

# Capstone Project-1

## Airbnb NYC Booking Exploratory Data Analysis

### Team Members

Amlana Jyoti Biswal

Varnit Chauhan

# CONTENTS

The logo consists of the letters 'AI' in white, bold, sans-serif font, centered within a solid red square.

- ❑ Introduction
- ❑ Data Summary
- ❑ Exploratory Data Analysis
  - ❑ Statistics on Host
  - ❑ Availability of Airbnb
  - ❑ Statistics on Price
  - ❑ Statistics on Neighbourhood and Neighbourhood Group.
  - ❑ Statistics on Location
  - ❑ Geospatial visualisation of top 200 Airbnbs.
- ❑ Conclusion and Business Insights .

# Introduction

- ❑ Airbnb is an American company that operates online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. 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 Gebia.
- ❑ Here we are going to do an Exploratory Data Analysis on the dataset of Airbnb NYC Bookings.

# Data Summary

- **airbnb** : This dataset contains the data of Airbnb NYC 2019, it includes information such as the location, neighbourhood of the properties, room type, price, minimum nights etc.
- **max\_listing\_hosts**: This is a sub-dataset of airbnb dataset, it contains top 10 hosts with their listing counts.
- **most\_reviewed\_hosts**: This is a sub-dataset of airbnb dataset, it contains top 10 hosts with maximum count of reviews per month.
- **room\_availability** : This dataset shows the average availability of each room type.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                     48895 non-null  int64
1   name                                  48879 non-null  object
2   host_id                               48895 non-null  int64
3   host_name                             48874 non-null  object
4   neighbourhood_group                   48895 non-null  object
5   neighbourhood                         48895 non-null  object
6   latitude                             48895 non-null  float64
7   longitude                             48895 non-null  float64
8   room_type                             48895 non-null  object
9   price                                 48895 non-null  int64
10  minimum_nights                        48895 non-null  int64
11  number_of_reviews                     48895 non-null  int64
12  last_review                           38843 non-null  object
13  reviews_per_month                     38843 non-null  float64
14  calculated_host_listings_count        48895 non-null  int64
15  availability_365                       48895 non-null  int64
dtypes: float64(3), int64(7), object(6)
memory usage: 6.0+ MB
```

## Data Summary continued....

- **availability\_neighbour\_df** : This dataset shows the average availability of Airbnb in neighbourhood groups.
- **numeric\_df** : This dataset shows the correlation of price with independent variables.
- **avg\_stay**: This dataset contains the average stay night groupby neighbourhood groups and then room type.
- **hotel\_location**: This dataset contains the latitude and longitude of top 200 Airbnb by their number of reviews.

# Exploratory Data Analysis

## Stats on Hosts

A red square logo with the white letters 'AI' inside.

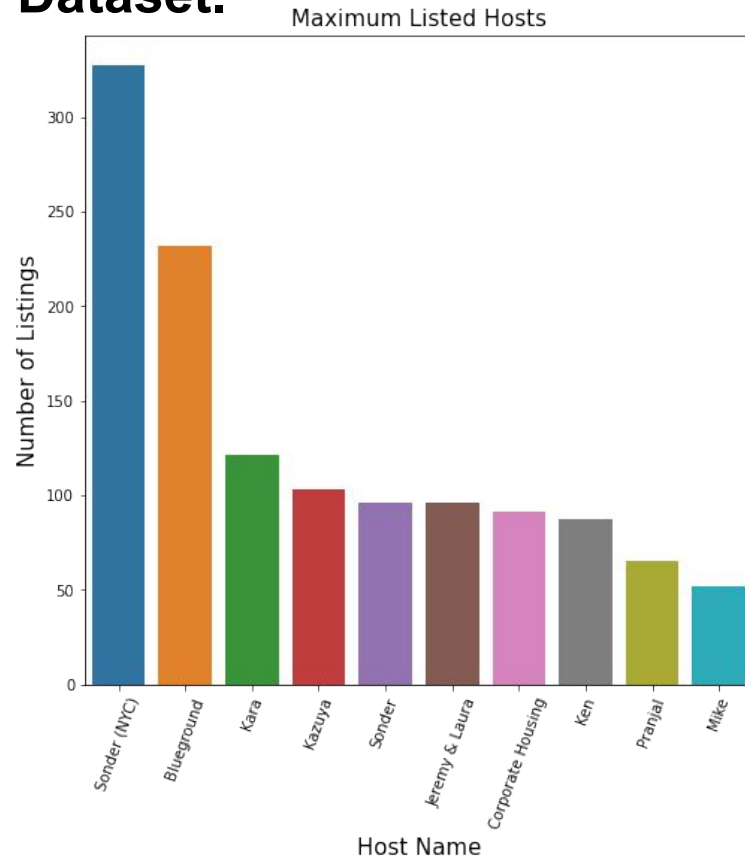
In this section we will analyze the data on host column, and will get the answer for the following questions :-

1. Top 10 hosts with maximum listing in Dataset.
2. Which host is busiest ?
3. Top 10 hosts vs room type.
4. Who is the most reviewed hosts among all.

## EDA

## Top 10 hosts with maximum listing in Dataset.

- Sonder (NYC) is maximum listed host with 300+ listing.
- Blueground stands at second with approx 230 listing
- Kara stands at third with approx 120 listing.
- Mike is least listed host among top 10 hosts in terms of most listed hosts.

