

Capstone Project

Airbnb Bookings Analysis

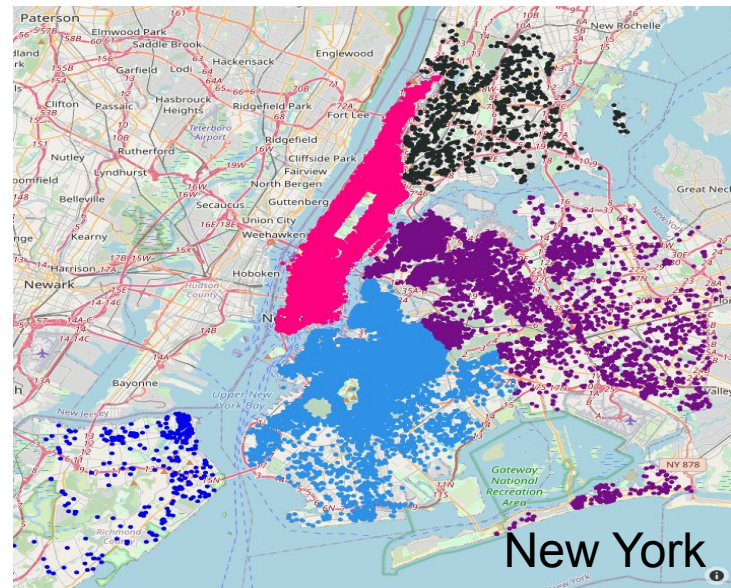
Mind Benders Team Members

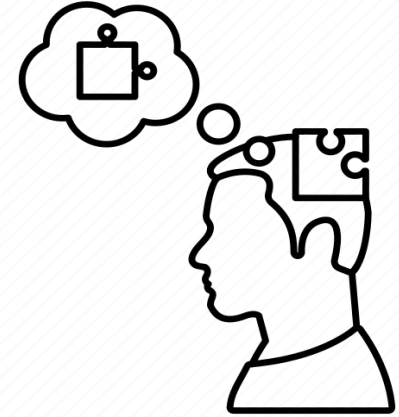
Abdul Aziz

Pooja Yadav

G M Sravya Sree

Abdullah Bin Mohammed





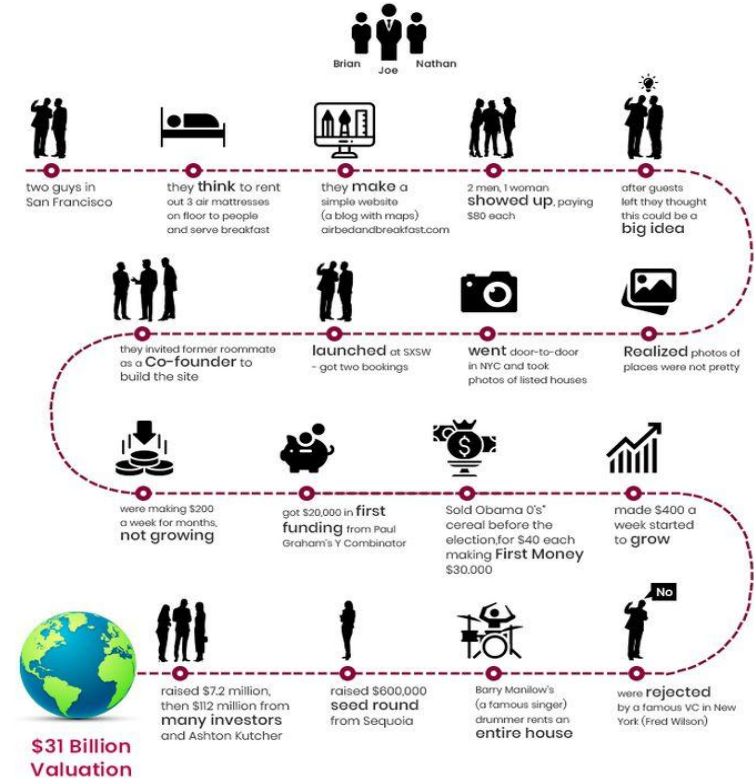
- Introduction to the Subject of Analysis - Airbnb.
- Motivation in Undertaking Airbnb for Analysis.
- Problem Statement.
- Presenting Dataset Sample.
- Exploratory Data Analysis Work Flow.
- Exercising Correlations, Patterns between various attributes and Visualization of the same.
- Inferences and Conclusions.
- Future Work.

About Airbnb



HOW AIRBNB STARTED

Or How 3 Guys Went From Renting Air Mattresses To
A \$31 Billion Dollar Company



Motivation

Millions of listings are generated around the world each year, near half a million of guests checks in per day.

Analyzing these data facilitates:

- ✓ Data driven business decisions.
- ✓ Assessing hosts behavior and performance.
- ✓ Understanding geographical influence.
- ✓ Helping customers in providing best accommodations.

Problem Statement

- Airbnb wants to draw more investors and hosts by advertising them the profitable investment opportunities based on the location.
- Airbnb also planned to facilitate its customers to find cheaper and better listings as per their needs.
- Analyzing the behaviour of top hosts of Airbnb.
- Geographical distribution of different room types across the neighbourhoods of New York.

Exploratory Data Analysis Work Flow

- 1.Data Preparation
- 2.Statistical Analysis of Data
- 3.Explore and Plot Data Relationships
- 4.Draw Useful Inferences and Make Data Driven Decisions



Data preparation

Observations: 48,895

Features: 16

Numerical	Categorical
LATITUDE	NEIGHBOURHOOD_GROUP
LONGITUDE	NEIGHBOURHOOD
PRICE	ROOM_TYPE
MINIMUM_NIGHTS	
NUMBER_OF_REVIEWS	
AVAILABILITY_365	

Columns dropped: **ID, NAME, HOST_NAME, LAST_REVIEW**

Null Values treatment: **REVIEWS_PER_MONTH** Replaced with zeros implying no reviews provided.

