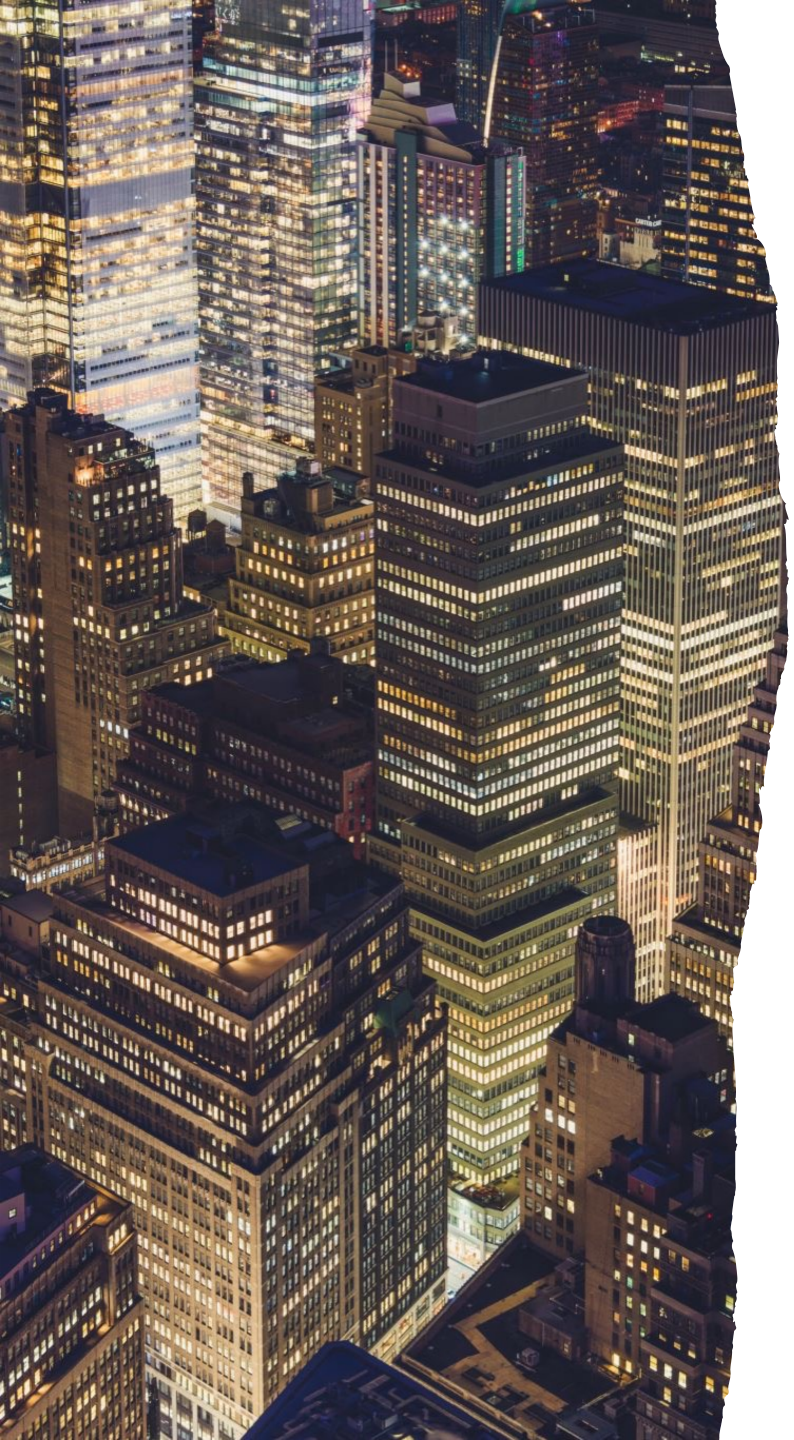
Exploring New York City Airbnb Data: Insights and Opportunities

Introduction

New York City is a vibrant metropolis with a thriving Airbnb market, offering a diverse range of listings and unique experiences for both hosts and guests. In this presentation, I will analyze the New York City Airbnb dataset, uncover valuable insights, and explore the exciting opportunities that arise from this wealth of information.

