

Evaluating Airbnb as a Side Hustle in NYC:

Insights into Profitability, Market Trends, and Strategies

Author:

Rahma Shahid

Abstract

Airbnb has revolutionized the hospitality industry, offering individuals an opportunity to earn supplemental income by hosting travelers. This study evaluates the profitability of Airbnb listings in New York City, examining market trends, pricing patterns, and competitive strategies. Using a dataset of over 48,000 listings, we explored key factors such as neighborhood demand, room types, and review trends. Data preprocessing and cleaning addressed issues like missing values and outliers, while visualizations highlighted significant pricing disparities across neighborhoods and room types. Predictive models, including Random Forest and Logistic Regression, were employed to forecast pricing and booking likelihood. Insights reveal that Manhattan dominates in profitability, but budget-conscious travelers prefer Staten Island and the Bronx. This report concludes with recommendations for maximizing earnings and an interactive dashboard to assist prospective hosts.

Introduction

Background and Motivation

The rise of Airbnb has transformed the lodging industry, enabling individuals to monetize their properties. In NYC, a competitive yet lucrative market, hosting on Airbnb offers significant potential but requires data-driven decision-making. This project examines whether Airbnb hosting is a viable side hustle in NYC, identifying profitable neighborhoods and strategies to optimize revenue.

Objectives

1. Analyze profitability by neighborhood and room type.
2. Predict potential earnings using data-driven models.
3. Develop interactive dashboards to guide hosts in making informed decisions.

Predicted Outcome

Listings in high-demand areas with favorable attributes (e.g., flexible stays, positive reviews) are expected to yield higher profitability.

Data Description

Dataset Overview

The dataset comprises 48,895 Airbnb listings in NYC, with 16 features, including:

- **Location Attributes:** Neighborhood, borough, latitude, and longitude.
- **Listing Features:** Room type, price, minimum stay, availability.
- **Host Metrics:** Number of reviews, calculated host listings count.

Data Issues and Resolutions

1. Missing Values:

- reviews_per_month and last_review were filled with zeros or placeholders like "No Review."
- Host names and property names were filled with "Unknown" and "No Name" to retain data integrity.

2. Outliers:

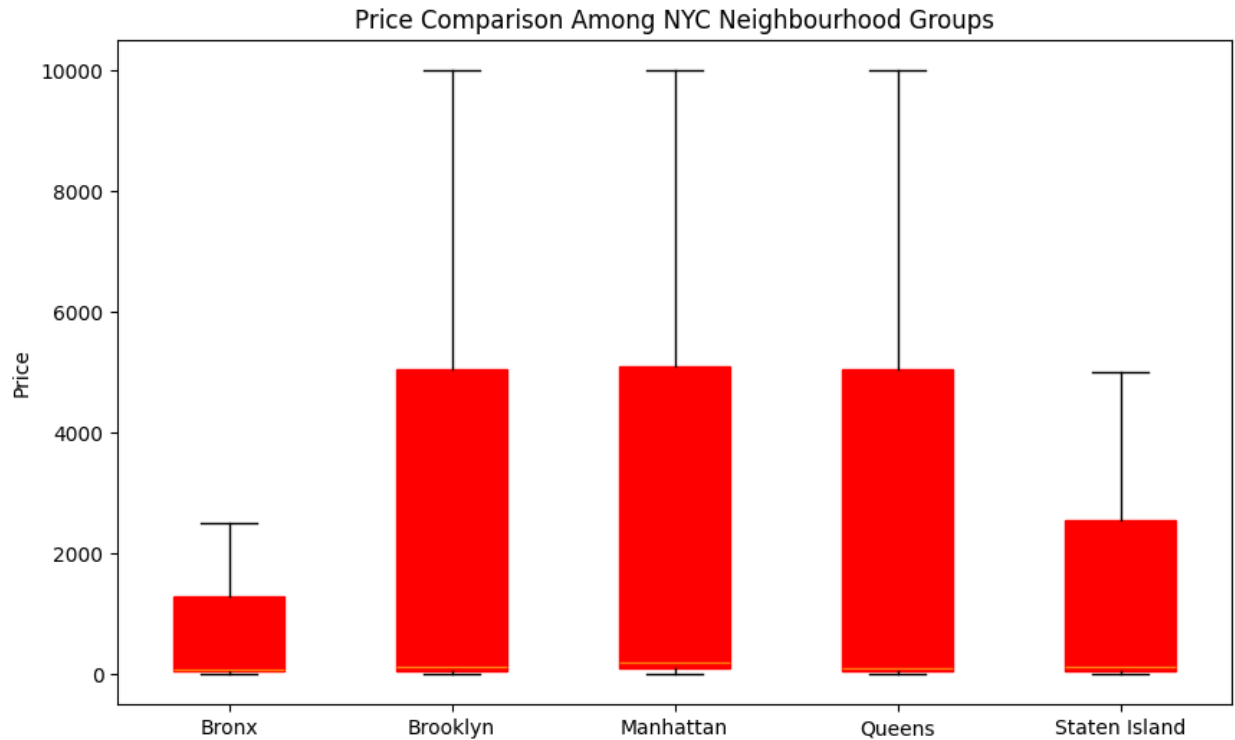
- Prices exceeding \$1,000 were identified as outliers using the Interquartile Range (IQR) method and removed.

