Machine Learning Module

Development Team Project

Group 1

**October 30, 2023**

Comprehensive Analysis of Airbnb Listing Dynamics in New York City

Table of Contents

[Table of Figures 3](#_Toc164602965)

[Introduction 4](#_Toc164602966)

[Business Analytic Question 4](#_Toc164602967)

[Methodology 5](#_Toc164602968)

[Data Preprocessing 5](#_Toc164602969)

[Data Analysis 7](#_Toc164602970)

[Discussion and Recommendation 13](#_Toc164602971)

[References 14](#_Toc164602972)

Table of Figures

[Figure 1: Dataset Attributes 5](#_Toc164684671)

[Figure 2: Plot for missing values. 6](#_Toc164684672)

[Figure 3: Density Plots and Histograms for Numerical Features 7](#_Toc164684673)

[Figure 4: Box Plots for Numerical Features 8](#_Toc164684674)

[Figure 5: Distribution of Listings across New York 10](#_Toc164684675)

[Figure 6: Scatter Plot showing the distribution of listing across New York City 10](#_Toc164684676)

[Figure 7: KDE plots from price and minimum nights across different neighbourhoods 11](#_Toc164684677)

[Figure 8: KDE plots across different neighbourhoods 11](#_Toc164684678)

[Figure 9: Bar Charts 1.1 12](#_Toc164684679)

[Figure 10: Bar Charts 1.1 13](#_Toc164684680)

Introduction

Since its inception sixteen years ago, Airbnb has signed up more than five million landlords, and more than a billion and a half stays have been arranged in this way, covering most countries (Airbnb Newsroom, 2024). In New York City, one of Airbnb's most vibrant markets, the variety and complexity of listings demand a thorough examination to understand how different factors such as location, price, and room type interact and influence guest choices and host earnings. By dissecting the dataset, this analysis aims to uncover underlying patterns that can inform strategic decisions and enhance both guest satisfaction and host profitability in the evolving market landscape.

Business Analytic Question

The central question guiding this analysis is: "How do various factors such as neighbourhood, room type, and pricing interact to affect the dynamics of Airbnb listings in New York City?" This seeks to dissect the multifaceted relationship between geographic and economic variables and their collective impact on the attractiveness and performance of listings. The insights from this analysis can be used to recommend prices for hosts, propose deals for guests, allocate advertising resources to different localities, and financially forecast based on activity in different areas.

Methodology

Data Preprocessing

The analysis utilises the AB\_NYC\_2019.csv dataset available on Kaggle, which contains detailed information about Airbnb listings in New York City for the year 2019 (Kaggle, 2019). This dataset includes essential attributes such as geographical coordinates, prices, room types, availability as shown in the figure below:

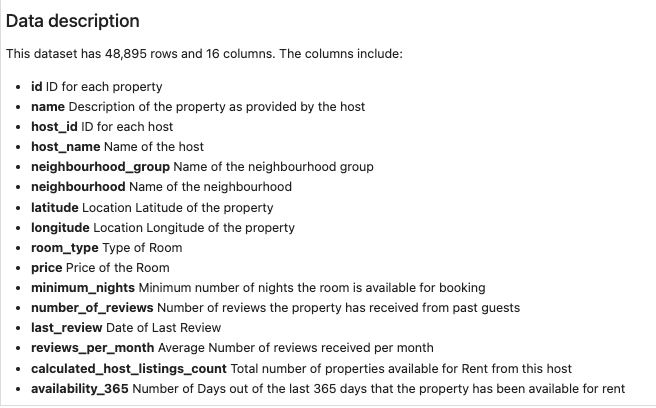


Figure 1: Dataset Attributes

Prior to analysis the dataset underwent preprocessing which included the dropping of irrelevant columns such as such as ‘host \_name’ and ‘last\_review’ and handling of missing values. Missing values within the dataset were handled by, either imputation or removal, depending on their impact on overall dataset integrity.  Figure 2 below shows the missing values present in ‘review\_per\_month’ and ‘name’. The missing values in the 'name' column were removed, and those in the 'reviews\_per\_month' column were imputed with 0, due to the clear relationship between the number of reviews column being 0 and missing values in this column.

