Capstone Project

Airbnb Bookings Analysis

Performed by: Shubham Naik

Presentation Overview:

- Introduction
- Data Information
- Data Processing
- Analysis & Visualizations
- Conclusion

Introduction:

Dataset used – Airbnb NYC (2011-19)

- Airbnb, Inc. 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. –Wikipedia
- Provided dataset is of Airbnb's host listing across New York state.
- The Dataset have ~50k rows of various Airbnb hosts with 16 columns of various attributes like host name, boroughs, neighbors, price etc.
- Performing EDA on this dataset can help the company to know about the various host listings in different boroughs of NYC and their price and availability with rating and reviews.

Data Information:

- **Attributes**

1. neighbourhood_group: Consists data which tells us about locations in New-York state.

2. neighbourhood: This attributes tells us about neighborhood areas from the major cities of New-York state.

3. room_type: Consists data which tells us about all the available room type among various cities of New-York state.

4. price: Contains prices for each Airbnb among various city locations.

Data Information:

- **Attributes**

5. minimum_nights: Tells us the number of minimum nights people were stayed in that Airbnb.

6. availability_365: Tells us about the number of days when listing is available for booking.

7. number_of_reviews: Contains data which shows the number of reviews given by their customer to that particular Airbnb.

8. latitude/longitude: Shows the geographical location of each host spread over the state of New-York.

Data Information:

- **Attributes**

9. host_id/host_name: Contains list of names and ids of hosts, by whom the house was rented from.

10. id/name: Contains list of names and ids of customers, to whom the house was rented.

11. review_per_month: Contains number of reviews given to specific room in each Month.

12. calculated_host_listings_count: It contains the amount of listing per host.

Data Processing:

File name : Airbnb NYC 2019.csv — original unmodified file

#	Column name	Non-Null count	Dtype
0	id	48895 non-null	int64
1	name	48879 non-null	object
2	host_id	48895 non-null	int64
3	host_name	48874 non-null	object
4	neighbourhood_group	48895 non-null	object
5	neighbourhood	48895 non-null	object
6	latitude	48895 non-null	float64
7	longitude	48895 non-null	float64
8	room_type	48895 non-null	object
9	price	48895 non-null	int64
10	minimum_nights	48895 non-null	int64
11	number_of_reviews	48895 non-null	int64
12	last_review	38843 non-null	object
13	reviews_per_month	38843 non-null	float64
14	calculated_host_listings_count	48895 non-null	int64
15	availability_365	48895 non-null	int64

- **48895 entries, 16 columns**

- Data processing is an important aspect of EDA. In the table you can see that the number of observation for each attributes are not same, which can affect our data while analyzing.

Data Processing:

- So there are total 16 columns and 48895 rows.

- 4 columns have missing values :

#	Column	Null value Count
1	name	16 null
3	host_name	21 null
12	last_review	10035 null
13	review_per_month	10035 null

- Treatment :
  - Since name and host_name have very few null values we can drop their entire row of total 37 observations. It wont affect the data significantly.
  - last_review : (sample : 2018-10-19) is a date column. So the missing date values can be filled by last date of the dataset year i.e. 2019-12-31.
  - Review_per_month : (sample : 0.21). Similarly the missing values can be filled by 0.0 .

Data Processing:

Modified dataset :

