CAPSTONE PROJECT - 1

EDA on AirBnb Bookings

airbnb

Presented by : data_hacker

Diptiranjan Bal
Rudra Prasad Sahoo
Pranay Bhuyan

About AirBnb :

AirBnb Inc. is an American San Francisco based company founded in 2008 by Brian Chesky, Nathan Blecharczyk and Joe Gebbia, operating on online Marketplace for short-term homestays and experiences. It is a worldwide recognized company for its user-friendly services. Guests and Hosts have used Airbnb to expand on travelling possibility and rental homes to present an unique and personalized way of experiencing the World.

Problem statement :

To perform Exploratory Data Analysis (EDA) on the AirBnb Bookings Analysis in order to understand the bookings of various properties, their hosts, customer interest, Revenue generated etc. and to help AirBnb to know its weak points and to improve those keys for the future growth of the Company.

Data Summary :

The dataset consists of observations of bookings done in New york city (NYC) till 2019.

Shape of the dataset : (48895, 16)

Columns Present in the Dataset :

1. Id
2. name
3. host_id
4. host_name
5. neighbourhood_group
6. neighbourhood
7. latitude
8. longitude
9. room_type
10. price
11. minimum_nights
12. number_of _reviews
13. last_review
14. reviews _per_month
15. calculated_host_listings_count
16. availability_365

Agenda (Page 01) :

1. Different Host and Areas they host.
2. Busiest Host and why ?
3. Areas Top 5 busiest hosts are hosting.
4. Which host gives AirBnb the most revenue ?
5. Most number of reviews given by people of different neighbourhood groups.
6. Which are the most preferred room types with respect to Neighbourhood Group ?
7. Revenue generated by different room type(Entire home/apt, Private room, Shared room) in each Neighbourhood Group.
8. Density of different types of rooms across the map.
9. Number of reviews with respect to room type.
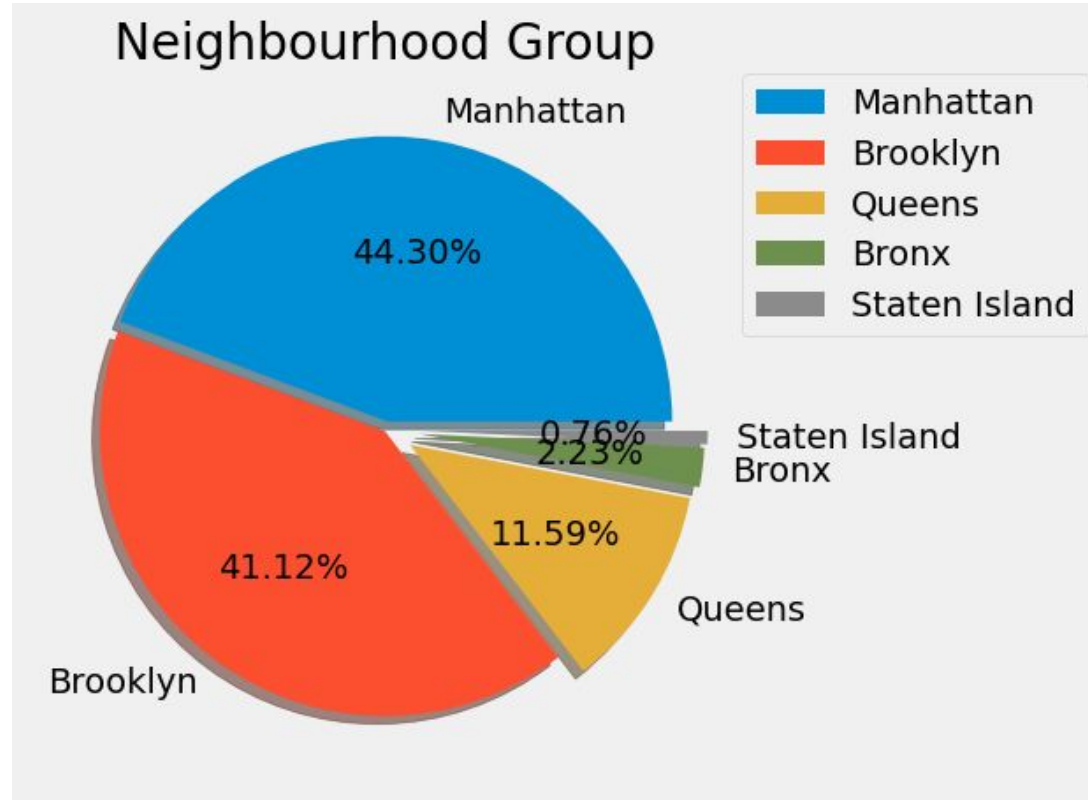
Agenda (Page 02) :

10. How price influence people ?
11. Observing bookings in different Neighbourhood Groups by using the price we got from Quartiles from Univariate analysis of Price.
12. Revenue generated by the room category(Cheap, affordable, expensive, luxurious).
13. Locations of different categories of rooms(Cheap, affordable, expensive, luxurious) over the map.
14. Noticeable difference of Traffic among different areas.
15. Percentage of the times people reviewed a hotel/apt in different Neighbourhood Group.
16. Average availability of nights for different "Neighbourhood Group".
17. Correlation between different columns present in our data set.

Map of New york city :

Bookings in Neighbourhood Group :

- **Manhattan** and **Brooklyn** have the First and Second highest hotel/room bookings among the Neighbourhood Group respectively.
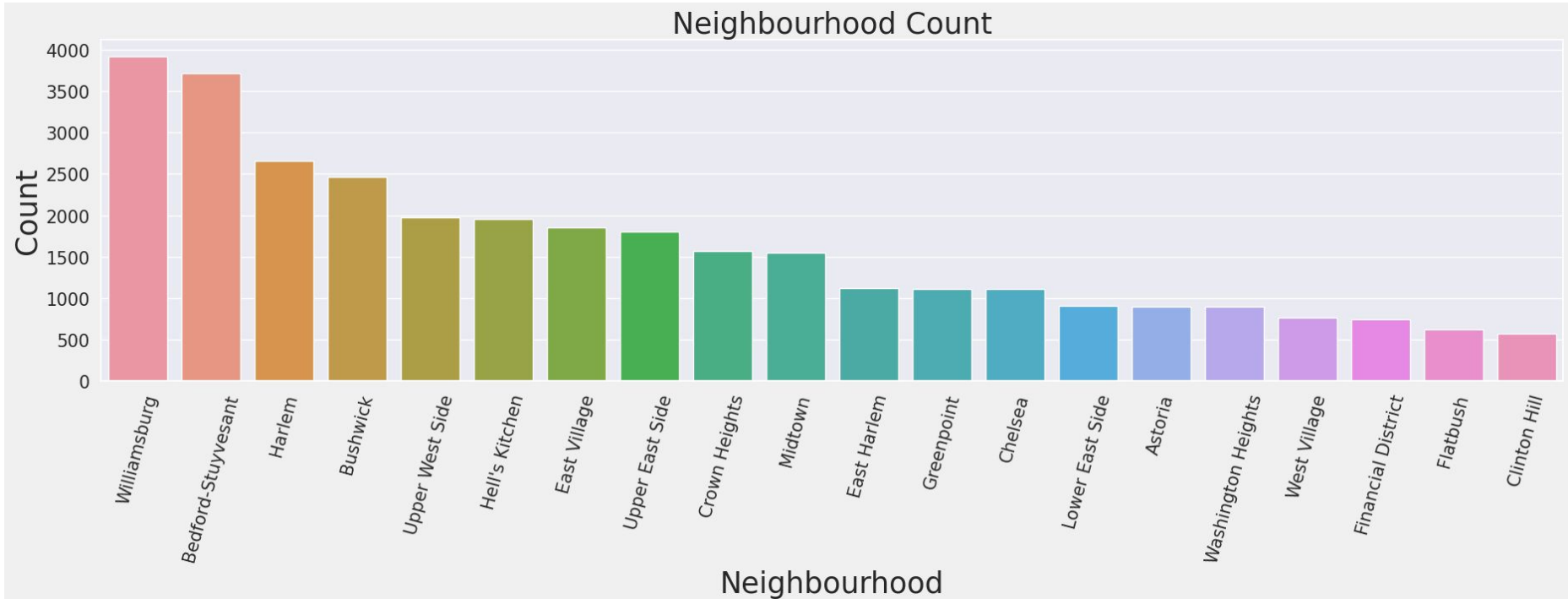

Neighbourhood Group

- Manhattan
- Brooklyn
- Queens
- Bronx
- Staten Island

Manhattan 44.30%
Brooklyn 41.12%
Queens 11.59%
Bronx 2.23%
Staten Island 0.76%

Count of Different Types of Room :

● In the New York city the Entire Home / Apt is the most prevailing room type.

