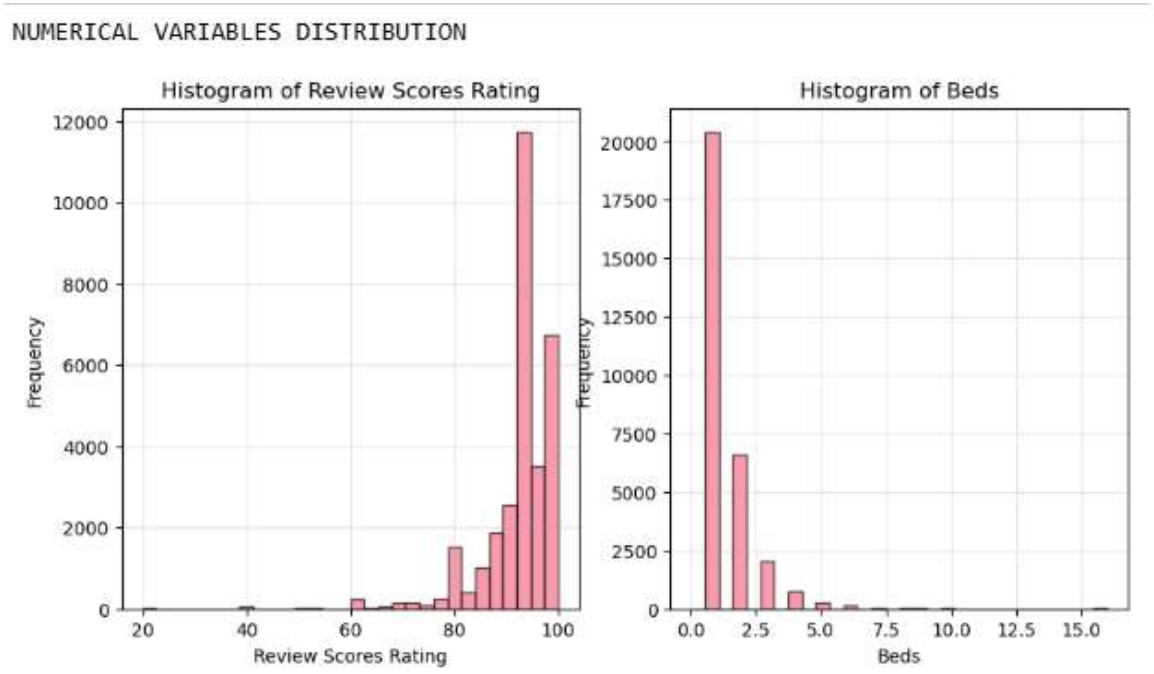
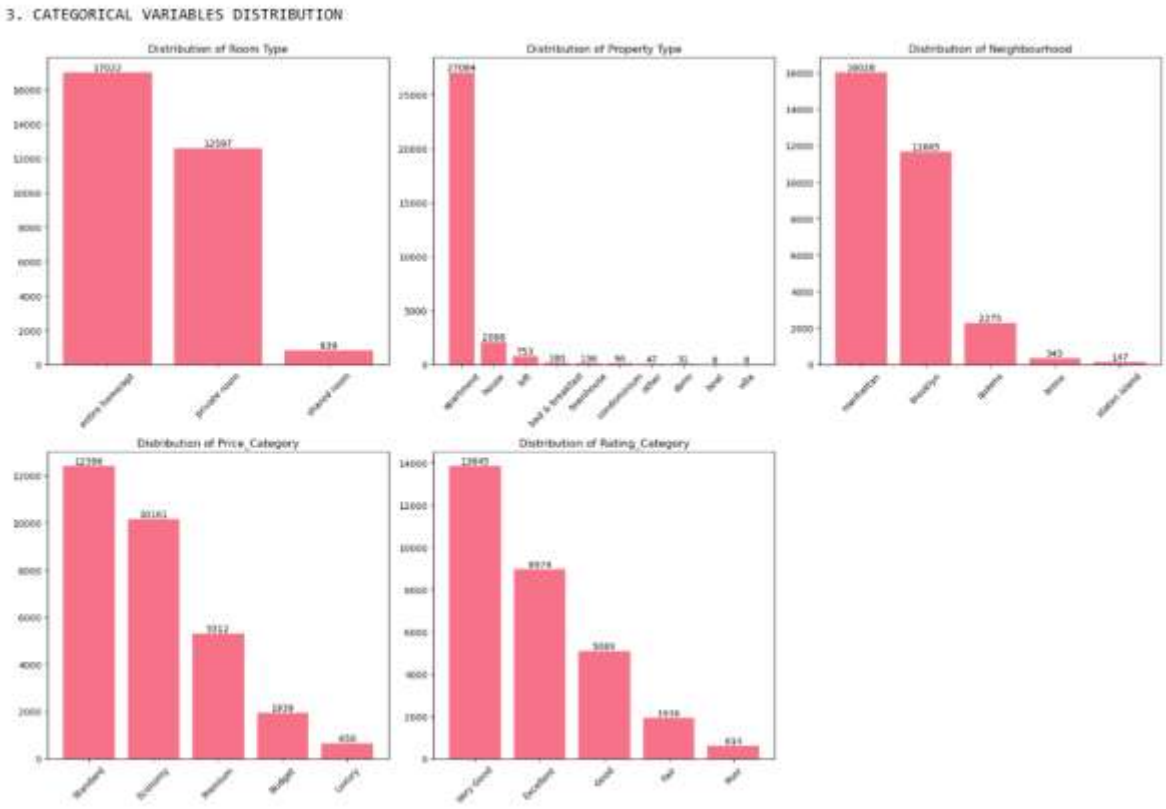
Airbnb Listing Analysis:

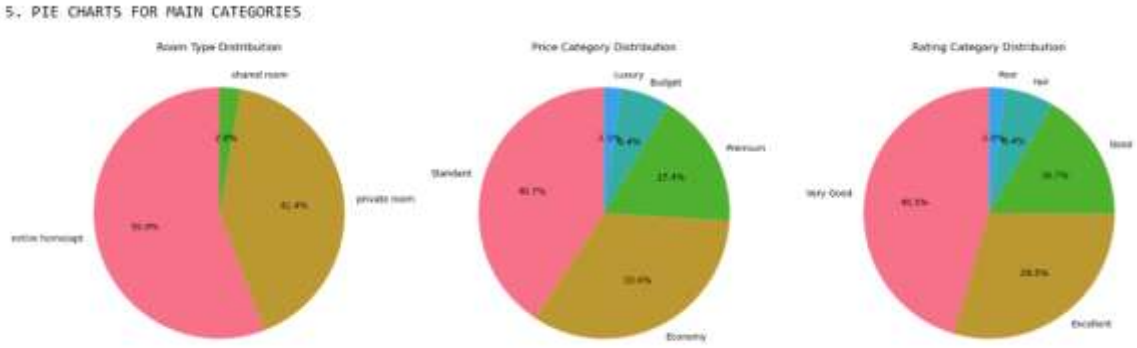


For understanding data, histograms of a few numerical columns were visualized. The 1st histogram tells us about the frequency of review score rating, and high number of ratings are above 90 as shown in image. The 2nd histogram shows the frequency of number od beds that will be available, it shows that more than 20,000 places will have 1 bead and as the number of beds increases their frequency also decreases.