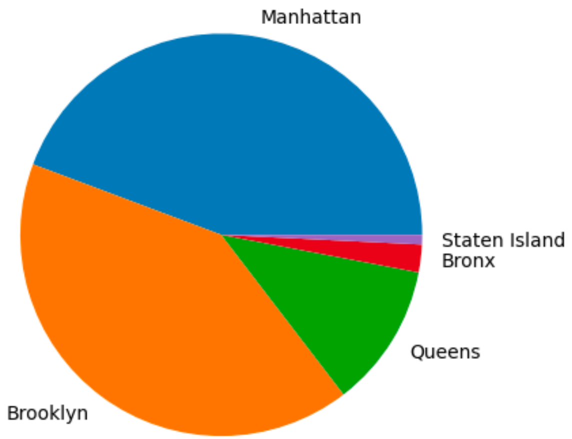


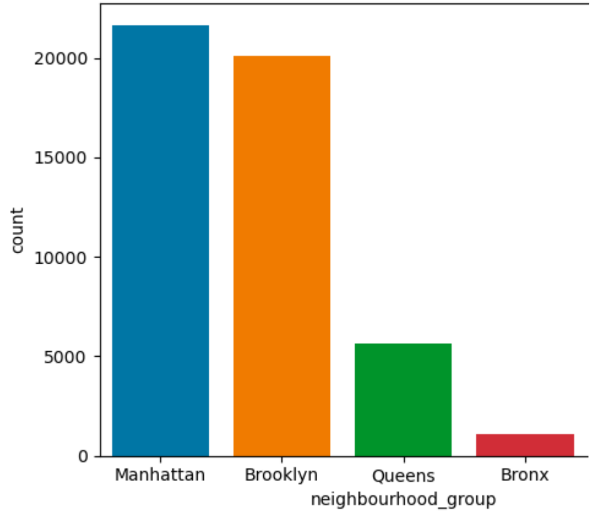
Objective

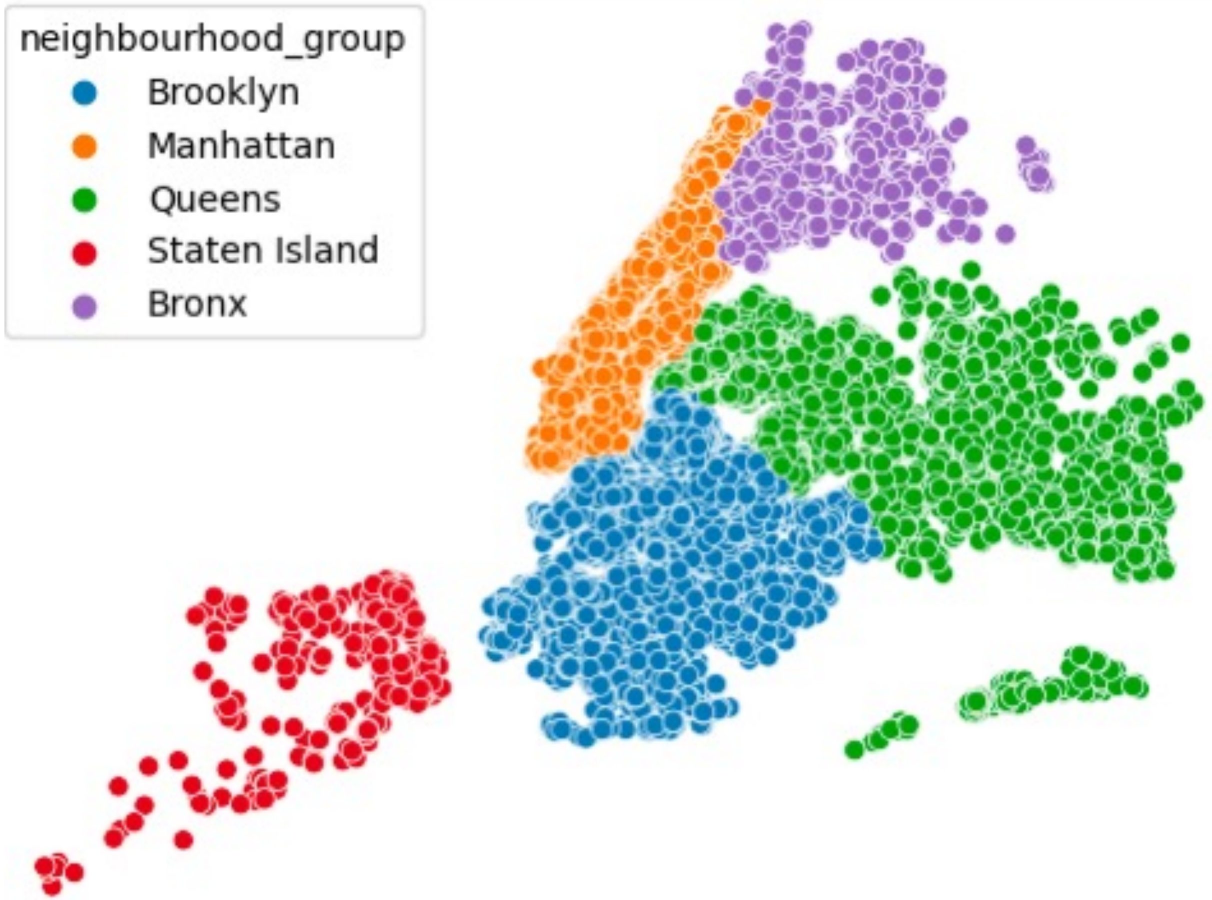
- To analyze the New York City Airbnb dataset and uncover key insights.
- To explore trends, patterns, and factors influencing prices and demand.
- To provide actionable recommendations for hosts and guests in the Airbnb market.

