Dov Bechhofer

Prof. Vesselinov

DATA 334

May 16, 2024

Impact of Location on Pricing in NYC Airbnb Listings

Introduction:

Being an avid traveler from New York, I have been in a variety of Airbnb lodgings, from private rooms with all the solitude you could want to share spaces with other people. Even though these experiences have been enriching, these events have never taken place within my hometown of New York City. The underlying economics of Airbnb postings in New York City are not as simple as the answer would seem. This sparks my curiosity about a more important research question: What variables affect Airbnb prices in New York City, and how important is location compared to other possible variables?

It is no secret that New York is a Mecca for tourists. Additionally, the NYC area has three major airports, LaGuardia, Newark, and JFK. Many domestic and international travelers opt to stay in an Airbnb for a connecting flight. Considering that tourists usually use short-term rentals what do they value most? Is it location, access to public transportation, attractions nearby, accommodation type, or verified host? Knowing the price patterns of short-term rental homes may provide insights into urban economics, hospitality trends, and real estate management in a metropolitan as diverse and dynamic as New York. It can also shed light on the potential and difficulties encountered by locals, legislators, and companies operating in the gig economy. Furthermore, as New Yorkers Airbnb profoundly impacts our lives. Most Airbnb's are rented short-term, maybe just a few days. A recent New York City Housing and Vacancy Survey (NYCHVS) found that vacancy rate has dropped to 1.4 percent [1]. NYC does not meet housing demand and supply is extremely limited which inherently leads to higher rent costs for all of us. Airbnb listings could be used as long-term rentals, but landlords choose to prioritize profits and cash in on tourists.

Methods:

Each variable chosen for analysis either directly influences pricing or provides contextual understanding critical for a comprehensive market analysis. The description of these variables sets the stage for the exploration of their relationships through various visualization techniques, aiming to uncover patterns that answer the research question regarding the impact of location and other factors on Airbnb pricing in New York City.

After cleaning, the dataset contains 100,337 observations and 19 variables/columns.

Variables:

- *'borough':* Categorical variable representing the borough in which the listing is located. This includes Manhattan, Brooklyn, Queens, Bronx, and Staten Island.
- *'neighborhood'*: More specific than borough, this categorical variable indicates the neighborhood of the listing.
- *'lat and long':* These are the latitude and longitude coordinates of the Airbnb listing. They provide precise geographical positioning that can be used for mapping and spatial analysis.
- *'instant_bookable':* This boolean categorical variable shows whether a listing can be booked without host approval ('True' or 'False').
- *'cancellation_policy':* Describes the type of cancellation policy set by the host, such as flexible, moderate, or strict.
- *'room type'*: The type of room offered, categorized as Entire home/apt, Private room, or Shared room.
- *'Construction year':* The year the property was constructed.
- *'price':* The cost per night to book the listing (USD).
- *'service fee':* Additional fees charged by hosts, not included in the nightly price (USD).
- *'Minimum Nights':* The minimum number of nights a guest can book.
- *'number of reviews':* A count of how many reviews the listing has received.
- *'review rate number'*: A derived metric indicating the overall review score or rating.
- *'calculated host listings count':* The total number of listings managed by the host.
- *'availability 365':* The number of X days in the year the listing is available for booking.

- *'price_to_service_fee_ratio':* This variable represents the ratio of the listing price to the service fee associated with the listing (created by me).
- *'mean_price':* This variable indicates the average price for listings within a specific category, such as a borough (USD-created by me).
- *'price_index':* A calculated index that compares the price of each listing to the mean price of all listings within the same category (borough). A price index value of 1 indicates that the listing is priced at the average rate for its category, values above 1 indicate a price higher than the average, and values below 1 indicate a price lower than the average (created by me).
- *'price_category'*: A categorical designation derived from the price_index that classifies listings into three categories based on their price relative to the average: 'Expensive' (index > 1), 'Neutral' (index = 1), and 'Good Value' (index < 1)

Visualization:

1) Barplot of Listings by Borough
This barplot provides a clear visual representation of the count of Airbnb listings across the five boroughs of NYC, highlighting areas with the highest and lowest densities of listings. Understanding this distribution is crucial for further analysis.

2) Barplot of the Count of Accommodation Type
This plot categorizes listings into types such as Entire homes/apts, Private rooms, etc., offering insights into the most popular types of accommodations on Airbnb. This helps in understanding the makeup of the dataset and consumer preferences.

3) Stacked Barplot by Borough and Accommodation Type

By stacking accommodation types within each borough, this visualization illustrates the composition of the market in each area, providing a nuanced view of how different types of listings are distributed across boroughs.

4) Boxplot of Prices by Borough

This boxplot reveals the distribution of prices within each borough, highlighting the median, quartiles, and outliers. It provides a visual explanation of pricing variability between boroughs.

5) Boxplot of Prices by Accommodation Type

Similar to the borough analysis, this boxplot compares price distributions across different types of accommodations, showing how room types might influence pricing.

6) Boxplot of Prices by Borough and Accommodation Type

Combining borough and accommodation types, this dual-variable boxplot allows for a direct comparison of price distributions, enhancing our understanding of the interplay between location and accommodation type on pricing.

7) FacetGrid Histograms of Prices by Borough

These histograms show the frequency distribution of prices within each borough, highlighting where prices cluster and revealing any skewness or anomalies in pricing across different areas.

8) Line plot of Average Price by Construction Year

This line plot explores trends over time, analyzing whether newer constructions command higher prices.

9) Area Plot of Average Price by Construction Year

The area plot builds on the line plot by visually accumulating values, providing a clearer picture of pricing trends over time and highlighting any significant shifts or trends.

10) Line plot of Average Price by Review Rate Number

This plot tests the hypothesis that higher-rated listings can charge more, examining the relationship between guest satisfaction (indicated by review rates) and pricing.

11) Scatterplot of Price and Number of Reviews

Analyzing the relationship between price and the number of reviews a listing has. Can potentially indicate if a listing has many reviews will it impact prices.

12) Histogram of Price to Service Fee Ratio

This histogram assesses how much of the listing price is taken up by service fees, which is crucial for understanding the cost structure faced by guests.

