

# Capstone Project

## Airbnb analysis



**Presented By :-**  
Sarthak Gupta



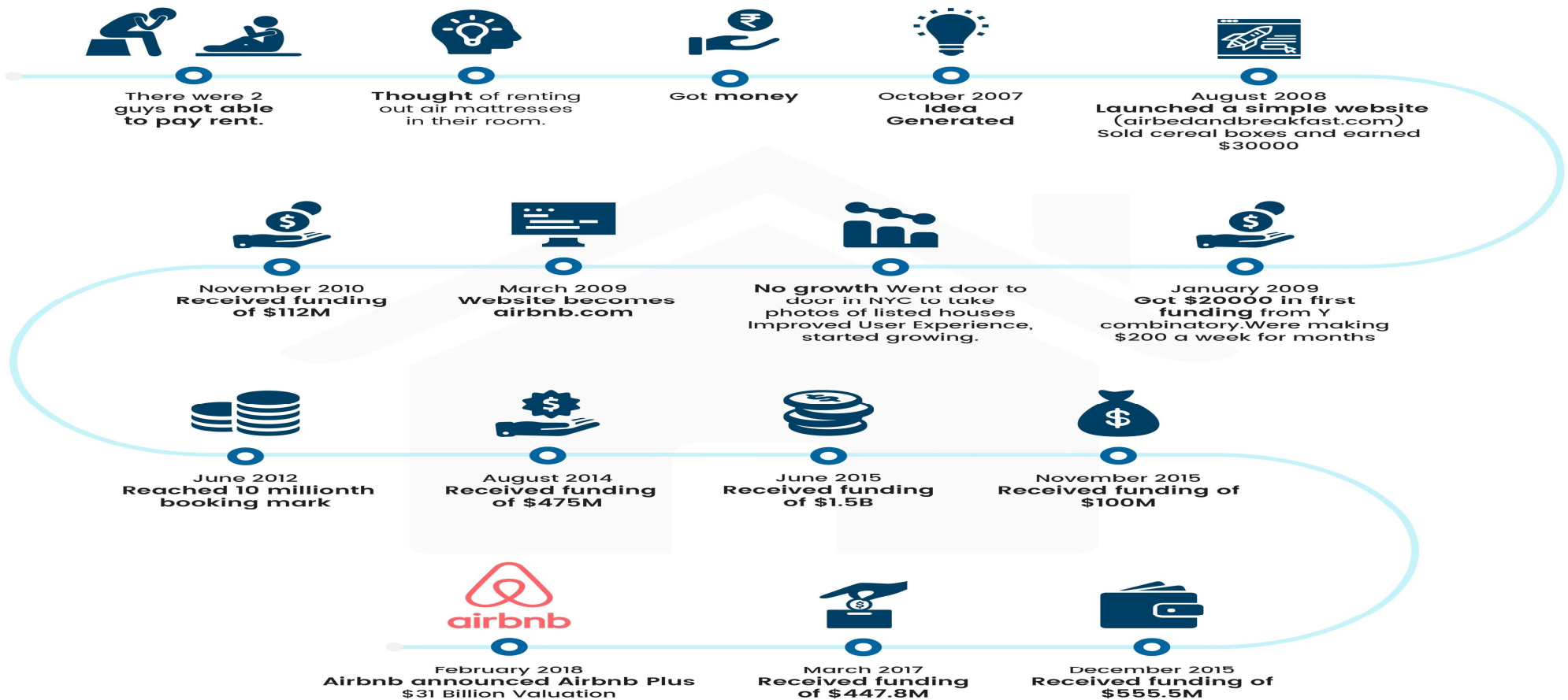
# Air Bed and Breakfast

- **Airbnb is an online rental business platform that stands for the Air, bed, and breakfast.**
- **An online marketplace that involves the renting of property to travelers.**
- **Airbnb does not own any of the properties. It provides a platform on which people can rent out their properties or spare rooms to guests.**
- **Prices are set by the property owners and payments are collected via the Airbnb app.**
- **There are many different types of Airbnbs. You can rent a room in someone's house or a whole island and everything in between.**
- **Airbnb operates in more than 65000 cities and 191 countries.**

# How did it start?



## Airbnb SuccessTimeline



## How Airbnb Makes Money?

**After checking on how Airbnb works, let's get into how it makes money.**

**Mainly, travelers act as the fundamental source here. For them, Airbnb charges 6-12% of the total booking amount. More the size of booking, lesser the expense, and higher is the savings for a traveler. The commission charge acquired is for keeping the room check-in prepared.**

**Without a doubt, the first reason behind Airbnb success is the less expensive methods they offer for hosts. Just a mere 3% booking charge is collected from hosts when the customer checks out.**

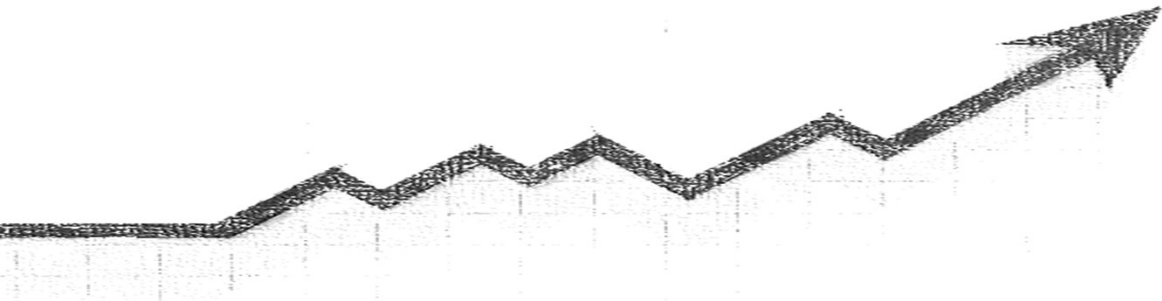
**Majorly, Airbnb generates revenue utilizing the above-mentioned strategies. Also, they utilize and follow other techniques like account arbitrage and corporate travel.**

# Exploratory Data Analysis



## City of interest: New York City

- Dataset provided for Analysis is belong to New York City.
- The City is divided into 5 Boroughs namely **Manhattan, Bronx, Queens, Brooklyn, and Staten Island.**
- These Boroughs are further divided into distinctive neighborhoods.
- Today New York is USA's largest Short-term Rental market, with more than 30K hosts.



## Discussion Topics :

1. How many bookings were made in each borough.
2. Number of Hosts in each Borough of NYC.
3. Types of rooms each borough has to offer.
4. Neighborhood with most bookings, which Hosts has most bookings.
5. Based on room type, an average number of nights the customer must stay.
6. Availability of room in each borough. Top reviewed Borough, Room type.
7. Price variation in each borough, based on the room types.
8. Most, a least expensive neighborhood in NYC, the average cost of a room in NYC.
9. Bird-eye view of the density of bookings, room type, price variation across NYC.
10. Words Host use most to attract the customers for bookings.

## Data Analysis Steps:

### **Import Libraries:**

In this part, we had imported the required libraries to perform Exploratory Data Analysis for the Airbnb dataset.

### **Descriptive Statistics:**

In this part, we start by looking at descriptive statistic parameters for the dataset. We will use `describe()` for this.

