Capstone Project –1

Airbnb Booking Analysis

By- Pratik Chaudhari

Aishwarye Gangwar

Abhinav Anand

Manish Gupta

Rutul Dhote

Point of Discussion

- About Airbnb
- Data summary
- Problems in Data set
- Building a correlational graph
- Data cleaning
- Information about hosts and neighbourhood
- Prediction on the bases of our Data set
- Busiest host of entire Data set
- Traffic across the area and reasons
- Most preferable properties
- Properties availabilities
- Airbnb across the whole neighbourhood
- Conclusion

About Airbnb

- Airbnb is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities.
- Based in San Francisco, California, the platform is accessible via website and mobile app.
- Airbnb does not own any of the listed properties instead, it profits by receiving commission from each booking.
- The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com.

Data summary

df Airbnb = In our data set, we have 16 columns and 48895 rows.

Columns

- [id] = Every Airbnb has an unique number that represented as id in this data frame.
- [name] = For specification of each Airbnb property a name was assigned which is represented by this column.
- [host id] = Every host can be identified by a unique id which is represented as host id.
- [host name] = Every host assigned by a specific name which is represented by column host name
- [neighbourhood group] = This column represent global area like whole city.
- [neighbourhood] = This column represent surrounding local area around the property.
- [availability_365] = Availability of each Airbnb in a year throughout represented by this column.

Data summary

- [latitude] = The angular distance of each Airbnb with respect to earth equator along horizontal direction represented by this column.
- [longitude]= The angular distance of each Airbnb with respect to Greenwich meridian from left to right represented in this column.
- [room type] = Each Airbnb has a specification of room that is described as room type column.
- [price] = Every Airbnb has a standard price value according to their property represented in this column.
- [minimum nights] = Minimum night halt by a person in an Airbnb represented by this column.
- [number of reviews] = Every Airbnb has certain review received by person up to date is represented in this column.
- [last review] =A specific date represented by Airbnb which shows last review receive date.
- [reviews per month] =Number of reviews per month Airbnb received by public represented by this column
- [calculated host listing count] = This column represent the number of hosts has been listed on the property.

Problems in Dataset

- Null value in data frame represent missing value which is a problem to analyse the dataset. So in our dataset very few null values in the name and host name columns but on the other side we have a large number of null values in a column last review and reviews number per month.

- In the price column there are few 0 values and that can not be possible because there is no free booking of Airbnb. so this is some kind of data entry problem

Building a correlational graph

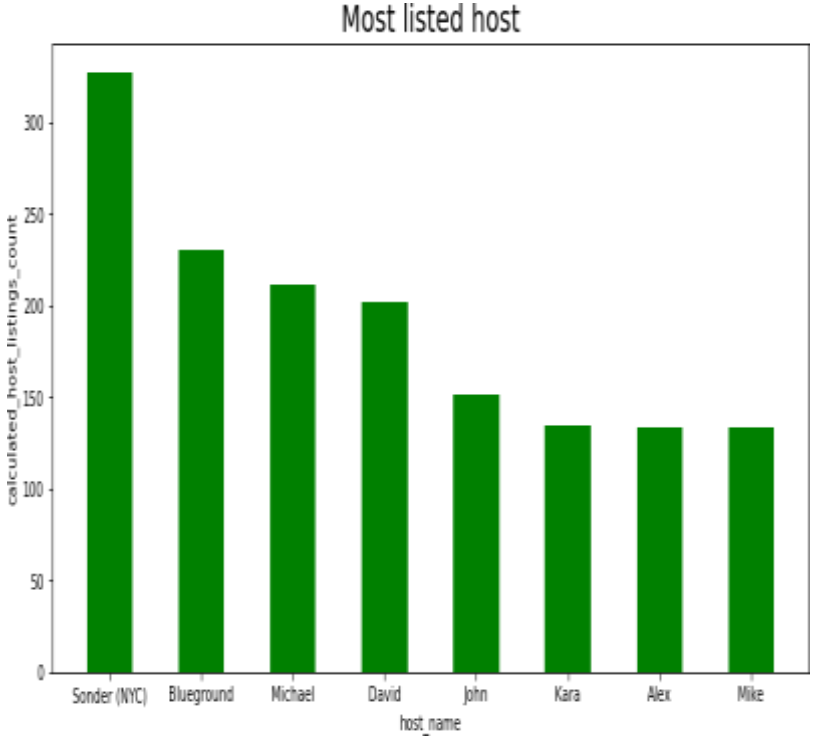
- Here we are using a heatmap for defining the correlation between various values.
- We can find various relations with every aspect of the data frame.
- Some of them are heavily correlated and some of them are mutually apart from each other.

