

Capstone Project – 1

(EDA)

Airbnb Booking Analysis

By

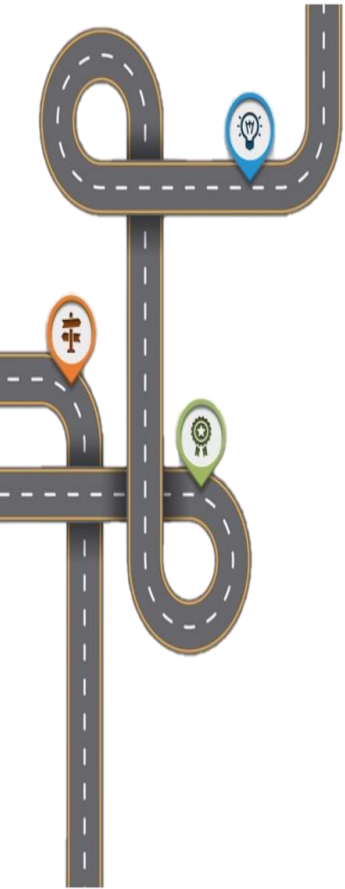
Team- The Scouts

Ishan Barway

Jasmine Sulekha Nag

Nitin M Sakhare

Roshan Alte



Index

1. About Airbnb
2. Problem Statements
3. Data Exploration
4. Data Featuring
5. Exploratory Data Analysis
6. Conclusion
7. Challenges

About Airbnb

- **Airbnb, Inc.** was founded in 2008 in San Francisco, California.
- It's original name, Air_Bed_and_Breakfast.com.
- It mainly based on online marketplace.
- It operates as a broker for hotel booking, home & apartment for rentals, based on requirements.
- Company charges a commission from each booking.



Problem Statements

Questions.

With the help of exploratory data analysis techniques', try to answer the following problem statements,

- What can we learn about different hosts and areas?
- What can we learn from predictions? (ex: locations, prices, reviews, etc.)
- Which hosts are busiest and why?
- Which room type is preferred in most popular neighbourhood?
- Is there any noticeable difference of traffic among different areas and what could be the reason for it?
- Where is the highest room availability in neighborhood group ?



Data Exploration

Data set- Airbnb NYC 2019 contains observation of bookings done in 5 neighbourhood groups in New York City.

- **Rows** – 48,895
- **Columns** – 16
- **Room types** - Entire homes/apt, private rooms, shared rooms
- **Neighbourhood groups** - Manhattan, Queens, Staten Island, Bronx, Brooklyn



Data Exploration

Data set contains null values.

```
airbnb_data.isnull().sum()
```

```
id          0
name        0
host_id     0
host_name   0
neighbourhood_group  0
neighbourhood  0
latitude    0
longitude   0
room_type   0
price       0
minimum_nights  0
number_of_reviews  0
last_review  0
reviews_per_month  0
calculated_host_listings_count  0
availability_365  0
dtype: int64
```

```
#checking null values
airbnb_data.isnull().sum()
```

```
id          0
name        16
host_id     0
host_name   21
neighbourhood_group  0
neighbourhood  0
latitude    0
longitude   0
room_type   0
price       0
minimum_nights  0
number_of_reviews  0
last_review  10052
reviews_per_month  10052
calculated_host_listings_count  0
availability_365  0
dtype: int64
```

Replacing all the null values.



Data Featuring

Given dataset consists the booking information of Airbnb from 2008 till 2019 in a form of csv file.

The features in the dataset can be described as follows:

List Of Fields	
o id	o latitude
o name	o longitude
o host_id	o price
o host_name	o minimum_nights
o neighbourhood_group	o number_of_reviews
o neighbourhood	o last_review
o room_type	o reviews_per_month
o calculated_host_listings_count	o availability_365



Data Featuring

- id - identity number of the property listed.
- name - name of the property listed.
- host_id - customer id number.
- host_name - customer name.
- neighbourhood_group - different areas in NYC.
- neighbourhood - sub areas (locality) in every area.
- latitude - latitude coordinates of the property listed.
- longitude - longitude coordinates of the property listed.
- room_type - property type listed.
- price - rent/day in USD .
- minimum_nights - number of nights customer rented the property.
- number_of_reviews - no. of customer review.
- last_review - date when the property was last reviewed.
- reviews_per_month - count of reviews per month.
- calculated_host_listings_count - no. of time customer visited.
- availability_365 - number of days the property is available among 365 days.



