

# **Capstone Project**

# **Airbnb Bookings Analysis**

## **Team Members:**

**Riya Sharma** – [riya272001@gmail.com](mailto:riya272001@gmail.com)

**Suchi Raghuvanshi** – [suchiraghuvanshi1996@gmail.com](mailto:suchiraghuvanshi1996@gmail.com)

**Akshit Saxena** – [akshit.saxena78@gmail.com](mailto:akshit.saxena78@gmail.com)

**GitHub Commit (Riya)** - <https://github.com/Sharmariyaa/Airbnb-data-analysis.git>

**GitHub Commit (Suchi)** - <https://github.com/sui16/airbnb-data-anaylsis.git>

**GitHub Commit (Akshit)** - <https://github.com/Akkii830/AIRBNB-DATA-ANALYSIS.git>

# Airbnb Bookings

## Introduction

We are analyzing Airbnb Dataset provided by the Almabetter for the capstone project. Brief introduction of Airbnb-Airbnb, Inc.is an American company that operates an online marketplace for lodging, primarily homestays 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. Airbnb is a shortened version of its original name i.e. **Air Bed and Breakfast.**

# Air Bed and Breakfast

## 1. Data Exploration

Explore the dataset using different functions like head, tail, info, describe, columns, shape, etc. Here we got some idea about the dataset.

## 2. Data Cleaning

Checked the null values followed by duplicated entries, after detection filled the null values with 'Unknow' and 'No Name' also dropped feature having most of the null records. Also detected outliers from some features.

## 3. Data Visualization

Plotted different bar charts , plots , scatterplots for different features. Got insights from that.

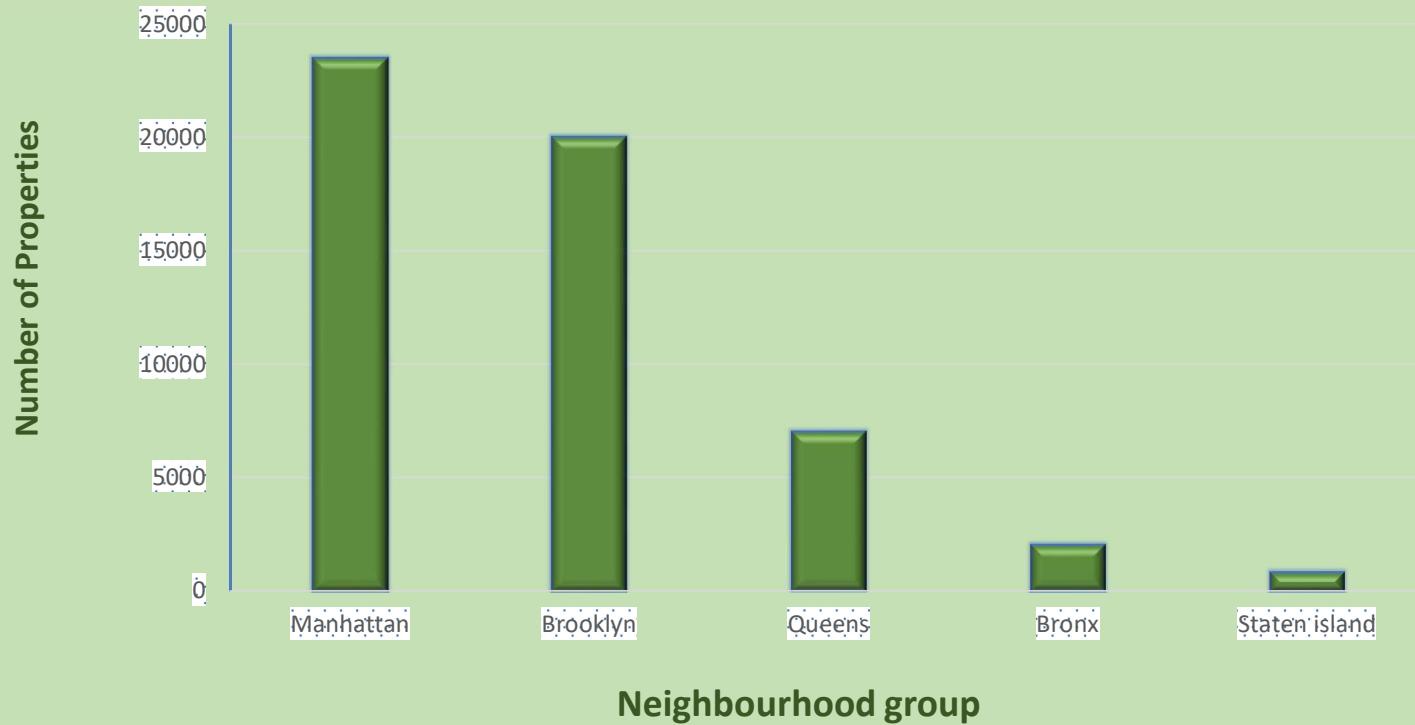
## 4. Conclusion

Concluded the analysis for main categorical variables.

# Define Column Names

- **Id** : It is the unique Id of the listings.
- **Name** : Name of the property of listing.
- **Host Id** : Unique Id of the host.
- **Host Name** : Name of the host of the property.
- **Neighborhood group** : Location of the property.
- **Neighborhood** : Area in the location of the property.
- **Latitude** : Horizontal coordinates of the area.
- **Longitude** : Vertical coordinates of the area.
- **Room type** : kind of room(Private room , Entire home/Apt , Shared room).
- **Price** : Price that is charge by the host.
- **Minimum nights** : Minimum number of nights must to stay in the property.
- **Number of reviews** : Total number of reviews that host had.
- **Last review** : Date of the last review that customer give.
- **Reviews per month** : It is the mean of the reviews that is given to the host per month.
- **Calculated host listings count** : Total listings of that particular host.
- **Availability 365** : It is the number of days the host available throughout the year.

