Capstone Project - 1
EDA on Airbnb Booking

By- Rajvee Sharma

Background of Airbnb –

Airbnb is a two-sided marketplace that sought to match people that owned real estate properties with people interested in renting short-term lodging. The company was founded in August 2008 and based in San Francisco. Joe Gebbia, Brian Chesky and Nathan Blecharzyk are the founders of Airbnb.

Airbnb allow travelers to rent their spaces from host or people who have vacant spaces in their houses. Airbnb's current worth is around $25.5 billion. Over the few years Airbnb have generated much attention from customers. Furthermore, there have been around 2 million listings worldwide.

Business Goals –

1. Airbnb has strong competitive position in a rapid growth industry and it should continue this strong competitive position to increase its growth rate.

2. To achieve market growth rate with an overall 80% market share from the existing market.

3. The diversity in related business to develop new market and penetrate the existing ones.

Opportunities –

Airbnb has an existing trust network that cooks could tap into to find dining guests for their meals .
Their target audience of people searching for authentic experience serve the user target by the problem statement.
Additionally their existing payment, messaging, home booking and rating platform could be utilized to facilitate in home dining events.

Airbnb’s great community is one of the company’s greatest strengths. The nature of Airbnb experience, the personal touch of hosts and the interaction with them help guests easily engage with Airbnb community

Map of New York

Problem Statement-

1. In this project I'm going to analyse Airbnb's New York City (NYC) data. New York is the most famous city in the world and also the top global destination for visitors.

2. My main aim is to find out key parameters that influence the listing of properties on Airbnb's platform. For this, I'll explore and visualize the dataset from Airbnb in NYC using EDA(Exploratory Data Analysis) techniques.

3. I'll be finding out the distribution of every Airbnb listing based on their important factors like location, price range, room type, listing count etc. It will be absolutely be the crucial factor for company growth.

Competitive Analysis

Data Understanding

There are 49000 rows with 16 columns in our dataset.
List of field areas are given below :

Id, Name , Host_id, Host_name, Neighbourhood, Neighbourhood_group, Latitude, Longitude, Room_type, Price, Minimum_nights, Number_of_reviews, Review_per_month, Calculated_host_listing_count, Availability_365

We will try to answer the following question for Airbnb project

1. What is the average price preferred by customers OR what we can learn from our predictions ?

2. Which is the most demanded/busiest host of Airbnb

3. Where most of the host focussed to own property ?

4. Finding top 20 host with most listing.

5. Find total number of nights spend per room type.

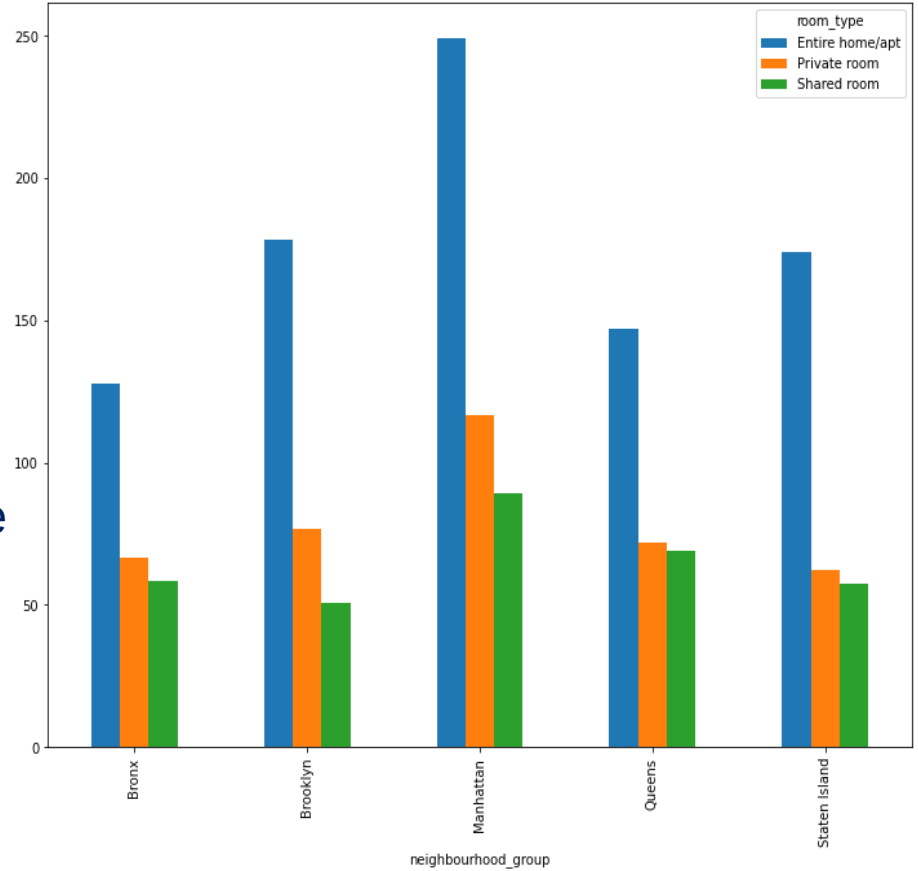
1) What is the average price preferred by our customers according to location ?

