

Capstone Project – 1

EDA On Airbnb Booking Analysis

By

SHAIK AHMAD BASHA

What is Airbnb ?

Airbnb, as in “Air Bed and Breakfast,” is a service that lets property owners rent out their spaces to travelers looking for a place to stay. Travelers can rent a space for multiple people to share, a shared space with private rooms, or the entire property for themselves.

Airbnb is based on a peer-to-peer business model. This makes it simple, easy to use, and tends to be more profitable for both parties. The model also gives you the opportunity to customize and personalize your guests' experience the way you want.

Problem Statement

For this project, we will be analyzing Airbnb's New York City(NYC) data of 2019.

This dataset contains listings information such as listing name, host name, room types, minimum night stays, availability, area, reviews etc.

Our main objective behind this project is to explore and analyze the data to discover the key understandings. For this, we will explore and visualize the dataset from Airbnb in NYC using basic exploratory data analysis techniques.

Work Flow:



Data Understanding

After collecting data its very important to understand the data. Our data has 48895 rows and 16 columns. Those 16 columns are:

- **id** : Unique id of listing
- **name** : Name of the listing
- **host_id** : Unique id of host
- **host_name** : Name of the host
- **neighbourhood_group** : Location
- **neighbourhood** : Area
- **latitude** : Latitude of listing
- **longitude** : Longitude of listing
- **room_type** : Type of rooms
- **price** : Price of listing
- **minimum_nights** : Minumum number of nights people stay
- **number_of_reviews** : No. of reviews given for listing
- **last_review** : Latest review given for listing
- **reviews_per_month** : No.of review given per month
- **calculated_host_listings_count** : Total No. of listings for host
- **availability_365** : No. of days listing available

Data Cleaning and Manipulation

Our data has 4 columns with null values.

```
# Checking Null values in the data
data.isnull().sum().sort_values(ascending = False)

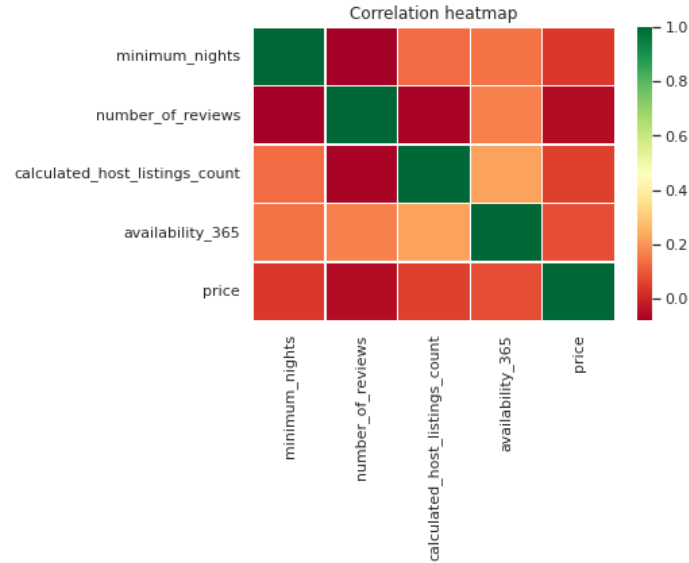
last_review          10052
reviews_per_month    10052
host_name             21
name                 16
```

- ❖ last_review and review_per_month columns have more null values, so we can remove those columns.
- ❖ host_name and name columns have 21 and 16 null values respectively, so we can remove those null values.
- ❖ we can also remove latitude, longitude columns as these columns don't effect much for our analysis.

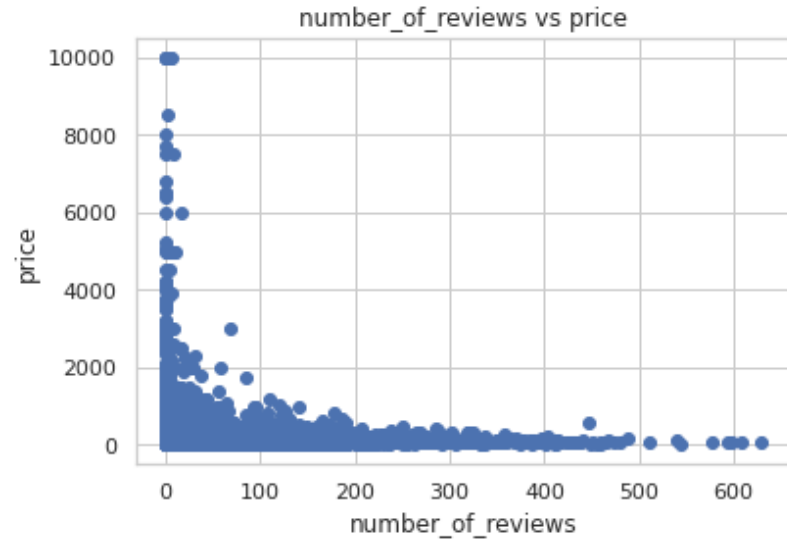
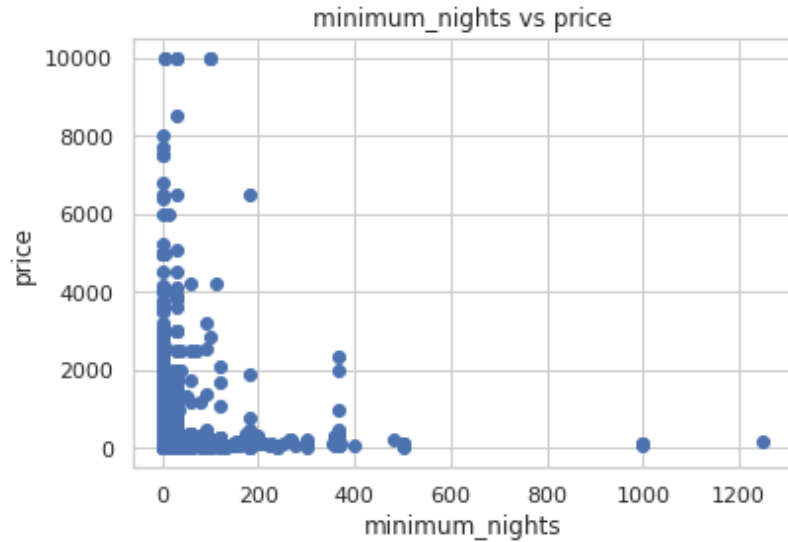
Exploratory Data Analysis