Dataset Overview

- The New York City Airbnb dataset contains information about Airbnb listings in the vibrant city of New York.
- It comprises 48895 records, each representing a unique listing.
- The dataset includes various features such as price, neighborhood, room type, availability, and more.
- The data was sourced from Kaggle.

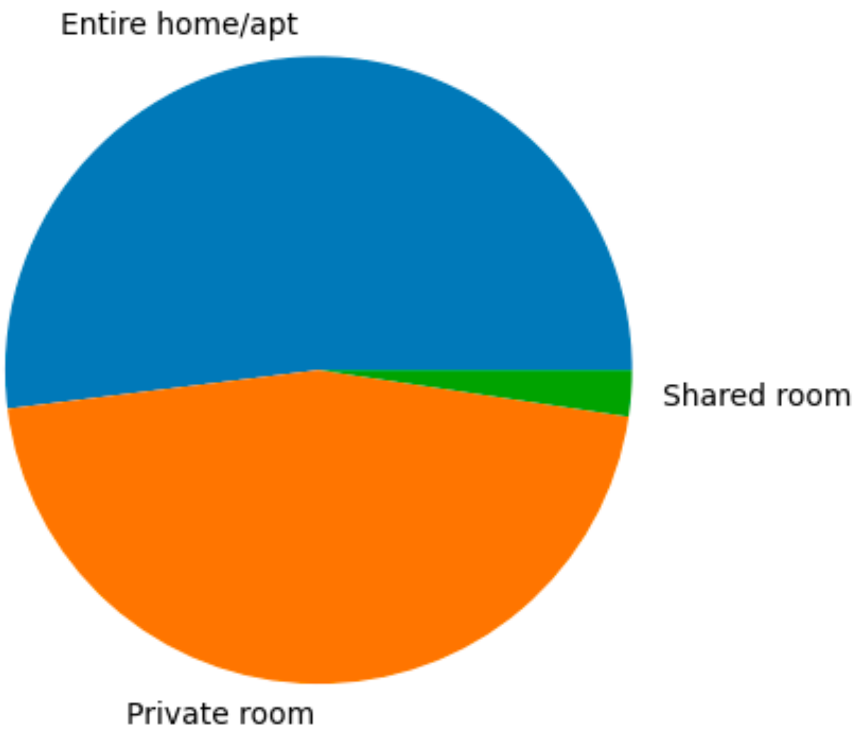


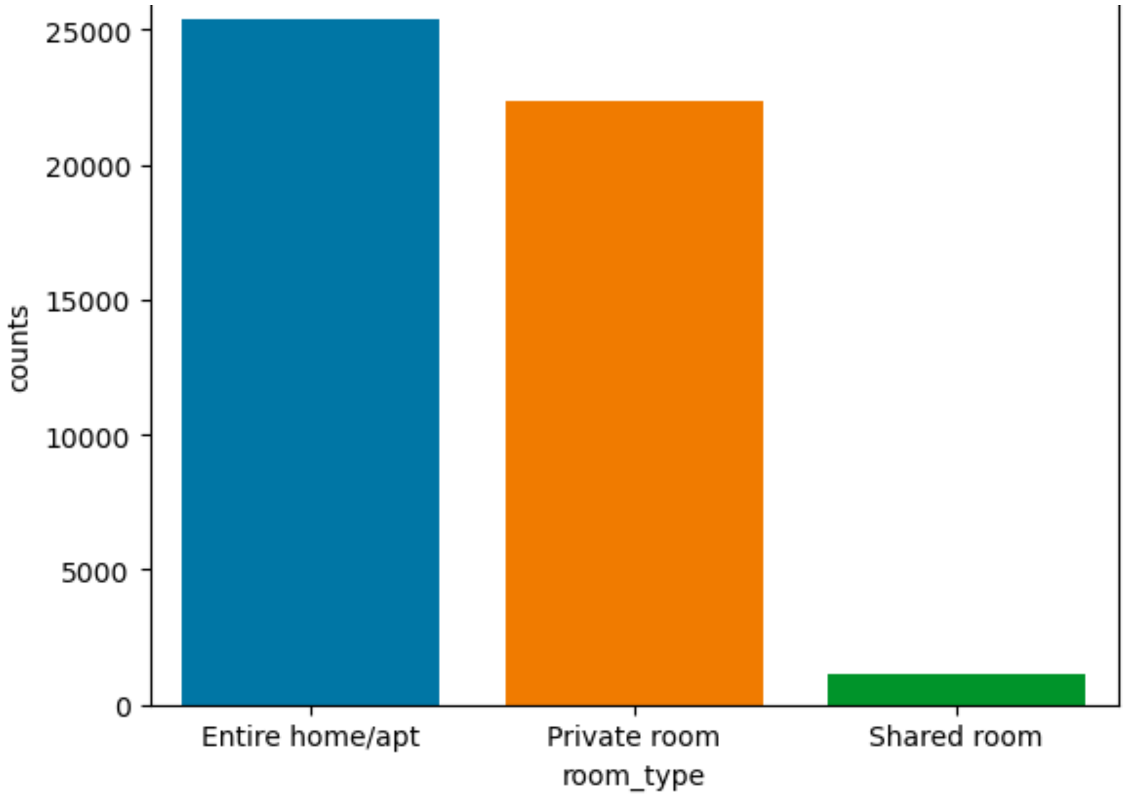




Neighborhood Group Distribution

- Manhattan: 21,661 listings (44.30%)
- Brooklyn: 20,104 listings (41.12%)
- Queens: 5,666 listings (11.59%)
- Bronx: 1,091 listings (2.23%)
- Staten Island: 373 listings (0.76%)