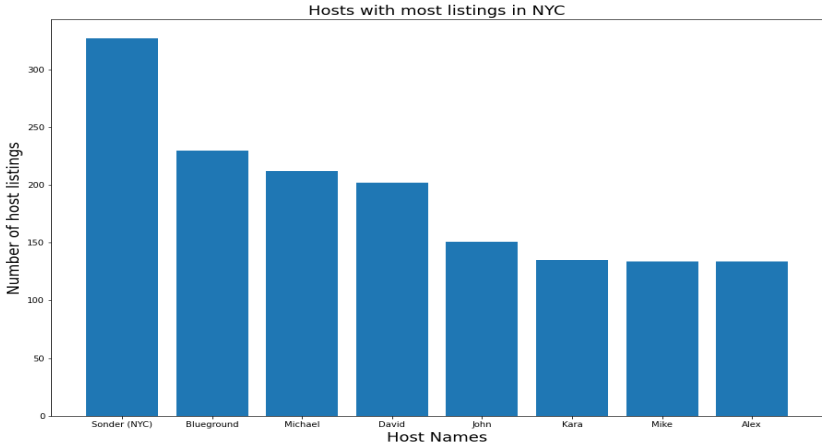
Data Analysis



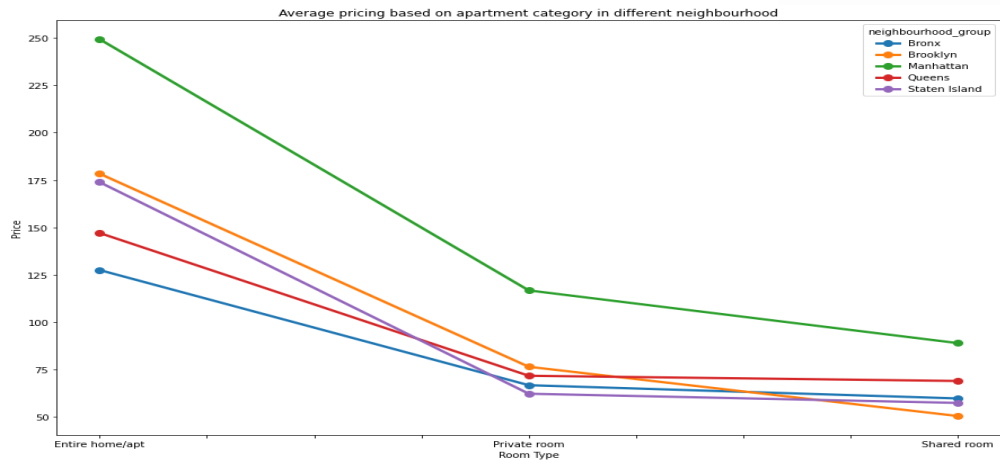
Host with most listed count in the NYC.

- **Sonder** has the highest count i.e. 327 in the Manhattan area .
- **Manhattan & Brooklyn** are the highest host count area.

host_name	neighbourhood_group	calculated_host_listings_count
Sonder (NYC)	Manhattan	327
Blueground	Manhattan	230
Michael	Manhattan	212
David	Manhattan	202
Michael	Brooklyn	159
John	Manhattan	151
David	Brooklyn	142
Kara	Manhattan	135
Mike	Manhattan	134
Alex	Manhattan	134



Data Analysis

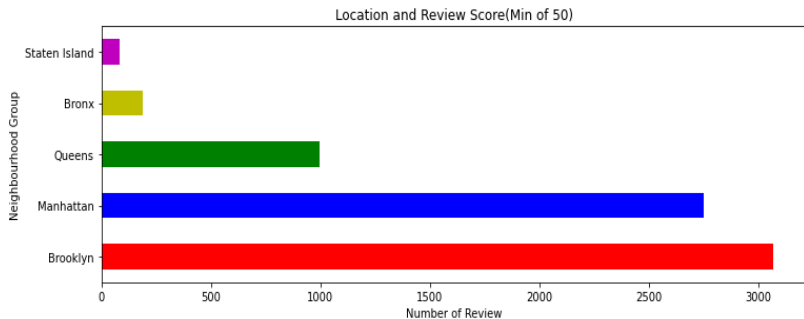


Price vs Room type on different areas.

