Capstone Project- 1

Airbnb Bookings Analysis

By Team GodSpeed

Members:

Akash Jaiswal

Jagdish Mahato

Abhishek Rana

Kanika Raj

What is Airbnb?

- Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily home stays 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. The company was **founded** in **2008** by ***Brian Chesky***, ***Nathan Blecharczyk*** and ***Joe Gebbia***.

- This Airbnb is a shortened version of its original name, AirBedandBreakfast.com.

History of Airbnb

After moving to San Francisco in October 2007, roommates and former schoolmates ***Brian Chesky*** and ***Joe Gebbia*** came up with the idea of putting an **air mattress** in their living room and turning it into a **bed** and **breakfast**.

In February 2008, ***Nathan Blecharczyk***, Chesky's former roommate, joined as the Chief- Technology Officer and the third co-founder of the new venture, which they named **AirBed & Breakfast**. They put together a website that offered short-term living quarters and breakfast for those who were unable to book a hotel in the saturated market.

The site Airbedandbreakfast.com officially launched on **August 11, 2008**. The founders had their **first customers** in town in the **summer** of **2008**, during the Industrial Design Conference held by Industrial Designers Society of America, where travelers had a hard time finding lodging in the city.

About Airbnb Booking Data

- 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 analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

- This dataset has around **49,000** observations in it with **16 columns** and it is a mix between categorical and numeric values.

Problem Statement Analysis

- For this project we are doing analysis of Airbnb's New York City(NYC) data of 2019. NYC is not only the most famous city in the world but also one of the top global destination for visitors drawn to its museums, entertainment, restaurants and commerce.

- Our main objective is to find out the key metrics that influence the listing of properties on the platform. For this, we will explore and visualize the dataset from Airbnb in NYC using basic exploratory data analysis (EDA) techniques.

- Data analysis on thousands of listings provided through Airbnb is a crucial factor for the company.

- We will be finding out the distribution of every Airbnb listing based on their location, including their price range, room type, listing name, and other related factors.

❖ Understanding the Data

- There are 49,000 observations with various types of field in our dataset.
- List of field:

- Id
- Name
- Host_id
- Host_name
- Neighbourhood_group
- Neighbourhood
- Latitude
- Longitude
- Room_type

- Price
- Minimum_nights
- Number_of_reviews
- Last_review
- Reviews_per_month
- Calculated_host_listing_count
- availabilty_365

Agenda

➔ **We try to answer following questions for Airbnb:**

- ❑ Number of active host per location.
- ❑ What can we learn from predictions? (ex: locations, prices, reviews, etc.)
- ❑ Which hosts are the busiest and why?
- ❑ Is there any noticeable difference of traffic among different areas. and what could be the reason for it?
- ❑ Finding Relation between neighborhood group and availability of rooms.

Agenda(Cont...)

- Find top 10 hosts with most listings.
- Find total no. of nights spend per location.
- Total no. of nights spends per room types.
- Top 10 highest listing neighborhood.
- What is the average preferred price by customers according to the location?
- Find the total count of each room type.
- Which is the most preferred place by hosts to do their business.

Map of New York City(NYC)

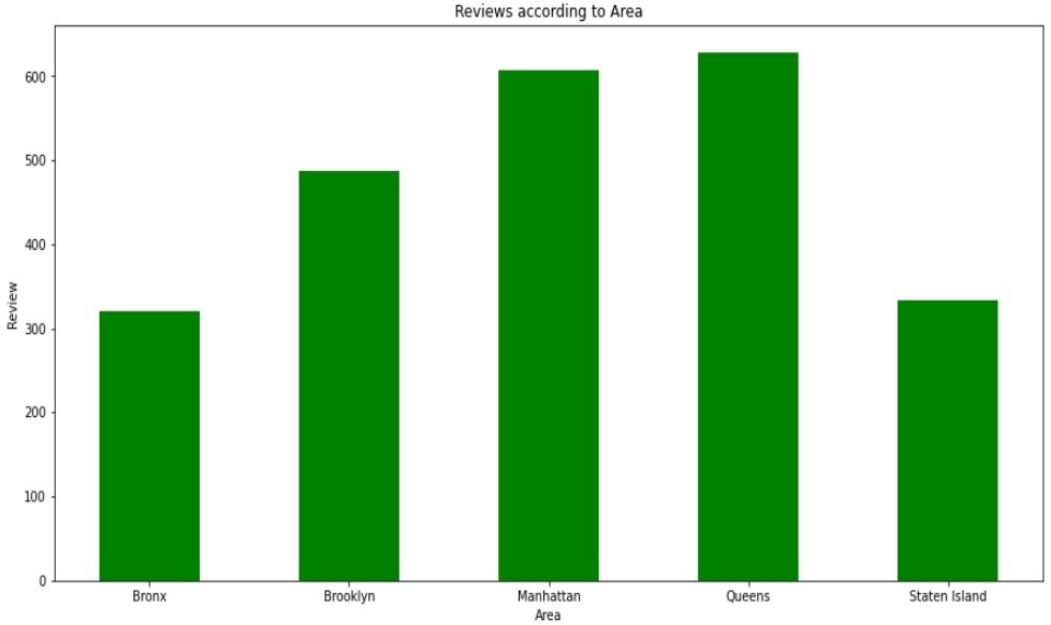
1. What can we learn about different hosts and areas?