Room Type Distribution

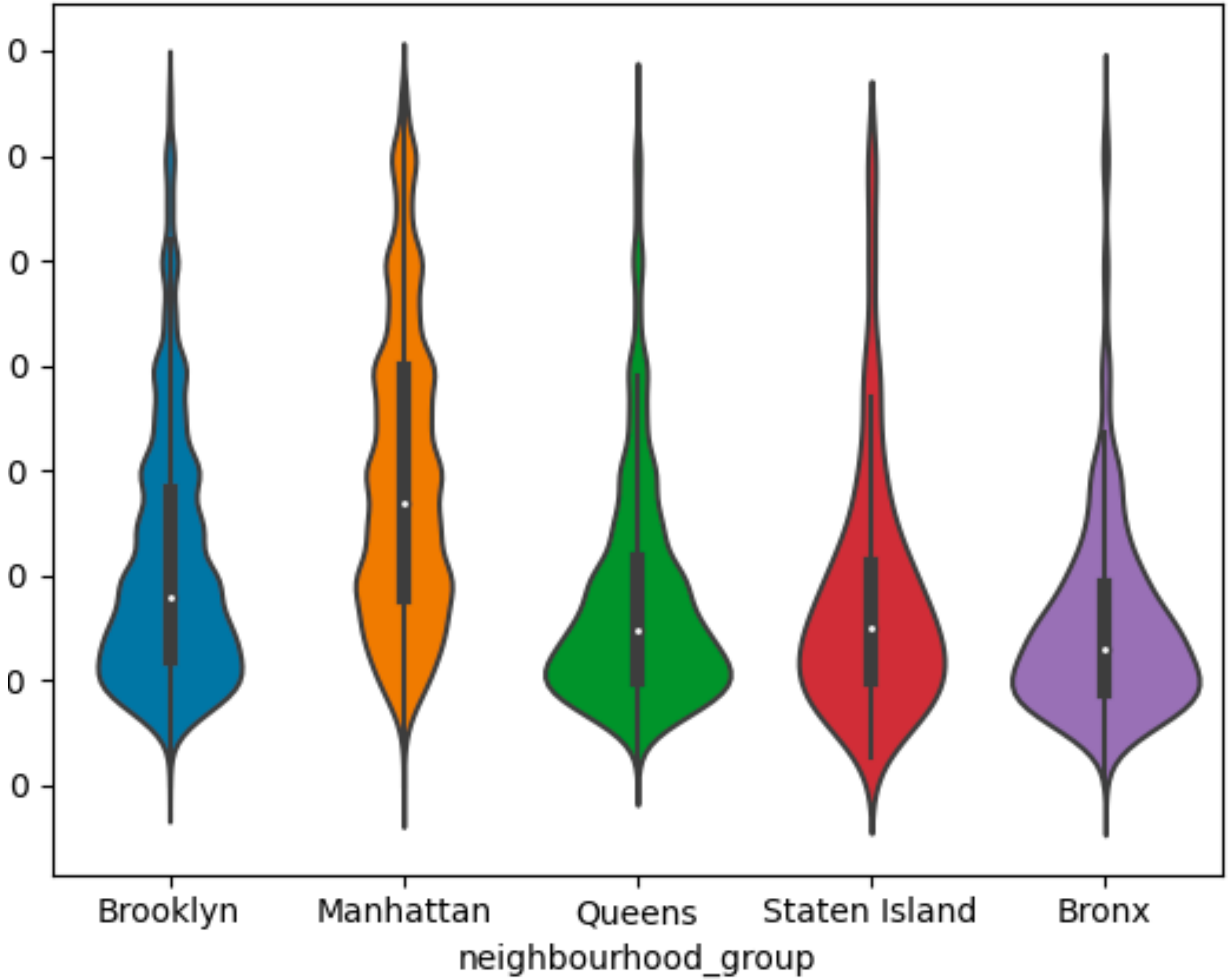
This analysis reveals that the majority of listings of 51.97% offer entire homes or apartments, while private rooms account for 45.66% of the listings. Shared rooms make up a smaller proportion at 2.37%.

Price Distribution