### **Missing Value Imputation:**

We will now check for missing values in our dataset. In case there are any missing entries, we will impute them with appropriate values.

### **Graphical Representation:**

We will start with Univariate Analysis. We will be using a **bar graph** for this purpose.

## Attributes of each variable:

- **id**: Unique for each listing available in the dataset. Total 48895 unique id's are available in data.
- **name**: Name of the listings available in the dataset. Total 47905 named hotels available in NYC.
- **host\_id**: Unique for each host who has listed rooms for rent on Airbnb. Total 37457 hosts with their id's.
- **host\_name**: Name of Host. Total unique 11452 host\_names in data set.
- **neighbourhood\_group**: Names of each borough in NYC city. In data all available hotel locations *Manhattan, Brooklyn, Queens, Bronx, State Island*.
- **neighborhood**: Locale area in each borough of NYC. Total 221 neighborhoods.
- **latitude, longitude**: Co-ordinated for each listing in the dataset. These are in between 40.72 and -73.95.
- **room\_type**: Each type of room available for listing. In all areas *Private rooms, Entire home/apt, Shared room types are available*.
- **price**: Price of each listing in the dataset. Price between 0 to 10000.
- **minimum\_night**: Mandatory number of nights room needs to be booked
- **number\_of\_review**: Number of reviews for each listed property. In 0 reviews and max 629 reviews.
- **last\_review**: Last date the listing was reviewed.
- **review\_per\_month**: Number of reviews received per month.
- **Calculated\_host\_listing\_counts**: Number of Listings each host owns.
- **Availability\_365**: Number of days in the year the listing is available for booking. Availability zero to 365 days.



## Sorting variable types:

### **Categorical Variables:**

name, host\_name, neighbourhood\_group, neighbour, room\_type.

### **Numeric Variables :**

id, latitude, longitude, price, minimum\_nights, number\_of\_reviews, last\_review, review\_per\_month, calculated\_host\_listing\_counts, availability\_365.

## Insights on variables:

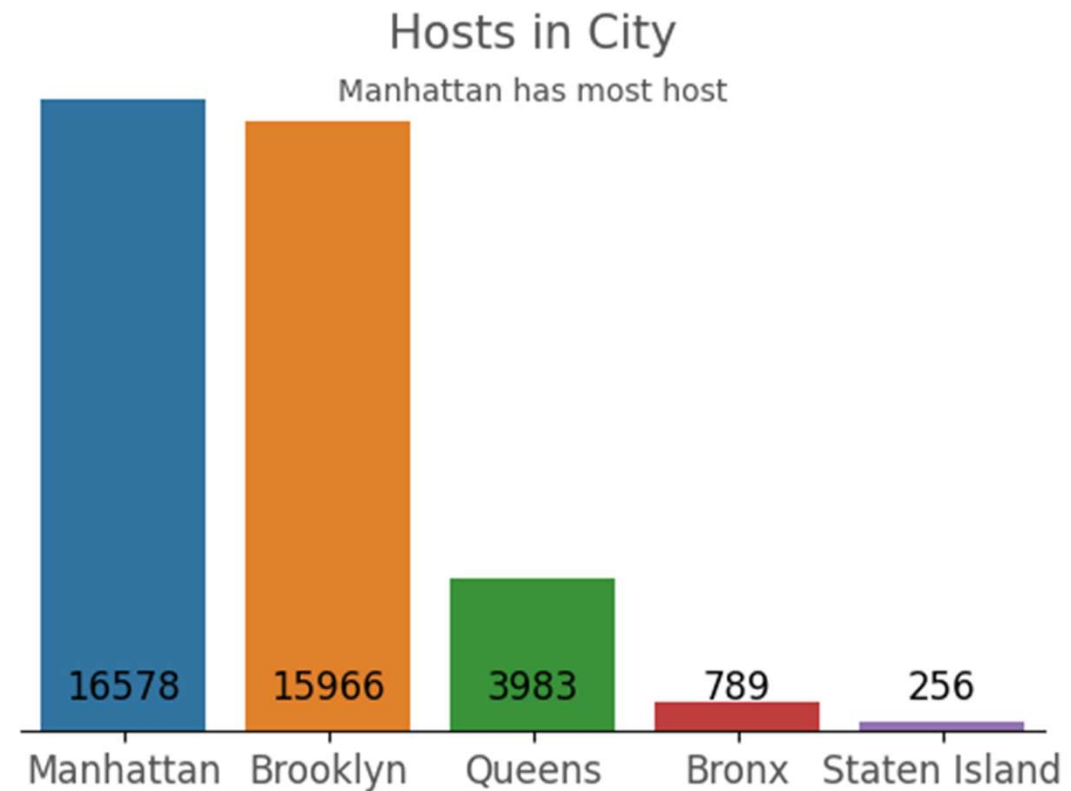
- Each host\_id could have multiple listings in the neighborhood, but each listing can have one and only one host\_id.

- *Begin with an examination of the host\_id*



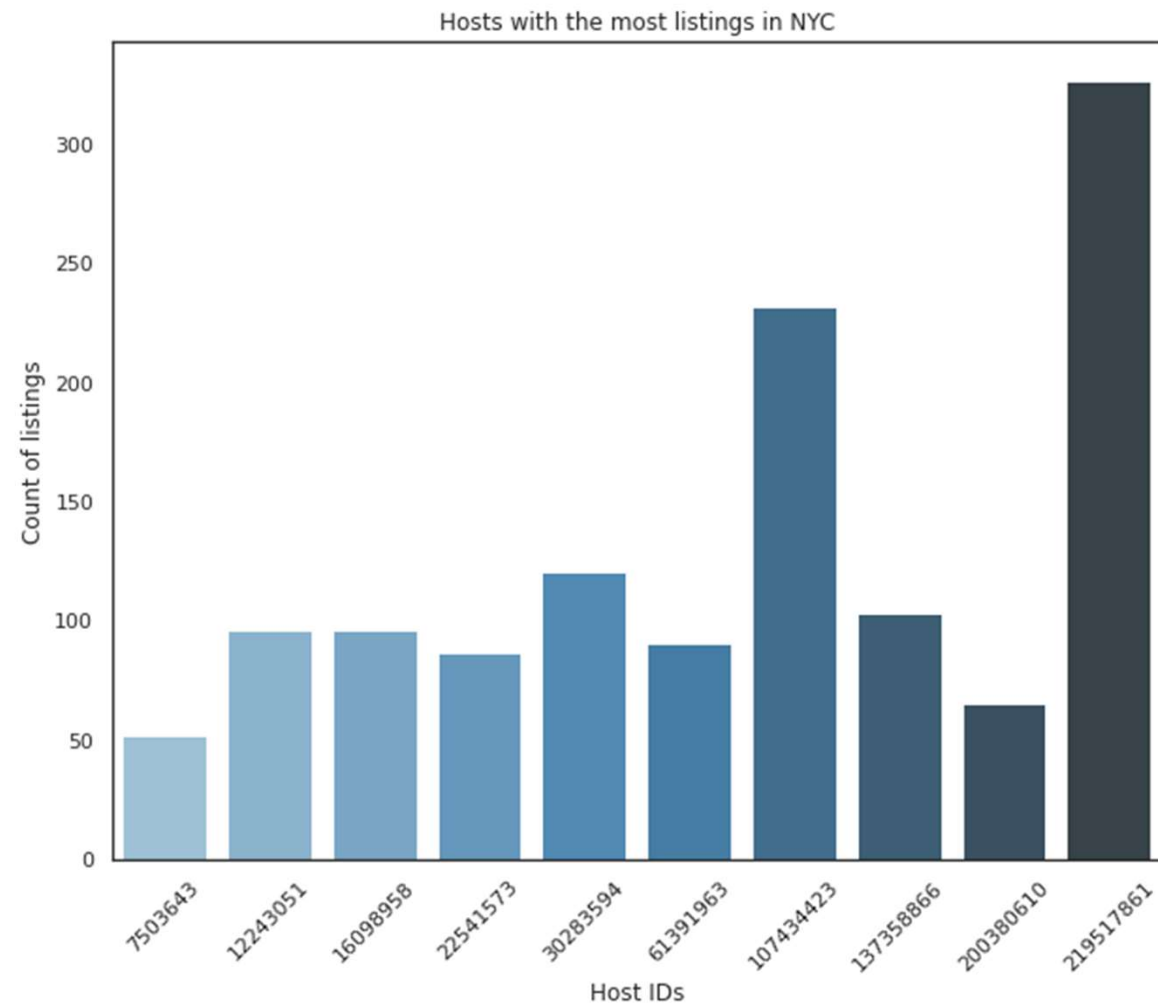
## Number of Hosts in each Borough of NYC