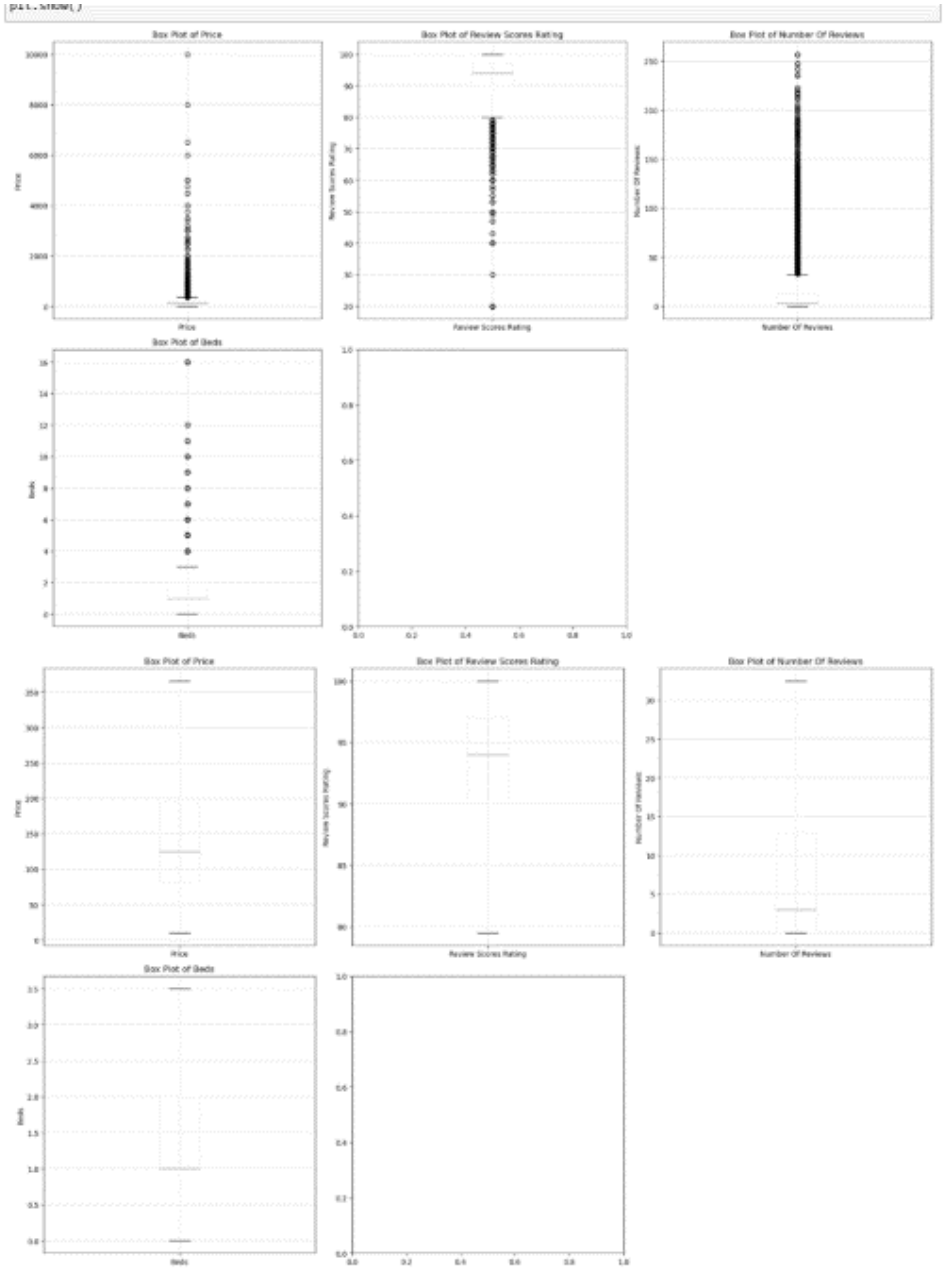
Same goes for categorical columns:

- 1st, the distribution of room types is shown with the help of bar chart, the highest frequency is of entire home/apt and lowest is of shared room.
- In 2nd graph, the distribution of property type is shown with the help of bar chart, the highest frequency is of apartment and lowest is of villa.
- In 3rd graph, the distribution of neighborhood is shown with the help of bar chart, the highest frequency is of Manhattan and lowest is of Staten Island.
- In 4th graph, the distribution of price category is shown with the help of bar chart, the highest frequency is of standard and lowest is of luxury.
- In the 5th graph, distribution of rating category is shown with the help of bar chart, the highest frequency is of very good and lowest is of poor.
-

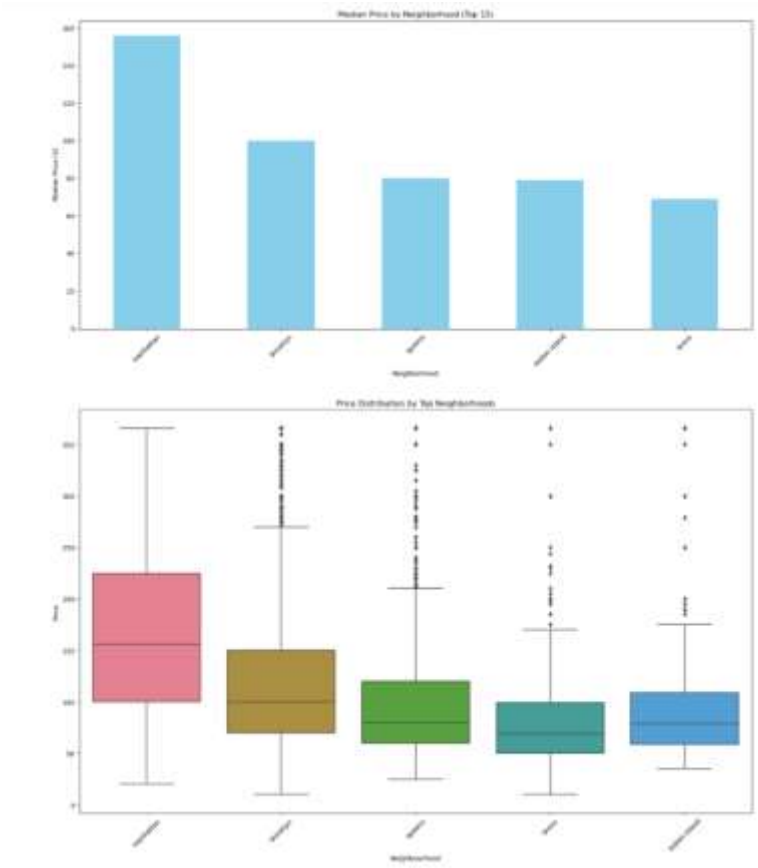


The above visualization shows the distribution of main categorical columns present in the data.