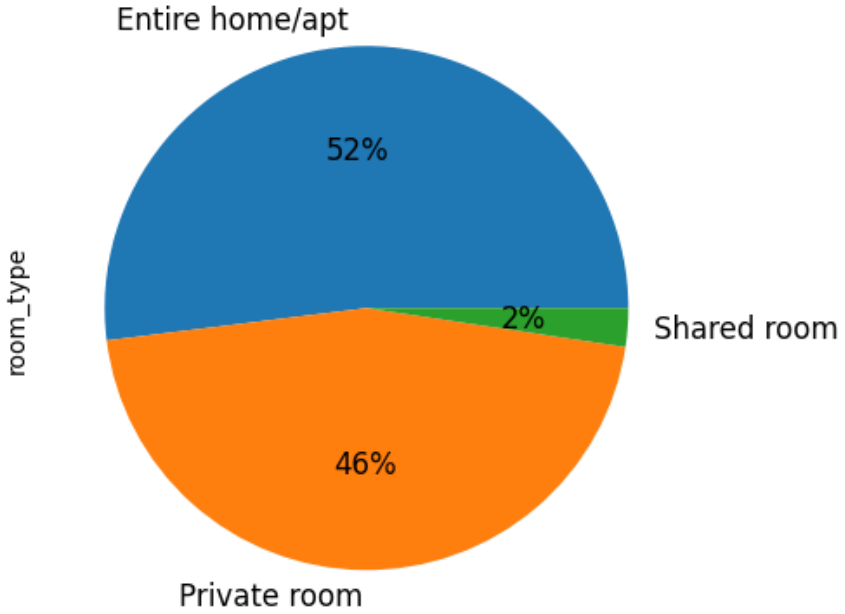
#	Column	Non-Null count	Dtype
0	id	48858 non-null	int64
1	name	48858 non-null	object
2	host_id	48858 non-null	int64
3	host_name	48858 non-null	object
4	neighbourhood_group	48858 non-null	object
5	neighbourhood	48858 non-null	object
6	latitude	48858 non-null	float64
7	longitude	48858 non-null	float64
8	room_type	48858 non-null	object
9	price	48858 non-null	int64
10	minimum_nights	48858 non-null	int64
11	number_of_reviews	48858 non-null	int64
12	last_review	48858 non-null	object
13	reviews_per_month	48858 non-null	float64
14	calculated_host_listings_count	48858 non-null	int64
15	availability_365	48858 non-null	int64

- Dataset is now optimized and we can proceed with analysis and visualization.

Analysis & Visualizations:

- **Analysis – Room types**

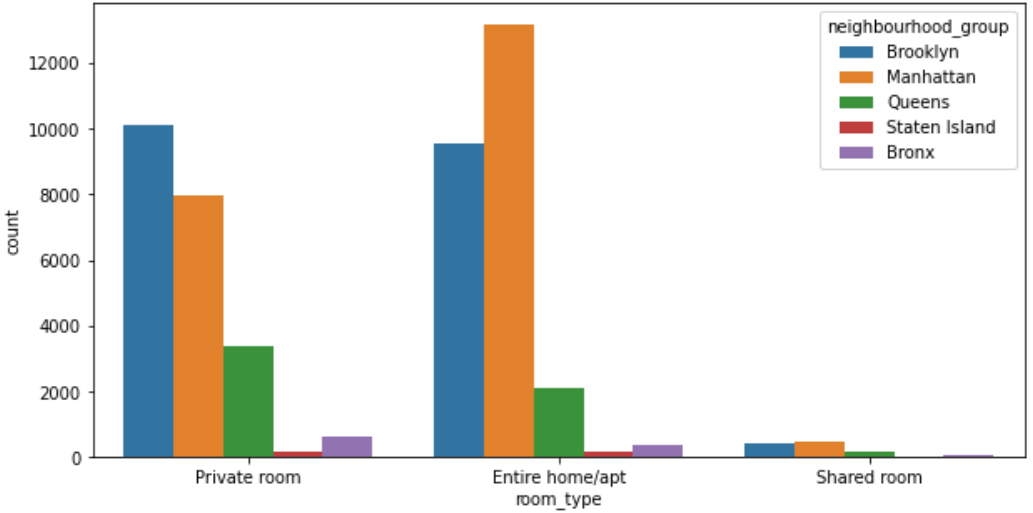
By analyzing the room types from cities; Manhattan, Brooklyn, Queens, Bronx, Staten Island, we can conclude that the majority of room types that were available for rents were 'Entire home/apartment' with occupancy of 52%, where as 'Private rooms' occupies 46% of all room types, and only 2% of total room types is occupied by 'Shared room'.



Analysis & Visualizations:

- **Analysis – Neighborhood Group & Room type**

From brief analysis we came to know that Brooklyn has the highest number of ‘private room’ hosts, whereas Manhattan has the highest number of ‘Entire room/apartment’ and ‘shared room’ hosts.



The Staten island and Bronx has the least number of rooms hosts.

