Capstone Project

Airbnb Bookings Analysis

Presented by

SANCHIT PATIL

RAMAN MAGAR

PREMANAND GAIKWAD

OSAMA ANSARI

Contents :

What Is Airbnb?

- Air, Bed And Breakfast forms a name as Airbnb.
- It is American company that provides an online marketplace platform for hospitality services such as lodging, primarily home stays for tourism and vacation rentals.
- The company was founded in 2008 by Brian Cesky, Nathan Blecharczyk and Joe Gebbia.
- Airbnb doesn’t have any own property, it generates revenue from receiving commission from each bookings.
- Bookings platform of Airbnb is accessible via website and mobile app.

Problem Statement :

- Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities a nd present a more unique, personalized way of experiencing the world.
- Nowadays, Airbnb became great platform for hospitality services.
- Data analysis helps for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing init iatives, implementation of innovative additional services and much more.
- We have data of Airbnb bookings in New York city.

- This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.
- What can we learn about different hosts and areas?
- 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?

Data Summary

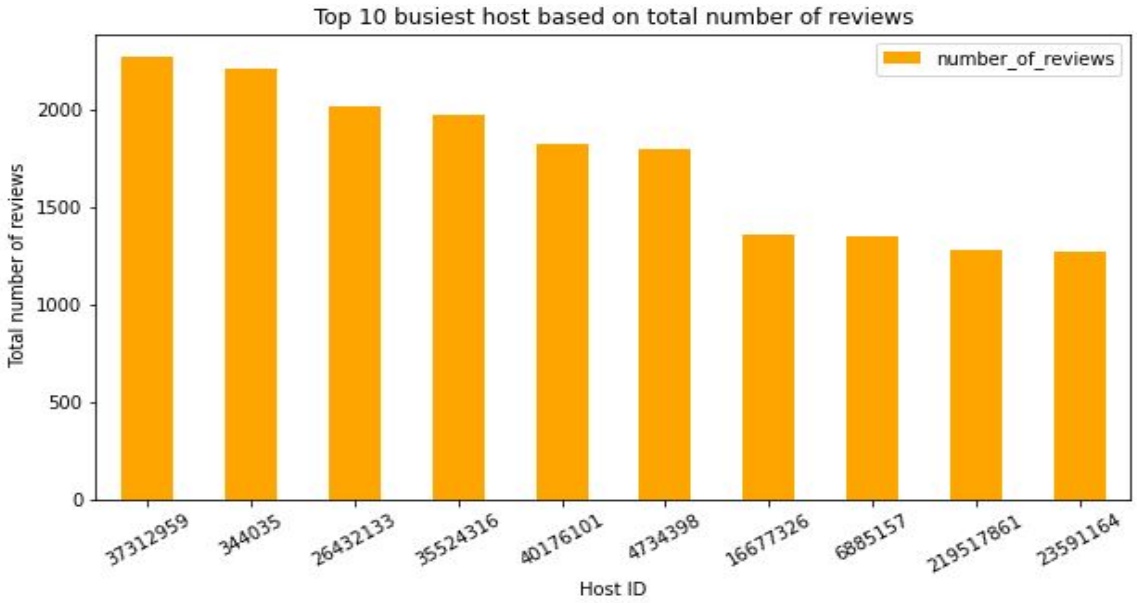
- We were provided sixteen dataset of New York with deep information about the prices, locations, customers preference to stay, properties available at different locations across the various cities, properties owners with there names and Identity number, so analysis on this listing data is the crucial factor
    - Host Name: Name of the host who is owner of the Property.
    - Neighbourhood Group: Group of residents or property owners who advocate organize activities within a neighborhood.
    - Neighbourhood: A district or community within a town or city. so analysis on this listing data is the crucial factor

Exploratory Data Analysis (EDA)

We have divided EDA into dataset wise different part.