Room Count
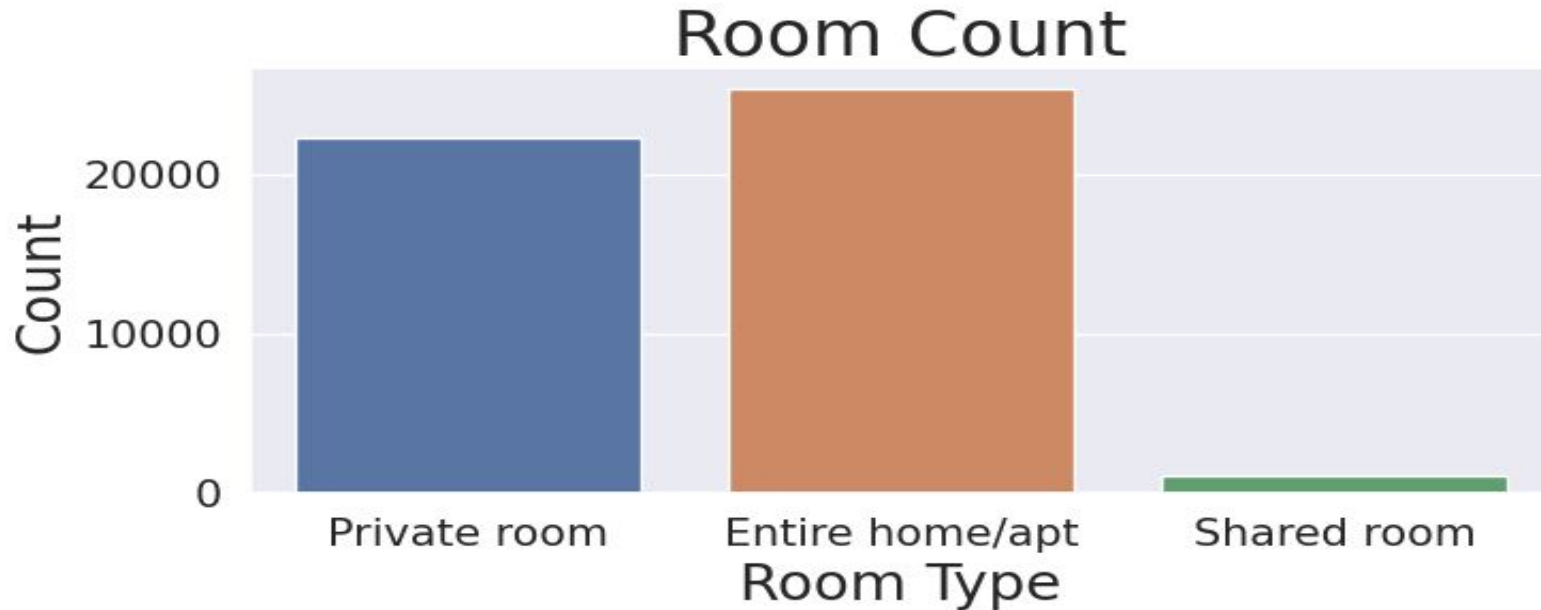
Most Common Neighbourhood :

- **Williamsburg and Bedford-Stuyvesant** are the two that are prefered comparatively high.
- Commonly People prefer to stay at these neighbourhoods.


Neighbourhood Count

Different Host and Areas they host :

- Most number of listings are created by the host named **Sonder (NYC)** from **Manhattan** neighbourhood group followed by the host named **Blueground** from both **Manhattan** and **Brooklyn** neighbourhood groups.
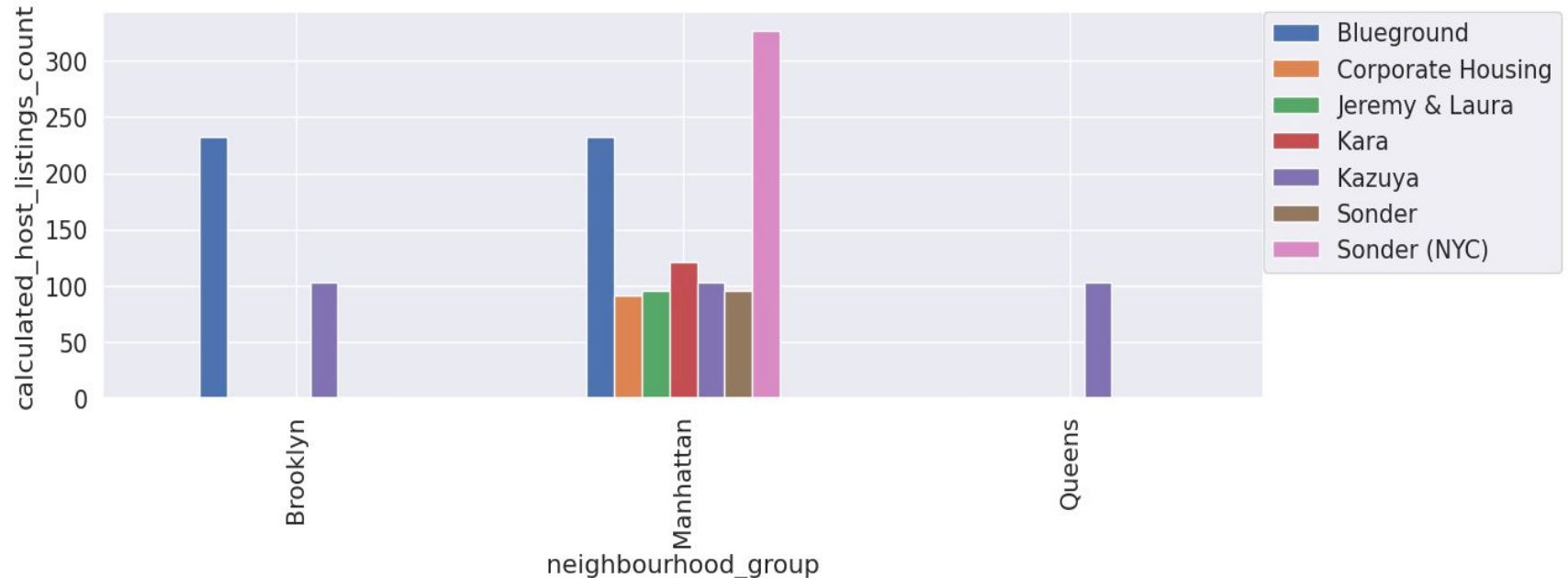
Busiest Host and why ?

- **Top 15 busiest hosts** on AirBnb NYC on the basis of the **number of reviews**.
- **Michael** is the busiest host with the max number of reviews, followed by **David, John, Jason, Alex.**
- It also has been observed that most of the busiest hosts **deals** with **Private rooms and Entire Home/apt**.