13) Violin Plot of Price to Service Fee Ratio by Borough

Expanding on the histogram, the violin plot adds a layer by comparing these ratios across boroughs, offering insights into how service fees vary in relation to location.

14) Boxplot of Price Index by Borough

This plot evaluates the price index, which normalizes prices within each borough to assess relative affordability and market positioning.

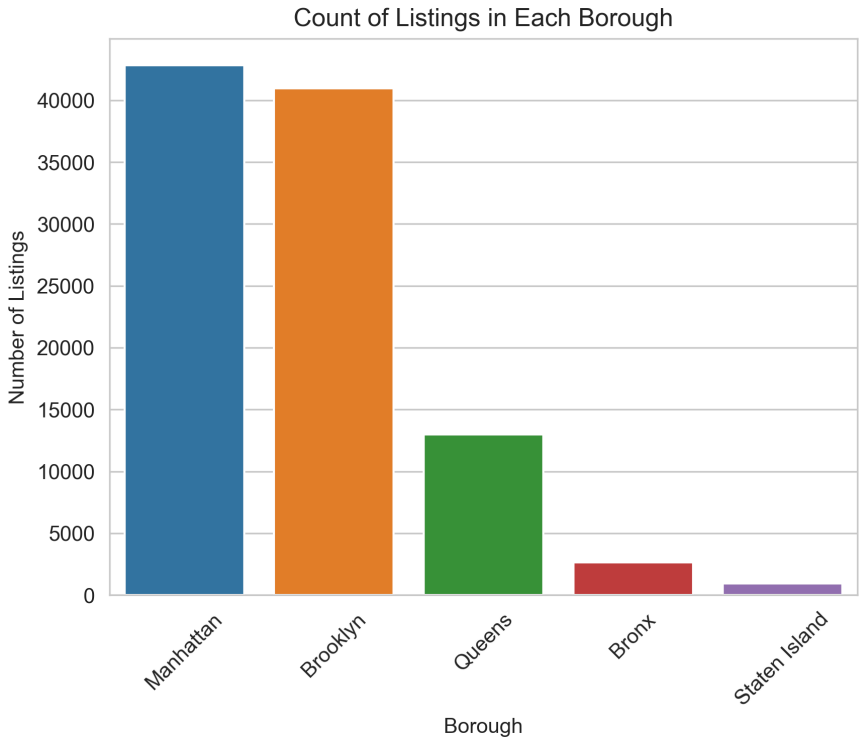
15) Correlation Matrix Heat Map

By visualizing the correlation coefficients among several numerical variables, this heat map identifies potential relationships and dependencies, informing more complex analyses.

16) Heat Map of Average Price Index by Borough and Room Type

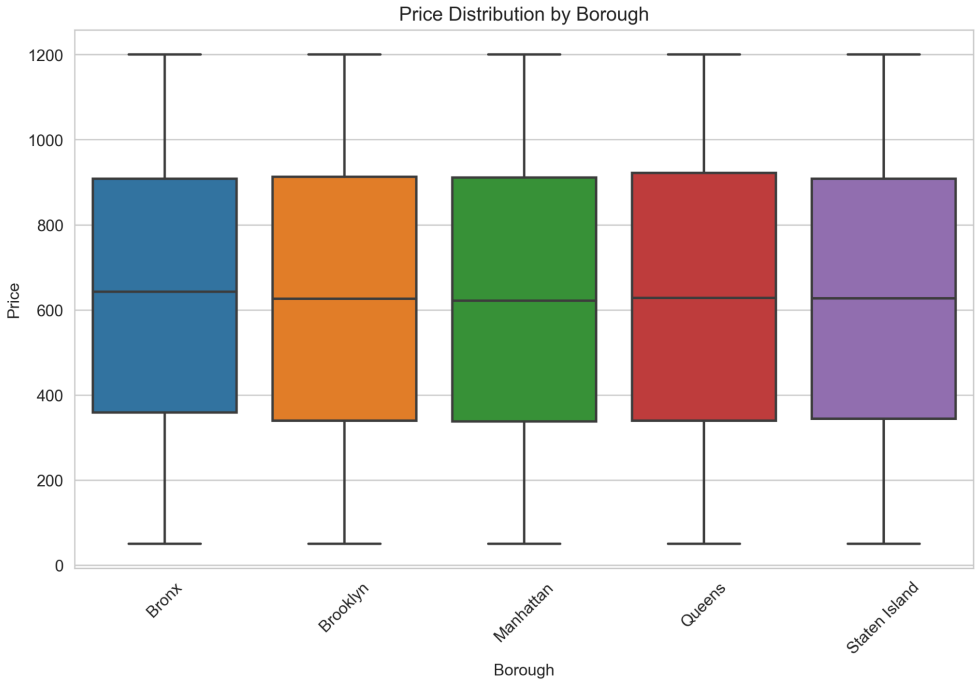
This heat map delves into the interaction effects between borough and room type on the price index, providing a multifaceted view of pricing influences.

Results:

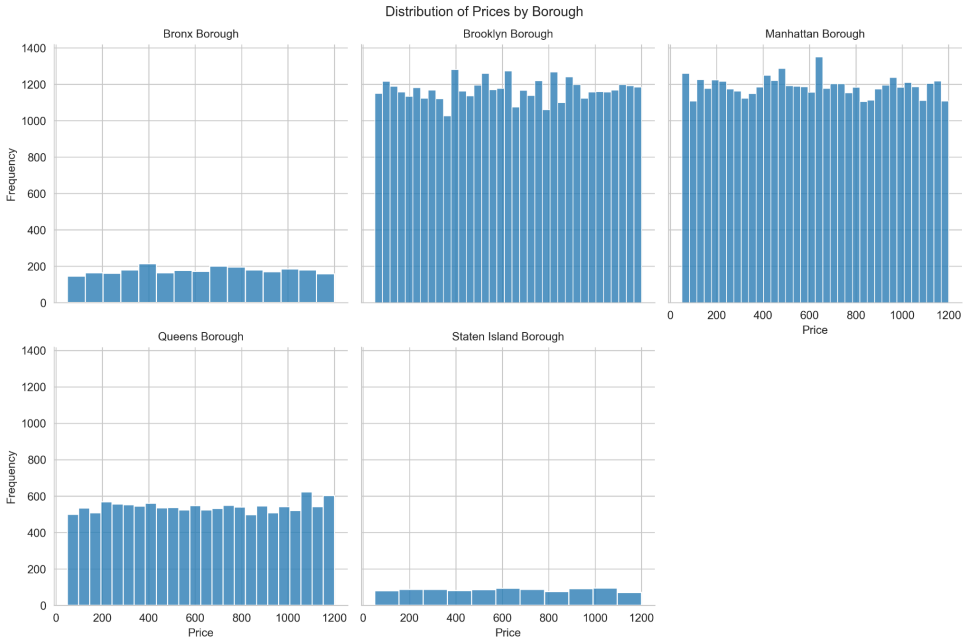


This bar plot indicates that Manhattan (42,813) and Brooklyn (40,946) have the highest number of Airbnb listings, suggesting a high demand or popularity in these areas. The relatively lower number of listings in Queens (12,992), Bronx (2650), and Staten Island (936) may indicate less tourist traffic or fewer Airbnb hosts, which could influence the average pricing in these less saturated markets.

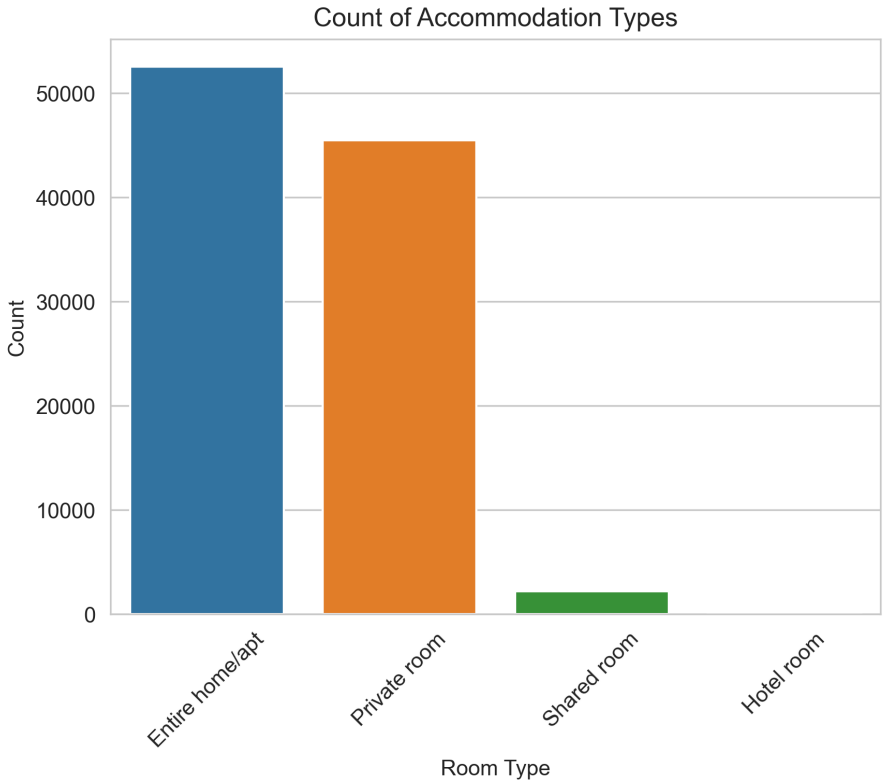
The overall average price per night is $625, yet the boxplot shows minimal variation in median prices between boroughs. Median prices in all boroughs are at or near the average of $625. This contradicts the hypothesis that location significantly impacts pricing, with central areas commanding higher rates.



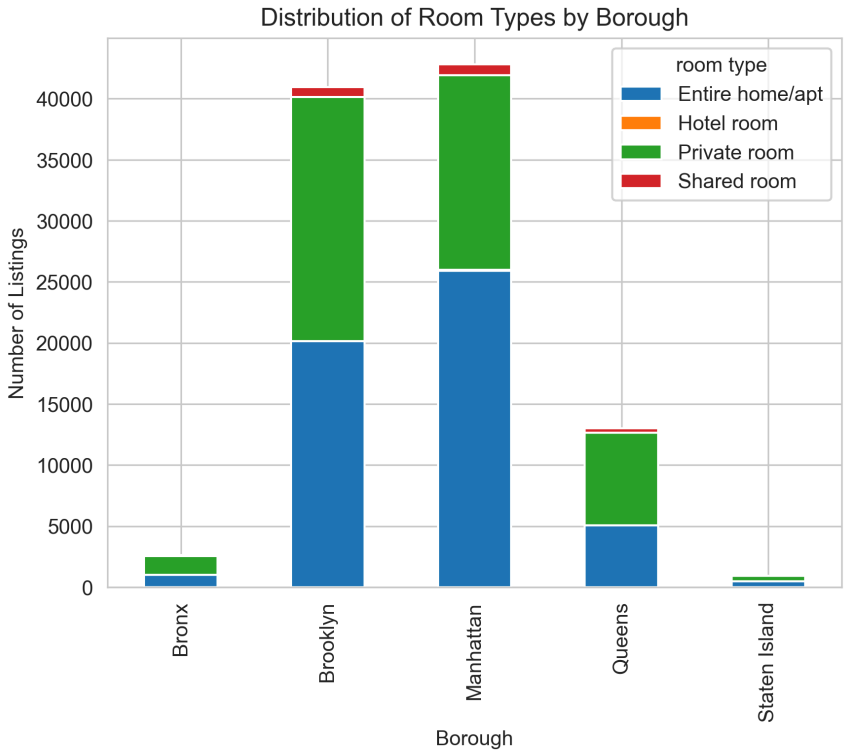
These histograms provide a deeper look into the price distribution



within each borough. Prices for the most part are distributed evenly in each borough and there is no concentration of listings that impacts the average price.