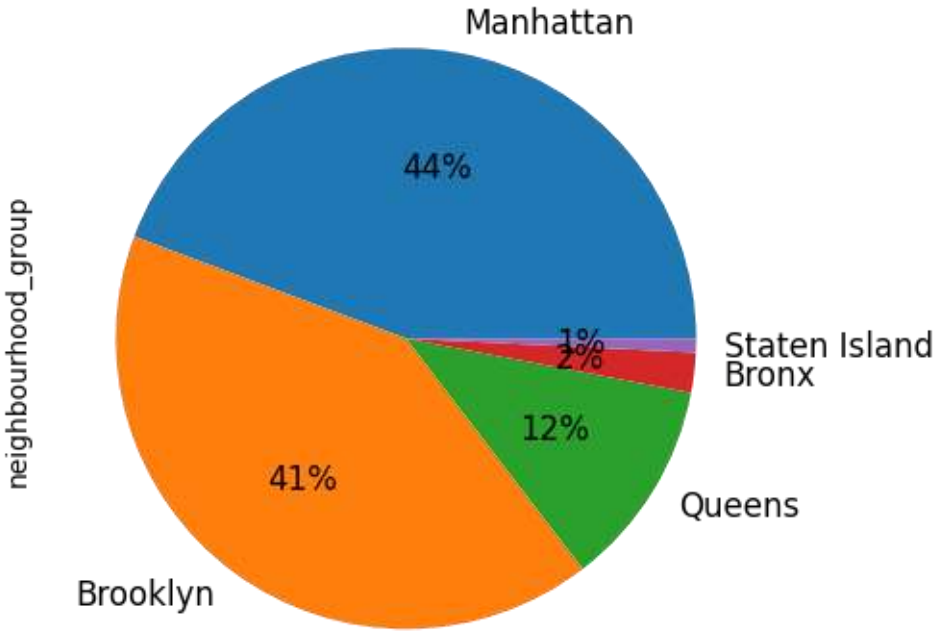
Analysis & Visualizations:

- **Analysis – Busiest Neighborhood Group**

This visualization helps us to understand which neighbourhood group was most busiest and widely chosen by the hosts.

Manhattan and Brooklyn seems to be the only one which were most widely occupied by hosts.

Staten Island and Bronx were on the backfoot in terms of popularity.

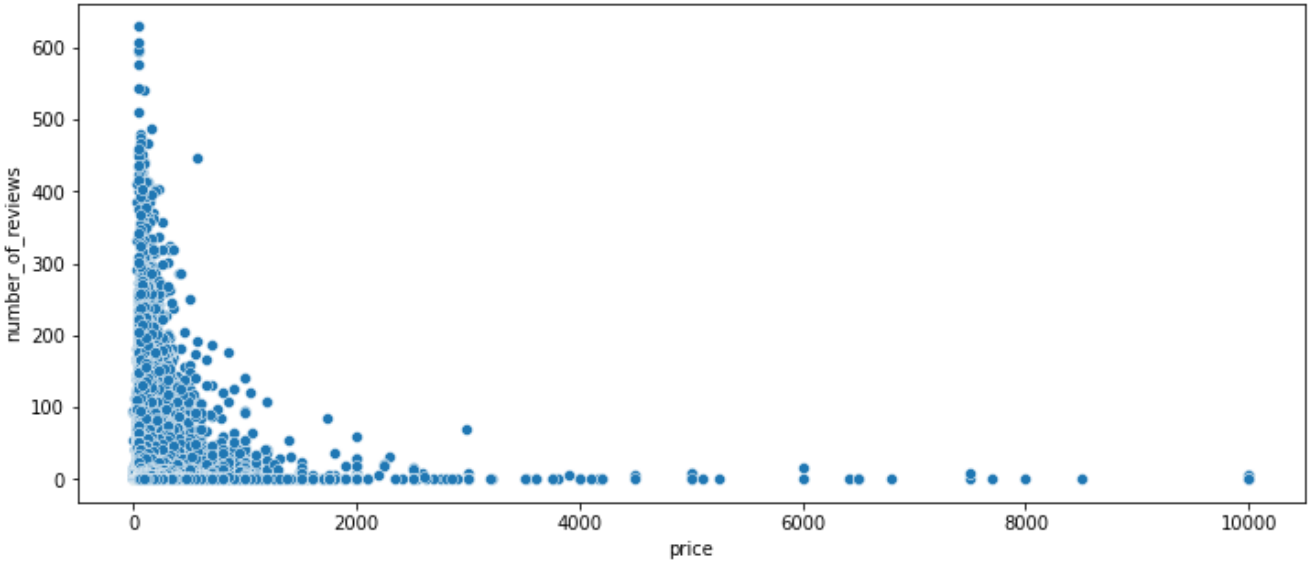


Analysis & Visualizations:

- **Analysis – Price vs Number of Reviews**

This analysis will give us brief insights on pricing of rooms based on reviews.

By analyzing the scatter plot given below we can say that expensive listings have fewer number of reviews whereas less expensive listings have more reviews.

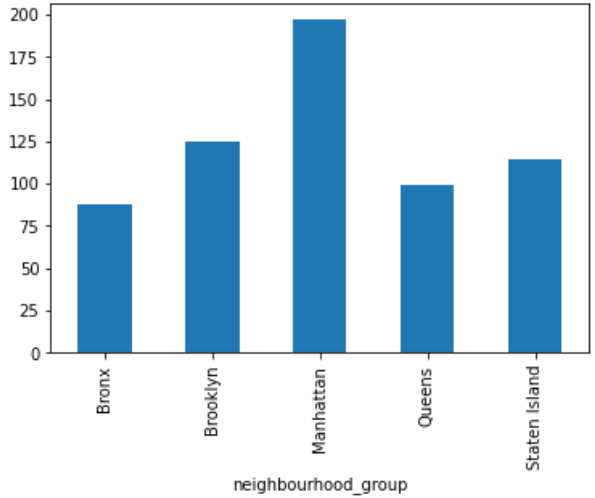


Analysis & Visualizations:

- ## Analysis – Average Prices

By analyzing prices for each room type, the average price for any room type in Manhattan is around $196, in Brooklyn it is $124, for Staten Island it is $114, in Queens it is around $100 and in Bronx it is $87.

Neighbourhood group	Entire home/apartment	Private room	Shared room
Manhattan	$249	$116	$89
Brooklyn	$178	$76	$50
Queens	$147	$71	$69
Bronx	$127	$66	$58
Staten Island	$173	$62	$57

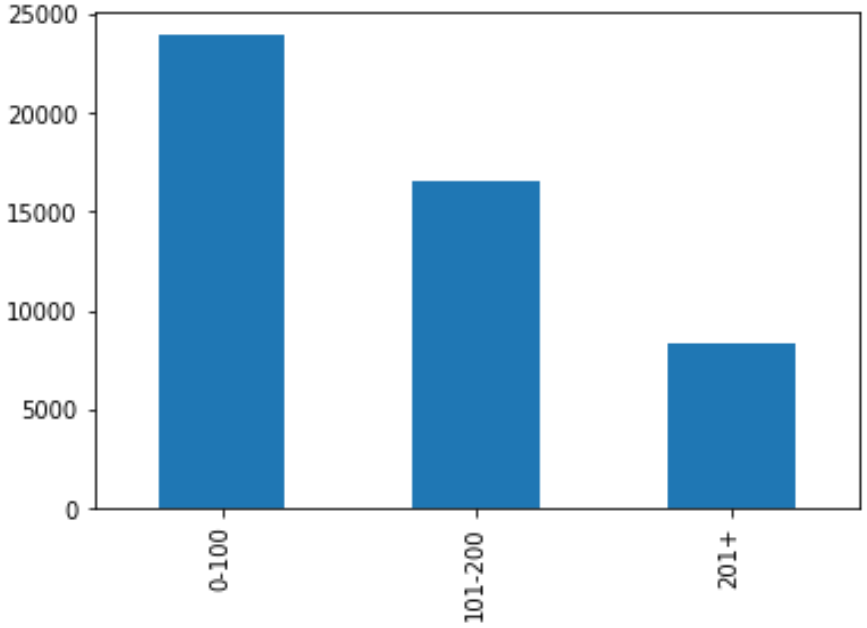


Analysis & Visualizations:

- **Analysis – Price Range of Booked Airbnb**

This analysis helps us to understand the number of room with price range.