Numerical Features	Categorical Features
id	name
host_id	host_name
price	neighbourhood_group
minimum_nights	neighbourhood
number_of_reviews	room_type
calculated_host_listings_count	
availability_365	

Correlation between numerical features:

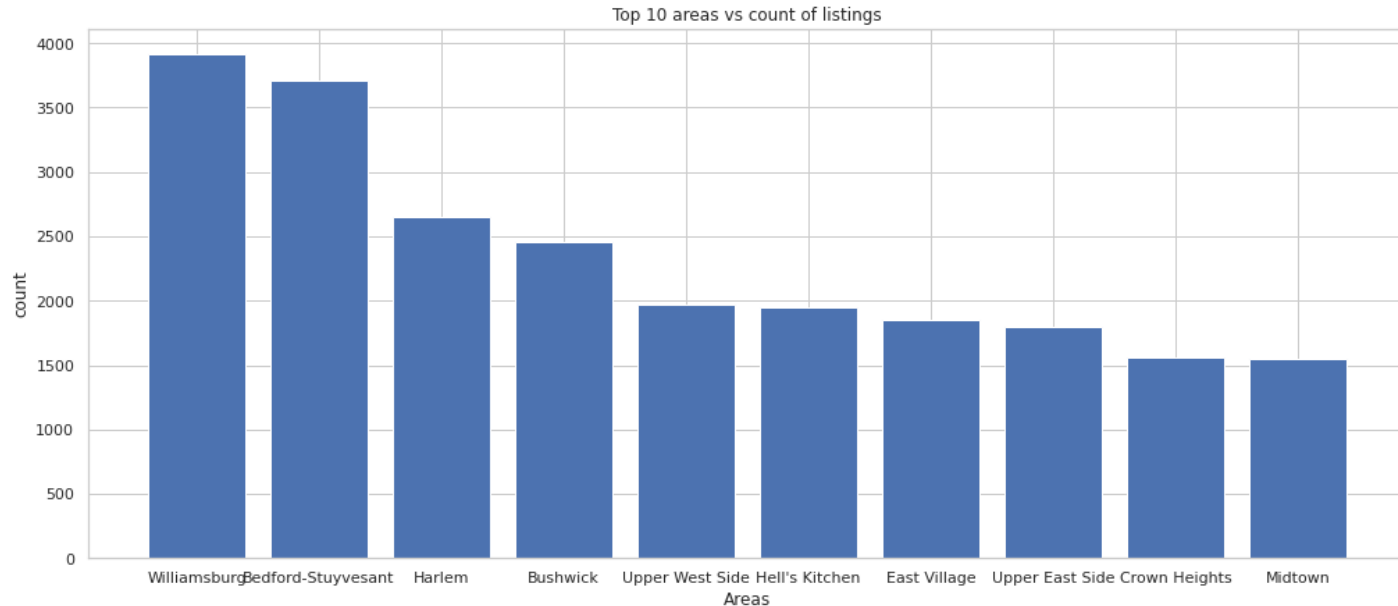


From the above heatmap, we can say that there are no variables with multicollinearity in the data

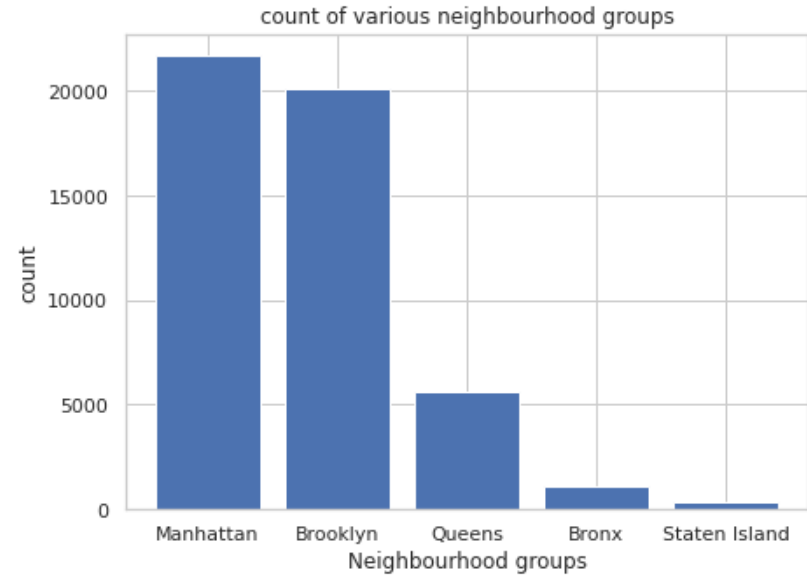
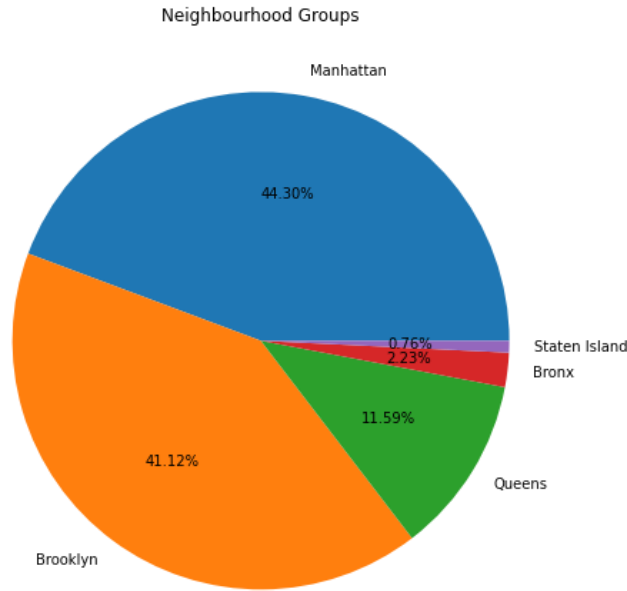


From the above scatter plots we can say that most people are preferring listings which has low prices.