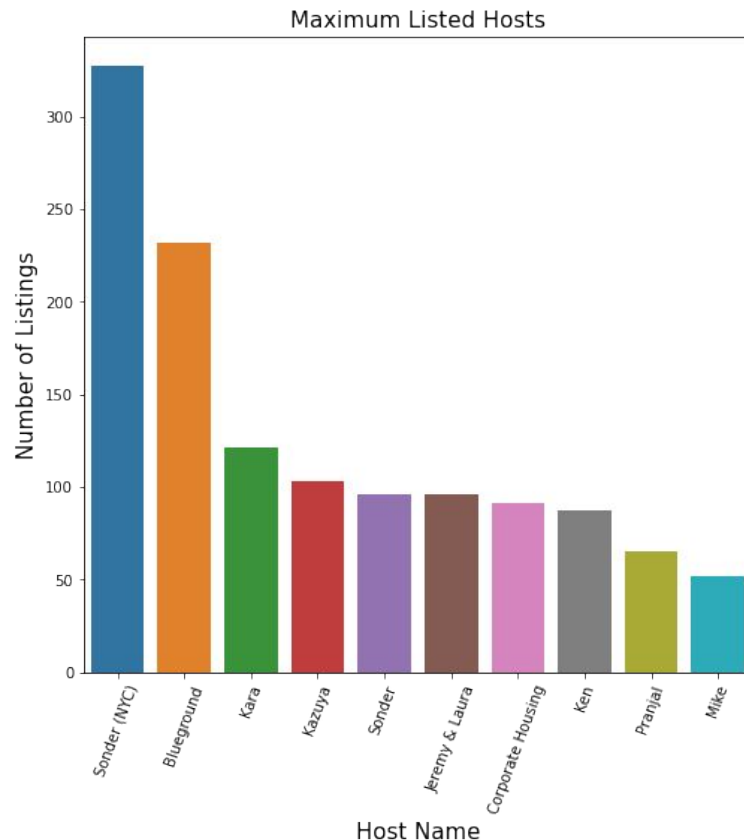


## EDA

## Which host is busiest ?

In this bar plot we can clearly see that Sonder (NYC) is maximum listed host with 300+ listing. Maximum listing means it is the most busiest host among all.

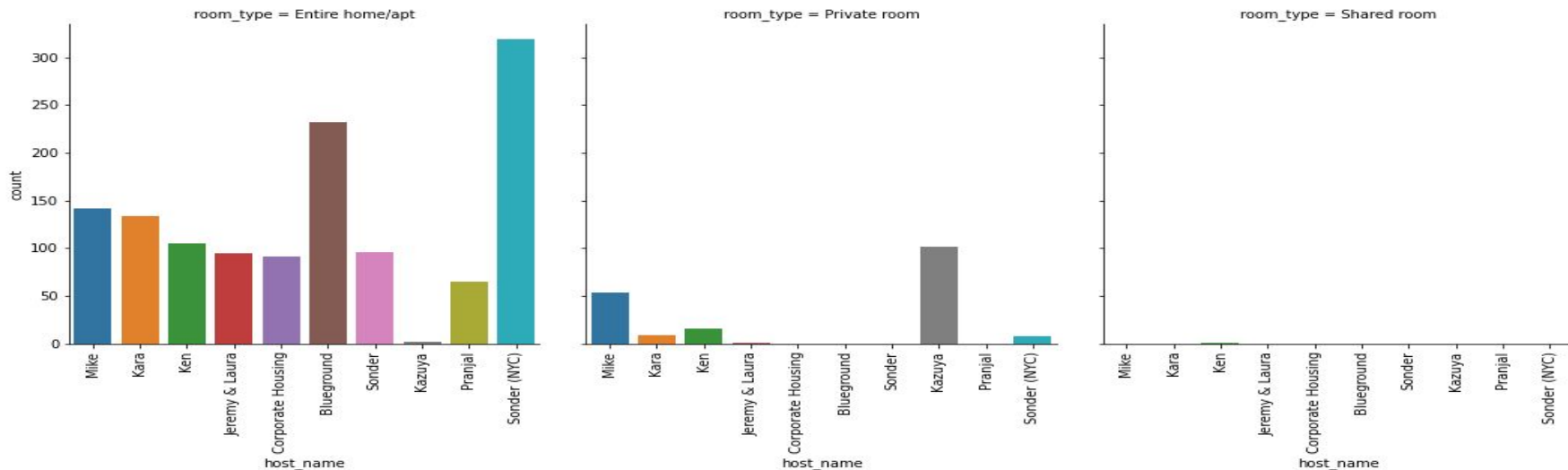
So, Sonder (NYC) is most busiest host.





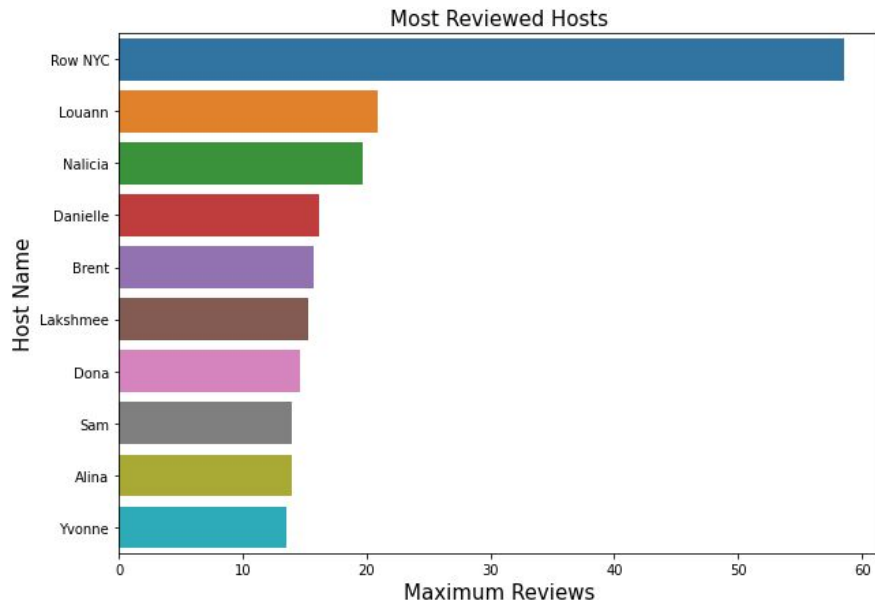
## Top 10 Hosts vs their Room type

- From the below plots we can conclude that Sonder(NYC) has the maximum number of counts in Entire home/apt room type.
- kazuya has most number of listings in private room category, and shared room category deals with least number of bookings.



## Who is the most reviewed host among all ?

- In the bar plot of hosts and maximum number of reviews per month, we can see that Room NYC has maximum number of reviews, We can say that this might be the popular Airbnb among guests.
- But we can not say that for sure because we do not know the nature of reviews, are they positive or negative reviews.