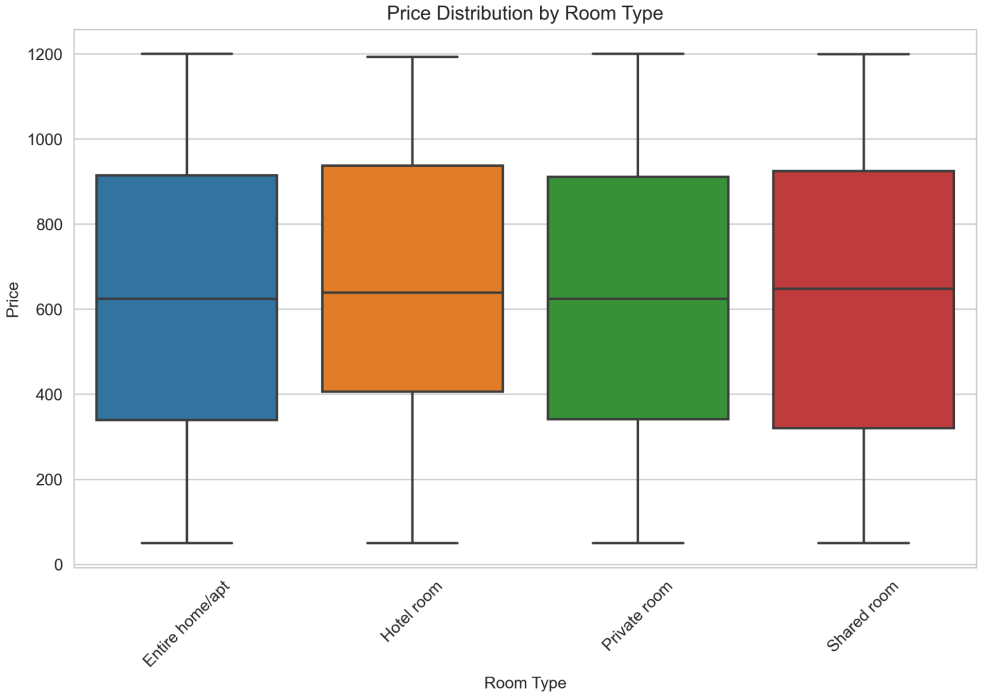


This bar chart reveals the prevalence of different types of Airbnb accommodations. Entire homes/apartments (52,551) and private rooms (45,498) dominate the listings, indicating a higher demand or availability for these types of accommodations. The relatively low count of hotel rooms (112) and shared rooms (2,176) may reflect niche markets or less demand.

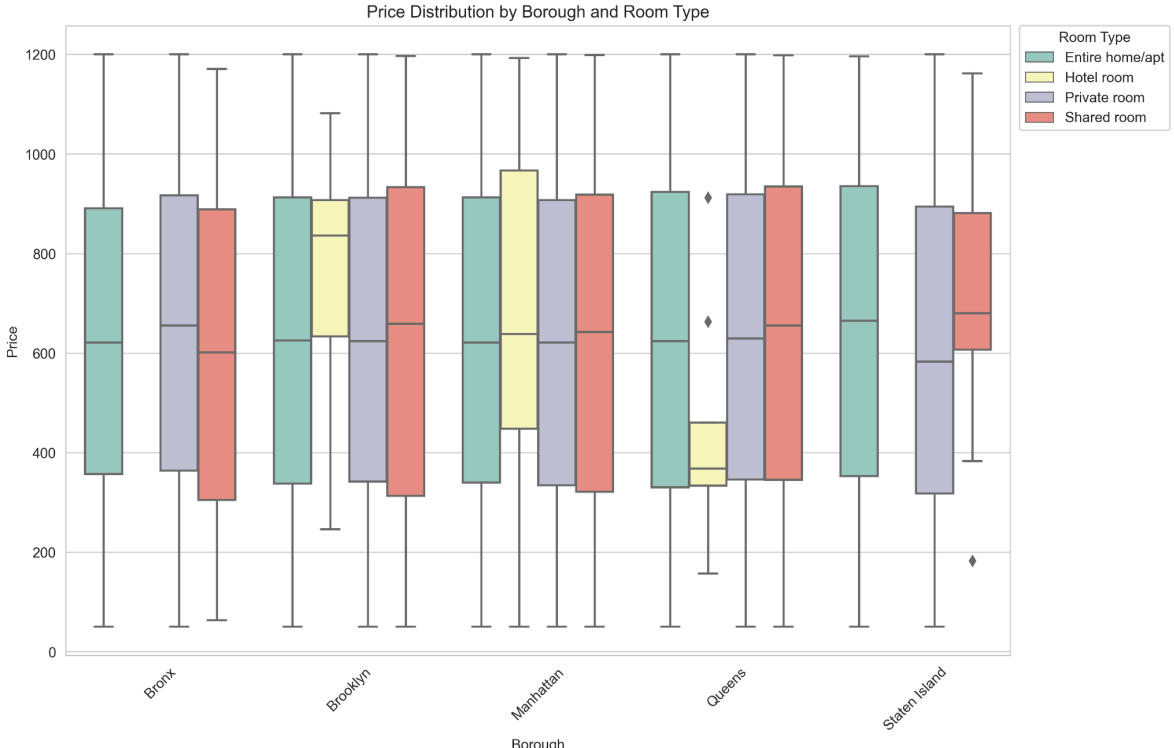


This stacked bar chart shows the composition of room types within each borough. It illustrates how each borough caters to different accommodation preferences, with entire homes/apartments and private rooms being more prevalent. This visualization underscores the diverse

accommodation landscape across boroughs.



