Data Story Telling Case Study

BY
ASHITHA SWAROOP
RASHI GUPTA
RISHIKESH PATIL

Objectives

Identify

Analyze historical data to identify customer behavior of Airbnb in respect of their preferences.

Investigate

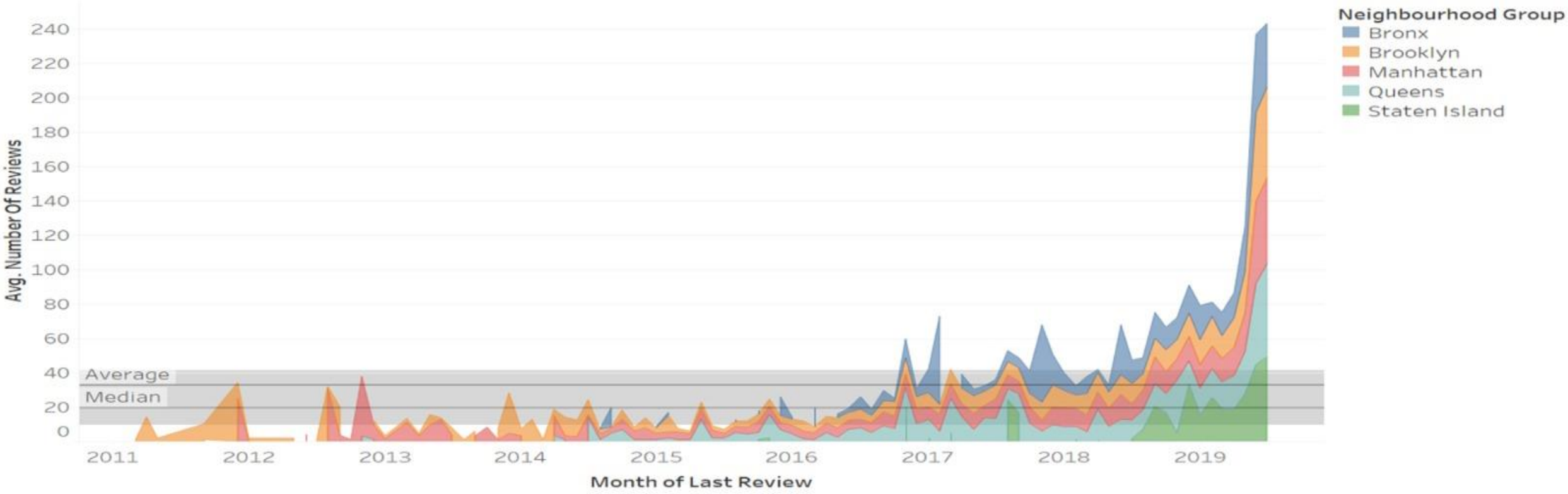
Investigate potential factors such as price, location preferences, and type of rooms that influences the customer

Suggest

Utilize historical data on bookings, pricing, and customer preferences to build accurate and reliable revenue forecasting models.

Seasonal Pattern of Reviews by Location

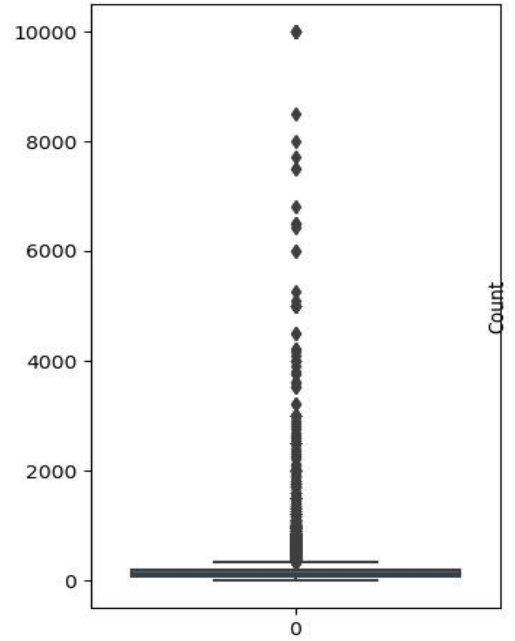
- The data reveals a noteworthy surge in reviews across all Neighborhood Groups during July 2019.
- Review activity demonstrates a seasonal pattern with peak periods observed in the winter months (December, January, February) and the summer months (April, May, June, and July).



