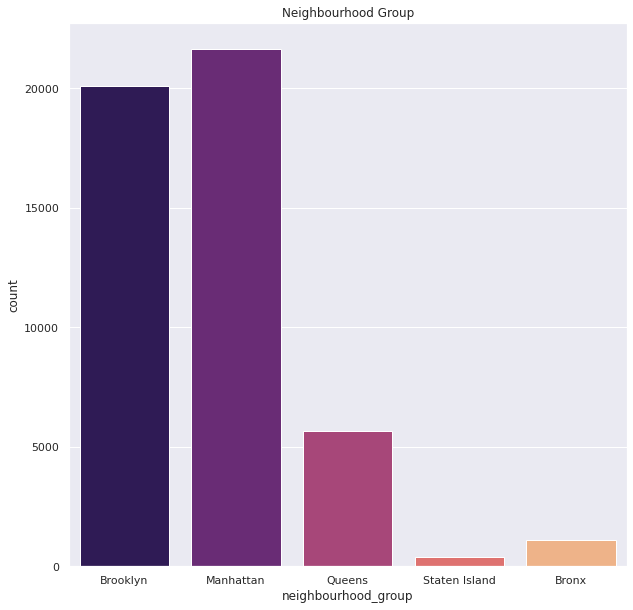
* id : Column id is a unique column in the dataset
* name : This column contains the name of the listing
* host\_id : This column contains the host IDs of the various hosts. Each host has a unique host ID
* host\_name : This column contains the name of the hosts for a listing
* neighbourhood\_group : It is an categorical column containg different neighbourhood groups
* neighbourhood : It is an categorical column containg the various neighbourhoods of a listing
* lattitude : It is an numerical column containg the latitude of the geographical location of the listing
* longitude : It is an 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 listings
* 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 neighbourhood and neighbourhood group

1. Count of Airbnb listings across different neighbourhood groups



1. Count of Airbnb listings across neighbourhoods in different neighbourhood groups:-
2. Brooklyn

Chart, waterfall chart

Description automatically generated

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

We answered following questions to get better understanding of the data and concluions :

# Question 1:-

# What can we learn about different hosts and areas?

**Answer :-**



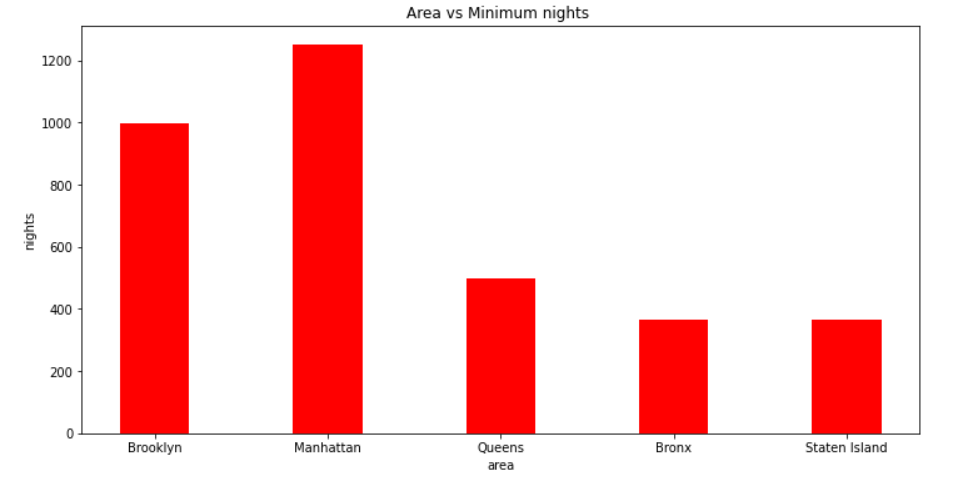
As we can see most number of listings are from Manhattan created sonder(NYC)