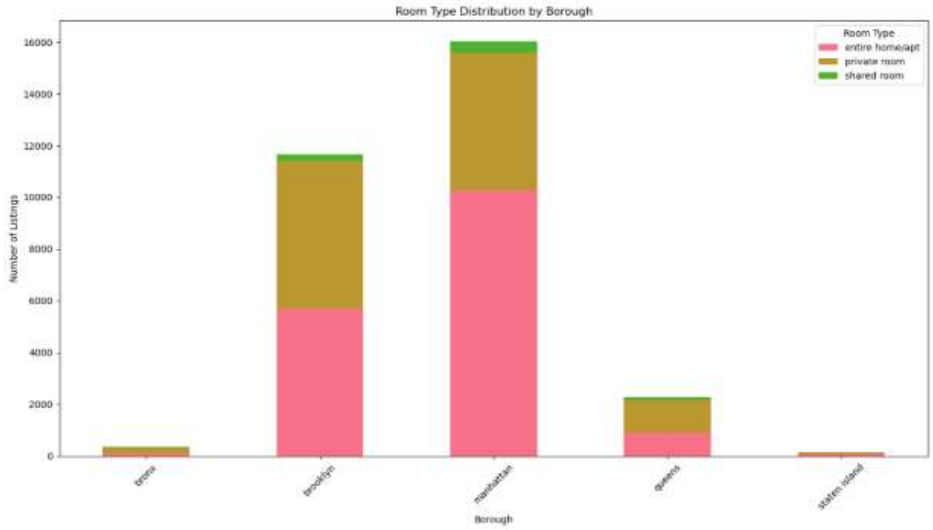
- The 1st pie chart shows distribution of room type, where entire home/apt is having majority part, as shown through pie chart.
- The 2nd pie chart shows distribution of price, where standard is having majority part, as shown through pie chart.
- The 3rd pie chart shows distribution of rating category, where ‘very good’ is having majority part, as shown through pie chart.



In order to perform further analysis on data, we have to remove the outliers. To get an idea of outliers we are using boxplots before and after so that we can verify our results. The line in between the box represents the mean and the upper line represents the highest value with data, same goes for the line that is at the bottom of box (for representing the lowest value within data). The box upper line represents 1st quartile and lower boundary represents 3rd quartile within data.



This visualization is bivariate and represents Pricing trends across different neighborhoods, the histogram. Shows that Manhattan has highest prices and lowest are of Bronx, Staten island and queens have almost same prices distribution. Hence by seeing this we can get an idea at glance about neighborhood and their respective prices



In graph 1:

Room Type:

Room Type

entire home/apt 17022

private room 12596

shared room 839

Name: count, dtype: int64

This is the given data in xlsx file and our stacked bar chart reflects it correctly! Representing the relationship between