1. Host wise analysis:

1.1 By total number of customers/reviews



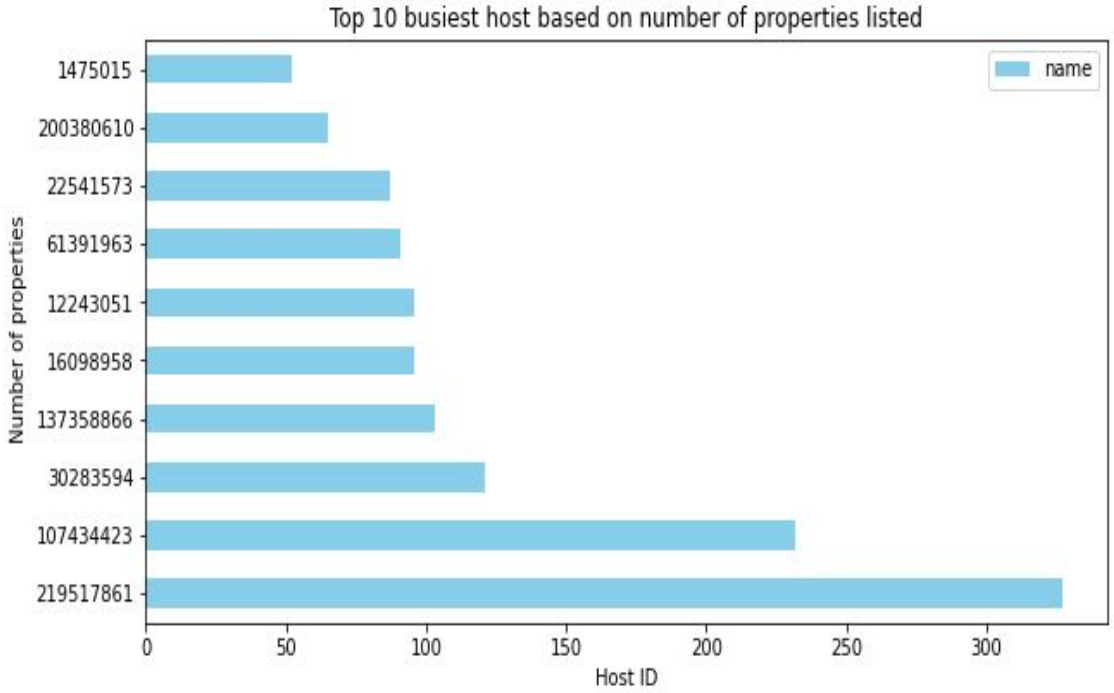
Busiest Host According to Number of Reviews

Host ID – 37312959

Host name – Maya

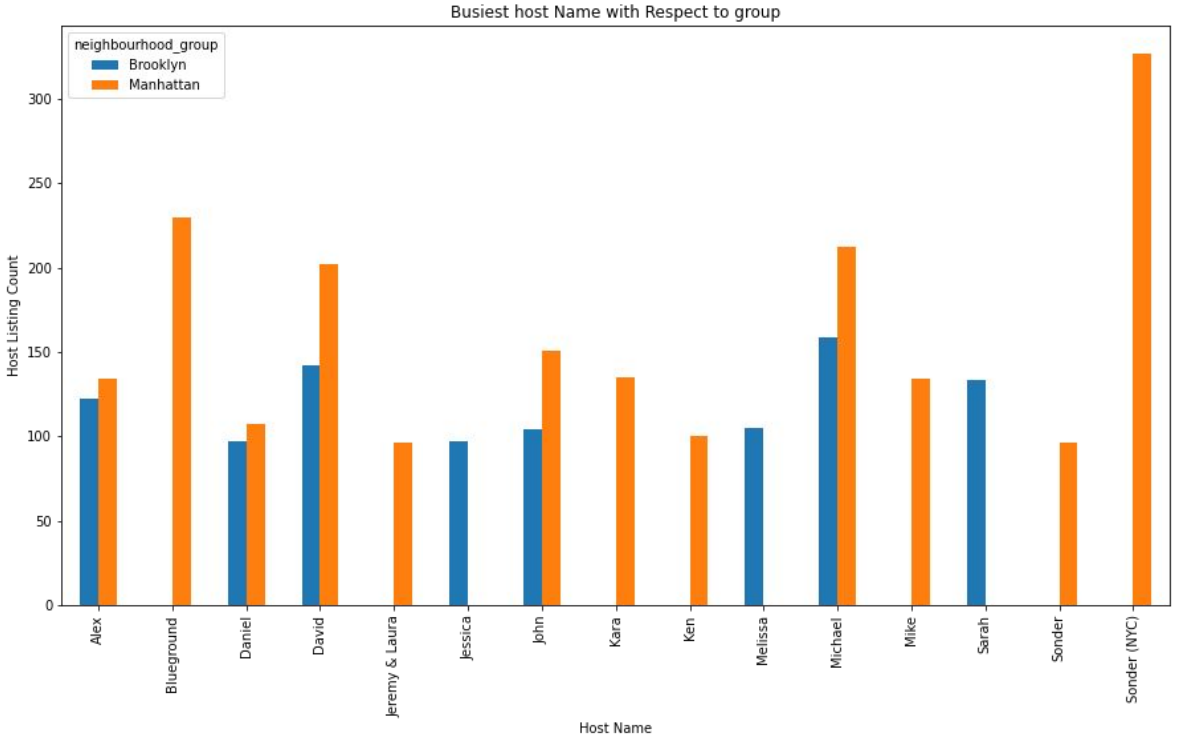
is the busiest host by **2273** number of customers.

1.2 By Number of Properties Listed



Among the top 10 host by number of properties listed
Host ID – 219517861
Host Name – Sonder (NYC)
Is the busiest host by **327** number of properties listing.