Understanding Data

Descriptive statistics on Airbnb prices

- Price data is heavily skewed.
- Third Quartile value is 175\$.
- Maximum price is 1000\$
- Price at 98th percentile is 550\$.

A convenient value for analysis can be 550\$.

count	48895.00
mean	152.72
std	240.15
min	0.00
25%	69.00
50%	106.00
75%	175.00
max	10000.00

Density plot for price Distribution.



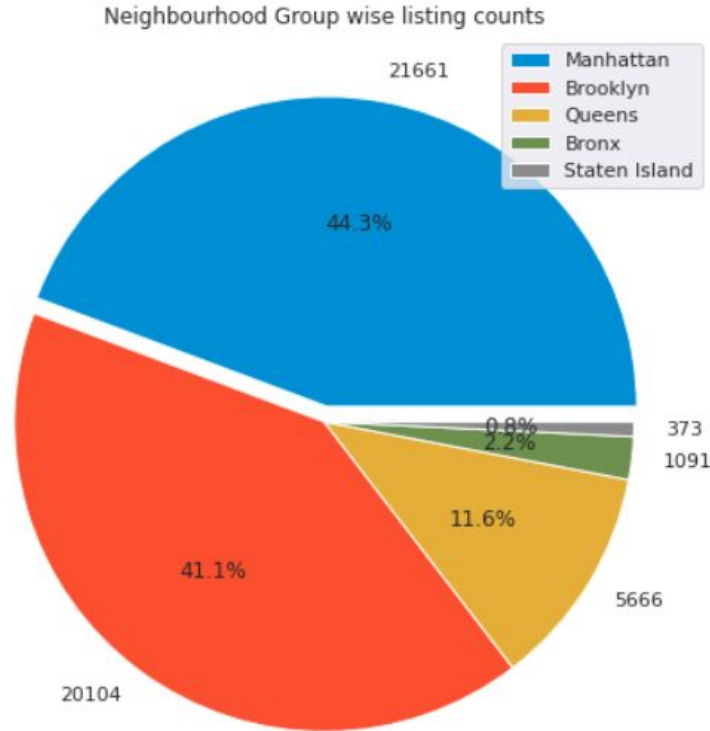
Neighborhood Groups
Manhattan
Brooklyn
Queens
Staten Island
Bronx

Room Type
Entire home
Private rooms
Shared rooms

Unique Neighborhoods
221

Maximum Price
550\$

Distribution of listings across different neighbourhoods.

