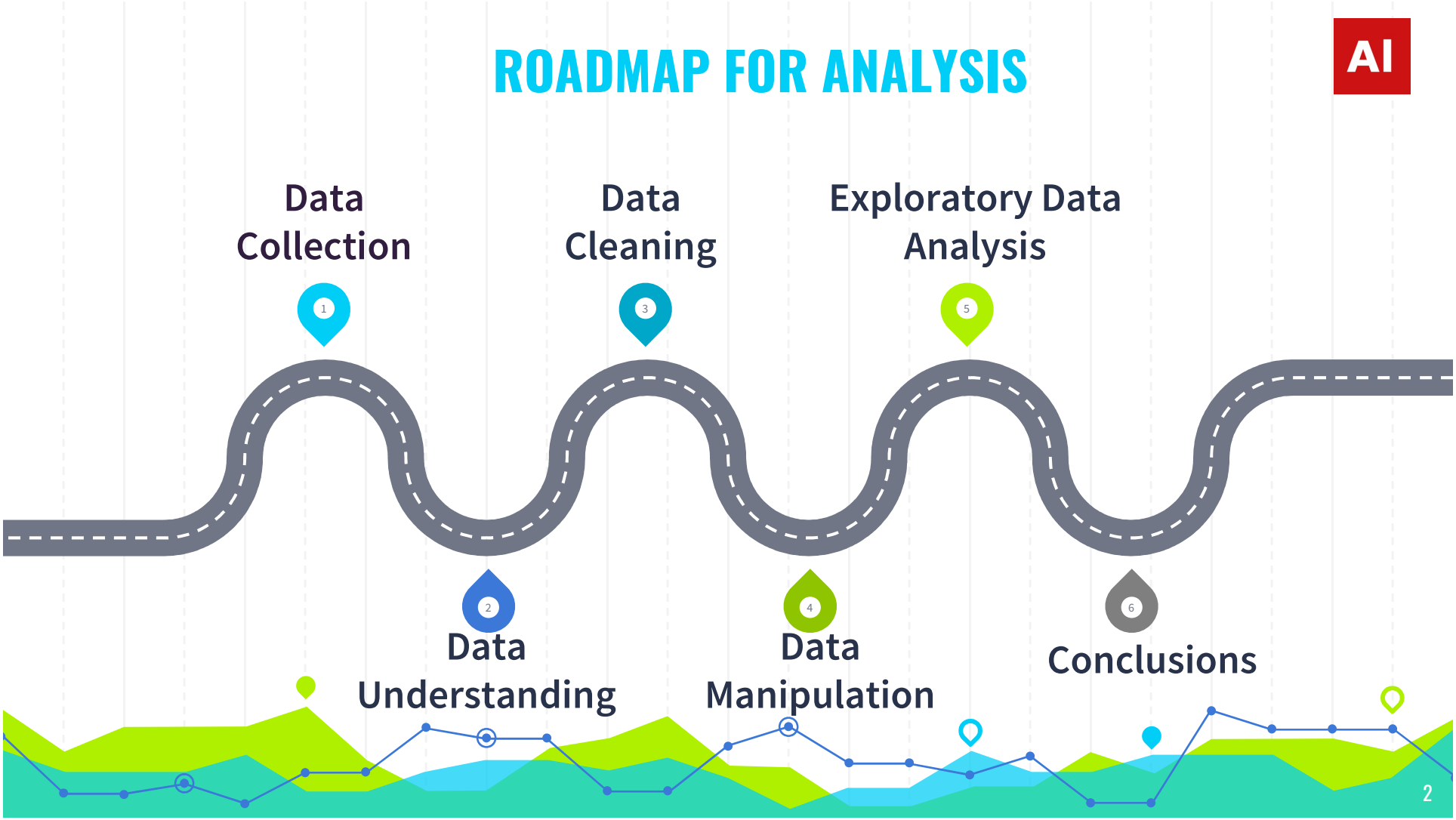
CAPSTONE PROJECT -1

Airbnb Booking Analysis

Team Members

Mritunjay Sarkar
Akash Choudary
Kajal Mahajan
Adarsh Gaurav
Vivek Raikwar

ROADMAP FOR ANALYSIS

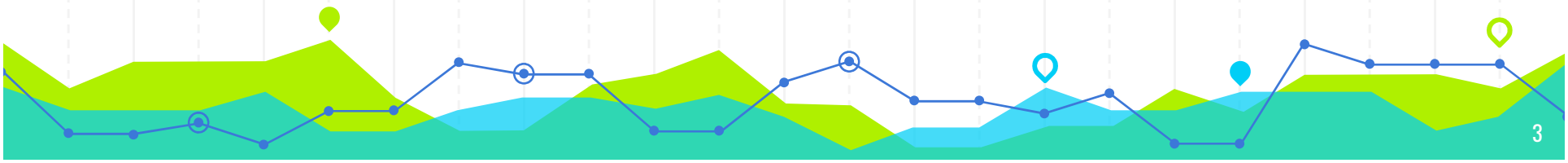


What is Airbnb ?