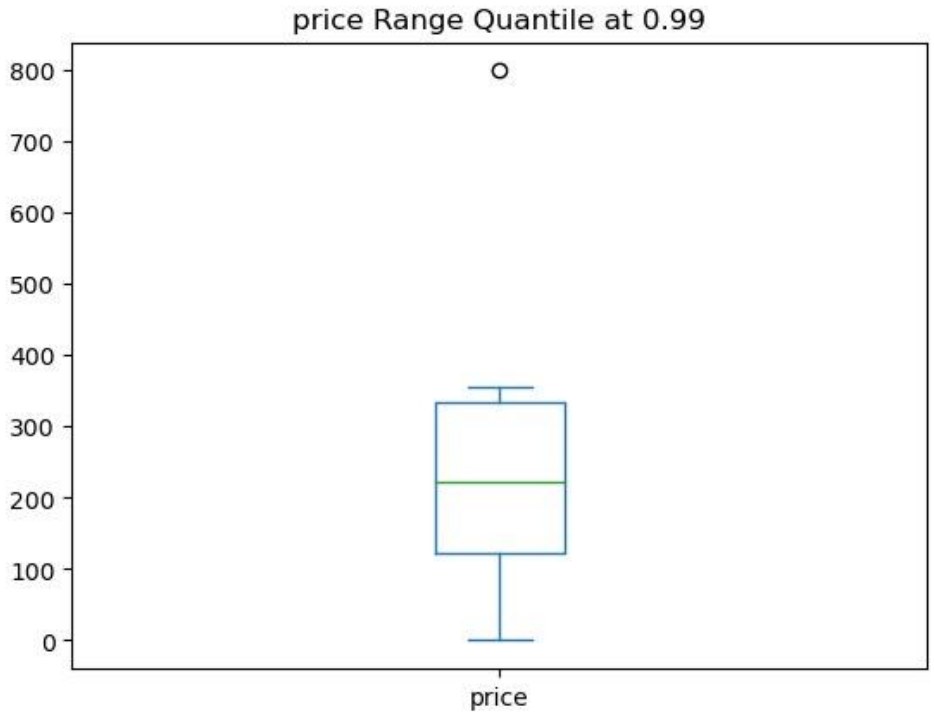
The plot of average of Number Of Reviews for Last Review Month. Colour shows details about Neighbourhood Group.The data is filtered on Last Review Year, which excludes 2010.

Univariate analysis of Price

- By examining the statistics, it is evident that the minimum value for the column is 0, while the maximum value is 10000. This indicates a significant range of values within the dataset.
- The notable disparity between the mean and median values suggests the existence of an outlier in the dataset, potentially skewing the average value.
- The distribution of the column as a whole is left-skewed, implying that there are fewer hosts with a substantial number of listings compared to those with fewer listings.
- When analyzing and quantifying the data at the 0.99 percentile, we observe a notable difference in the values, further highlighting the presence of extreme outliers.