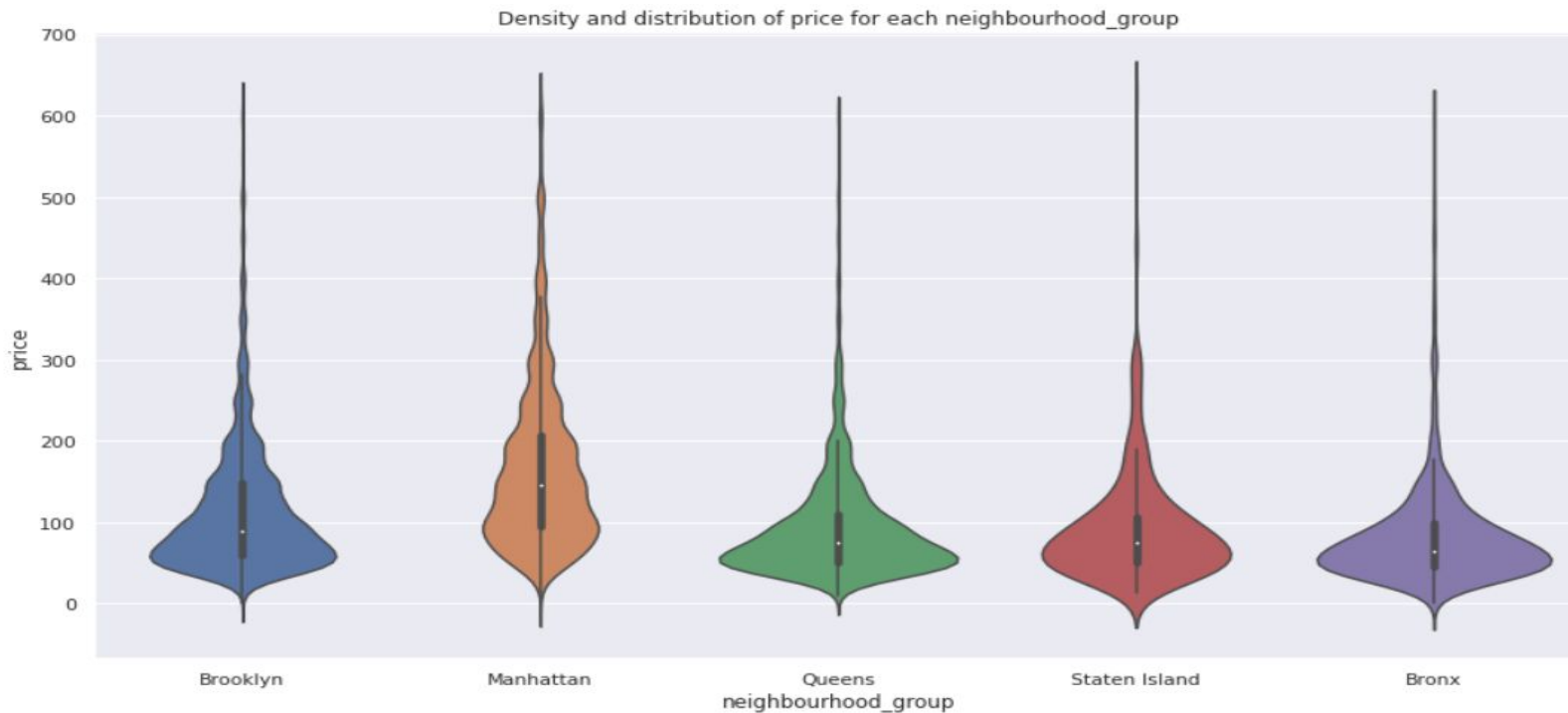


EDA

**Analyzing the prices across
neighborhoods**

EDA Violin Plot

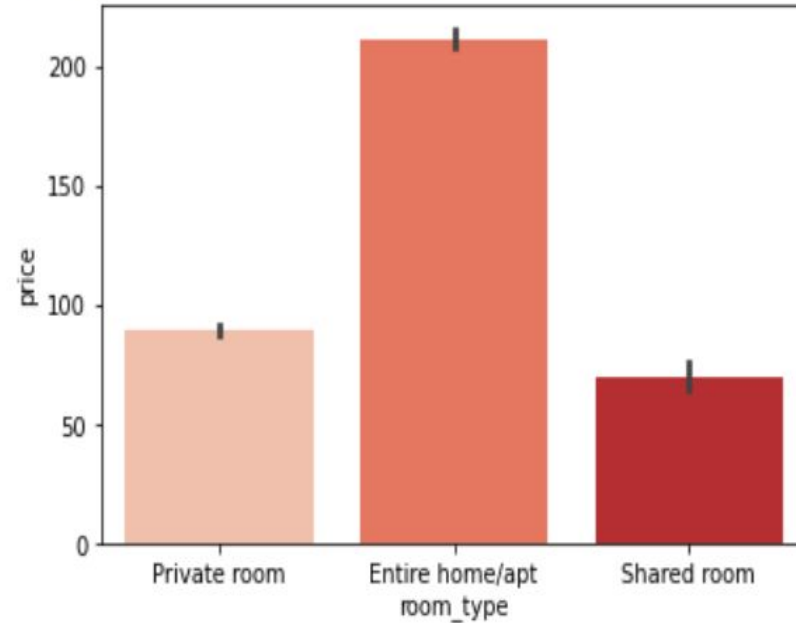
Price of Airbnb across neighborhood groups



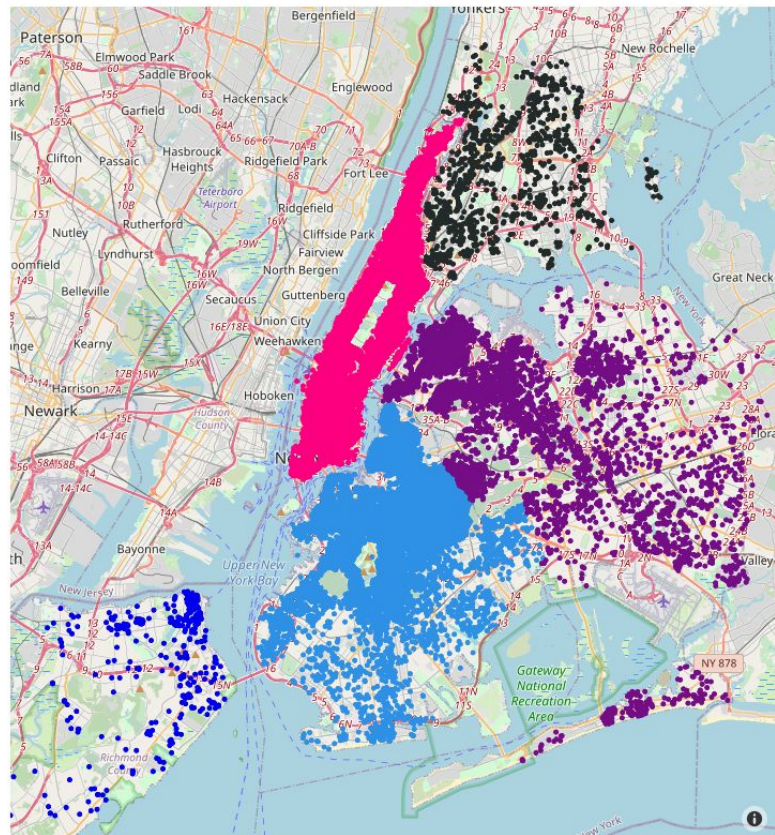
- Data skewness makes median price a better estimate for comparisons.
- Mean value would be influenced by the outliers.

Type of rooms available in the listings

<matplotlib.axes._subplots.AxesSubplot at 0x7f485c4d0e10>



EDA(continued)