- From this table we can see that neighborhood group ‘Manhattan’ has the highest number of calculated listing counts followed by host name blue ground.
- Kara and Kazuya hosts have the least number of calculated listing counts.

	host_name	neighbourhood_group	calculated_host_listings_count
13217	Sonder (NYC)	Manhattan	327
1834	Blueground	Manhattan	232
1833	Blueground	Brooklyn	232
7275	Kara	Manhattan	121
7480	Kazuya	Queens	103

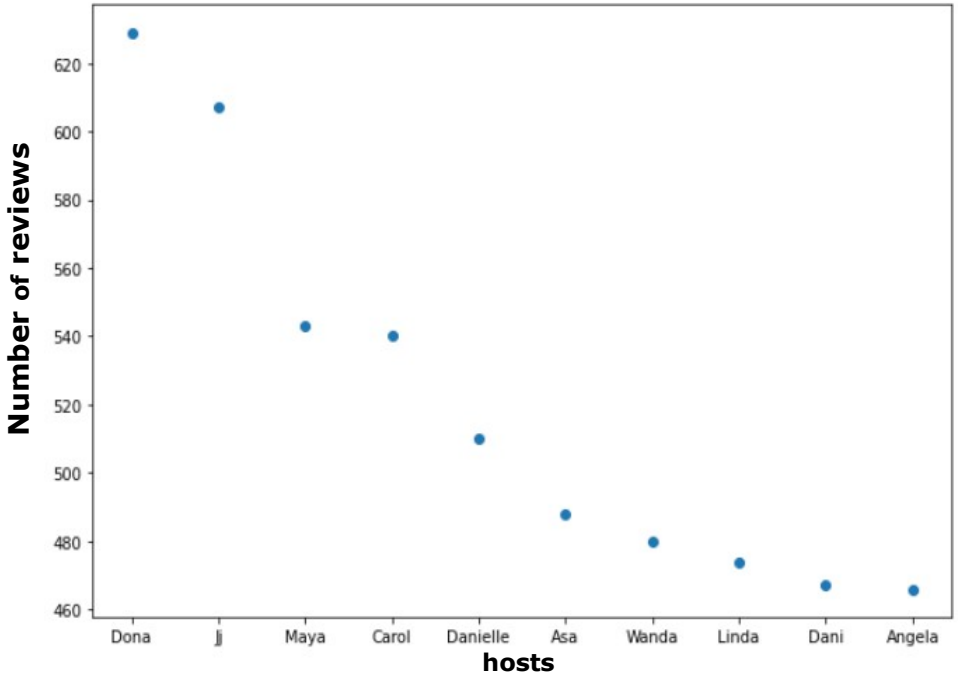
2.What can we learn from predictions? (ex: locations, prices, reviews, etc.)

- This analysis concludes that most people prefer to stay where number of reviews are more.
- Queens got the highest number of reviews with the total of 629 reviews.
- Bronx neighborhood group consist of lowest number of reviews.