- **Manhattan** has the most hosts, 16578.
- Brooklyn is the second-highest with 15996 bookings.
- Most customer stays explains the high number of hosts Manhattan has.
- Most customer stays are directly related to the number of hosts each borough has.



## Top ten hosts of the NYC city

- Top ten hosts with the most listing have a good dispersion, which is interesting to see.
- Host with host\_id 219517861 has hosted most customers **327**.
- Host, with host\_id 107434423 is the second-highest with a total of 232 customers hosted.
- Host, with host\_id 30283594 has hosted 121 customers.

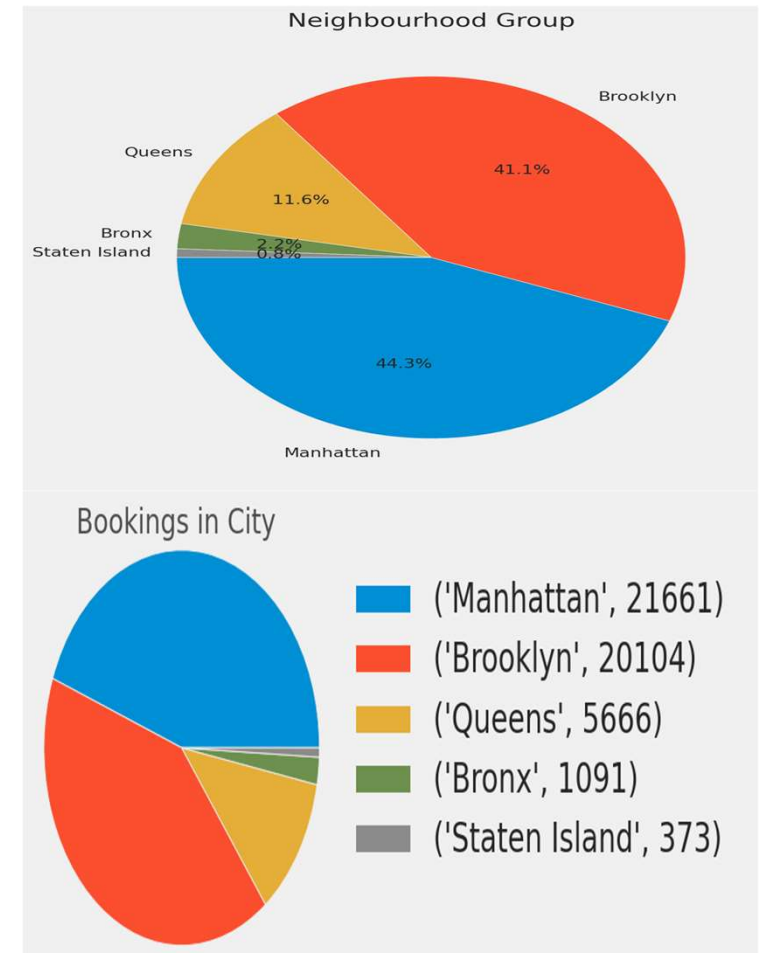


## *Examination of Neighbourhood\_group*



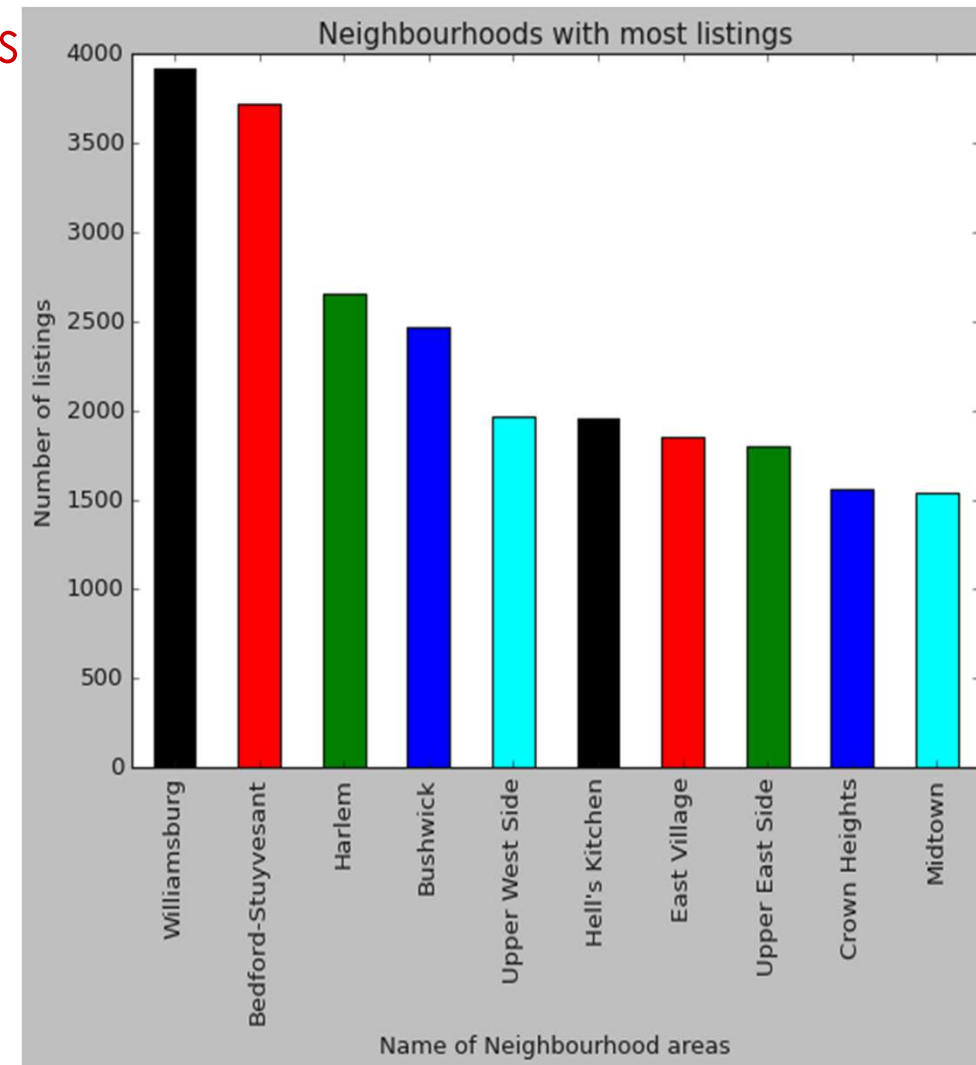
How many bookings were made in each borough.