- **Manhattan** has the highest room price among all the room type.
- **Bronx** has lowest Entire home/apt price as compared to other areas.
- **Staten Island** has lowest private room price as compared to other areas.
- **Brooklyn** has lowest shared room price as compared to other areas.

neighbourhood_group	Bronx	Brooklyn	Manhattan	Queens	Staten Island
room_type					
Entire home/apt	127.506596	178.327545	249.239109	147.050573	173.846591
Private room	66.788344	76.500099	116.776622	71.762456	62.292553
Shared room	59.800000	50.527845	88.977083	69.020202	57.444444

Data Analysis



neighbourhood_group	
Brooklyn	3065
Manhattan	2751
Queens	997
Bronx	187
Staten Island	81

Number of reviews vs Neighbourhood

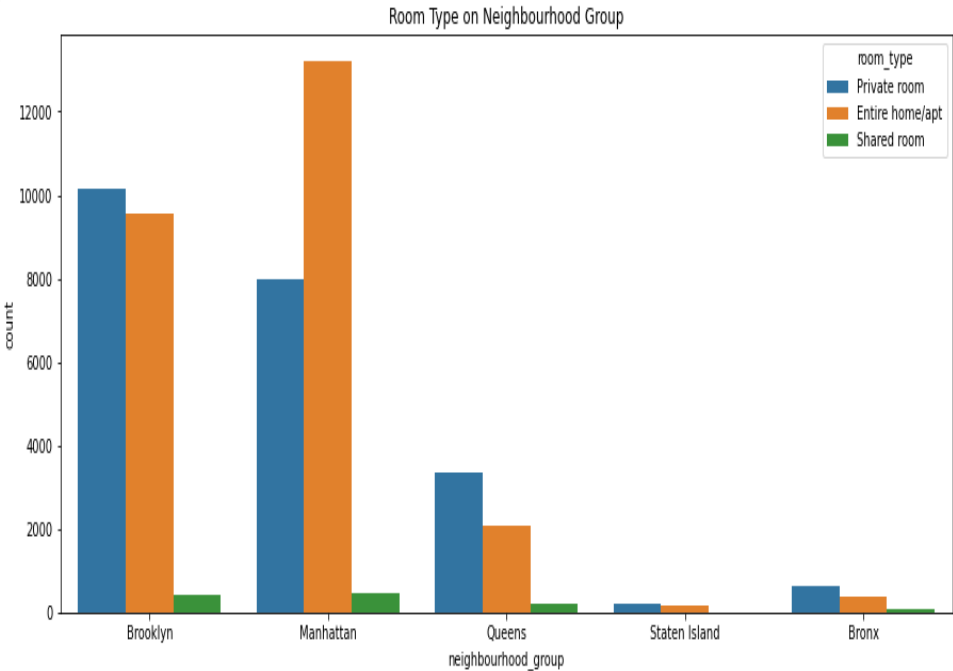
- **Brooklyn** has the highest reviews among all the neighbourhood, followed by **Manhattan**.
- **Staten Island** has lowest reviews among all the neighbourhood.



Data Analysis

Neighbourhood vs Room type

- **Manhattan** has the highest Entire home/apt.
- **Brooklyn** has the highest Private rooms.
- **Staten Island** has lowest room count in all categories of room.
- **Shared rooms** are very few in all the neighbourhoods



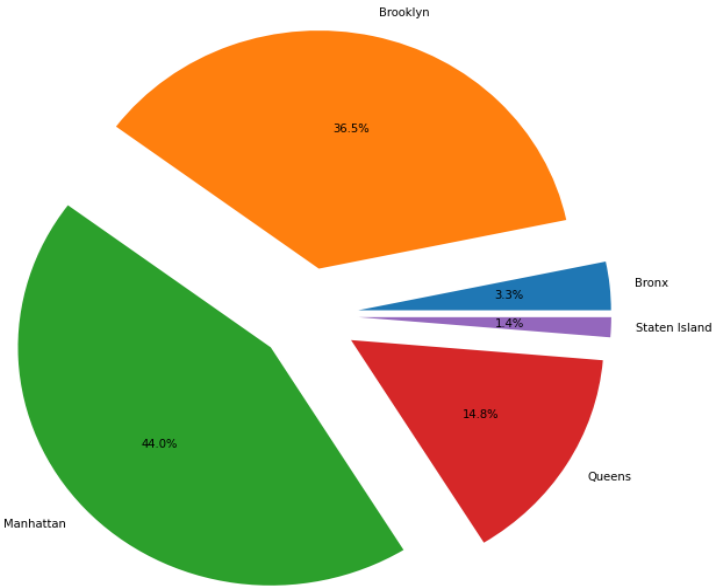
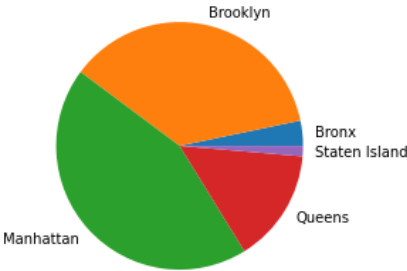
Data Analysis