3. Which hosts are the busiest and why?

- **Calculating the busiest hosts and rooms preferred by people.**
- **Dona and ji are the most busiest hosts.**
- **These hosts are the most busiest because they have listed entire home and private rooms.**
- **Private rooms and entire homes are mostly preferred by the people.**



Scatter plot

Agenda(Jagdish)

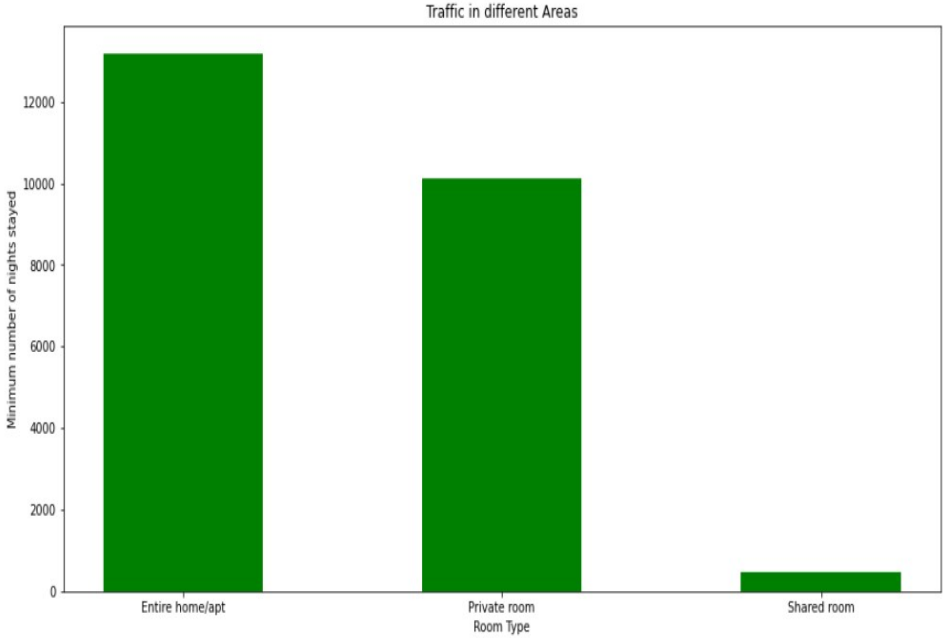
- ❖ **Is there any noticeable difference of traffic among different areas and what could be the reason for it?**
  - ➢ Finding the different type of areas in airbnb dataset.
  - ➢ Calculating the highest and lowest number of 'night spend by customers' on various room types.

- ❖ **Room types and their relation with availability in different neighborhood groups?**
  - ➢ Finding the different type of neighbourhood group in dataset.
  - ➢ Count the value of each neighbourhood group and type of room in the dataset.
  - ➢ Analysis of neighbourhood groups with the highest and lowest number of 'availability of rooms' during 365 days in a year.

- ❖ **Top 10 hosts with most number of listings.**
  - ➢ Calculate the counting value of each host_id.
  - ➢ Which host_id has the maximum listing value.
  - ➢ Which host_id has the least counting value in the dataset.

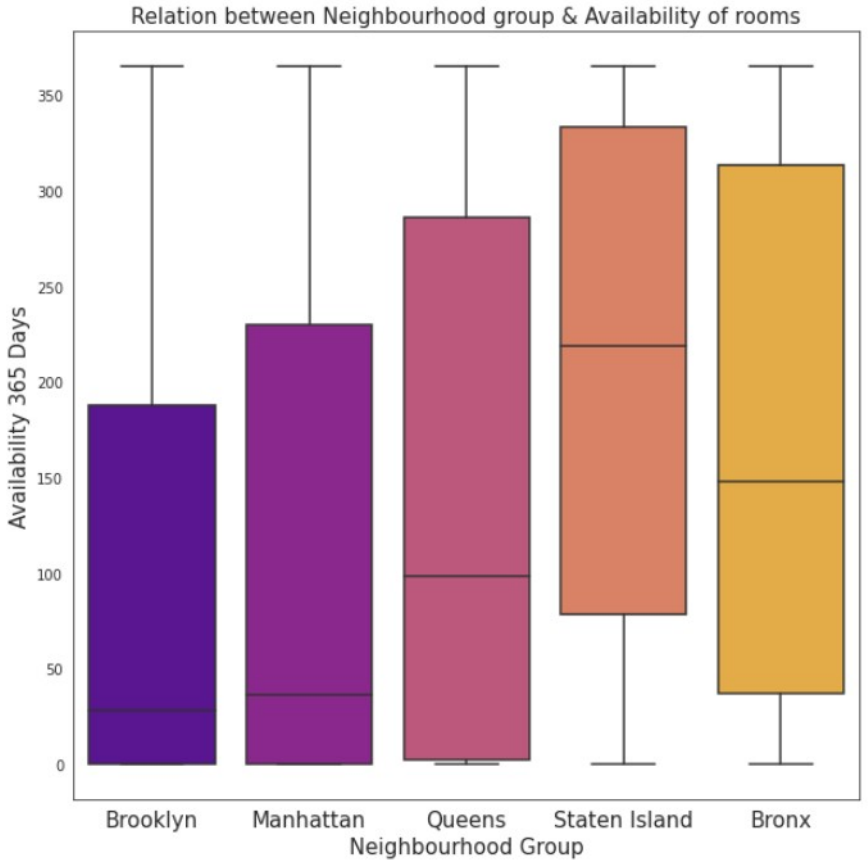
1. Is there any noticeable difference of traffic among different areas and what could be the reason for it?