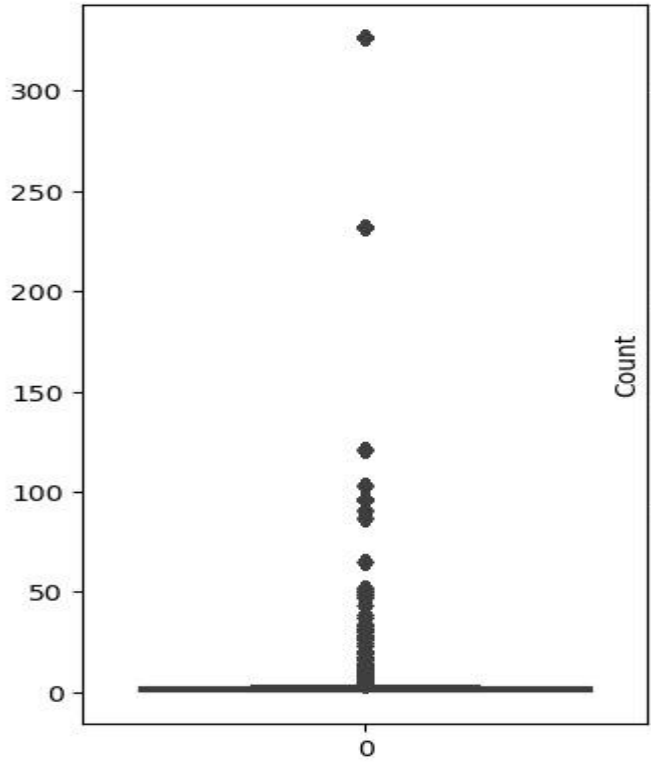


The trend of average of Number Of Reviews for Price (bin).

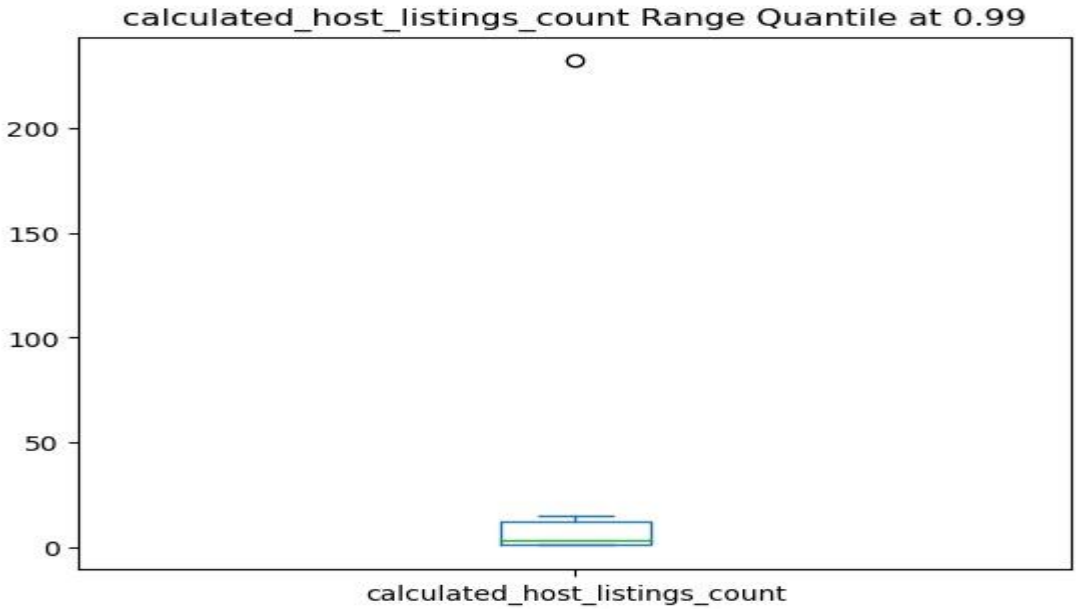


Univariate analysis of calculated_host_listings

- The data reveals that hosts on Airbnb have listed a varying number of properties, with the lowest being 1 and the highest being 327. On average, hosts have listed 7 properties.
- Furthermore, the 75th quantile indicates that 75% of hosts have 2 or fewer properties listed, suggesting that the majority of hosts have a relatively small number of listings.
- Overall, the column distribution is left-skewed, indicating that there are relatively fewer hosts with a large number of listings compared to hosts with fewer listings.