# PROPERTY COUNT



## Analyzing The Dataset

- From the given dataset we see that there are 5 unique locations are there named by neighbourhood\_group. Each neighbourhood group have various number of properties that are rented by different hosts. All these properties are listed in the Airbnb .
- In the previous slide we see a Bar graph in which the x axis denoted the unique name of the neighbourhood groups and the y axis denoted the total number of properties each neighbourhood group have.

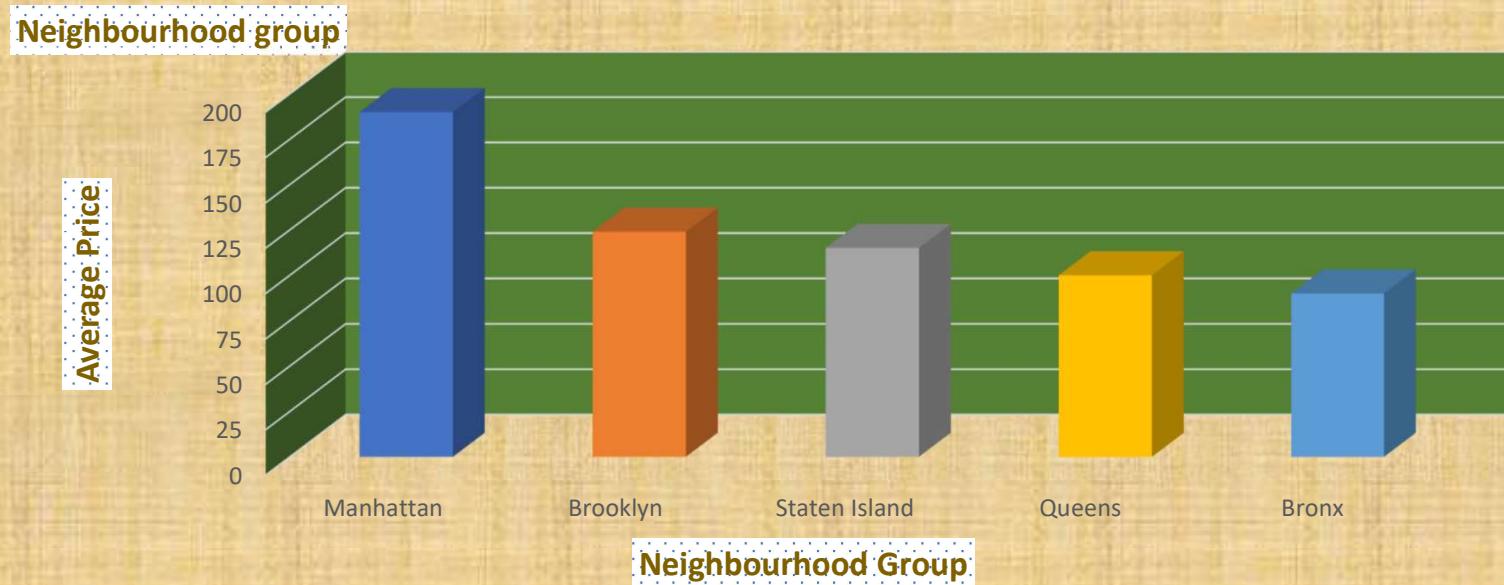
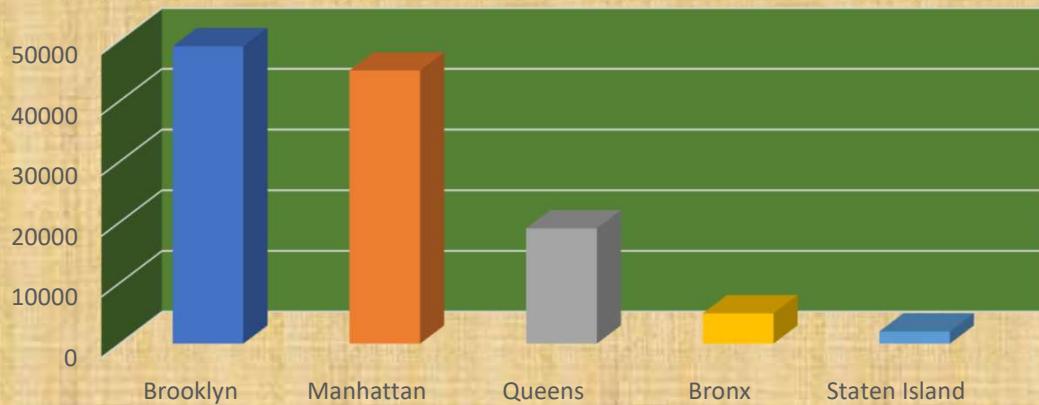
- **Observation :**

We see that Manhattan and Brooklyn have the maximum number of properties .

- **Why?**

The reason behind of this is may be these locations are more developed so because of this more number of people are used to travelled here for the meetings and for the vacations, etc. These all factors increases the demands in these particular locations. So that is may be the reason they have more properties .

# Total Reviews and Average Price VS Neighbourhood Group



# Reviews and Average Price

- In our given dataset the column named number of reviews give the total reviews that each host had and the Price column give the information of the price that charged by the host from customers.
- Reviews are given by the customers based on the behavior of the host , the facilities provided by the host and the hospitality of the host.

## **Observation :**

- In first plot we see that Brooklyn and Manhattan have the maximum number of reviews while the Staten Island has least number of reviews.

In another plot we see that in Manhattan average price of the properties are very high.

## **Why?**

- In Manhattan and Brooklyn the demand is more that is more people are used to visit in these locations . As more people come so more culture are exchange between the people. People get new experiences. So these things get the customer to attached with the hosts and leave the places with good experience. That all factors are responsible for getting the more reviews.
- Because of all these factors the average price at the places are higher.

# Property Count For Each Room Type In Different Neighbourhood Group

From the dataset we see that there are three type of rooms are available i.e.

1. Entire room/apt(apartment)
2. Private room
3. Shared room

Now we want to see how much properties are there for each kind of room at different locations(neighbourhood group).