Listings of Neighborhood and the room type within them

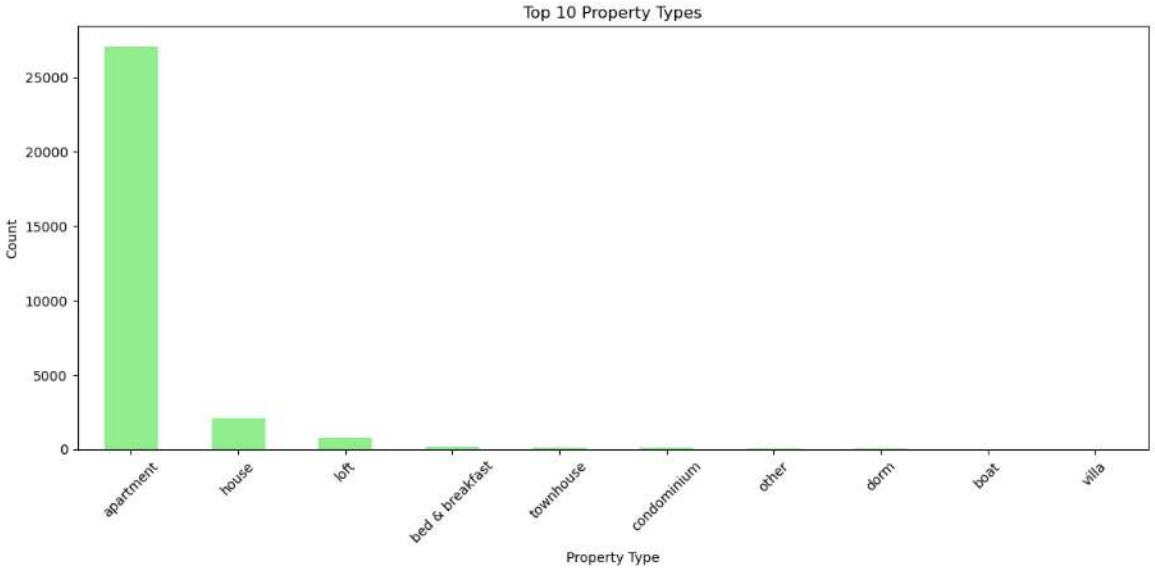
manhattan 16028

brooklyn 11665

queens 2274

bronx 343

staten island 147



In graph 2 :given property types

Property Type

apartment 27084

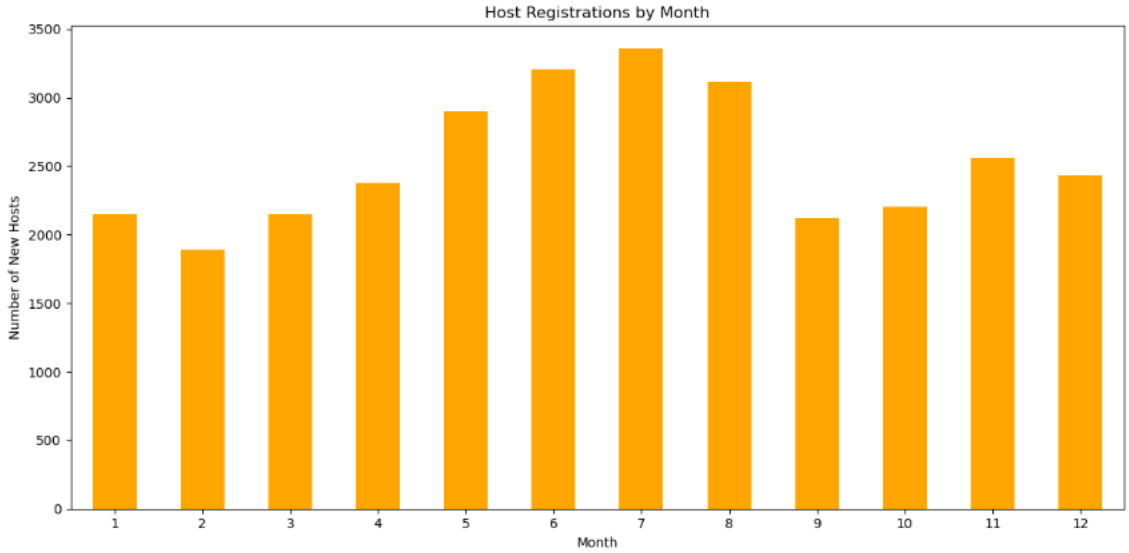
house 2087

loft 753

bed & breakfast 180

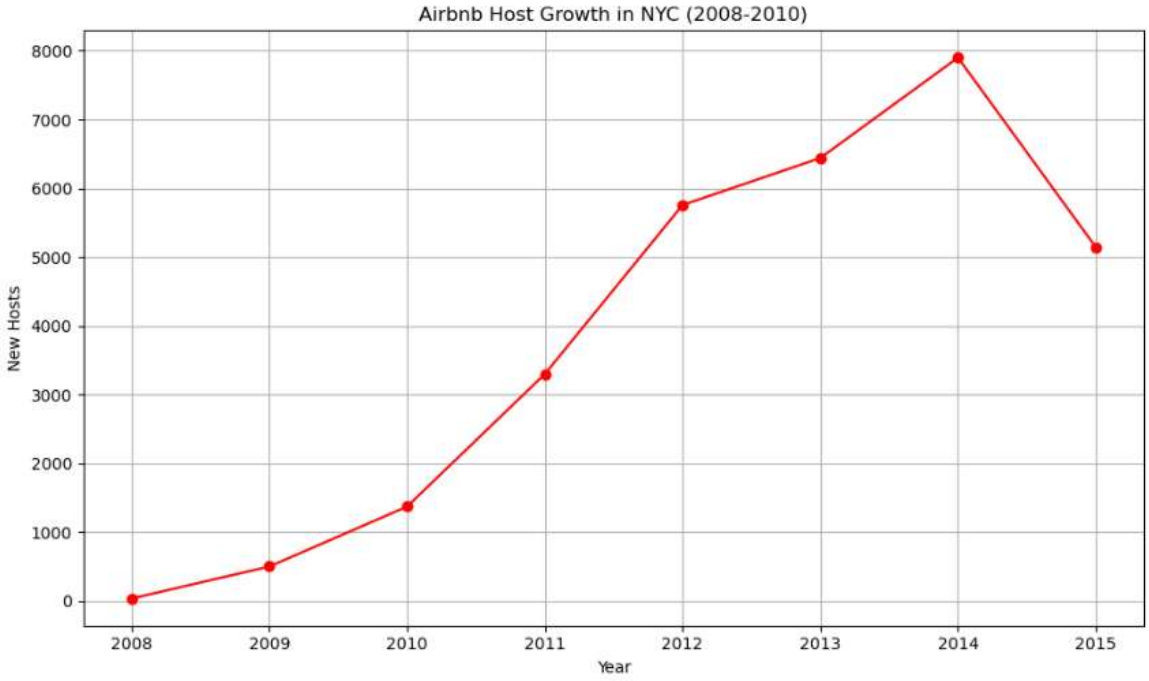
townhouse 136

and as shown by graph also



Graph1: Shows the monthly variation of hosts, where hosts are on y axis and months by x-axis. in this bivariate visualization

We can see hosts are increased the most in the 7th month as shown by the bar chart.



In Graph2:

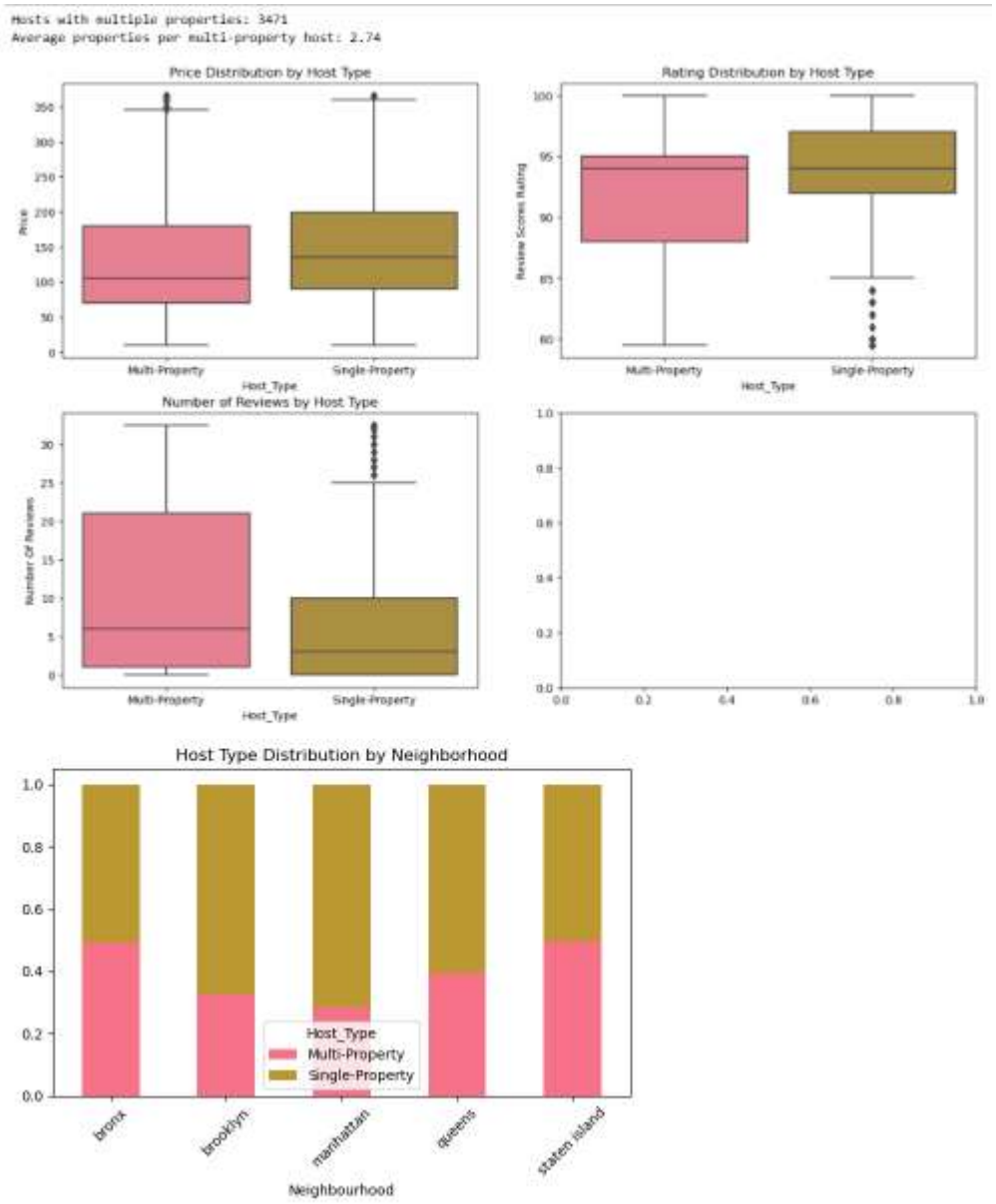
We are having a bivariate relationship between years and hosts