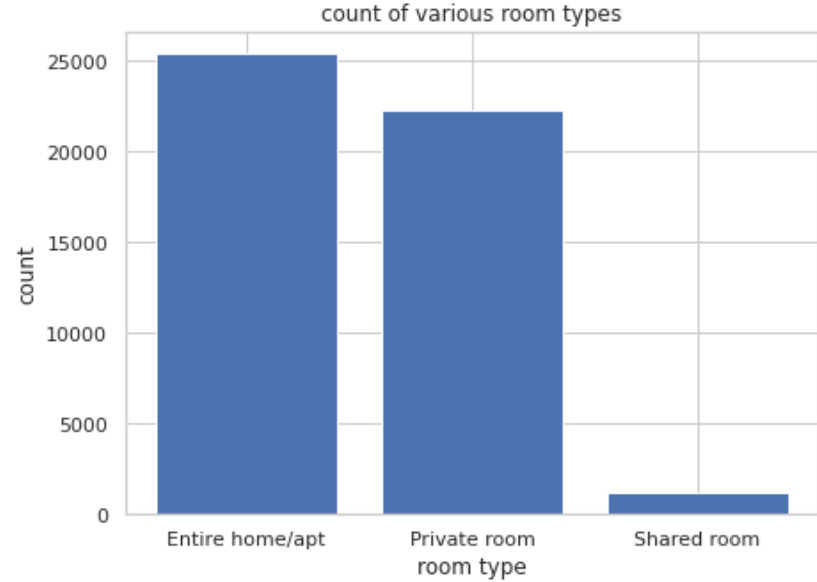
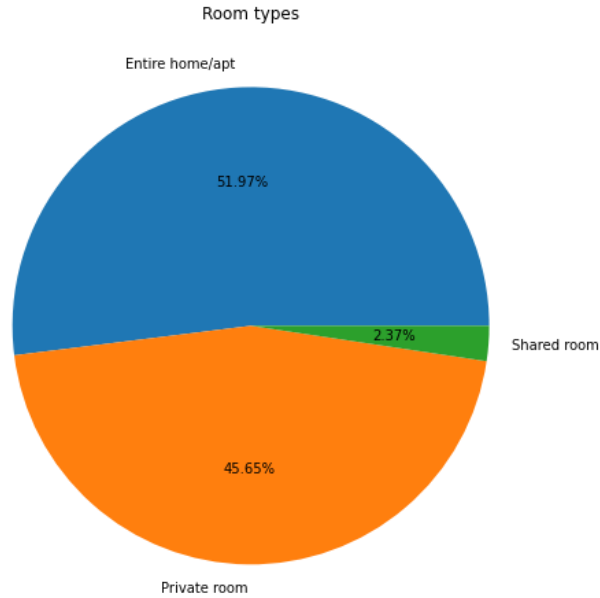


- ❖ In the data, there are total of 221 unique areas
- ❖ The areas with most number of listings are ['Williamsburg' 'Bedford-Stuyvesant' 'Harlem' 'Bushwick' 'Upper West Side' "Hell's Kitchen" 'East Village' 'Upper East Side' 'Crown Heights' 'Midtown']

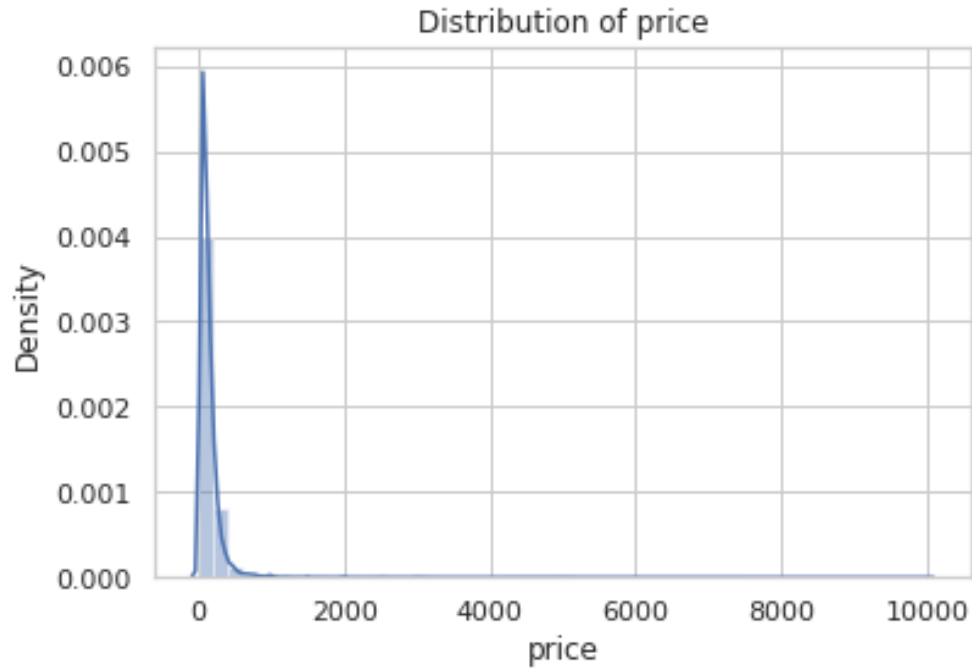


From the above charts, we can say that most number of listings are from Manhattan with 44.3% and Brooklyn with 41.12%