So, for that we plot a bar graph that gives us the data of the number of properties for each room type at the various locations.

X axis represent the neighbourhood group and Y axis represent the number of properties.

Bar 1 represent Entire home/apt

Bar 2 represent Private room

Bar 3 represent Shared room

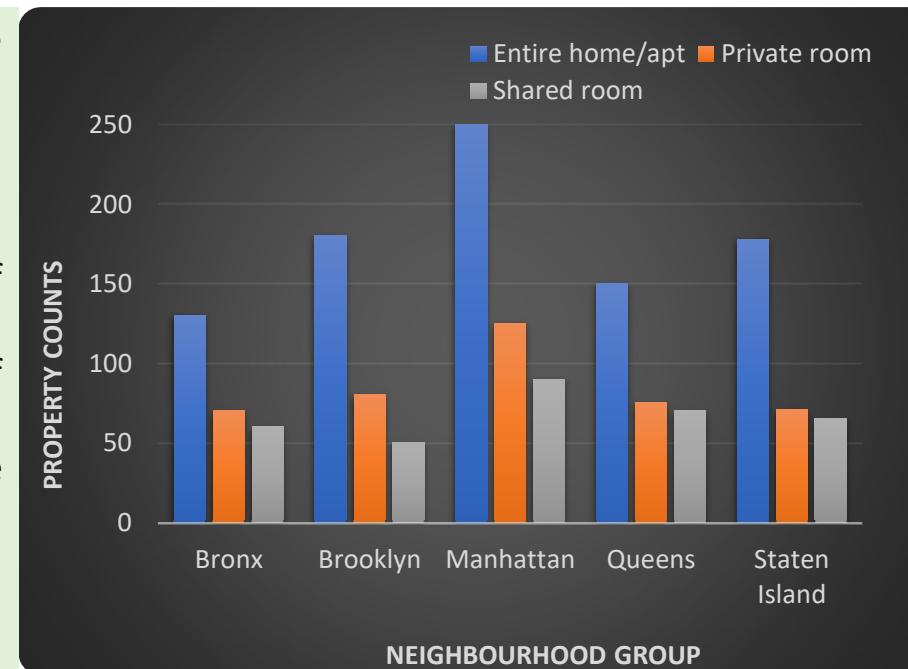
## **Observation :**

We get to know that maximum number of people prefer Entire room/apt and least number of people prefer Shared room.

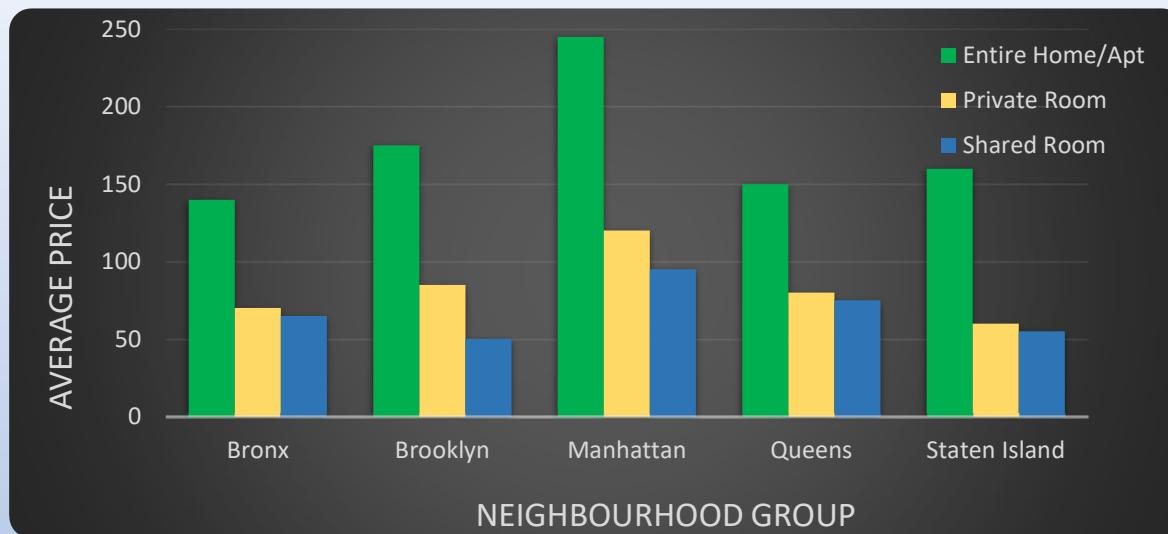
## **Why?**

This is because mostly people are traveling with their family and for the business meetings so more host provide entire room/apt facilities.

While very few people travel individually and less number of people prefer to shared the rooms. That's why very few properties are available that provide shared room facilities.



## Average Price For Each Room Type In Different Neighbourhood Group



In above graph it gives the data of the mean price charged by the host for each kind of room at various locations.

- **Observation :**

We see that average price of entire room type is high in all the neighbourhood groups. While the average price of the shared room type is less.

- **Why?**

The reason behind of this is more demand for entire room so the rate is high while less demand for shared room so the rate is low.

# Minimum Night Stay Category VS Number of Host



# Analyses The Minimum Night Stay Category

- There are numbers of host are listed in AIRBNB and they have their own allowability.
- Some hosts allow their customers for 2 ,4,5,10 days etc. as minimum number of nights to stay in their property.
- So we made a minimum night stay category column that gives the data like –

1 day=One day	7 days = 1 Week	30-180 days = More than a month
2 and 3 days = 2-3 days	7-29 days = Less than a month	180-365 days = More than half year
4 -6 days = Few days	30 days = 1 Month	>365 days = More than a year
- Now we plot a graph that gives the information of the total number of hosts that provide each minimum night category.

