Insights from Airbnb Analysis from Pre-COVID Period

Ameya Shukla
Rohit Ram

AGENDA

- Background
- Objective
- Data Preparations and Cleaning
- Data Dictionary
- Insights
- Appendix:
  - Data Methodology
  - Data Model Assumptions

BACKGROUND

- For the past few months, Airbnb has seen a major decline in revenue due to COVID-19.
- Now that the restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change to recover the losses.
- The different leaders at Airbnb want to understand some important insights based on various attributes in the dataset so as to increase the revenue

OBJECTIVE

- Suggest strategies for the business after pandemic.
- Find and understand important patterns and insights from pre-pandemic data of Airbnb.
- Make use of the insights to make further business decisions to improve the business.

DATA PREPARATION AND CLEANING

- Missing values were treated in few columns of the dataset.
- Simple statistical analysis of dataset was performed on the data using Python. There are categorical, numerical, and geographical variables in the dataset
- Correlation of different features in the dataset were evaluated. We could not find significant relationships among the features using heat map and pair plot.
- Created visualisations using the Tableau to find and understand key insights.

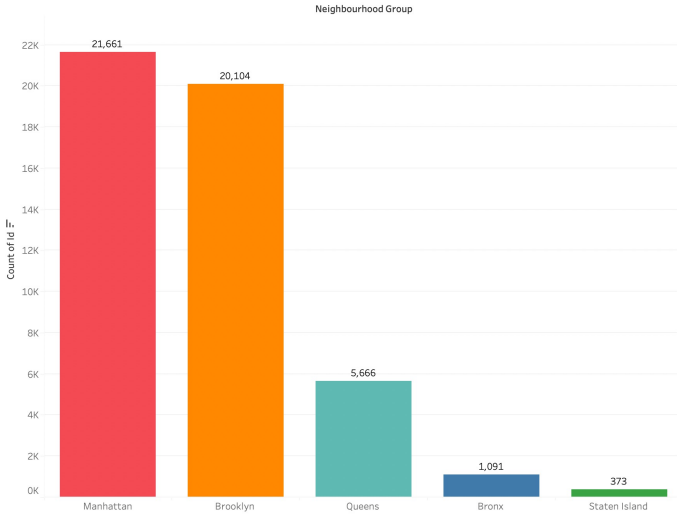
Data Dictionary

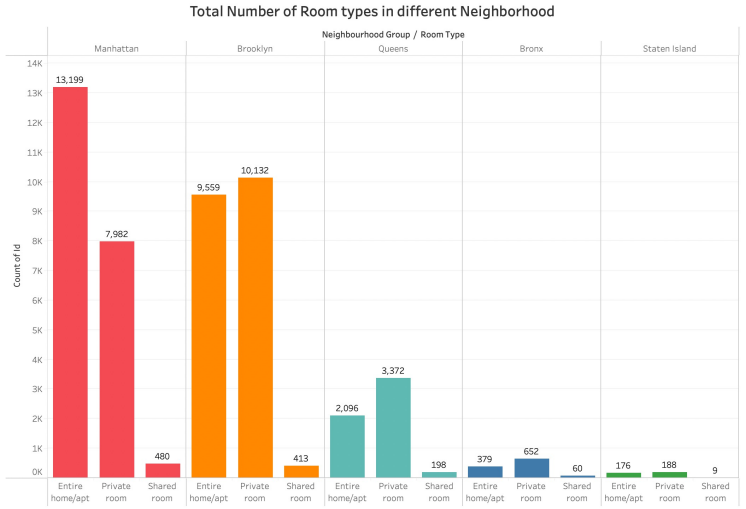
Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

```
Categorical Variables:
    - room_type
    - neighbourhood_group
    - neighbourhood

Continous Variables(Numerical):
    - Price
    - minimum_nights
    - number_of_reviews
    - reviews_per_month
    - calculated_host_listings_count
    - availability_365
- Continous Variables could be binned in to groups too

Location Varibles:
    - latitude
    - longitude

Time Varibale:
    - last_review
```

TOTAL NUMBER OF ROOMS TYPES IN DIFFERENT NEIGHBORHOOD

- Manhattan has the highest number of Entire home/apt, whereas, Brooklyn has the most Private Rooms.
- Shared Rooms are the lowest in count in every neighborhood group.
- Staten Island is having the lowest count of all of the room type.