# Exploratory Data Analysis

## Availability in Airbnb

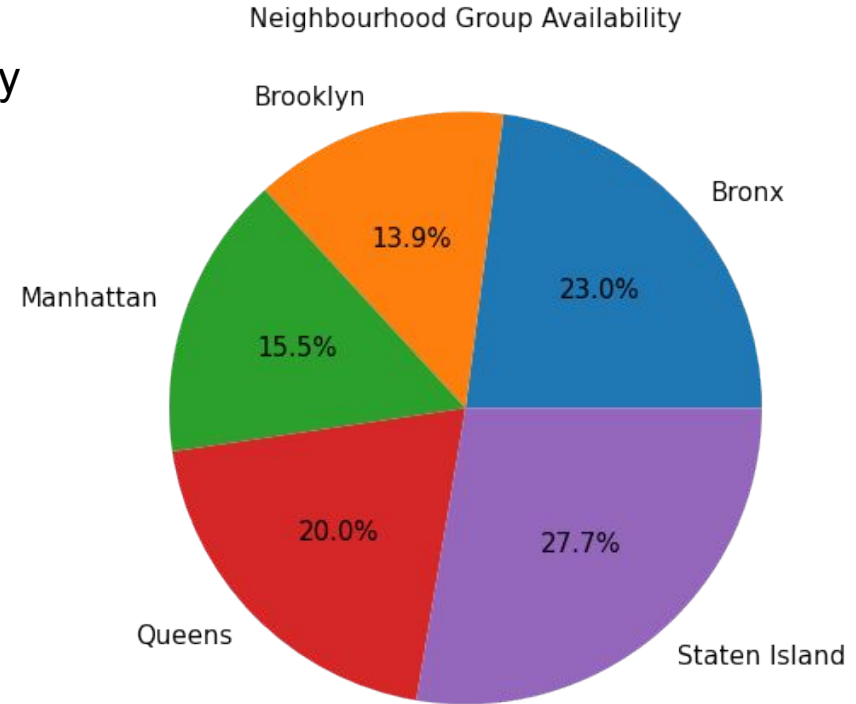
In this section we will analyze the data of availability of room, neighbourhood etc. and will get the answer for the following questions :-

1. Availability of Airbnb in neighbourhood groups.
2. Box plot showing room availability in neighbourhood group.
3. Availability of different room types.
4. Average stay of neighbourhood group in each room type.

## Availability of neighbourhood groups

Here the graph is plotted on average availability of neighbourhood group per year.

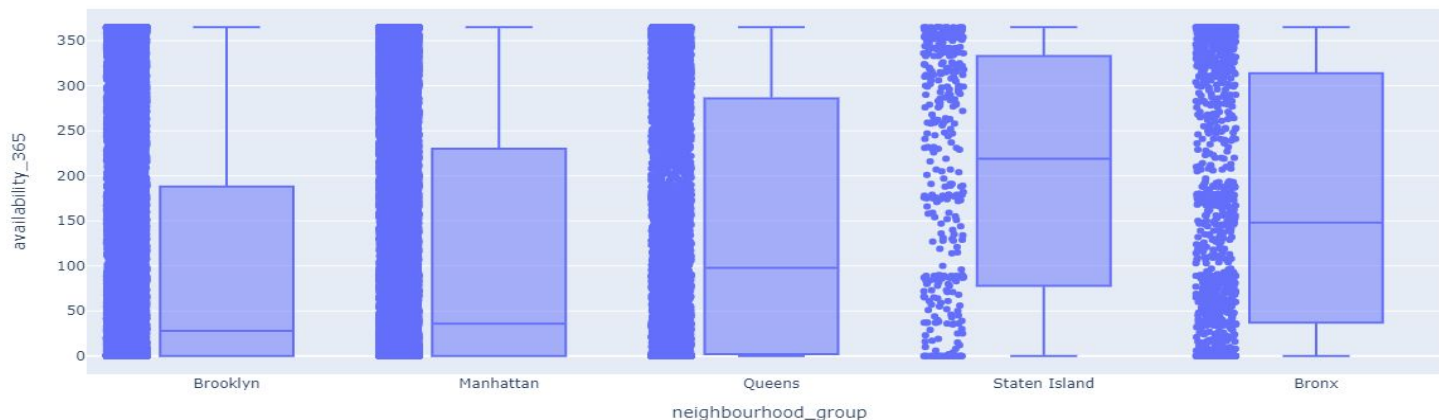
- Staten Island is the most available group among all with 27.7% availability.
- Brooklyn is least available group with approx 14% of availability.
- Manhattan is available 15.5%.
- Bronx and Queens are available 23% and 20% respectively.



## Box plot showing room availability in neighbourhood group

- We can get the value of Q1, Q3 and Median from the plot. The points beside each Box plot shows the rooms in that area.
- From this, we can conclude that room availability is maximum for the Staten Island, followed by Bronx. Room availability is lower for Brooklyn means it is one of the busiest areas. The demand for rooms in Brooklyn is higher than in others. This can help in deciding where to invest more, and acquire new hotels for Airbnb. This will help in taking critical business decisions.