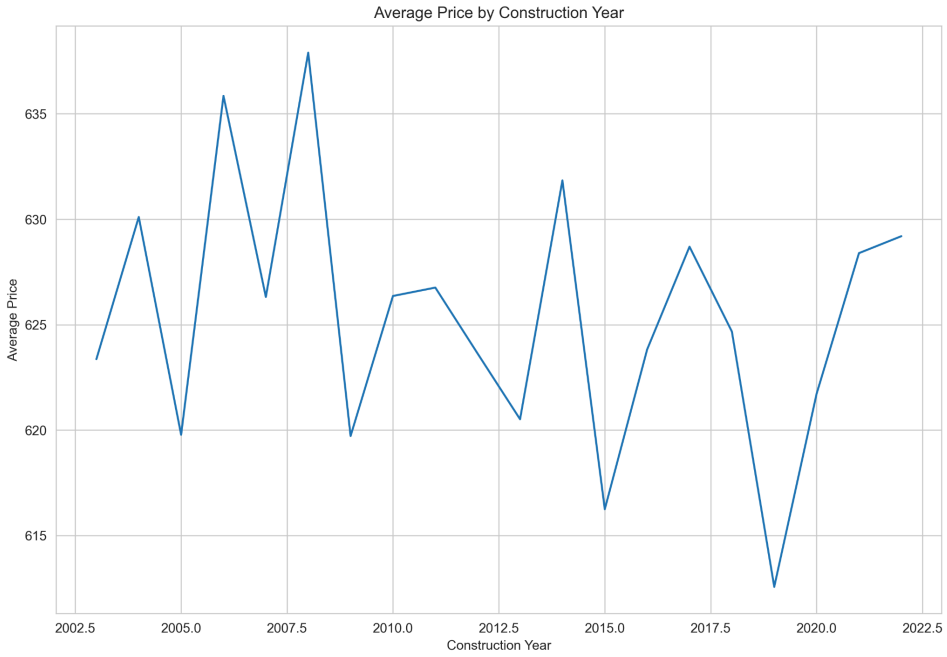
This boxplot compares the price ranges for different types of room accommodations. It shows minimal to no difference in median prices between room types. Again, each boxplot is at or near the average of $625 per night, so we can rule out that different accommodation types impact pricing.



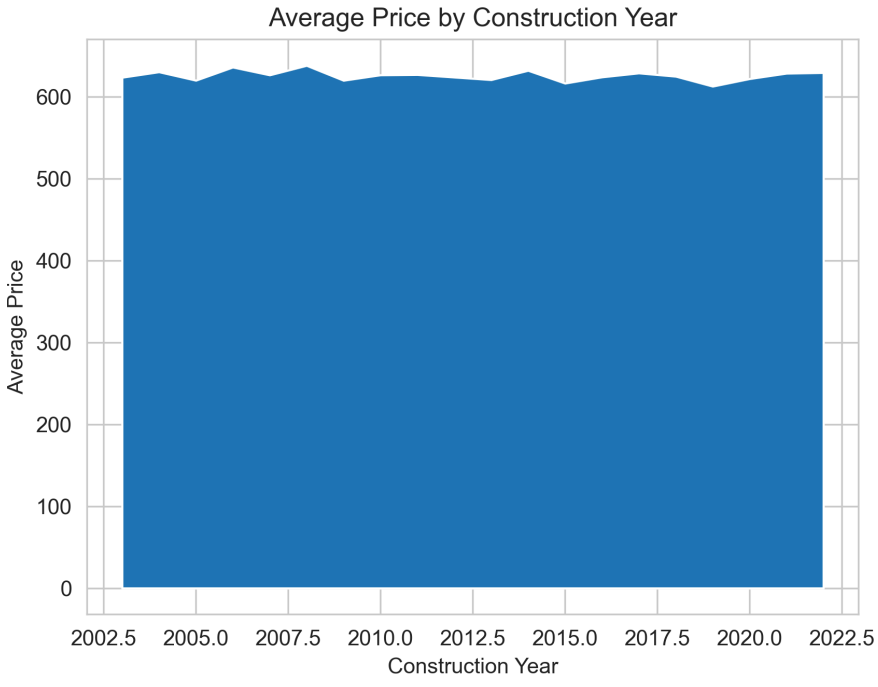
This boxplot displays the price distribution for different types of accommodations across boroughs. It shows variations in median prices

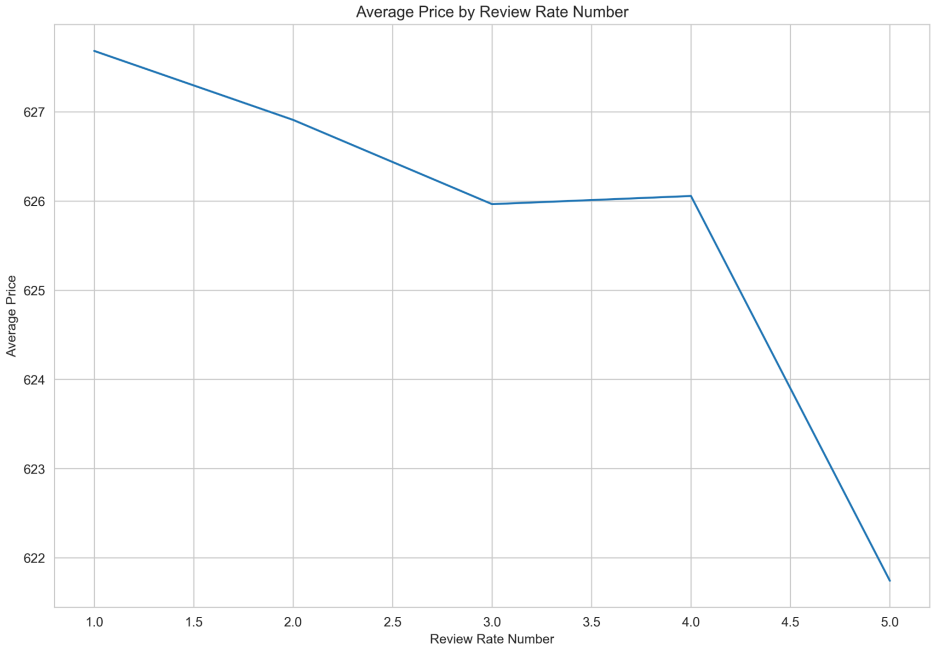
and the range of prices within each borough for different room types. Important to note that shared rooms and hotel rooms have significantly lower sample sizes compared to the other room types and this may impact our analysis (ex: hotel room).

The line plot of average price by construction year shows fluctuations over time but does not display a clear trend of newer properties being more expensive per night. Thus, we can rule out that construction year has any impact on pricing.

