# Capstone Project

## Airbnb bookings analysis

**Team Member:**
**Roshan Tile**
**Rasik Jain**
**Sakshi Juneja**
**Nikhil Dutt**

# Tables Of Contents

# About AirBnb:

1. Also known as Air Bed and Breakfast.
2. It is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities.
3.It is accessible through website as well as mobile app.
4. It is based in San Francisco, California and founded in 2008.
5.Airbnb provides a platform for hosts to accommodate guests with short-term lodging and tourism-related activities
6. Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking.

# Problem statement

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 contd.

**Explore and analyze the data to discover key understandings (not limited to these) such as :**

- **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?**

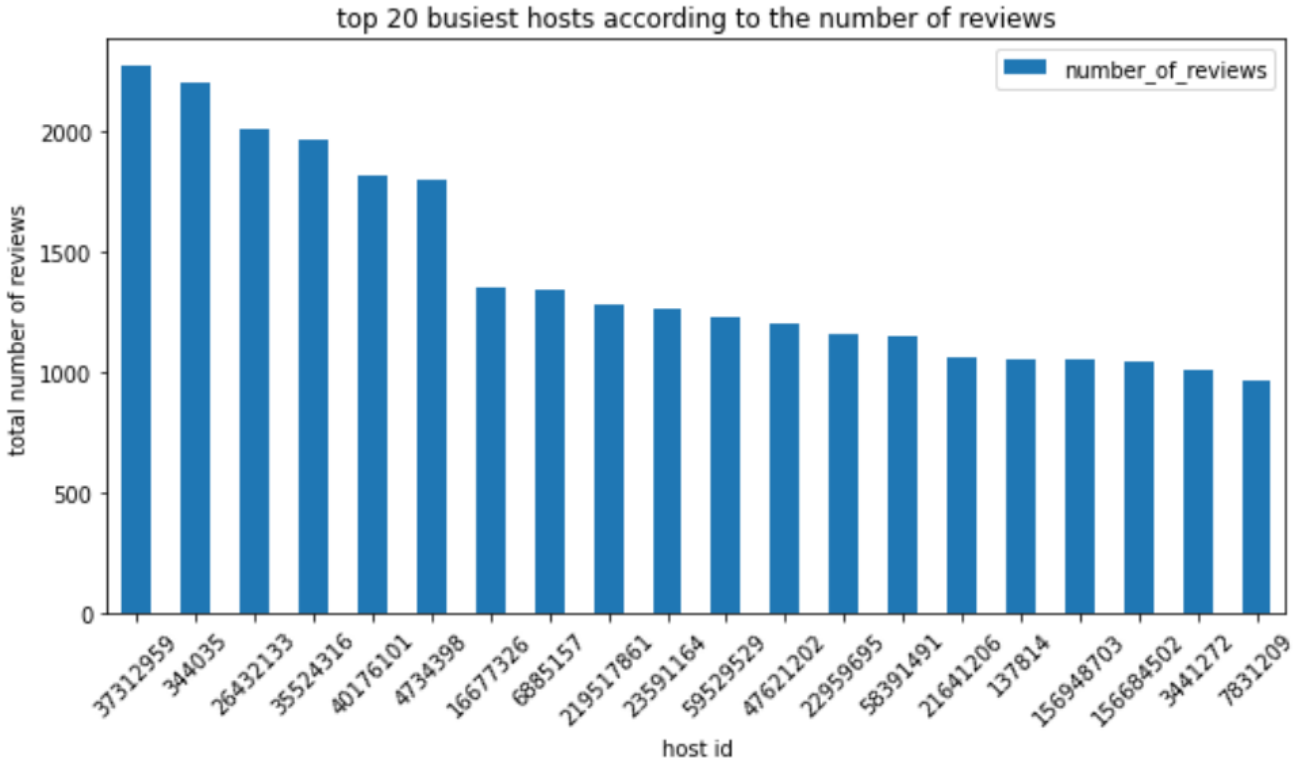
# Data Summary

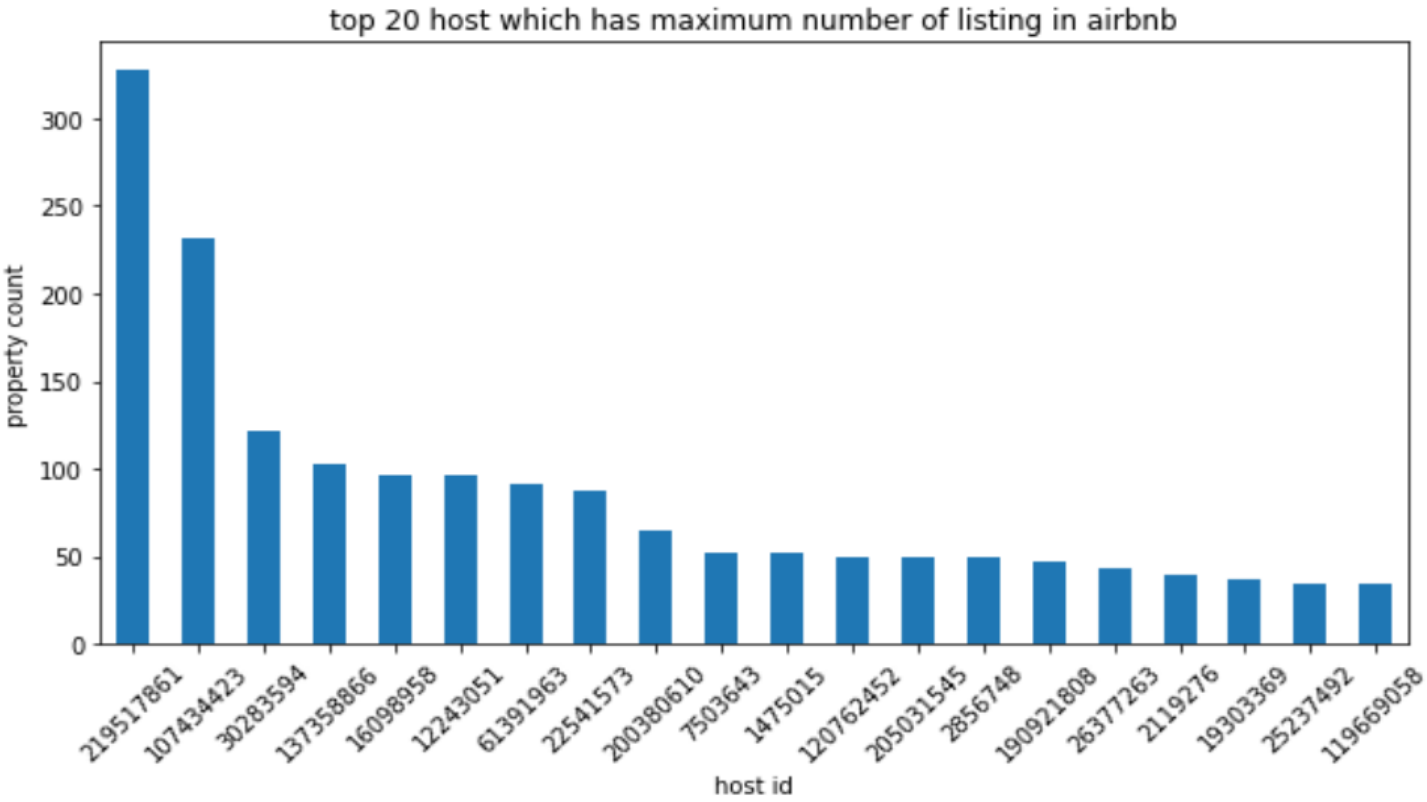
- Airbnb dataset provides deep insights for locations, people's preference and average rates of different properties across various regions of New York. We first perform a basic check for our data to know how many records and column present in our dataset. Then begin our EDA by looking out for different column to select the most relevant column for our analysis. Then, we perform data cleaning by checking and replacing for any null value present in our dataset which can interfere with our analysis.
- Assumption: For our analysis purpose we have assumed that the number of reviews corresponds to number of peoples stays in Airbnb.