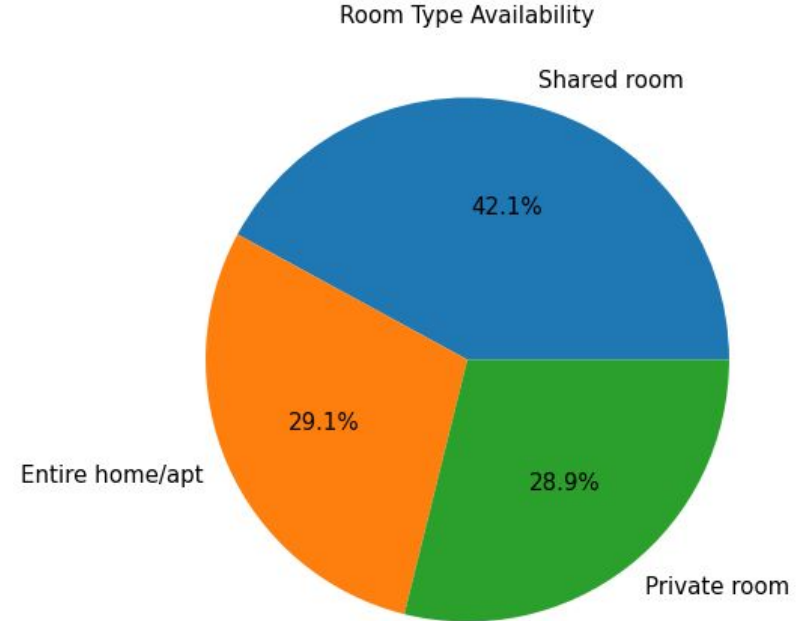
Neighbourhood Vs Room Availability



## Availability of different room types

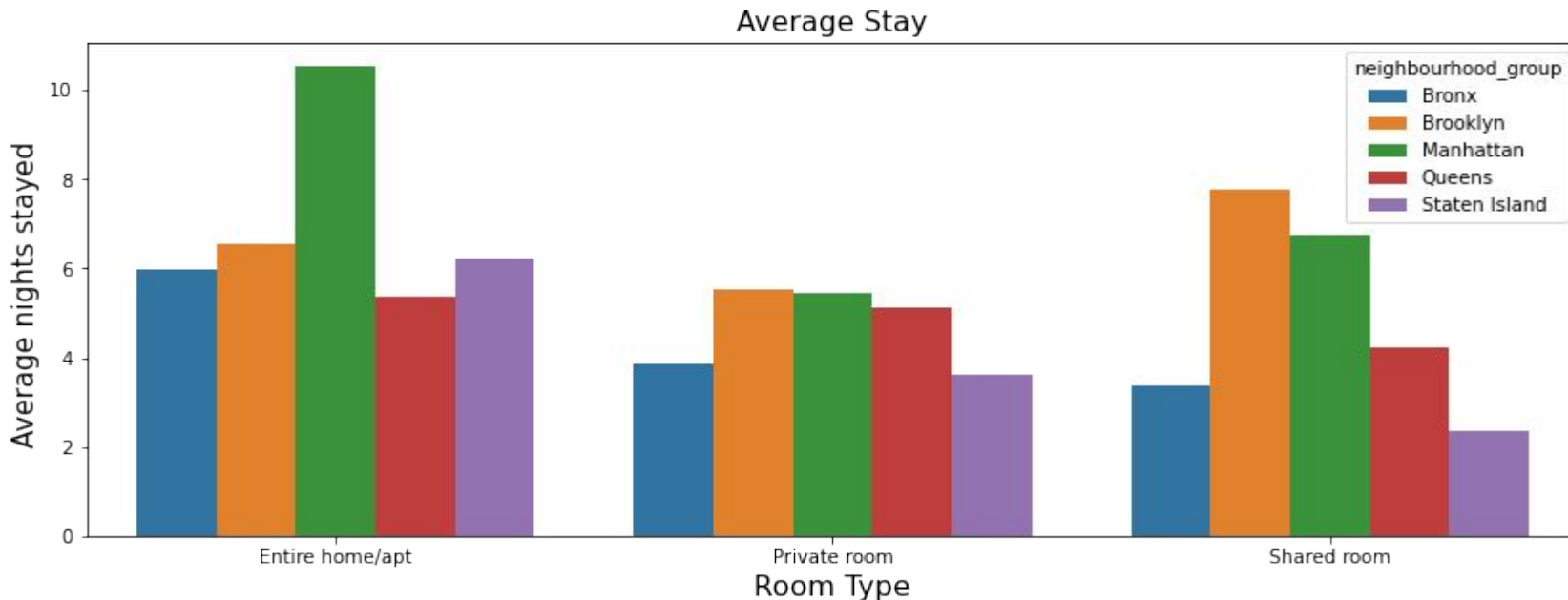
Here we have taken Mean availability for plotting

- Shared rooms have availability of about 42% which makes it the most available room type throughout the year.
- Both Entire home/apartment and private rooms have almost similar availability with 29.1% and 28.9% respectively.
- This suggests that Entire Home/apartment and Private rooms have high demand among guests.



## Average Stay of neighbourhood group in each Room Type

- Most people likely to stay longer in entire homes/apartments with an average of 8 nights
- The next thing guests prefer to stay longer are the shared rooms



# Exploratory Data Analysis

## Statistics of Price

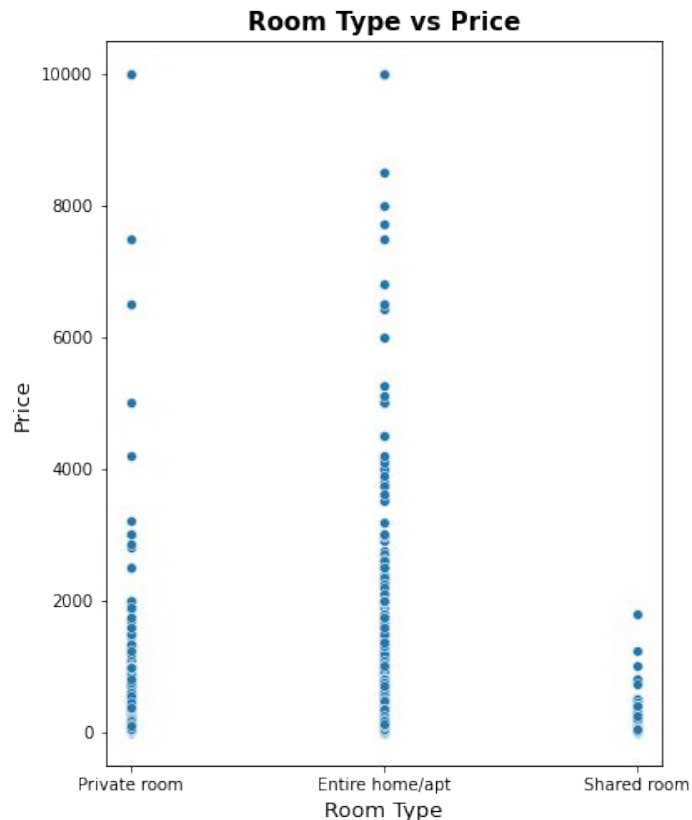
In this section we will analyze how price is related to the different room types, neighbourhood groups and median price comparison of different datas.