- Count: 48,895 listings
- Mean: $152.72
- Standard Deviation: $240.15
- Minimum Price: $0
- 25th Percentile: $69
- Median (50th Percentile): $106
- 75th Percentile: $175
- Maximum Price: $10,000

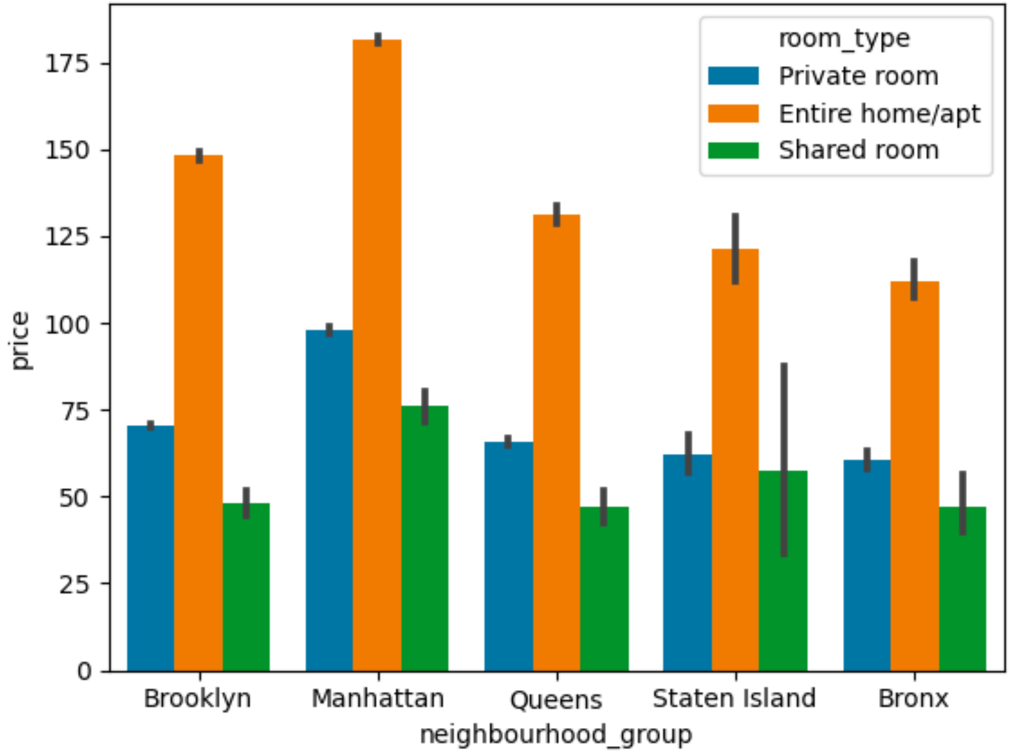
Price Distribution After Remove Outliners