Where years are on x axis and hosts number on y axis

from the shown line plot we can clearly see that in 2014 we have the highest number of hosts that is 8000

the hosts number started to increase in rapidly after 2010 and decline after 2014

the rapid decline as shown was of 8000 to 5000 between 2014 and 2015



In the above visualizations we are seeing different attributes related to airbnb categorized in two hosts types

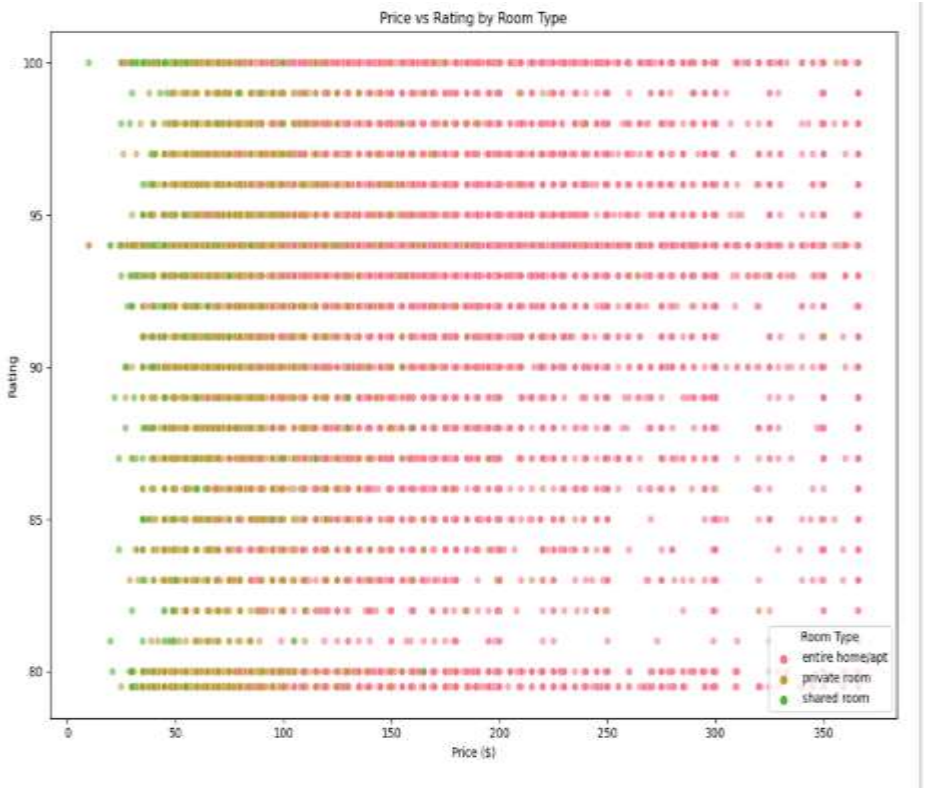
1. multi-property
2. single-property

-According to graph (bos plot) the single-property hosts are more high avg price than multi-property hosts

-According to graph(box ploy) the single-property hosts are more high ratings than multi-property hosts, but avg is almost same

-According to graph(box ploy) the single-property hosts have less no of reviews than multi-property hosts, also avg is of single-property hosts is low.

-The bar chart tells us about the distribution of multi-property hosts vs single-property hosts in different neighborhoods

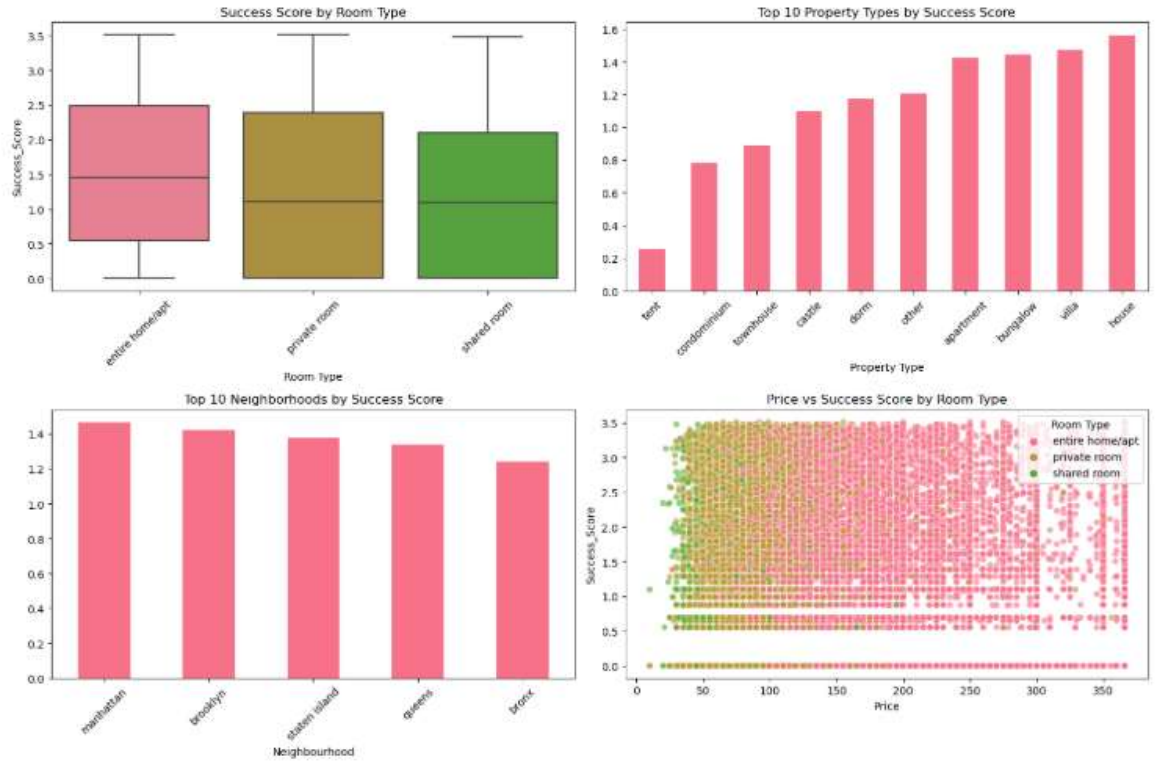


The Above visualization is a scatter plot between price – on x-axis and Ratings – on y-axis. With hue having room types. We can see that high prices (they also lie at avg) are of entire home/apt, and they have more ratings as they are more in number. And shared room are low in price while private room price lies between lowest and average prices. The ratings are almost propotional to their numbers

A Correlation Matrix between: Price, Rating, Reviews & Beds, along with their scale. This shows that direct relation between beds and prices which is valid here. Almost no relationship between beds and number of reviews, which is again acceptable and some other inverse relations are also shown through the matrix

The Above visualization is a scatter plot between price – on x-axis and Ratings – on y-axis. With hue having Neighborhoods. We can see that high prices are mostly of Manhattan and Brooklyn and some of queens. Low prices and average prices are spread between other two along with high price neighborhoods. But mostly Staten Island neighborhood is found at low prices.

a successful Airbnb listing (Multivariate Analysis)



For all graphs given above a success rate been calculated by considering the high ratings AND high number of reviews as success.