1. Room type vs range of price.
2. Median price comparison of 5 neighbourhood groups.
3. Median price comparison of room types.
4. Median price comparison of top 15 neighbourhoods.
5. Correlation of price with other independent variables.



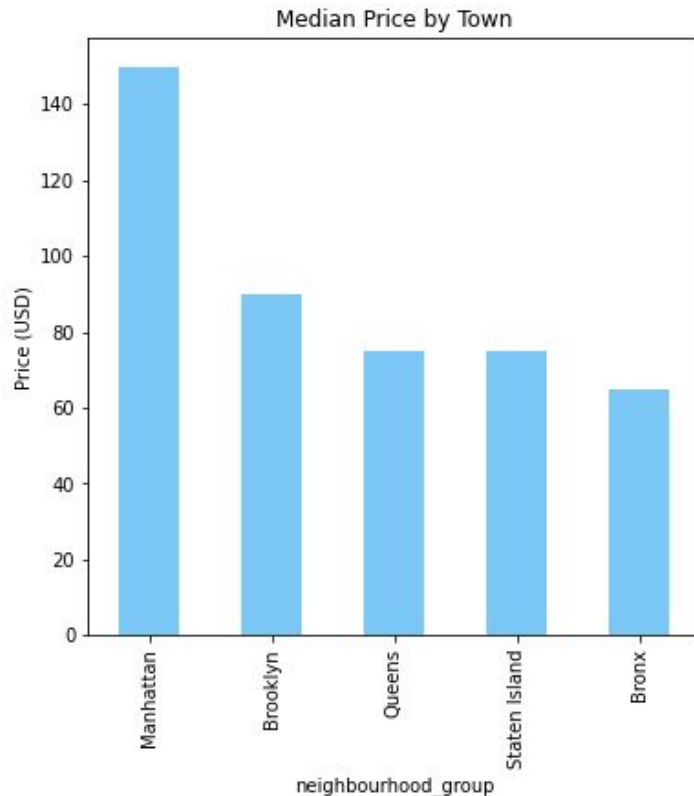
## Room type vs range of price

- Most expensive room of Private room is of \$10000.
- Most expensive room of Entire home/apt room is of \$10000.
- Most of the rooms of Entire home/apt room is at the range of \$2000.
- Entire home/apt have various type of range in price.
- Shared room have least variety in price.



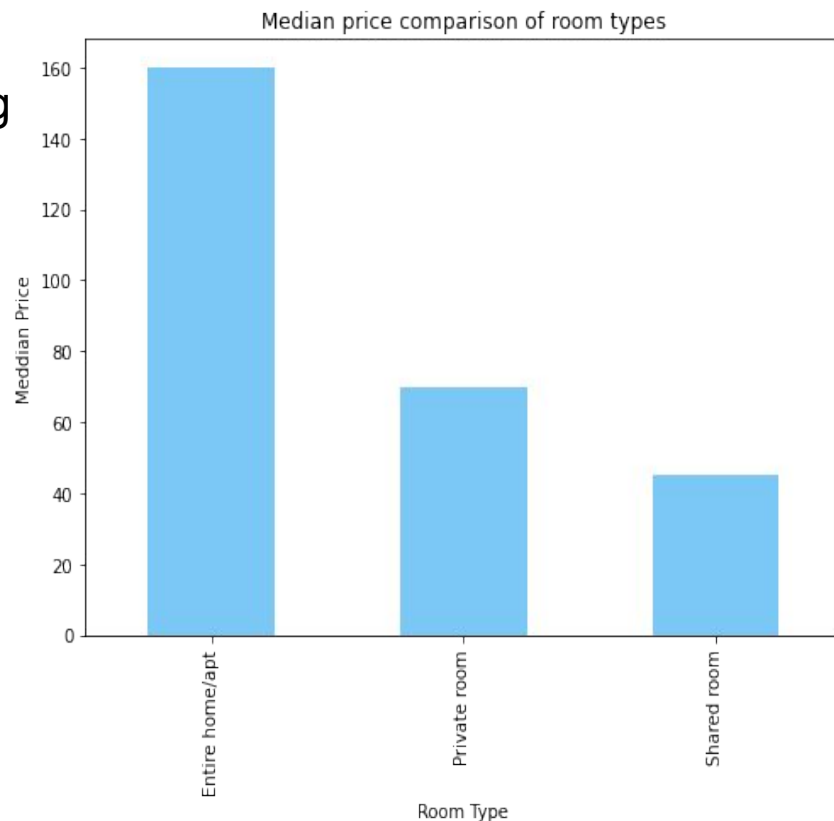
## Median Price Comparison of 5 Neighbourhood Group

- From the Bar graph, we can see that Manhattan has the highest median price of 150, followed by Brooklyn with a median price of 90.
- Queens and Staten Island have the same median price of 75.
- The Bronx is on the bottom of the list with a median price of 65.
- From these data we can say that Manhattan is the costliest neighbourhood among all.



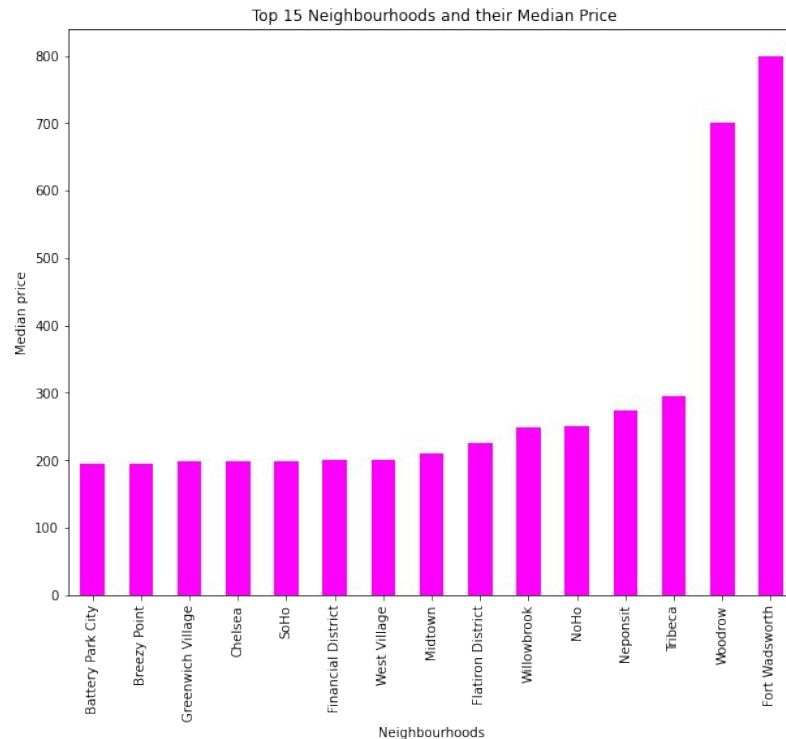
## Median Price Comparison of room types

- Entire Home/ apt is most expensive among all the room types, Median price for Entire home/apt is \$160.
- Median price for Private room is \$70.
- Median price for Shared room is approx \$45.