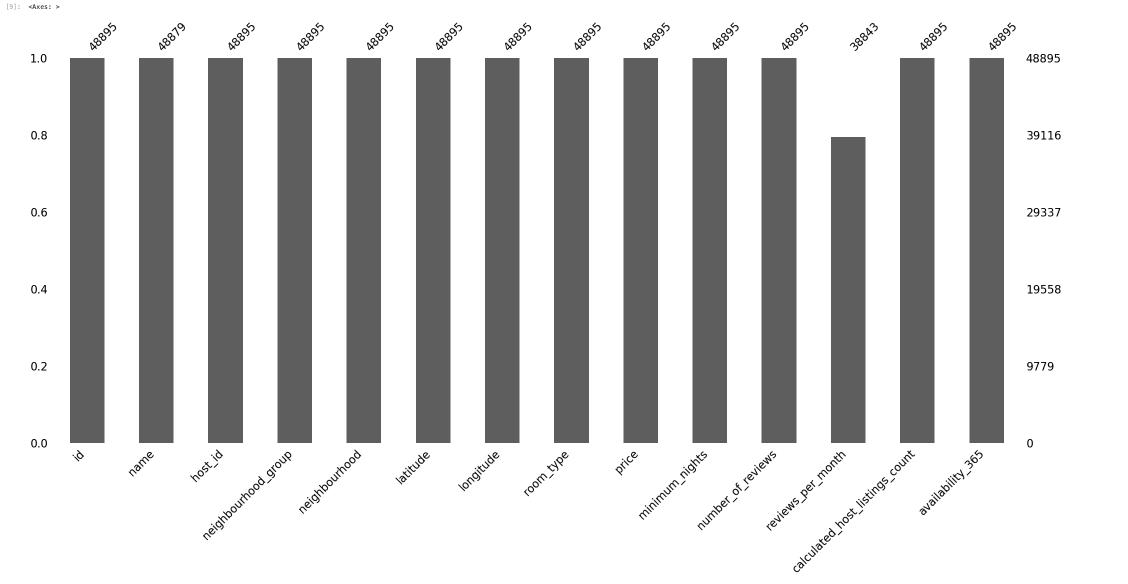


Figure 2: Plot for missing values.

Data Analysis

The distribution of the following features price, minimum nights, number of reviews, reviews per month, calculated host listings count, and availability 365 was investigated using density histograms and box plots to identify median trends and detect outliers as shown in figures 3 and 4 below:

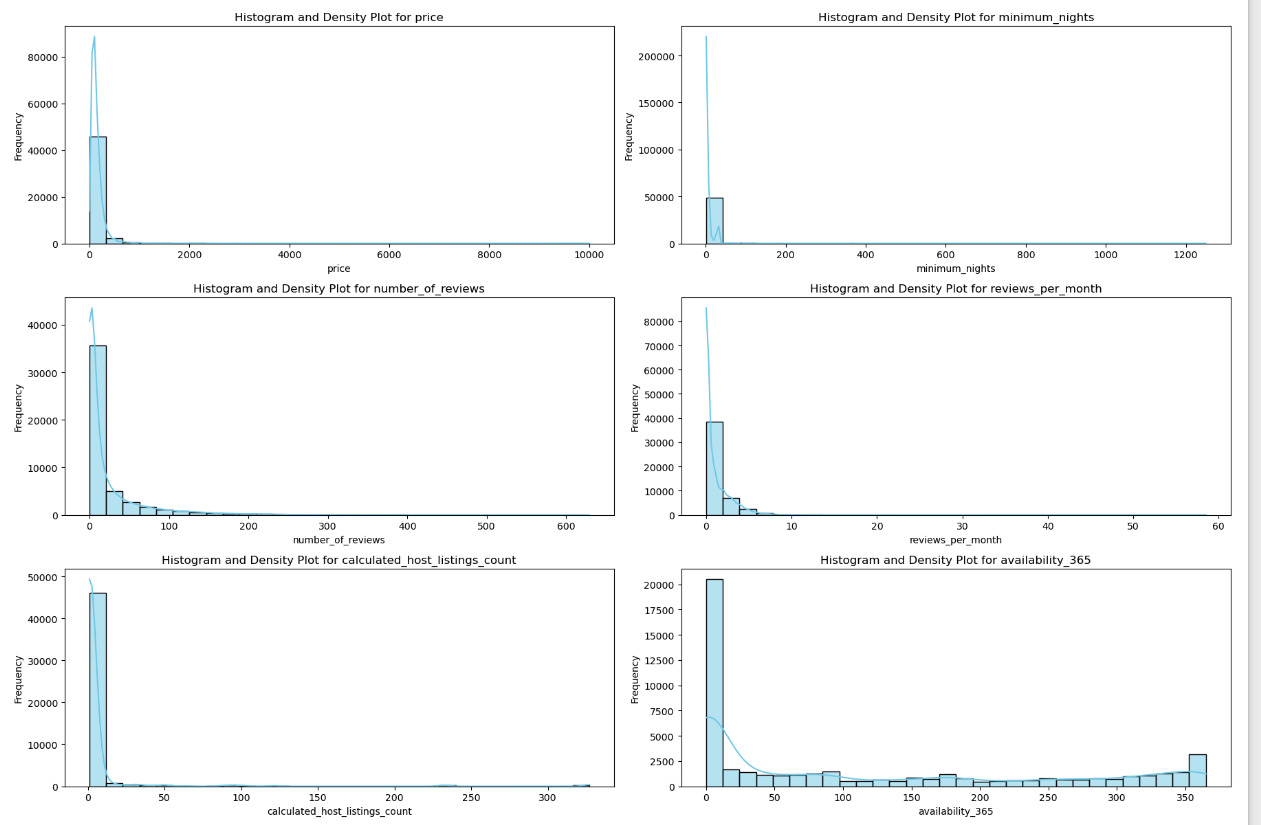


Figure 3: Density Plots and Histograms for Numerical Features

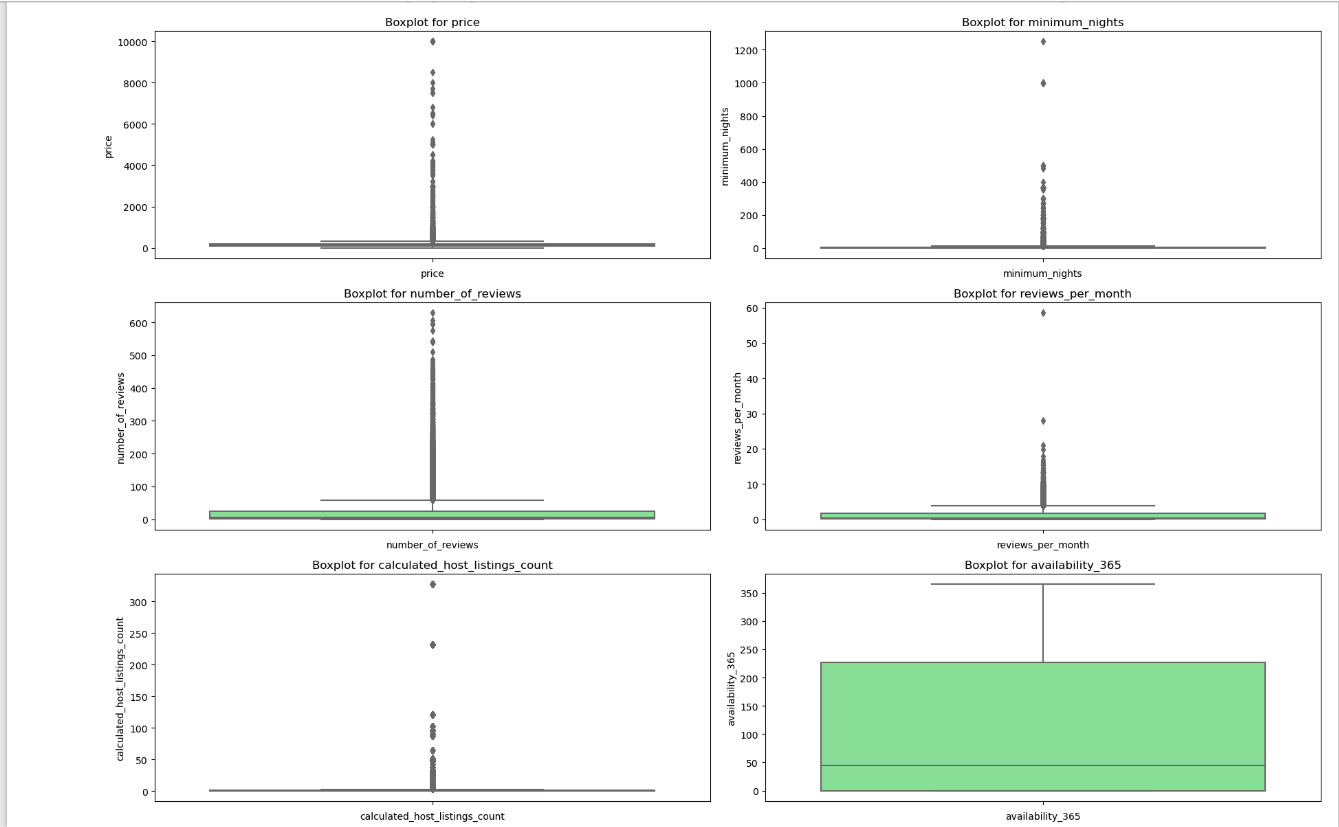


Figure 4: Box Plots for Numerical Features

Density histograms revealed that most listings are priced at the lower end of the market, indicating a concentration of affordable options. Similarly, most listings require only a few minimum nights’ stay, suggesting a market geared towards short-term accommodations. The reviews-related features generally showed a skew towards fewer reviews, with many listings having no reviews at all. Additionally, the 'calculated host listings count' histogram showed that many hosts manage just one or a small collection of properties, illustrating the platform's foundational role in supporting individual and small-scale hosts. Nonetheless, the presence of some hosts with multiple listings points towards professional or semi-professional hosts who may be leveraging Airbnb for broader property management purposes.

The 'availability 365' feature, which denotes the number of days a listing is available throughout the year, presented a wide-ranging distribution. Many properties showed availability for fewer days in the year, potentially reflecting the hosts' preference to rent out their properties only part-time or seasonally. On the other hand, a substantial number of listings were available most days of the year, which may correspond to dedicated rental properties aimed at consistent income generation through Airbnb.

The box plots complemented these findings by highlighting the median values for the features, and pinpointing outliers, particularly for listings with higher prices and longer minimum stays, potentially indicative of premium accommodations. These findings reveal a diverse Airbnb host community in New York City, ranging from those offering private spaces for extra income to those managing several listings as a business. The variation in availability also indicates differing host strategies or constraints, from those who offer their spaces as a full-time business to those who do so casually or seasonally. This diversity in hosting patterns has implications for market competitiveness and customer choice within the Airbnb ecosystem in New York City.

Delving deeper into the geographic distribution, as displayed in the bar chart in Figure 5, It was noted that Manhattan and Brooklyn boast the highest concentration of listings, echoing their appeal and high demand. In contrast, the Bronx and Staten Island exhibit fewer listings, which could correspond to lesser tourism or emerging Airbnb markets. Figure 6's scatter plot further clarifies this geographic delineation, with a dense accumulation of listings in Manhattan highlighting it as a bustling epicenter of Airbnb activity, while Brooklyn enjoys a substantial spread of listings. The sparser distribution in Queens, Staten Island, and the Bronx implies a less condensed market presence.