AI

Airbnb is an online marketplace that connects people who want to rent out their homes with people looking for accommodations in that locale. NYC is the most populous city in the United States, and one of the most popular tourism and business places globally.Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world

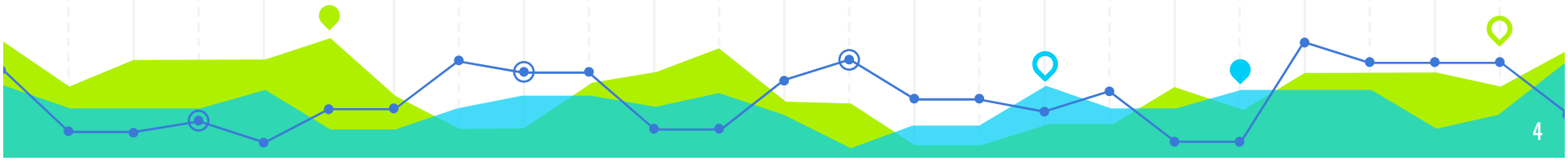
★ **In this project we have taken *Airbnb* booking data for New York City of the year 2019 (henceforth will be represented as AIRBNB_NYC_2019).**



❖ This dataset contain around 49000 observations, distributed among 16 columns.

Data Features

- Id
- Name
- Host name
- Neighbourhood group
- Neighbourhood
- Latitude
- Longitude
- Room type
- Price
- Number of reviews
- Minimum nights
- Last review
- Reviews per month
- Calculated host listings
- Availability
- Host id



Data Loading and Exploration

1.Data loading:loading data from google drive and reading in notebook

2.Data exploration:checking data in different columns

3.Numerical columns and categorical columns

4.Information and data types of column

AI

Data Wrangling

Data wrangling is a process of cleaning and unifying messy and complex data sets for easy access analysis