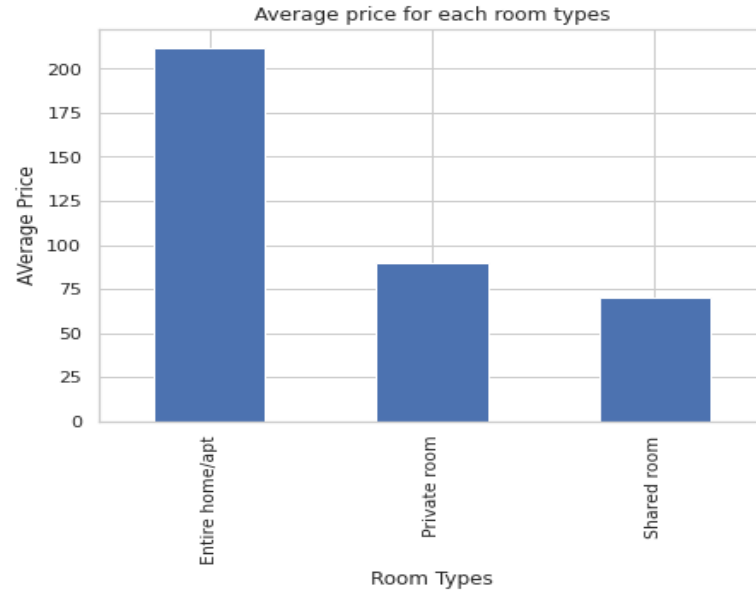
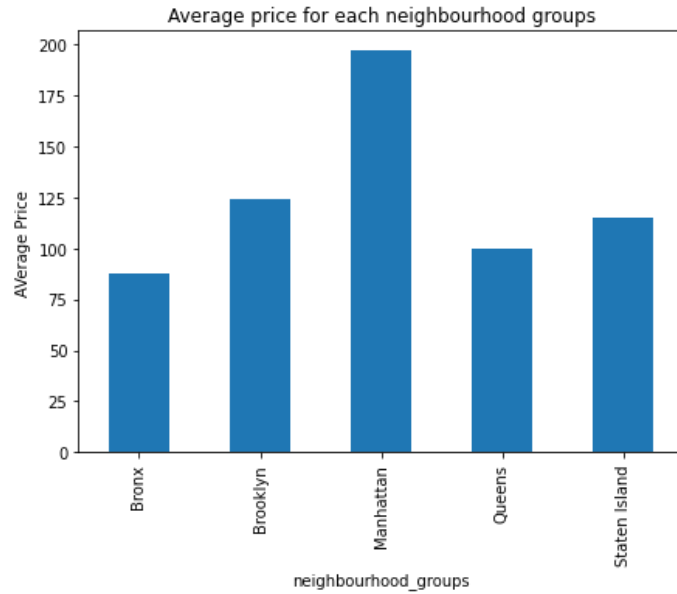
The least number of listings are from Staten Island with 0.76%



From the above charts, we can say that most number of listings are in category of Entire home/apt room type than others.

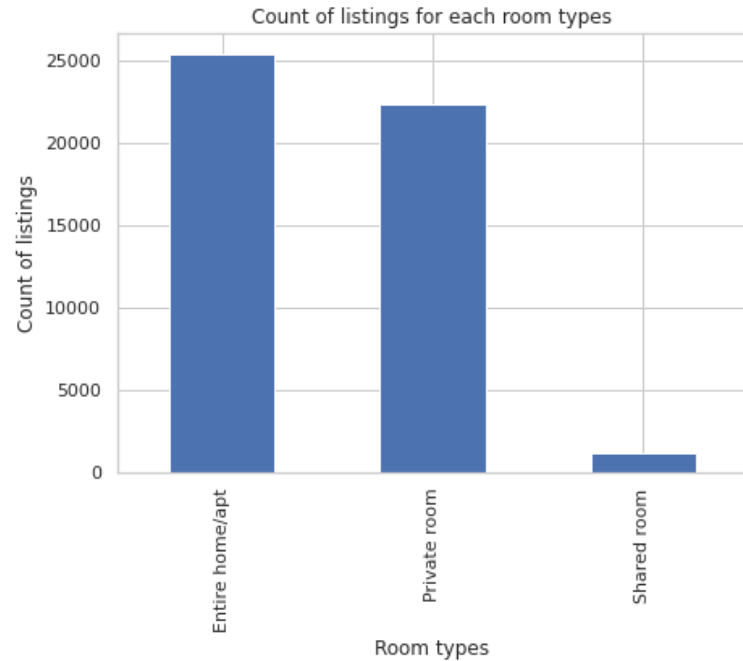
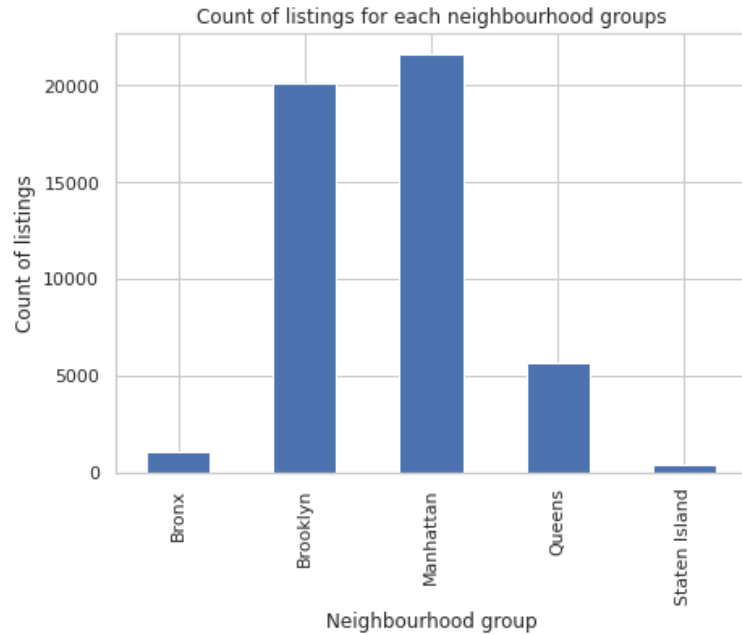


From the above plot we can observe that price is mostly in the range of 0 to 1000 and a few are in range of 1000 to 10000