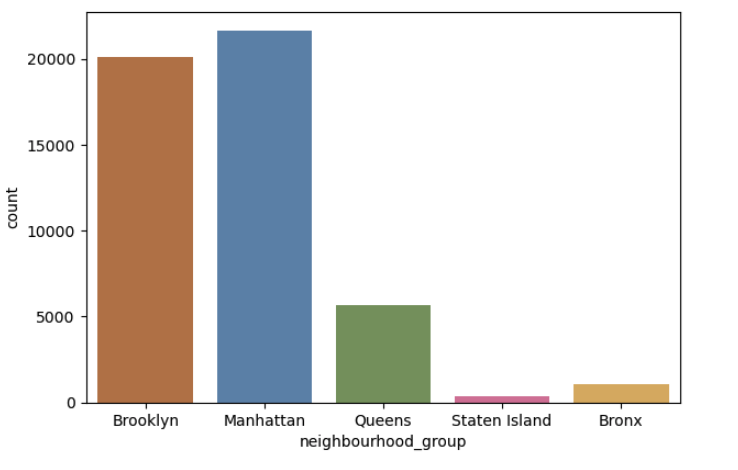


Figure 5: Distribution of Listings across New York

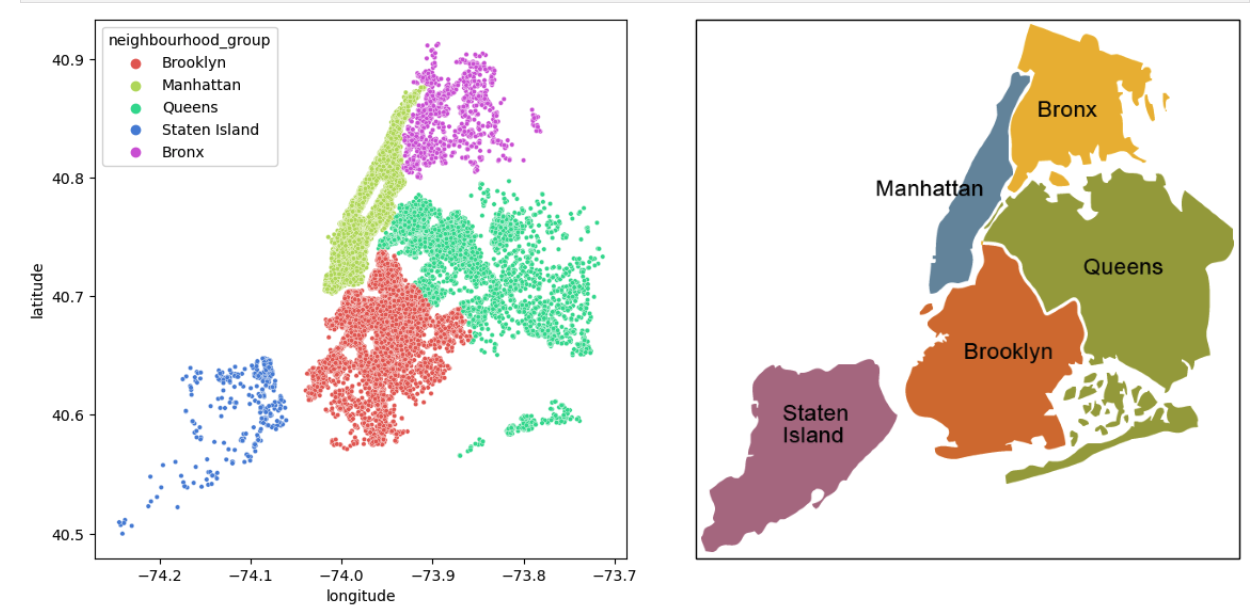


Figure 6: Scatter Plot showing the distribution of listing across New York City

The kernel density estimation (KDE) plots in figures 7 and 8 below reveal NYC Airbnb listings skew towards affordability, especially in Brooklyn and Queens, while Manhattan's diverse offerings span from budget to luxury, with a hint of premium properties. Short-term stays dominate, reflecting a market catering to tourists and brief visits. Review distributions suggest higher turnover in Manhattan and Brooklyn, with Staten Island and the Bronx possibly featuring newer or less frequented listings. Most hosts have few properties, yet Manhattan has more multi-listing hosts, indicating professional operations. Availability trends point to a mix of strategies, from year-round rentals to occasional hosting, showcasing the market's varied dynamics.