Data cleaning

- As per our requirement to analyse data those data which not required in our data frame that to be clean out and further need to analysis the data of our dataframe.
- So as we discussed earlier in our previous slide that there are a few 0 value in price columns which is need to be correct because there are no free rooms at Airbnb, so the 0 values on the price column replaced with the median to the rest of the non-zero values in the columns. so with this step we’ll eliminate the zero values from price columns

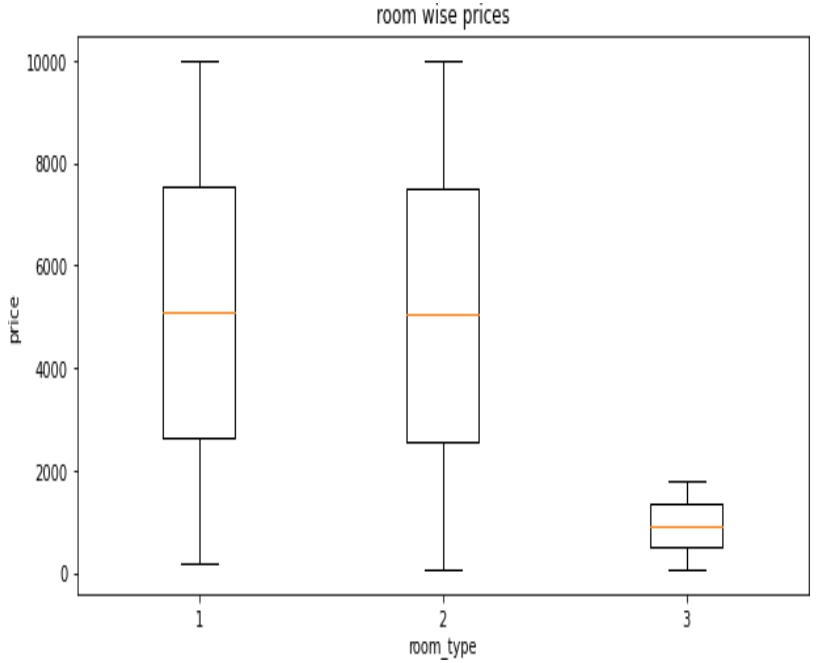
Information about hosts and areas

- From our analysis we found out that the Most listed host of our data frame is Sonder(NYC) then followed by Blueground which has the second highest listed on the Airbnb data frame.
- Interestingly both of them have been hosting in Manhattan
- Manhattan has the most number of hosting that's why whoever hosting in Manhattan will have more listed Airbnb.