Checking room availability based on neighbourhood

- **Manhattan** has the highest room availability which is **44%**, followed by **Brooklyn** with **36.5 %**.
- **Staten Island** has lowest room availability with **1.4%**.

neighbourhood_group	availability_365
Manhattan	21661
Brooklyn	20104
Queens	5666
Bronx	1091
Staten Island	373



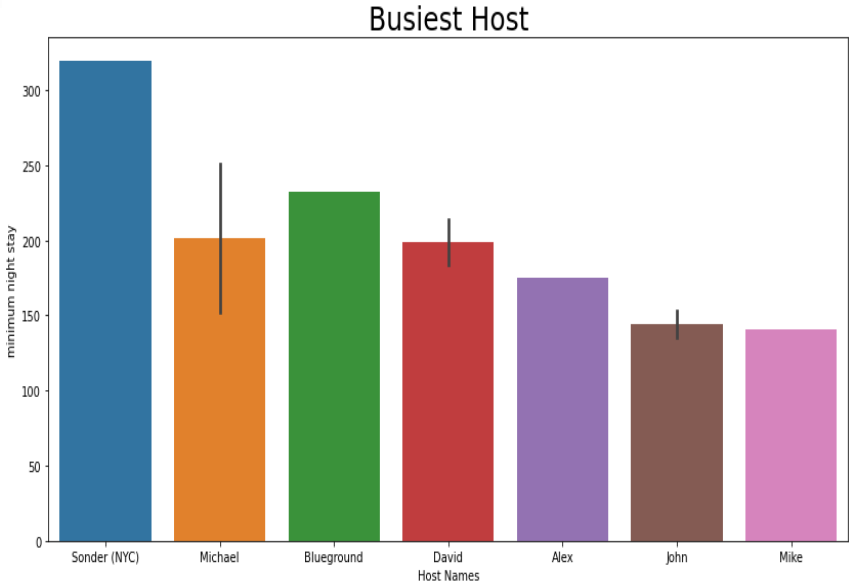
Data Analysis



Busiest Host

- **Sonder** prefer **entire room/apt** & he is the most busiest host followed by **Michael & Blueground**.
- **Michael, David & John** are all on the top 10 but they preference is **Shared** and **Entire room/apt** both.

host_name	room_type	minimum_nights
Sonder (NYC)	Entire home/apt	319
Michael	Entire home/apt	251
Blueground	Entire home/apt	232
David	Entire home/apt	214
David	Private room	184
Alex	Entire home/apt	175
John	Private room	153
Michael	Private room	152
Mike	Entire home/apt	141
John	Entire home/apt	135



Data Analysis

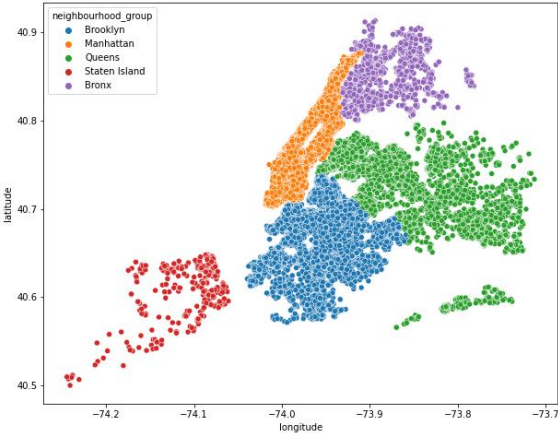


Location Mapping of Hotels.

- Total rooms

Manhattan	21661
Brooklyn	20104
Queens	5666
Bronx	1091
Staten Island	373

- **Manhattan** is very densely populated .



Conclusion

- Manhattan has the highest number of listings, followed by Brooklyn and Queens. Staten Island has least number of listings.
- Sonder(NYC) is the most successful host, who mostly chooses Manhattan to stay.
- Brooklyn and Manhattan are most liked neighbourhood groups by people.
- People of all the neighbourhood prefer Entire room/apt more then private room or shared one.
- Manhattan is one of the most expensive neighbourhood.
- Entire home/apt is the most expensive and yet get the most traffic.
- Although Manhattan has the highest no of hotel present still Brooklyn has the highest reviews among all the neighbourhood.
- Staten Island has lowest reviews among all the neighbourhood.



Conclusion

- Manhattan has a lead on Entire home/apt, while Brooklyn has the highest Private rooms. Shared rooms has the lowest customer count.
- Staten Island has lowest room count in all categories of room.
- Manhattan has the highest room availability which is 44%, followed by Brooklyn with 36.5 %.
- Manhattan is very densely populated.



Challenges

- Huge chunk of data was to be handled keeping in mind not to miss anything which is even of little relevance.
- Computation time.



Thank You

