

Capstone Project – 1

E.D.A ON Airbnb booking

by-

Team zlearners :-

**Akshit Singh
Mishra Om**

Agenda

- Look at the Problem statements
- Study the Airbnb dataset
- Handling the null values
- Look at the correlation among variables
- Coordinates of various listings
- Popular neighbourhoods and room types
- Analysis summary

Problem Statement

- What can we learn about different hosts and areas?
- What is the availability of room types for different neighbourhood groups?
- What is the minimum number of nights people stay in different places?
- What is the average price for different neighbourhood groups?
- What is the relationship between price and the number of reviews?
- Which neighbourhood group has got the most reviews?
- Which hosts are the busiest and why?
- Is there any noticeable difference in traffic among different areas?

Data Summary

- There are around 49,000 observations with various types of field in our Dataset.
- List of columns:-

id	room_type
name	price
host_id	minimum_nights
host_name	number_of_reviews
neighbourhood_group	last_reviews
neighbourhood	review_per_month
latitude	calculated_host_listings_count
longitude	availability_365

Studying the columns of the dataset

- Dataset contains 6 columns of categorical data.
- There are 10 columns of numeric data.
- last_review and reviews_per_month contains approximately 10 thousand null values.

```
# studying the missing values of columns  
airbnb_df.info()
```

```
<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
```

Handling the null values

- Used the fillna() method to fill the null values in the dataset.
- Filled the reviews_per_month column with 0.
- Filled the rest of the columns with 'not_present'.

```
# Handling the null values
airbnb_df['reviews_per_month'].fillna(0, inplace = True)
airbnb_df.fillna('not_present', inplace = True)

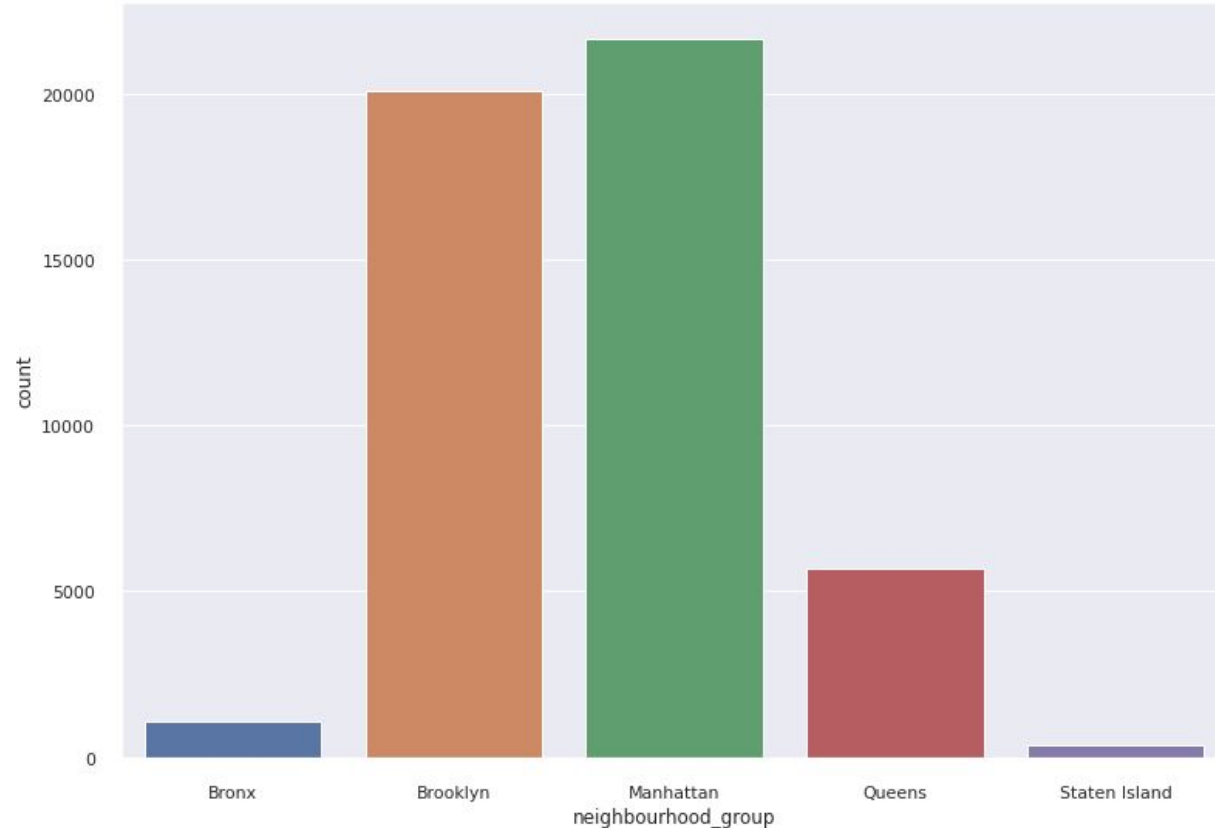
# Now checking if any null values are present in our dataset
count_of_null_values = airbnb_df.isnull().sum()
count_of_null_values
```

id	0
name	0
host_id	0
host_name	0
neighbourhood_group	0
neighbourhood	0
latitude	0
longitude	0
room_type	0
price	0
minimum_nights	0
number_of_reviews	0
last_review	0
reviews_per_month	0
calculated_host_listings_count	0
availability_365	0
dtype: int64	