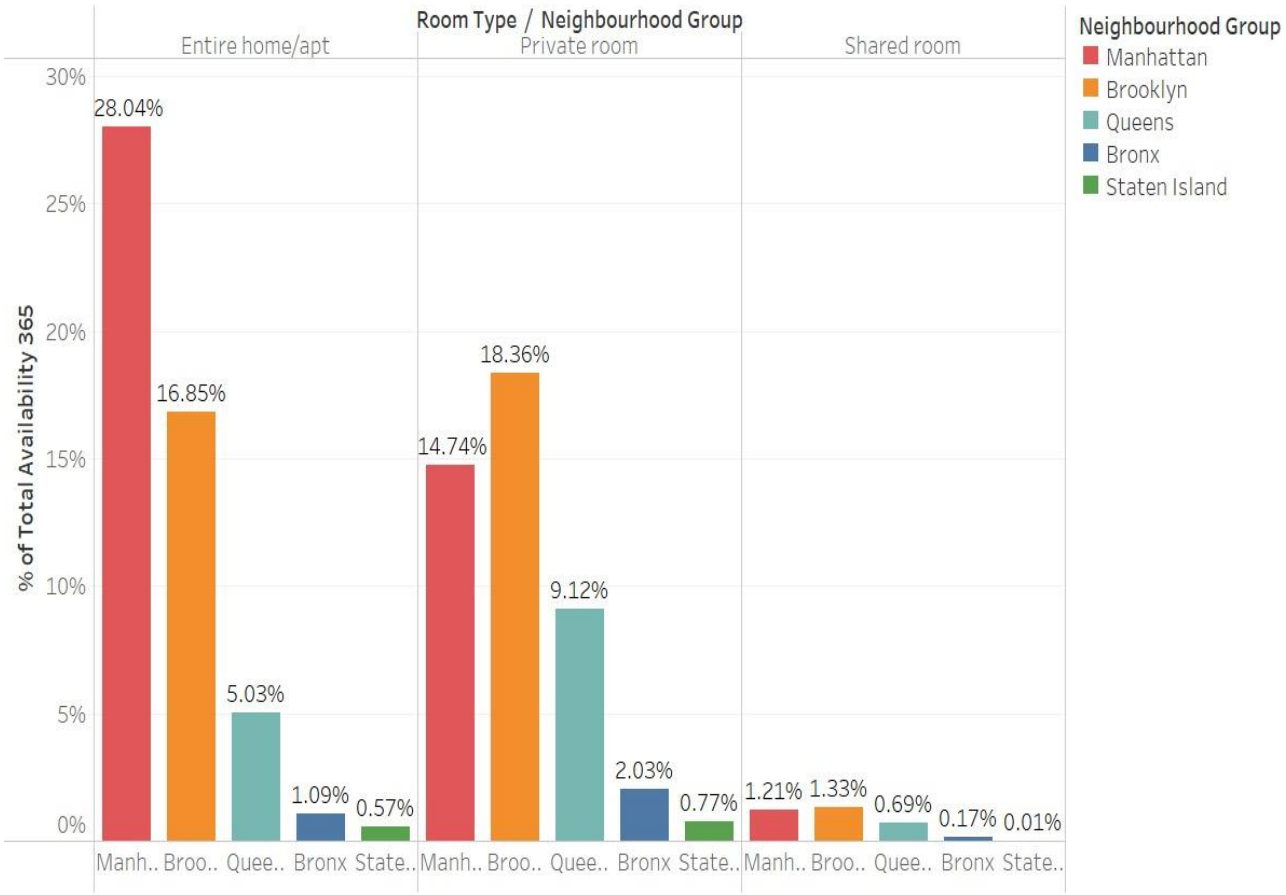

```
**********************************************************
Analysing outliers for  calculated_host_listings_count
**********************************************************
count    48895.000000
mean         7.143982
std         32.952519
min          1.000000
25%          1.000000
50%          1.000000
75%          2.000000
max        327.000000
Name: calculated_host_listings_count, dtype: object
**********************************************************
```