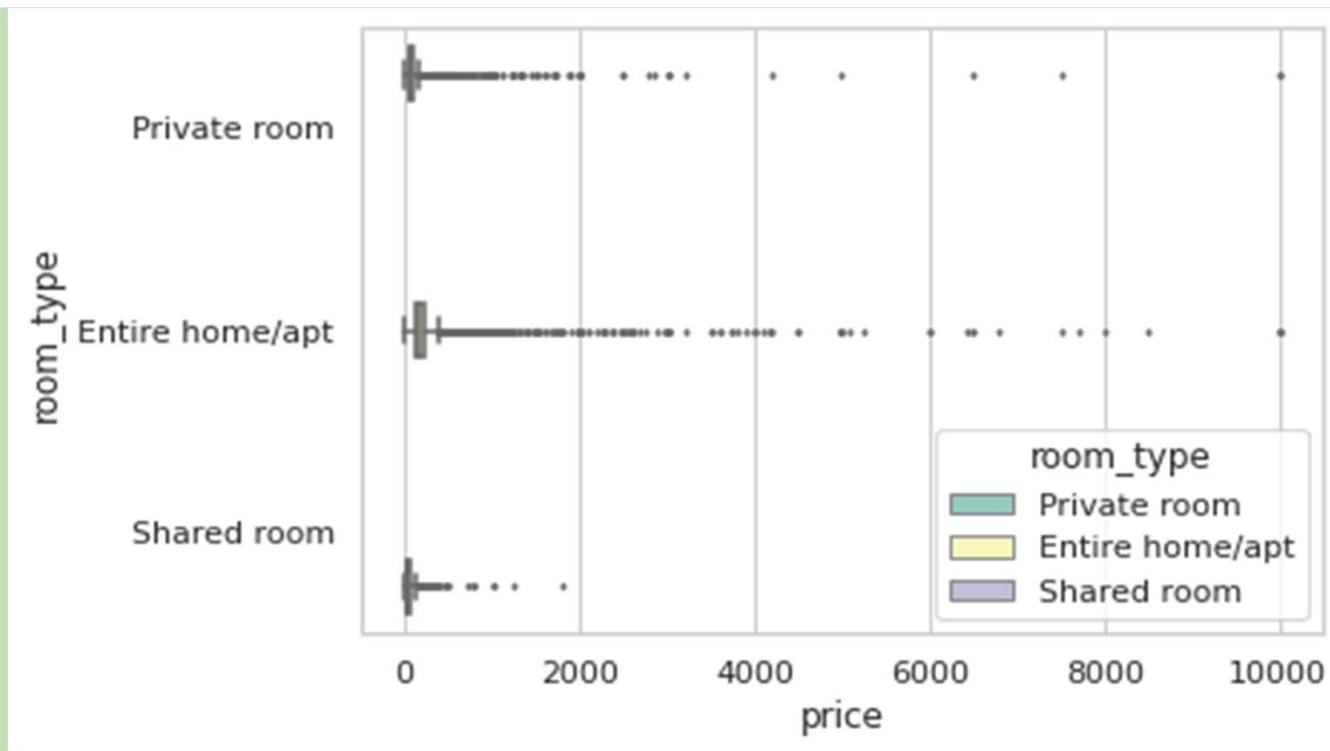
## **Observation :**

- We see that maximum host provides 2-3 nights to stay as minimum. And very few provide more than a year nights to stay.

## **Why?**

- This is because mostly people came for travelling and business meetings that completed in almost 2-3 days.

# Price Column Over Room Type



## **Observation:**

- From the above box plot we can see there are so many outliers in price.
- So that is the reason of having so much high prices.

# Highest Price With Availability, Reviews Per Month, Last Reviews & Neighbourhood Group

```
rbnb_data_df[airbnb_data_df['price']==airbnb_data_df['price'].max()][['host_name','reviews_per_month','last_review','availability_365','price','neighbourhood_group']]
```

	host_name	reviews_per_month	last_review	availability_365	price	neighbourhood_group
9151	Kathrine	0.04	2016-02-13	0	10000	Queens
17692	Erin	0.16	2017-07-27	0	10000	Brooklyn
29238	Jelena	0.00	NaN	83	10000	Manhattan

## Observation:

- Kathrine and Erin have high price and there is also low availability.
- The reviews are also less as many people might not be able to stay due to high price or may be because there are no availability.

## Top 5 Hosts that obtained highest number of reviews

	host_id	host_name	number_of_reviews
21304	37312959	Maya	2273
1052	344035	Brooklyn& Breakfast -Len-	2205
18626	26432133	Danielle	2017
20872	35524316	Yasu & Akiko	1971
21921	40176101	Brady	1818

## Top 5 neighbourhood having highest reviews per month

	neighbourhood	reviews_per_month
0	Theater District	58.50
1	Rosedale	20.94
2	Springfield Gardens	19.75
3	East Elmhurst	16.22
4	Jamaica	15.32

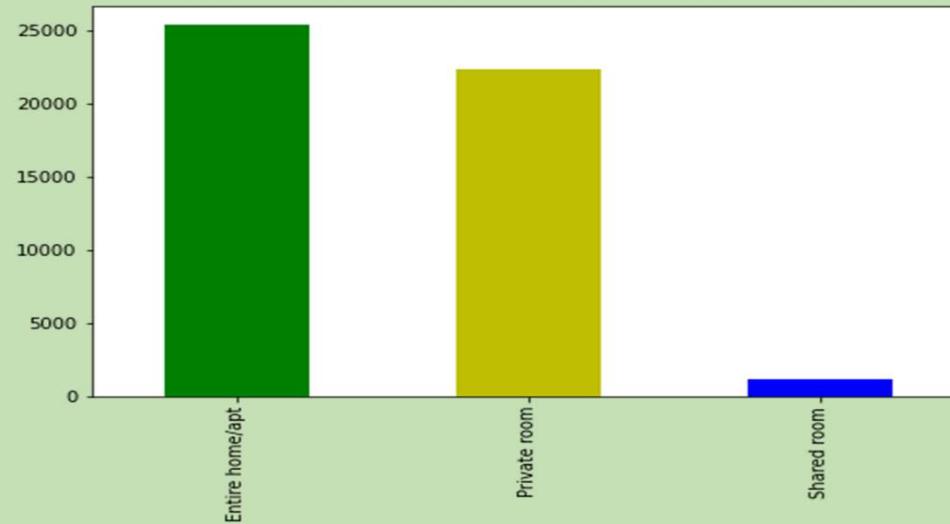
# Different Room Types And Which Is The Most Common Type

neighbourhood_group	room_type	Entire home/apt	Private room	Shared room
Bronx	Entire home/apt	379	652	60
Brooklyn	Entire home/apt	9559	10132	413
Manhattan	Entire home/apt	13199	7982	480
Queens	Entire home/apt	2096	3372	198
Staten Island	Entire home/apt	176	188	9