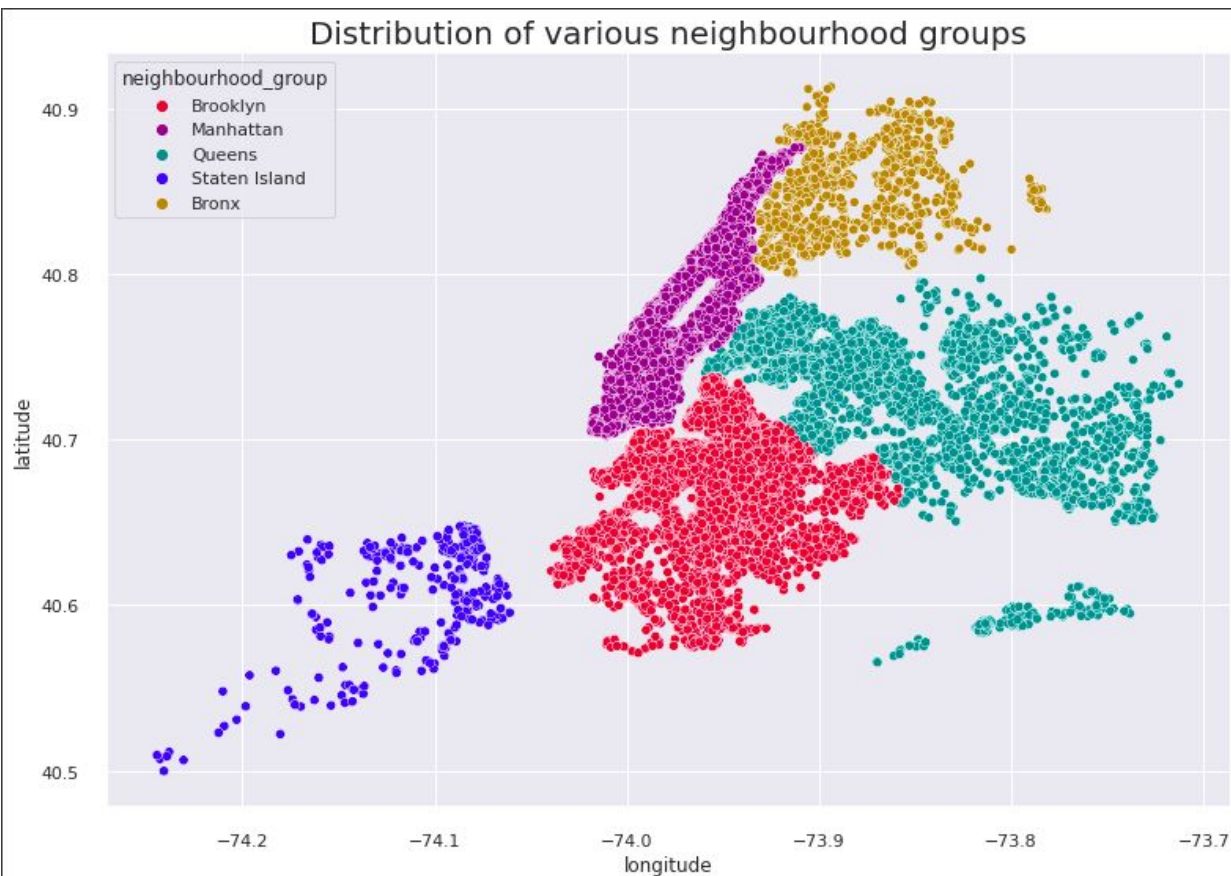
Airbnb listings in different neighbourhood groups

- Manhattan has the most no. of Airbnb listings.
- Staten Island has the least no. of Airbnb listings.

```
neighbourhood_group  count  
Bronx                1091  
Brooklyn             20104  
Manhattan            21661  
Queens               5666  
Staten Island        373
```



Coordinates of rooms in different neighbourhood groups

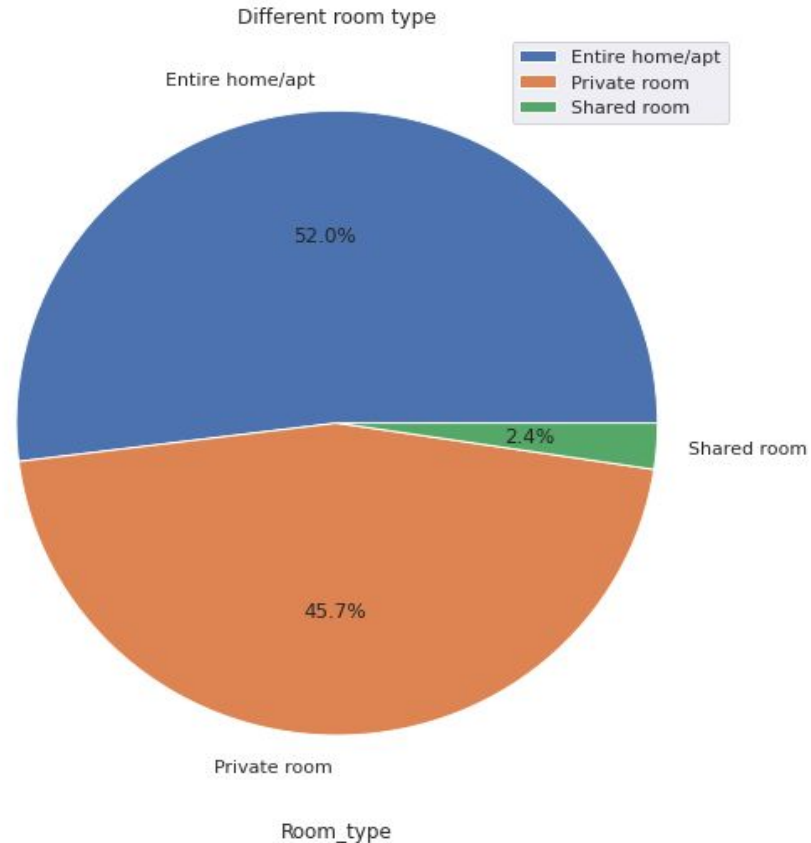


- Staten Island has the least number of rooms.
- Manhattan and Brooklyn have a fairly good amount of listings.
- Listings of Queens are widely spread.