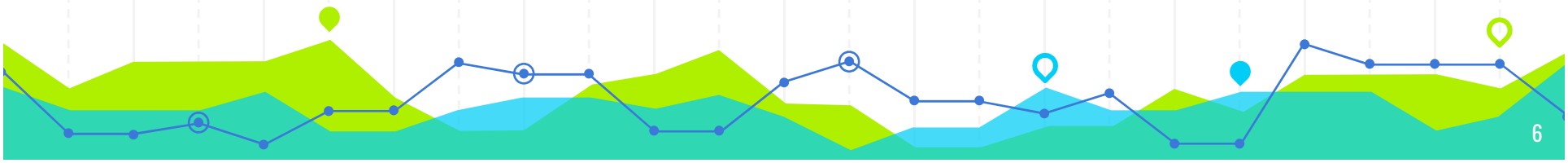
It Includes Following Steps

STEP 1
Handling Missing Values

STEP 2
Removing Duplicate Data

STEP 3
converting column to proper dtype format

STEP 4
Adding and removing columns for analysis

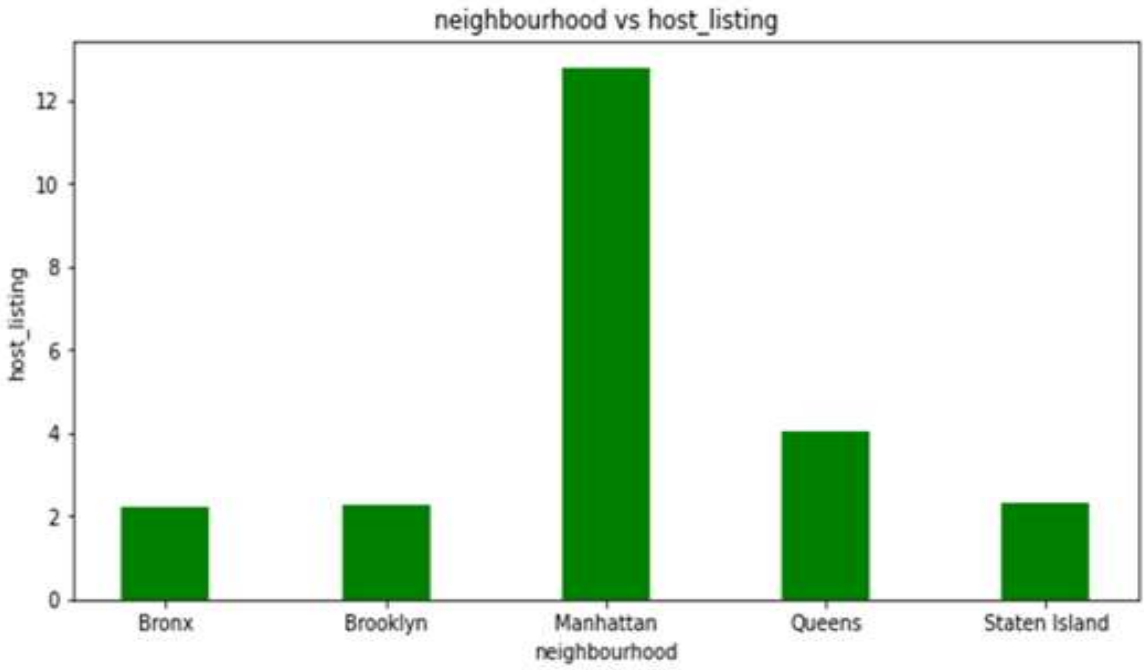


Problem Statements For EDA

1. What can we learn about different hosts and areas?
2. What can we learn from predictions? (ex: locations, prices, reviews, etc)
3. Which hosts are the busiest and why?
4. Is there any noticeable difference of traffic among different areas and what could be the reason for it?
5. Depending upon the booking, can we segregate the neighborhood among various tires?
6. What is the exact postal address for the best reviewed hotel?
7. Which room has the lowest reviews ?
8. Relationship between number of reviews and price ?

What can we learn about different hosts and areas?

we have done a twofold analysis, first: to assess which host is most popular and second: in general, which neighborhood is most trusted.