- ‘Entire home/apt’ consists of highest number of night spend.
- Traffic was the highest in the area of ‘entire home/apt’.
- Minimum number of nights were spend in ‘shared room’ as it has the lowest traffic.



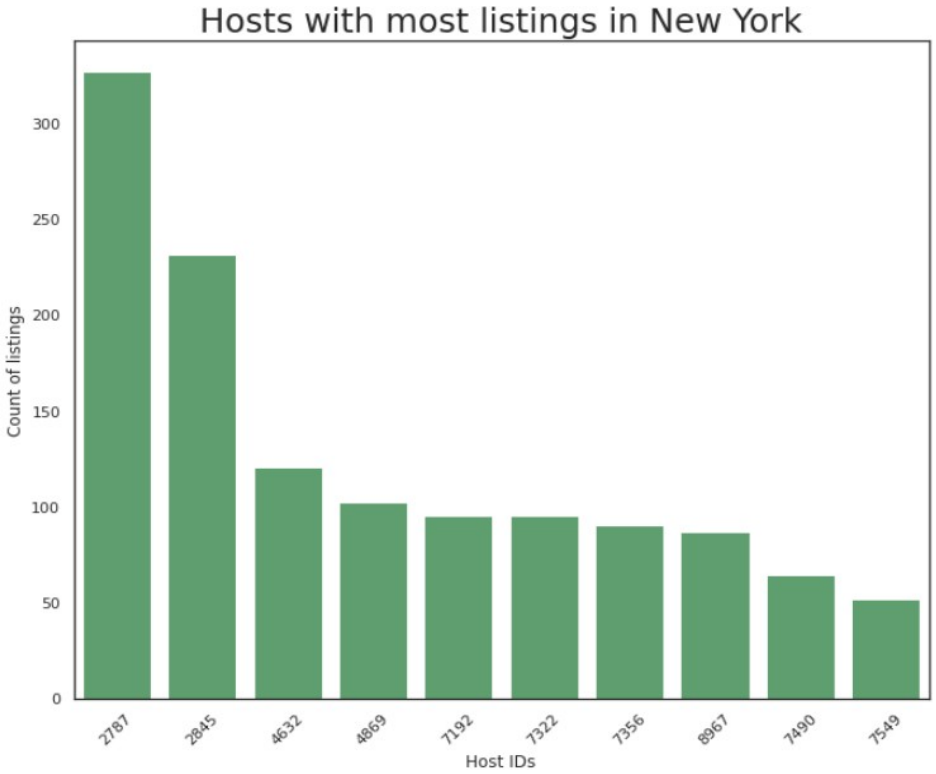
2. Room types and their relation with availability in different neighbourhood groups?

- Brooklyn and Manhattan have the least availability of rooms overall, as low as 0 days.
- Staten Island and Bronx has the highest availability rate overall at around 300 days.
- Form this analysis we can say that people stay for longer duration of time in Private rooms in Brooklyn and Manhattan.



3. Top 10 hosts with most number of listings.

- From the chart, we can see that count of listing by top 10 hosts is almost 2.5%(1270 listings) of the whole dataset.
- Even one of the hosts has more than 300 listings!