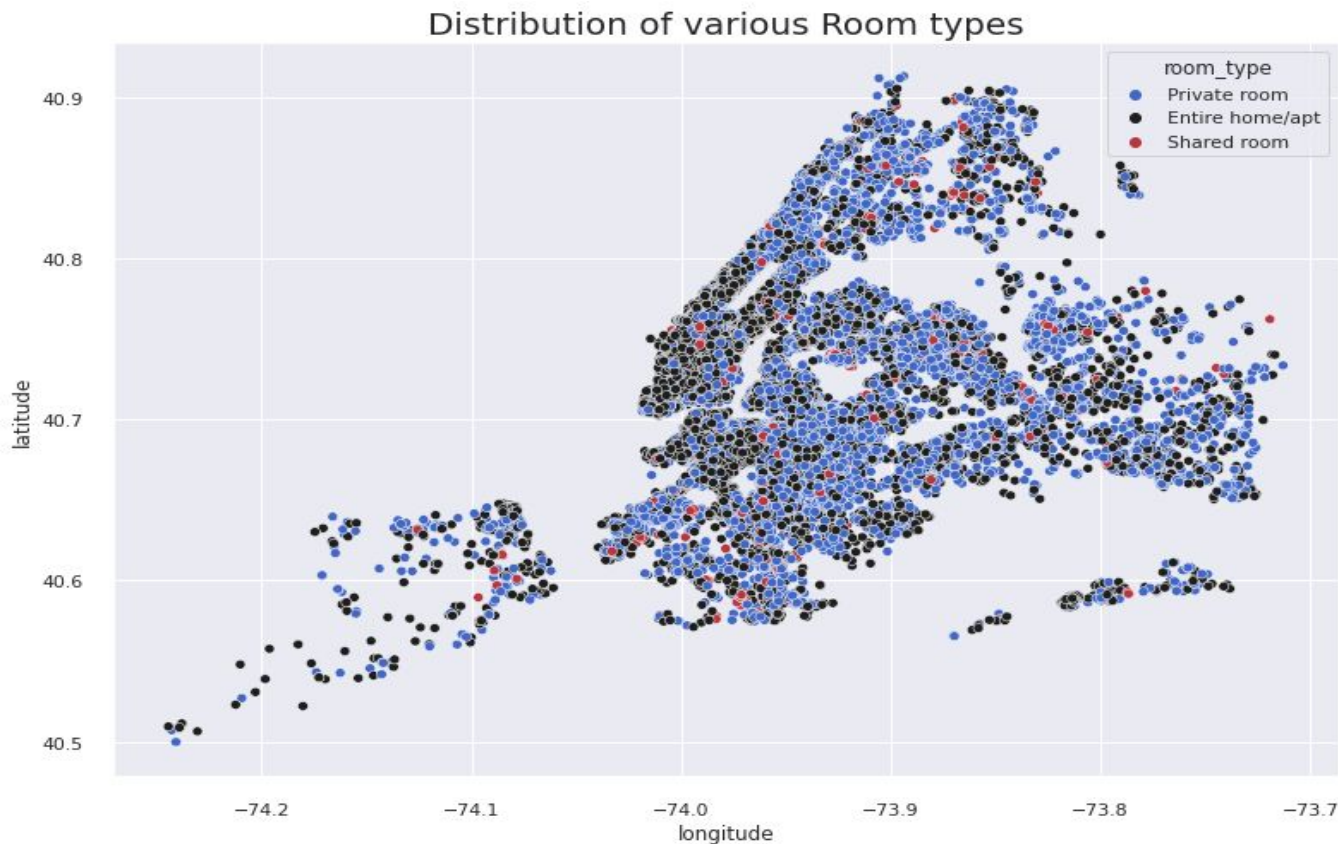
Distribution of Room types

- Shared room has the least number of listings.
- Entire Home/apt has the highest number of listings.
- Private room has a decent no. of listings.

room_type	count
Entire home/apt	25409
Private room	22326
Shared room	1160



Coordinates of various rooms for different room types

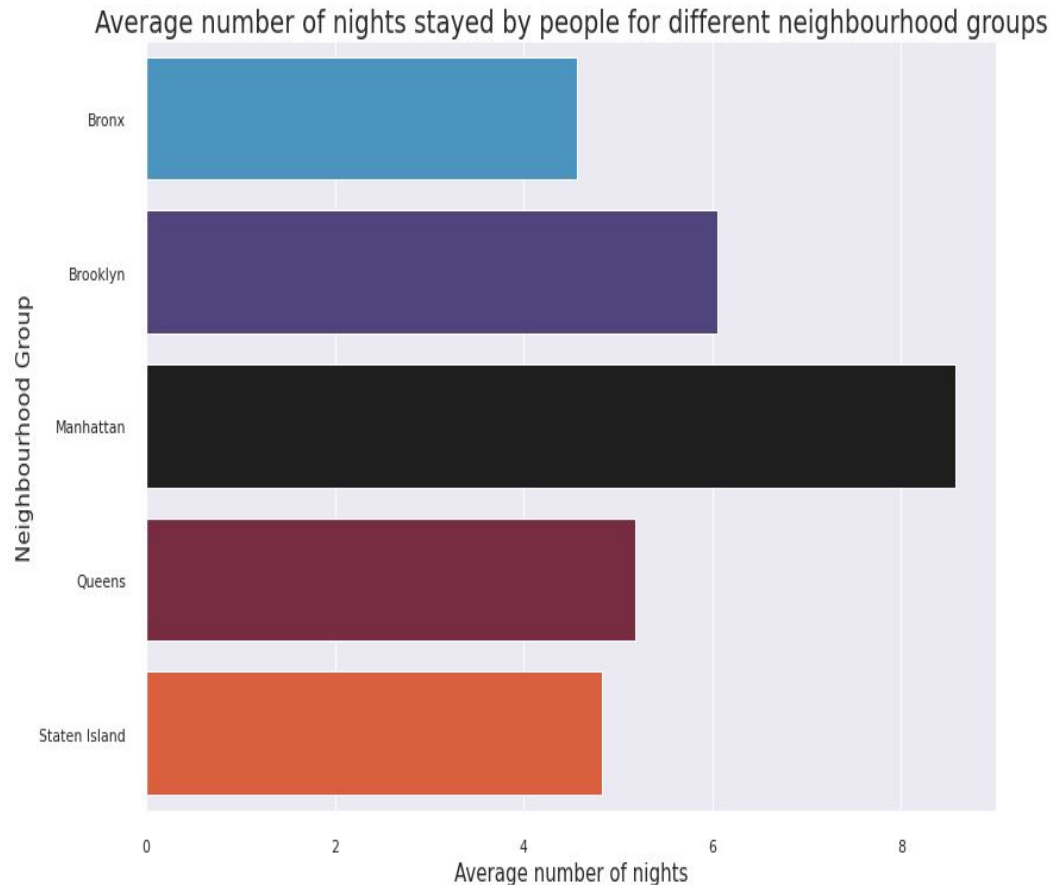


→ Shared rooms are least in number.