- **Manhattan** has the most bookings with 21661.
- **Brooklyn** has the second-highest with 20104 bookings, **Queens** has the third highest bookings with 5666.
- **The Bronx** and **Staten Island** trail the pack with 1091 and 373 bookings.



## Top Neighborhoods with most bookings

- There are 221 unique neighborhoods in the given dataset. Furthermore, top ten are mentioned in this bar chart.
- **Williamsburg** is the top neighborhood with 3920 bookings.
- **Bedford-Stuyvesant** has the second most bookings with 3714 bookings.
- **Harlem** has the third most bookings with 2658 bookings.
- Rest count has mentioned in the chart.



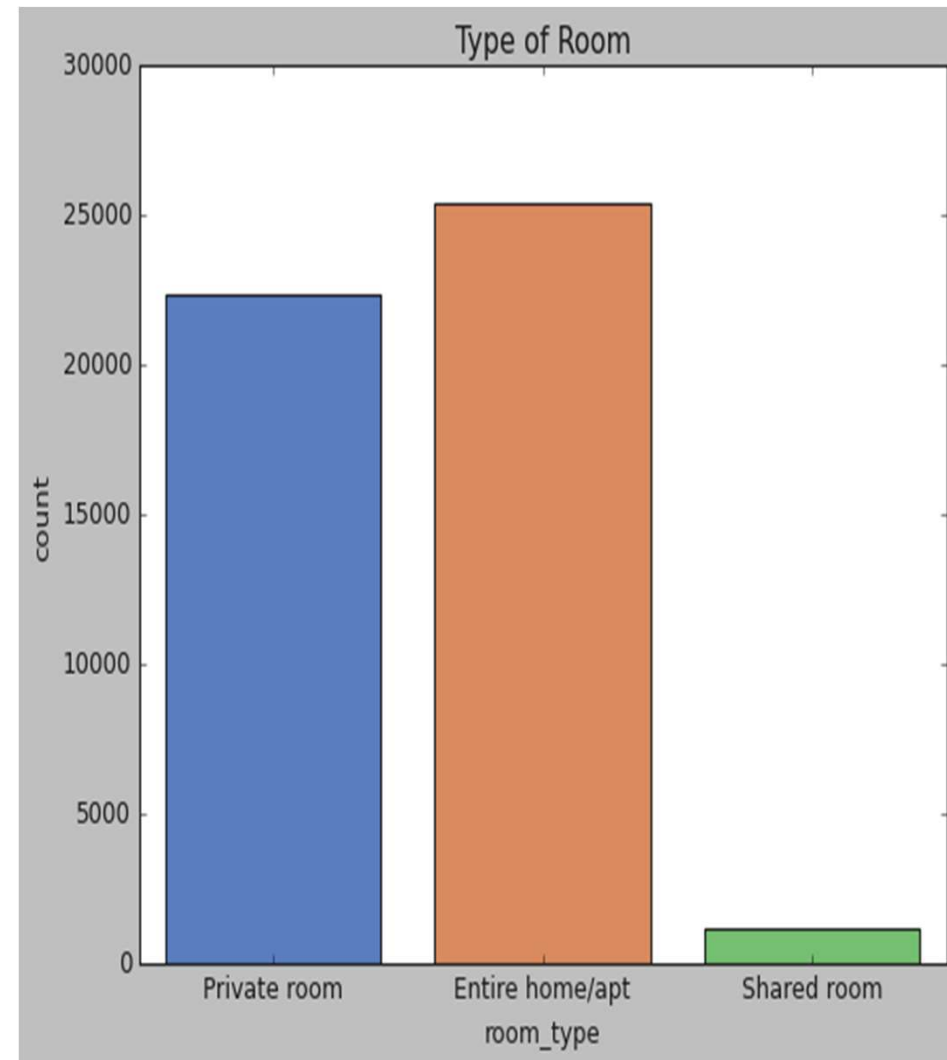
## *Types of room*





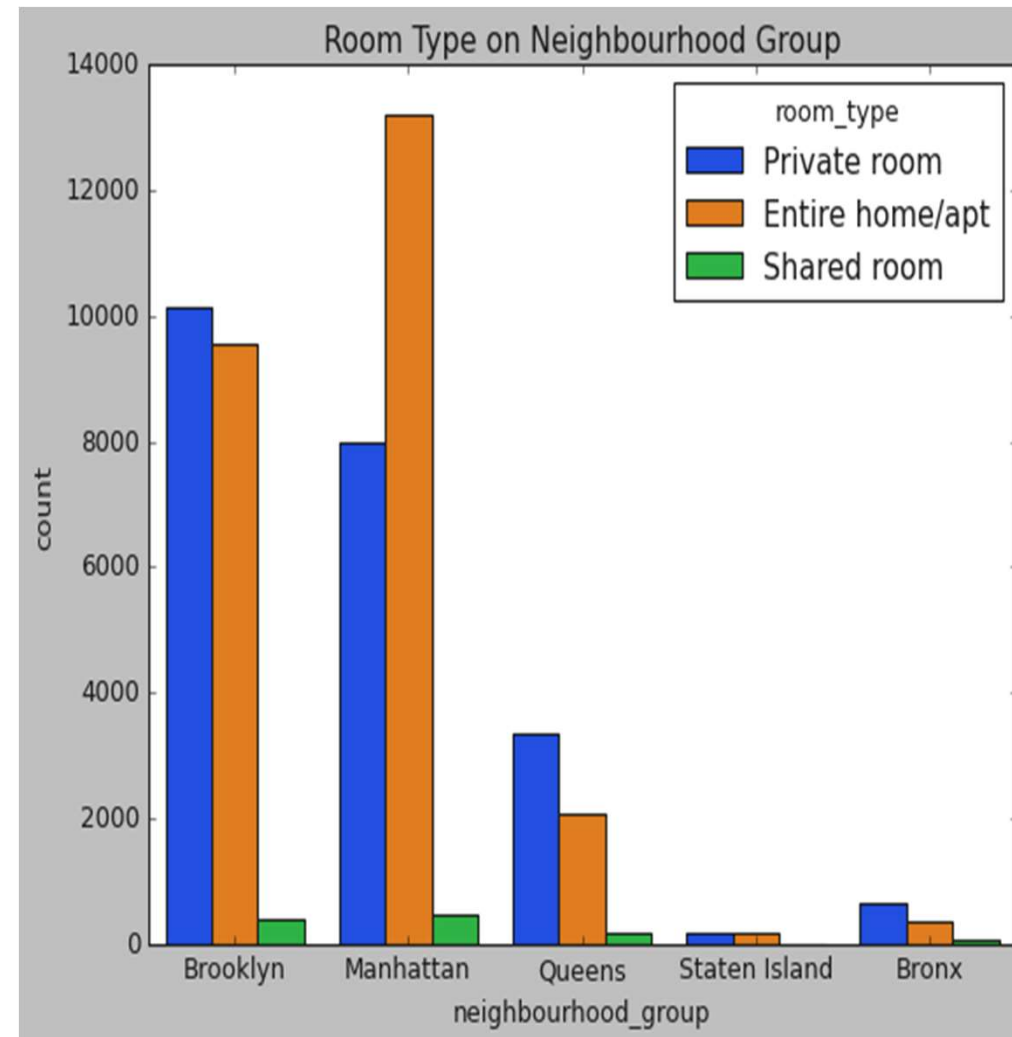
## Types of room

- **Private room**
- **Entire home/apt room\_type**
- **Shared room**
- According to the findings, the **entire house** is preferred over a **private room**. Furthermore, in New York City, **shared rooms** are rarely preferred.