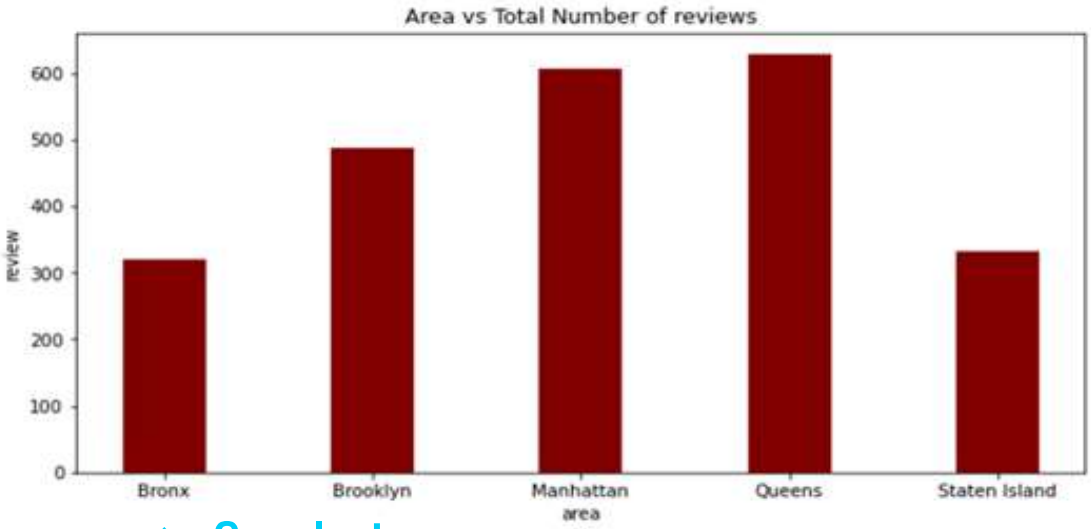


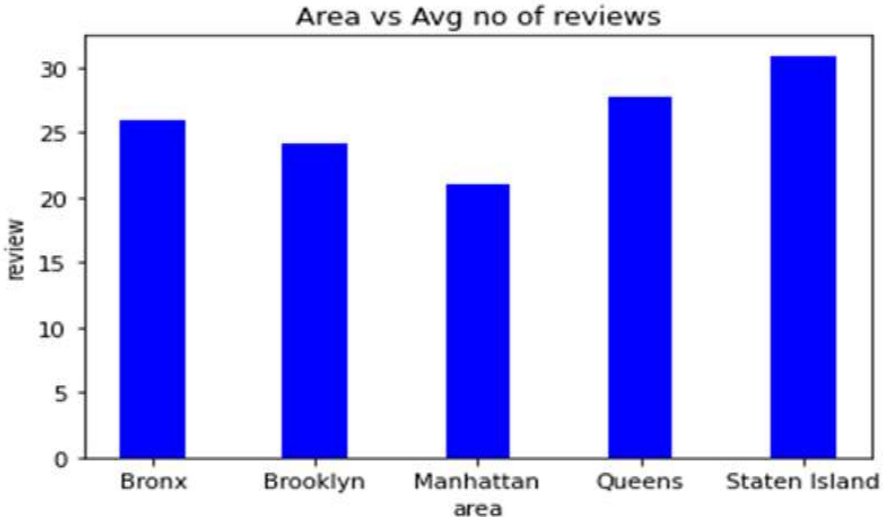
❖ Conclusion

- SONDER(NYC),Blueground,Kara and Kazuya comes out tobe best host
- Manhattan is most preferred by guest , followed by Queens, Staten Island,Brooklyn and Bronx

What can we learn from predictions? (ex: locations, prices, reviews, etc)

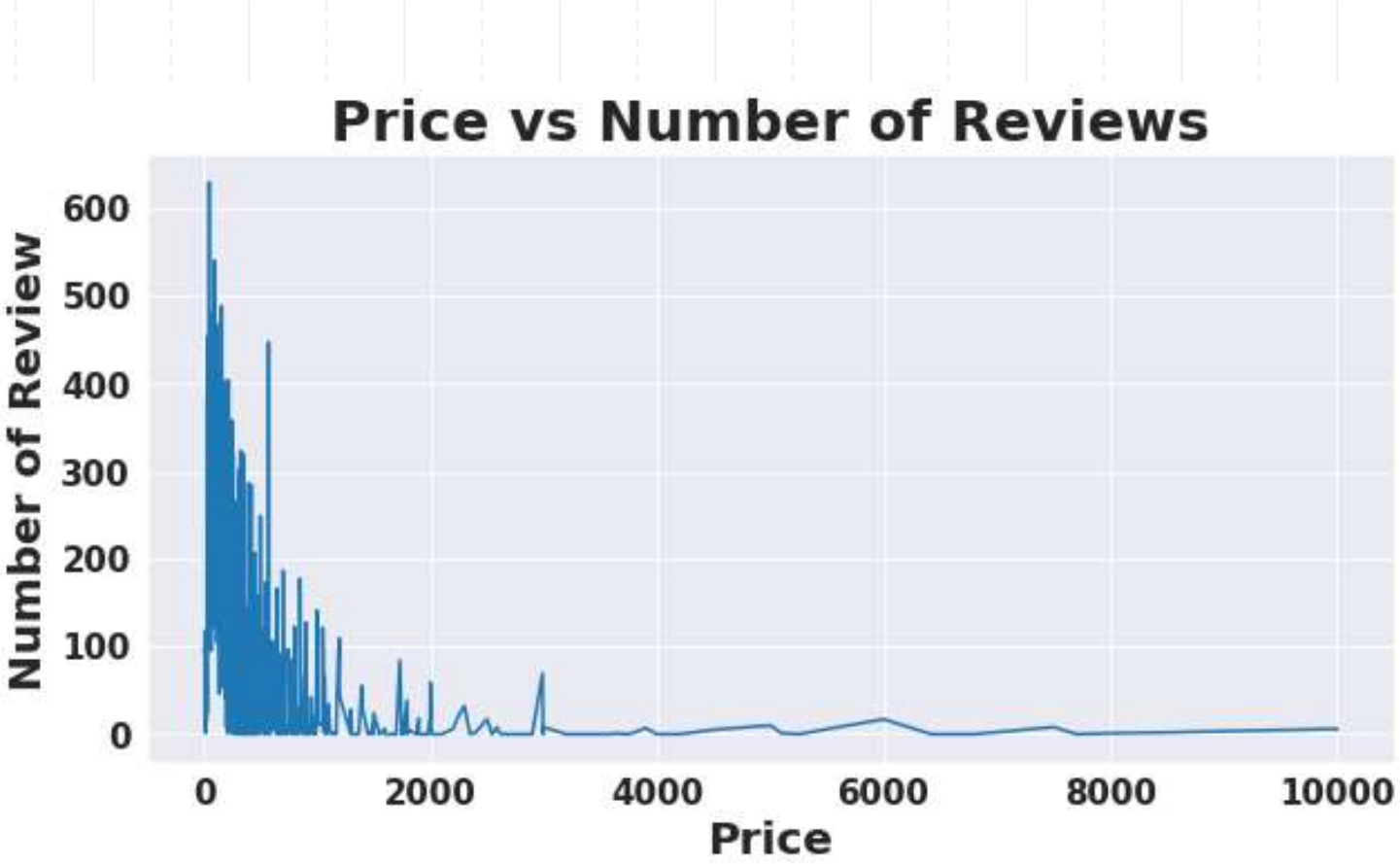
For location wise review analysis, we have considered both overall review and average review