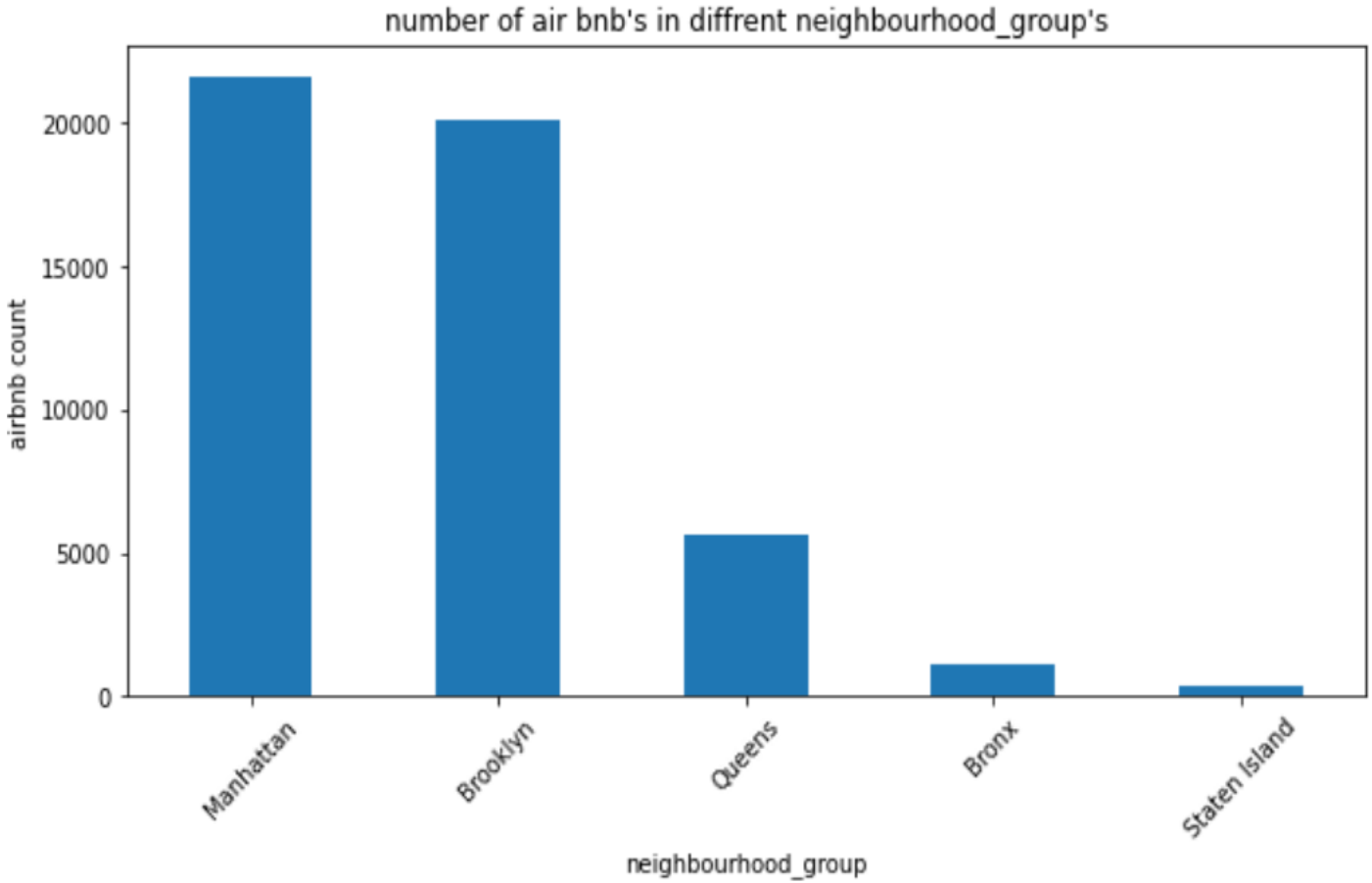
# Data Summary

Looking out at columns in our dataset, we observe that the data is divided into multiple categorical variable based on geographic coordinates, regional groups, regions, room types, host Id as one host may have multiple properties. We also observed columns which gives information about price of stay. Minimum nights for stay, availability and number of reviews.