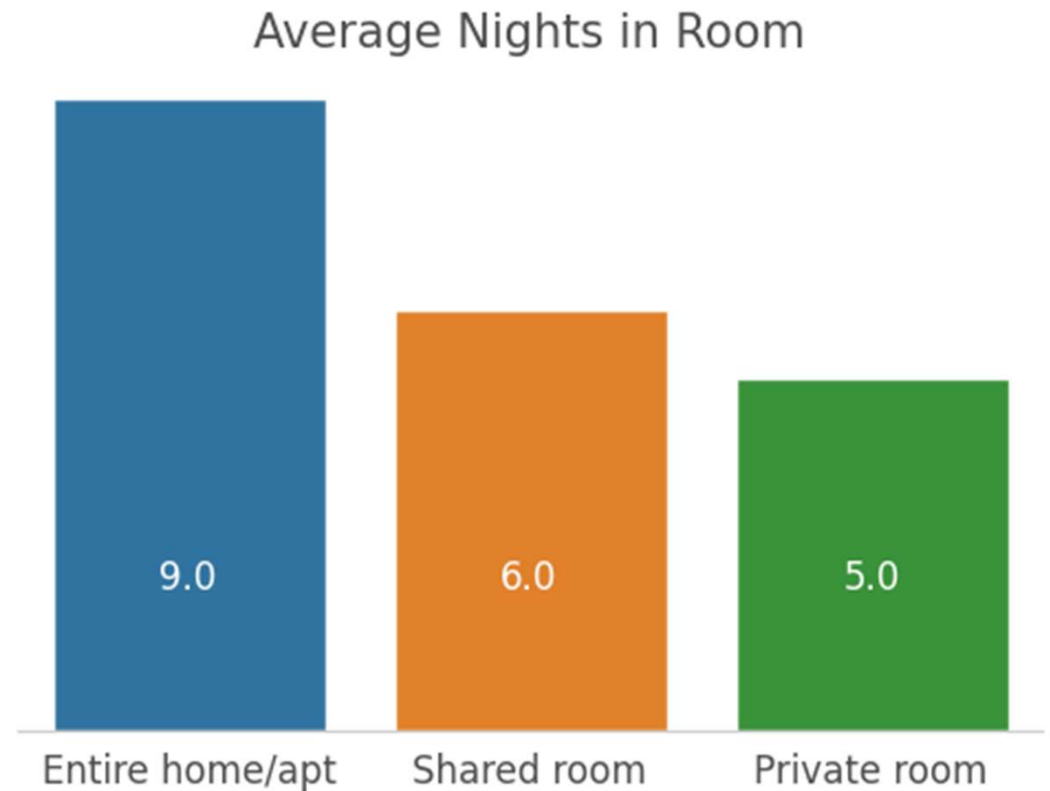
## Types of rooms each borough has to offer

- **Manhattan** leads the group with the most rooms bookings, Customer prefer **Entire homes** more compared to Private rooms in Manhattan.
- Customer in Brooklyn prefer **Private rooms** more than **Entire rooms**.
- Except in Manhattan all the other boroughs have more bookings for a private room.
- Shared rooms are the least preferred type across all the boroughs of NYC.



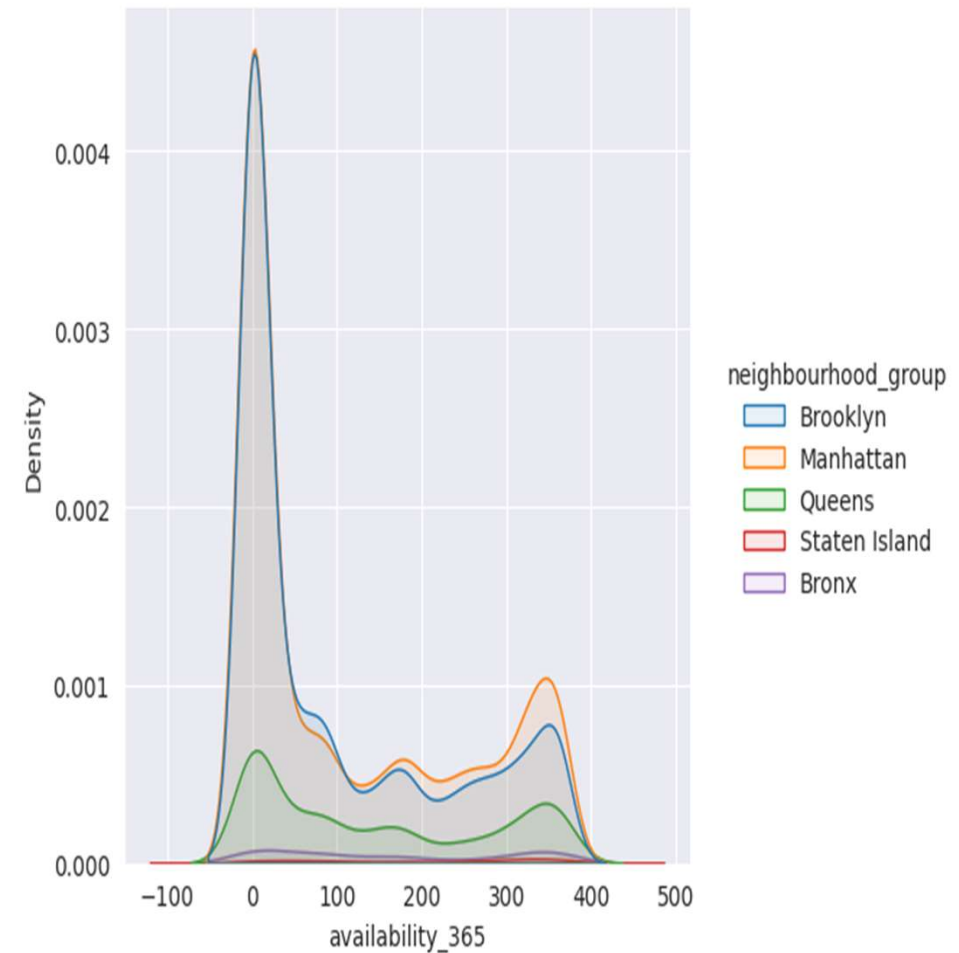
## Average of Minimum number of Nights for each room type

- Average 9 nights are mandatory for stay for booking of Entire home
- Shared room is the second highest with average 6 nights stays mandatory
- Private room is third place with average minimum 5 nights.