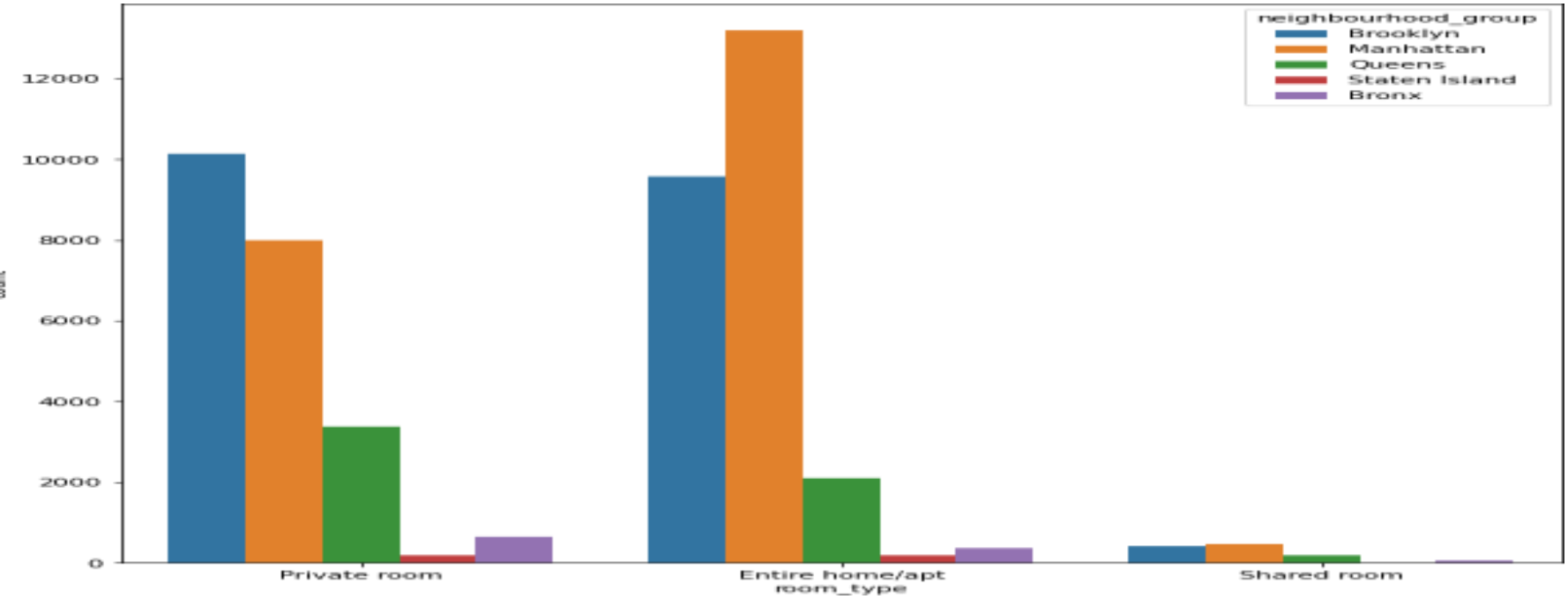


Prediction on the bases of our Data set

- First we work with the prices column.
- Here we found out that there are a few 0 values so we have to replace them with the median by the rest of the values.
- Then we analysis that most of the high prices of a private room and entire apartment cause most of the people wants to stay in private rooms and the entire home