❖ Agenda(Kanika)

➔ **I try to answer following questions for Airbnb:**

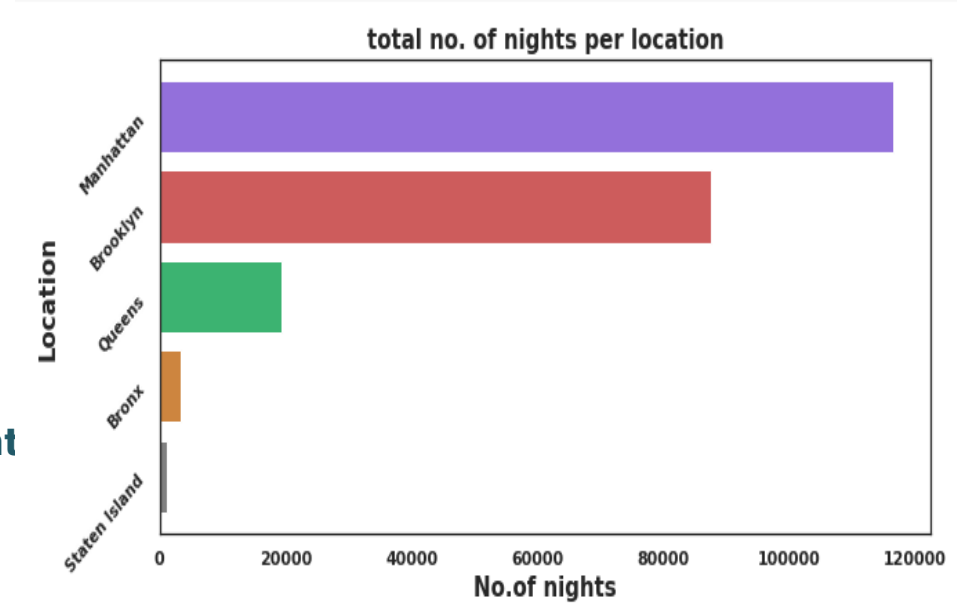
- **What is the total number of nights spending by customer across different location.**
  - Find the values of all neighbourhood and minimum nights in airbnb dataset.
  - Which is the most preferred neighbourhood /Location by the customer.

- **What is the total no. of nights spends per room types.**
  - How many unique types of room in airbnb dataset.
  - Find the maximum(minimum nights) spending by customer across different room type.
  - Calculate the percent value of each room type w.r.t to minimum/maximum nights spend by customer.

- **Which are the top 10 highest listing neighbourhood/Location.**
  - Find the listing values of all neighbourhood in airbnb dataset.
  - Which particular neighbourhood has the highest listing value.
  - Which neighbourhood has the lowest listing value.

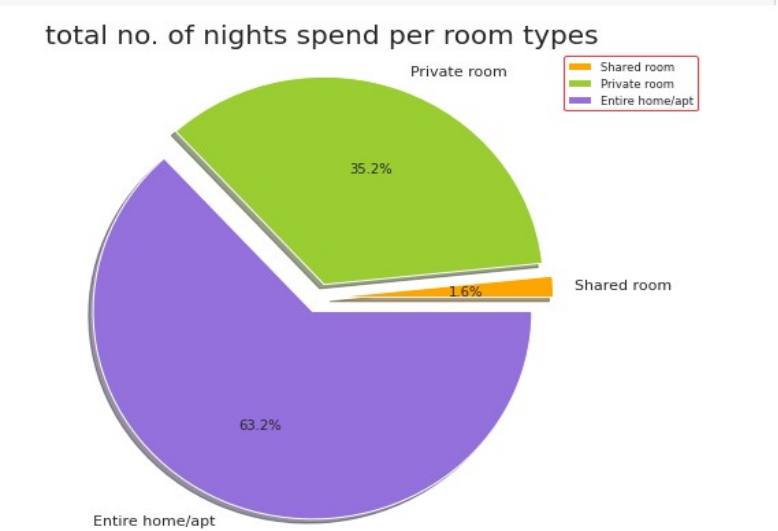
1. Find total no. of nights spend as per location

- The locations are categorized based on its total number of nights spend by customer.
- From this chart we can get an overall idea of which location is mostly preferred by the customer.
- We can state that more customers prefer Manhattan and Brooklyn for night stay as compared to other locations.