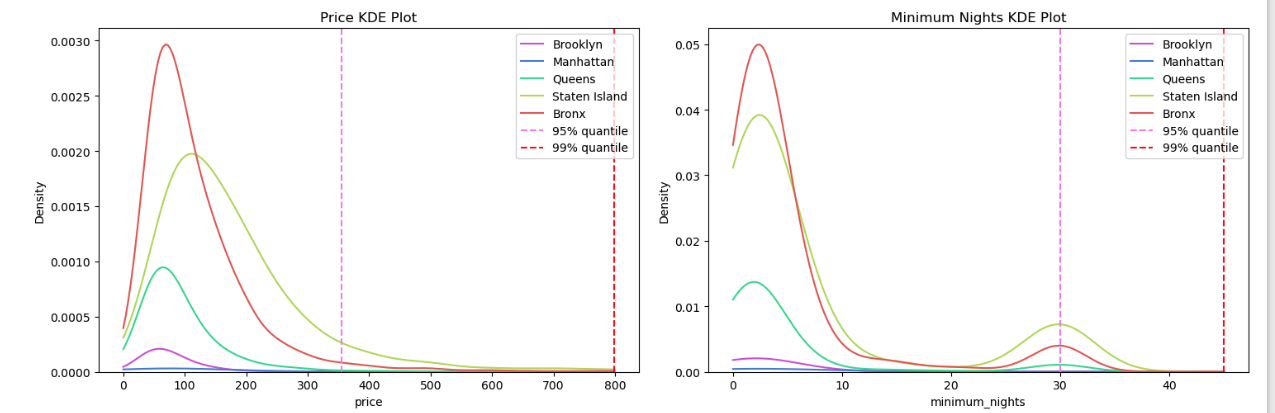


Figure 7: KDE plots from price and minimum nights across different neighbourhoods