Busiest Hosts
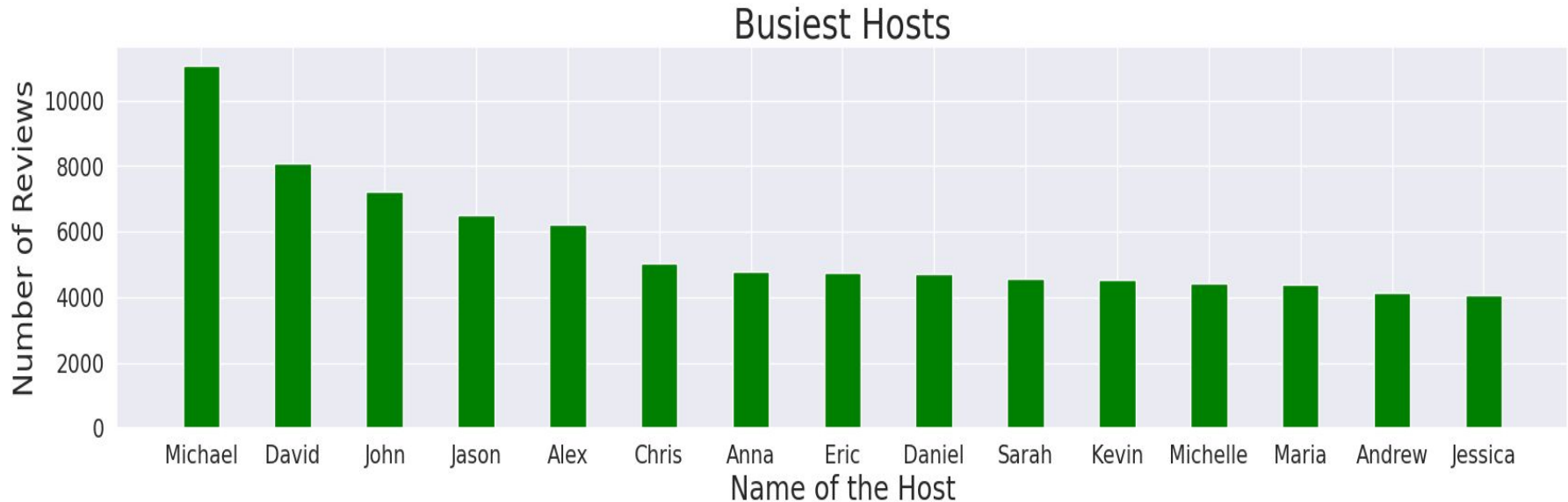
Density of different types of rooms across the map :

- **Private rooms** and **Entire home/apartments** room types are more densely spread across the map of New york city.

Room type location per Neighbourhood Group
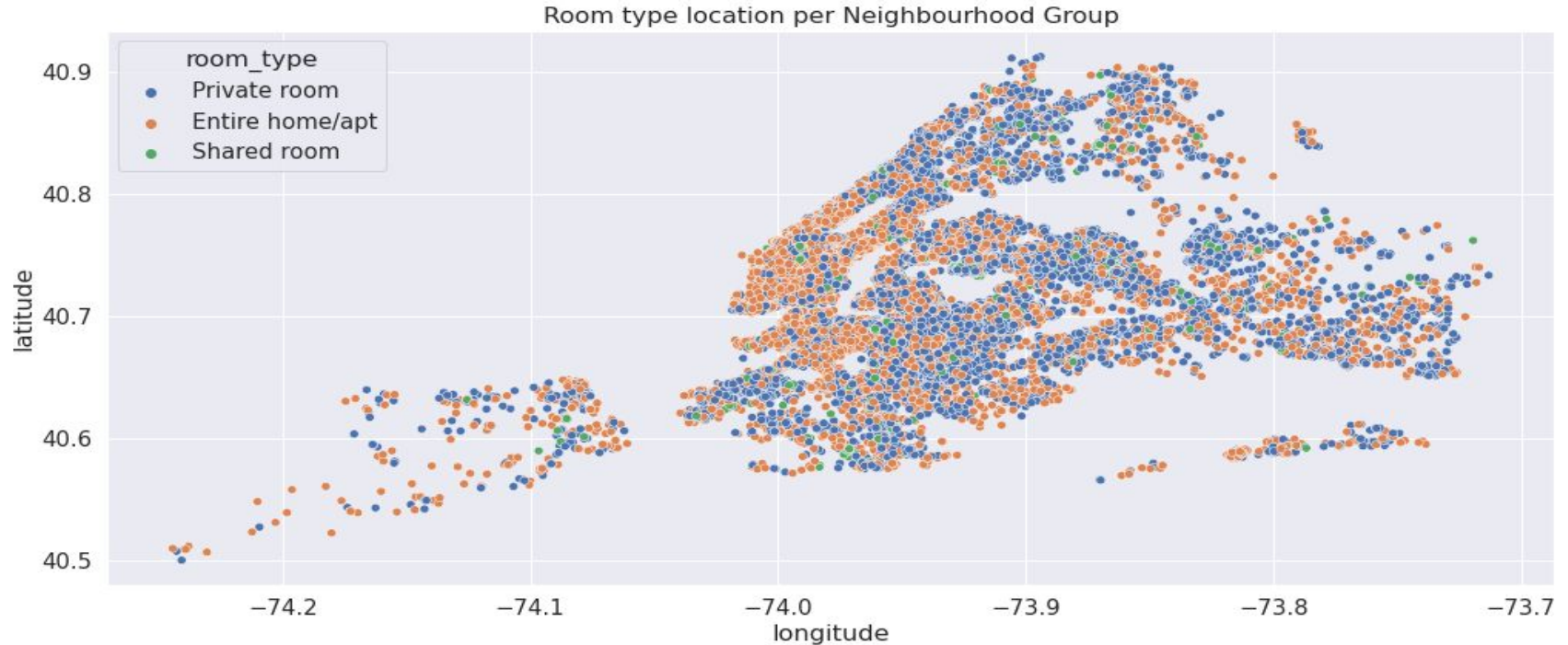
Which host gives AirBnb the most revenue ? :

- We can see that the host named **Blueground** is leading in the list with most revenue generated around $ 2.2 million followed by **Jenny, Kara, Amy, Katherine** and others.


Revenue Generated by Hosts
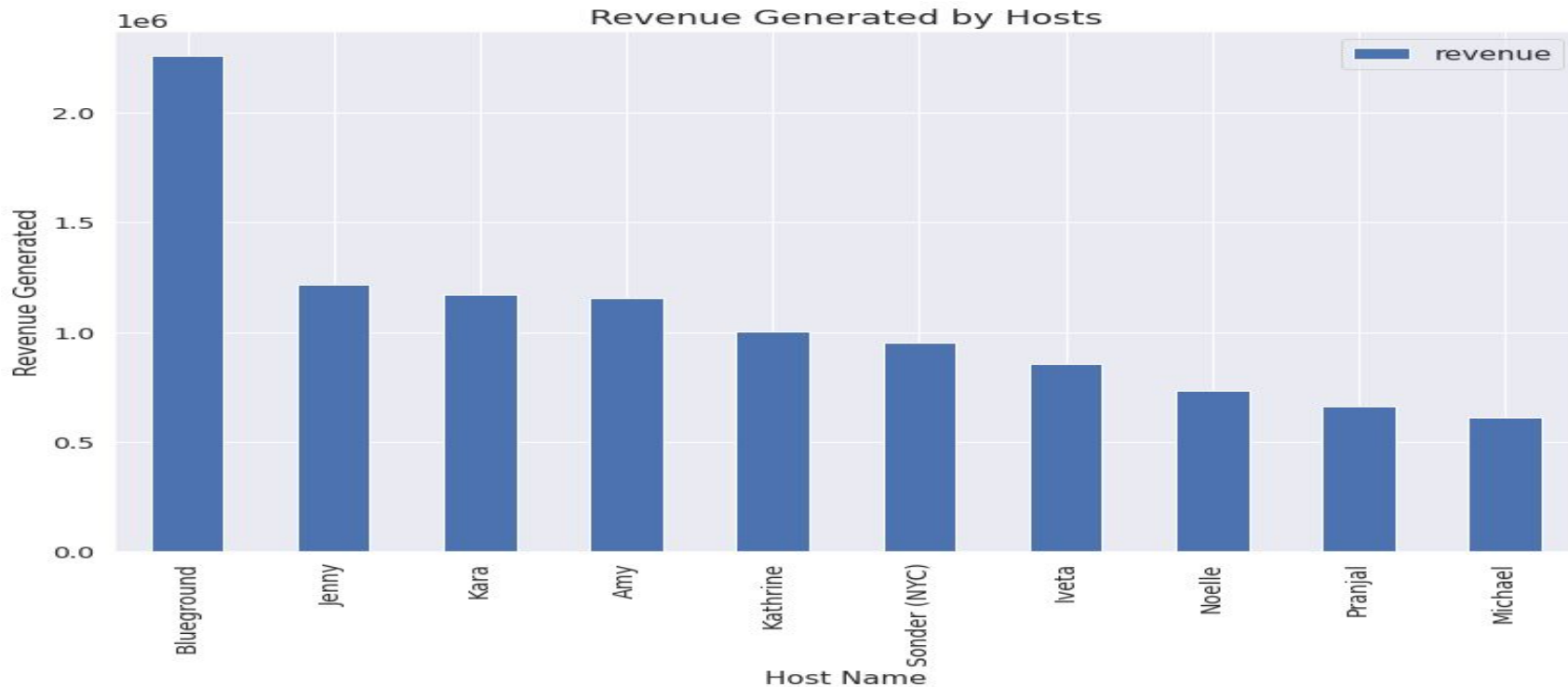
Most number of reviews given by people of different neighbourhood groups

- **Brooklyn and Manhattan** are the neighbourhood groups that gets the most reviews.
- **Queens, Bronx and Staten island** gets less reviews comparatively.