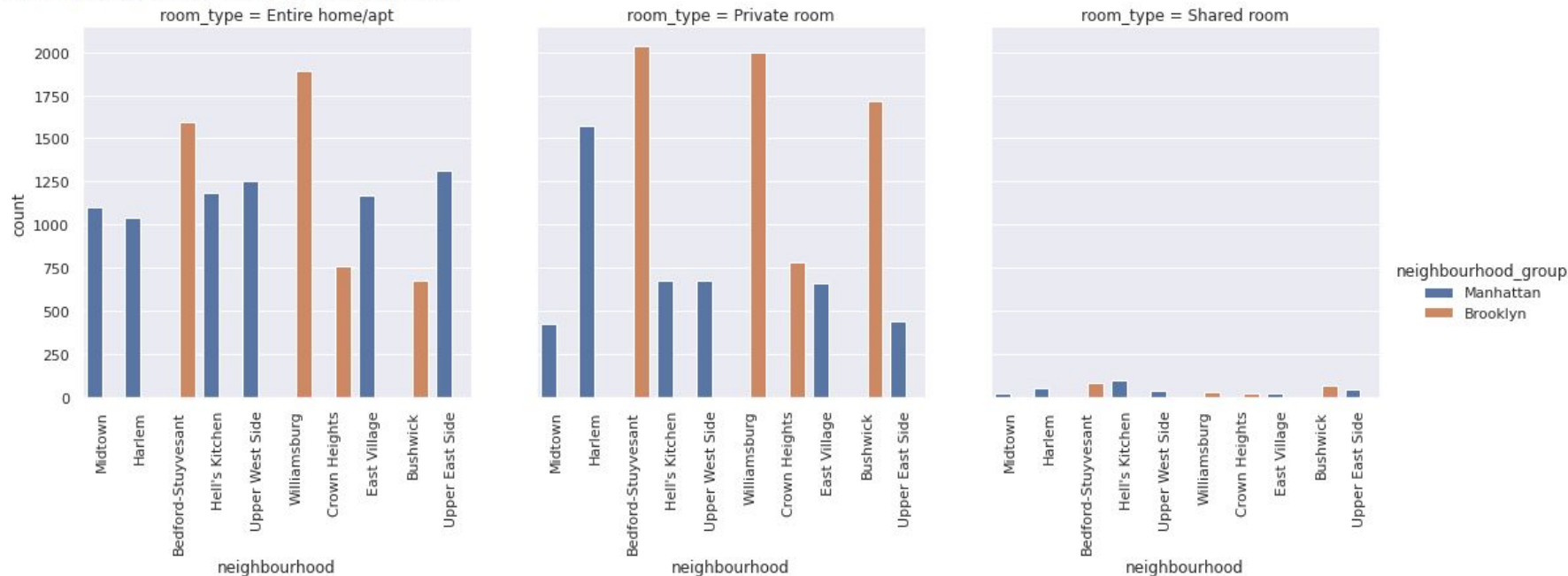
- neighbourhood_group=Brooklyn
- neighbourhood_group=Manhattan
- neighbourhood_group=Queens
- neighbourhood_group=Staten Island
- neighbourhood_group=Bronx

- Airbnbs are considerably denser in 'Manhattan' and 'Brooklyn'.
- This defines the higher demand of Airbnbs in those boroughs.
- Could be reason for the higher pricing of Airbnb there.

EDA(continued)

Airbnb distribution across all neighborhoods

```
<seaborn.axisgrid.FacetGrid at 0x7f484fb6ec10>
```



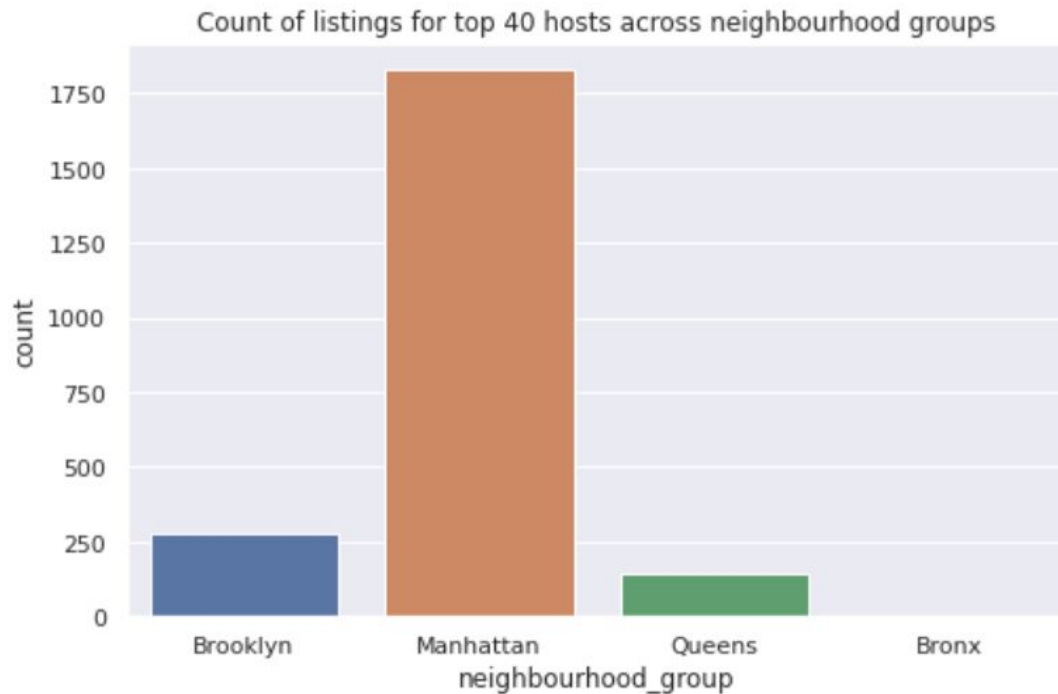
Understanding the behaviour of top hosts

EDA(continued)