3. Standardization:

- Room type values were converted to lowercase for consistency.
- Dates were standardized to datetime format.

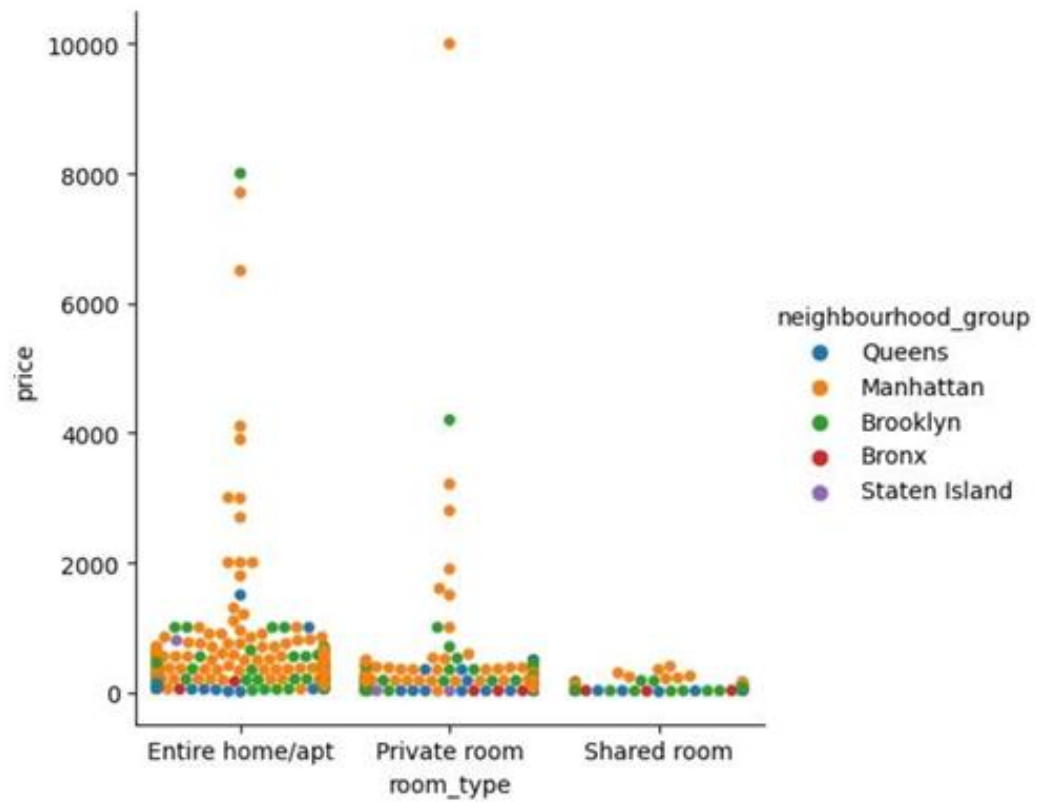
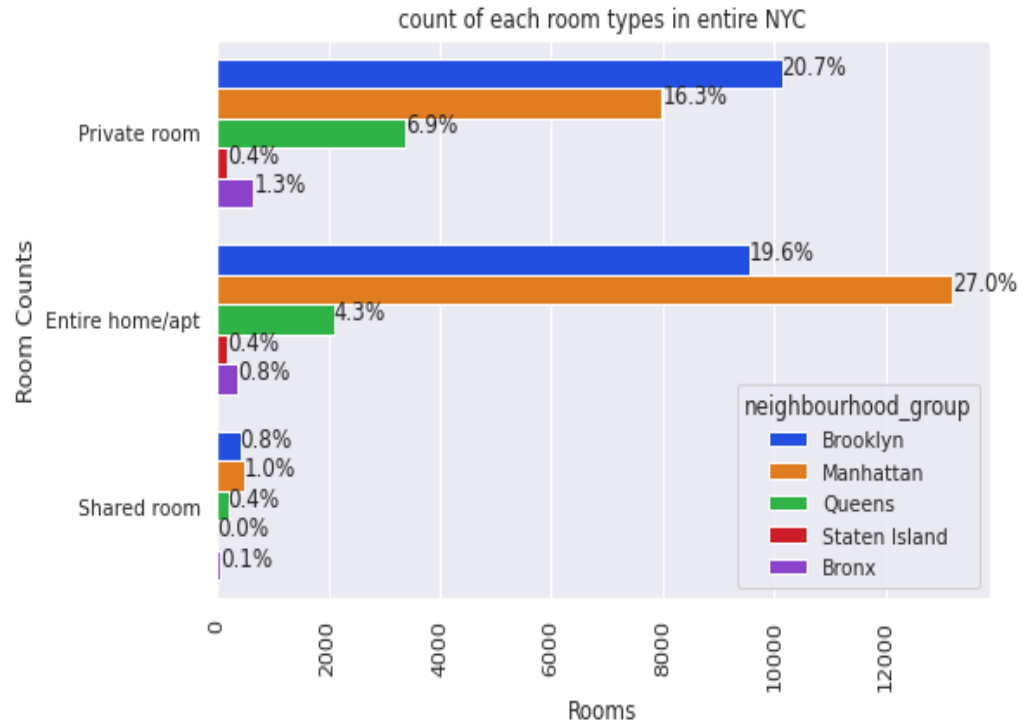
Exploratory Data Analysis (EDA)