The individual mean of every state is below or near to 100 while for Manhattan its greater (near to 150)



Price Analysis by Neighborhood Group and Room Type

- Almost every state share have equal number of private rooms and shared rooms except Manhattan and Brooklyn. Manhattan have maximum number of Entire home/apt typr of rooms followed by Brooklyn and staten island.
- Price in Manhattan is obviously higher than others.

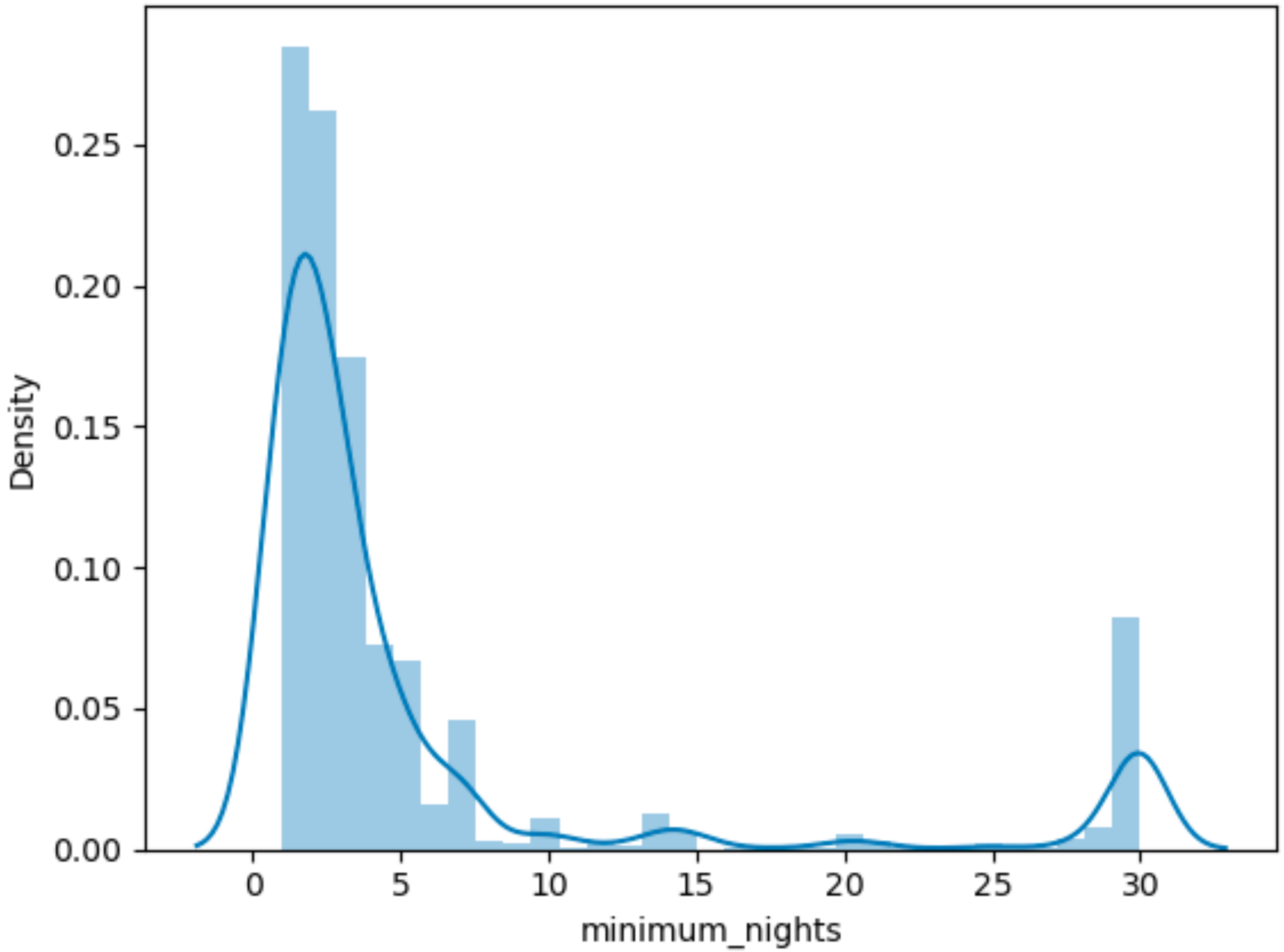


Minimum Nights Distribution

Count: 45,923 listings
Mean: 6.94 nights
Standard Deviation: 19.86 nights
Minimum: 1 night
25th Percentile: 1 night
Median (50th Percentile): 2 nights
75th Percentile: 5 nights
Maximum: 1,250 nights

Minimum Nights Distribution (30 Days Focus)

By narrowing my focus to a 30-day range, I gain insights specifically related to short-term stays. This information is valuable for guests who are looking for accommodations for shorter durations.



Availability Analysis

- Count: 44,858 listings
- Mean: 107.85 days
- Standard Deviation: 129.79 days
- Minimum: 0 days
- 25th Percentile: 0 days
- Median (50th Percentile): 37 days
- 75th Percentile: 212 days
- Maximum: 365 days