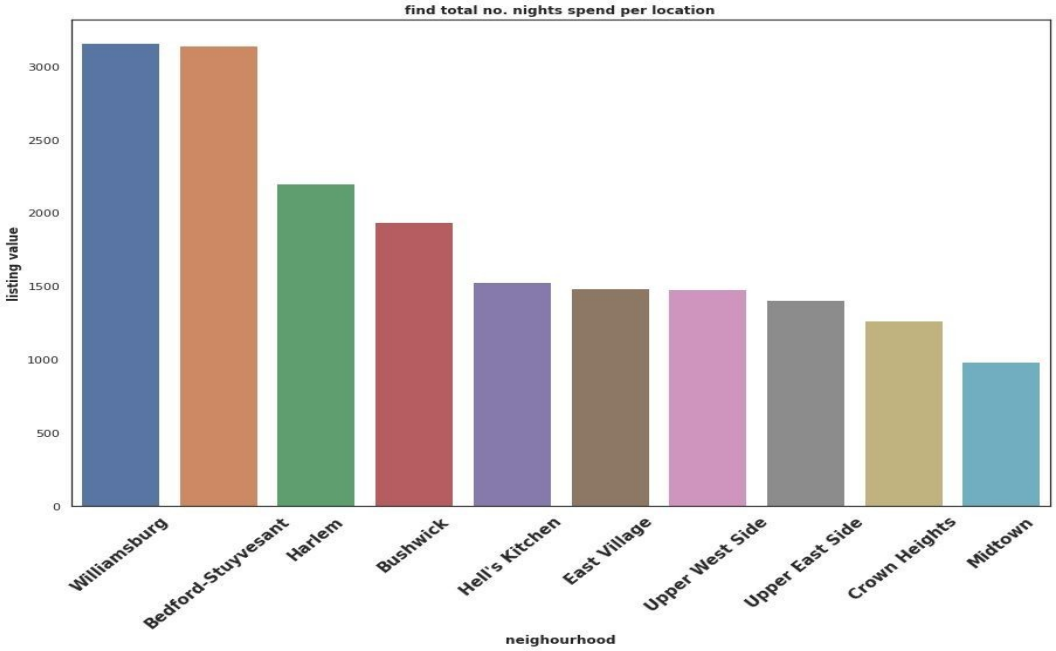
2. Total no. of nights spend per room types

- Here we can state that in which room type customers prefer for night stay.
- From pie chart we can conclude that 63.2% customers spend night in entire home/apt.
- Only 1.6% customers spend night in shared room.



3. Top ten highest listing neighbourhood

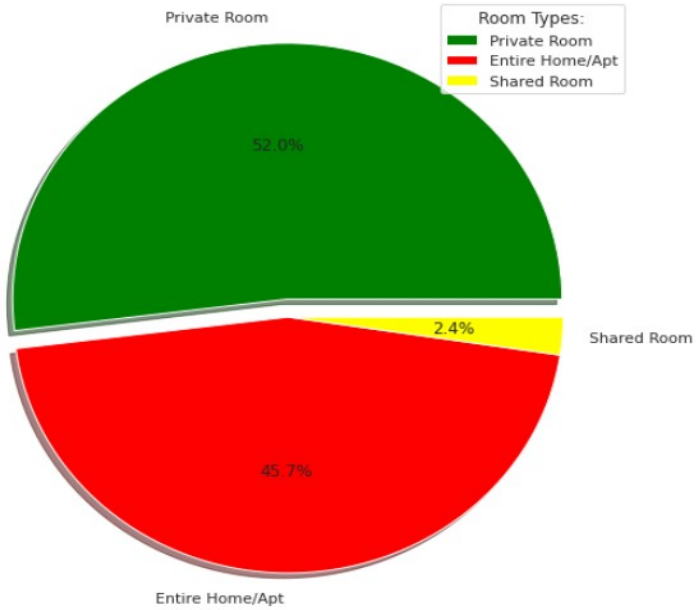
- Here we can state that Williamsburg has the highest listing.
- And almost same in Bedford-Stuyvesant.
- It’s a tourist attraction hence the number of listings could be higher.