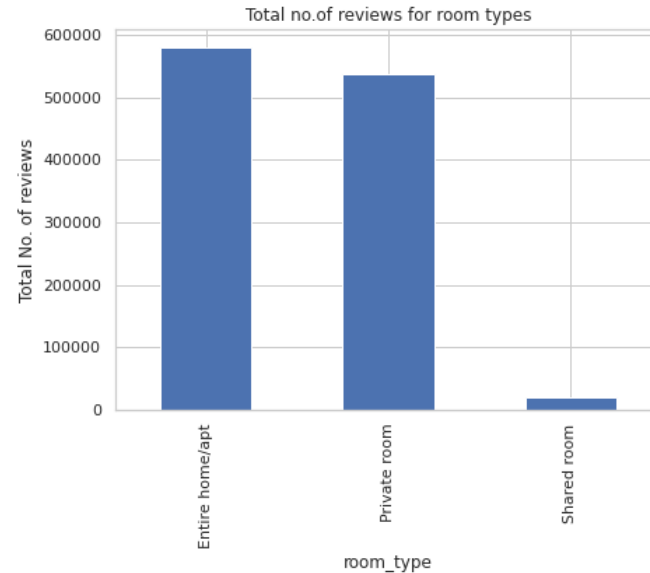
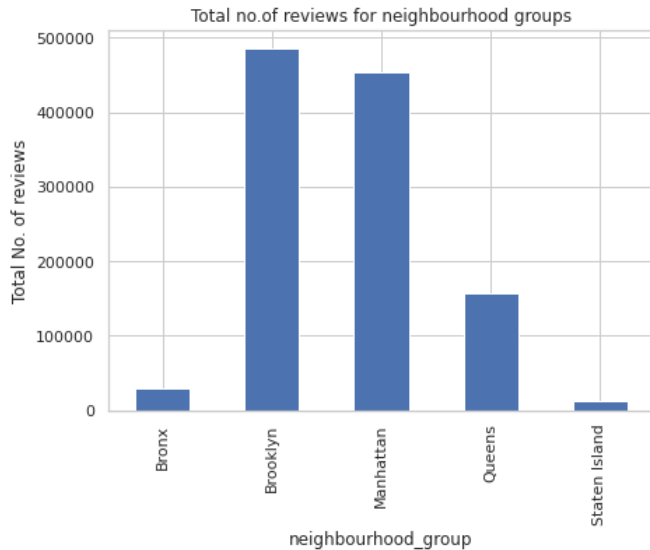
From the above charts we can observe that :

- ❖ Manhattan area has highest average price and least is Bronx
- ❖ Entire home/apt has highest average price than others



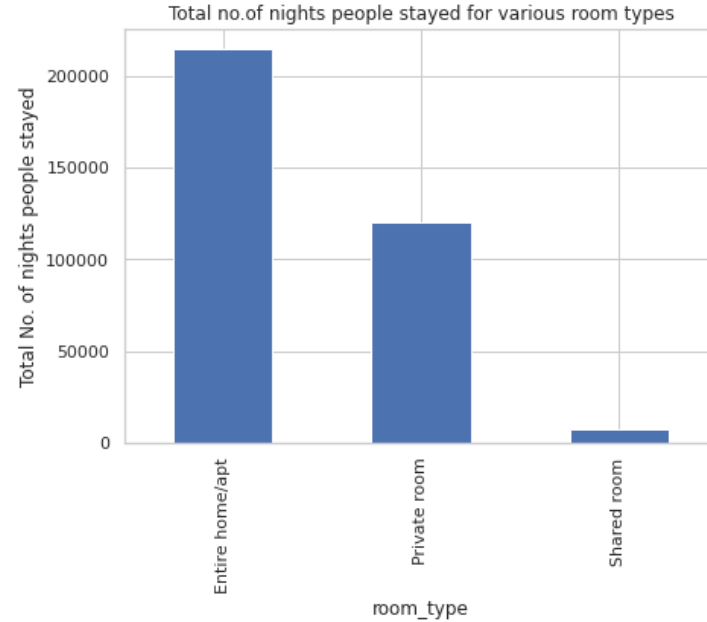
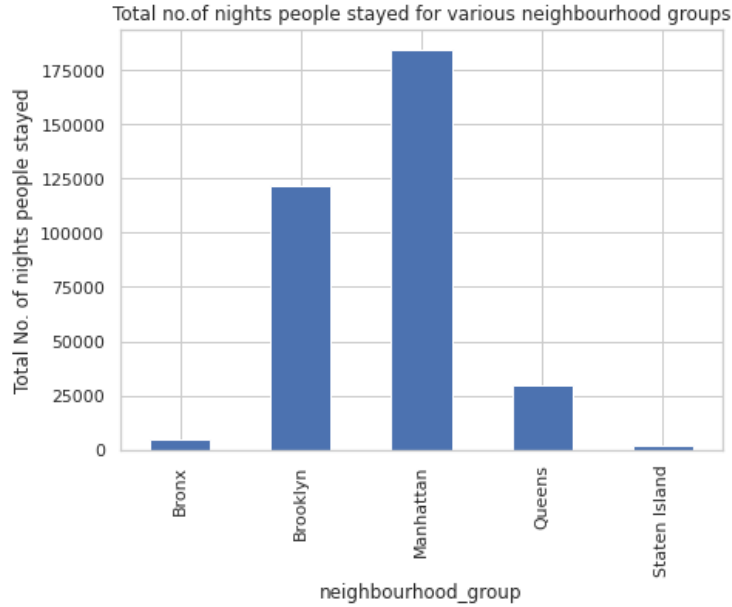
From the above chart we can observe that:

- ❖ **Manhattan has high count of listings and the least is Staten Island.**
- ❖ **Host listing count is higher for Entire home/apt followed by Private room**