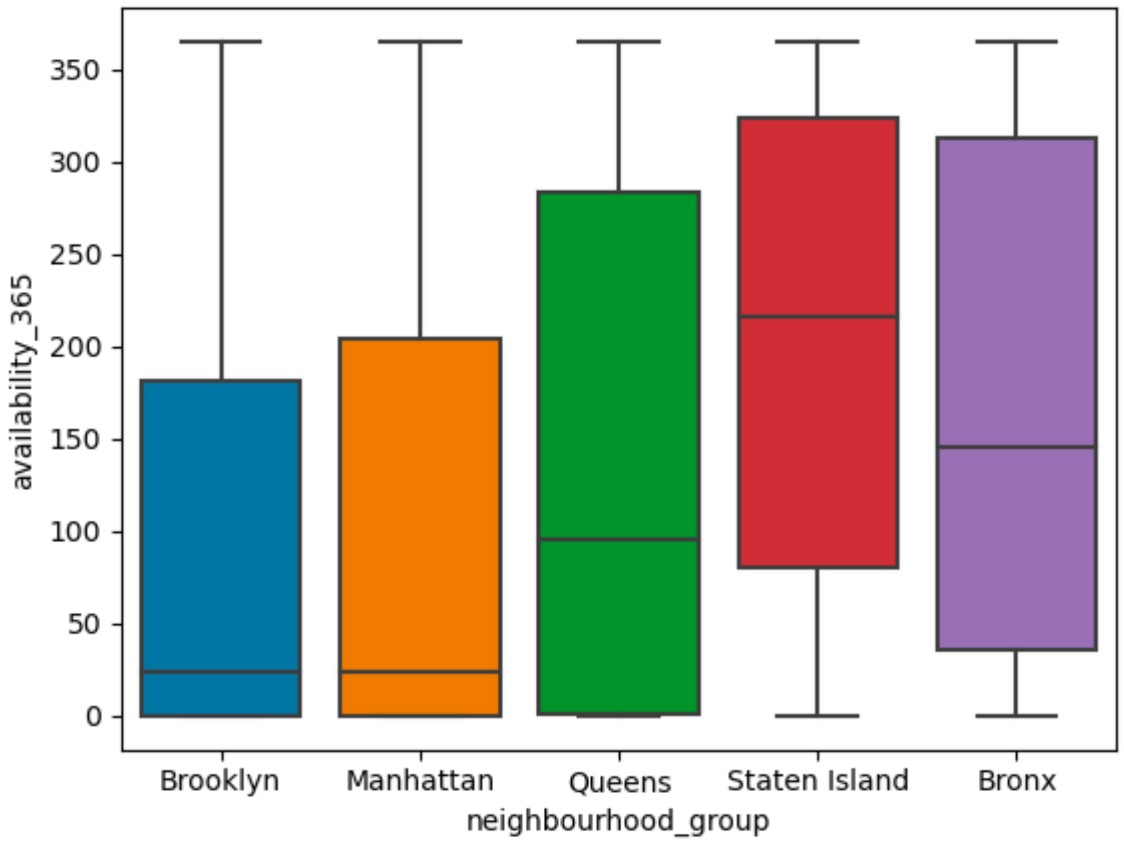


Table comparing model performance metrics

Model	Mean Squared Error (MSE)	R-squared
Dummy Regression	4468.08	-0.00003
Linear Regression	2204.77	0.5065
Random Forest Regressor	1871.59	0.5811
Gradient Boosting Regressor	1915.52	0.5713
XG Boost Regressor	1872.09	0.5810

Model Evaluation and Conclusion

- Dummy Regression: The baseline model had an MSE of 100% and an R-squared of 0%.
- Linear Regression: The linear regression model improved the MSE by 55.4% and achieved an R-squared of 50.7%.
- Random Forest Regressor: The Random Forest Regressor model achieved the best performance, reducing the MSE by 81.6% compared to the baseline and achieving an R-squared of 58.1%.
- Gradient Boosting Regressor: The Gradient Boosting Regressor model also performed well, reducing the MSE by 78.4% and achieving an R-squared of 57.1%.
- XG Boosting Regressor: The XG Boosting Regressor model showed similar performance to the Random Forest Regressor, reducing the MSE by 81.6% and achieving an R-squared of 58.1%.
- Based on these results, both the Random Forest Regressor and XG Boosting Regressor models demonstrated a significant improvement over the baseline model, with an approximately 81.6% reduction in MSE. The models' R-squared values indicate that they capture around 58.1% of the variance in the target variable. These models provide a solid foundation for further analysis and optimization to enhance predictive accuracy.

Interpretation of Findings

1. Price Variation:

The dataset showed a wide range of prices for Airbnb listings in New York City, with the minimum price at $0 and the maximum at $10,000. The mean price was approximately $152, indicating the average price per night for an Airbnb stay in the city.

2. Room Types:

The majority of Airbnb listings in New York City were categorized as "Entire home/apartment" (approximately 52%) and "Private room" (around 46%). "Shared room" listings accounted for a smaller proportion (about 2%) of the total.

Interpretation of Findings

3. Neighborhood Groups:

The dataset included listings from five main neighborhood groups in New York City. Manhattan had the highest number of listings (around 44%), followed by Brooklyn (approximately 41%). Queens, Bronx, and Staten Island accounted for smaller proportions of the listings.