1.3 By Neighbourhood group

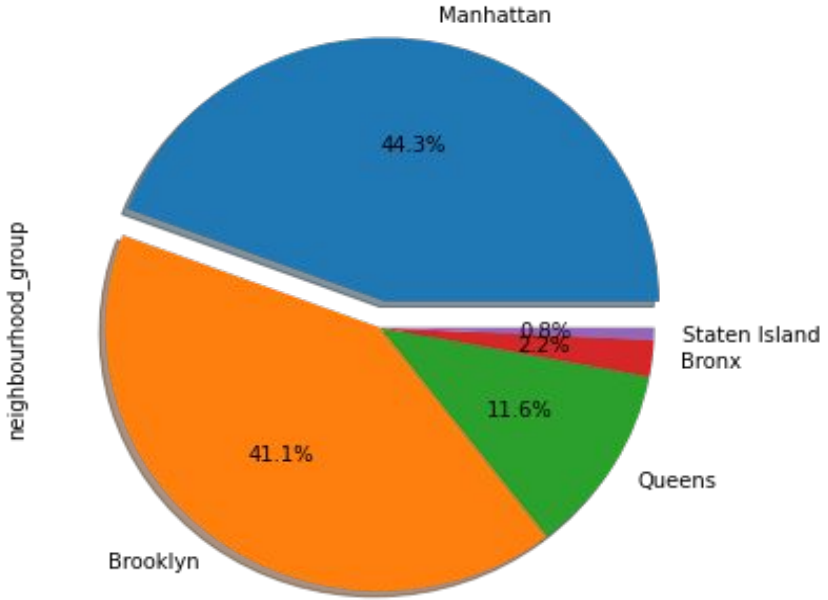


- ✔ Top 10 host are from the Brooklyn and Manhattan group
- ✔ In Manhattan group of cities **Sonder(NYC)** is the busiest host by **327** number of properties listing.
- ✔ In Brooklyn group of cities **Michael** is the top host by **159** number of properties listing.

2. Neighbourhood Group wise :

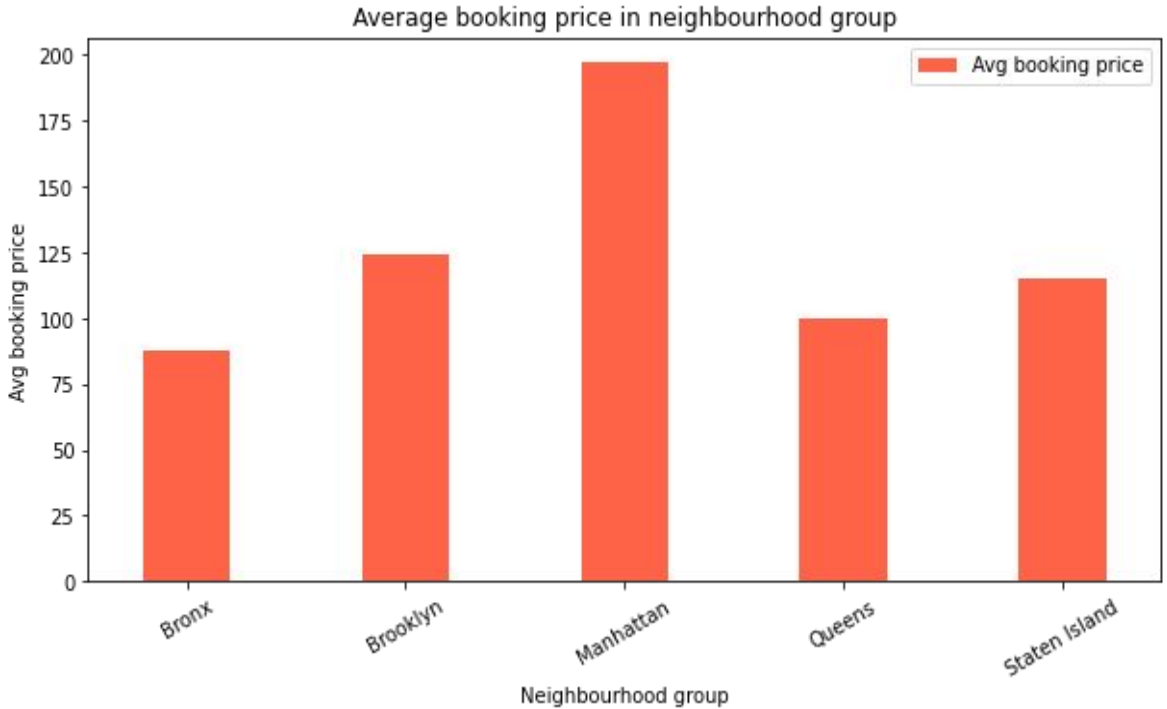
2.1 Neighbourhood group which has max number of customers/bookings

From neighbourhood group of cities it is seen that **Manhattan** has max number of customers or bookings by **44.3%** share the total bookings in New York 2019 data.