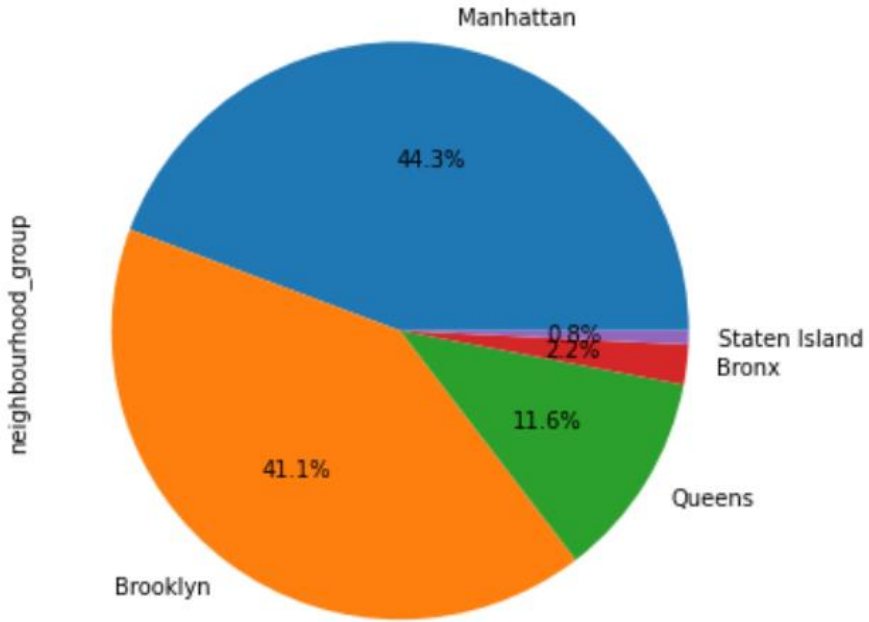
# Exploratory Data Analysis (EDA)

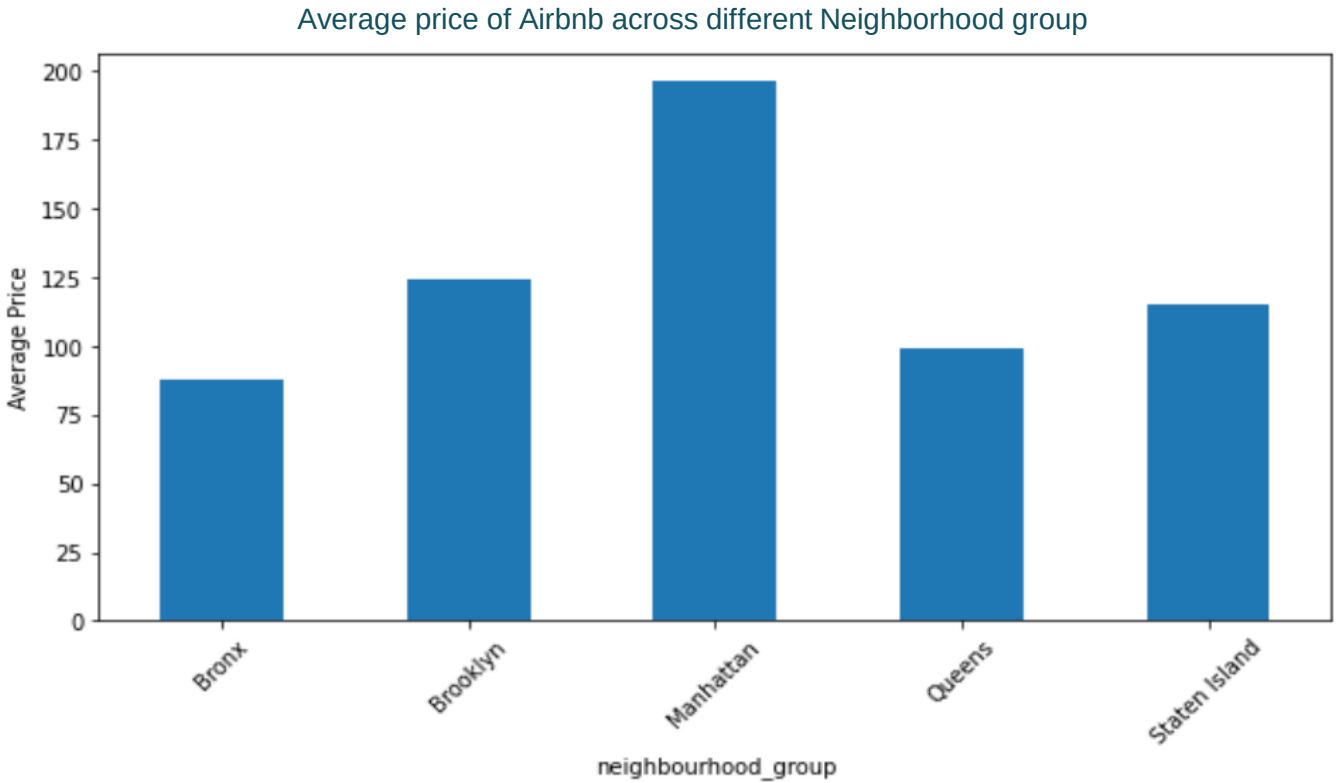



top 20 host which has maximum number of listing in airbnb
property count
0
50
100
150
200
250
300
host id
219517861
107434423
30283594
137358866
16098958
12243051
61391963
22541573
200380610
7503643
1475015
120762452
205031545
2856748
190921808
26377263
2119276
19303369
25237492
119669058