1. Total count of each room type as per listing.

➢ **Based on the Analysis we found that:**

- **Entire home/apt has the highest number of listing of 52% among other room types.**
- **Private room has 45.7% of listing among other room types.**
- **Shared Room is the least listed room type at only 2.4% in total.**



❖ Agenda(Abhishek)

I try to answer following questions for Airbnb:

➢ **What is the total count of each room type as per listing.**

- Finding unique values from column room type in Airbnb dataset.
- Which room type has the highest number of listing.
- Which room type has the lowest number of listing.

➢ **What is the average preferred price by customers according to the location.**

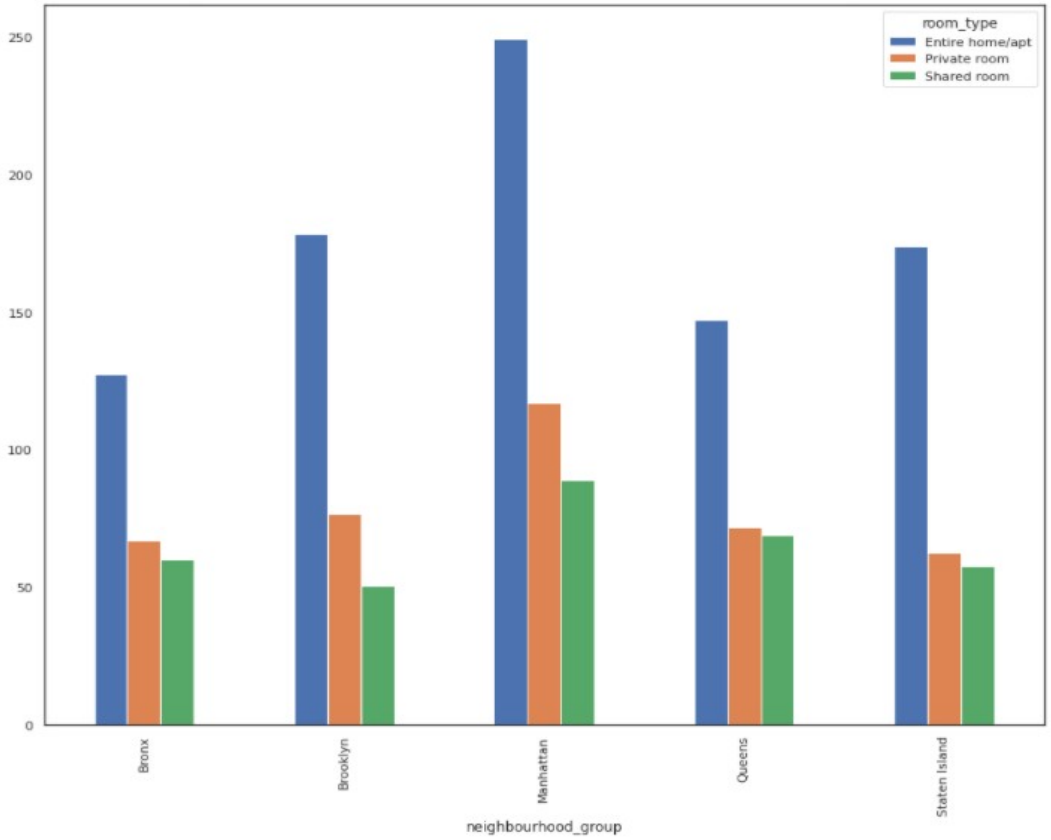
- Finding average price of different room types according to location.
- Which location has the highest and lowest average price for the room types.

➢ **Which is the most preferred place by hosts to do their business.**

- Finding number of hosts according to locations.
- Which is the most and least preferred place by hosts.