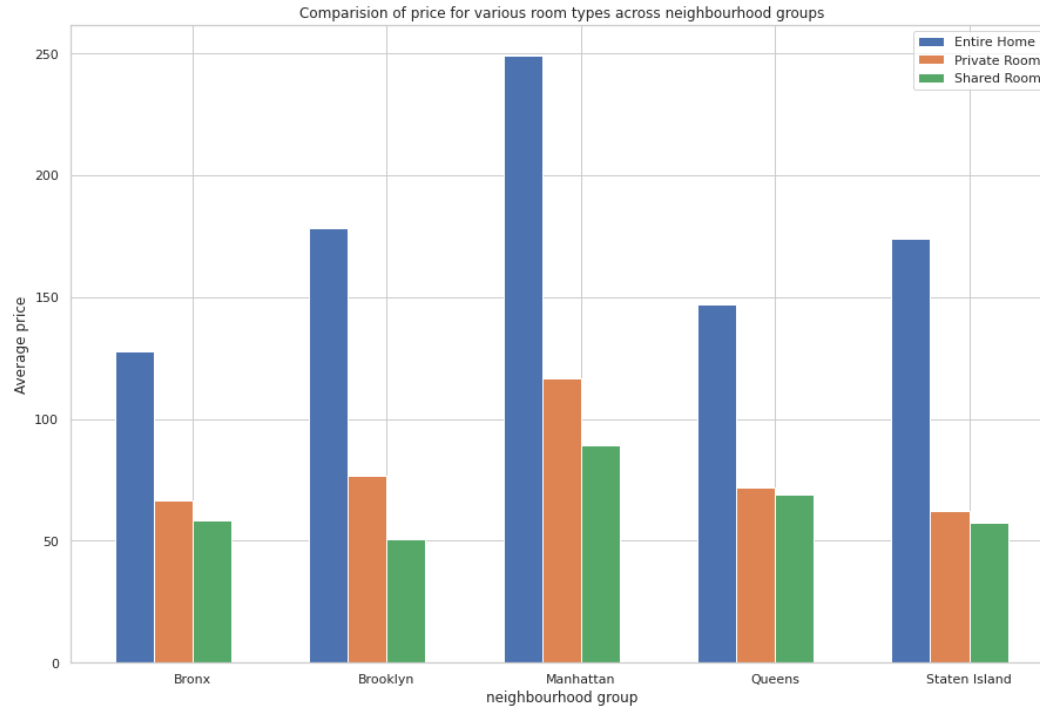
From the above chart we can say that :

- ❖ Brooklyn has highest number of reviews followed by Manhattan and least is Staten Island
- ❖ Entire home/apt has highest number of reviews than others.



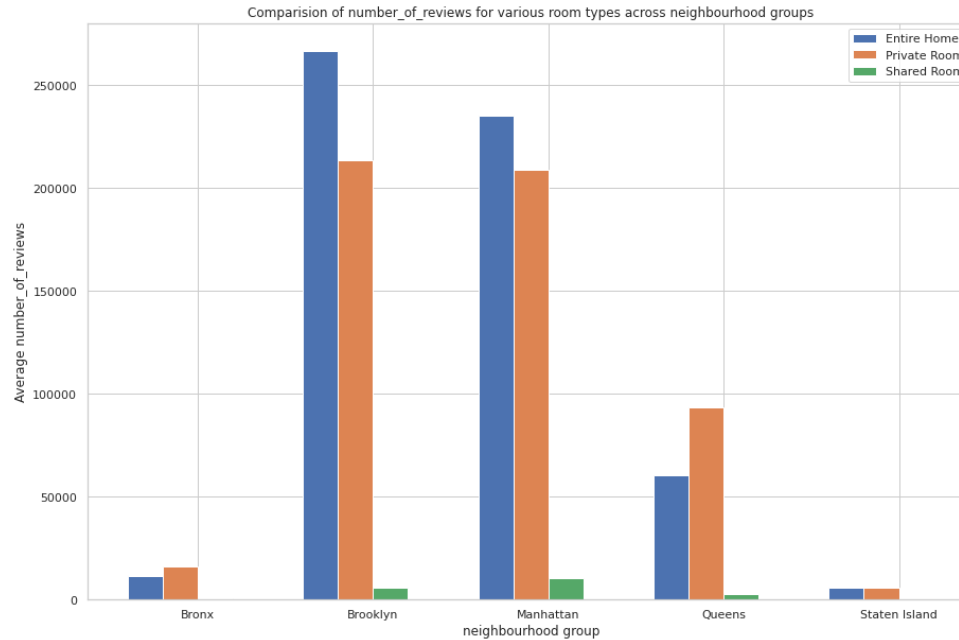
From the above chart we can say that :

- ❖ **Manhattan has more night stays**
- ❖ **Entire home/apt has more night stays and least is Shared room**



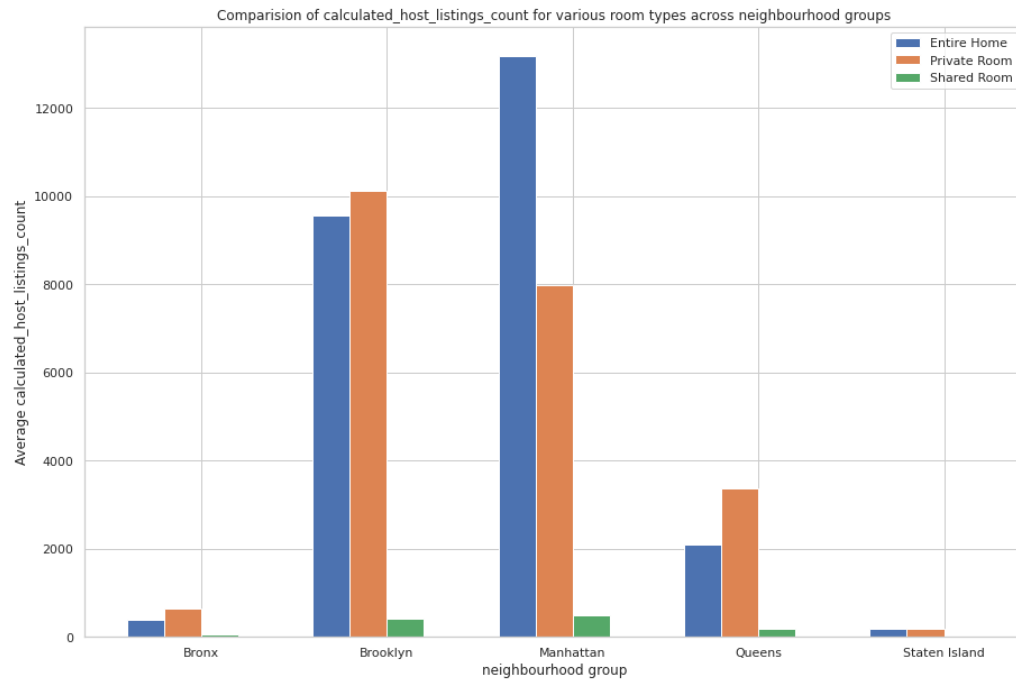
As we know Manhattan has highest average price. In Manhattan, entire home/apt has highest average price.