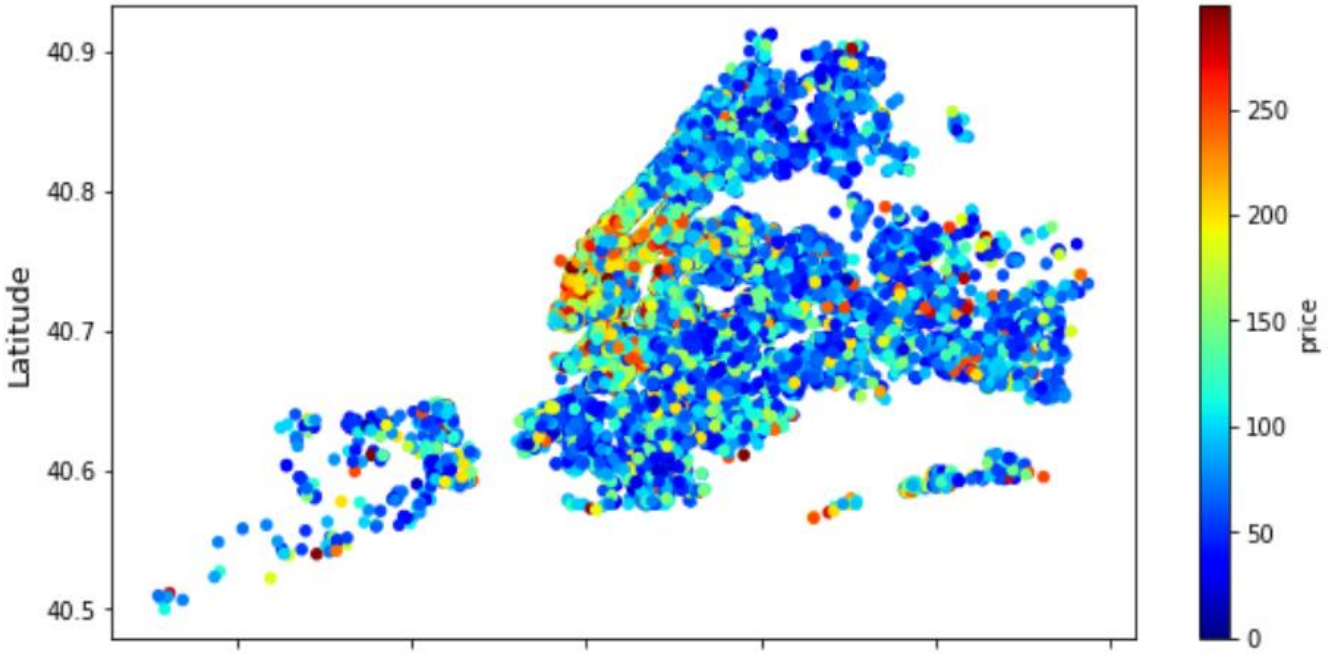

number of air bnb's in diffrent neighbourhood_group's
airbnb count
20000
15000
10000
5000
0
Manhattan
Brooklyn
Queens
Bronx
Staten Island
neighbourhood_group