> As i can see my observation and concluded that average price of entire home/apt is maximum in Manhattan in all.

> Average price of shared room is lowest in Brooklyn as compared to others.

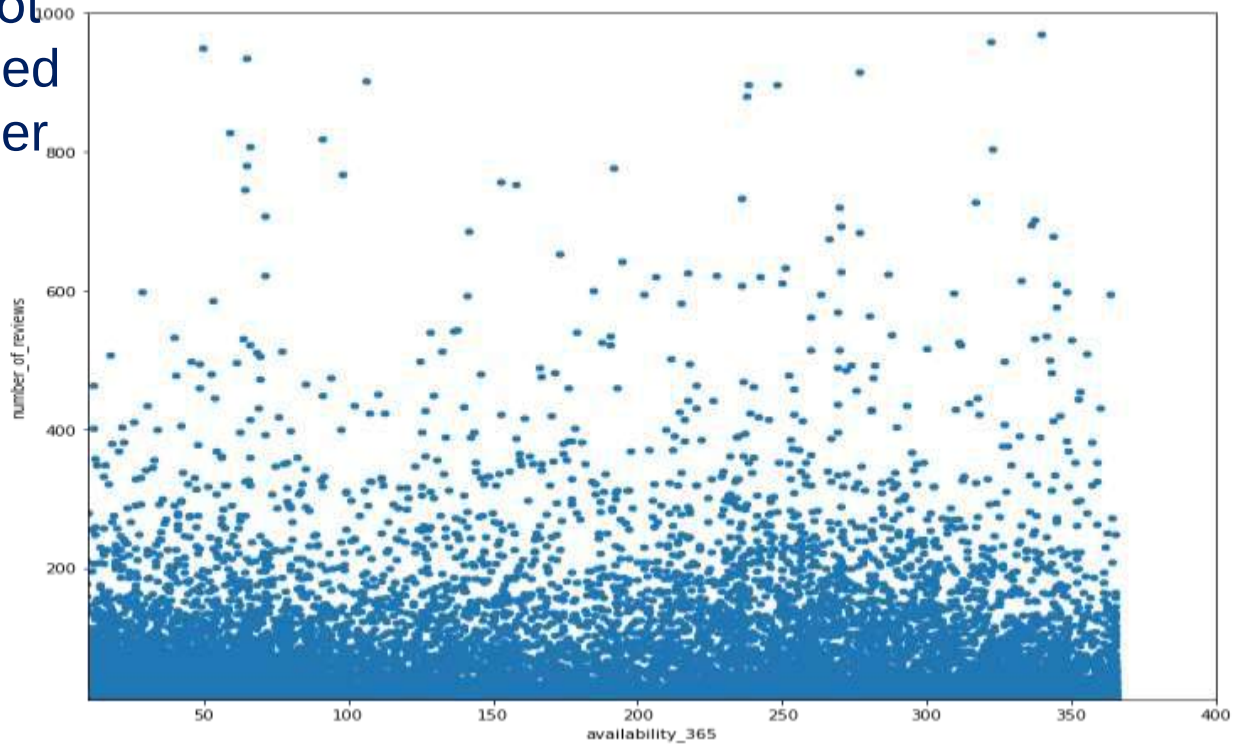
> Queens, Staten Island and Bronx shared the almost same price i.e, $70 for shared and private room.

> Staten Island and Brooklyn also have shared almost same price for entire home/apt.



2) Which is the most demanded/busiest host of Airbnb ?

> **H**ere I'm using scatter plot to show the most demanded host on the basis of number of reviews and availability in a year.



Cont…..

In the table given below I displayed top 5 host of Airbnb in New York city.