Identifying the Hosts with most number of listings and their listed areas

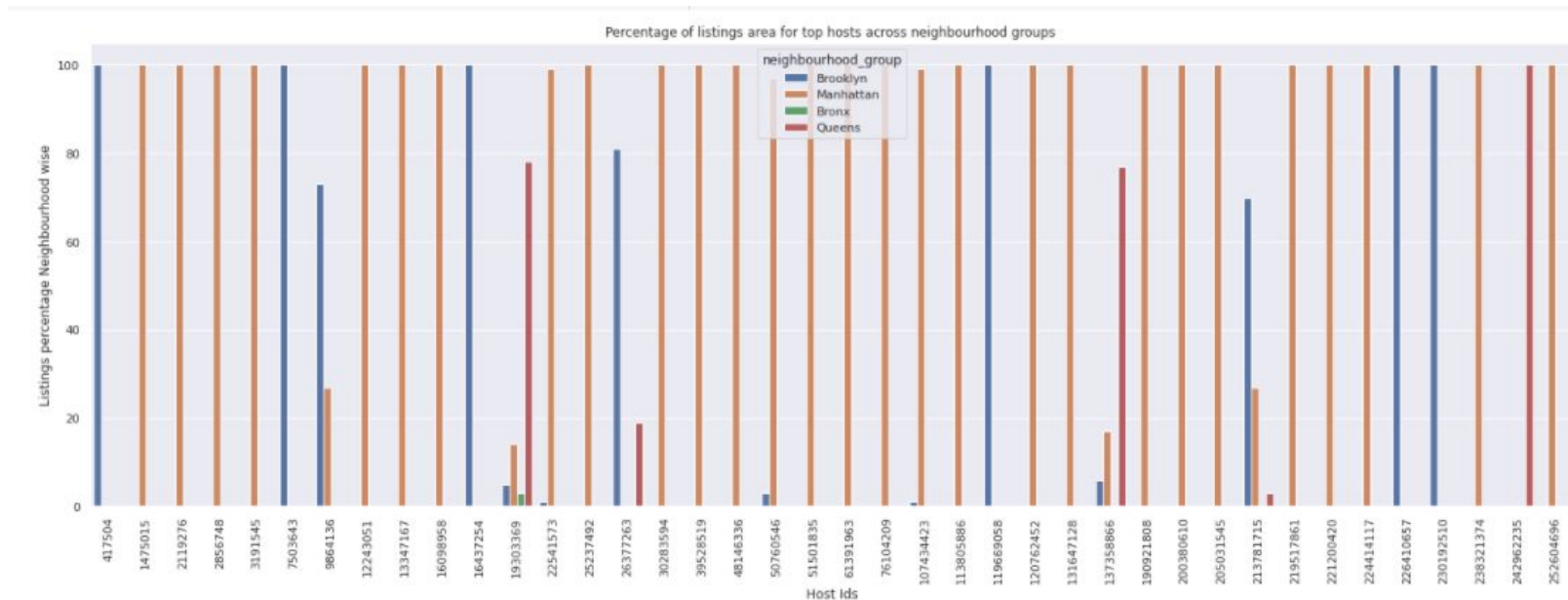
Top hosts have listed
mainly in Manhattan.

None of them have
apartment in Staten Island.



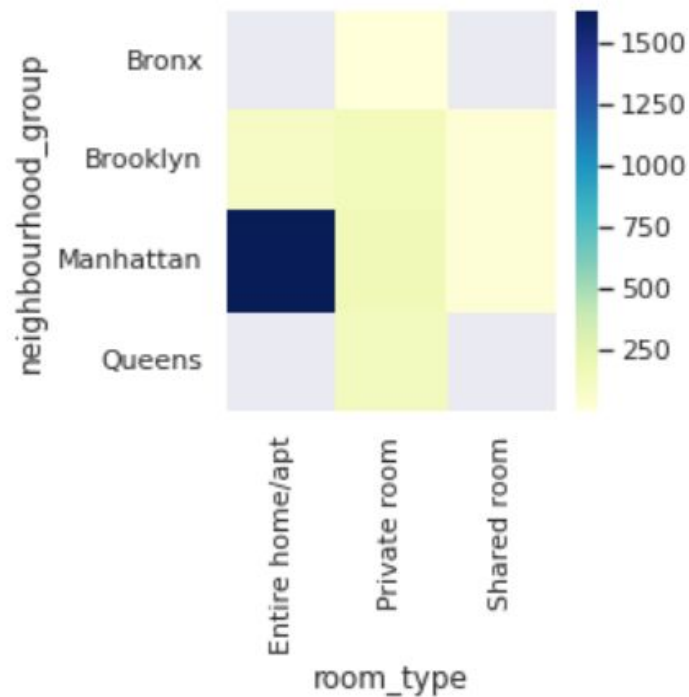
EDA(continued)