Percentage wise Airbnb distribution across various neighborhood groups

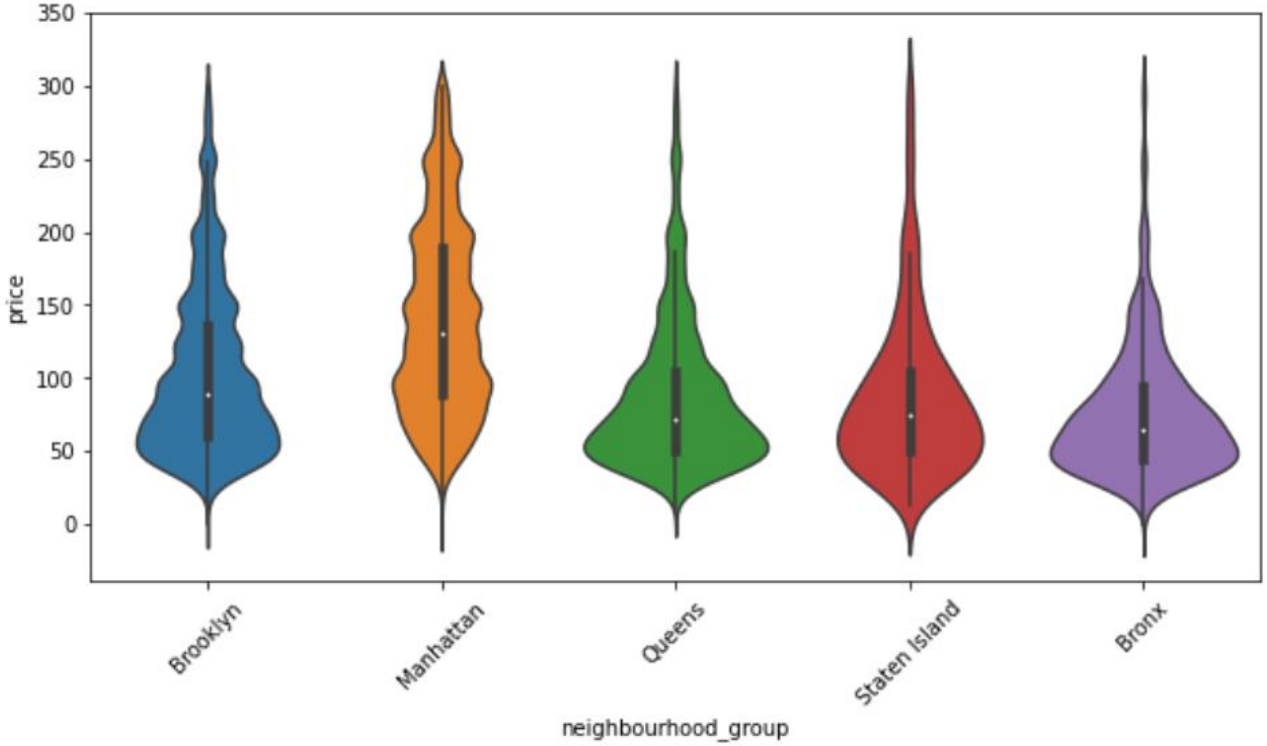


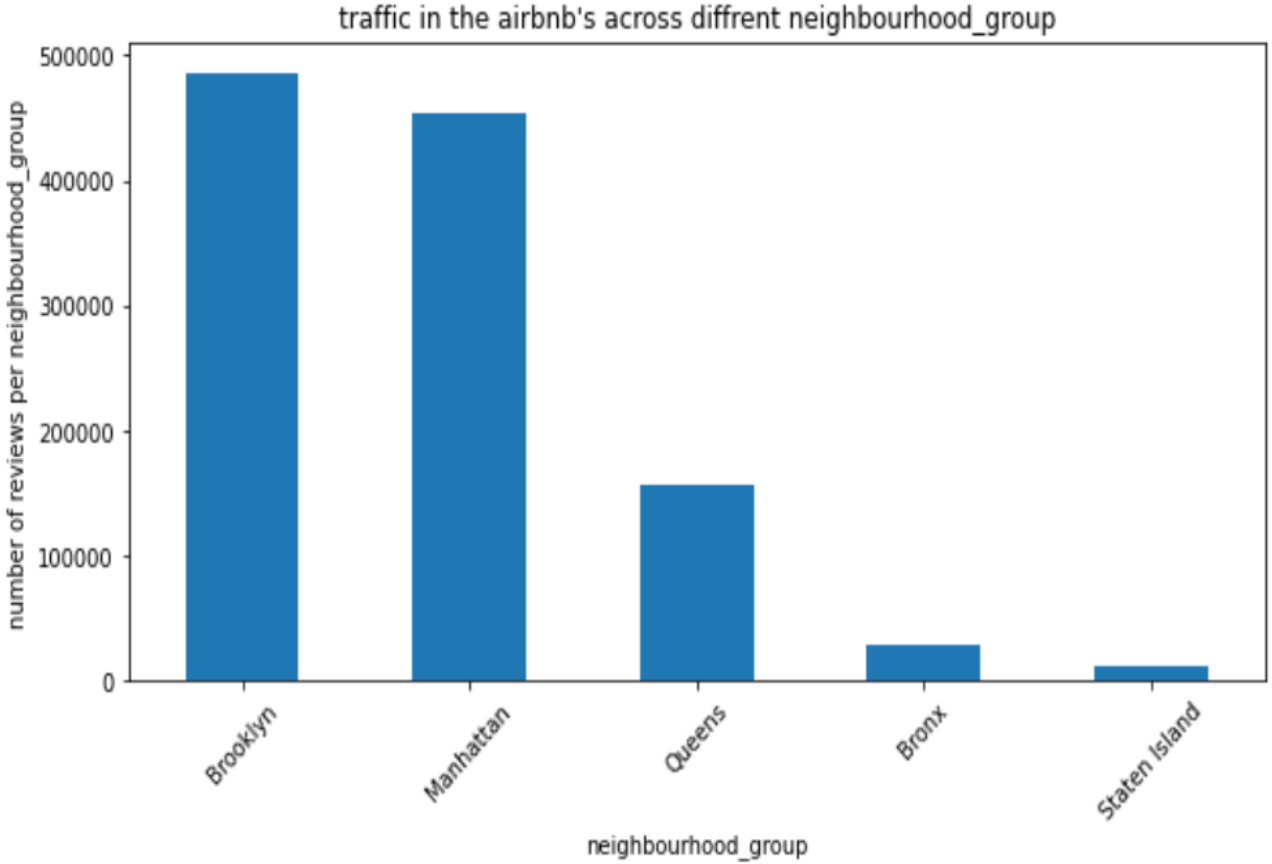

Average price of Airbnb across different Neighborhood group
Average Price
200
175
150
125
100
75
50
25
0
Bronx
Brooklyn
Manhattan
Queens
Staten Island
neighbourhood_group