→ Entire home/apt and Private Rooms have a good amount of listings.

What can we learn about different hosts and areas?

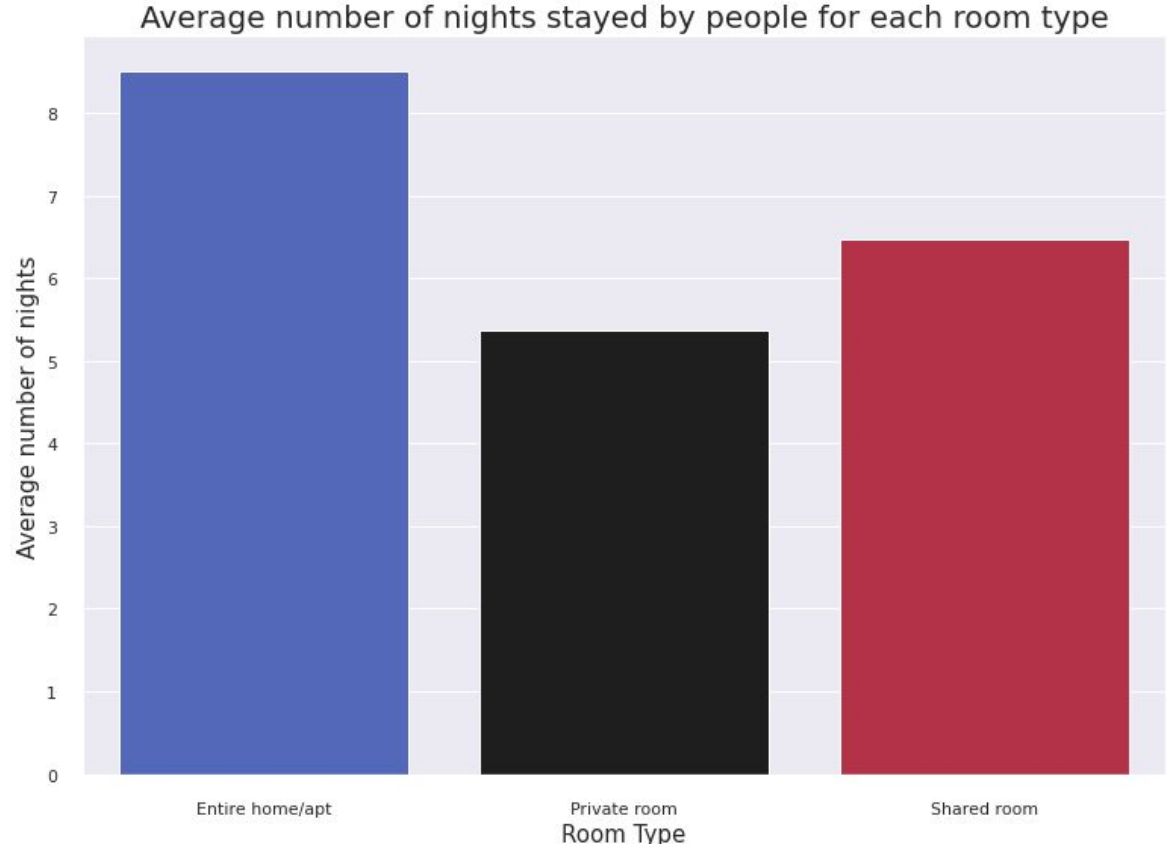
- People stayed for more nights (on an average of 8-9 nights) in Manhattan.
- People have stayed roughly the same number of nights in Bronx and Staten Island.