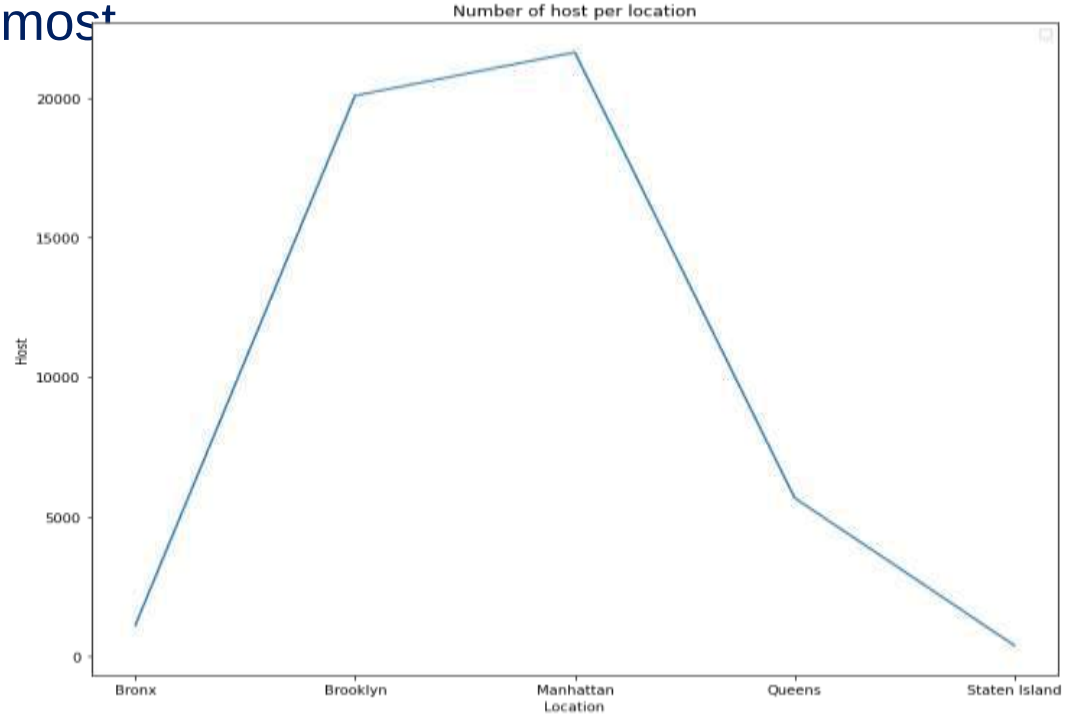
	Host_id	Host_name	Number_of_reviews	Availability_365
17263	22959695	Gurpreet Singh	1157	0.0
28792	99392252	Michael	732	0.0
30113	121391142	Deloris	693	0.0
1864	792159	Wanda	480	0.0
21398	37818581	Sofia	479	0.0

3. Where most of the host focussed to own property ?

> Manhattan and Brooklyn is the most preferred place by the hosts.

> Queens has average number of hosts as compare to others.

> Bronx and Staten Island is low traffic for host .