Variation of price with geographic location

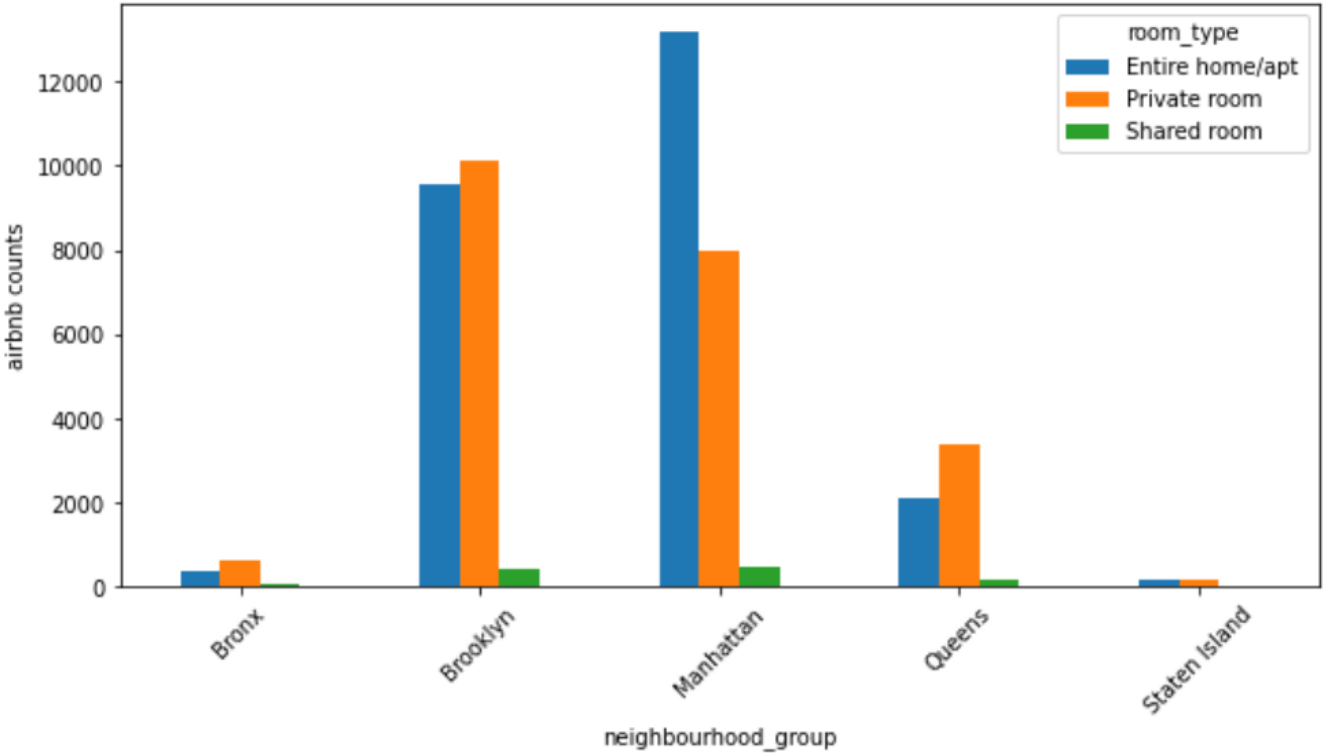


Price distribution across various neighborhood groups in New York

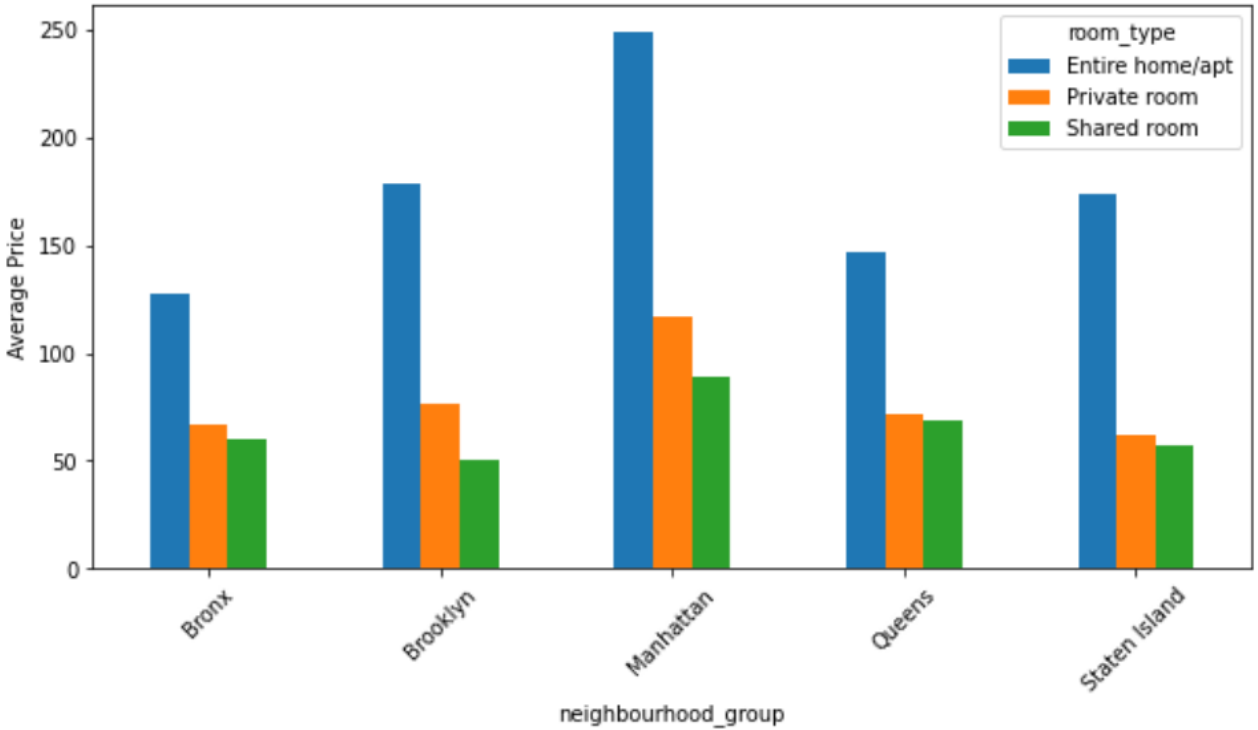


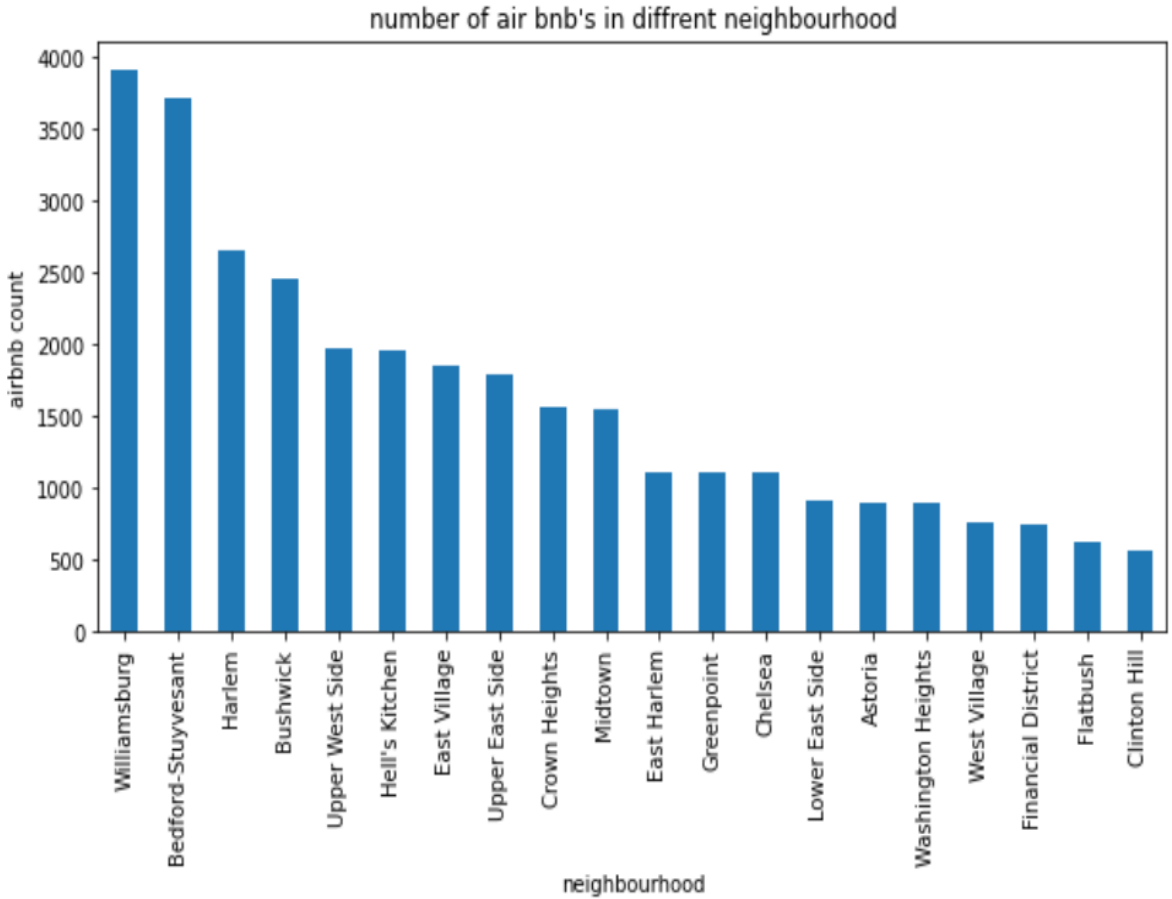

traffic in the airbnb's across diffrent neighbourhood_group
number of reviews per neighbourhood_group
500000
400000
300000
200000
100000
0
Brooklyn
Manhattan
Queens
Bronx
Staten Island
neighbourhood_group