Analysis of Room Type as per Neighborhood Groups

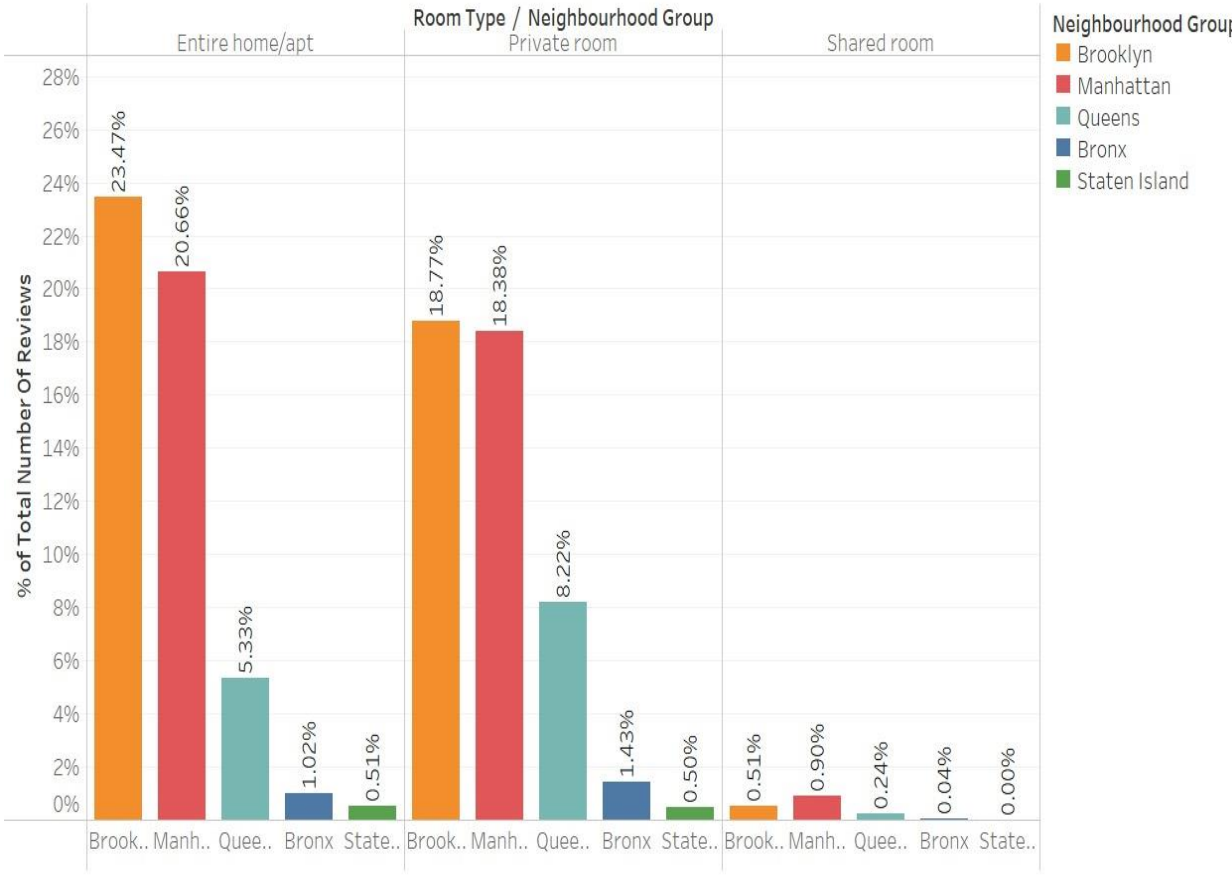
- Based on the data analysis, it is recommended that Airbnb focuses its promotional efforts on entire homes/apartments in Manhattan. This is because the data shows a higher availability of such accommodations in that area.
- The data suggests that Brooklyn is a popular location with limited availability of rooms. To capitalize on this demand, Airbnb should consider acquiring more properties in Brooklyn to meet the growing demand from potential guests.

Location Wise_Availablity of Room Type



% of Total Availability 365 for each Neighbourhood Group broken down by Room Type. Colour shows details about Neighbourhood Group. The marks are labelled by % of Total Availability 365.