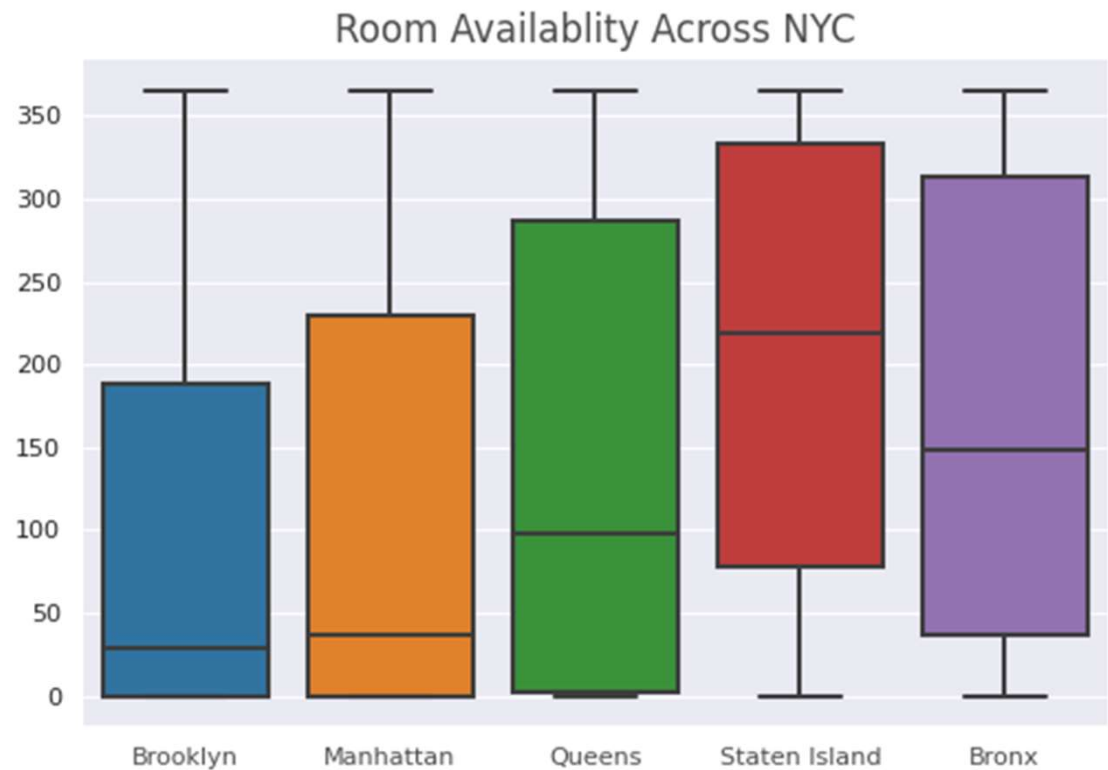
## Availability distribution across New York City

- Manhattan and Brooklyn have very similar distribution with peaks close to zero.
- This shows most of the rooms are mostly booked in Manhattan and Brooklyn.
- Availability of rooms is more evenly distributed compared to Brooklyn and Manhattan.

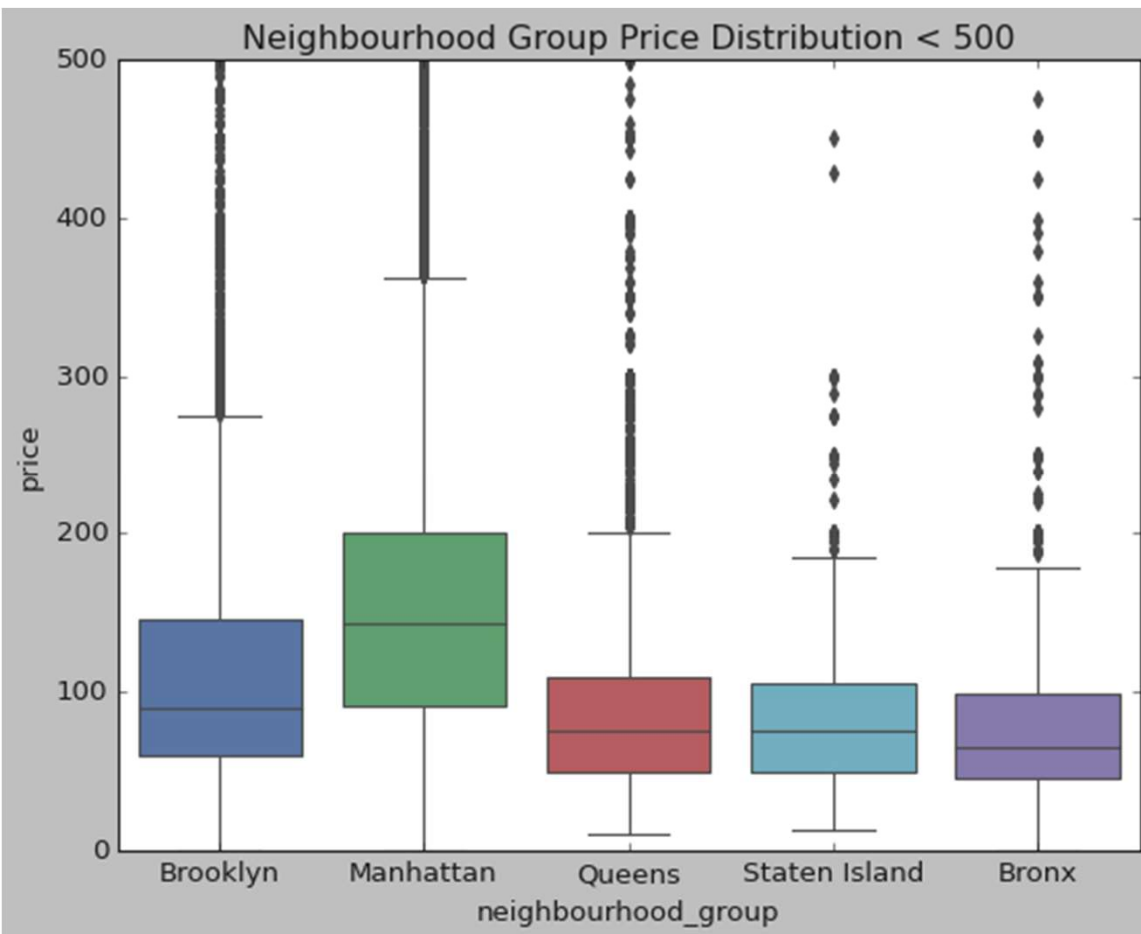


## Availability of each room in Borough

- Brooklyn has the least availability followed by Manhattan-based on the distribution of room availability.
- Staten island has the highest availability of rooms.
- Availability of rooms in the Bronx is more evenly distributed compared to other boroughs.