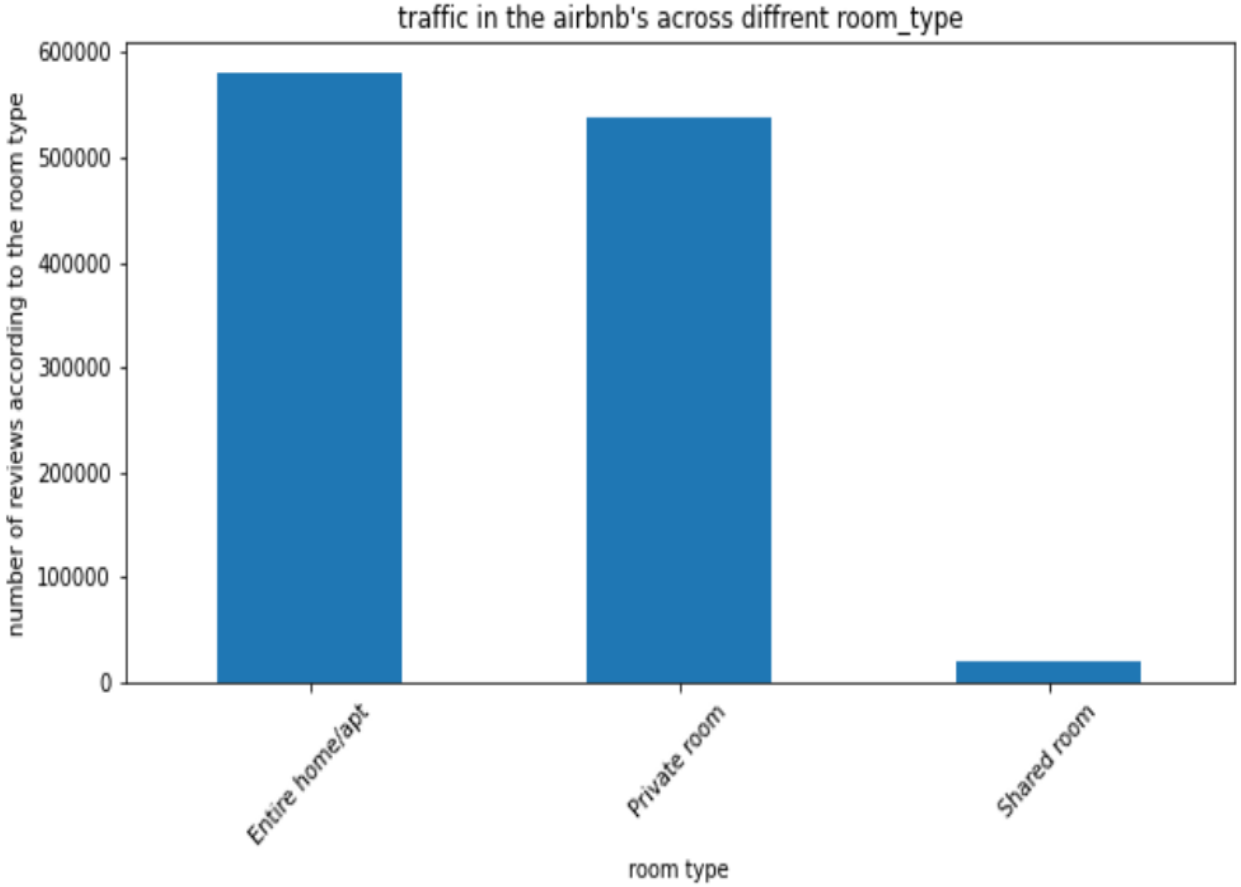
Airbnb count in neighborhood group according to room type