Maximum Number of reviews with respect to Neighbourhood Group
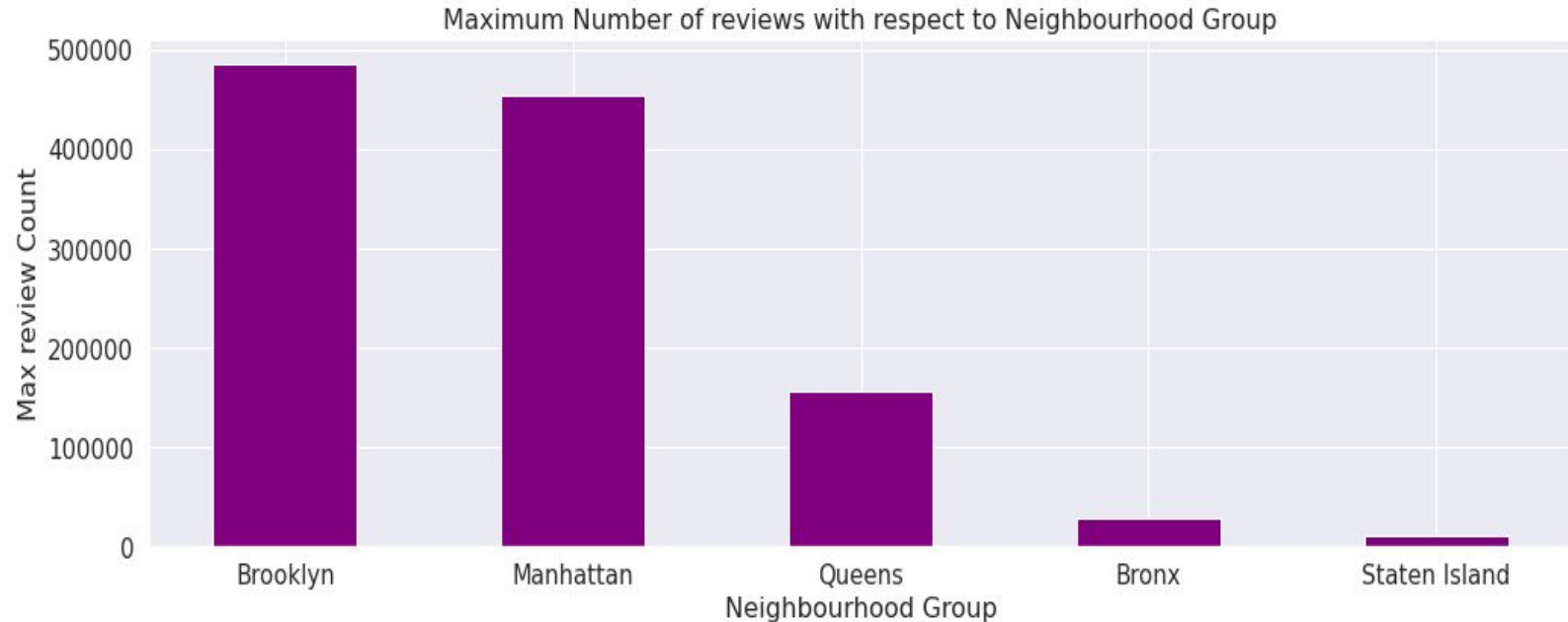
Which are the most preferred room types with respect to Neighbourhood Group ?

- **Manhattan** is the neighbourhood group where the e**ntire home/apt** has been booked mostly.
- As for **Brooklyn**, **Private rooms** have been preferred the most.
- In **Staten Island**, people like to live in **Entire home/apt** and **Private room**. But people barely prefer **Shared room.**
- Concluded that most of the people prefer **Entire home/apt** and **Private rooms** compared to **shared room**.

Most prefered room type with respect to Neighbourhood Group
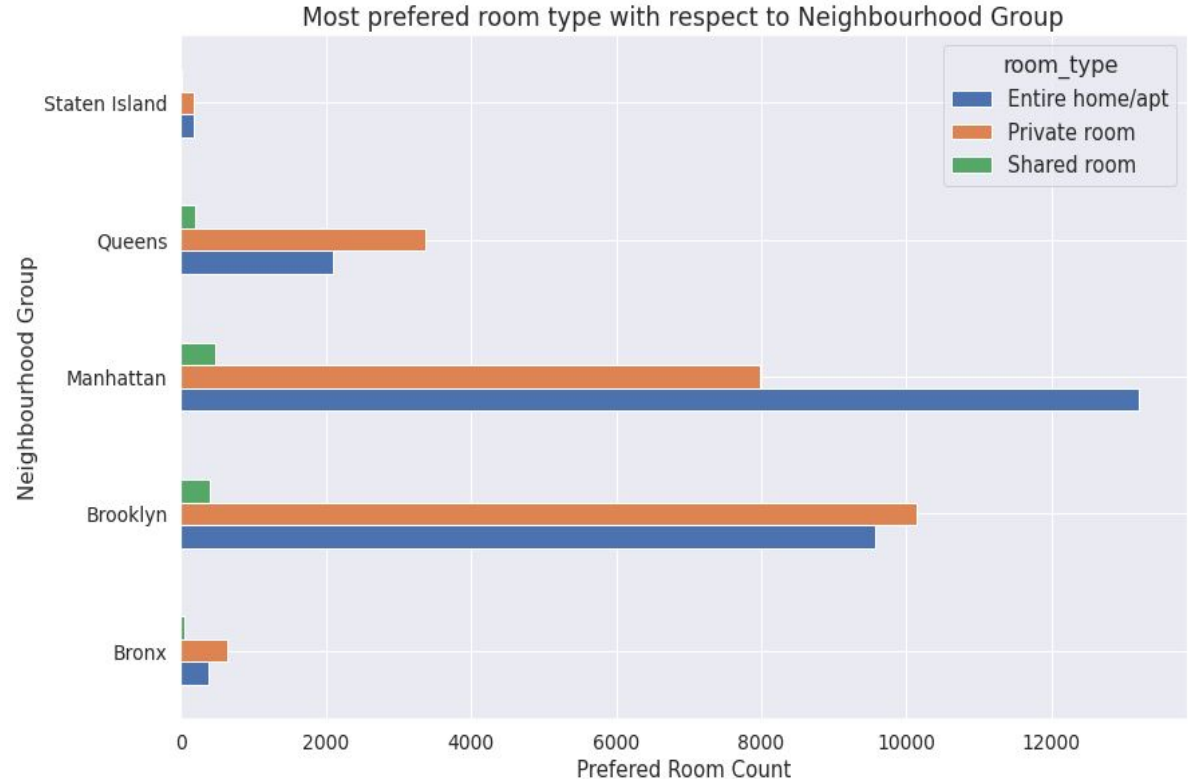
Number of reviews with respect to room type :

- This **pie chart** represents the percentage of **number of reviews** w.r.t the **room type**.
- The majority of the reviews are given for the **Entire Home/Apt** type of rooms with 51 %, while 47.3% of the reviews are given for **Private rooms**.
- **Shared rooms** get the least reviews with only 1.7% of the total reviews.