Minimum number of nights people live in different room types

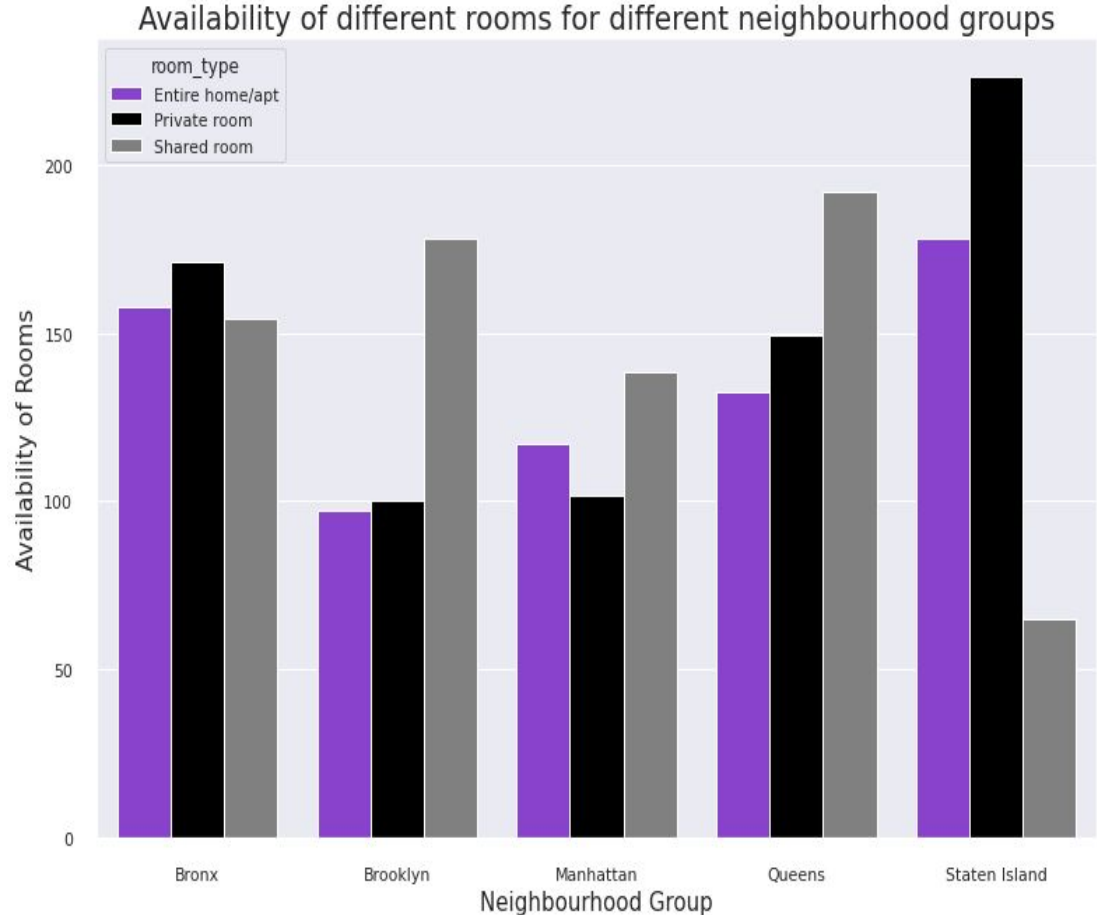


→ People typically prefer to stay longer in an entire home/apartment which is around 8-9 nights followed by shared rooms.



Availability of room types for different neighbourhood groups

- Staten Island has the highest availability of private rooms and entire home/apt.
- For Brooklyn, the availability of rooms is opposite to that of Staten Island.
- For Bronx and Manhattan the availability for all room types is roughly the same.



Count of last reviews per year

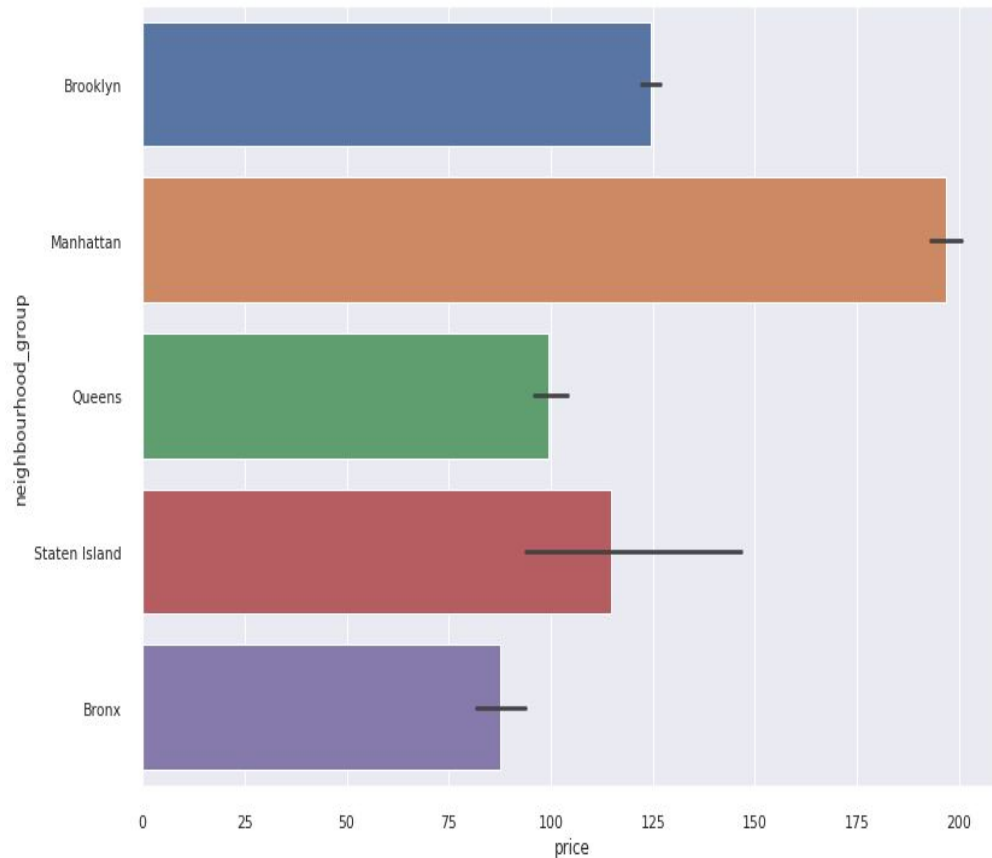
- Most people last reviewed in the year 2019
- Around 6,000 people last reviewed in 2018.

2019	25209
2018	6050
2017	3205
2016	2707
2015	1393
2014	199
2013	48
2012	25
2011	7