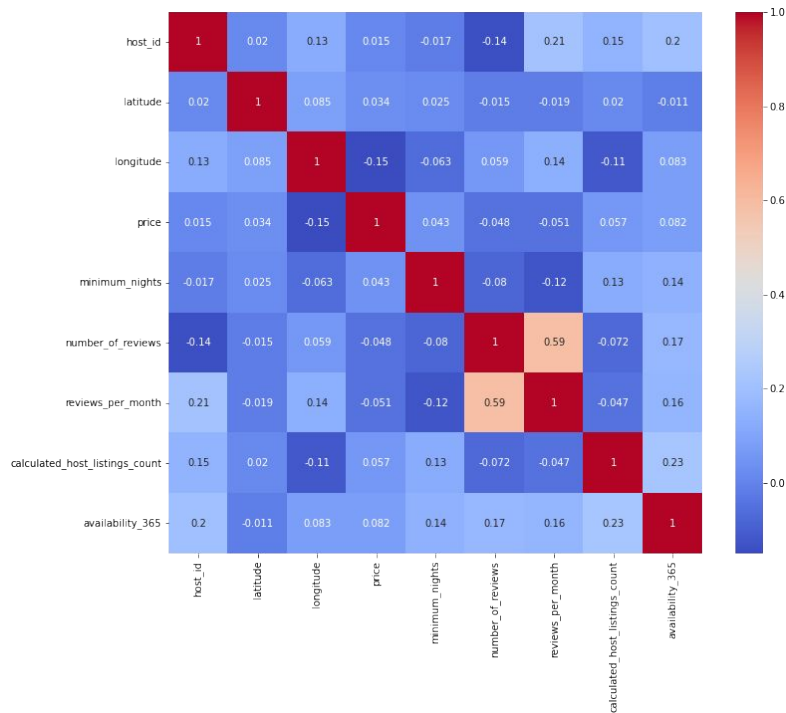
## Median Price Comparison of Top 15 Neighbourhoods

- From the Bar graph of median price comparison of the top 15 neighbourhoods, we can deduce that Fort Wadsworth has the highest median price of 800, and Battery Park City is at the bottom of the list.



## Correlation of price with other independent variables

- We can see that price does not have any strong relationship with any independent variables.
- Also, there is no relationship among independent variables.
- Thus we can streamline our analysis toward categorical data with independent variables.



# Exploratory Data Analysis

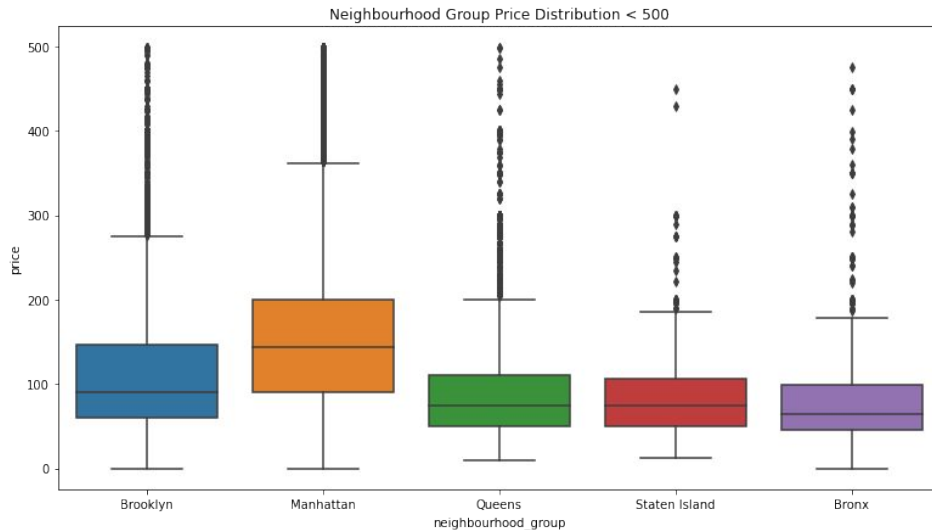
## Statistics on Neighbourhood and Neighbourhood Group

In this section we are going to study neighbourhood group price distribution, reviews and their listings.

1. Neighbourhood group price distribution  $< 500$ .
2. Neighbourhood group total price and reviews.
3. Neighbourhood listing.

## Neighbourhood group price distribution < 500

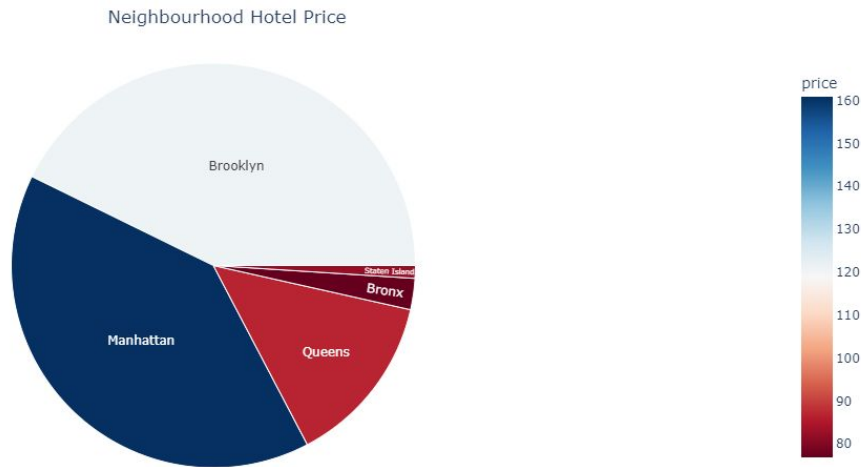
- We can state that Manhattan has the highest range price for the listings with about 140 as an average price, followed by Brooklyn with 90 per night.
- Queens and Staten Island seem to have a very similar distribution.
- The Bronx is the cheapest.



## Neighbourhood group total price and reviews

- In the pie chart size of the group shows the number of reviews they have and the colour is showing the price range.
- We can see that Manhattan is at the top of the price range which is followed by Brooklyn. The Bronx is at the bottom of the price chart, where the cheapest Airbnbs are located.