❖ Conclusion

- Manhattan area fetched most number of reviews
- Other places have got less reviews but there average review is better than Manhattan
- Manhattan has more number of host/hotels compared to others
- Guest are bound to stay in Manhattan because of lesser no of host in other neighbourhood



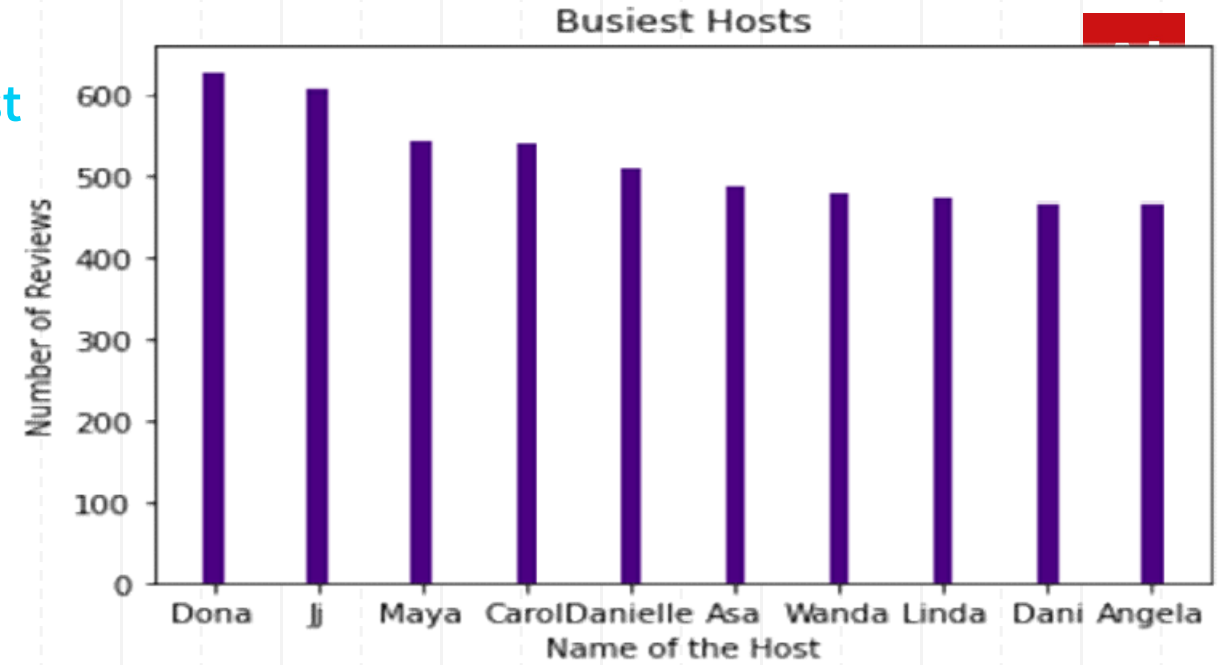
AI

Price Vs No of reviews

The figure shows that hosts, who are charging more, have a very poor review. So the guests generally choose cheaper hosts, which is quite obvious.

Which hosts are the busiest and why ?

Our quests in search of Busiest Hosts are fulfilled by the total number of reviews with the help of number of reviews we will find busiest host