In graph 1, we can see that success is determined by room type, and entire home/apt has high success rate, the average of private and shared rooms are almost equal. Shared room has low success rate as compared to private room.

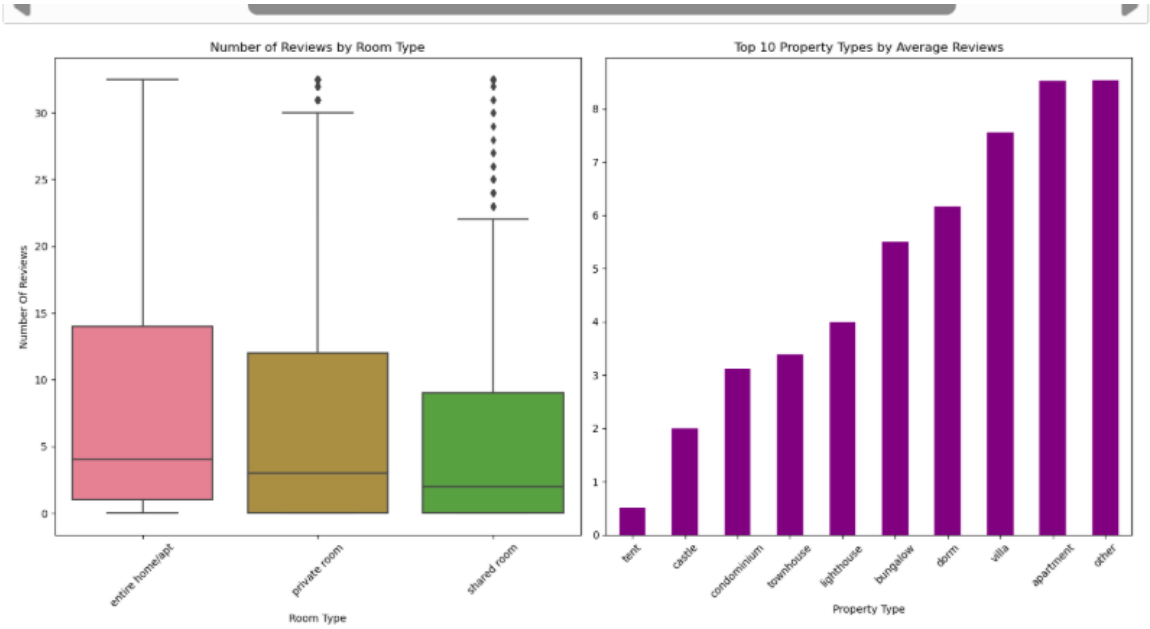
In graph 2, success rate is determined by property type, where top ten property types are on x-axis and the success rate on y-axis and the success per property is represented in form of bar chart. The house has the highest success rate and tent has lowest, as shown in bar chart.

In graph 3, the neighborhood determines the success rate, , where neighborhoods are on x-axis and the success rate on y-axis and the success per neighborhood is represented in form of bar chart. The manhattan has the highest success rate and Bronx has lowest , as shown in bar chart.

In graph 4, , the price determines the success rate, , where room price on x-axis and the success rate on y-axis and the success is represented in form of scatterplot. We are having hue of room type here as well.

CONSUMER BEHAVIOR

What types of properties get more reviews (Bivariate Analysis)

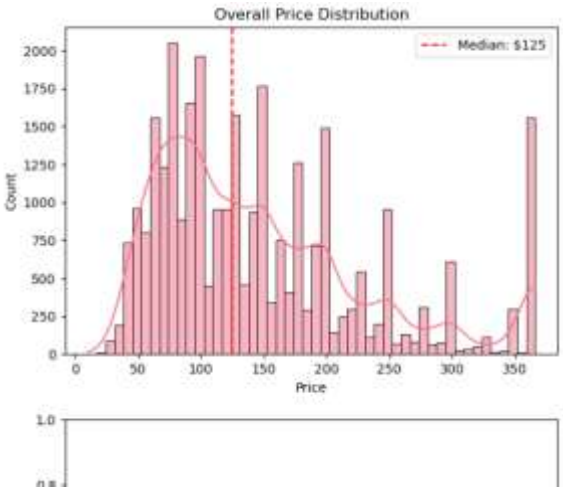


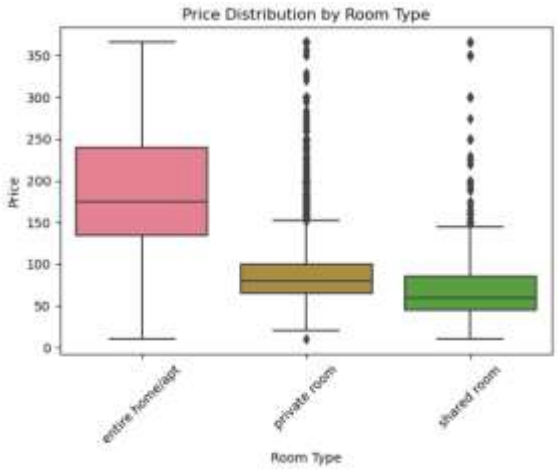
Now we will be discussing the consumer behavior based on the number of reviews.

In graph 1, the room type determines the customer behavior. Where room type is on x-axis and the box plot represents the behavior. We can see clearly that entire home/apt has more no.of reviews and then we have private room and again at lowest we have shared room.

In grap 2, we are detetrming the customer behavior on the basis of property type. Having property type on x-axis and the no.of reviews as shown in the form of bars. We can conclude from the visualization that tenst has less number of reviews and the apartment and other has highest number of reviews and all rest lie in between these two.