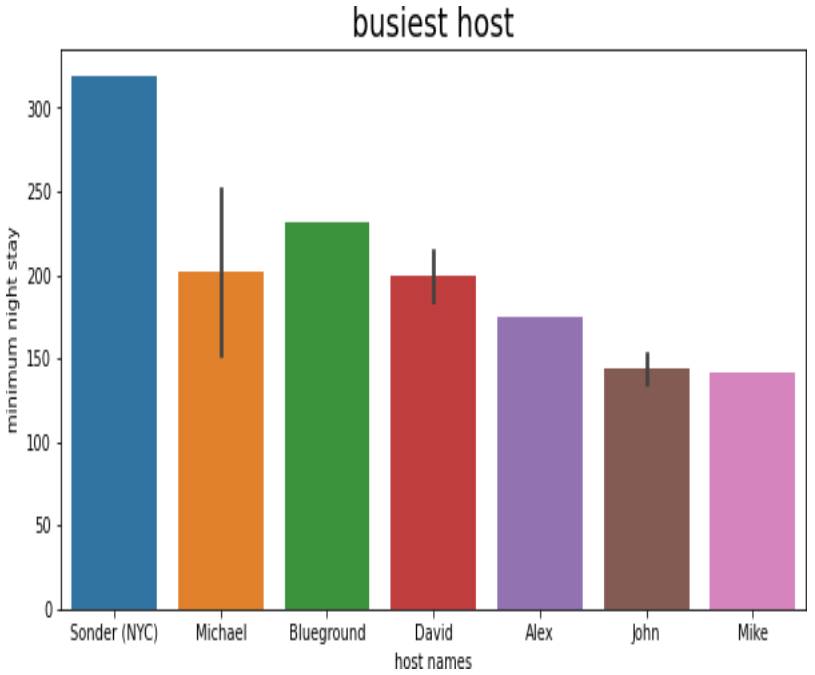


Prediction on the bases of our Data set



Busiest host

- Which host has the highest minimum night stay in their Airbnb he or she will be the busiest host cause there'll be so much work to do in every day.
- According to our analysis we found out that sonder(NYC) has the highest minimum night stay and after that Michael has the second busiest host

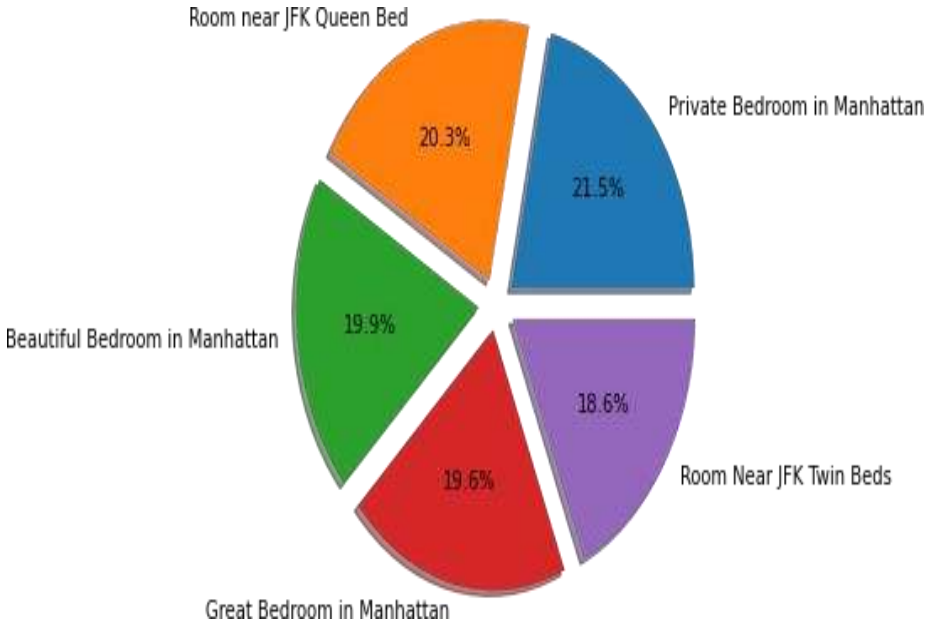


Traffic across the area and reasons

- In this context traffic is defined as how many people are going to that place to stay on Airbnb. We can say that busiest area across the whole data set.
- According to our analysis, we found out that Manhattan is having the most traffic across the whole data set because Manhattan has most 1250 minimum night stay.
- On the second, we have Brooklyn which have 999 minimum night stay.

