

Capstone Project - 1

AirBnB Bookings Analysis

Content

Distribution

- Listings across neighbourhood
- Prices of listings across neighbourhoods

Findings

- Demand and Supply
- Busiest hosts and the reason
- Traffic and Price across area
- Frequent words in names

Problem Statement

- What can we learn about different hosts and areas?
- 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?
- What can we learn from predictions? (ex: locations, prices, reviews, etc)

Data Summary:

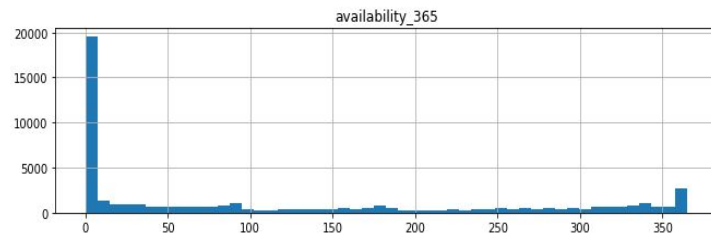
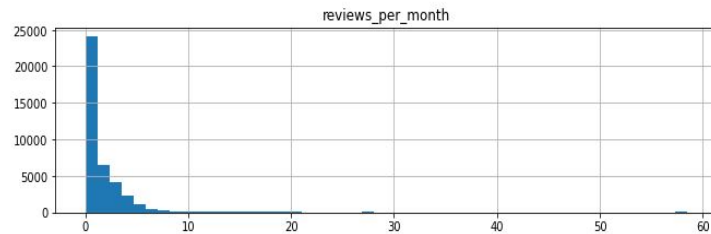
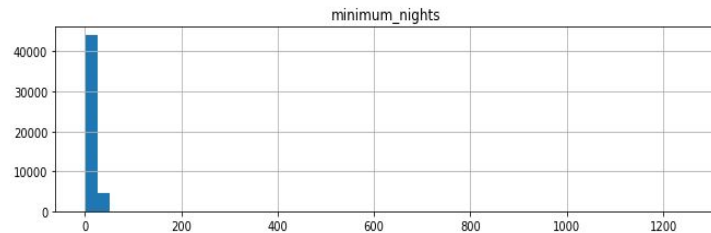
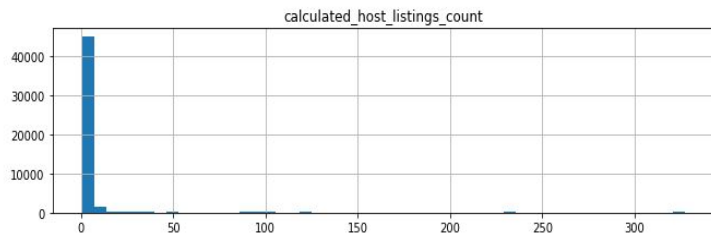
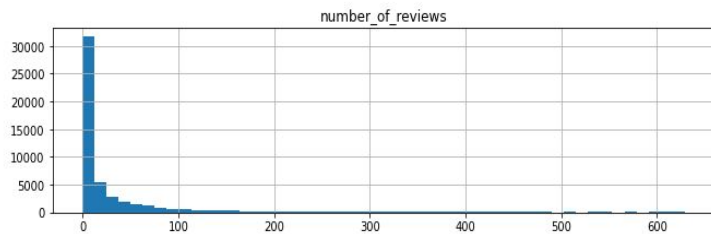
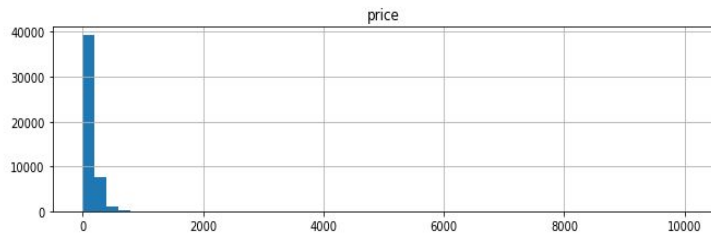
Data set name-- Airbnb NYC 2019