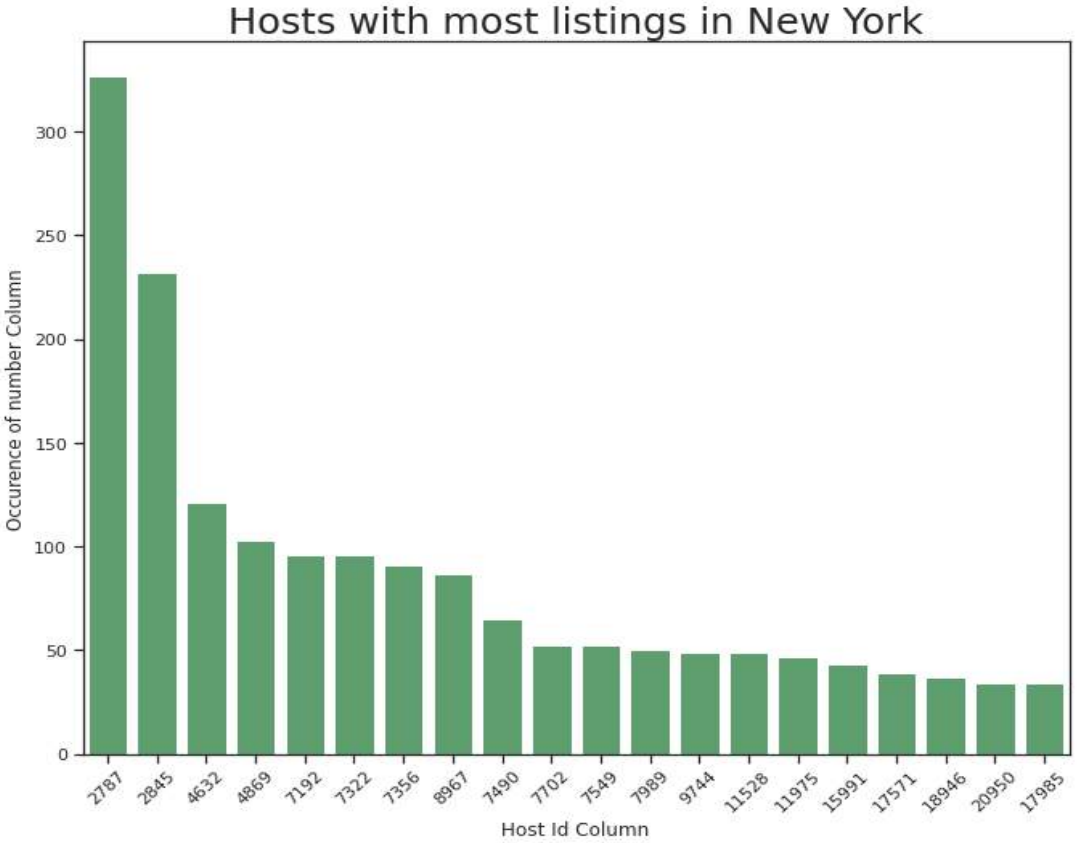


4) Finding top 20 hosts with most listing

> Host id 2787 has highest listing in New York city.

> This host has more than 300 listing.

> From the given graph I can see count of listing by top 20 host is almost 5.0% from whole dataset.

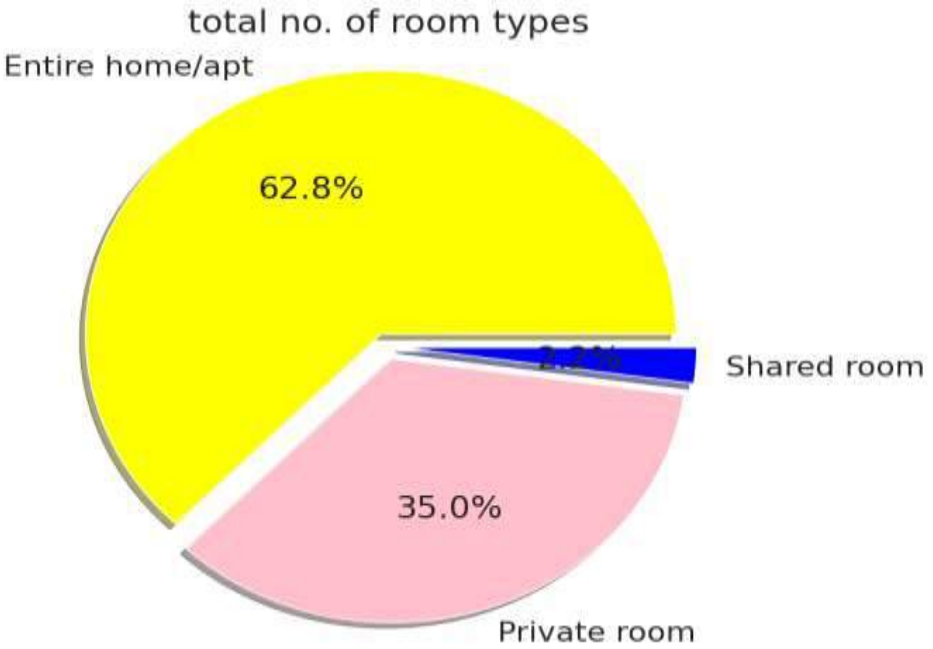


5) Find the total number of nights spend per room type.

> Here I’m analysing the types of room preferred by customers most.

> From the given pie chart conclude that 62.8% customers spend their night in entire home/apt.

> Only 2.2% customer spends night in shared room.



Challenges Faced

> Firstly, go through with whole dataset and understand the meaning of columns what they actually trying to interpret.

> Secondly, understand the whole business model of Airbnb that how they work.

> Removing null values, duplicates and NaN values.

> I have taken 5 insights from Airbnb dataset and designing multiple visualization to summarize the information given in particular dataset.

> After this finally communicate the results and trends which helps in business growth.

Conclusion –

> Manhattan is the most popular place in NYC for hosts to run their business.

> Average price preferred by customers in Entire home/apt is $155 and for shared room it is $60 and for private room it is $73.

> Average number of people spend night in entire home/apt = 62.8% and in private rooms= 35.0% and in shared room= 2.2%

> People stay for longer duration of time in private rooms in Manhattan and Brooklyn.

Cont....

> More customer preferred Manhattan location for night stay then Brooklyn 62.8% customer spend night in Entire home and 2.2% spend night in Shared room.

> Count of listing by top 20 host is 5.0% of whole dataset.

> Maximum customers preferred Manhattan location over Brooklyn for night stay.

THANK YOU

BY – Rajvee Sharma