CAPSTONE PROJECT -1

Airbnb Booking Analysis

- Data collection
- Data understanding
- Data cleaning
- Data manipulation
- Exploratory data analysis
- Conclusion

AIRBNB

- What is Airbnb ?

- 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).

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.
- Information and data types of column.

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

Data Features

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

Data Wrangling

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

It Includes Following Steps

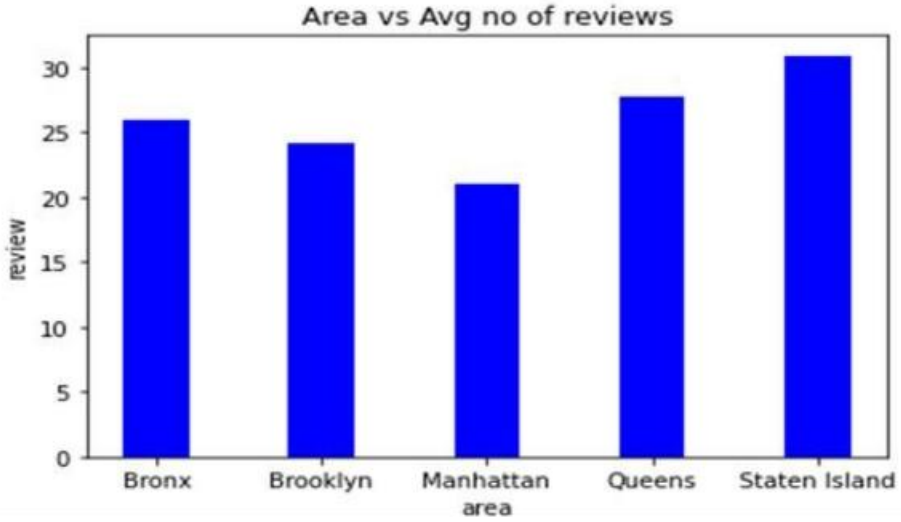
- STEP 1 :
- Handling Missing Values
- Step 2 :
- Removing duplicate data
- Step 3 :
- Converting column to proper Dtype format
- Step 4 :
- Adding and removing columns for analysis

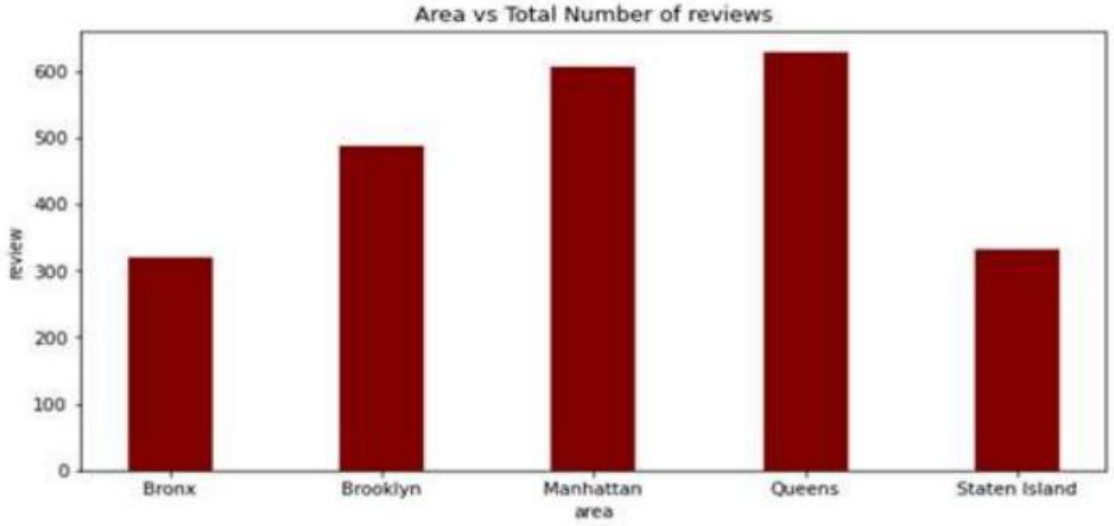
Problem Statements For EDA

- 1.What can we learn from predictions? (ex: locations, prices, reviews, etc)
- 2.Which hosts are the busiest and why?
- 3.What can we learn about different host and area ?
- 4.Relationship between number of reviews and price ?
- 5Which room has the lowest review ?

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

- For location wise review analysis, we have considered both overall





- **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 neighborhood.