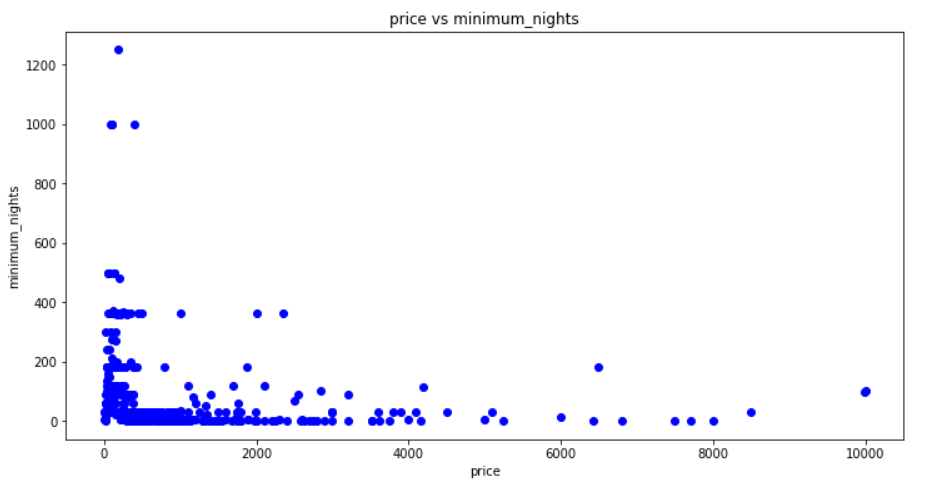
## **Question 2 :-**

## What can we learn from predictions? (ex: locations, prices, reviews, etc.)

**Answer :-**

Area of minimum nights:-

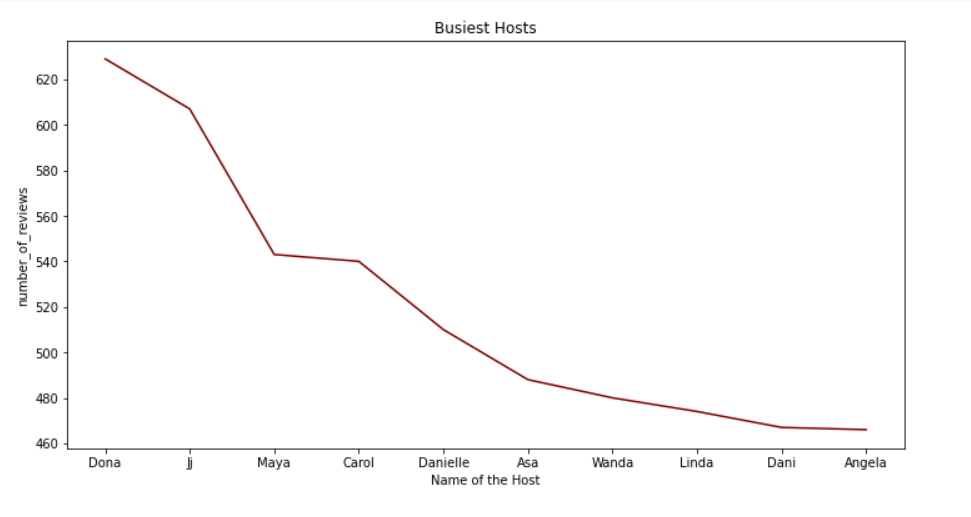




# From the above Analysis we can say that most people prefer to stay in nights where price is less.

# Question 3:- Which hosts are the busiest and why?

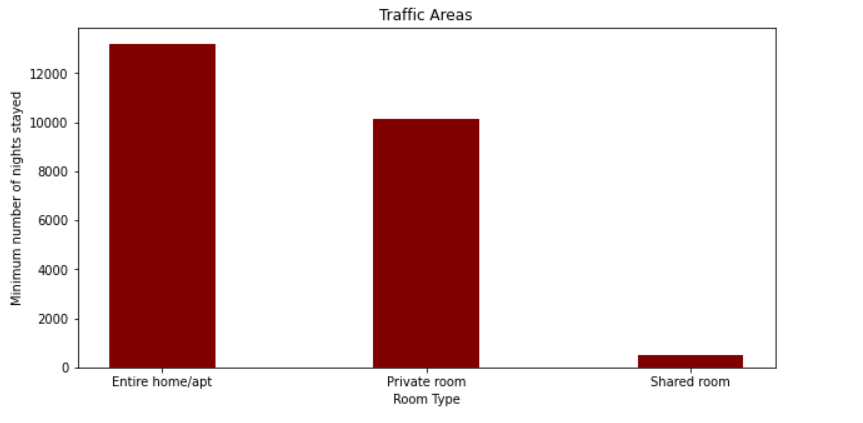
# Answer :-

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

# Question 4:-

# Is there any noticeable difference of traffic among different areas?

**Answer :-**

****

# From the Above Analysis We can Stay 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.

# Question 5:-

# Which neighbourhood group has the highest price for the listings as an average price?

# Answer:-

# We can state that Manhattan has the highest range price for the listings with about $ 140 as an average price

# Description: C:\Users\Levovo\Pictures\Capture.PNG

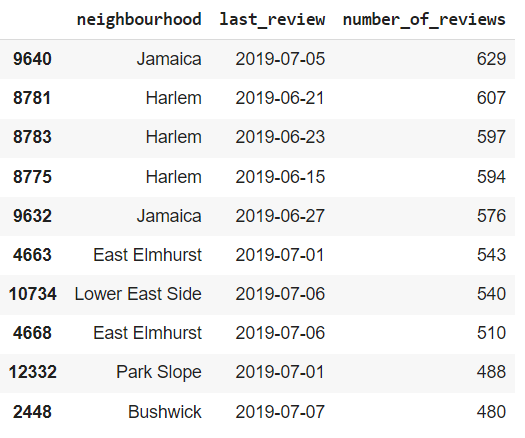