Price ranges guests are willing to pay - Bivariate Analysis



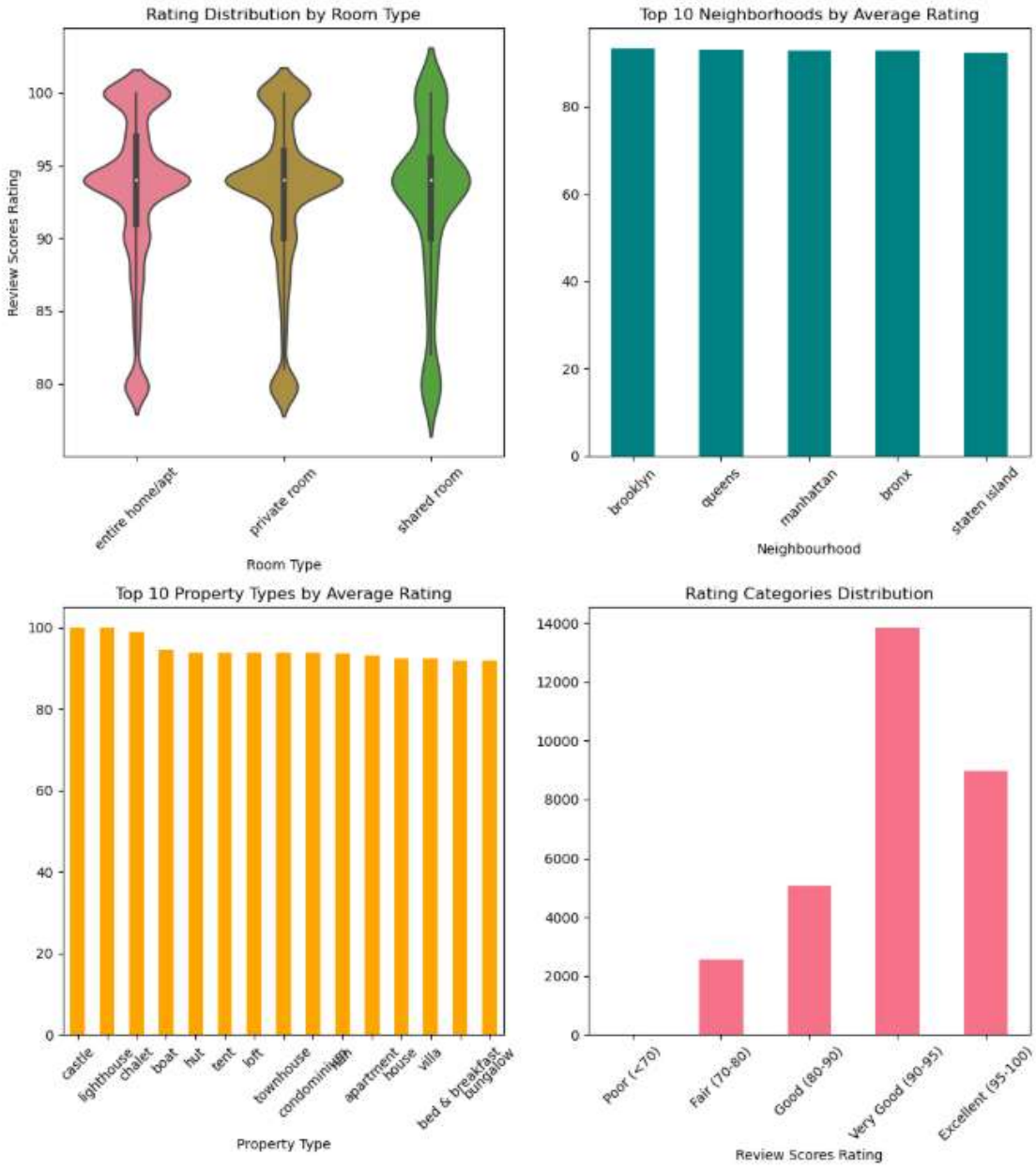


In this graph we will be discussion the price ranges that guests are willing to pay, the first graph shows the overall distribution of prices along with their respective count, we are having median of 125$ and highest price as of 350$ with the count of above 1500. Similarly all prices are given with their respective count in form of histogram.

The second graph represents the price distribution by room type. It is clear from graph that entire home/apt has highest price, then we have private room and lastly shared room as the cheapest option. The averages are also visualized in the form of box plots.

The graph given above gives a visualization of price ranges along with the number of listings, in the form of bar chart. We can observe that the price range of 50-100$ have the highest listings and of below 50$ and above 500$ have the lowest listing.

Rating patterns and guest satisfaction - Multivariate Analysis

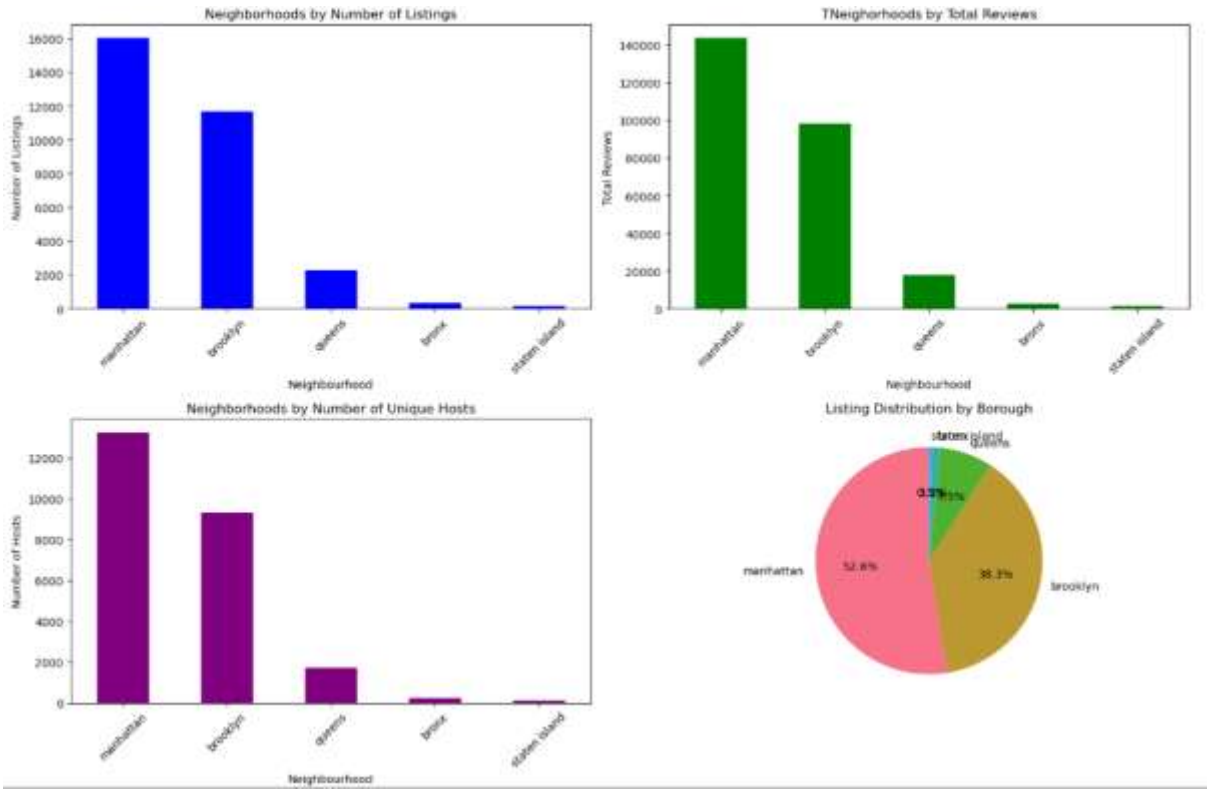


In all above plots we are discussing the Rating patterns and guest satisfaction.