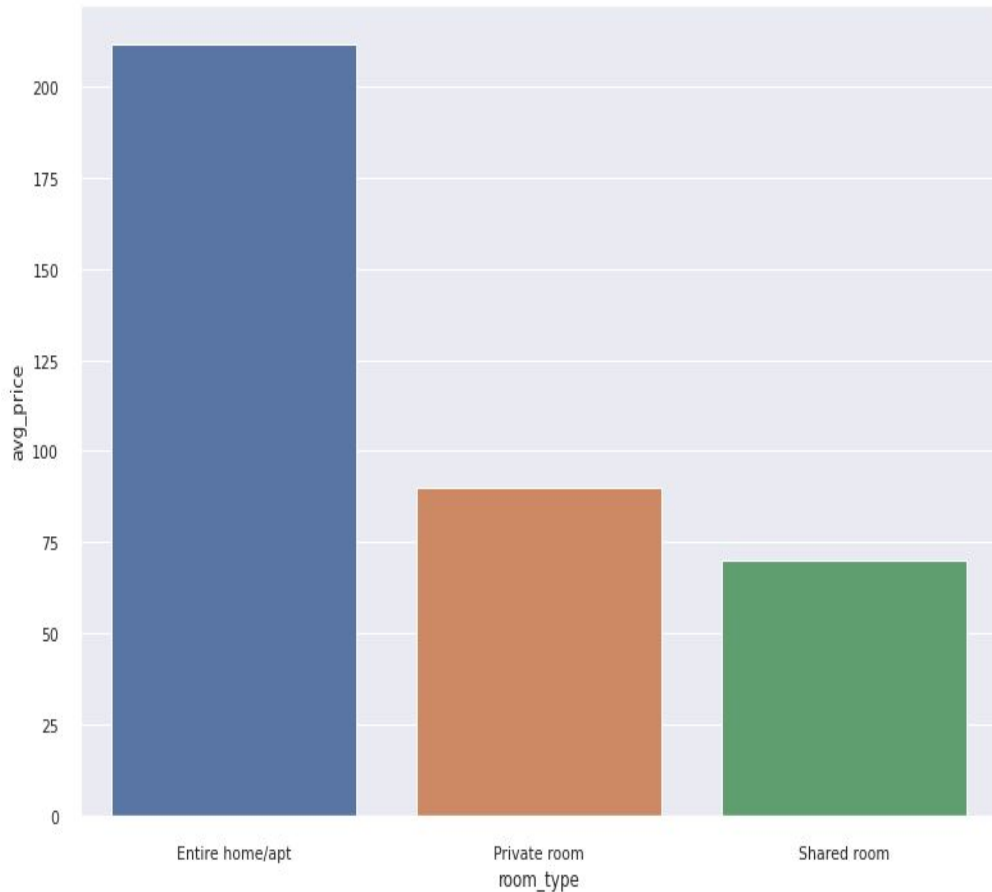
What is the average price of Airbnb with respect to different neighbourhood group ?

- Manhattan has the highest average price among all areas.
- Brooklyn has the second highest average price.
- Bronx has the lowest average price among all areas.



What is the average price for each room type?

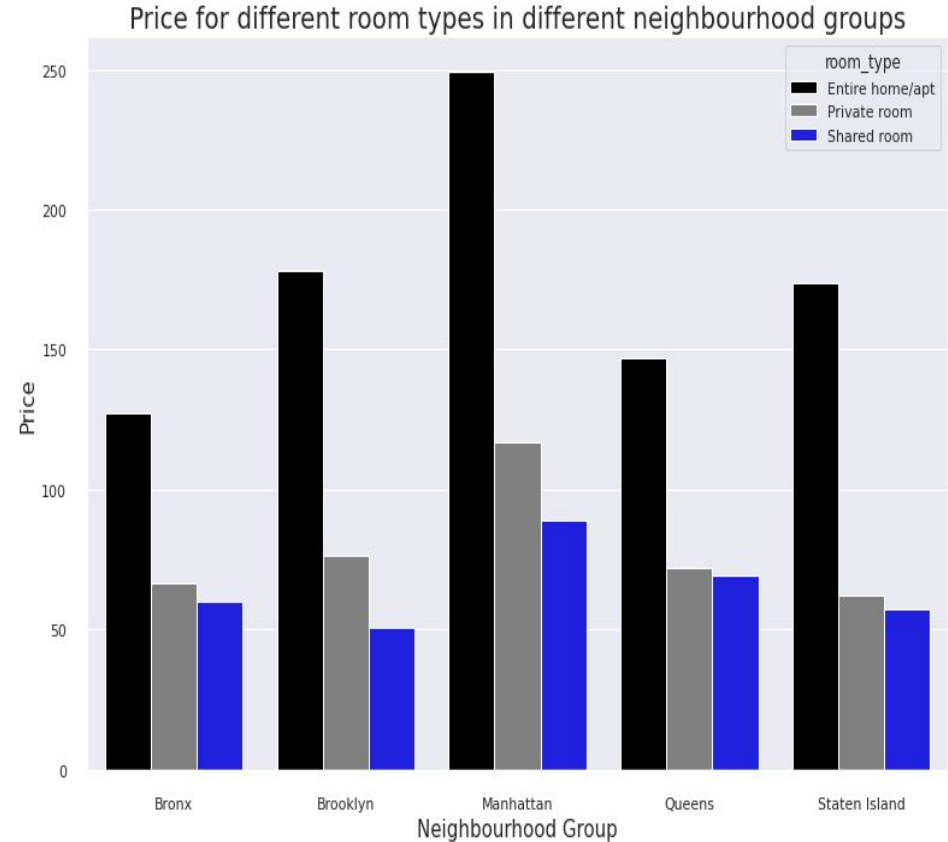
- For the Entire home/apt the avg. rate is approx. \$ 211.
- For the Private room the avg. rate is approx. \$ 89.
- For the Shared room the avg. rate is approx. \$ 70.