Percentage and Number of Reviews with respect to Room Type

Entire home/apt

51.0%
(580403)

1.7%
(19256)

Shared room

47.3%
(538346)

Private room
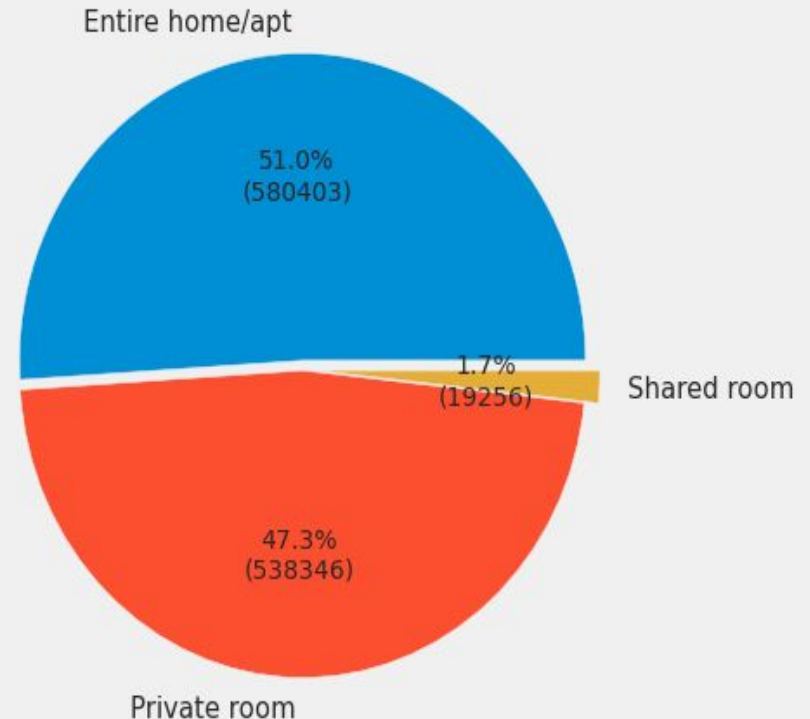
How price influence people ?

- We have analysed that price is one of the most important factor in influencing people to go for a booking and the review after that.
- Most of the people have preferred a low price room i.e below 1000$.

Number of Reviews with respect to Price
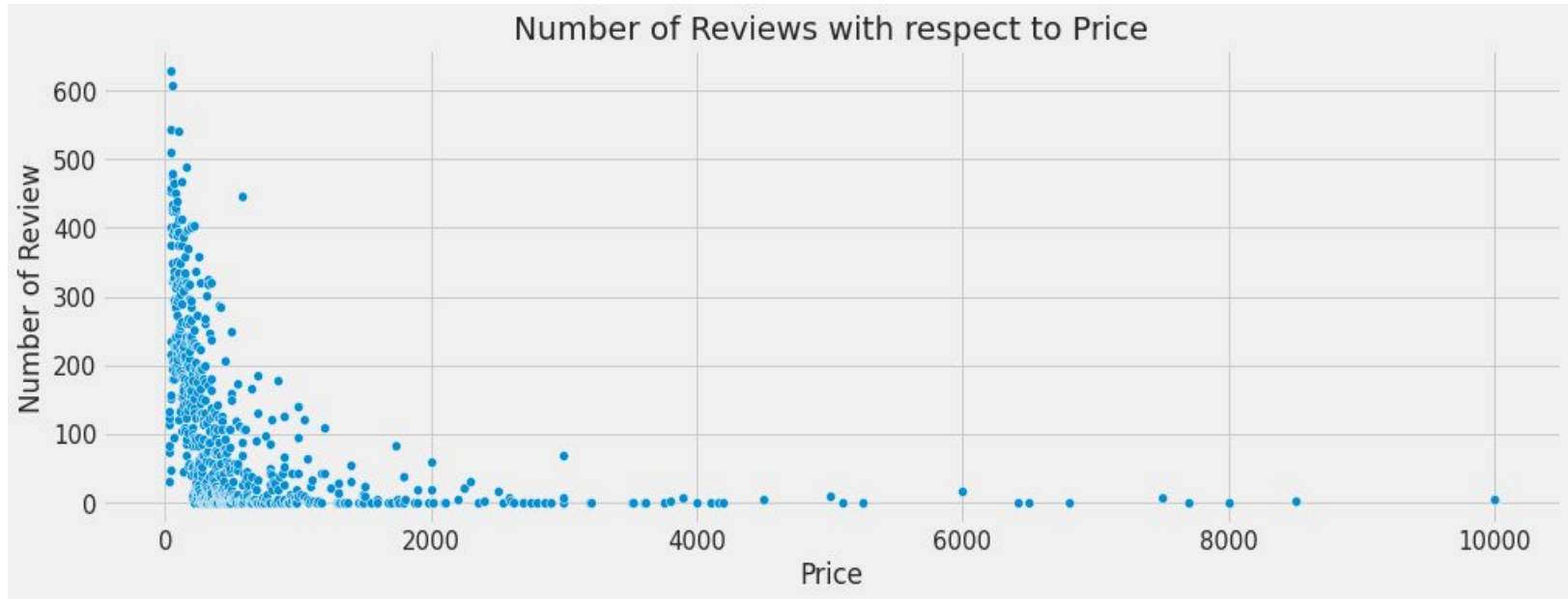
Revenue generated by the room category (Cheap, affordable, expensive, luxurious) :
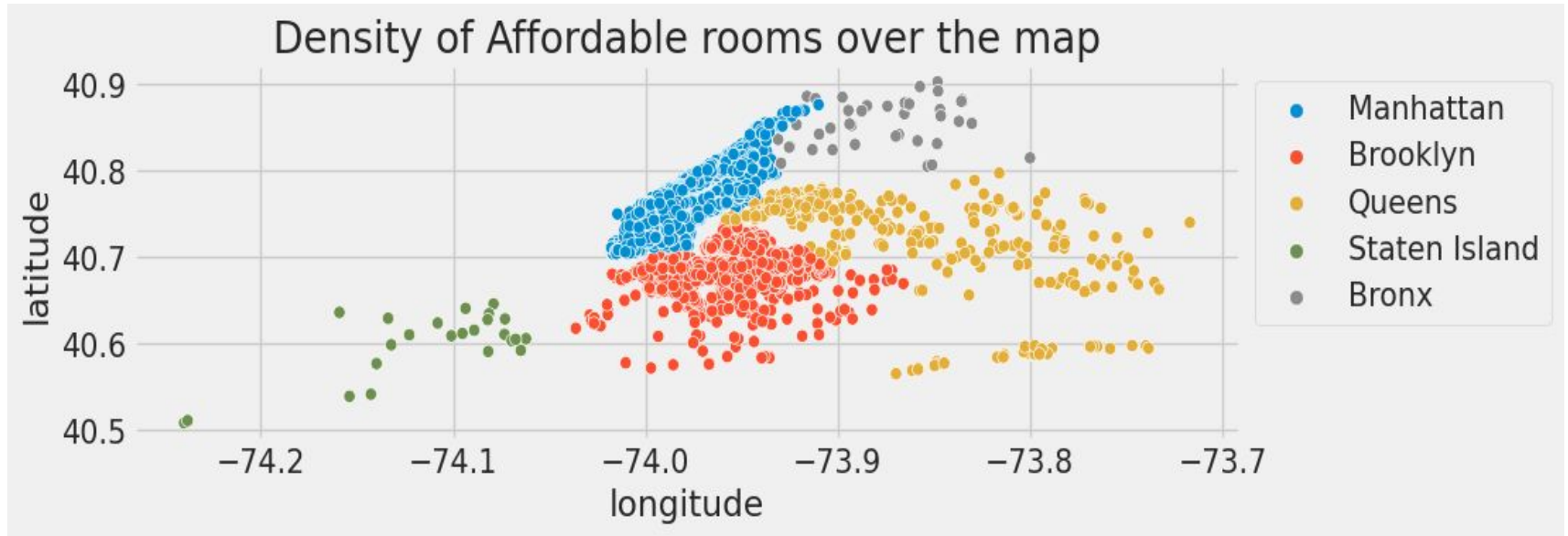
- Differentiated the room category on the basis of their prices and derived which category is generating the most reviews.
- **Cheap rooms** have generated the most revenue while **Expensive rooms** have generated the least.
- **Cheap-** < $200
- **Affordable-** $(200-500)
- **Expensive-** $(500-1000)
- **Luxurious-** > $1000
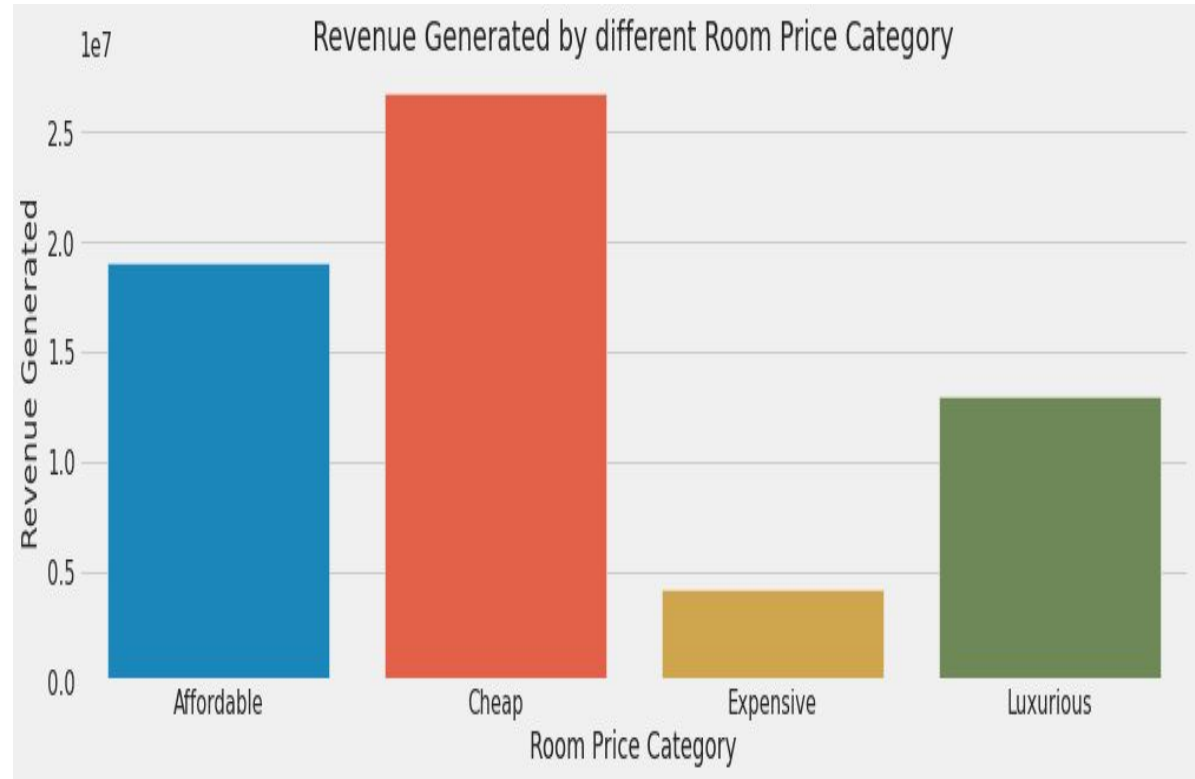
Revenue Generated by different Room Price Category