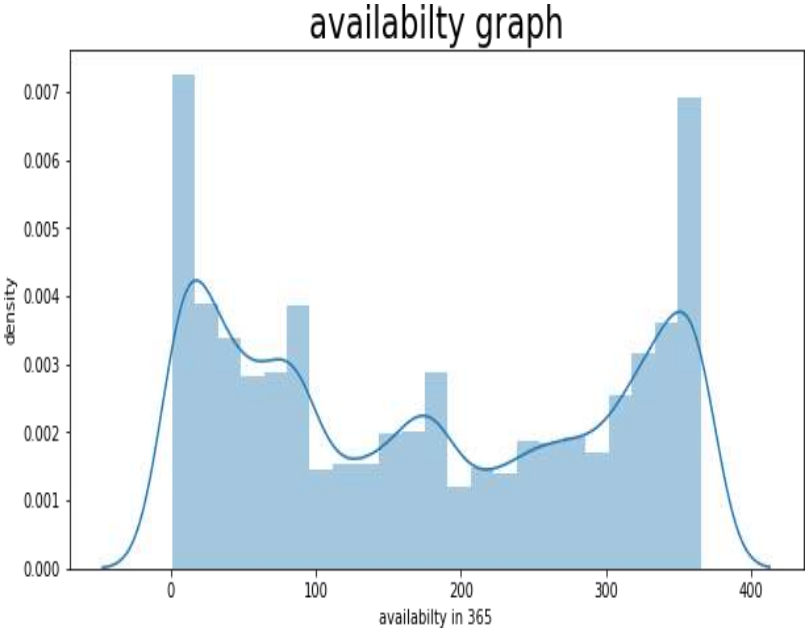
Most preferable properties

- We can easily predict the most likable properties by counting the number of reviews and properties gets.
- So according to our analysis, we found out that a private bedroom in Manhattan has the most reviews so this Airbnb is the most preferable .and other preferable properties are defined in a pie chart.



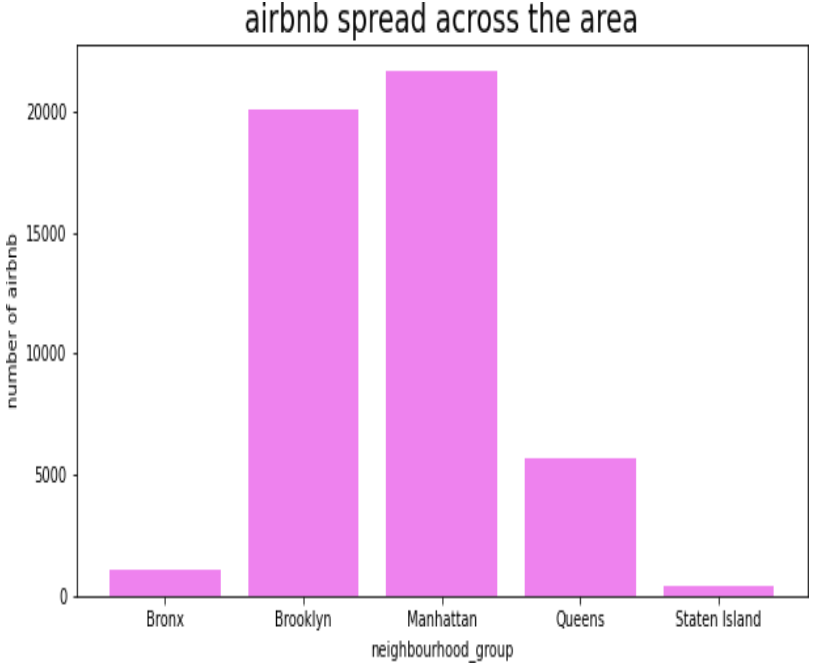
Properties availabilities

- Due to unavailability of various properties throughout the year we found out the availabilities of every Airbnb in this analysis.
- So we set a mark of 200 days to watch how many properties are available for more than 200 days.
- According to our analysis we found out that 42.62 % of properties are only available for more than 200 days.
- So the rest 57.38% of properties are not available for more than 200 days.



Airbnb across whole neighbourhood

- In this analysis we tried to find out how many Airbnb are there in each neighborhood group.
- According to our analysis, we found out that Manhattan has the most number of Airbnb and Brooklyn has the second most number of Airbnb.
- Staten Island and the Bronx have the least amount of Airbnb.



Conclusion

- So according to our analysis some various important point are listed below

1. From our analysis we found out that Sonder(NYC) is the most number of listed hosts among the whole data set.
2. The Most listed neighbourhoods in Manhattan and Brooklyn.
3. Prices of private rooms and entire apartments are on the higher side because these types room are preferred by most people.
4. Busies host are Sonder(NYC),Blueground and Michael.

Conclusion

5.Most people want to visit Manhattan and Brooklyn because of that these neighborhoods have the highest amount of traffic.

6. The Most preferable property of this data set is a private bedroom in Manhattan.

7. Only 42.64 % of properties are available for more than 200 days.

8. Manhattan has the most number of Airbnb.

THANK YOU