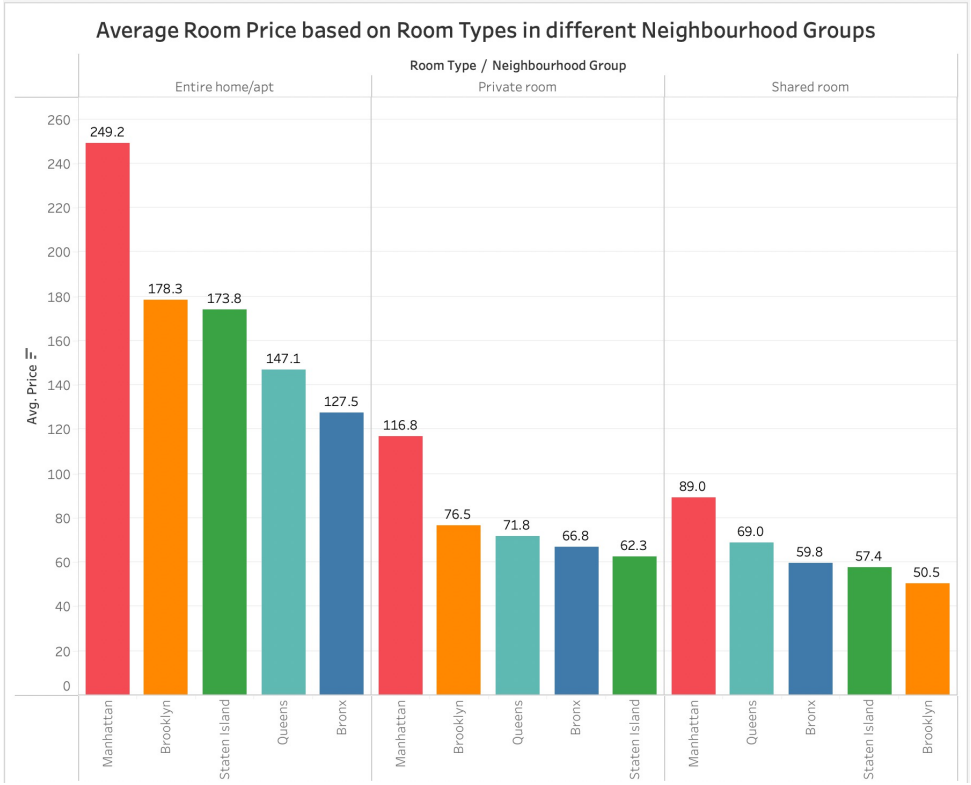




AVERAGE PRICE OF ROOM-TYPES IN DIFFERENT NEIGHBOURHOOD GROUPS

- Shared and private rooms are lower in terms of average price, whereas prices of Entire home/apt are on the higher side irrespective of the neighborhood group.

- Manhattan is the most expensive neighborhood group with higher average price in all three types of rooms as compared to others. This might be because Manhattan has higher number of rooms as compared to others.



ROOM TYPE PEOPLE PREFERENCES BASED ON TOTAL REVIEWS AND AVERAGE PRICE

- Entire home/apt rooms had the highest number of reviews along with having a trend of higher average price of rooms.
- Manhattan is the neighborhood group having the highest average price in all the three types of rooms however, Brooklyn has highest total number of reviews.
- Shared rooms have lower prices along with low total number of reviews.

APPENDIX : DATA ASSUMPTIONS

- Price column has been assumed to represent price per night.
- The average price and total number of reviews has been assumed to be a fundamental measure for finding customer preferences.
- For analysis related to price, average of the price is used.
- For total number of reviews, the sum of the number of reviews has been considered.
- For Dual Axis plots, both the axes have not been synchronized as values for both the features have been ranging widely in magnitude.
- Excluded last_review column from analysis.

APPENDIX : DATA METHODOLOGY

- We performed detailed exploratory data analysis on Airbnb data.
- The data cleaning and processing was done using Python and the visualisations were created using Tableau.
- The process included the following:
  - Using Python, found missing values and performed data cleaning.
  - Performed sanity check on columns Id and Price to check whether all Id's are unique and found that there are no negative price present.
  - Using Tableau, created visualisations to identify customer preferences based on
    - Neighborhood Group
    - Room Type
    - Average Price
    - Number of Reviews
  - Finally, found important insights based on the each of the created charts.

THANK YOU