Locations of different categories of rooms (Cheap, affordable, expensive, luxurious) over the map :

Density of Cheap rooms over the map

- Brooklyn
- Manhattan
- Queens
- Staten Island
- Bronx
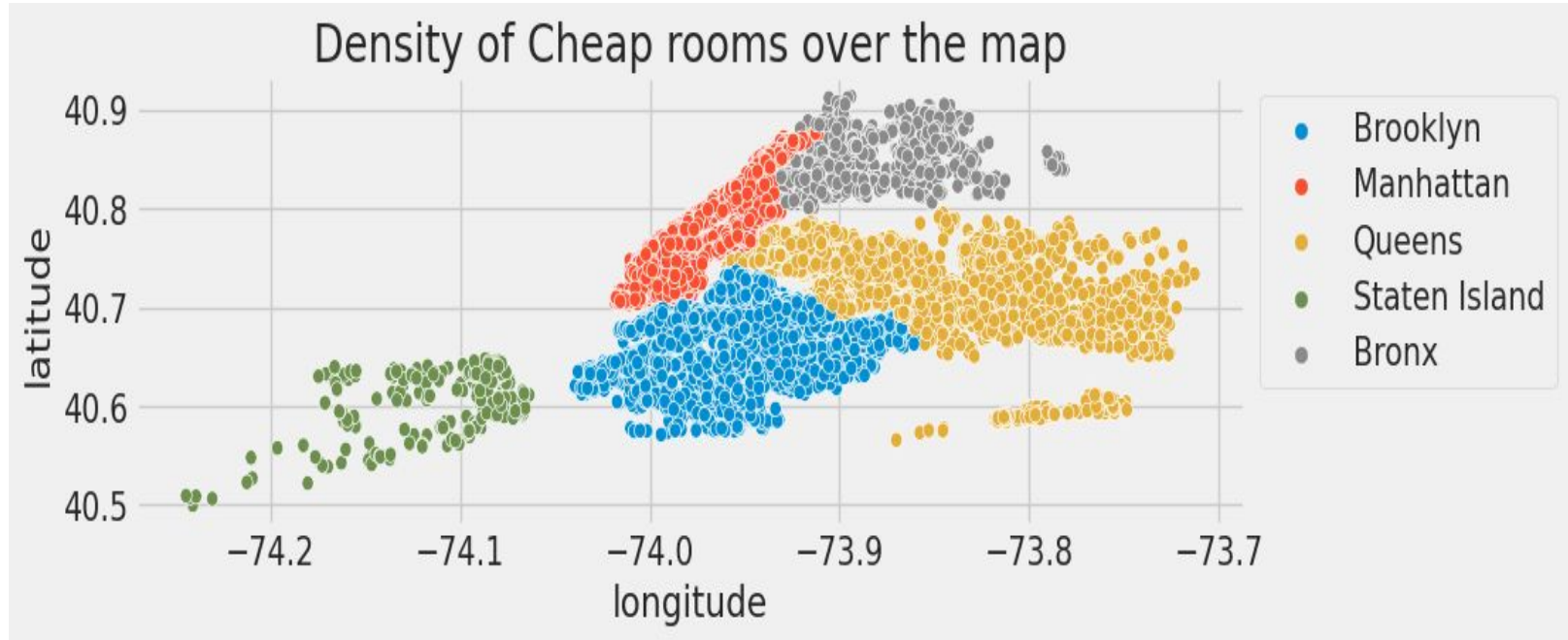
- The number of **Cheap** rooms are more in **Brooklyn** compared to others.

Density of Affordable rooms over the map

- The number of **Affordable** rooms are more in **Manhattan** compared to others.

Density of Expensive rooms over the map
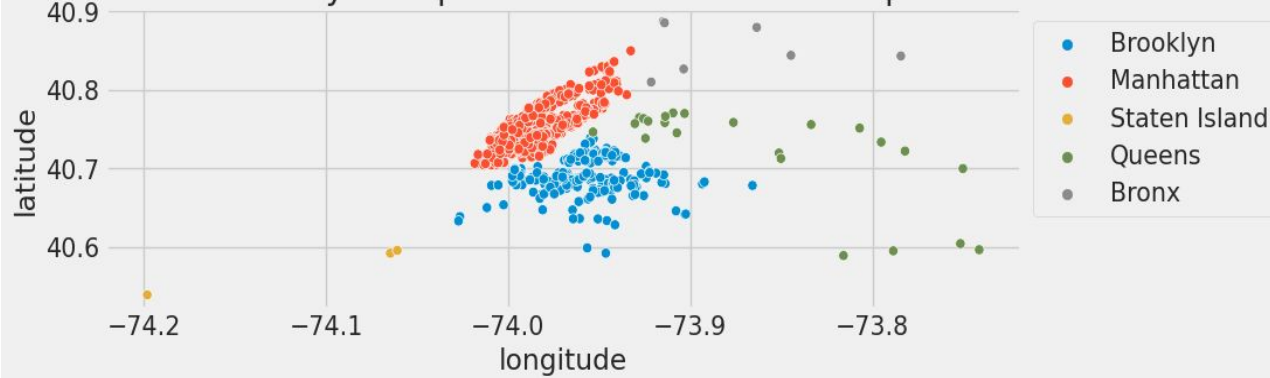
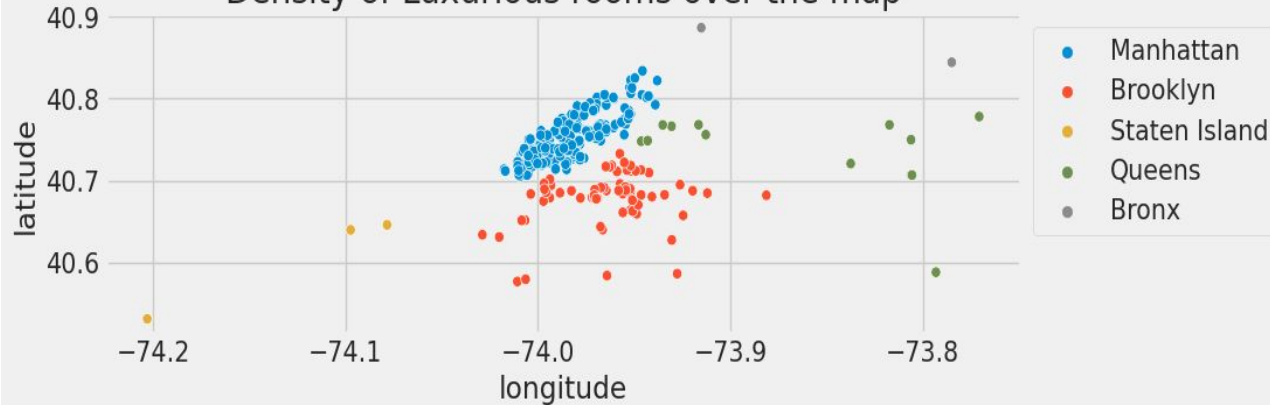Density of Luxurious rooms over the map

- The number of **Expensive** rooms are more in **Manhattan** compared to others.
- The number of **Luxurious** rooms are more in **Manhattan** compared to others.