In 1st plot we are having violin plots that help us visualize the 3 room types and their review score ratings. The 2nd plot is a bar plot giving us idea about the neighborhood and the average ratings, it shows that all neighborhood has almost same ratings with minor differences. The 3rd again a bar chart shows different property types and their ratings, With castle having heighest rating. The 4th gives us the Rating categories distribution with the help of bar chart, and it seems as if most of guests are satisfied from their services cause most of the ratings are "very good".

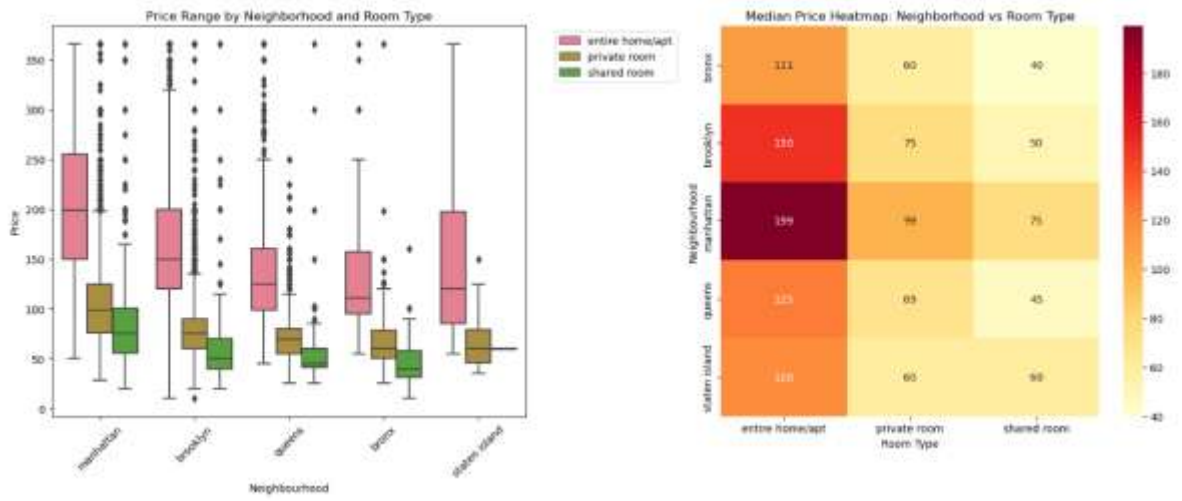
NEIGHBORHOOD COMPARISONS

Which areas have more Airbnb activity - Bivariate Analysis



The above analysis are done in order to find out which neighborhood has the highest activity. The results by four different graphs Neighborhoods by Number of Listings, Neighborhoods by Total Reviews, Neighborhoods by Number of Unique Hosts and by one pie chart it is clearly shown that our answer is manhattan, then we have Brooklyn neighborhood and at lowest we are having staten island.

Price variations by location - Multivariate Analysis



For price variation with respect to locations we are having two visualization here. The 1st one shows the boxplot of neighborhood vs price and as hue we are having room type. The graph shows high prices of

entire home/apt at every neighborhood, with manhattan being the most expensive. Same results are depicted the Heatmap, shown as the 2nd image to the right of boxplots.

Significant Findings & Patterns:

- **Market Dominance of Entire Homes/Apartments:** "Entire home/apt" is the most common room type and also has the highest success rate (defined by high ratings and high number of reviews), highest number of reviews, and commands the highest prices.
- **Geographical Concentration:** Airbnb activity is heavily concentrated in **Manhattan** and **Brooklyn**. These neighborhoods have the highest number of listings, hosts, total reviews, and success rates. Manhattan is consistently the most expensive location.
- **Price Determinants:** There is a clear, direct relationship between the number of beds and the price. Location (Neighborhood) and Room Type are the strongest determinants of price.
- **High Overall Satisfaction:** The vast majority of ratings are above 90, categorized as "very good," indicating generally high guest satisfaction across the platform, with minor variations between neighborhoods and property types.
- **Host Structure:** The market is dominated by "single-property" hosts. Interestingly, these hosts tend to have higher average prices and ratings but fewer reviews compared to "multi-property" hosts.
- **Price Sensitivity:** The market shows high demand for mid-range prices. The price range with the most listings is $50-$100, suggesting this is the most competitive and popular segment.

Anomalies & Unexpected Findings:

- **Host Growth Anomaly:** There was a rapid and significant decline in the number of hosts from 2014 (8,000) to 2015 (5,000). This sharp drop is unusual for a growing company and could indicate a major market event, such as a regulatory change or data collection issue.
- **Success Rate vs. Property Type:** While "Apartment" is the most numerous property type, "House" is reported to have the highest success rate. This is counter-intuitive and suggests that less common property types can be more desirable, potentially due to space or unique characteristics.
- **Single vs. Multi-Property Hosts:** The finding that single-property hosts charge higher prices and achieve similar or higher ratings is noteworthy. It contradicts the assumption that professional, multi-property hosts might be more efficient. This could be due to single-property hosts offering more personalized service or unique properties.
- **Rating-Price Relationship:** The scatter plots show that high prices do not guarantee high ratings, and high ratings are achieved across all price points. The relationship is not strongly positive, indicating guests' value is based on more than just price.

Summary of Key Insights:

The analysis reveals a mature and stratified Airbnb market in NYC. The platform is dominated by entire homes and apartments, primarily located in Manhattan and Brooklyn, which are also the most expensive and successful areas. Guests are highly satisfied overall and are most active in the mid-price range ($50-

$100). The host community is primarily composed of single-property hosts who, despite having fewer reviews, maintain competitive prices and high ratings.

Initial Hypothesis:

"Listings in more central neighborhoods (like Manhattan) and entire homes will be more successful and expensive."

The visual analysis strongly supports this hypothesis**.** The data consistently shows Manhattan and "Entire home/apt" as leaders in price, quantity, and success metrics.

Discussion of Unexpected Findings:

- The sharp decline in hosts in 2014-2015 is the most significant anomaly. Its potential implication is major; it could reflect the impact of new legislation (like the 2016 Illegal Hotel Law), a market correction, or a data anomaly that warrants immediate investigation.
- The higher performance of "Houses" over "Apartments" in success rate implies that guests may value space and privacy highly, offering a strategic insight for potential hosts.
- The performance of single-property hosts suggests that professionalism (in terms of managing multiple properties) does not necessarily lead to a price or quality premium, which is an important nuance for understanding the host ecosystem.

Recommendations for Further Analysis

- **Investigate the Host Decline:** Perform a deep dive into the data and external events (local news, law changes) around 2014-2015 to determine the cause of the sharp drop in unique hosts. This is critical for understanding market stability.
- **Incorporate Time-Series Analysis:** Analyze booking data over time to identify seasonal trends, the impact of specific events (holidays, concerts), and to see if the success metrics are stable or changing.
- **Explore Host-Specific Variables:** Analyze how host response time, acceptance rate, and superhost status impact pricing, occupancy, and success. This could explain the performance difference between single and multi-property hosts.
- **Conduct Competitive Analysis:** Compare the performance of listings that are adjacent to major tourist attractions versus those in residential areas to better understand the drivers of demand and price.