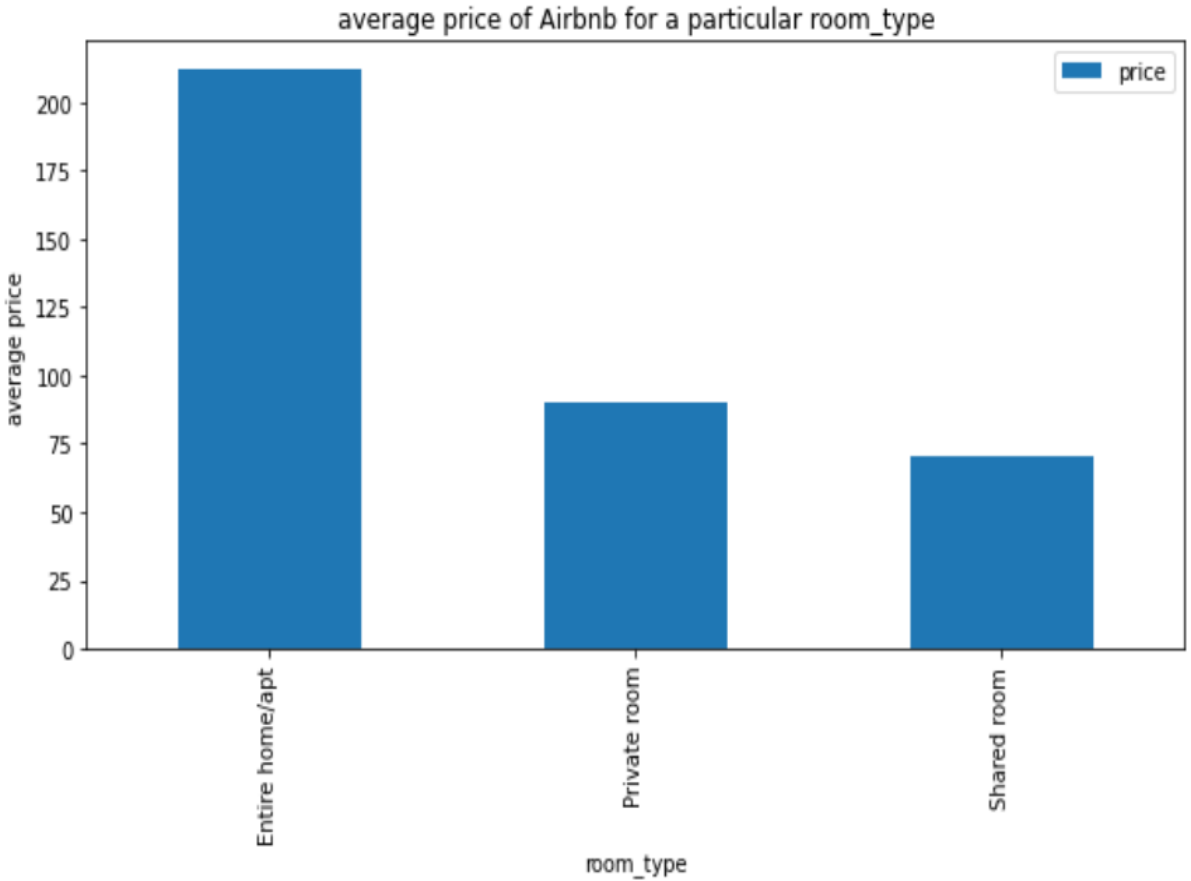


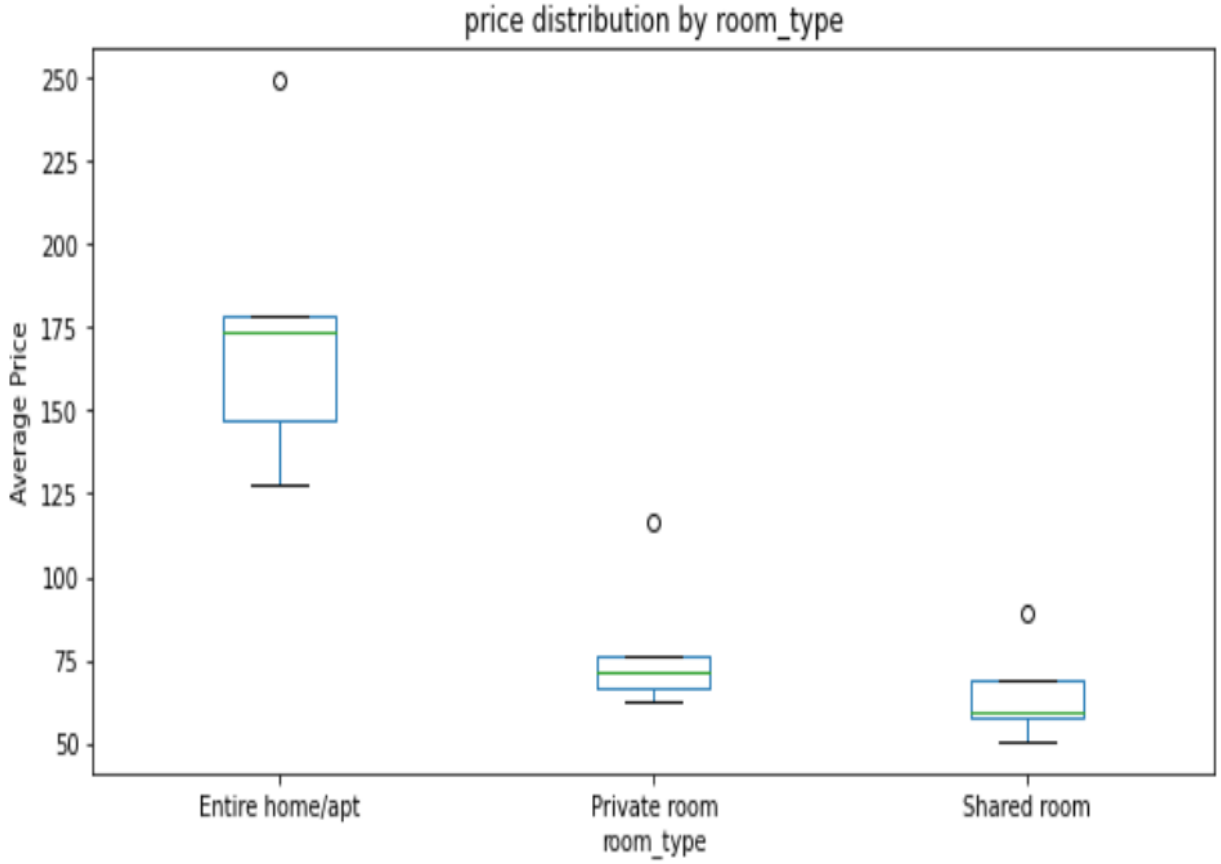
Average price of Airbnb in neighborhood group according to room type




number of air bnb's in diffrent neighbourhood
airbnb count
0
500
1000
1500
2000
2500
3000
3500
4000
Williamsburg
Bedford-Stuyvesant
Harlem
Bushwick
Upper West Side
Hell's Kitchen
East Village
Upper East Side
Crown Heights
Midtown
East Harlem
Greenpoint
Chelsea
Lower East Side
Astoria
Washington Heights
West Village
Financial District
Flatbush
Clinton Hill
neighbourhood


traffic in the airbnb's across diffrent room_type
number of reviews according to the room type
600000
500000
400000
300000
200000
100000
0
Entire home/apt
Private room
Shared room
room type


average price of Airbnb for a particular room_type
price
average price
200
175
150
125
100
75
50
25
0
Entire home/apt
Private room
Shared room
room_type


price distribution by room_type
Average Price
250
225
200
175
150
125
100
75
50
Entire home/apt
Private room
Shared room
room_type

Correlation between price and number of reviews

# Conclusion

- First, we checked for hosts which has the maximum number of properties and hosts which has the maximum number of customer. By taking the number of reviews into account. we found out that hostname MAYA(ID-37312959) and SONDER(ID-219517861) has the maximum number of customers (2273) and properties (327) respectively.
- After that, we explore which neighborhood group has the maximum number of listing and its corresponding average price. we found out 'Manhattan' has 44.3% of the total Airbnb listing and higher average listing price among others and has a maximum number of apartment. This is because it has a good amount of tourist attraction and companies located.

We do further analysis on the number of reviews across different zones and found out that Brooklyn has the maximum number of reviews that corresponds to more number of the customer. We plotted a categorial bar graph to dig out the reasoning and found that Brooklyn has the maximum number of private rooms available and the average price is lower compared to Manhattan. So, most singles and student preferred this zone. Also, we plotted neighborhood plots for Airbnb counts and found that Williamsburg and 'Bedford-Stuyvesant' has the highest Airbnb counts and it belongs to the Brooklyn zone.

We performed some additional analysis on people's preferences for different properties types and correlation between price and number of reviews and found that people preferred apartments then private or shared rooms in New York. this result in a higher average price of an apartment than its intrinsic value.

Also, the correlation between price and reviews is negative indicating higher price having fewer people preference. overall, we developed a good insight from the Airbnb dataset. This analysis helps us to perform a better business decision, future marketing strategy and much more.

# Challenges

- To select out which column will be relevant which to discard for analysis involves lots of brainstorming
- To filter out the data having NAN values

# Reference


- Google.com
- Kaggle.com
- Stackoverflow.com

# GitHub Repository

https://github.com/sakshiii819/Airbnb-Booking-Analysis-Sakshi-Juneja_

https://github.com/roshan10597/Airbnb-booking-analysis--Roshan-Tile.git

https://github.com/jainrasik/Airbnb-Booking-Analysis-Rasik-Jain.git

https://github.com/24nik/Airbnb-Booking-Analysis-Nikhil-Dutt.git