Average price of different room types for each neighbourhood group

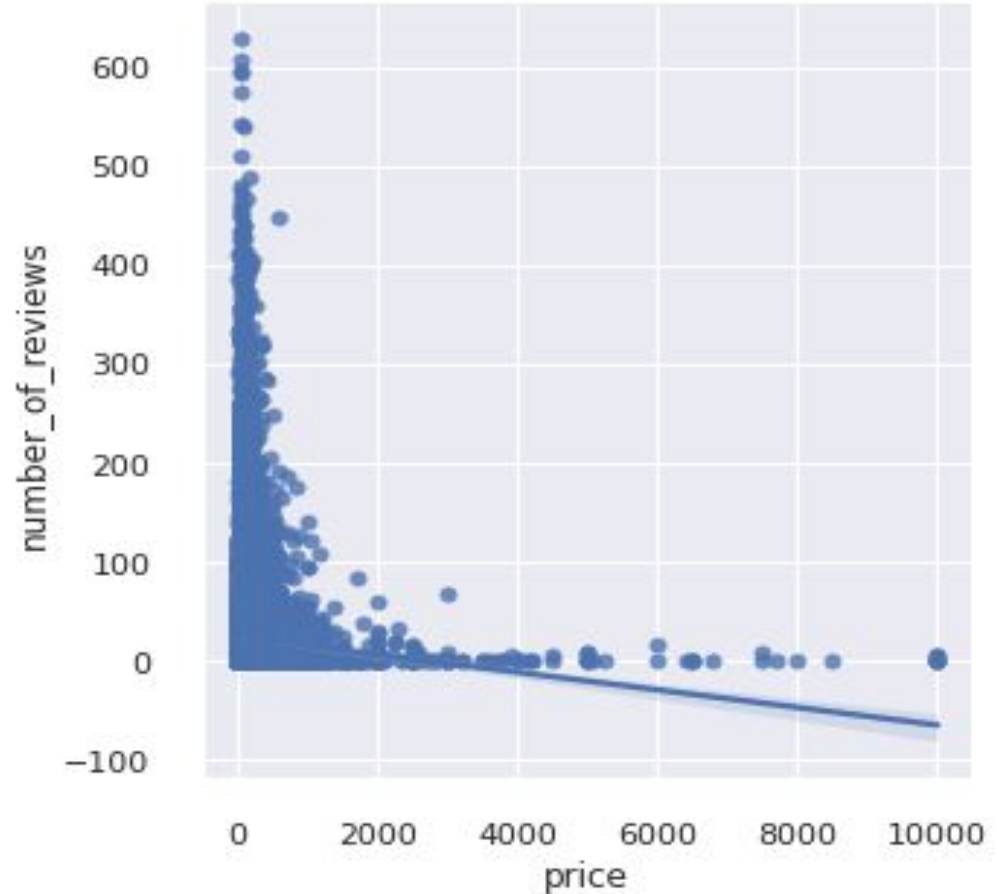


- Manhattan has the highest rate for Entire room/apt and Private rooms with respect to other locations.
- Staten Island and Queens have the approx. same rates for the private room and shared room.
- Brooklyn has the lowest rate for shared rooms compared to others.



Relationship between price v/s numbers of reviews ?

→ By looking at the graph we can conclude that people mostly prefer to stay where the price is minimum.

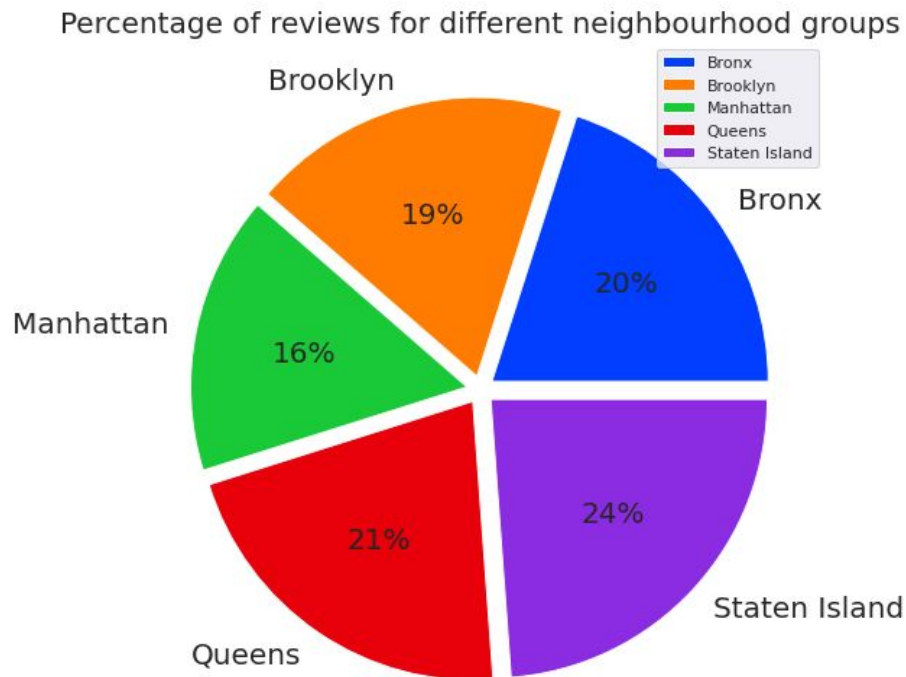


Which neighbourhood group has got the most reviews?

→ Staten island has the most numbers of reviews.

→ Manhattan has the least number of reviews.