Noticeable difference of Traffic among different areas

- People in **Manhattan and Brooklyn** like to stay longer in **Entire home/apt** as compared to **Private rooms**.
- People in **Queens** prefer to stay in **Private rooms** as compared to **Entire home/apt**.
- In each Neighbourhood group **Shared rooms** are least prefered to stay.
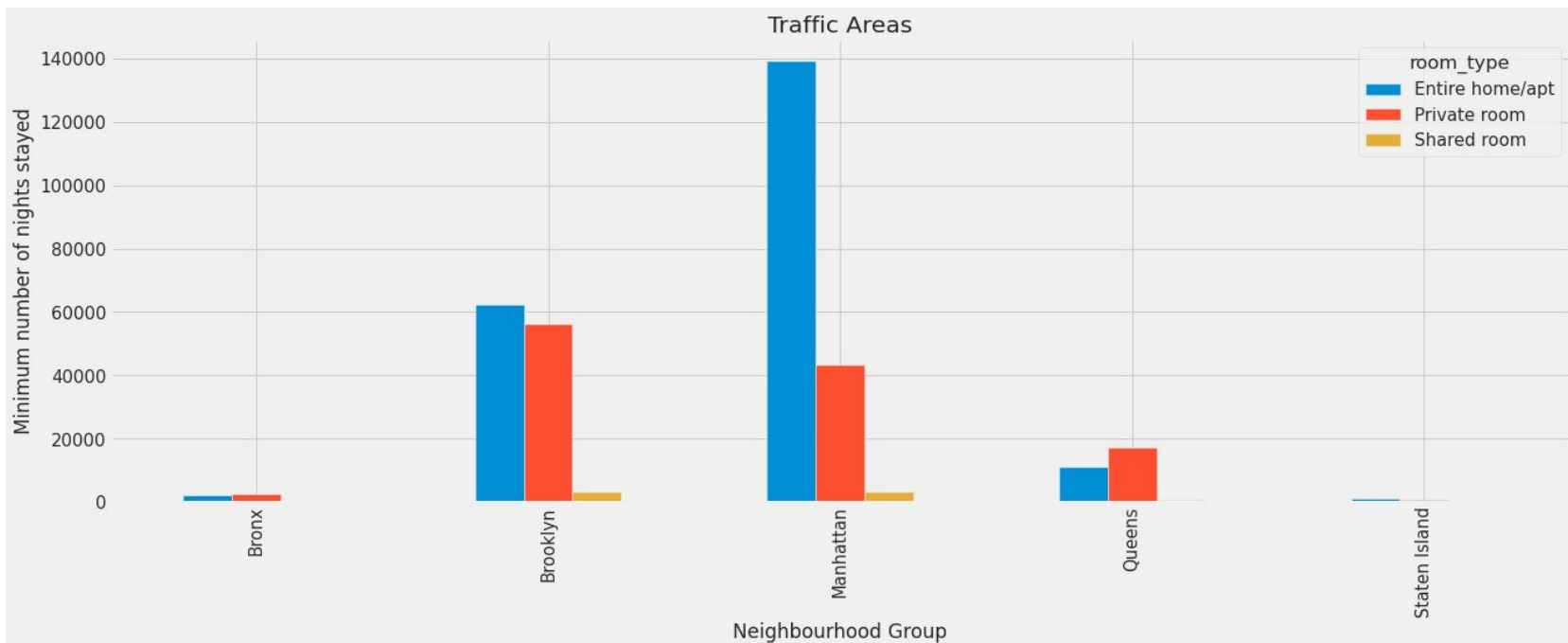
Percentage of the times people reviewed a hotel/apt in different Neighbourhood Group :

- In **Manhattan** percentage of people, who are **not reviewing** the rooms are **more** compared to others.
- In **Staten Island** the percentage of people have **reviewed** the rooms the **most.**

Percentage of Review with respect to Neighbourhood Group
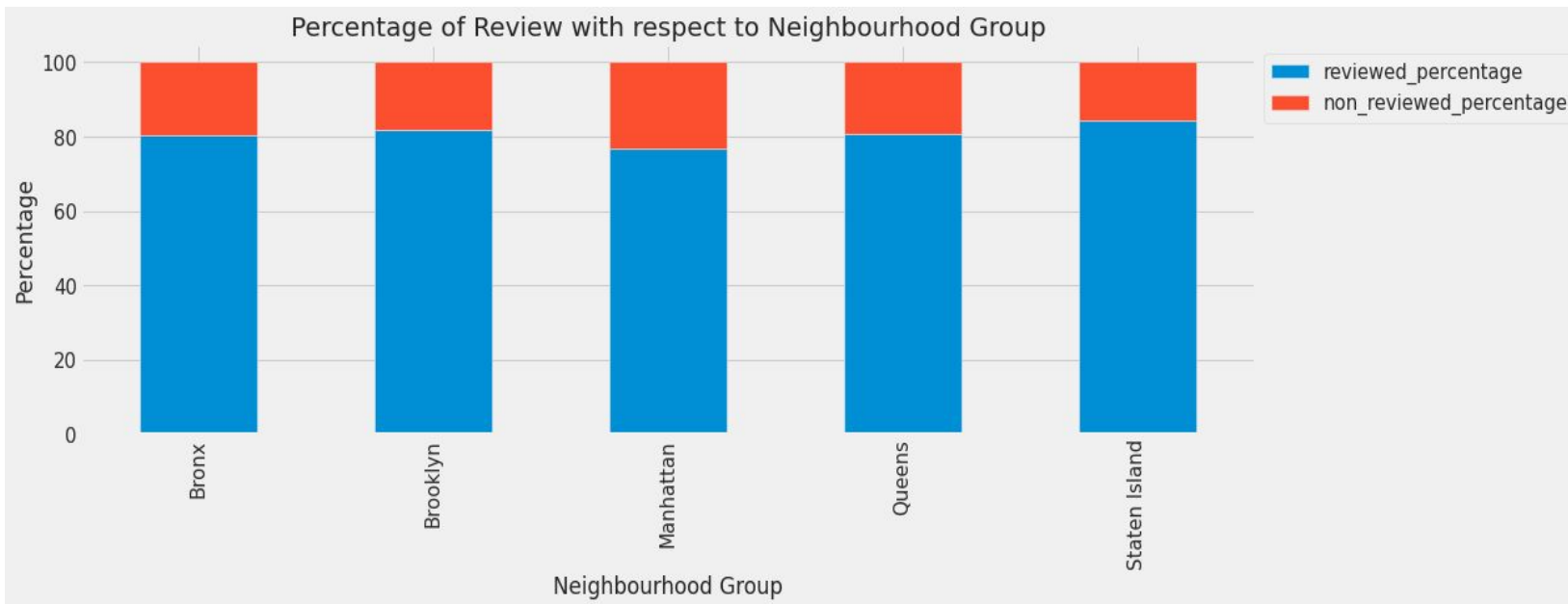
# Average availability of nights for different "Neighbourhood Group" :

- The Bar plot shows the **availability of rooms** with respect to the **neighbourhood group**.
- **Staten Island** have the **highest** availability followed by **Bronx**.
- **Brooklyn** has the **least** availability. That means it is the busiest neighbourhood group in NYC.