Shape--

- Rows -- 48,895
- Columns--16

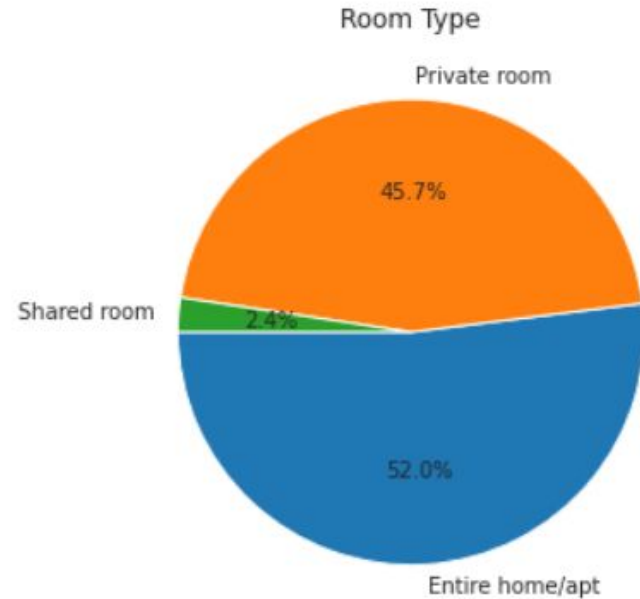
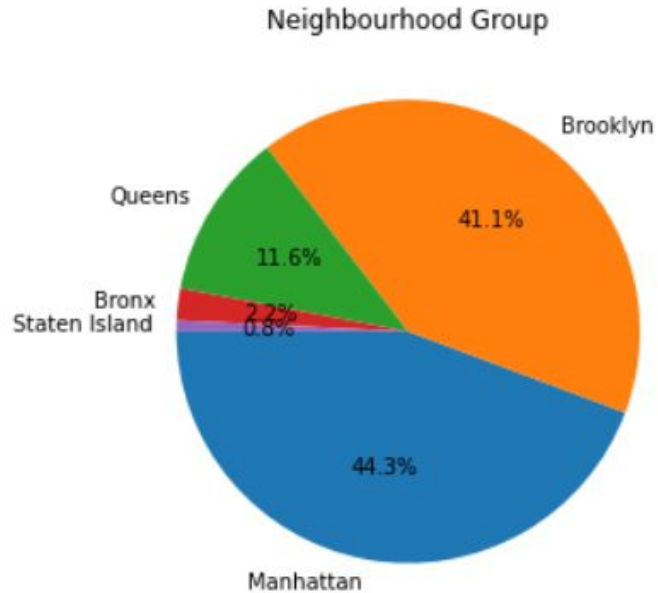
Columns-- ['id', 'name', 'host_id', 'host_name', 'neighbourhood_group', 'neighbourhood', 'latitude', 'longitude', 'room_type', 'price', 'minimum_nights', 'number_of_reviews', 'last_review', 'reviews_per_month', 'calculated_host_listings_count', 'availability_365']

Distribution of numerical features:



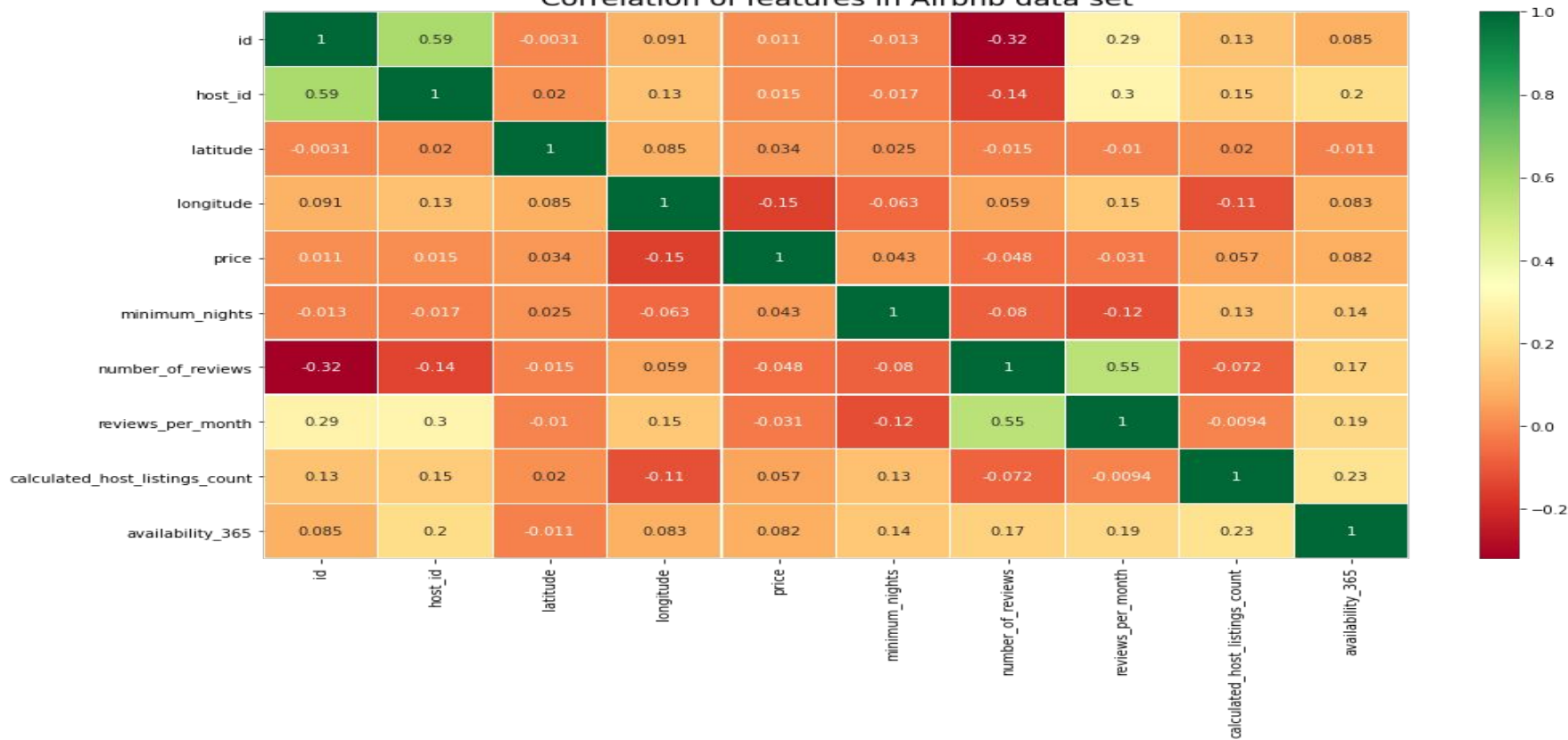
<Figure size 432x288 with 0 Axes>

Distribution of listings across categories:

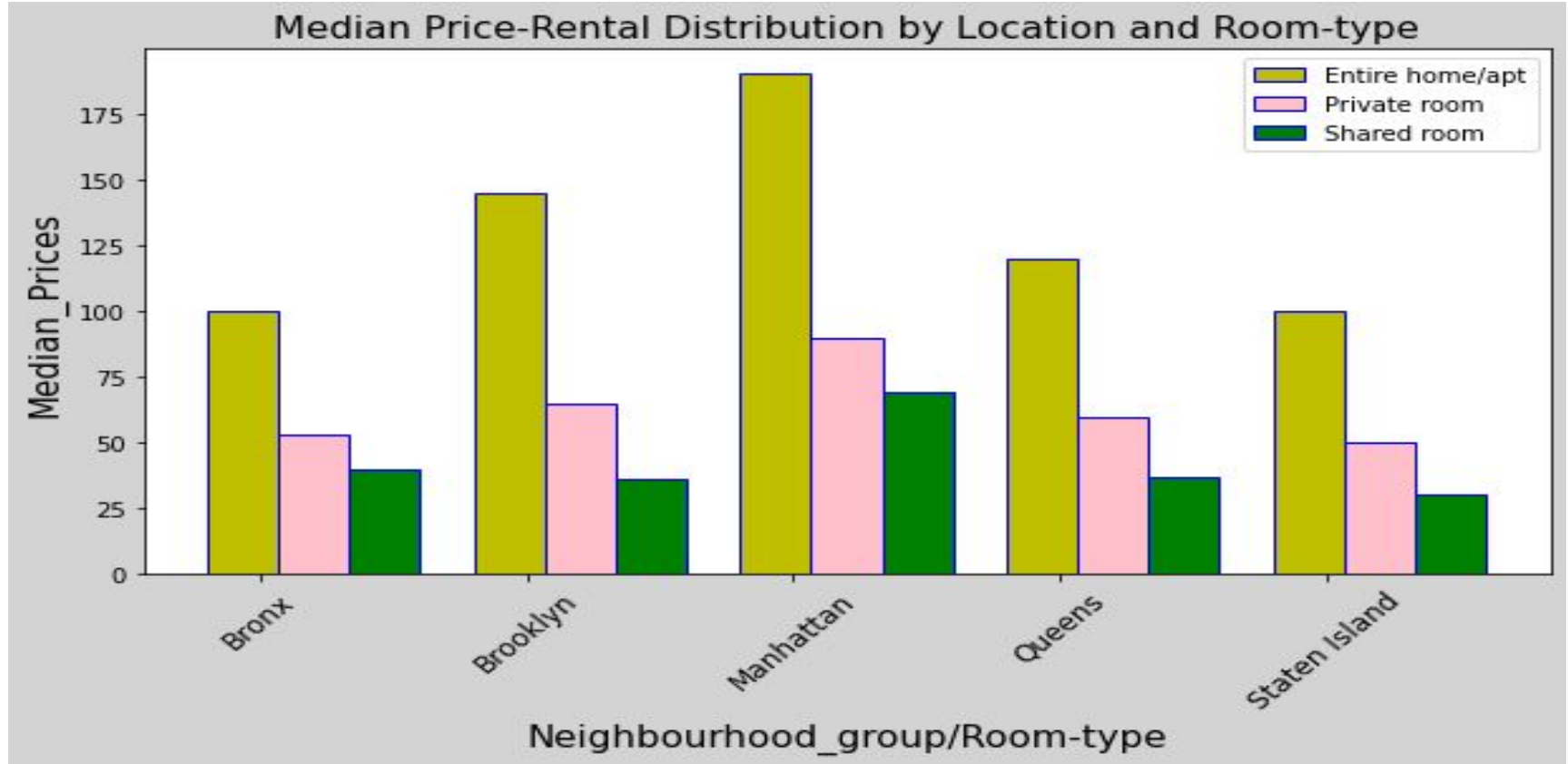


Correlation:

Correlation of features in Airbnb data set

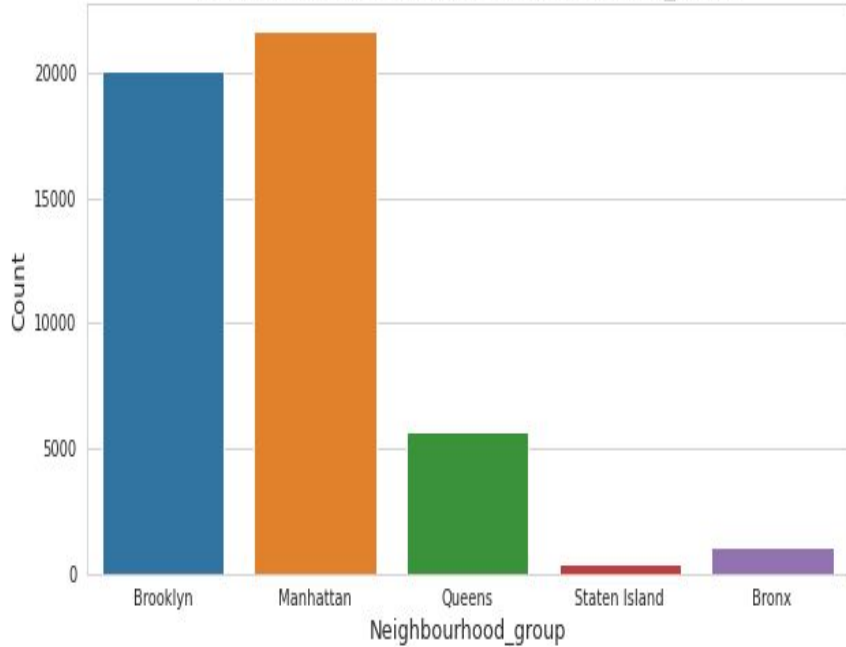


Price Distribution by Location and Room-type:

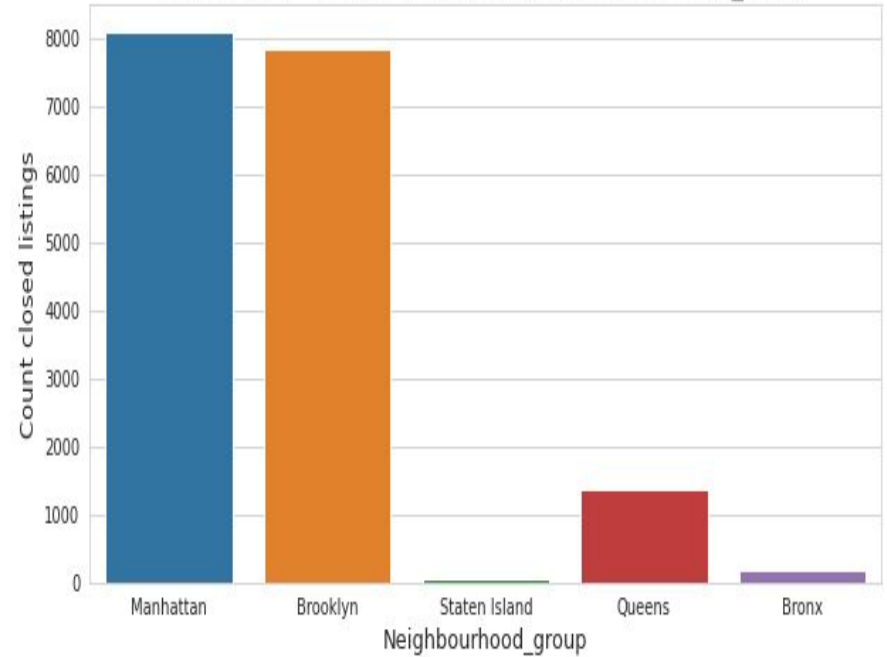


Total and closed listings:

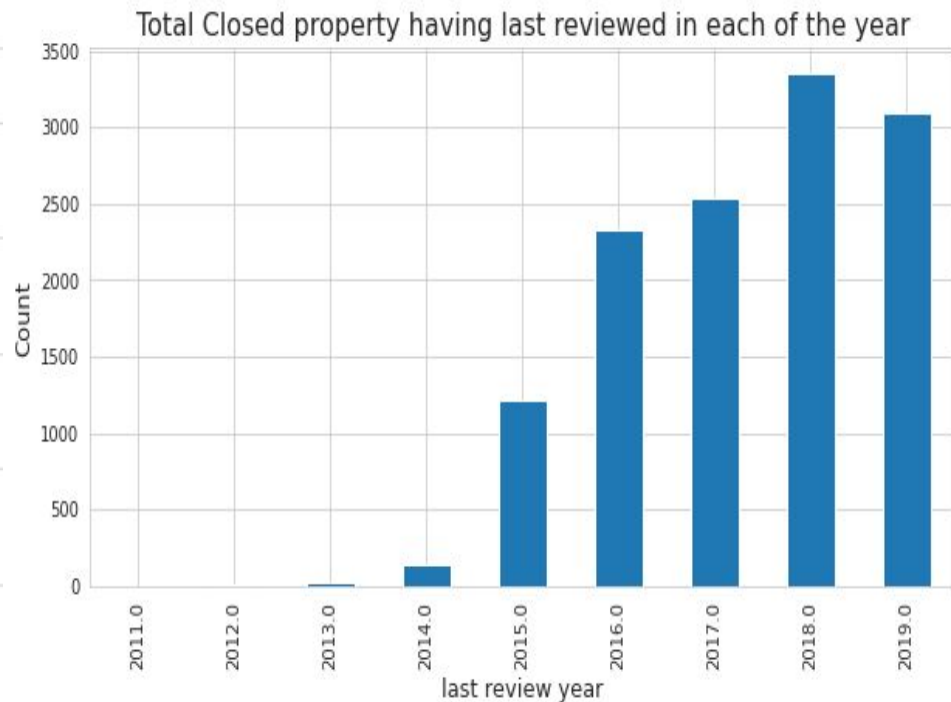
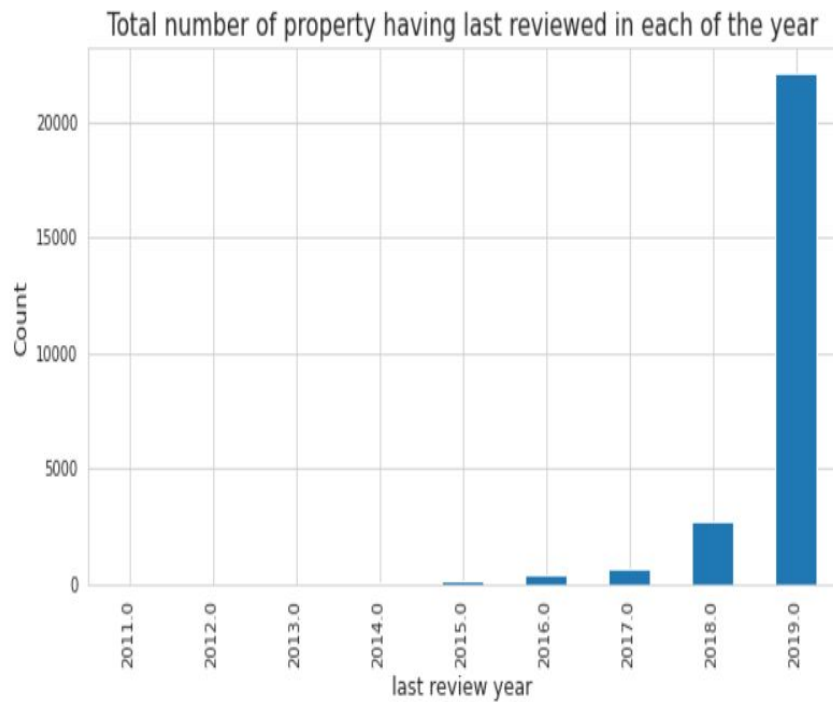
Total listings in each of the Neighbourhood_group