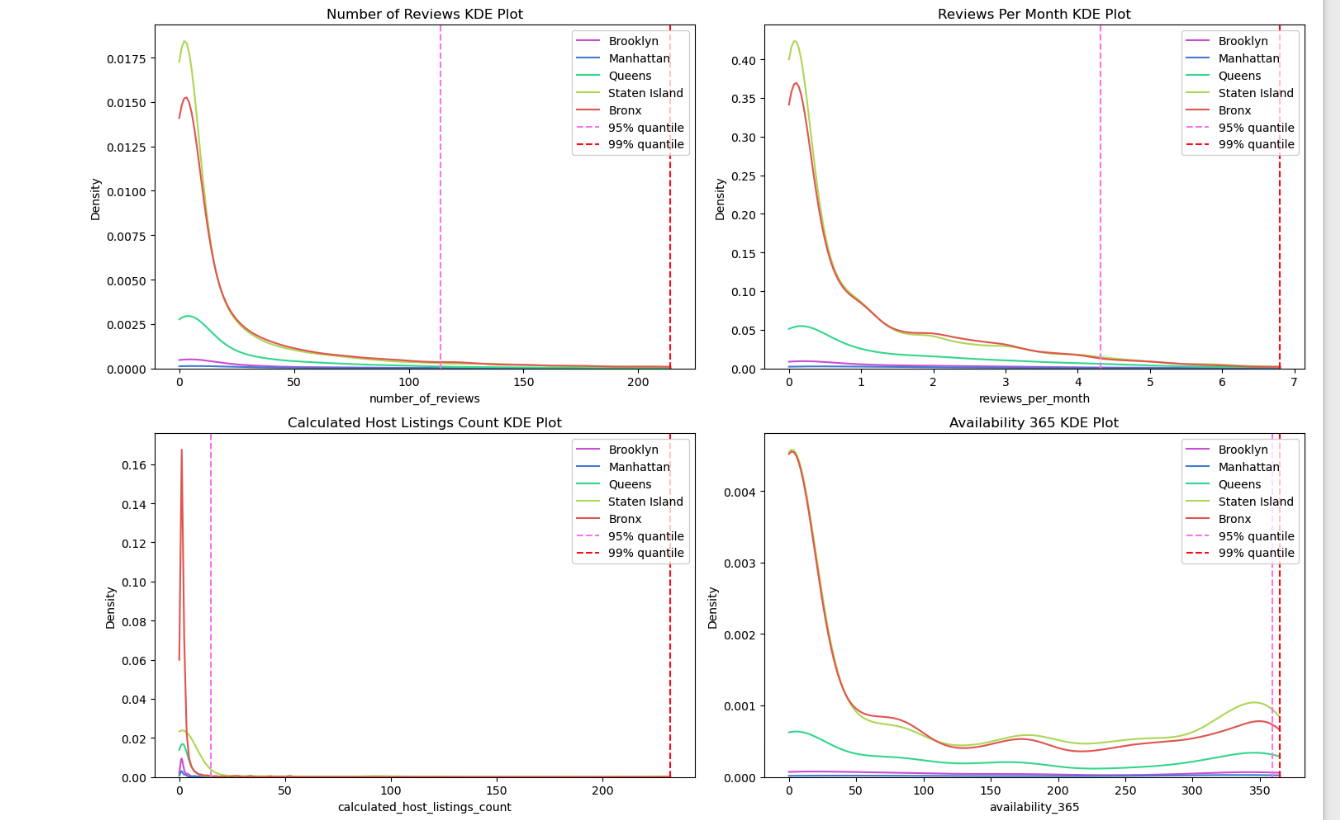


Figure 8: KDE plots across different neighbourhoods

Lastly, the set of bar charts in figures 9 and 10 reveal nuanced view of Airbnb activity in New York City by neighborhood and room type. The 'Price' and 'Minimum Nights' charts reveal price segmentation and stay requirements, with Manhattan showing the highest average prices across all room types, consistent with its status as a prime location. In contrast, Staten Island and the Bronx show lower prices, which could reflect location preferences or lesser demand. The reviews charts show higher engagement with listings in more frequented areas like Manhattan and Brooklyn, possibly due to a higher turnover of guests. This is further confirmed by the 'Calculated Host Listings Count' chart, where Manhattan stands out with a higher average number of listings per host, indicating the presence of more experienced or professional hosts in this borough. The 'Availability 365' chart indicates that listings in Manhattan and Staten Island have higher availability throughout the year, while other boroughs show varied availability, which may correlate with the hosts' strategies or external factors affecting rental patterns.