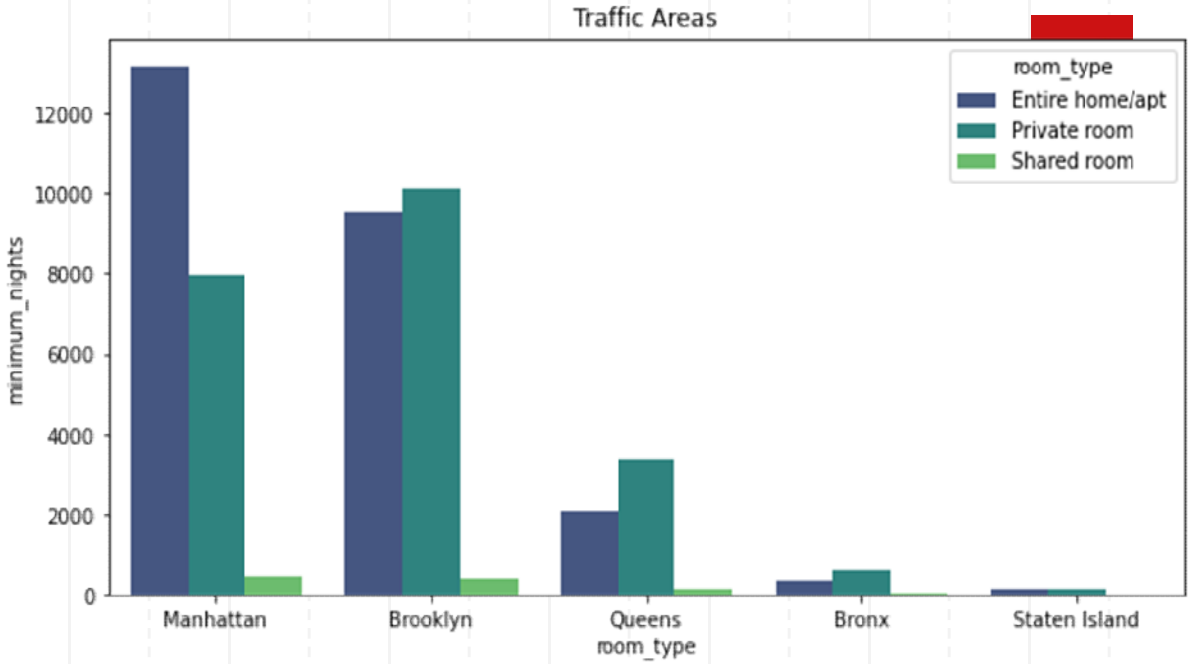


❖ Conclusion

- Top 10 busiest host are Dona,Jj,Maya,Carol,Danielle,Asa,Wanda,Linda,Dani and Angela
- Dona’s hospitality is best

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

We have find noticeable difference of traffic among different areas by plotting the barplot

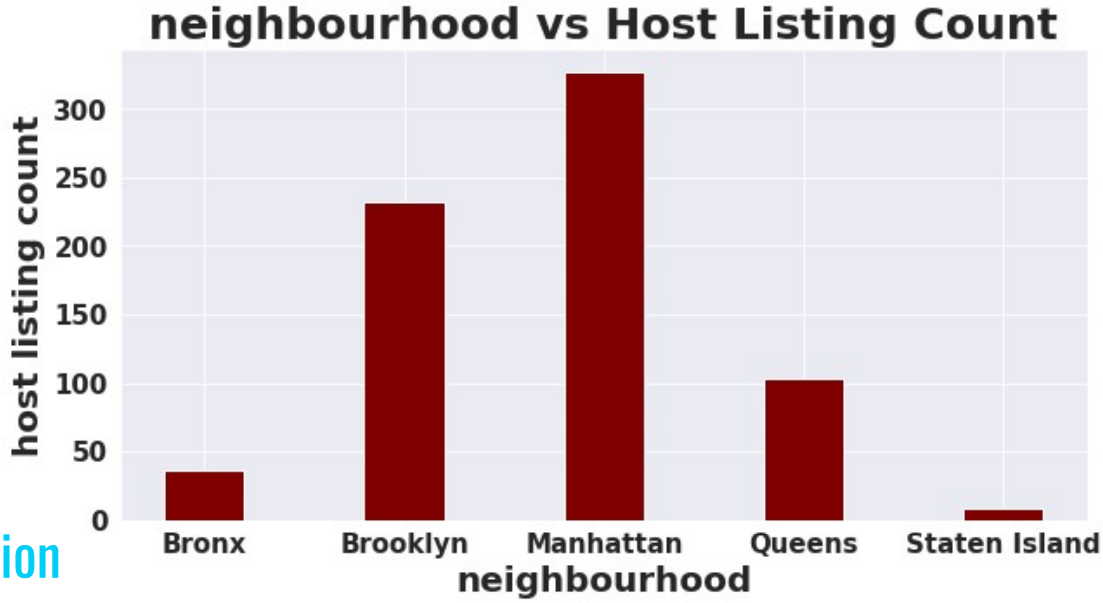


❖ Conclusion

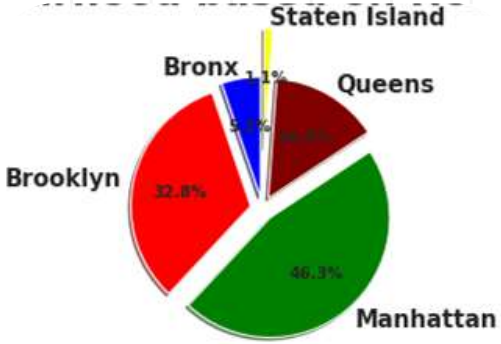
- People are preferring entire home/apt which is present in Manhattan
- Entire home/apt has the highest number of listings among other room type
- Private rooms are more in Brooklyn
- The total number of shared room is way more less than other room type

Depending upon the booking, can we segregate the neighborhood among various tires?