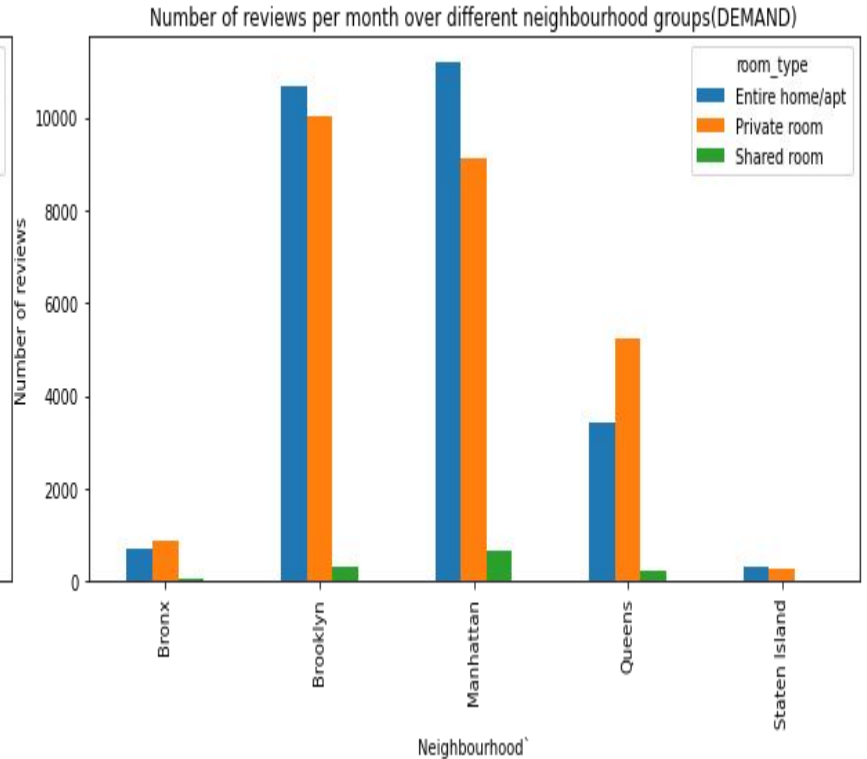
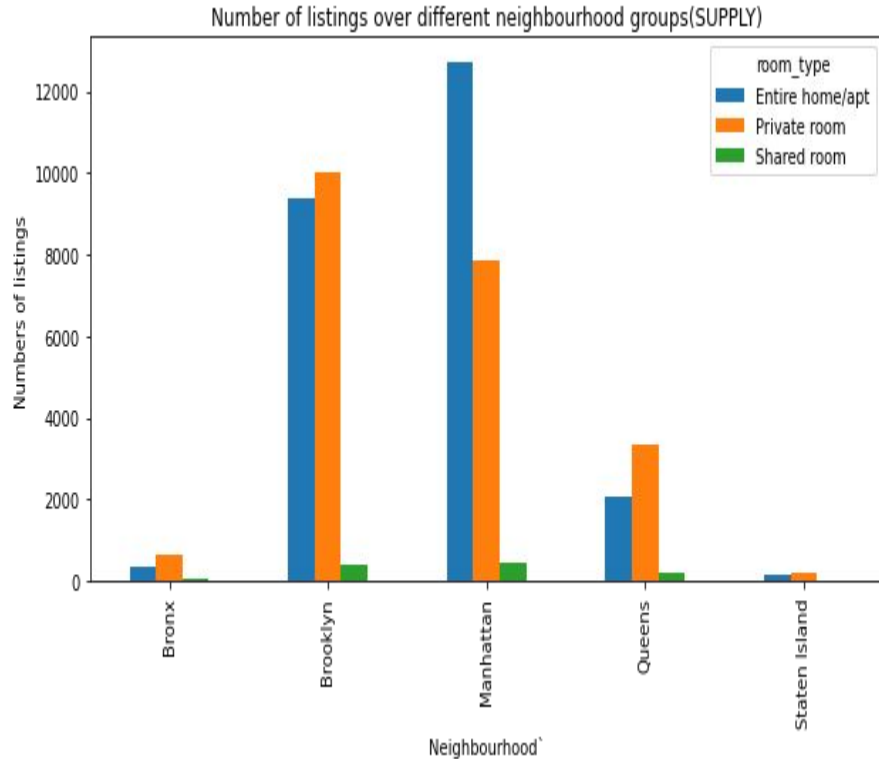
Total Closed listings in each of the Neighbourhood_group



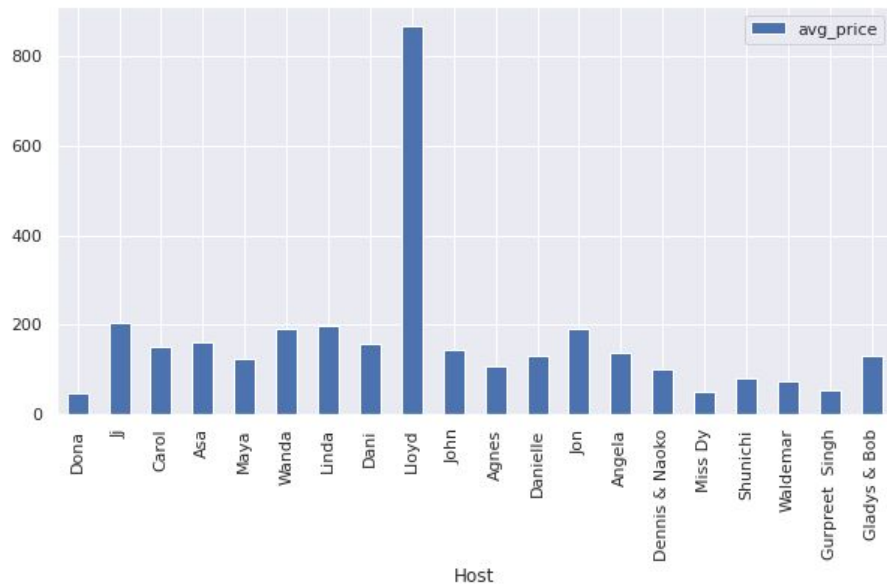
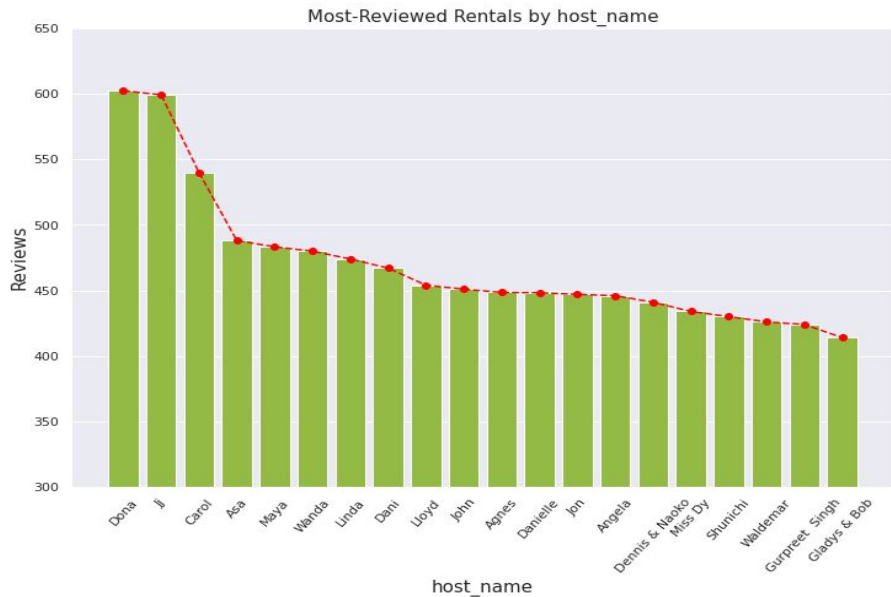
Total and Closed listing's number of reviews:



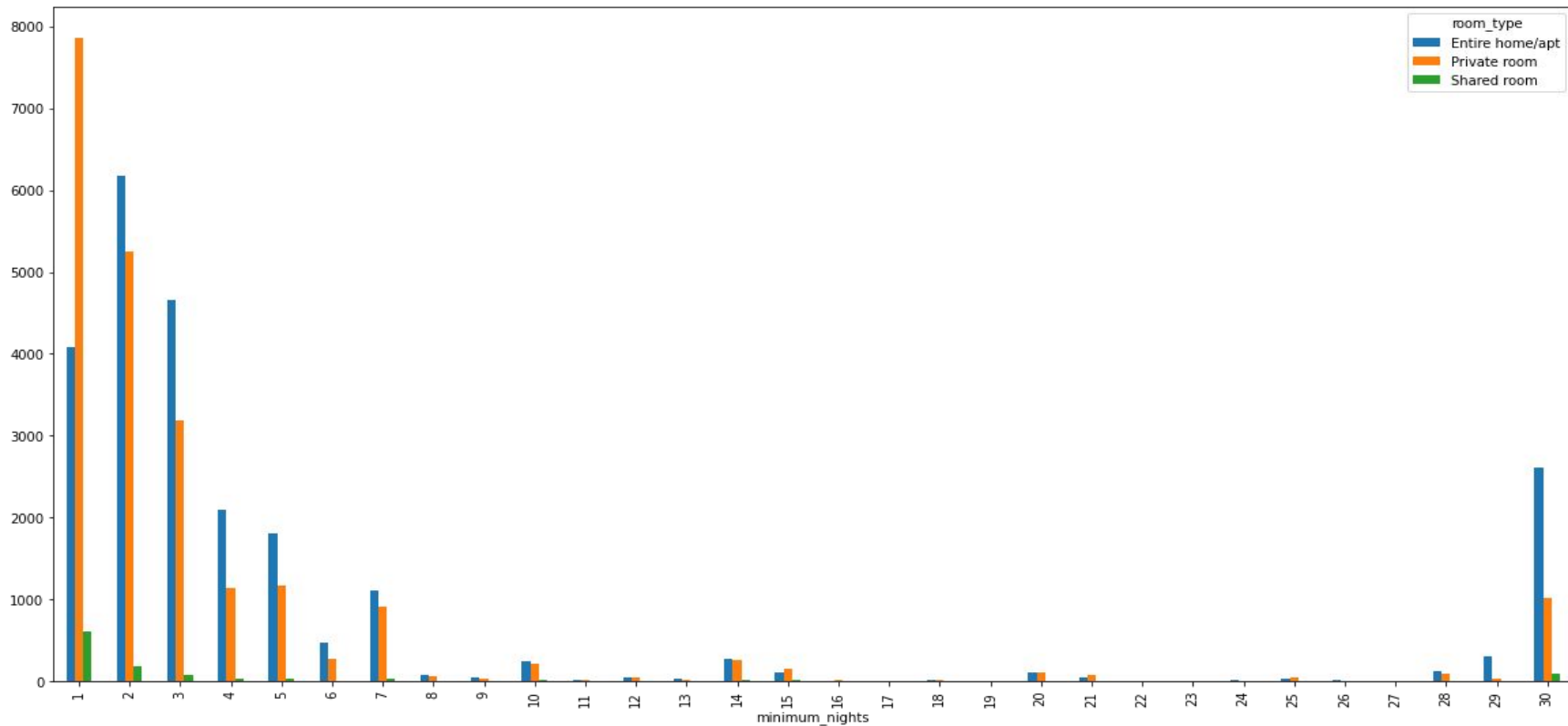
Demand and Supply:



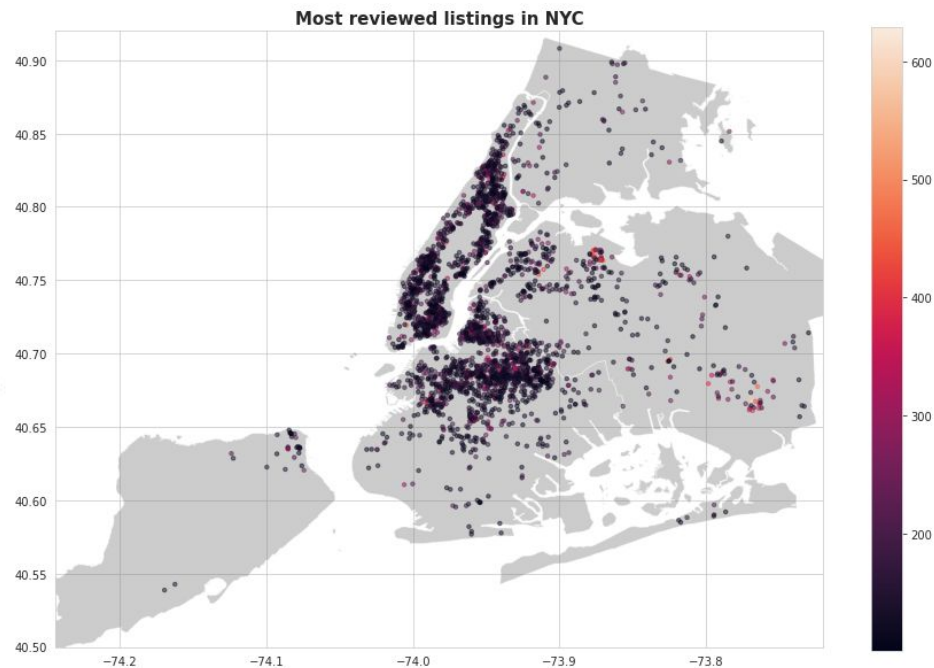
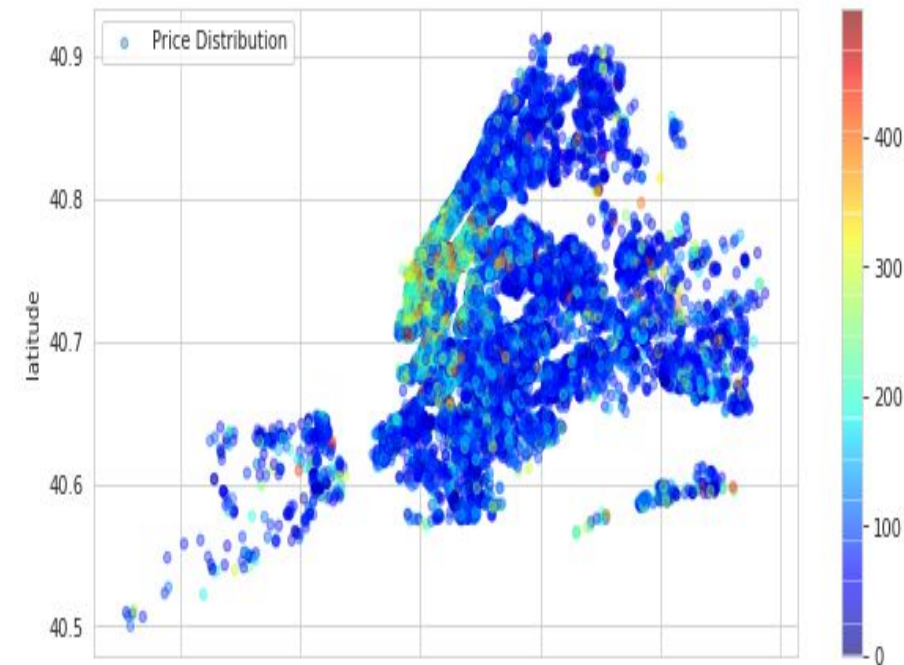
Busiest hosts and the reason behind that:



Listings with minimum nights restriction



Neighbourhood Price and Traffic



Frequent words in names of listings:





Challenges

- Lots of NaN values in the dataset.
- Difficulties while loading packages for maps.
- Cleaning the texts of name column.
- Dealing with the outliers.
- Faced some issues while dealing with availability column as nearly 30 % of those contains 0.
- Data visualization.

Conclusion

- People of New York City are mostly preferring entire-home/apt for longer stays and private rooms for shorter stays.
- Manhattan neighborhood-group have higher number of listings as well as it is the costliest as it has more number of notable places in New York like Central Park and Time Square.
- Most popular hosts have their mean listings price in the range of 50\$ to 200\$.
- Traffic of listings directly correlates with some specific keywords in the description.
- Prices of listings depends on the neighbourhood area and the room types and doesn't depend much on available number of listings.
- Most of the listings having minimum-night restrictions as 1 are private or shared and as the number of minimum nights is increasing the entire-home/apt is becoming more preferable.

Q & A