2.2 Neighbourhood group wise average booking prices

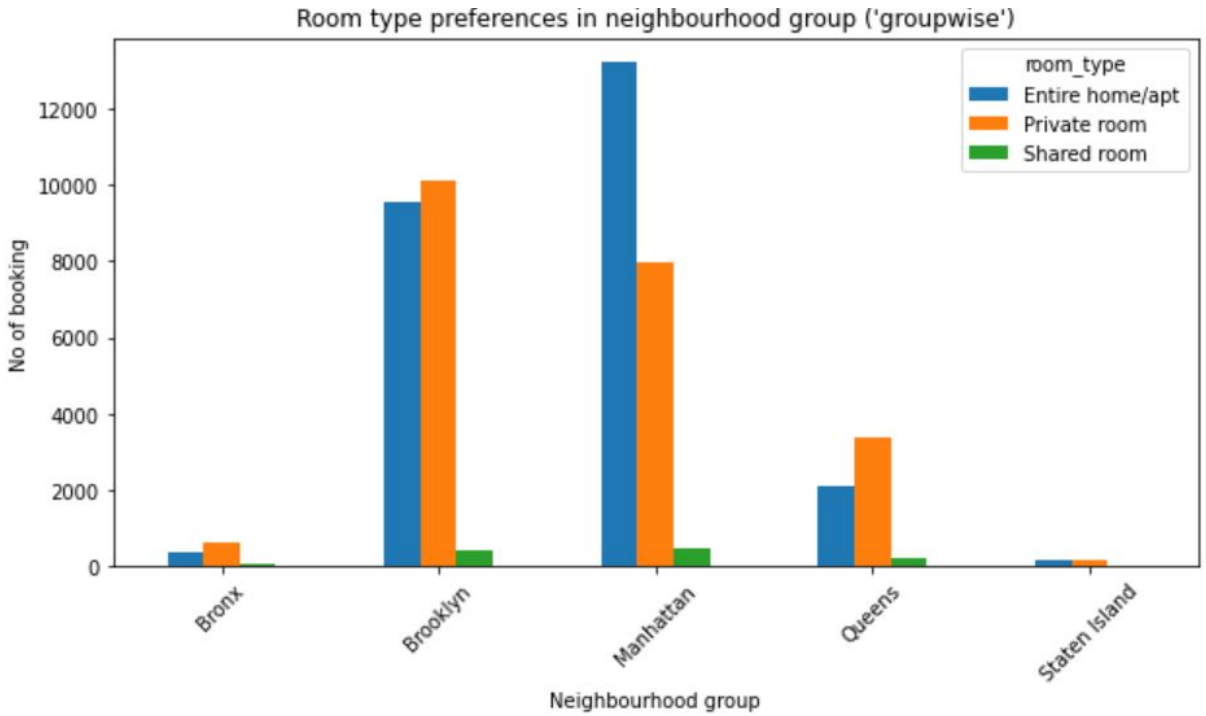
Average booking price in Manhattan group of cities is maximum amongst the other group i.e. in **Manhattan** stay is expensive than other. Maximum average booking price is **196.8 $.**



2.3 Room type preferences in the neighborhood group

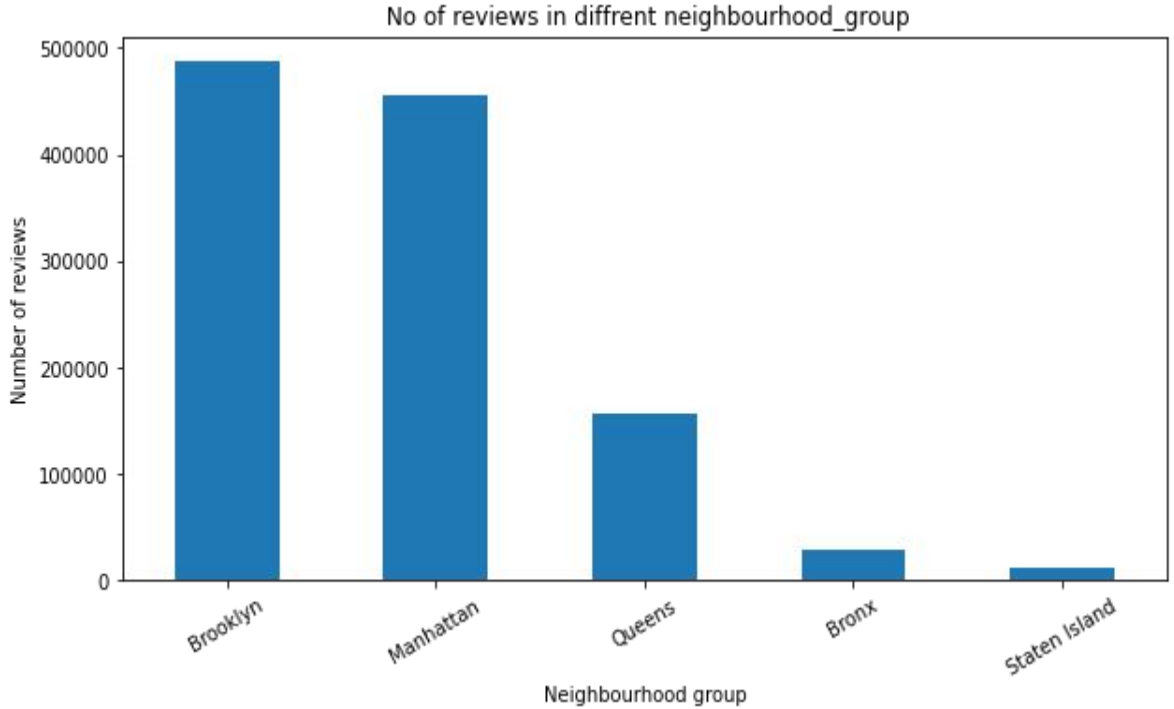
Room type preferences in group of cities:

- ✔ **Entire Home/apt** are mostly preferred in **Manhattan by 13199** number of customers.
- ✔ **Private room** are mostly preferred in **Brooklyn** by **10132** number of customers.
- ✔ **Shared rooms** are also mostly preferred in **Manhattan** by **480** number of customers



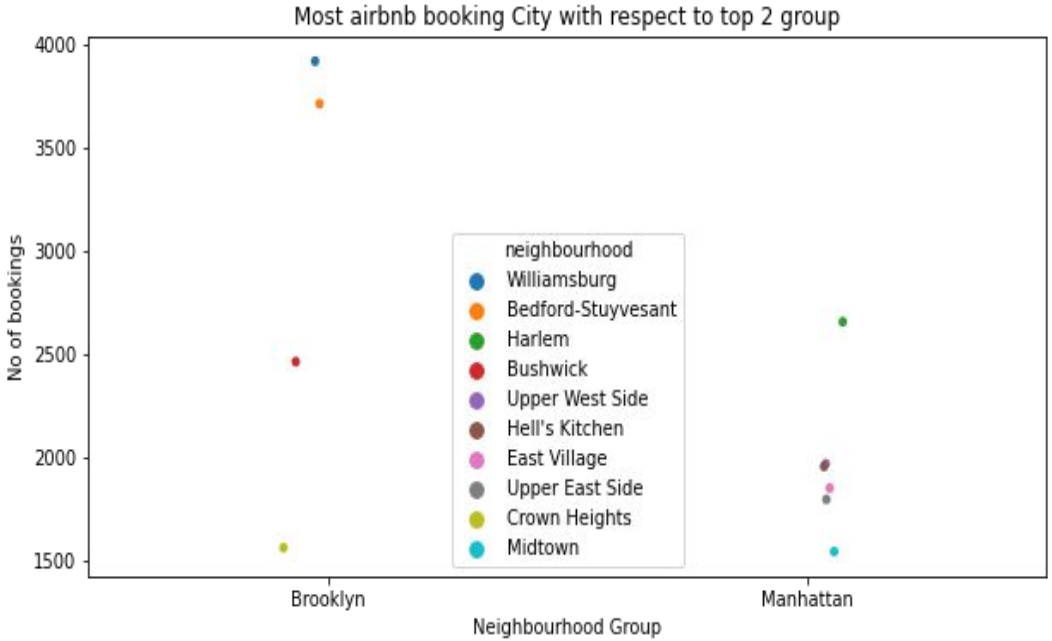
2.4 Neighbourhood groups based on the number of reviews

Number of reviews for **Brooklyn** group of cities is maximum amongst the other groups which is **486574** reviews.