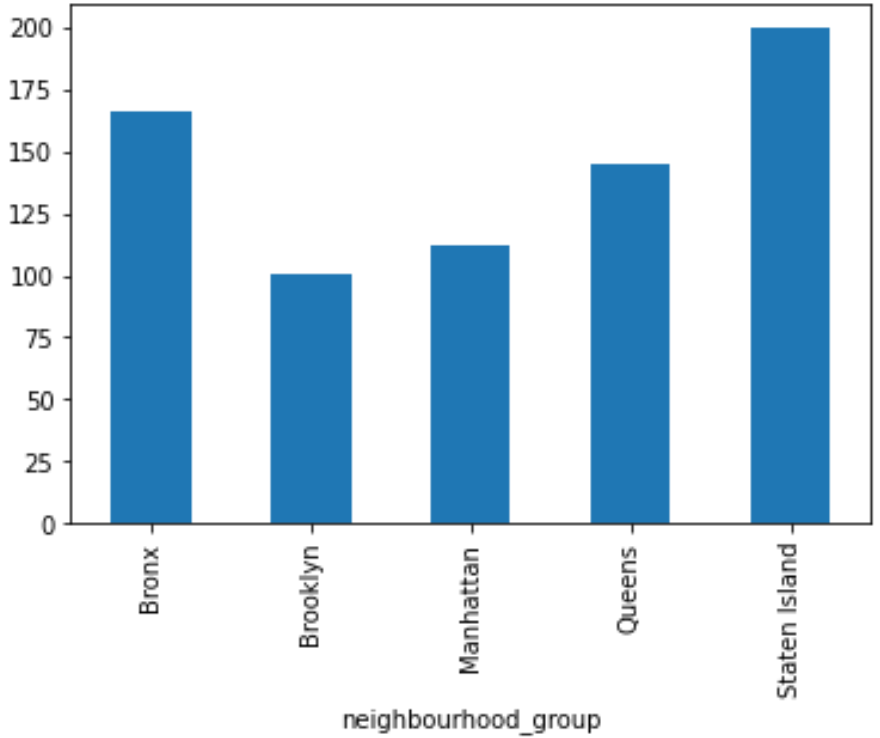
The number of rooms with price less than or equal to $100 are above 23,000, for rooms with price greater than $100 but less than $200 are over 16,000 and the rooms with price over $200 are just under 10,000, 8378 to be more specific.



Analysis & Visualizations:

- **Analysis – Availability out of 365 days**

This analysis helps us to understand the availability of each listings in each city.
According to the statistics, listings in Manhattan & Brooklyn were available for just over 100 days, whereas for Queens & Bronx listings were available for over 150 to 200 days, and for Staten Island the listings availability were over 200 days.
But from this statistics we can state that people were choosing 'Shared room' in Staten Island & Bronx which explains its low days availability, for Manhattan and Brooklyn 'Private room' were the optimal choice and in Queens 'Entire home/apt' was widely chosen.



Analysis & Visualizations:

- **Analysis – Busiest Month for Hosts**

This analysis states that the most busiest time of year for any host was between May and August.

This is the exact time when Spring and Summers begins in USA, which explains the exponential rise of demand for rental rooms & apartment in NYC, as these period are suitable for tourism.

Conclusion:

- Expensive listings has fewer reviews and less expensive listings has more reviews.
- By analyzing host_id it shows that there are more than 37,000 hosts available across NYC.
- Majority of bookings were done for Entire home/apt (25393), followed by private room (22306).
- Manhattan was the busiest neighbourhood group with 21643 entries, followed by Brooklyn with 20089 entries.
- Staten Island has the highest average availability of 199 days were Manhattan has the least average availability of 112 days.
- Manhattan has the highest average minimum nights of 8 days.
- Average price of neighbourhood group –
    - Bronx was the cheapest with average price of $87, Manhattan was the most expensive with average price of $196.
    - Average price of all neighbourhood group is $152.
- 23909 listings choose the Airbnb whose price were less than $100, only 8378 listings booked Airbnb for more than $200.