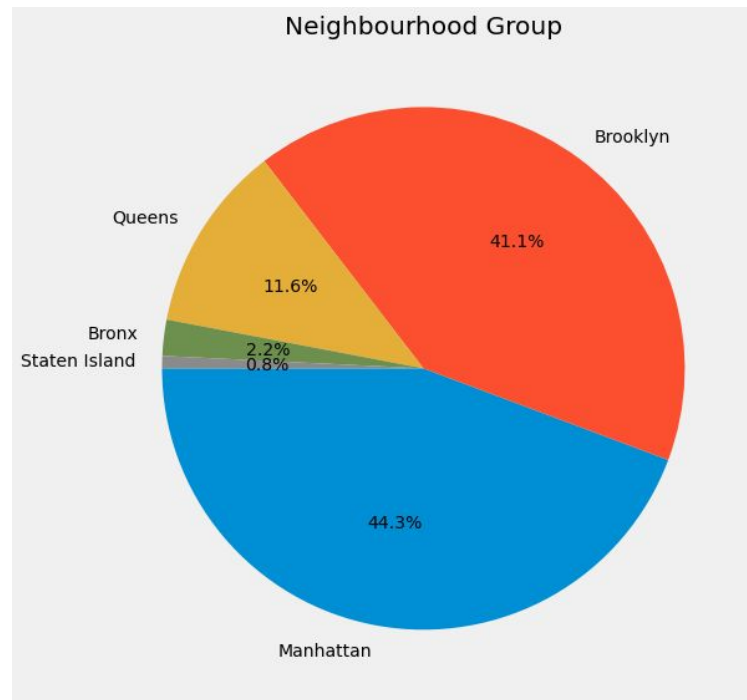
This is an interactive pie chart but we have attached the static image here.





## Neighbourhood listing

- This pie chart shows the percentage of Airbnb listings in the Neighbourhood group.
- From the pie chart, we can see that Brooklyn and Manhattan share the maximum number of listings.



# Exploratory Data Analysis Statistics on Location

A red square logo with the white letters 'AI' inside.

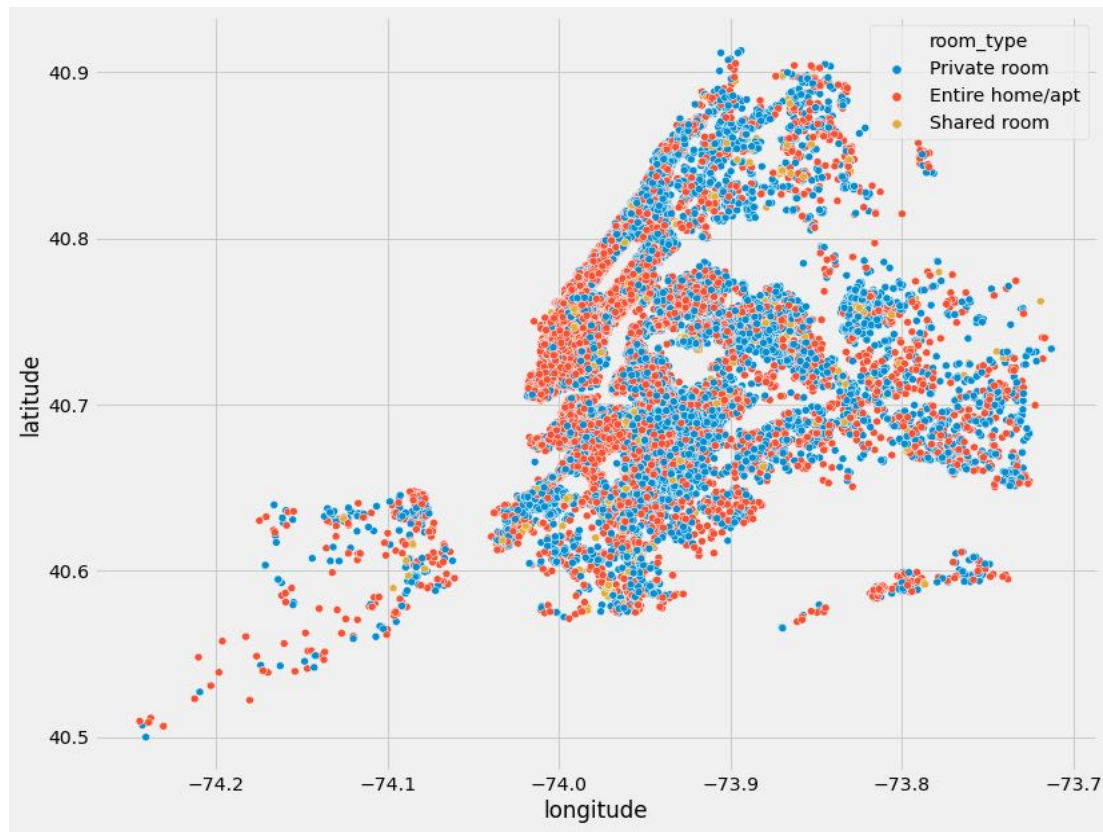
In this section we will see the relation of location and different parameters.