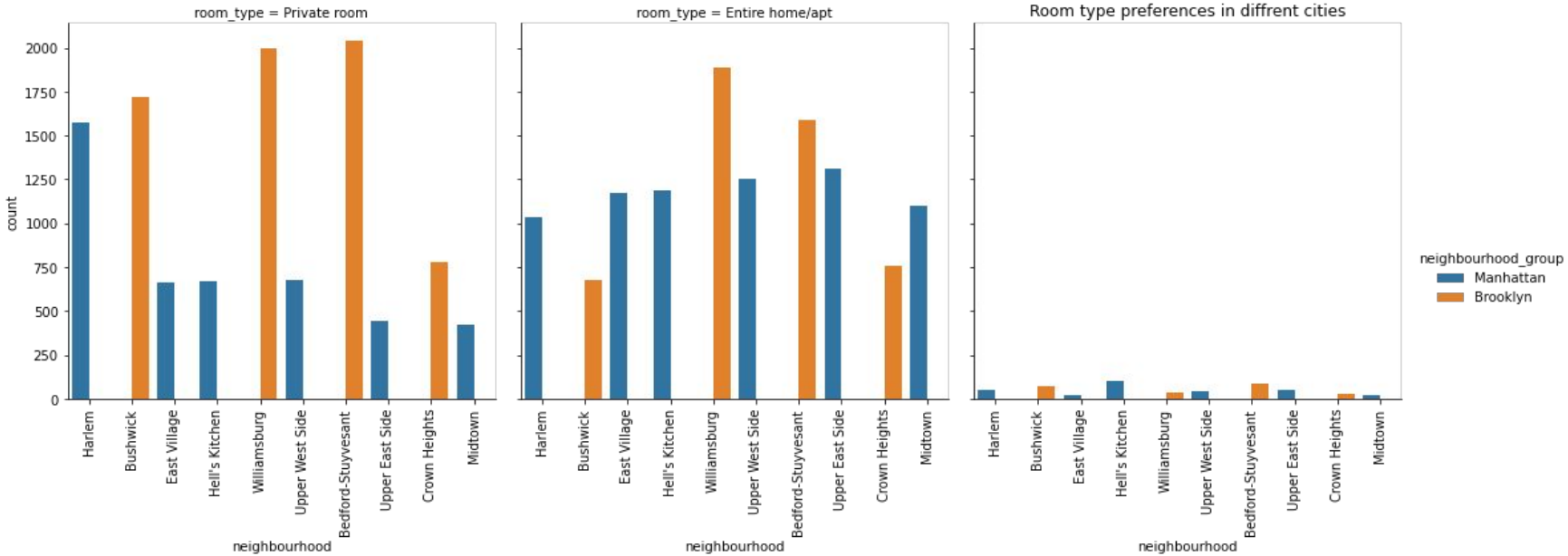
3. Neighbourhood or cities wise

3.1 Airbnb bookings in cities with respect to top 2 groups.



- ✔ Max number of bookings are from **Williamsburg** city of Brooklyn group, **3920** number of bookings.
- ✔ In Manhattan group **Harlem** city has maximum number of bookings, **2658** number of bookings.

3.2 Room type preferences in Neighbourhood or Cities



✔ Private rooms mostly prefer in '**Bedford-Stuyvesant**' city of Brooklyn.

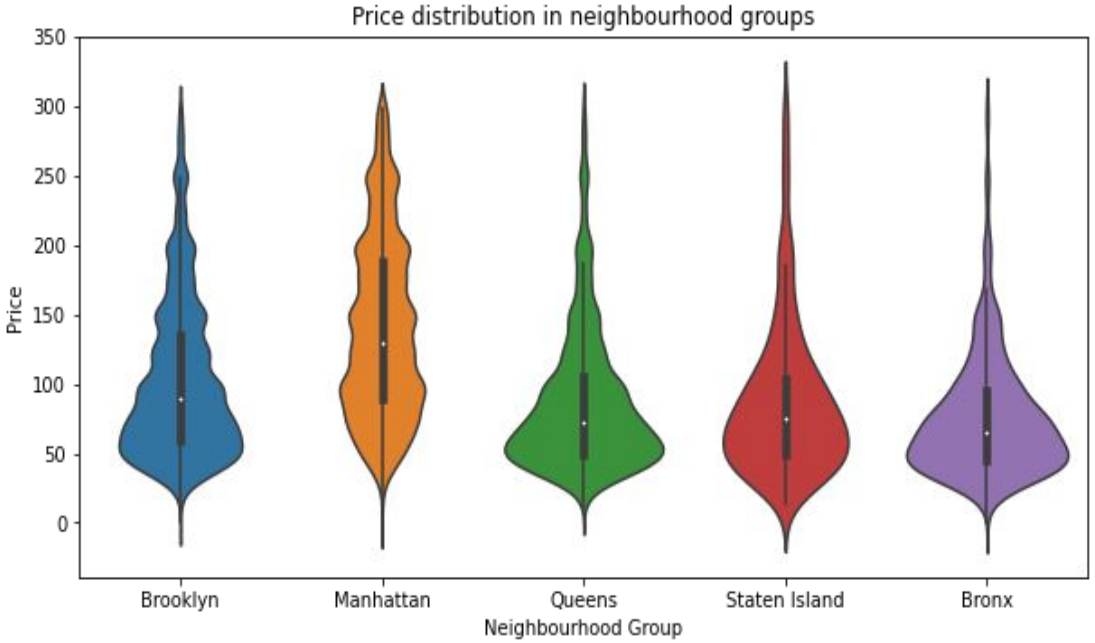
✔ Entire home/apt mostly prefer in '**Williamsburg**' city of Brooklyn.

✔ Shared rooms are prefer in **'Hell's Kitchen**' city of Manhattan.