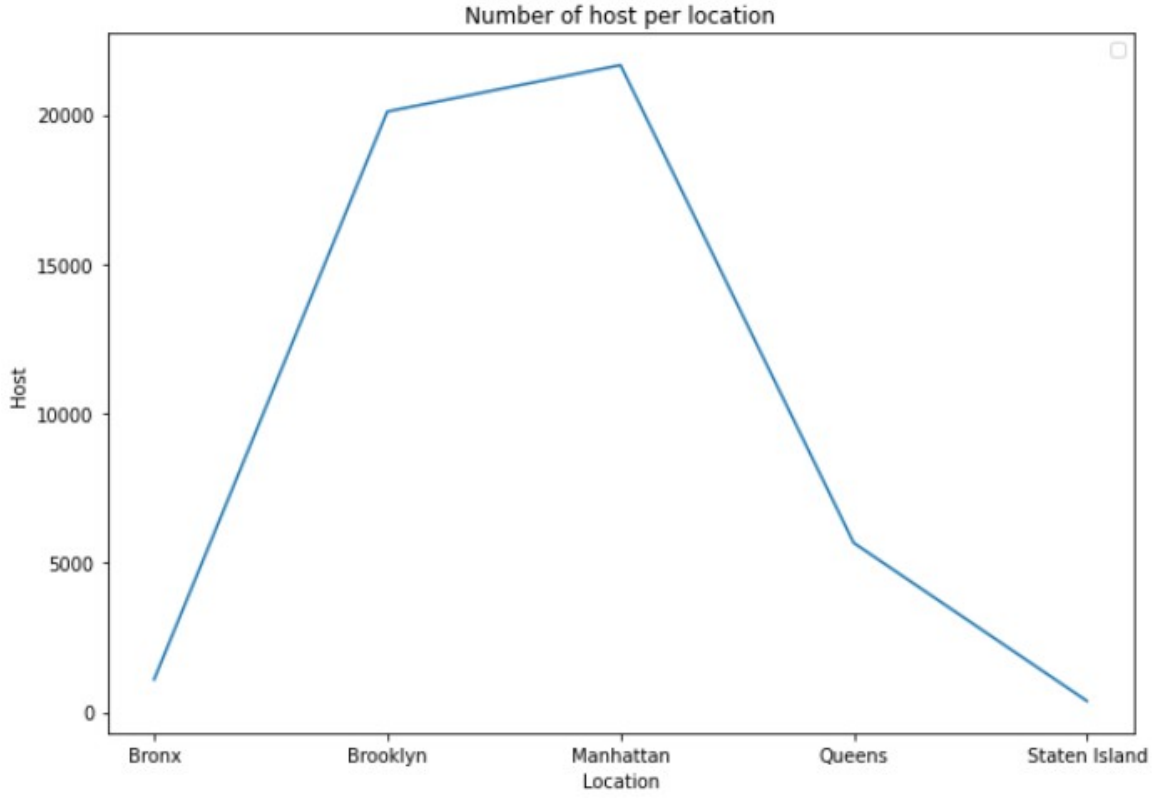
2. What is the average preferred price by customers according to the location?

- As we can see in Manhattan average price of Entire home/apt is highest, compared to other location.
- Average price of shared room is lowest in brooklyn.
- Queens, Staten Island and Bronx share almost the same price.



3.Which is the most preferred place by hosts to do their business.

- Manhattan and Brooklyn is the most preferred place by hosts.
- Bronx and Staten Island is low traffic for host as compared to other location
- Queens has a average number of host.



Challenges Faced

- For answering some of the questions we had to understand the business model of Airbnb that how they work.
- Reading the dataset and understanding the meaning of some columns.
- Handling Nan values, null values and duplicates.
- Designing multiple visualizations to summarize the information in the dataset and successfully communicate the results and trends to the reader.

What's factors makes Airbnb successful

- Any service provider is successful if it has:-
  - A high average user ratings.
  - Focus on User Experience.
  - A good number of positive reviews.
  - A good number of monthly average users.
  - High revenue per customer and so on.

Analysis Summary:

- In new York **Manhattan** is the most focused place for hosts to do their business.
- 'Entire home/apt' room type has the highest number of listing of **52%** and 'Shared Room' is the least listed room type at only **2.4%** in total.
- People stay for longer duration of time in Private rooms in **Brooklyn** and **Manhattan.**
- Count of listing by top 10 hosts is almost **2.5% (1270 listings)** of the whole dataset.
- More customer preferred **Manhattan** location for night stay than **Brooklyn.**
- Total of **63.2%** customer spend night in Entire home and **1.6%** spend night in Shared room.

Analysis Summary(Cont.)

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

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

- If there are **more** number of **Reviews** for particular Neighborhood group that means that place is a **tourist place.**

- If people are not staying more then **one night** means they are **travelers.**

Thank You

Team GodSpeed☺