1. Location of different room type according to their latitude and longitude.
2. Neighbourhood groups by location, price and availability.
3. Geospatial visualization of top 200 Airbnb according to their reviews.

## EDA

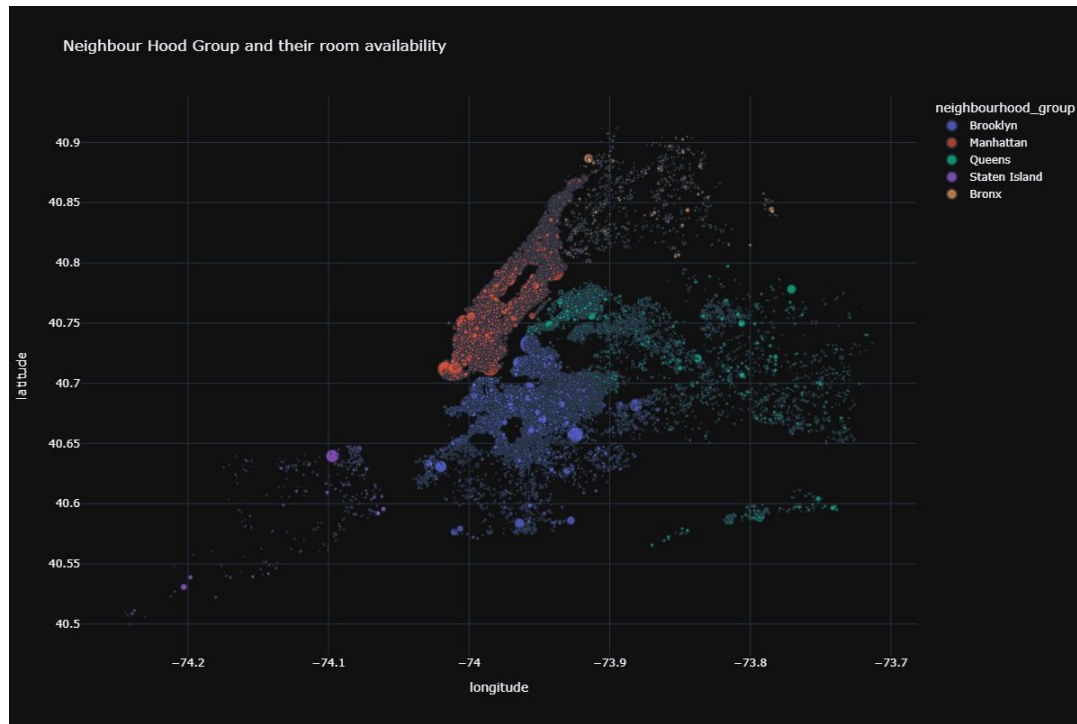
AI

# Location of different room type according to their latitude and longitude



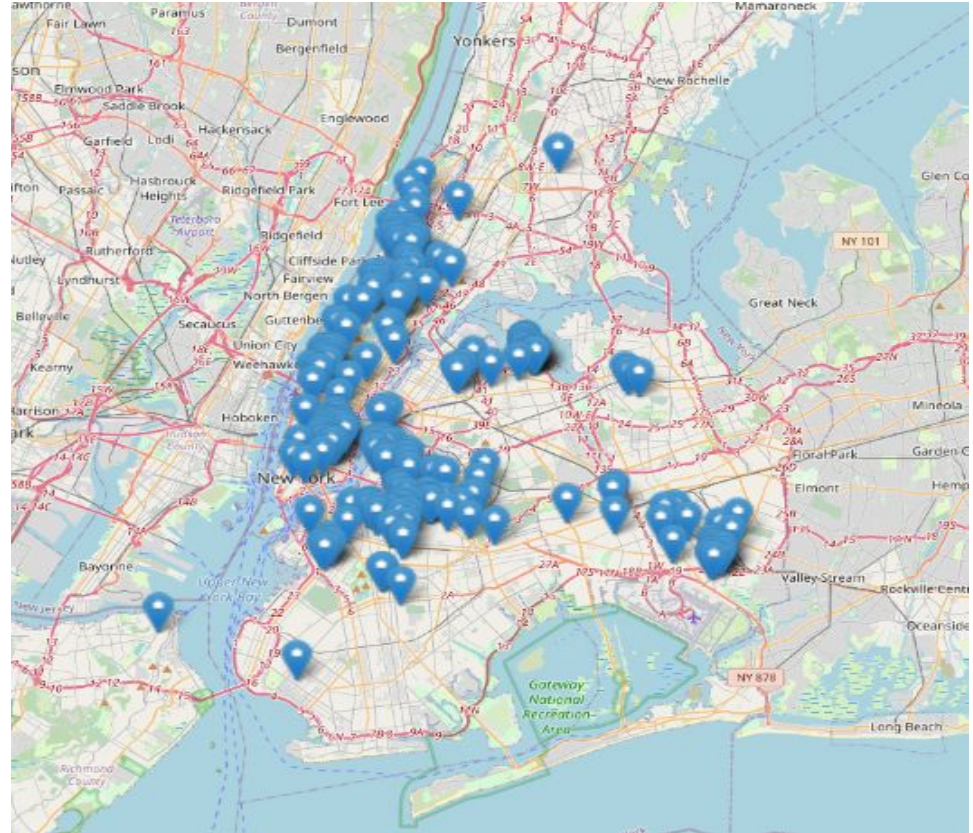
## Neighbourhood groups by location, price and availability

- We have plotted the scatter plot using the latitude and longitude of the Airbnbs. They are colour coded with different colours according to their Neighbourhood group.
- The size of the scatter points indicates their price, the larger the circle the more expensive it is.
- On hovering over the point we can get their name, availability, price, number of reviews, latitude and longitude.



## Geospatial visualization of top 200 Airbnb according to their reviews

- From the Map, we can conclude that most of the top Airbnb are located in the Manhattan area. Airbnb in this area most probably receives more customers than other Airbnbs. That's why they have more number of reviews.



# Conclusion



Following are the conclusion of our EDA :-

- Most expensive room are of Private room and Entire home/apt room type with price of \$10000.
- Sonder(NYC) is the most listed host with more than 300 listings.
- All the top 10 listed host are dealing in Entire home/apt room type and only host Ken is dealing in Shared room type with only listing of approx 1.
- Row NYC is the most reviewed host with approx 58 reviews per month.
- Shared room available 42% of the time which is highest in all room types.
- Staten Island is the most available neighbourhood group among all with 27.7% of availability.
- Manhattan is the most listed neighbourhood group with 44.3% of total listing.

## Conclusion continued...

- Manhattan has the highest median price of 150, followed by Brooklyn with median price of 90.
- Most people likely to stay longer in Entire home/apt with an average stay of 8 nights.
- Entire apartments, unsurprising, cost more and had more reviews compared to other room types.
- Manhattan and Brooklyn significantly had more listing than the other neighbourhood.
- These observations will help us in deciding where to invest more, and acquire new hotels for Airbnb. This will help in taking critical business decisions.
- Entire room/ Apartment and Private rooms are the most demanded room types among the guests, this might be due to the privacy issue and luxury. Because in shared rooms the chances of theft and burglary are greater than in others. That's the reason having more cost than other room types these are mostly preferred by the customers.

## Conclusion Continued.....

- From a business perspective, our suggestion would be, to invest more or acquire more hotels for Airbnb in Manhattan and Brooklyn area. The maximum number of room types should be Entire room/ Apartment and private rooms.
- Ultimately, location matters, one could argue, for Airbnb listings, but there can be several factors that may contribute such as tourist attractions, transportation facilities, nearest airport, serene geographical locations, the crime rate in the neighbourhood, nightlife in that area, other amenities like gymnasium, swimming pools, night clubs, bars, restaurants, food delivery services also affects the number of customers they receive. As these data are not available in the given data frame we can not comment on these topics.





Thank You