Percentage wise listings distribution in different areas for the top 40 hosts



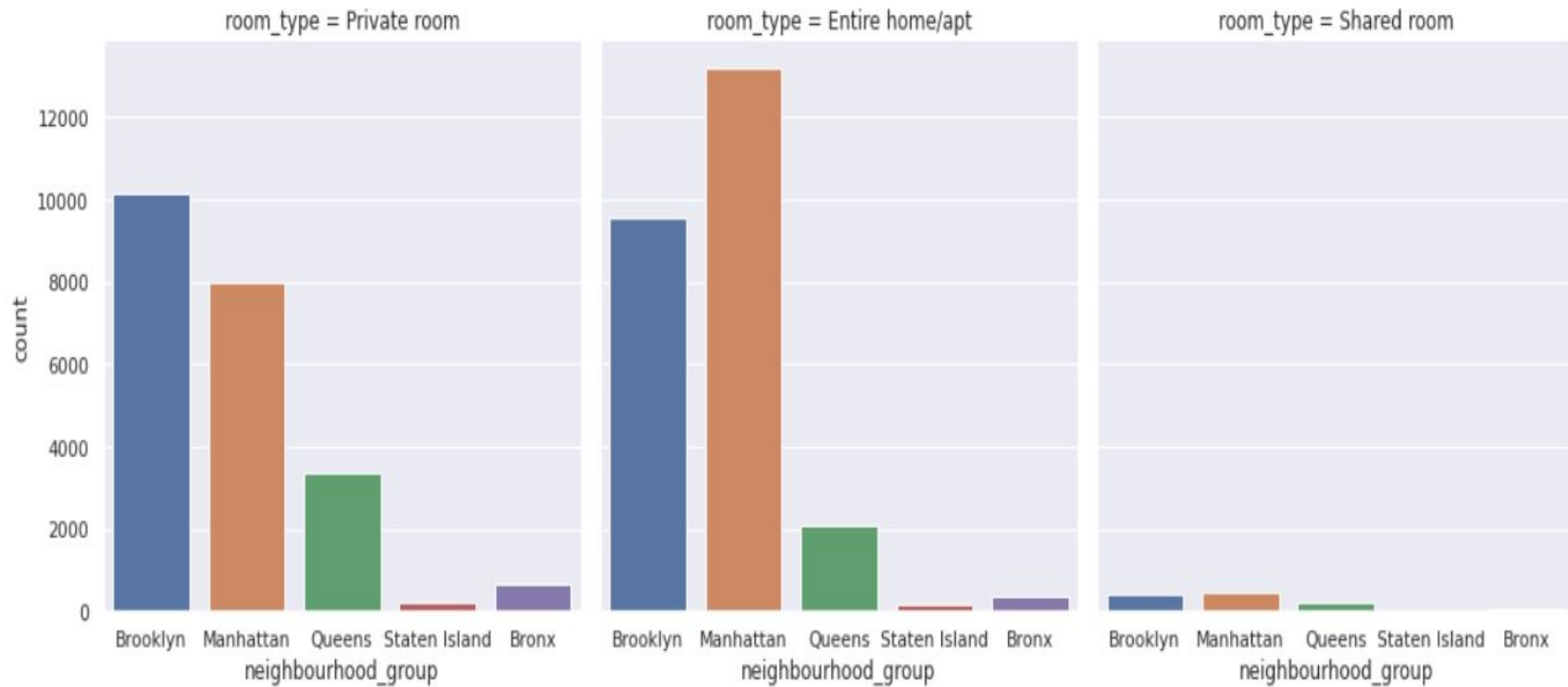
EDA(continued)

Heat Map for top hosts v/s neighbourhood group v/s room type



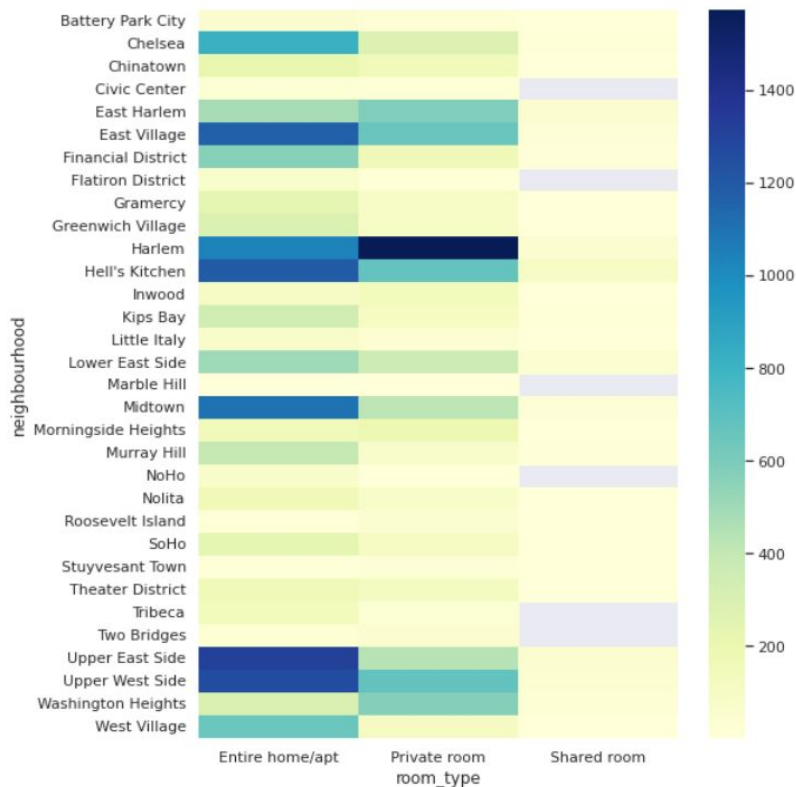
Analysis about the room types across neighborhood groups

EDA(continued)



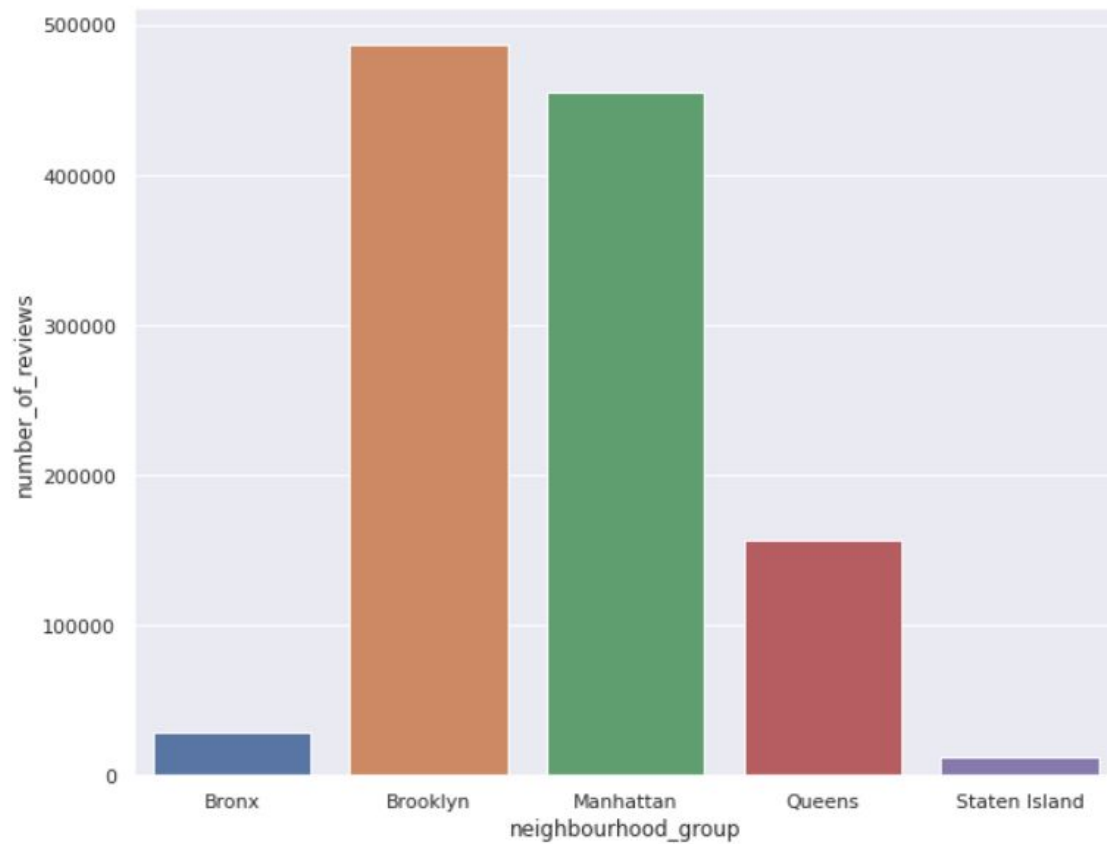
EDA(continued)