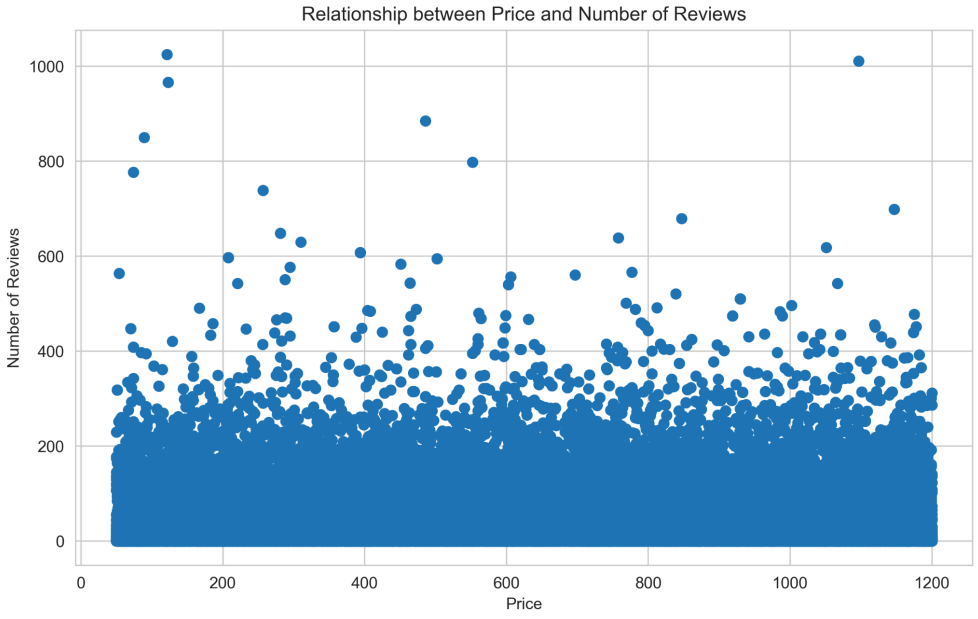


Similar to the line plot, this area plot emphasizes the fluctuations in pricing over different construction years without showing a definitive trend toward higher prices for newer or older listings.



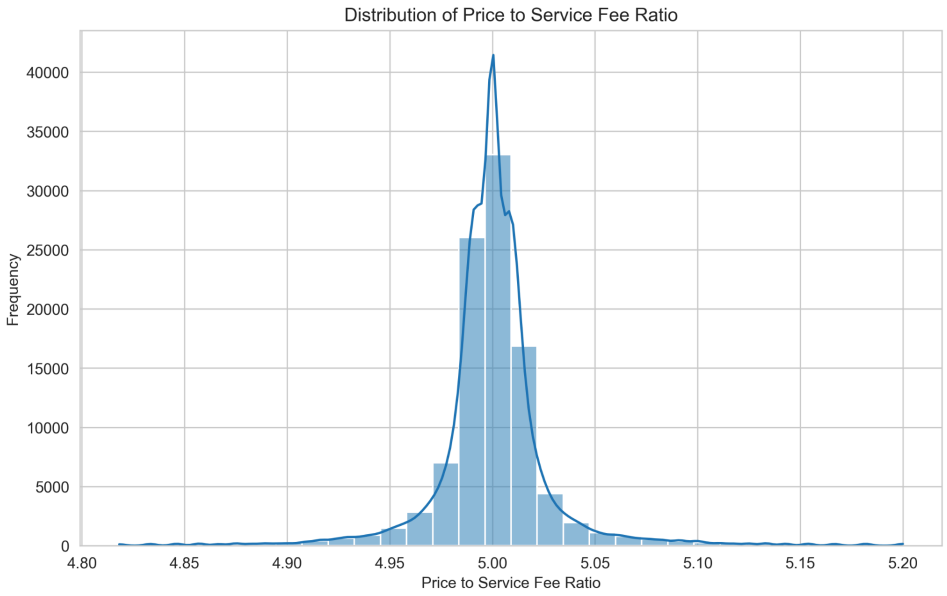


This line plot suggests that higher review rates (indicating higher guest satisfaction) do not correlate with higher prices. In fact, there's a slight decline at the highest review rate. To conclude, higher-rated listings aren't necessarily going to be more expansive.

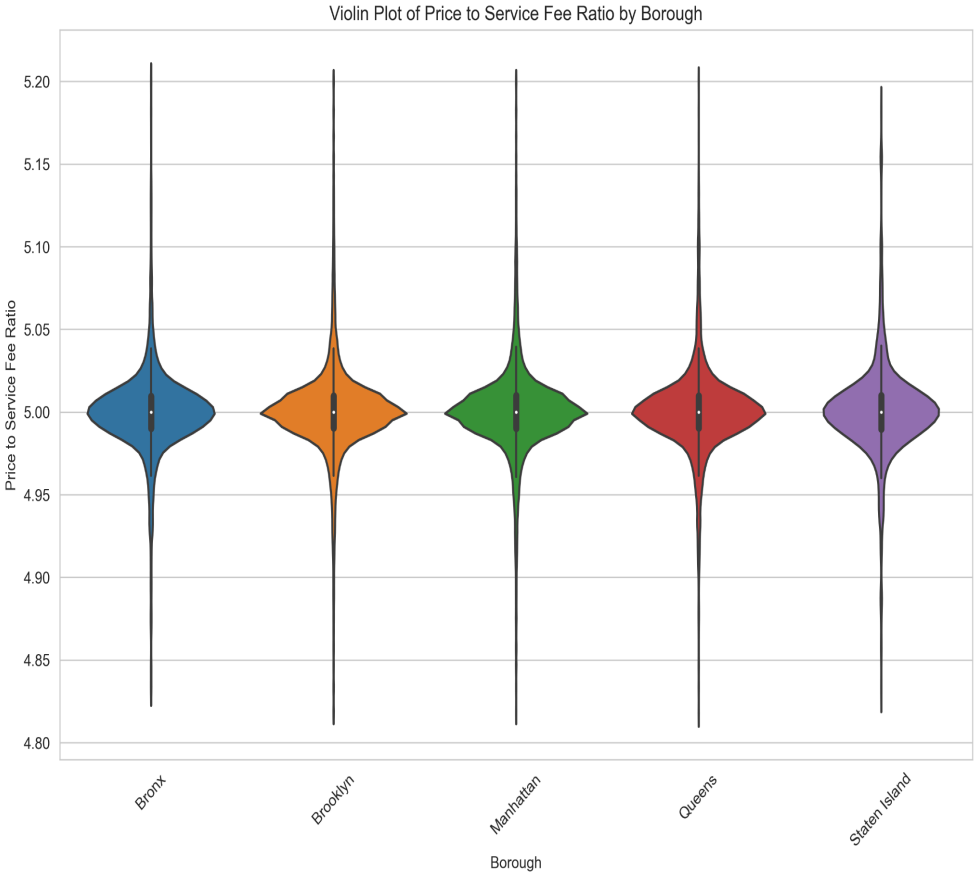


This scatterplot shows the relationship between the price of listings and the number of reviews they have received. The plot indicates that there is no strong correlation between higher prices and an increased number of reviews. Most listings, regardless of price,