# From the boxplot above,

# We can state that Manhattan has the highest range price for the listings with about $ 140 as an average price

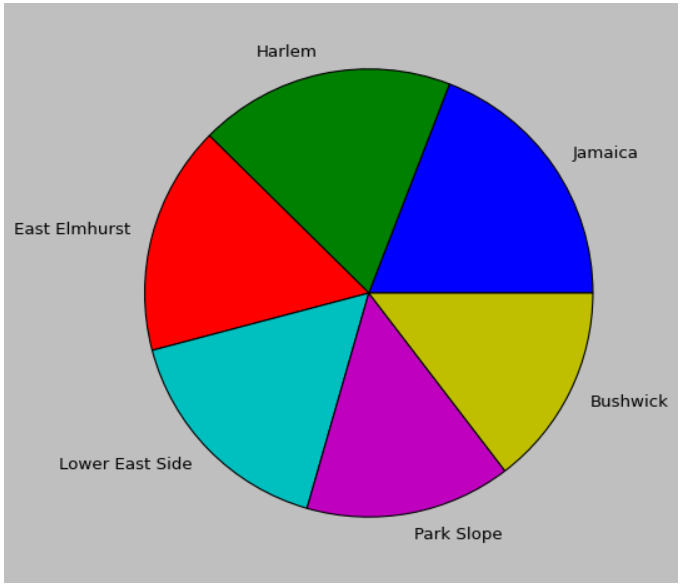
**Question 6:-**

Which neighbourhood has the highest last review on the basis of date.

**Answer:-**



number\_of\_reviews And neighbourhoods



we can see that jamaica has the highest number of reviews in comparison of all neighbourhood as per last reviews on date : 2019-07-05

* **Conclusion:-**
* AFTER DOING ANALYSIS OF AIRBNB BOOKING 2019 BOOKING DATA WE GET A CONCLUSION
* MANHATTAN HAS A MAXIMUM HOST LISTING COUNT.
* BROOKLYN NEIGHBOURHOOD GROUP HAS A MAXIMUN NIGHT .
* DONA IS A BUSIEST HOST IN PRIVATE ROOM TYPE.
* MANHATTAN GROUP HAS A HIGHEST TRAFFIC IN ENTIRE HOME/APT ROOM TYPE.
* MANHATTAN GROUP ALSO HAS A HIGHEST PRICE IN COMPARISON OF OTHER GROUP WITH MAXIMUM NIGHT STAYS.
* JAMAICA HAS A MAXIMUM NO. OF REVIEWS IN COMPARISON WITH OTHER .