Is there any pattern for room type distribution in different neighborhoods in Manhattan



Analysis based on reviews

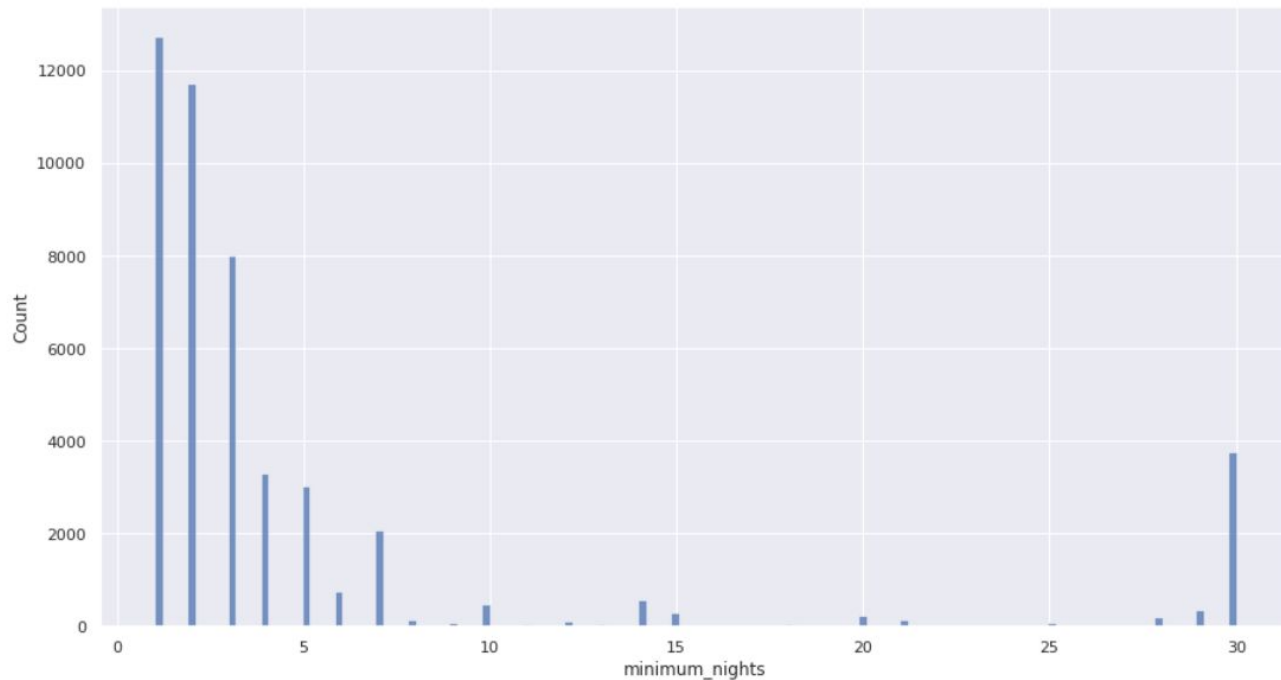
EDA(continued)



How the bookings for certain number of nights affect the prices?