## **Observation:**

From this we can get to know that there are 3 different types of room

- Entire home/apartment
- Private room
- Shared room



## **Observation:**

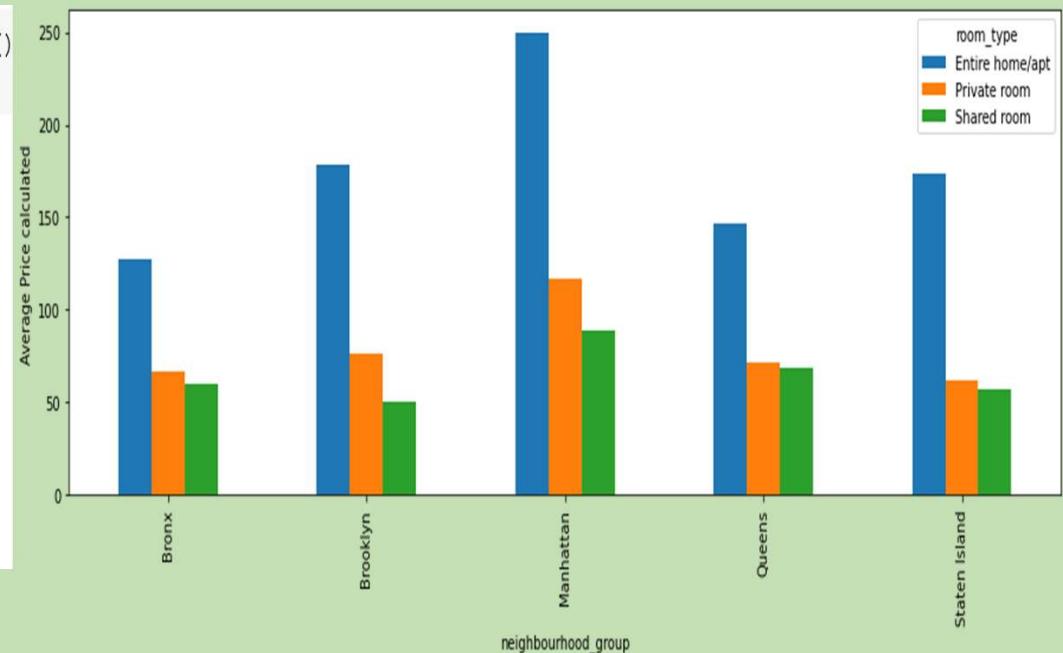
- The maximum numbers of room are Entire home/Apartment and Private room.
- There are only few shared rooms.
- So mostly host prefer to give Entire home/Apartment or Private Rooms rather than Shared rooms.

# Average Preferred Price By Customers

Here We have compared the price according to the neighborhood for each category of Room type

```
[ ] aaverage_price_df = airbnb_data_df.groupby(['neighbourhood_group','room_type'])['price'].mean().unstack()
```

	room_type	Entire home/apt	Private room	Shared room
neighbourhood_group				
Bronx	Entire home/apt	127.506596	66.788344	59.800000
Brooklyn	Entire home/apt	178.327545	76.500099	50.527845
Manhattan	Entire home/apt	249.239109	116.776622	88.977083
Queens	Entire home/apt	147.050573	71.762456	69.020202
Staten Island	Entire home/apt	173.846591	62.292553	57.444444



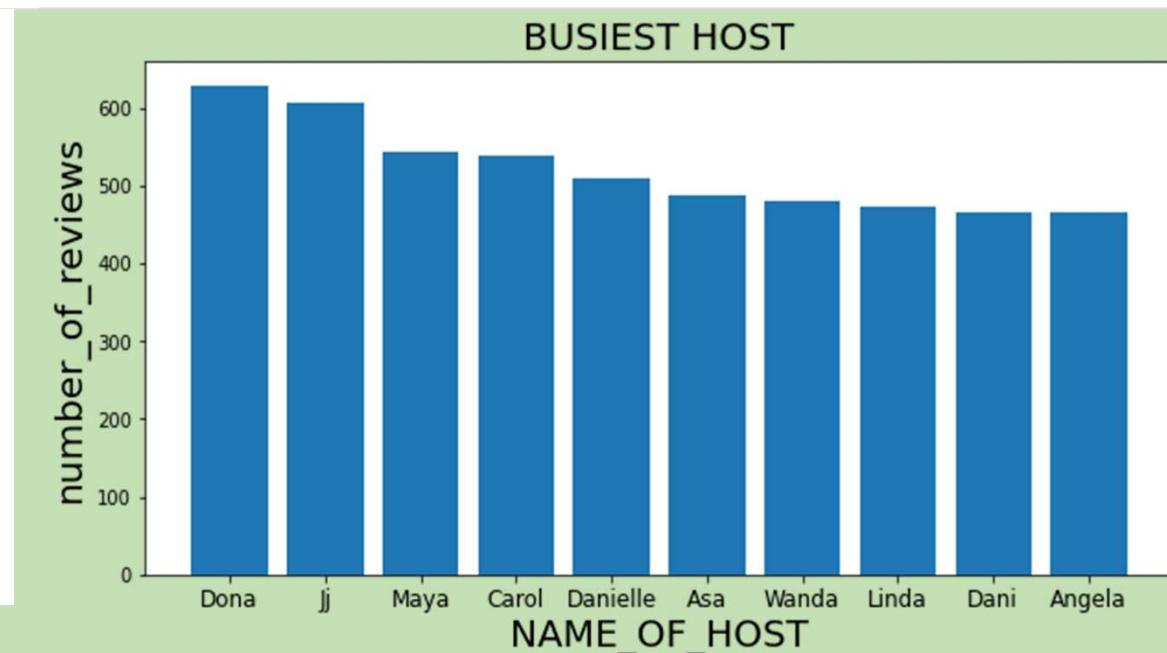
## Observation:

Here we can see the Manhattan is more costly compared to other neighbourhood group and Bronx is the cheapest in the neighbourhood group.