In this problem we have tried to divide neighbourhood in tiers by using there host listing count



❖ Conclusion



- Manhattan has the maximum host share with a whopping 46.3% market share
- Brooklyn,Queen ,Bronx and Staten Island with 32.8,%14.6%,5.2%and 1.1% of market share respectively
- Manhattan can be coined as Tier 1 followed by others Tier 2,Tier 3,TIer 4 and Tier 5

What is the exact postal address for the best reviewed hotel?

We will find the exact postal location for the most reviewed hosts. The idea behind this is to cater the premium guest only, who doesn't bother about price. By using the 'Google Map' feature of any smart phone, one can reach the corresponding hotel / host very easily.

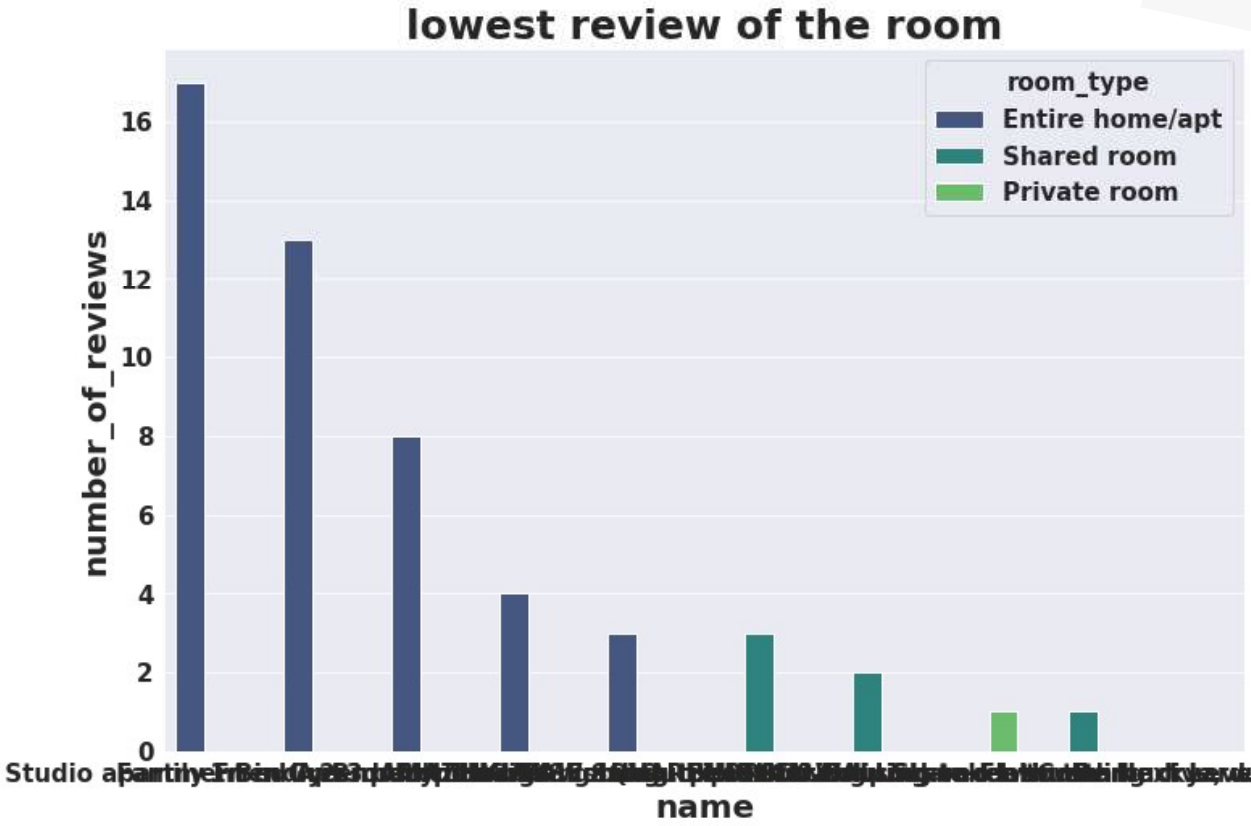
❖ Conclusion

The Exact Address is

Latitude: 40.75918
Longitude: -73.98801
Review: 58.5

Location of the maximum review is :
726, 8th Avenue, Theater District, Manhattan Community Board 5, New York County, New York, 10036, United States

Which room has the lowest reviews ?