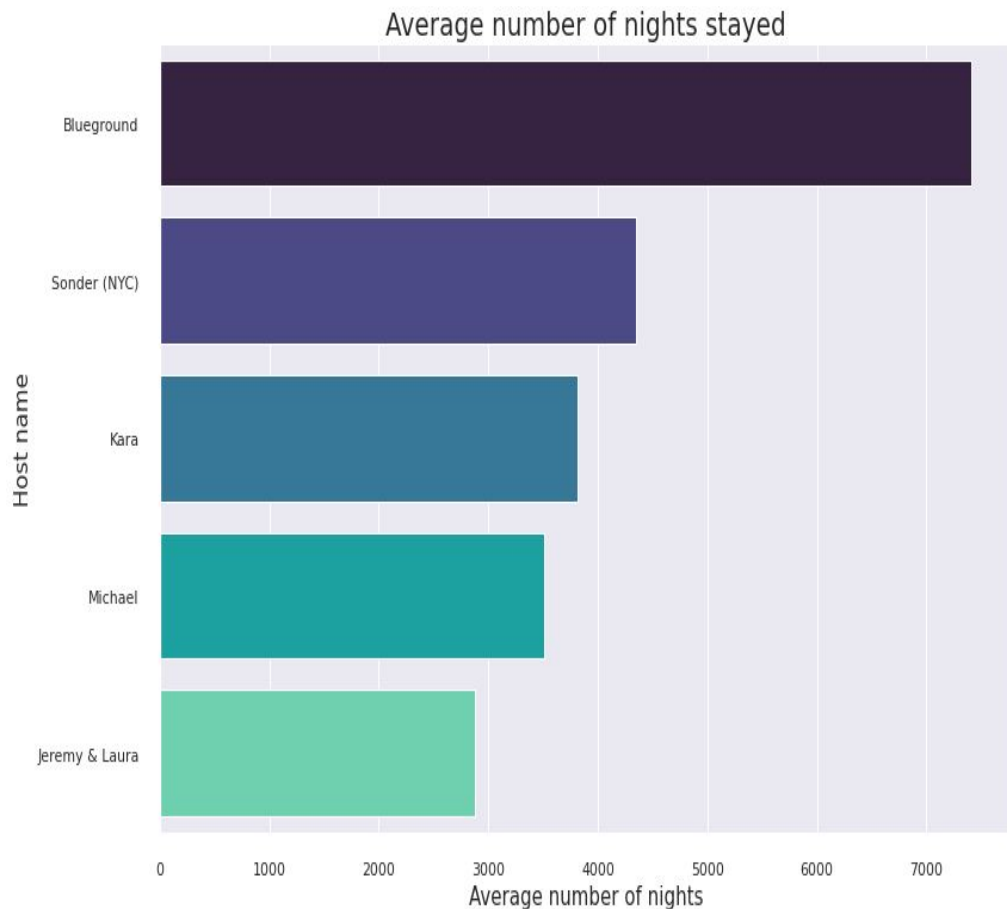
	neighbourhood_group	Reviews
0	Bronx	26.004583
1	Brooklyn	24.202845
2	Manhattan	20.985596
3	Queens	27.700318
4	Staten Island	30.941019



Which hosts are the busiest and why?

- For Blueground, people has stayed the most number of nights i.e. over 7000 in Manhattan as it one of the expensive place to live in.

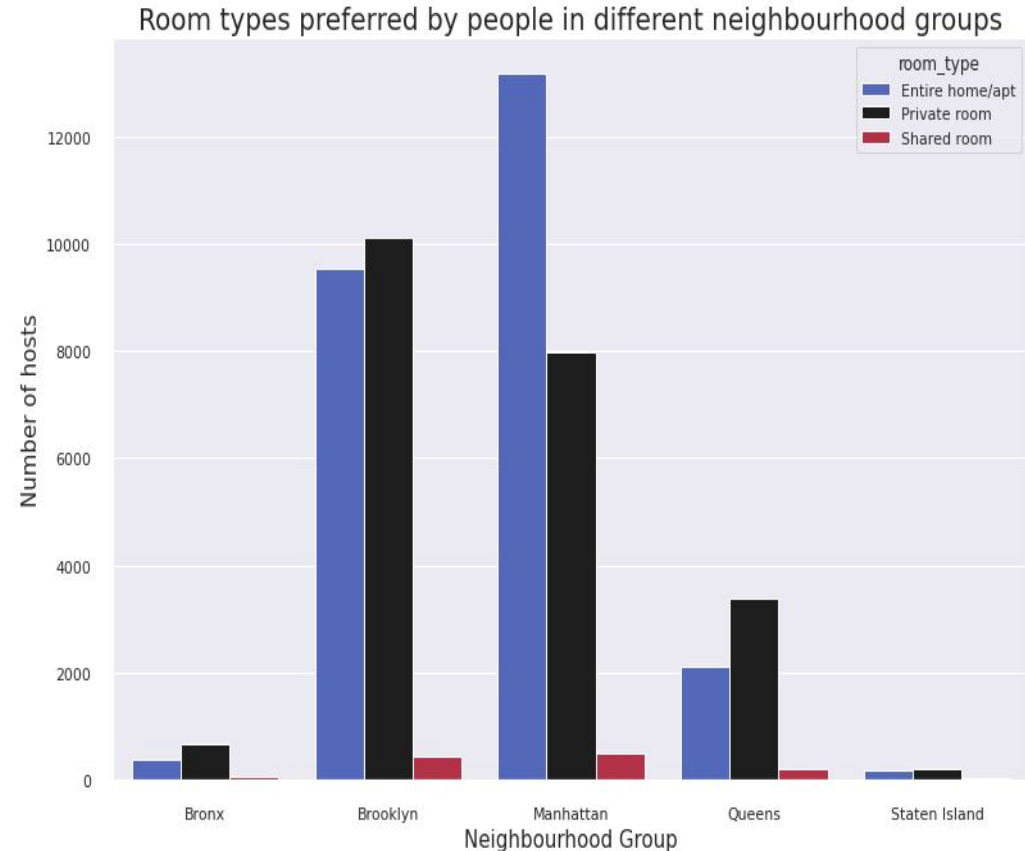
	host_name	neighbourhood_group	total_no_of_nights_stayed
1834	Blueground	Manhattan	7410
13221	Sonder (NYC)	Manhattan	4353
7275	Kara	Manhattan	3814
9742	Michael	Manhattan	3510
6540	Jeremy & Laura	Manhattan	2880



Is there any noticeable difference of traffic among different areas & what could be the reason for it?



- Entire home/apartment is mostly preferred by people in Manhattan.
- In Queens, people prefer private rooms
- For Bronx, Brooklyn and Staten Island, the overall preference for entire home/apartment and private room is same.
- Shared rooms are least preferred by the people.



Analysis summary

- People stay more nights in Manhattan in comparison to other places though it is one of the most expensive places to live.
- Although the average price in Manhattan is the highest, people prefer to stay in those rooms whose price is low.
- Among the 3 room types entire homes/apartments are in high demand.
- People who like to stay in the entire home or apartment tend to stay longer.
- Staten Island has the highest availability of rooms to other places.
- Staten Island has got the most reviews and people stay here for the least number of nights, which means it is a tourist attraction.