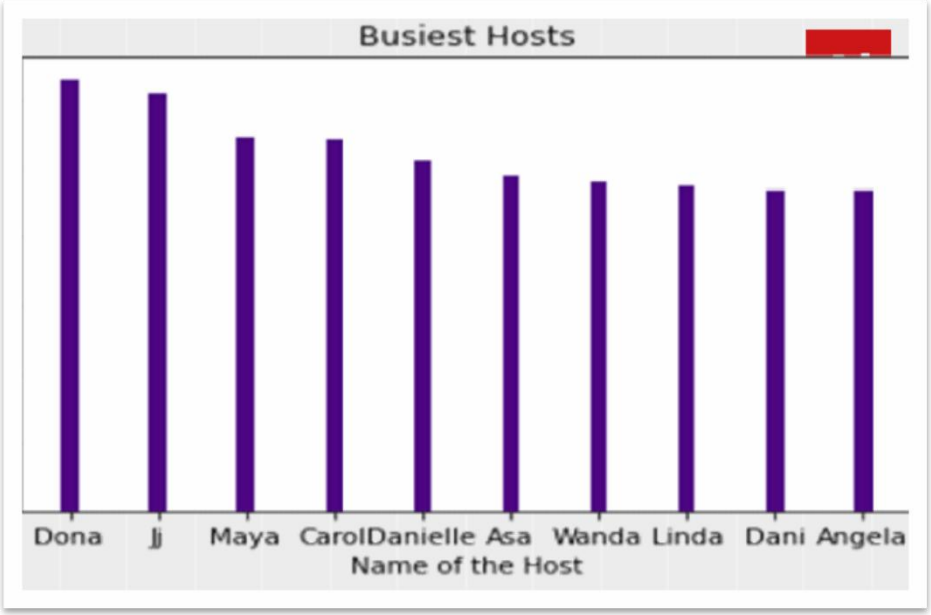
Price Vs No of reviews

- 

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

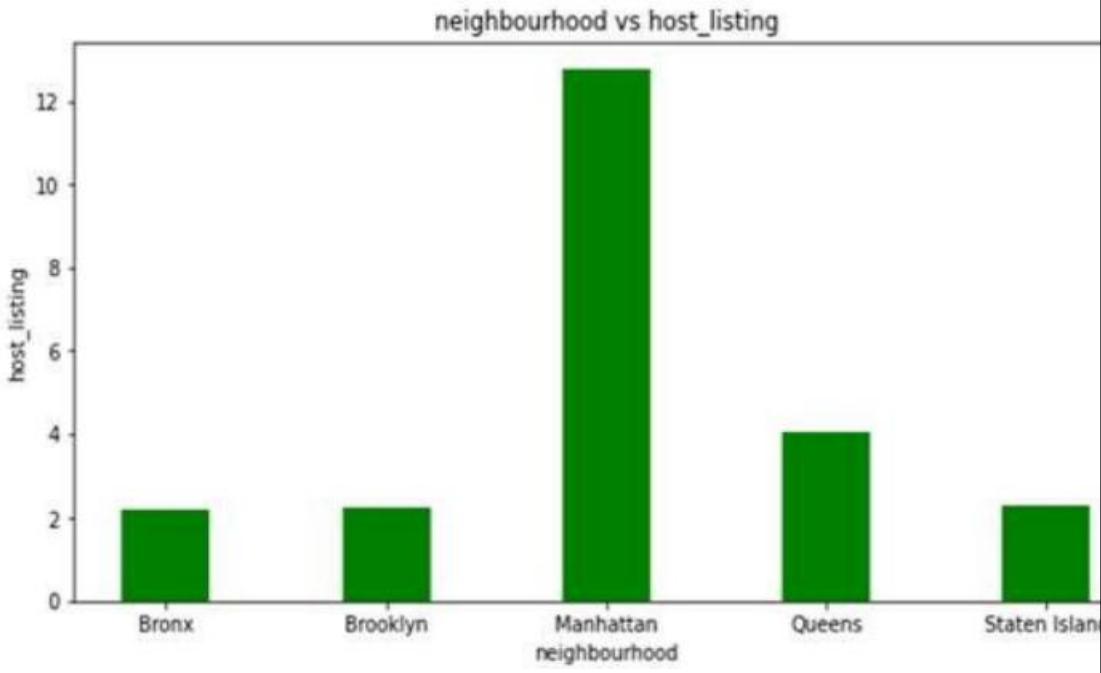


❖ Conclusion

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

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.