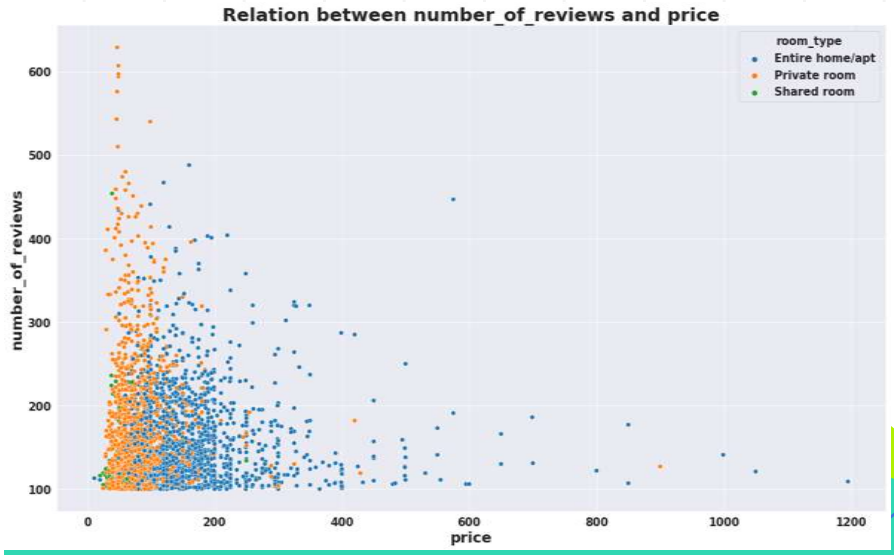


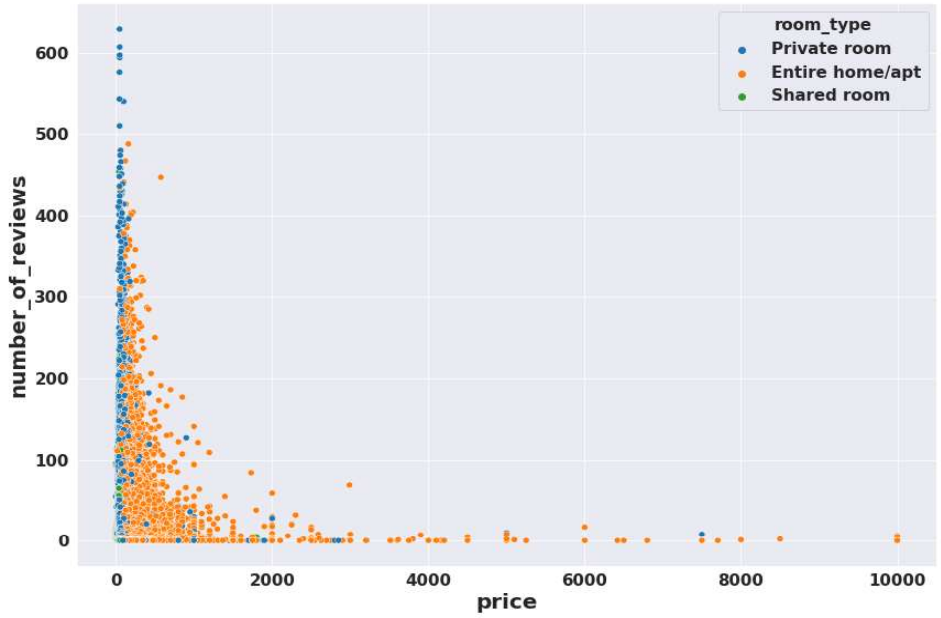
Conclusion

- Shared room have lowest number of reviews
- If price is more then no of reviews are less. It means people choose cheap hotels

Relationship between number of reviews and price ?

- Private rooms are cheap and are most favoured.
- Entire room is costlier, but still manages to get above average review, means people also choose them





- Shared rooms, although cheap, but fetched a poor review. Means people choose privacy over price

Conclusion

- Manhattan is most preferred by guests, followed by Queens, Staten Island Brooklyn and Bronx
- Average review however shows that guests are satisfied with other neighbourhood too. Maybe this is due to the fact that Manhattan has most hosts, thus fetching a high value of total reviewes
- the guests generally choose cheaper hosts, which is quite obvious
- People are preferring entire home/apt or private room which are present in Manhattan , Brooklyn, and Queens.
- As the price is increasing then number of reviews decreasing so we can say that more price is better dealing with customers.

★ Challenges Faced

- ❑ Reading the dataset and understanding the data the meaning of some columns
- ❑ Internet connectivity was an issue
- ❑ choosing visualization for different analysis
- ❑ Handling null values and finding a way to replace them with something meaningful so that it doesn't affect analysis.

AI

“

THANK YOU