tend to cluster around a lower to moderate number of reviews, suggesting that price is not a significant factor in the frequency or volume of reviews.



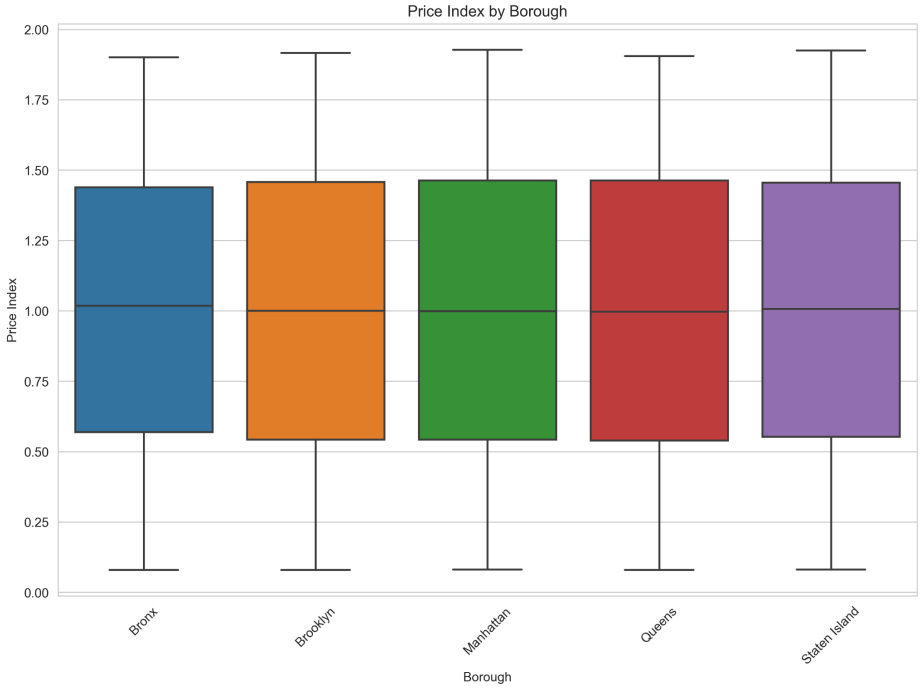
The histogram shows a bell-shaped distribution of the price-to-service fee ratio, the peak of the histogram is very sharp and close to 5, which suggests that for most listings, the service fee is roughly one-fifth (or 20%) of the total price. This suggests a standardized approach to how service fees are applied relative to listing prices across the dataset



The violin plot provides a deeper look at the distribution of the price-to-service fee ratio across boroughs. Each violin shows the density of the ratio, indicating variability and skewness. This plot helps identify if any boroughs have outliers or a wider

range of fees. We can see that it is practically the same in each borough.

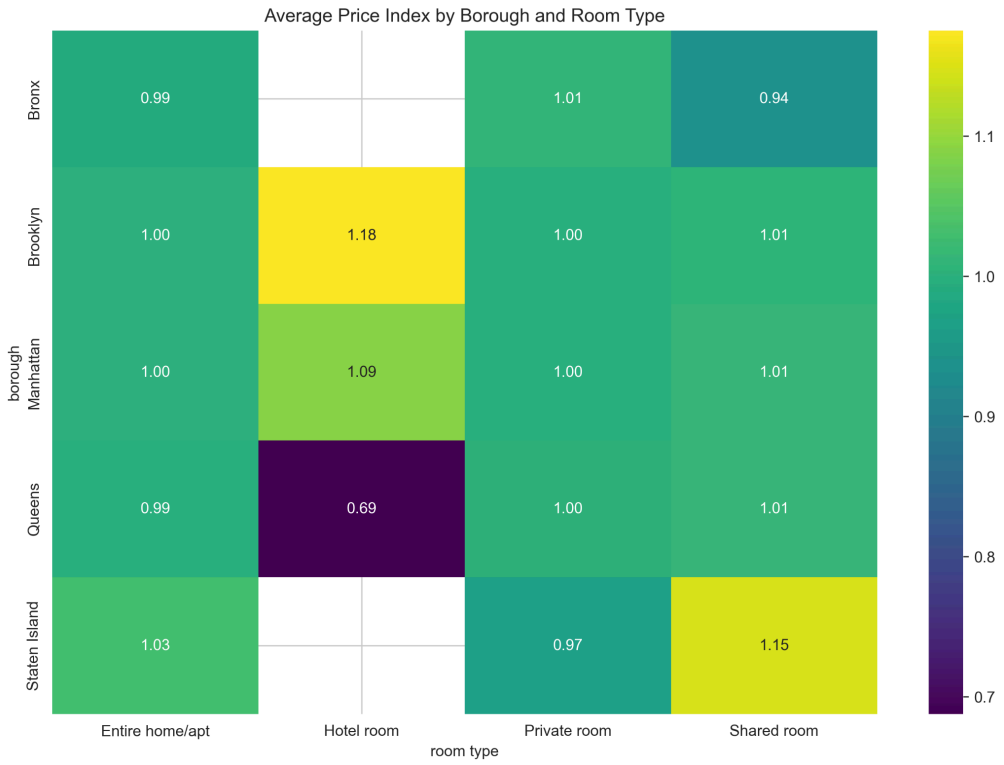
This histogram indicates that most listings have a price index close to 1, which implies that most prices are close to the borough average. The fairly normal distribution suggests there is not a significant deviation from the mean, pointing to a consistent pricing strategy across listings.