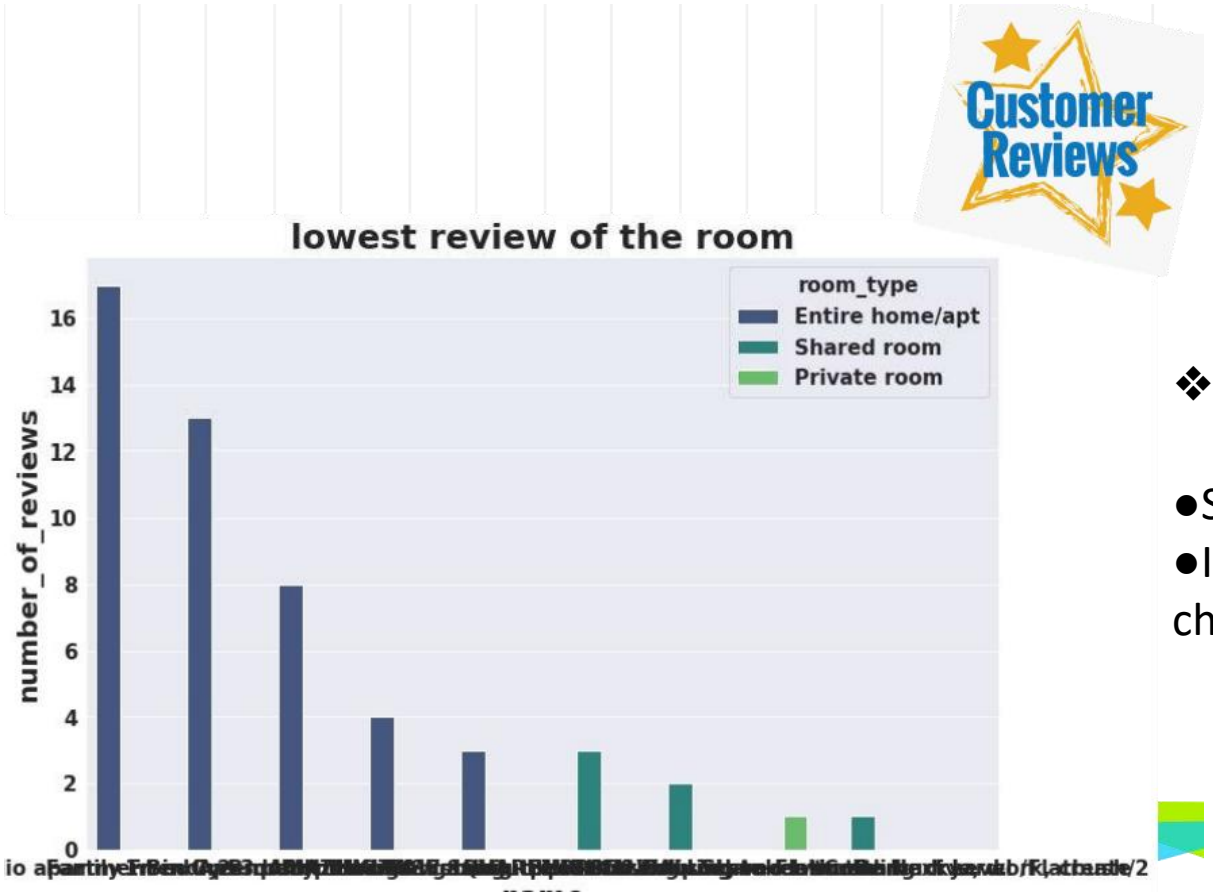


❖ Conclusion

●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

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 favorite.
- ●Entire room is costlier, but still manages to get above average review
- ●Shared rooms, although cheap, but fetched a poor review. Means people choose privacy over price, means people also choose them.

Conclusion

- 1. As the price is increasing then number of review decreasing so we can say that more price is better dealing with customers.
- 2. Manhattan is most preferred by guests, followed by Queens, staten island, Brooklyn and Bronx.
- 3. The guest generally choose cheaper hosts, which is quite obvious.
- 4. Average review however shows that guest area a satisfied with other neighbourhood too. May be this is due to the fact that Manhattan has most hosts, thus fetching a high value of total review.
- 5. People are preferring entire home/apt or private rooms which are present in Manhattan, Brooklyn and queens.

Challenges faces

- 1. Facing internet connectivity was an issue.
- 2. Reading the clean set and understanding the data the meaning of some column.
- 3. Choosing visualization different analysis.
- 4.I had made the project alone, so I had to face a lot of problem related PPT part.

THANK YOU