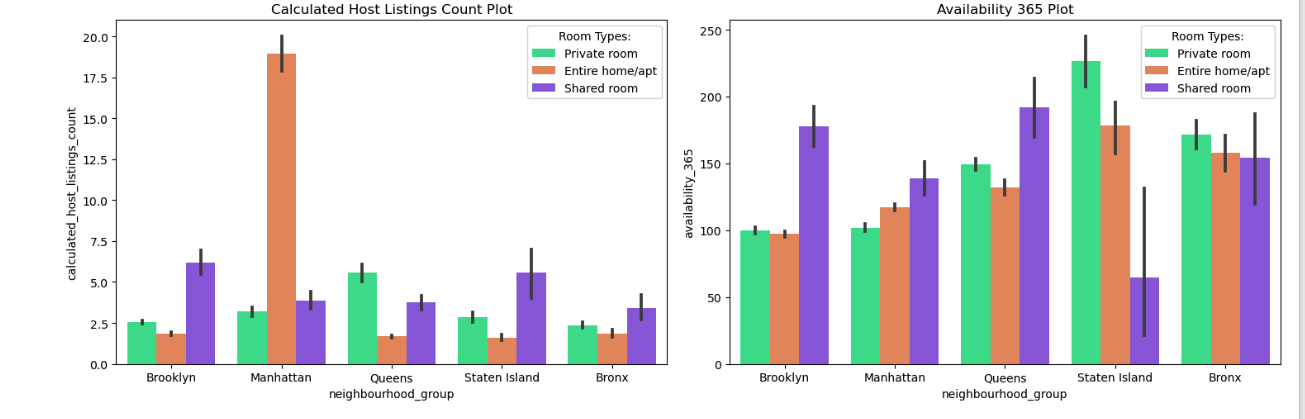


Figure 9: Bar Charts 1.1

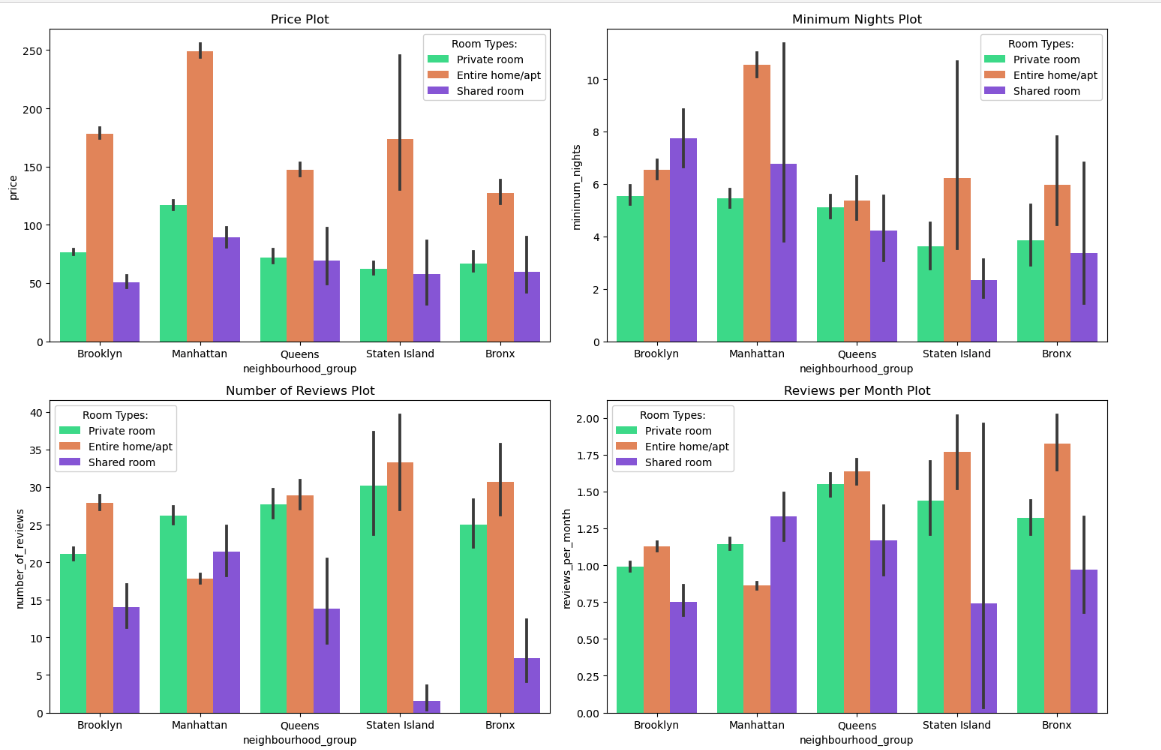


Figure 10: Bar Charts 1.1

Discussion and Recommendation

The analysis of Airbnb listings in New York City reveals a market brimming with diverse offerings and host engagement. Brooklyn and Queens cater to tourists through affordable and short-term accommodations, while Manhattan offers a spectrum from budget-friendly to premium, with a tendency toward professional hosting. Growth opportunities abound in the Bronx and Staten Island, where the market is less saturated, suggesting potential for new hosts. Year-round availability in these areas also indicates a possibility for a stable income for hosts and consistent choices for guests. The data and study insights point to the effectiveness of marketing strategies that highlight unique listing qualities, cleanliness, and value. This approach, alongside a diversified property portfolio, could enhance bookings and user experiences, tapping into the Bronx and Staten Island's emerging markets, and refining the Airbnb experience citywide.

References

Airbnb Newsroom. (2024). *About us - Airbnb Newsroom*. Available at: <https://news.airbnb.com/about-us/> [Accessed 9 April 2024]

Kaggle. (2019). *New York City Airbnb Open Data*. Available at: <https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data> [Accessed 7 April 2024]

Appendix: Python Code link

<https://github.com/Ngugi-Joy-Grace/airbnb-business-analysis/blob/main/analysis.ipynb>