# Top 10 Busiest Hosts

- Following are the top 10 busiest hosts.

	host_name	room_type	neighbourhood_group	number_of_reviews
4598	Dona	Private room	Queens	629
8445	Jj	Private room	Manhattan	607
12044	Maya	Private room	Queens	543
2869	Carol	Private room	Manhattan	540
3988	Danielle	Private room	Queens	510
1637	Asa	Entire home/apt	Brooklyn	488
18272	Wanda	Private room	Brooklyn	480
10578	Linda	Private room	Queens	474
3935	Dani	Entire home/apt	Brooklyn	467
1179	Angela	Private room	Queens	466



## **Observation:**

From the above data we can clearly see that the top 5 busiest host are Dona, Jj, Maya, Carol & Danielle.

# Why these hosts are the busiest?

- From the below dataset we can see that out of 10, 8 hosts has the similar room type which is 'Private Room' and 2 have 'Entire home/apt'.

```
#Get the value count for room_type
pd.DataFrame(imp_host['room_type'].value_counts())
```

room_type	
Private room	8
Entire home/apt	2

## Observation:

- So from the above data we can say that hosts which have more listing of 'Private Room' as room type are the busies hosts.

# Price Per Night:

## **Observation:**

Below are the minimum pricing of top 5 hosts who offers 'Private Room'.

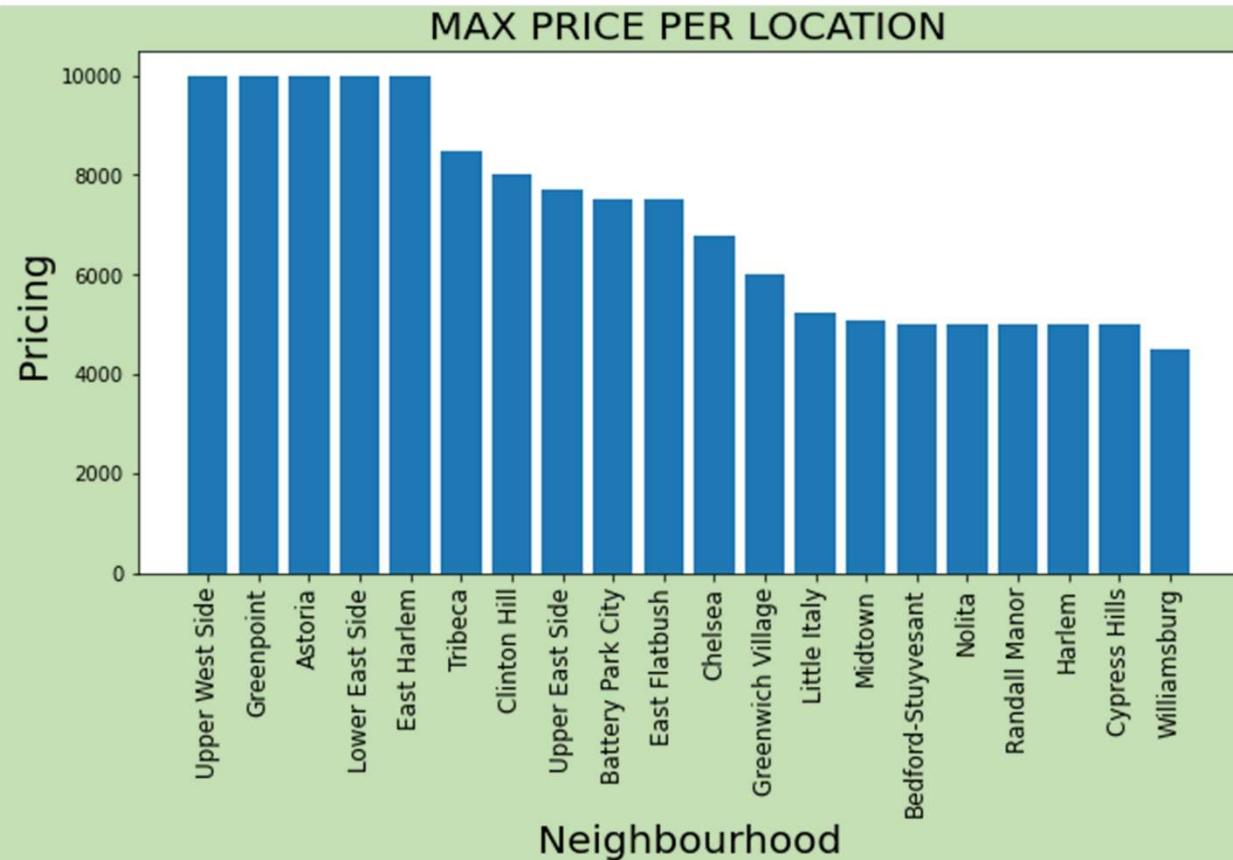
- Maya offers private room at price 32 per night in Queens.
- Carol offers private room at price 35 per night in Brooklyn.
- Danielle offers private room at price 38 per night in Queens.
- Dona offers private room at price 47 per night in Queens.
- Jj offers private room at price 49 per night in Manhattan.

	host_name	price	room_type	neighbourhood_group	minimum_nights
0	Maya	32	Private room	Queens	1
1	Carol	35	Private room	Brooklyn	1
2	Danielle	38	Shared room	Queens	1
7	Dona	47	Private room	Queens	1
11	Jj	49	Private room	Manhattan	1