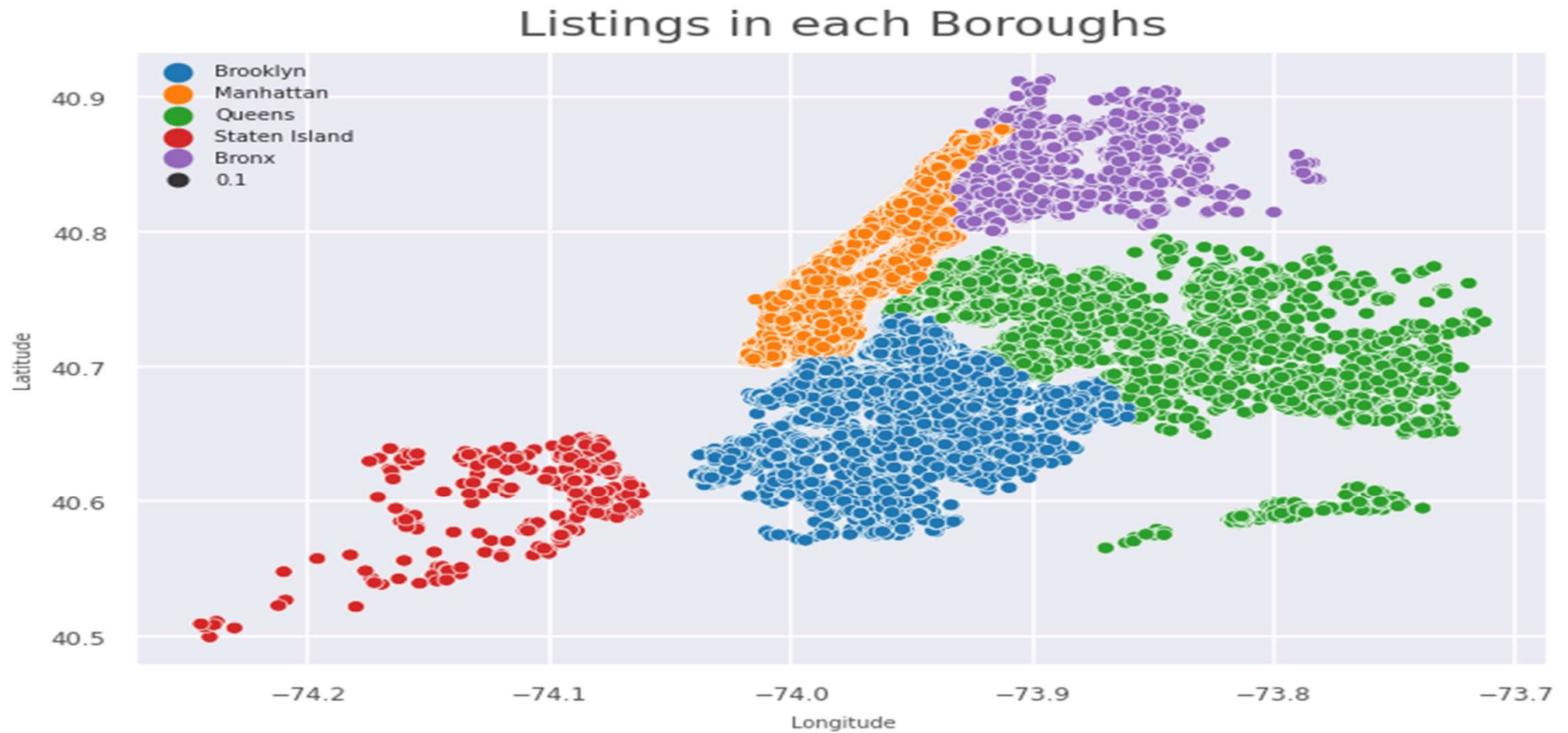


## Price distribution across New York City

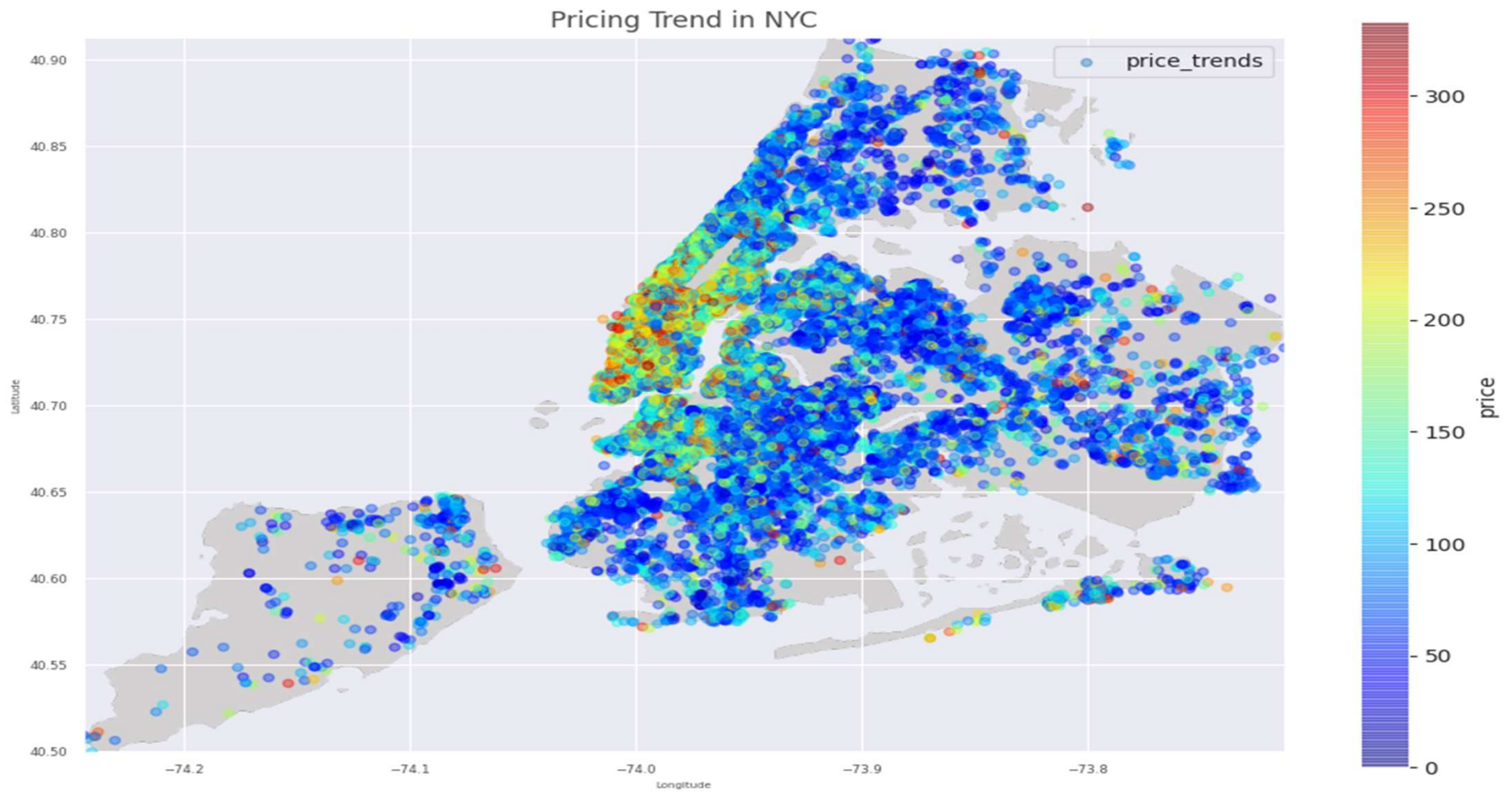


- **Manhattan** is the **most expensive** as the rental charges are more evenly distributed across all the **price ranges**.
  - Median price in **Manhattan** is approx. \$150 that's around double the Median price of **Bronx**.
  - The distributions in **Queens** and **Staten Island** appear to be very similar, while the Bronx appears to be the cheapest of the three.
- Note-** For instance, given that Manhattan is unquestionably one of the most expensive cities in the world to live in, and the Bronx appears to have lower living standards and the cheapest. Lastly, this price distribution and density were entirely predictable.