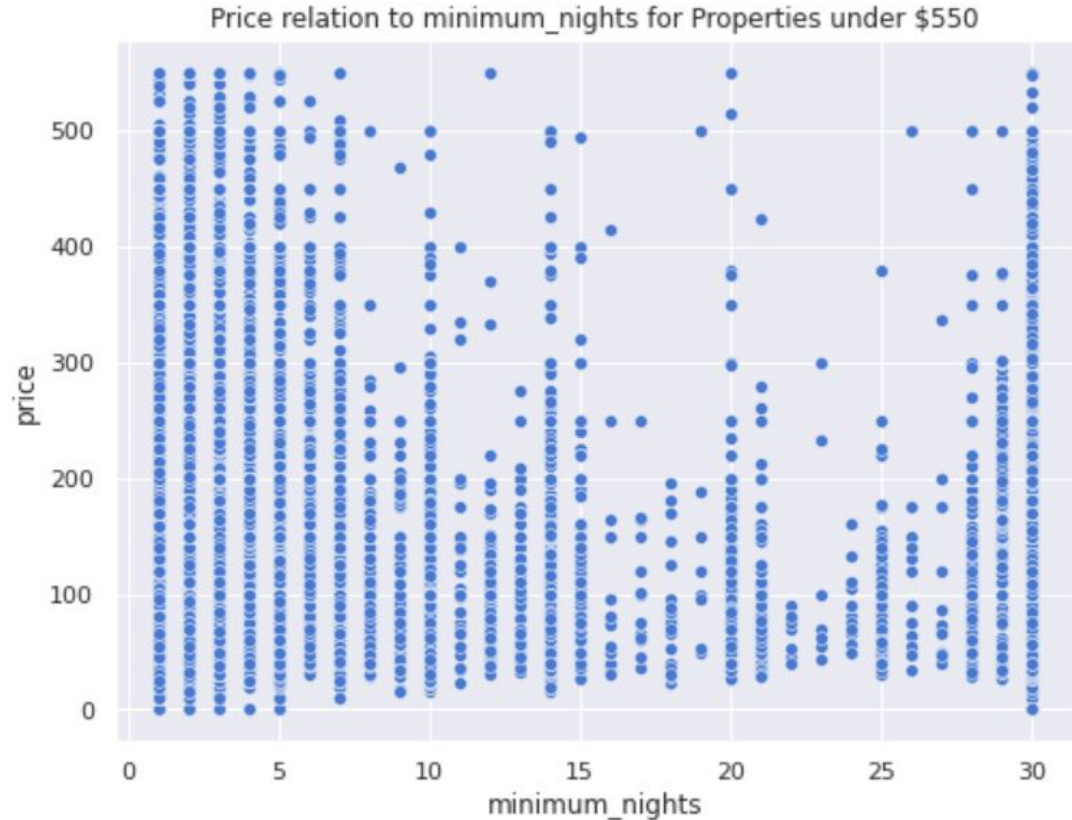
EDA(continued)

minimum_nights	
count	48895
mean	7
std	21
min	1
25%	1
50%	3
75%	5
max	1250



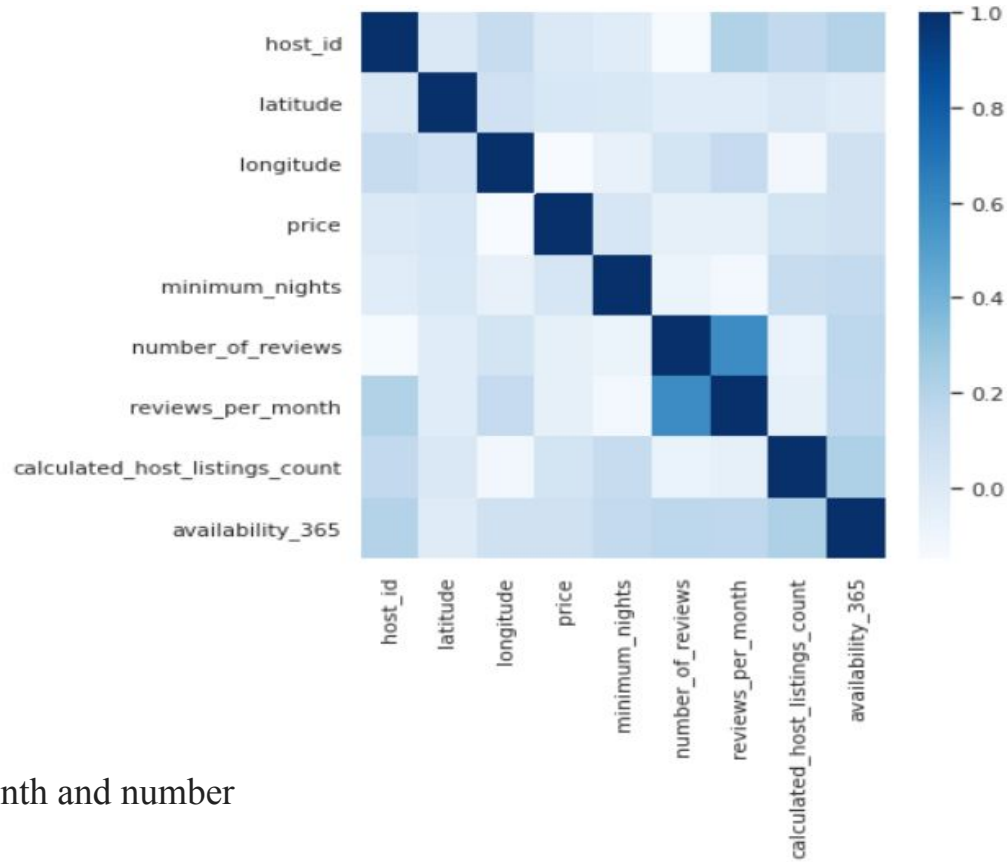
EDA(continued)

- First half the plot is more dense implying wide range of price options.
- Around 25 a shorter range signifies lesser rooms and also cheaper prices.



EDA(continued)

Correlation matrix



High correlation b/w reviews per month and number of reviews.

Conclusion:

- Evident that Manhattan and Brooklyn are the neighborhood groups with most number of Airbnb rooms.
- Room types listed mostly is either Entire home or Private rooms.
- Manhattan has the highest median price followed by Brooklyn.
- Reviews across Manhattan and Brooklyn further says they are most popular.
- Price ranges options based on nights can help plan his expenses for longer stay.
- Helpful for a new host who wants to enter into business with Airbnb.
- Visitors get to know the price range for different room types across neighborhoods.
- Demographic analysis lets visitors book rooms as per convenience.

Future Work

For better analysis further data can be collected like the reviews, categorized as positive and negative or provided with the ratings or stars on a scale of (0-5).

Inclusion of this feature might have extended the scope of analysis, that is to list the highly rated hosts or the listings.

Data can be to fit a Machine Learning regression model to predict the prices of Airbnb listings.

Thank You!!