# Highest Rent Location

- Following are the top 20 highest rent location.

	neighbourhood	price
202	Upper West Side	10000
91	Greenpoint	10000
4	Astoria	10000
119	Lower East Side	9999
61	East Harlem	9999
197	Tribeca	8500
41	Clinton Hill	8000
201	Upper East Side	7703
6	Battery Park City	7500
	60	East Flatbush
	34	Chelsea
	92	Greenwich Village
	115	Little Italy
	127	Midtown
	13	Bedford-Stuyvesant
	145	Nolita
	161	Randall Manor
	94	Harlem
	52	Cypress Hills
	214	Williamsburg



## Observation:

The neighbourhoods with the highest rents are:

- Upper West Side with a price of: 10000
- Greenpoint with a price of: 10000
- Astoria with a price of: 10000
- Lower East Side with a price of: 9999
- East Harlem with a price of: 9999

# Lowest Rent Location

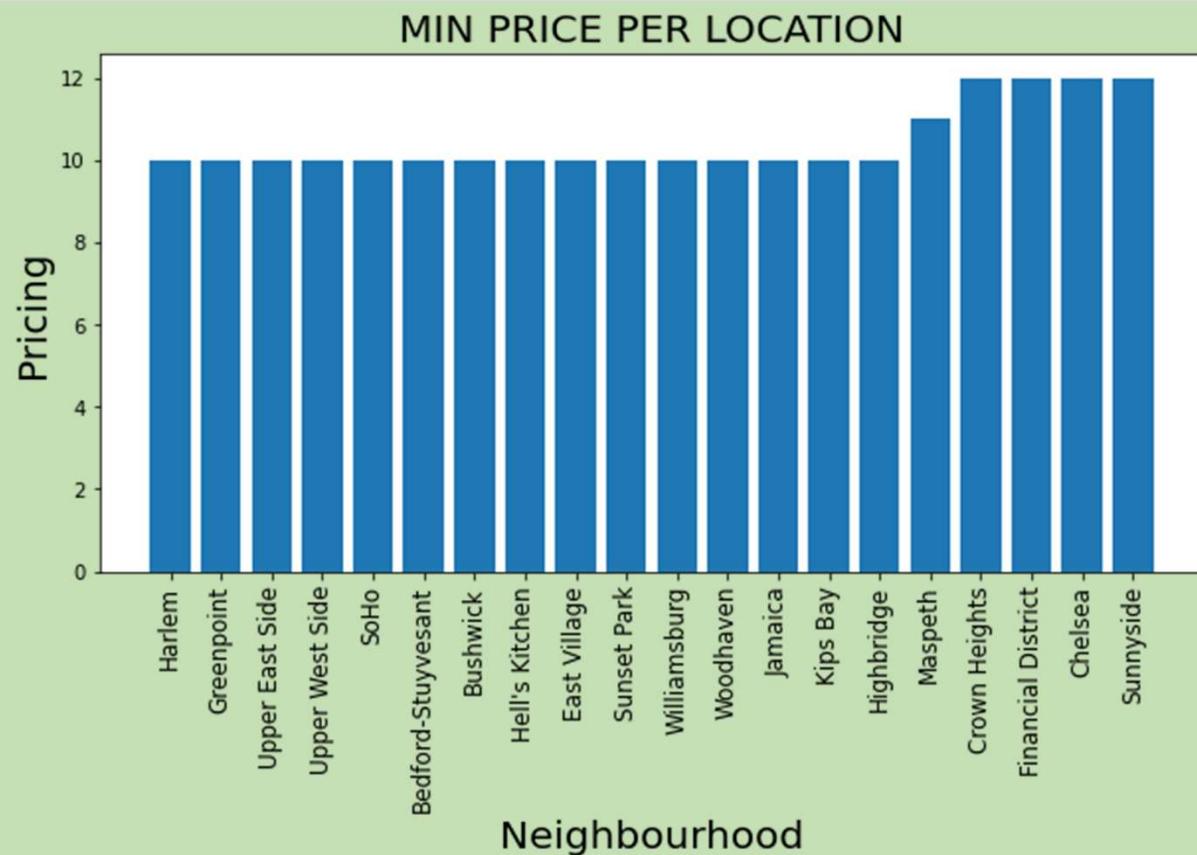
- Following are the top 20 lowest rent location.

	neighbourhood	price
94	Harlem	10
91	Greenpoint	10
201	Upper East Side	10
202	Upper West Side	10
178	SoHo	10
13	Bedford-Stuyvesant	10
28	Bushwick	10
95	Hell's Kitchen	10
64	East Village	10
190	Sunset Park	10
	214	Williamsburg
	217	Woodhaven
	105	Jamaica
	112	Kips Bay
	96	Highbridge
	123	Maspeth
	51	Crown Heights
	73	Financial District
	34	Chelsea
	189	Sunnyside

## Observation:

The neighbourhoods with the lowest rents are:

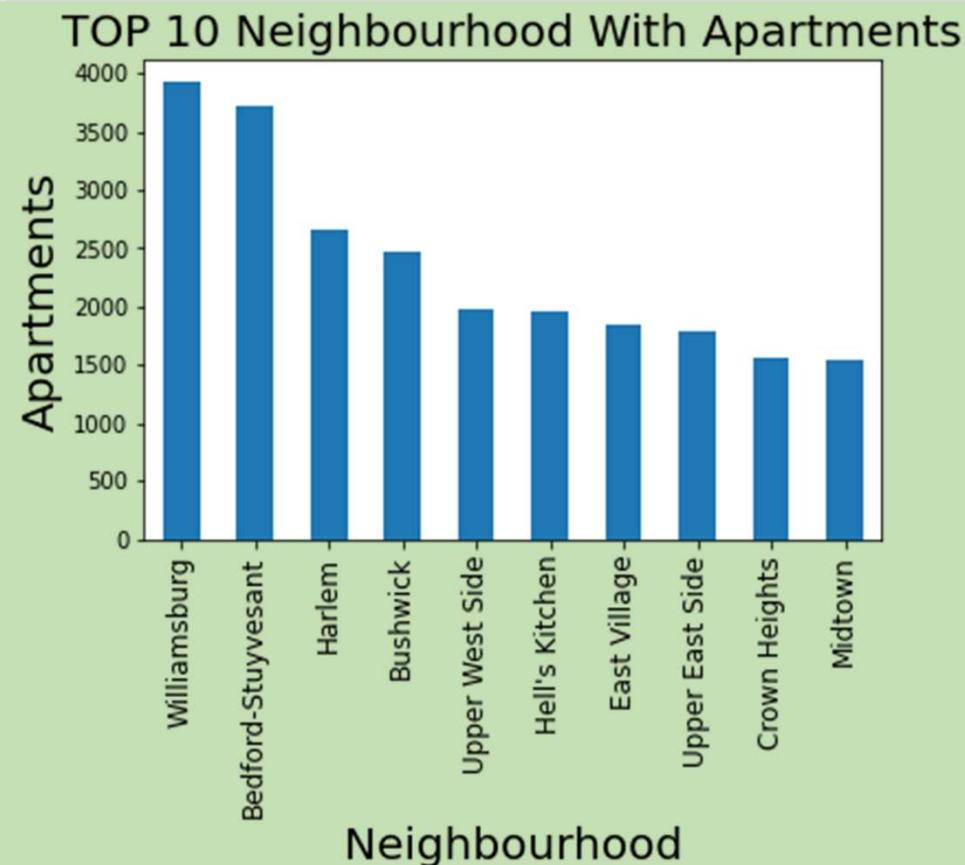
- Harlem with a price of: 10
- Greenpoint with a price of: 10
- Upper East Side with a price of: 10
- Upper West Side with a price of: 10
- SoHo with a price of: 10



# Maximum Number Of Apartments for Airbnb

- Lets see the top 10 neighbourhood with maximum number of apartments.

Williamsburg	3920
Bedford-Stuyvesant	3714
Harlem	2658
Bushwick	2465
Upper West Side	1971
Hell's Kitchen	1958
East Village	1853
Upper East Side	1798
Crown Heights	1564
Midtown	1545



## Observation:

From the given data & graph we get to know the top 10 neighbourhood with maximum number of apartments out of which these 5 are having the most number of apartments

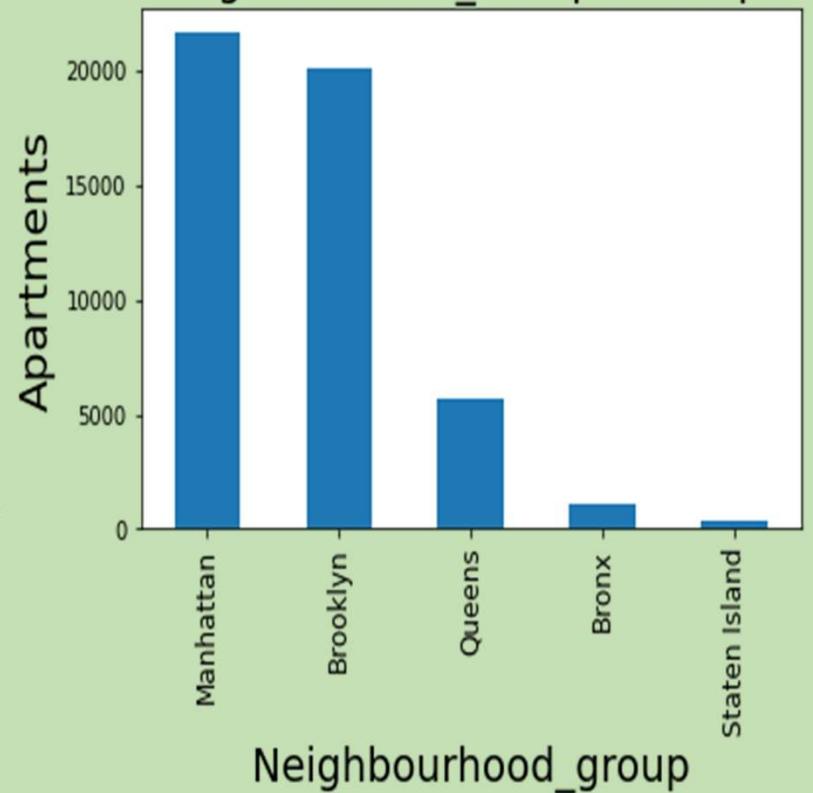
- Williamsburg
- Bedford-Stuyvesant
- Harlem
- Bushwick
- Upper West Side.

# Maximum Number Of Apartments for Airbnb

- Lets see the top 10 neighbourhood\_group with maximum number of apartments.

Manhattan	21661
Brooklyn	20104
Queens	5666
Bronx	1091
Staten Island	373

TOP 10 Neighbourhood\_Group With Apartments



## Observation:

From the given data & graph we get to know the top 10 neighbourhood\_group with maximum number of apartments out of which these 5 are having the most number of apartments.

- Manhattan
- Brooklyn
- Queens
- Bronx
- Stateb Island

# Conclusion

- Manhattan & Queens are the busiest hosts among all.
- Manhattan & Brooklyn have the maximum number of apartments.
- Upper West Sides, Greenpoint, Astoria have the highest rent which is 10000 as compared to others.
- Harlem, Greenpoint, Upper East Side have the lowest rent which is 10 as compared to others.
- Maximum people like to stay in ‘Private Room’, ‘Entire Home/Apartment’.
- People doesn’t like to stay in ‘Shared Room’.
- Maximum hosts provides 2-3 nights to stay as minimum and very few provide more than a year nights to stay.
- Kathrine and Erin have less number of reviews because the high price and low availability people might not like to stay there.
- Queens is the most famous neighbourhood where people get the private room at the affordable price.