Average Nights Availability in Neighbourhood Group
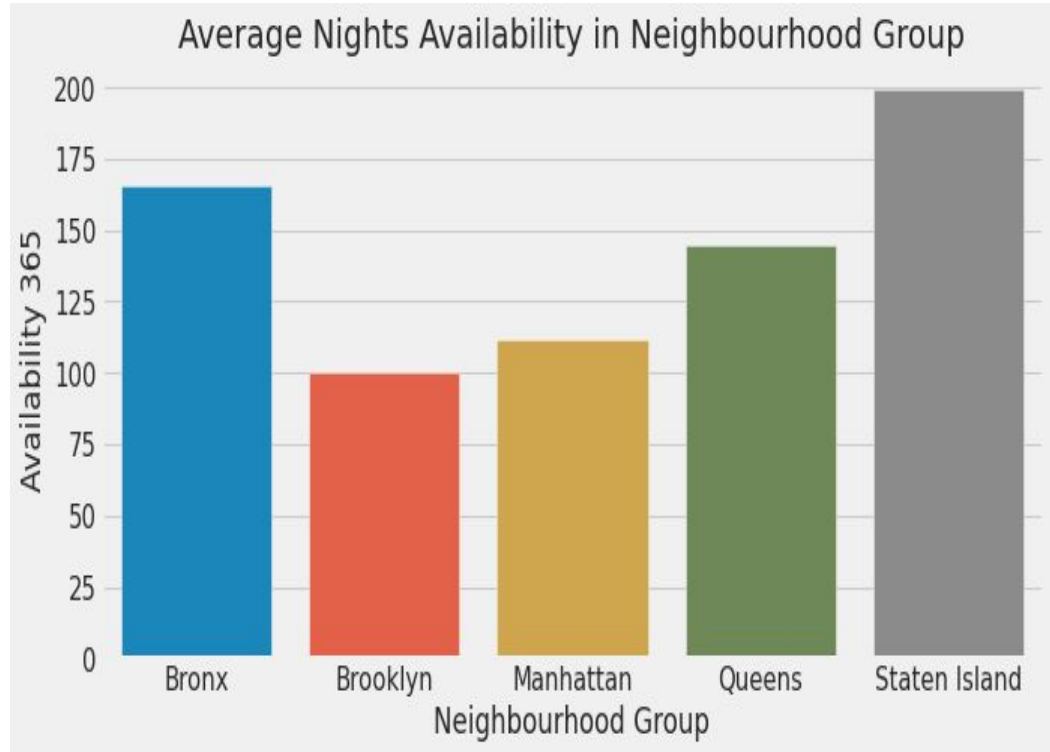
Correlation between different columns present in our data set :

- From the correlation map we can conclude that there is a **weak relation** between all the columns *except* **'reviews_per_month'** and **'number_of_reviews'**.

Correlation matrix of numerical variables
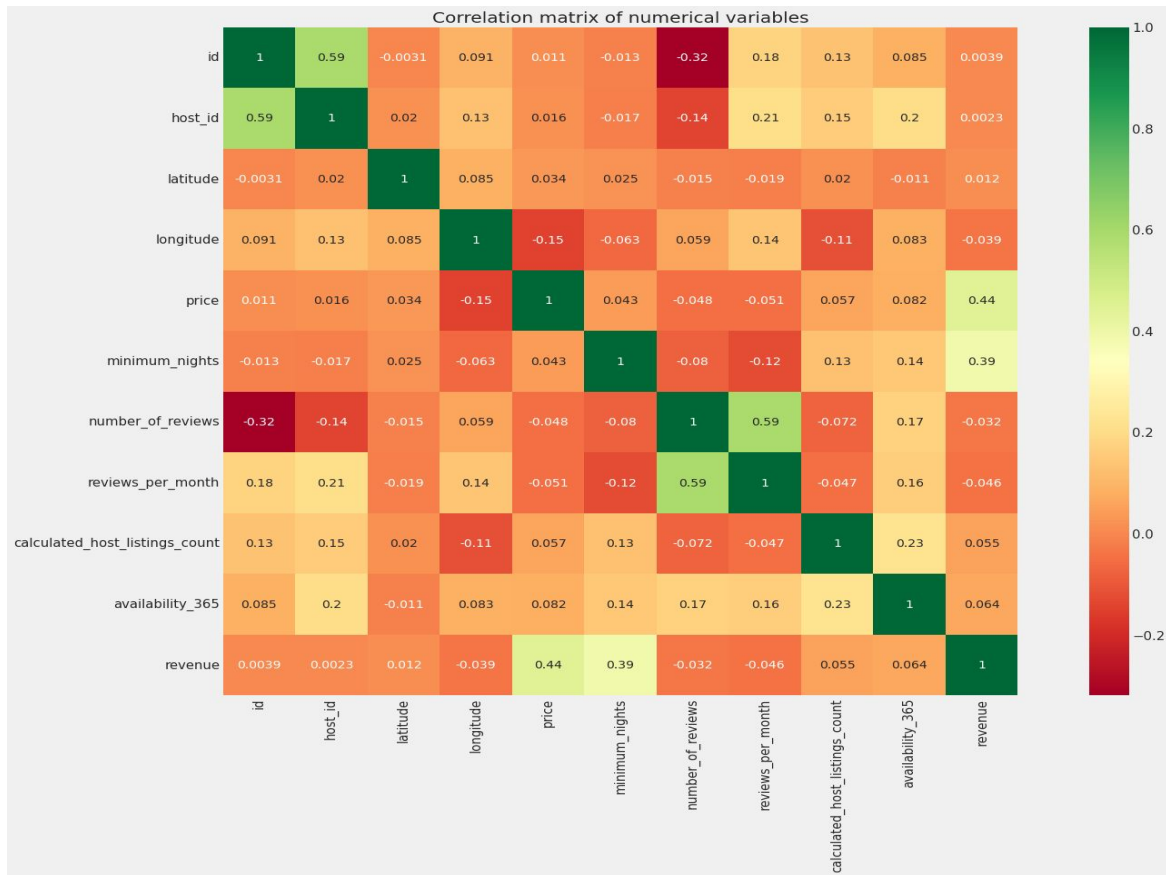
Challenges we faced :

- Understanding some Columns present in the Dataset.
- Handling the Null values present in the Dataset.
- Handling some exceptions present in the Data.
- Doing some further inspections to get the answer for the questions arose from insights.
- Deciding appropriate plots to represent the observations.

CONCLUSION :

- Sonder(NYC) and Blueground are the hosts with most number of listings on AirBnb NYC.
- Manhattan is the neighbourhood group which is most focused by the hosts.
- Private rooms and Entire home/apt have a higher demand and hence hosts who deals with them are comparatively busier.
- Blueground is the host who has generated the most revenue.
- Brooklyn and Manhattan are the neighbourhood groups with most reviews i.e most of the bookings are made for these two groups.
- People have preferred to stay in rooms with low prices and hence cheap rooms have generated the most revenue.

- The cost of living is higher in Manhattan as compared to other neighbourhood groups.
- Private room and Entire home/apt are the two most preferred room type in New York and also generate significantly more revenue. So AirBnb should prioritize these room type.
- The people who prefer to stay in Entire home/apt are going to stay a bit longer in that particular Neighbourhood.
- Around 80 % of the people are reviewing the rooms they have stayed. AirBnb should encourage people to review more in order to know the customer satisfaction.
- If a person or Family is a casual visitor or has an urgency to book a room, Staten Island is the perfect place for that as the average availability of rooms is high.

IMPROVEMENTS TO BE MADE :

- AirBnb should give extra support to those hosts who are tend to be busiest and host who generate significant amount of revenue. AirBnb should also encourage other hosts to improve their facility and service so that more and more customer would get attracted to their property as well.

- As of now Manhattan and Brooklyn are mostly liked by people because these two places are peak attraction of New York. Yet we can't neglect other cities. So AirBnb should collaborate with hosts to improve facilities and add some fancy ideas in the rooms so that people will get attracted towards the rooms of other cities as well.

- From lots of analysis we saw that Shared rooms are least preferred. So AirBnb should analyse the reason behind it and make improvements so that both host and company should get more profit.

- We saw that cheap rooms are mostly preferred. So AirBnb should reduce the margin of the price(not too much) so that more people could afford the rooms which will ultimately increase the profit.

- Reviews are the most important aspect of increasing the business. As of now there is no particular kind of review (good or bad) in our data set. So we suggest AirBnb to monitor the real time reviews so that it can bring some revolution in their service.

Thank you

Team data_hacker