4. Minimum Nights and Availability:

The average minimum nights required for an Airbnb stay was approximately 6.9, with a standard deviation of 19.9. This indicates variations in the minimum stay requirements among listings. The availability of listings throughout the year varied, with an average of 107.9 days available out of 365 days.

Interpretation of Findings

Based on these findings, I can conclude that the New York City Airbnb market offers diverse options in terms of price, room types, and neighborhoods. Understanding these factors can help hosts and guests make informed decisions. Further analysis and modeling have shown promise in predicting listing prices, with the Random Forest Regressor, Gradient Boosting Regressor, and XG Boosting Regressor models demonstrating strong performance.

Application

- One of the significant implications of my analysis is the potential for price prediction in the Airbnb market. By leveraging the insights gained from the dataset, it is possible to develop models and algorithms that can accurately predict the price of a listing based on various factors such as neighborhood group, room type, minimum nights, and availability.

- Price prediction models can be beneficial for both hosts and guests. For hosts, accurate price prediction can help them set competitive prices for their listings, ensuring maximum occupancy and revenue generation. By taking into account factors such as neighborhood popularity, room type demand, and seasonal variations, hosts can adjust their prices dynamically to maximize their earnings.

Application

- On the other hand, guests can benefit from price prediction by having a better understanding of the expected price range for listings in their preferred neighborhoods and room types. This can help them make informed decisions, plan their budget, and find the best value for their money.

- The development of robust price prediction models requires advanced machine learning techniques, data preprocessing, and feature engineering. By incorporating the insights derived from my analysis into the modeling process, it is possible to create accurate and reliable price prediction models that can contribute to the efficiency and transparency of the Airbnb marketplace.