Similarly, we can see comparison between the average price for room types across each neighborhood groups.



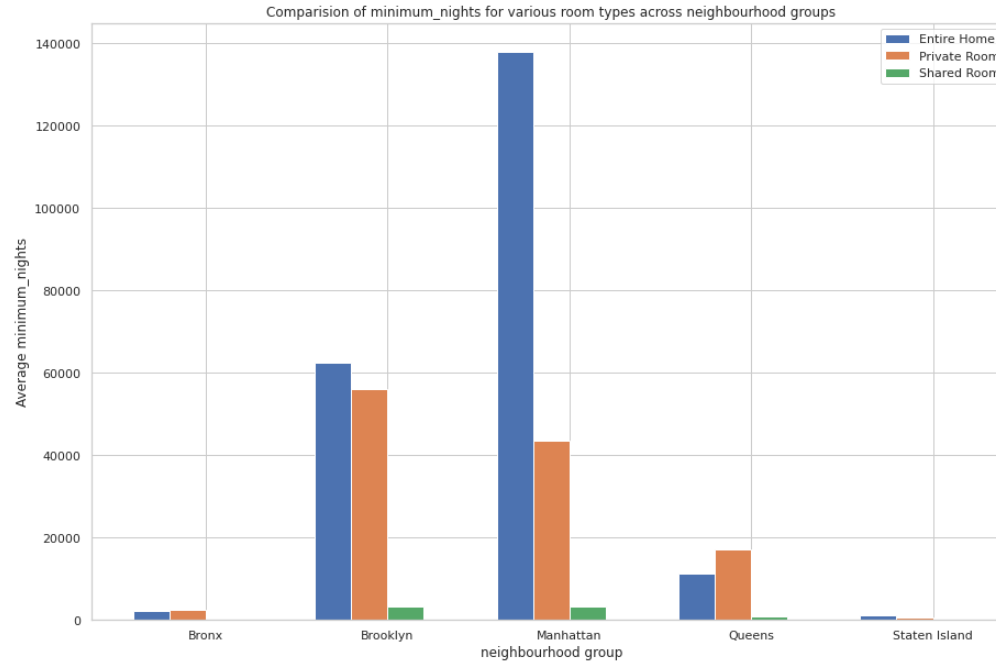
As we know Brooklyn has highest number of reviews. In Brooklyn, entire home/apt has highest number of reviews.

Similarly, we can see comparison between the number of reviews for room types across each neighborhood groups



As we know Manhattan has highest number of host listings. In Manhattan, entire home/apt has highest number of reviews.

Similarly, we can see comparison between the number of host listings for room types across each neighborhood groups

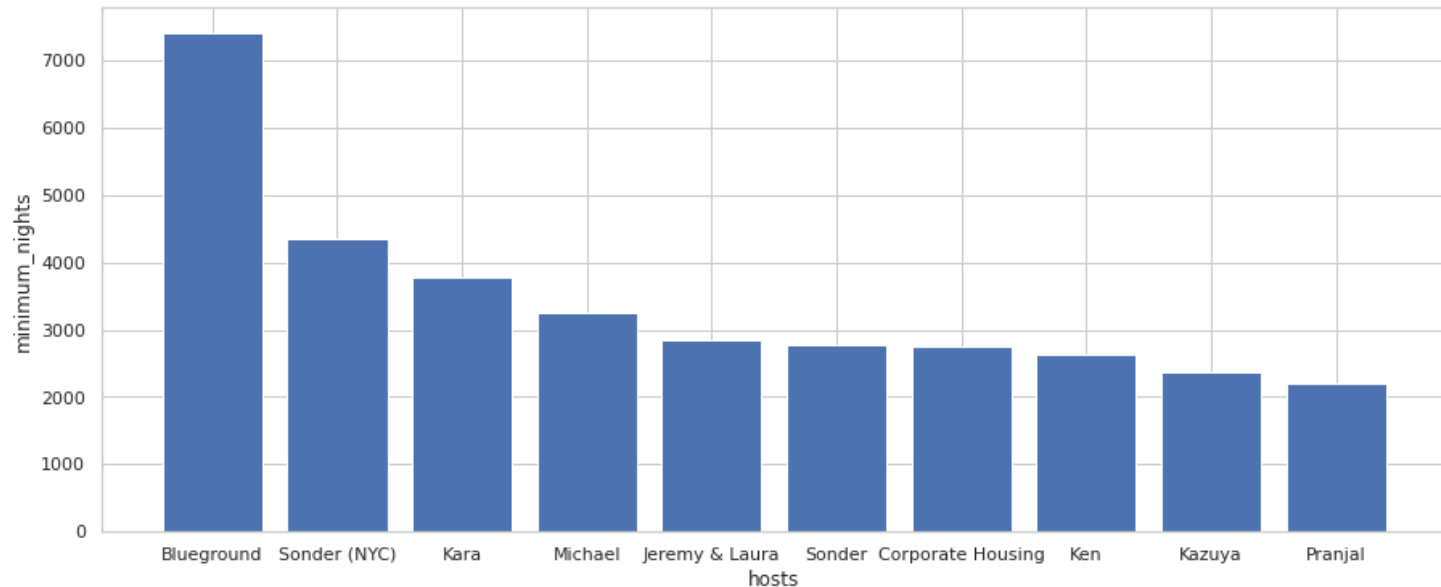


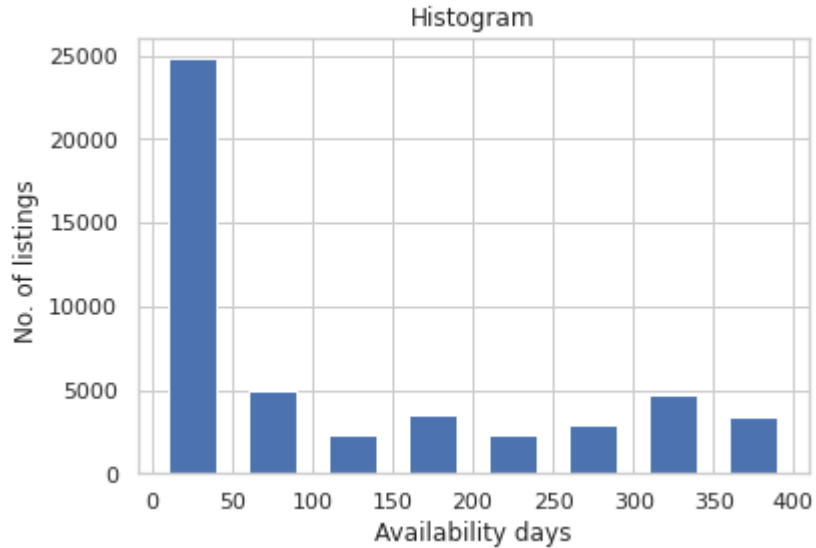
As we know Manhattan has highest number of night stays. In Manhattan, entire home/apt has highest number of night stays.