# Visualizing variables with Longitude and Latitude



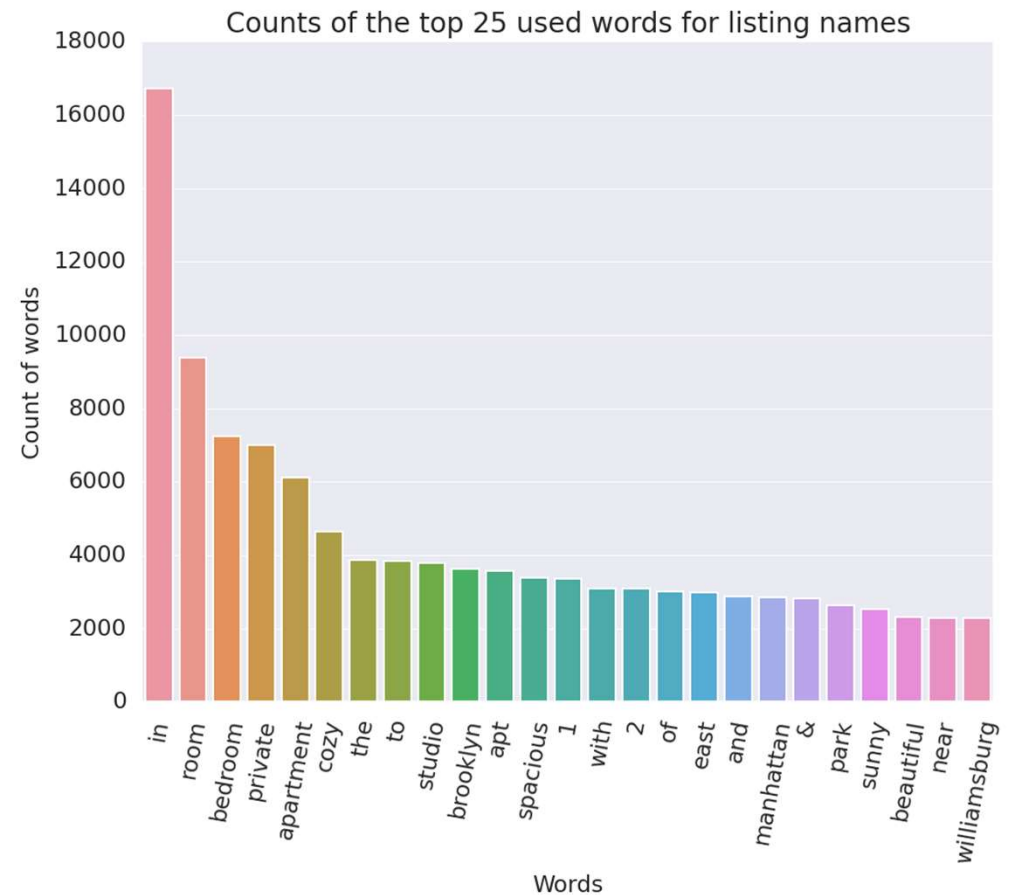
# Price distribution New York city





## Bar graph of top 25 words for listing the names

- Words like in, room, private, apartment, cozy are frequently used in the name of Airbnb listings so that a customer can find it easily as per their requirement.



## Listing names in each Neighborhood



- 
- Heatmap showing the correlation matrix for the features: id, host\_id, latitude, longitude, price, minimum\_nights, number\_of\_reviews, reviews\_per\_month, calculated\_host\_listings\_count, and availability\_365. The color scale ranges from -0.2 (dark purple) to 1.0 (yellow).



# Observations from Airbnb analysis

## Summarizing our findings, suggesting other features

This Airbnb ('AB\_NYC\_2019') dataset for the 2019 year appeared to be a very rich dataset with a variety of columns that allowed us to dive deep into each significant column presented.

To begin, firstly, we identified the data of top ten **host\_id** and we figured out that top host ID has **327** listings.

Secondly, we take "**Neighbourhood\_Group**", and we found that Airbnb listings in New York City are concentrated in **five** neighborhoods: "**Brooklyn**," "**Manhattan**," "**Queens**," "**Staten Island**," and "**Bronx**". Moreover, we also learned from this chart that "Manhattan" and "Brooklyn" have the most hotel properties. Then, we found that **Manhattan** is the **most expensive** as the rental charges are more evenly distributed across all the **price ranges**, median price in **Manhattan** is approx **\$150** that's around double the median price of **Bronx** and the distributions in Queens and Staten Island appear to be very similar, while the **Bronx** appears to be the **cheapest** of the three.

Thirdly, we take the data of "**room\_type**" and figured out that it is divided into three subcategories and we can observe that the Entire **Home/Apartment** has the **highest** share, followed by the **Private Room**, and the **least** preferred is **Shared Room**. Furthermore, entire Home/Apartment is listed most near **Manhattan**, while **Private Rooms** and Apartments Near **Brooklyn** are **Nearly equal**.

Fourthly, we put our **latitude** and **longitude** columns to good use by creating a **geographical map** of New York city which represents the location of all the areas with their latitude and longitude. In other map is **Color-coded** for **listing price of room as per the location**.

In addition, we returned to the **first column "name"** and found out the words from the hotel names, as well as the **count** for the most **frequently used words by hosts**. Hosts prefer to use **Private rooms, brooklyn, central park, modern, nyc** and **Beautiful** these words in their listing to seek customer attention.

Finally, we looked for the listings with the "**most reviews**". Count the rating of **top ten reviewed hotels**, and found out The top 10 most reviewed listings on Airbnb for NYC have an **average price of \$65 per night**, with the majority of them under \$50, and **9/10 of them are "Private Room" types**, with the top reviewed listing having **629** reviews.

# Conclusion

Customer Needs Rental space with **Cheaper rent, good reviews, lower minimum nights, proximity to transportation services, good space.**

Based on the Word plot, the list of Words that are used to seek Customer attention are:

***Private room, Brooklyn, Manhattan, Central Park, beautiful, spacious, subway, Williamsburg, Bushwick, times square***

## **EDA Validation:**

- Private room has lowest minimum nights avg, and cheaper compared to Entire room, and its prices are very similar to Share rooms when charges are compared on per night basis.
- Private rooms are the most reviewed across NYC
- Brooklyn has Cheaper per night charges for private rooms compared to other boroughs, 3 of the most booked neighborhoods are from Brooklyn, Williamsburg, Bushwick, Bedford-Stuyvesant.
- Manhattan is the hub of the center of activities like Central park, times square and has the busiest subway,
- Entire home provides more space, Entire home Charges in Manhattan are very close to Private room.

# THANK YOU

