Insights from Airbnb Analysis from Pre-COVID Period

-BY PRAKHAR & MILAN

Agenda

- Objectives
- Background
- Key Finding
- Recommendation
- Appendix:
  - Data Source
  - Data Methodology
  - Data Model Assumption

Objective

- Improve strategies to revive the business the business in the post-COVID period.
- Understand the key-insights from AirBnB NYC business in pre-COVID period.
- Estimate customer preferences for the post-COVID period travel.

Background

- The COVID-19 pandemic affected Airbnb business due to travel restriction.
- The revenue took the largest hit in NYC in the Q2 in 2020.
- Now the travel restriction are lifted, the business should be operated to recover the loss.

Objective of the analysis of the Airbnb NYC data

- Customer preference and rating of different hosts of Airbnb in were analyzed.
- The data was analyzed to derive key insights from the Pre-COVID period.
- The insights will be used to take decisions for the NYC Airbnb business for travel in post-COVID period.

Data Preparation And Cleaning

- The data was cleaned for any missing values and outliers using Python.
- The data was analysed using simple statistical analysis of the data.
- The visualization were derived using Tableau tool to understand key interference from the analysis.

2.2) Checking for 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
```

Here, we will be treating certain columns. Let's discuss which columns, how and why.

- `name` & `host_name` : Since most of our analysis would be based on the major demographics like location, price etc., we do not need actual names of the hosts and the properties. However, instead of removing the column, we will impute the null values in `name` column with **"$"** and in `host_name` with `#`.
- `reviews_per_month` : Here, since null values depict that there we no reviews for the property, we can safely replace null values with "0.0"
- `last_review` : This column has more than 10k null values and is not important for our analysis. Hence we will drop the column.