Similarly, we can see comparison between the number of night stays for room types across each neighborhood groups

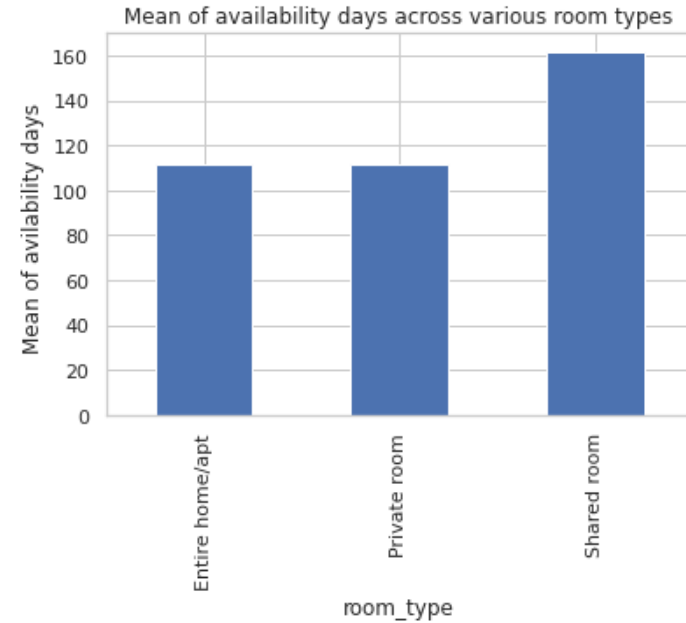
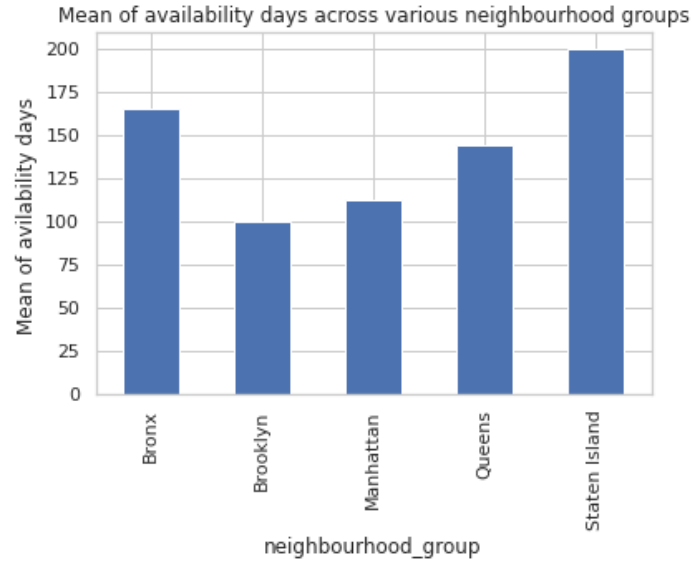
Busiest Hosts

The busiest hosts are Blueground, Sonder (NYC), Kara, Michael, Jeremy & Laura, Sonder, Corporate Housing, Ken, Kazuya, Pranjal





- ❖ From the above histogram, we can say that a most number of listings are available for only less than 50days
- ❖ There are total of 1294 listings that are available for 365 days



- ❖ From the above charts we can see that Staten Island has availability of rooms with more number of days and
- ❖ Shared room category has availability of rooms with more number of days

Final Conclusions :

- ❖ Most people are preferring Manhattan (44.3%) and Brooklyn(41.12%), and the least is Staten Island(0.76%).
- ❖ Most people are preferring listings which has low price.
- ❖ Most people are preferring Entire home/apt (51.97%) followed by Private room(45.65%) and least is Shared room(2.37%).
- ❖ The price of listings are mostly in the range of 0 to 1000 and a few listings are in range of 1000 to 10000
- ❖ The number of listings with price less than 1000 are 48619.
- ❖ The number of listings with price more than 1000 are 239.
- ❖ Manhattan area has the highest average price (196.87) and the least is Bronx(87.4967)
- ❖ Entire home/apt has highest average price (211.806) than others

- ❖ Manhattan has the high count of listings (21,643) and the least is Staten Island(373).
- ❖ Host listing count is higher for Entire home/apt (25,393) followed by Private room (22,306) and least is Shared room (11,59).
- ❖ Brooklyn has the highest number of reviews and the least is Staten Island.
- ❖ Entire home/apt has the highest number of reviews
- ❖ Manhattan has more night stays and the least is Staten Island
- ❖ Entire home/apt has more night stays and the least is Shared room
- ❖ The top hosts with most number_of_reviews are Dona, Jj, Maya, Carol, Danielle, Asa, Wanda
- ❖ The top 10 hosts with most minimum_nights are Genevieve, Glenn H., Angie, Meg, Aliya, John, Laura, Amanda, Meg, Shining

- ❖ The top hosts with most listing price are Kathrine, Erin, Jelena, Olson, Amy, Matt, Rum, Jessica, Sally, Jack.
- ❖ The top host with most calculated_host_listings_count is Sonder (NYC).
- ❖ The busiest hosts are Blueground, Sonder (NYC), Kara, Michael, Jeremy & Laura, Sonder, Corporate Housing, Ken, Kazuya, Pranjai.
- ❖ Staten Island has availability of rooms with more number of days
- ❖ The shared room category has availability of rooms with more number of days.
- ❖ There are total of 1294 listings that are available for 365 days.

THANK YOU