4. Price distribution analysis in different areas

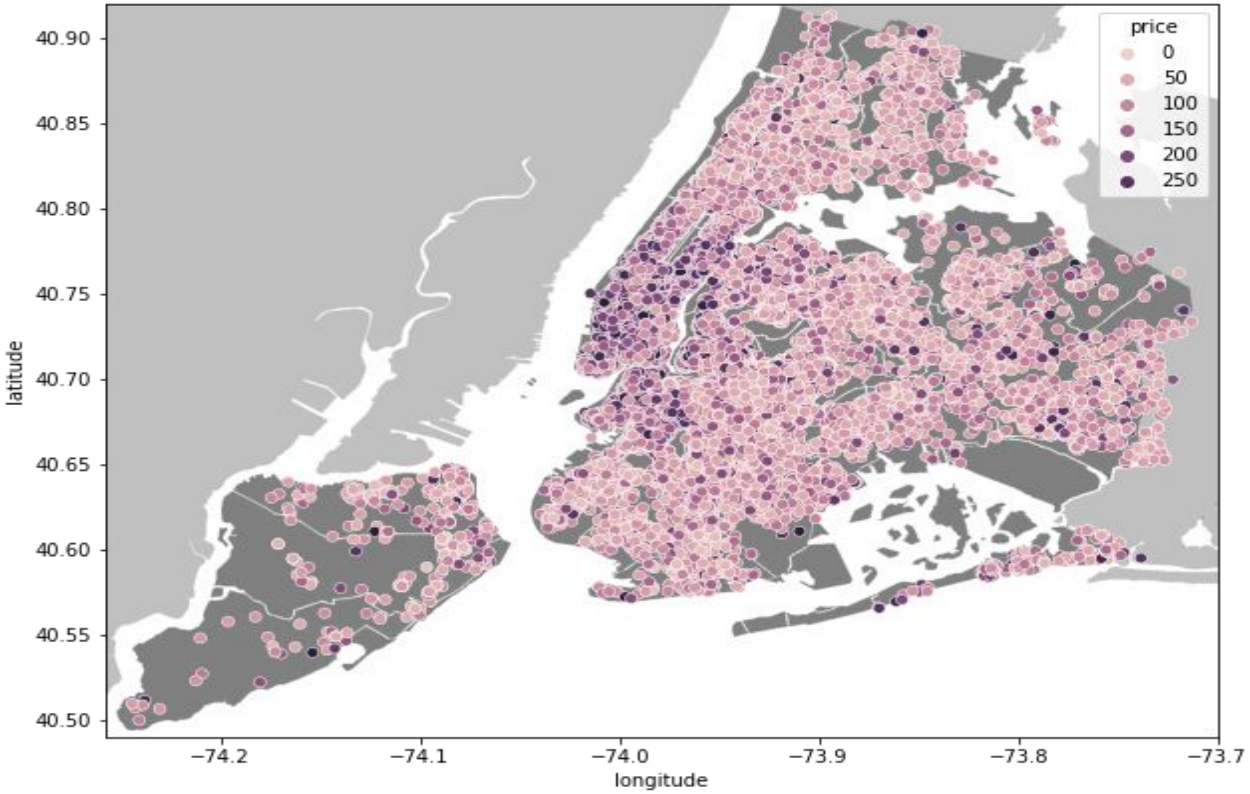
4.1 Price distribution in neighbourgood groups

- ✔ **Manhattan** has wide variety of price range with moderate bandwidth.
- ✔ **Bronx** has less price range with higher bandwidth.



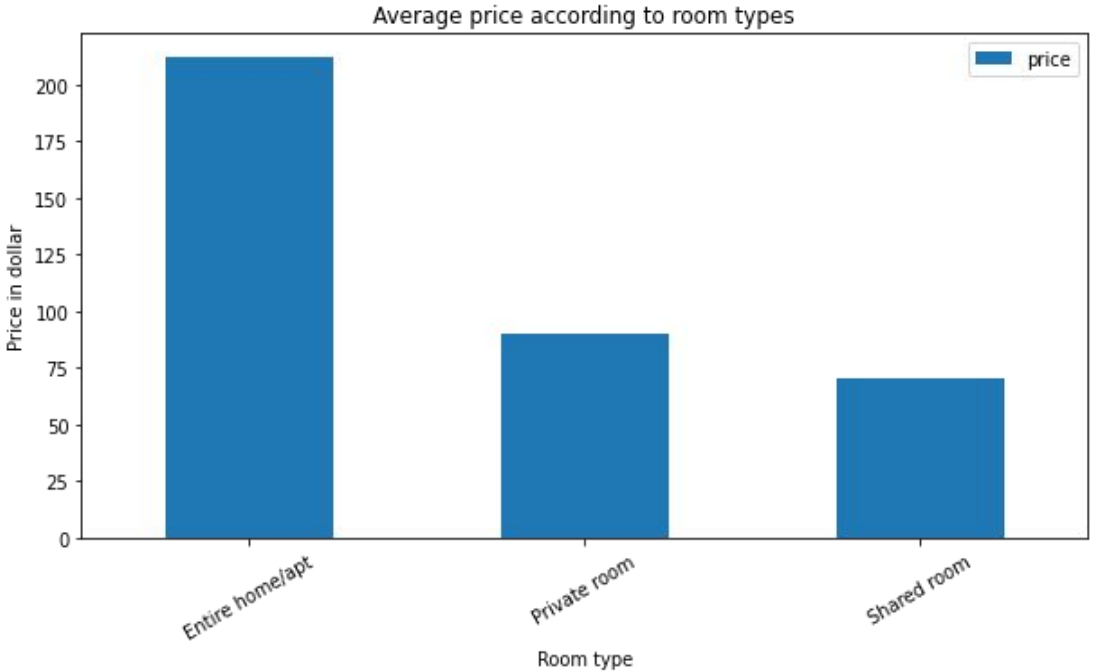
4.2 Prices distribution according to geometrical location