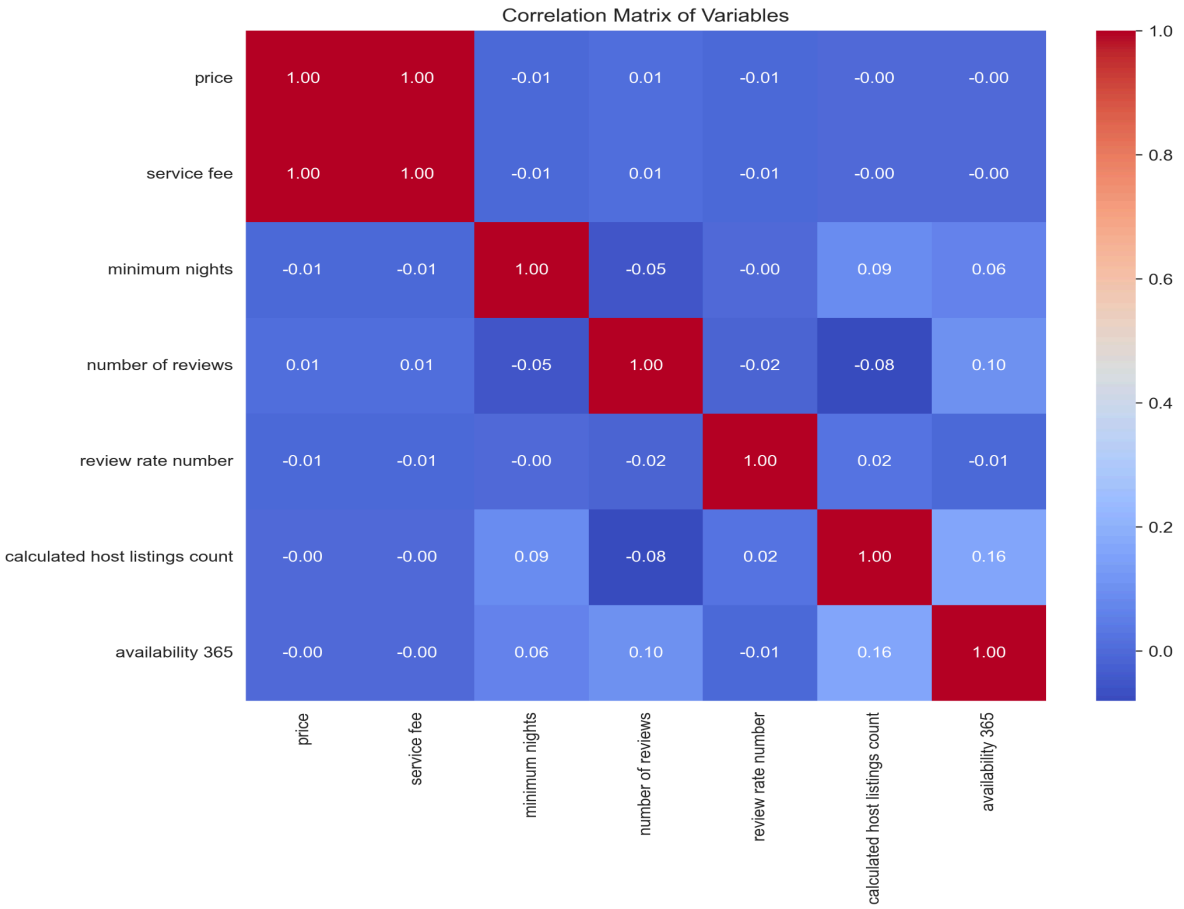
The price index boxplot across boroughs reveals that despite variations in actual prices, the relative affordability index is fairly consistent, with all boroughs hovering around the median index of 1. This suggests that while

absolute prices vary, the relative positioning within each borough is stable.

This heat map illustrates the average price index by borough and room type, showing how different combinations of location and accommodation type can affect pricing relative to the borough average. For example, hotel rooms in Manhattan are priced close to the borough's average, whereas shared rooms in Staten Island are significantly above the borough average. (again, important to note that hotel sample size is extremely small and there are no hotel rooms in our dataset in the Bronx and Staten Island)



The correlation matrix heat map shows a strong positive correlation between price and service fee, confirming that these two variables iare tied together together. The minimal correlation



between price and other factors like minimum nights or number of reviews suggests that these factors do not significantly influence pricing.

Conclusion:

In my analysis of the New York City Airbnb market, I found that, contrary to my assumptions, the borough alone does not strongly influence Airbnb pricing. Most listings, regardless of their location, tend to average around $625 per night. This indicates a surprising uniformity in pricing across a city known for its economic and geographic diversity. Additionally, factors such as room type, review rate number, and a few more factors did not substantially diverge from this average. The dataset did not yield the result I expected/predicted, yet I've learned so much about the short-term rental market and how it operates. Nevertheless, in a broader sense, my findings are important to NYC residents.

These findings are significant, especially considering New York City's desperate need for more long-term housing options amidst historically low vacancy rates. The data suggests that landlords currently operating Airbnb rentals have little financial incentive to convert their properties into long-term rental units, as the potential revenue from short-term rentals is lucrative across all boroughs. This situation exacerbates the housing shortage, making it difficult for residents to find affordable long-term accommodations.

Work Cited

"New York City's Vacancy Rate Reaches Historic Low of 1.4 Percent, Demanding Urgent Action." *The Official Website of the City of New York*, 8 Feb. 2024, www.nyc.gov/site/hpd/news/007-24/new-york-city-s-vacancy-rate-reaches-historic-low-1-4-percent-demanding-urgent-action-new#/0.

"Example Gallery#." *Example Gallery - Seaborn 0.13.2 Documentation*, seaborn.pydata.org/examples/index.html. Accessed 6 May 2024.

"How Do I Select a Subset of a DataFrame?#." *How Do I Select a Subset of a DataFrame? - Pandas 2.2.2 Documentation*, pandas.pydata.org/docs/getting_started/intro_tutorials/03_subset_data.html. Accessed 6 May 2024.