```
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
reviews_per_month                   0
calculated_host_listings_count      0
availability_365                    0
dtype: int64
```

Now are data is ready for further analysis!

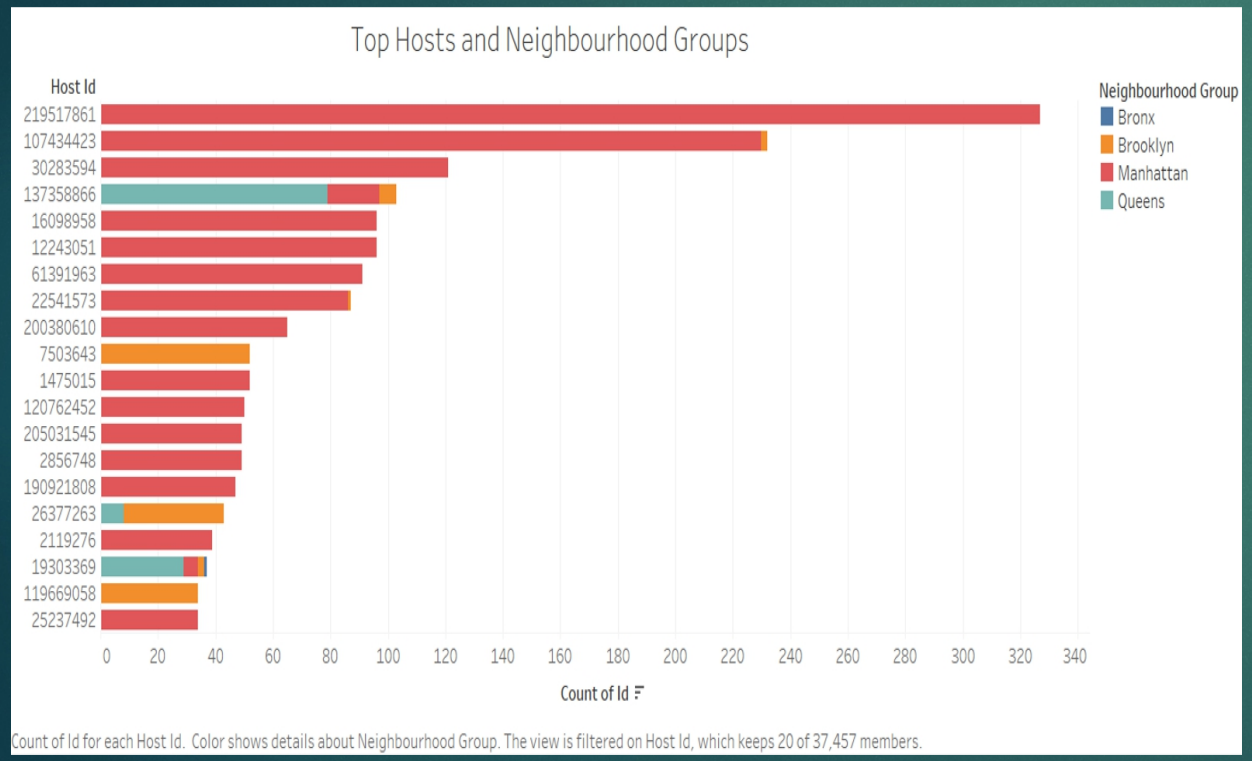
Hosts

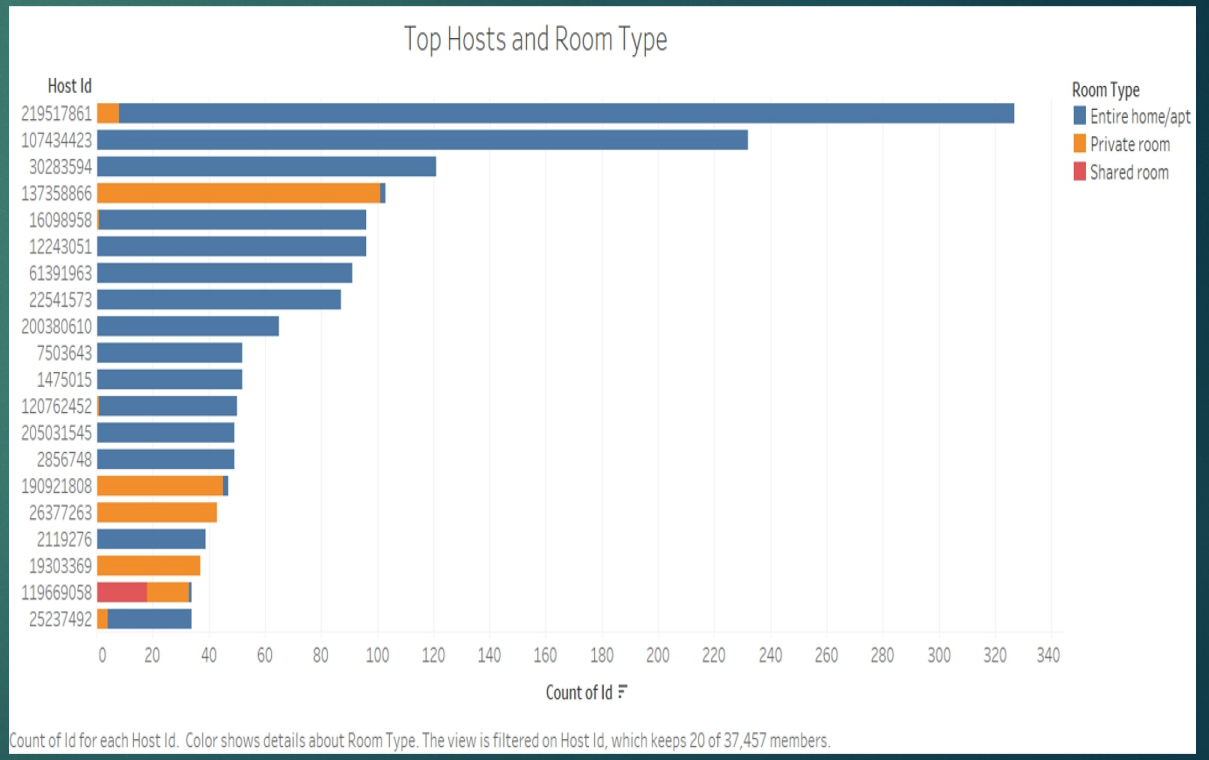
Top Hosts and Neighbourhood Groups:

Host who have maximum properties in Manhattan, Brooklyn and Queens. As evident, these areas generally see a high demand, hence more hosts from these areas should be acquired.

Top Hosts and Room Type:

Hosts with Entire home/apartment are the most in number followed by hosts with private rooms. These are evidently in the maximum demand, hence more hosts with entire apartments or private homes should be acquired.

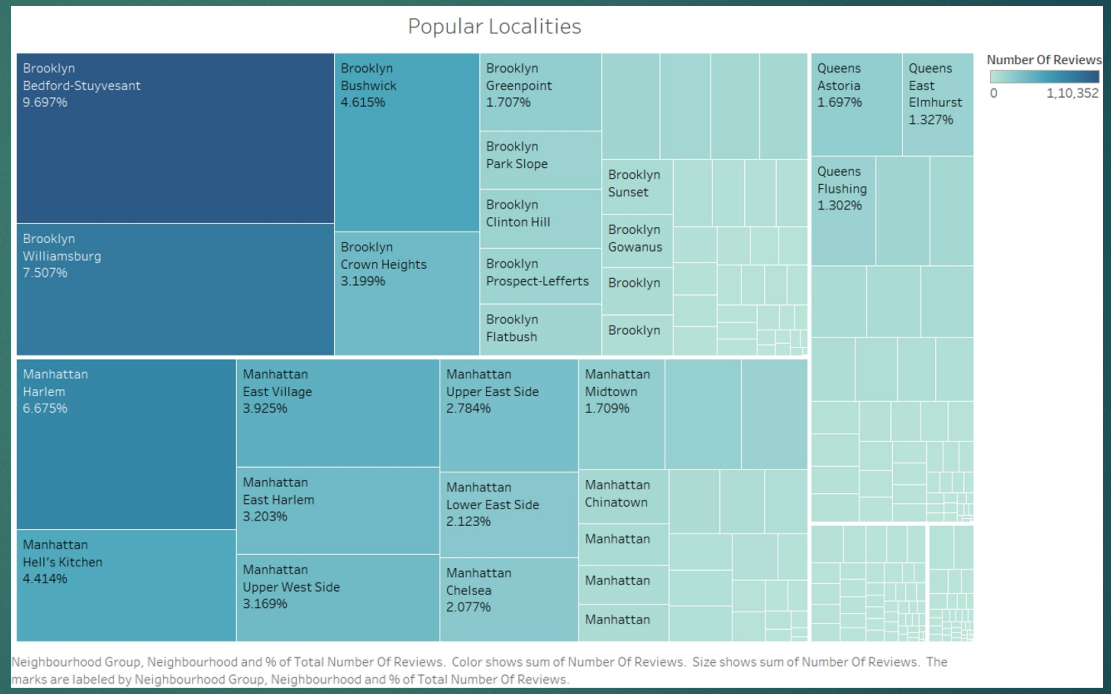




Categorization of customers based on their preferences

Neighbourhoods:

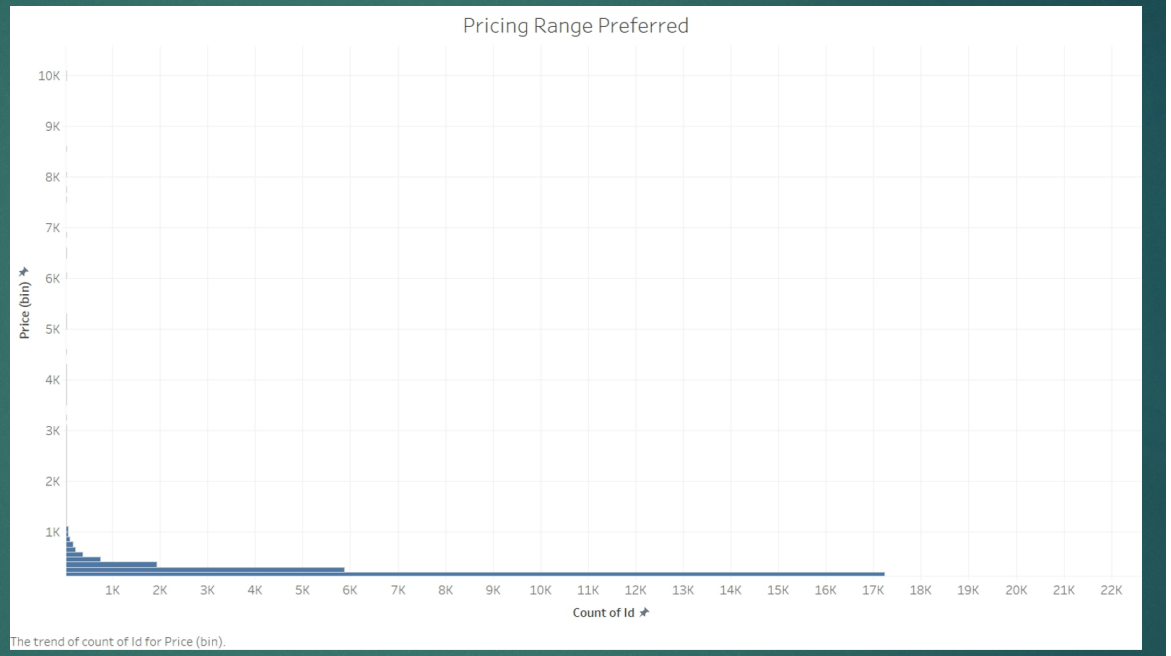
Customers seem to have preferred neighbourhoods mainly in the Manhattan (44% of total listings) and Brooklyn (41% of total listings) areas. Since, getting traction for the business is a key goal here, these highly preferred areas should be targeted.

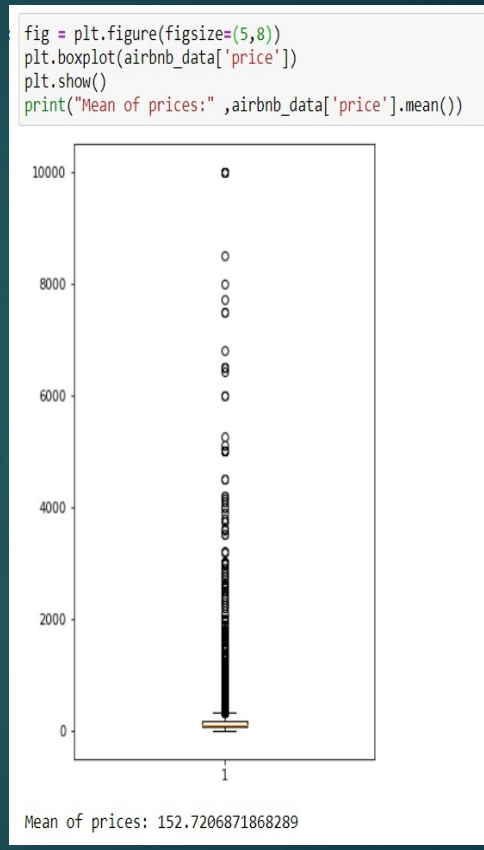


Categorization of customers based on their preferences

Price Preference:

As we can see from the boxplot, the mean of the prices is around $150 which is evident from the histogram which shows maximum properties in the range of $0 to $200. However, there are premium properties as well, with price going up till $10,000, but generally not more than 5 for each premium price range.

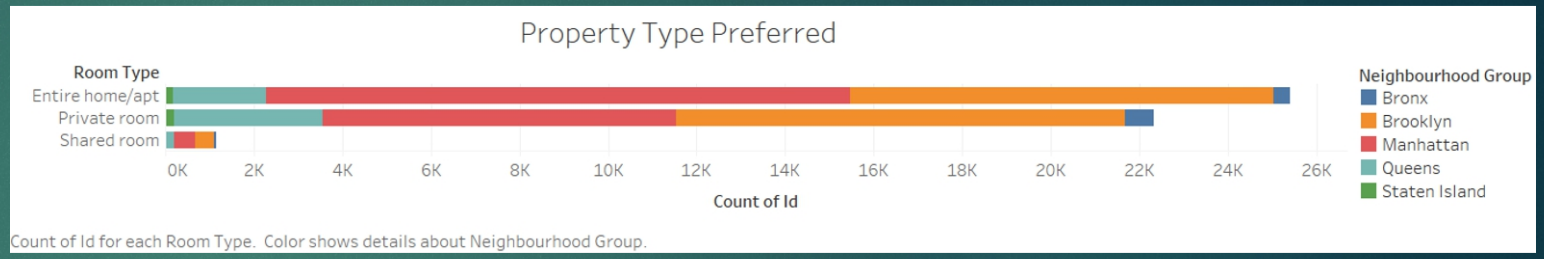




Categorization of customers based on their preferences

Properties w.r.t Customer Preferences:

While entire homes/apartments are the most preferred in general, among them, properties in the Manhattan area have been the most in demand. This is followed by private rooms where the Brooklyn area is most in demand.

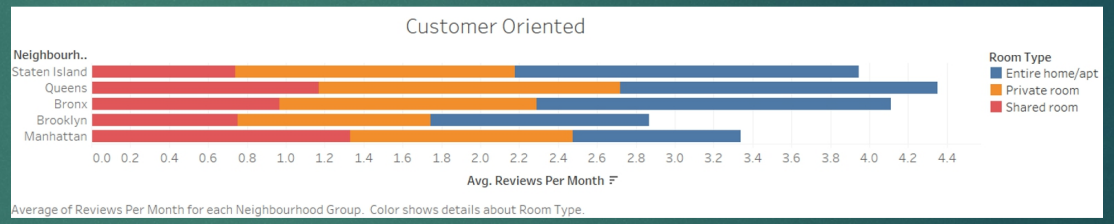


Categorization of customers based on their preferences

Adjustment in existing properties to make them customer oriented:

Although Manhattan and Brooklyn areas are in high demand, the average customer reviews per month are least for these. Low average monthly reviews show that these properties aren't functioning at their full potential. Since these are high potential areas, improvements should be made in the services for a more customer oriented experience.

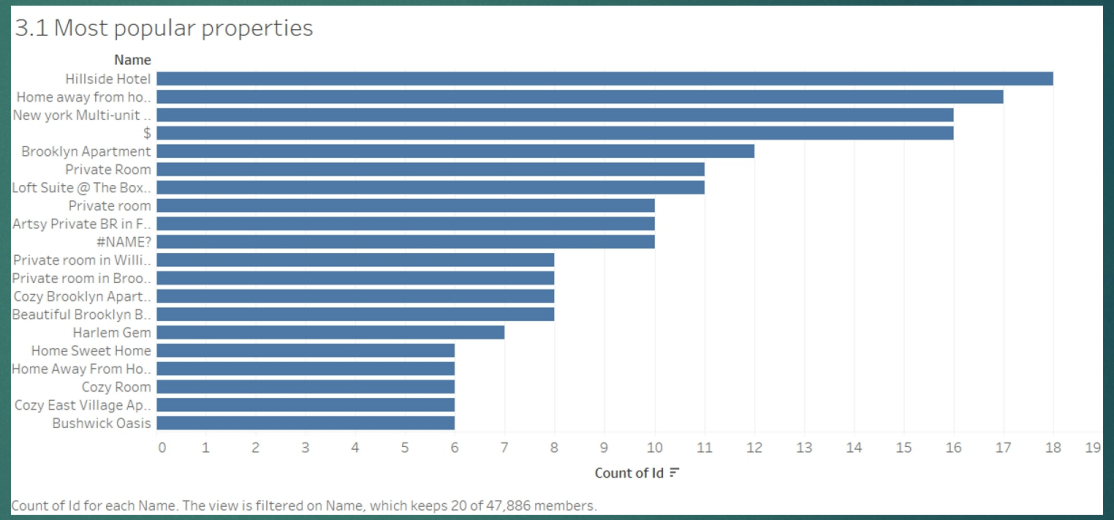
Also, entire homes/apartments are more customer oriented and provide a better experience as compared to private rooms.



Categorization of customers based on their preferences

Most popular localities:

Most popular localities are in Brooklyn and Manhattan. Bedford-Stuyvesant (10% of total reviews), Williamsburg (8% of total reviews) belong to the Brooklyn area, followed by Harlem in Manhattan (7% of total reviews) and fourth being Bushwick (5% of total reviews) again in Brooklyn.



Recommendation

- Manhattan and Brooklyn areas are preferred neighbourhoods area. Hence those areas should be targeted.
- Entire homes/apartments are the most preferred, hence they should be targeted.
- Entire homes/apartments are more customer oriented and provide a better experience as compared to private rooms, hence customer services for private room should be improved.

Appendix

- The important variable in the data dictionary are neighborhood_group, room_type, price , location & reviews.
- Data methodology has been included with the presentation in the file folder.
- We have used AirBnB data from the year 2019.

Thank You