Prices of stay are higher in Manhattan and area of Brooklyn near to the Manhattan.



5. Price distribution analysis according to room type

5.1 Average price according to room types

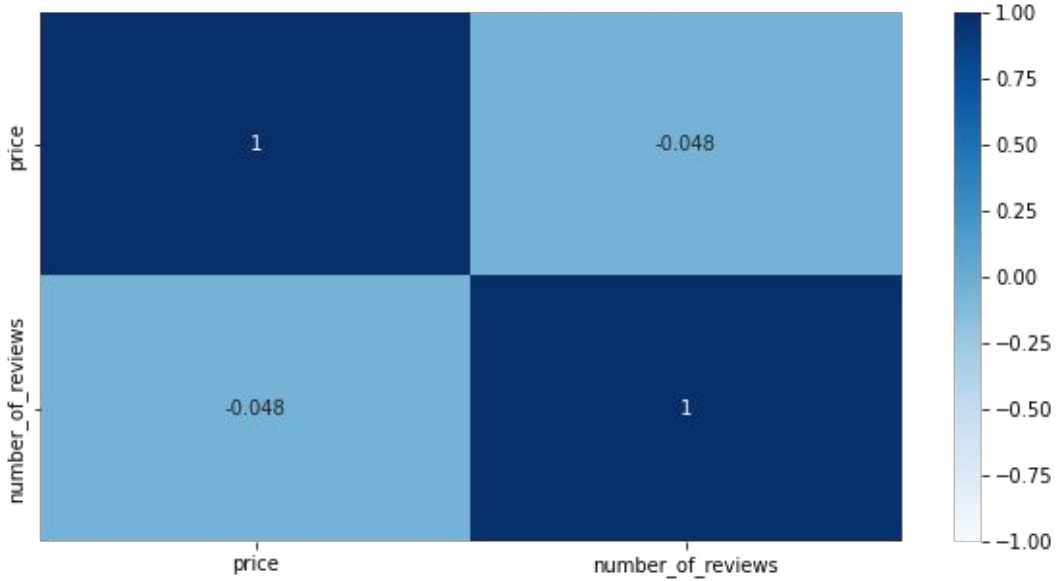


- ✔ Average prices for **entire home or apartment** is higher which is **211.8 $**
- ✔ Lowest average prices is for **shared room** which is **70.12 $**
- ✔ We seen in previous analysis customers are prefer to stay in entire room or apartment.

5.2 Price distribution on box plot to know most preferred price range

- ✔ For **entire home or apartment** most of the customer books stay between **150-175 $**
- ✔ For **private rooms** most of the customer books the stay between **65-75 $**
- ✔ For **Shared room** most of the customer books the stay between **60-70**

6. Correlation Between prices and number of reviews



Price to the number of reviews relation is the negative i.e. it is **inversely proportional.** As prices of the stay increases, less customer refer to stay

Conclusion :

From this we conclude that,

- Busiest host - Maya has 2273 number of Reviews, Sonder(NYC) has maximum number of properties(327) in Manhattan area and Michael is having (159) properties in Brooklyn.

- Manhattan and Brooklyn has the maximum share of customers amongst all other groups, which is 44.3% and 41.1% , respectively where most of the customers prefer to stay in an entire home or apartment.

- Brooklyn has the maximum number of reviews and the corresponding maximum number of customers as it is nearer to Manhattan and at a reasonable price.

- Entire home or apartment and private rooms were mostly preferred.

- Manhattan has a moderate bandwidth in higher price range and 'Bronx' group has a higher bandwidth with lower price range.

- In Brooklyn's group Williamsburg has the maximum number of customers (3920) and in Manhattan's group Harlem has the maximum number of customers (2658) bookings. On top of that entire home apt. were most booked in both the cities.

- We tried to dig out correlation between prices and reviews and we found that both are inversely proportional.

Challenges :

- Huge data was to be handle contain important records of bookings by keeping in mind that not to be miss any small record which is relevance.

- To select and discard dataset which is not required for analysis involve lots of brainstorming. Also to handle null values and process or replace them.

References :

- Stackoverflow.com
- geeksforgeeks.org
- w3resource.com
- Google.com
- Wikimedia.org

Thank you