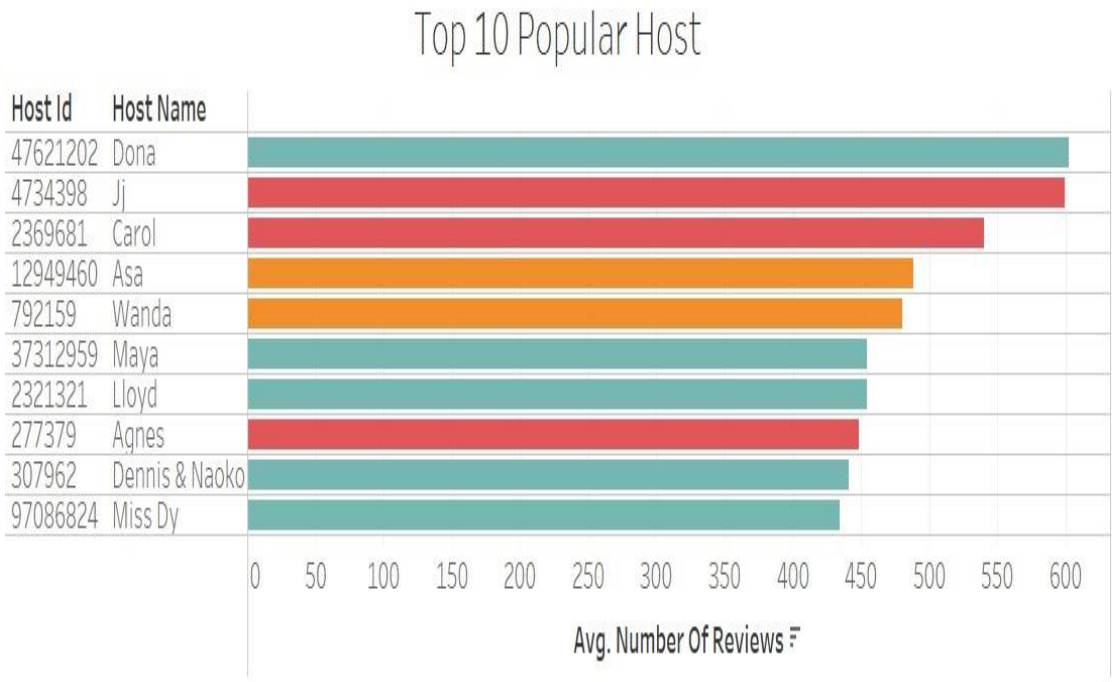
Location wise _Popular_Room Type



% of Total Number Of Reviews for each Neighbourhood Group broken down by Room Type. Colour shows details about Neighbourhood Group. The marks are labelled by % of Total Number Of Reviews.

Analysis of Top 10 hosts based on popularity and number of properties

- The majority of the Top 10 hosts are having properties in Manhattan.
- The majority of hosts in the list are located in Queens.
- Among the top hosts, Dona (ID: 47621202), Jj (ID: 4734398), and Carol (ID: 2369681) stand out with the highest number of reviews.



Average of Number Of Reviews for each Host Name broken down by Host Id. Colour shows details about Neighbourhood Group.The view is filtered on Host Id, which keeps 10 of 37,457 members.

Count of Calculated Host Listings Count for each Host Name broken down by Host Id. Colour shows details about Neighbourhood Group.The view is filtered on Host Id, which keeps 10 of 37,457 members.

APPENDIX - DATA SOURCES:

- Here is a snapshot of our data dictionary.
  - Different Airbnb listings along with their hosts, locations, prices, and other attributes.
  - Properties information such as location, price, reviews, type of Room.
- The team used the following data sources:
  - New York Airbnbs Dataset
  - Survey conducted in 2019

APPENDIX - DATA METHODOLOGY

- We conducted a thorough analysis of the AirBnb Data. The process included:
  - Cleaning and analysis done using Python.
  - Analysis done using Excel.
  - Visual Insight using Tableau.

Assumptions

- The number of reviews provided in the dataset is considered a positive measure to analyze customer preferences and satisfaction.
- We acknowledge the presence of properties with the existence of outliers. However, for the purpose of our analysis, we assume that these outliers will not significantly impact our findings and conclusions.
- Null values present in the dataset are assumed to have no significant impact on the analysis conducted

Suggestions

- Develop targeted promotional campaigns and pricing strategies to capitalize on the winter and summer peaks in review activity.
- Collaborate with marketing and operations teams to effectively communicate the value proposition to potential guests during high-review seasons.
- Clustering algorithms such as K-means can be used to identify groups of listings and customer behavior that have similar characteristics and plan targeted strategies for each cluster to optimize revenue.

Thank You

For Your Attention