Visualization Insights



1. Price Trends by Neighborhood Group:

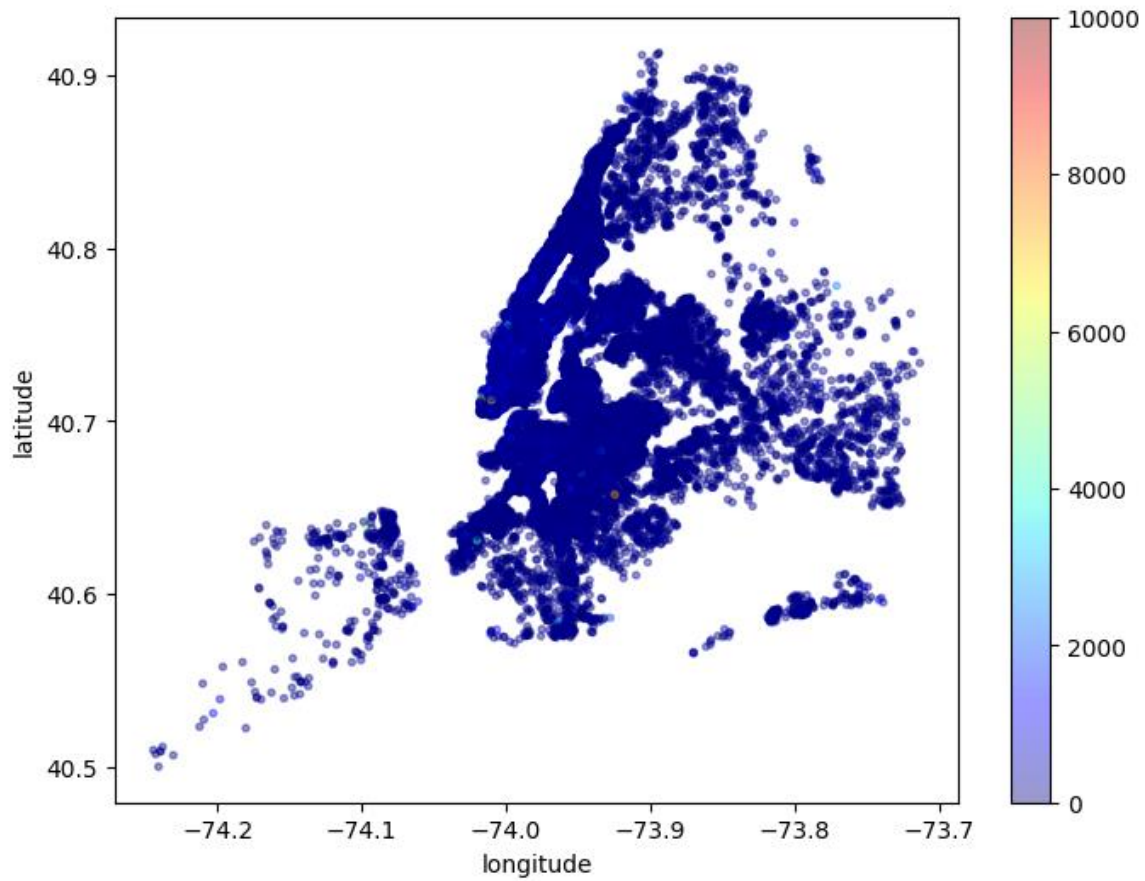
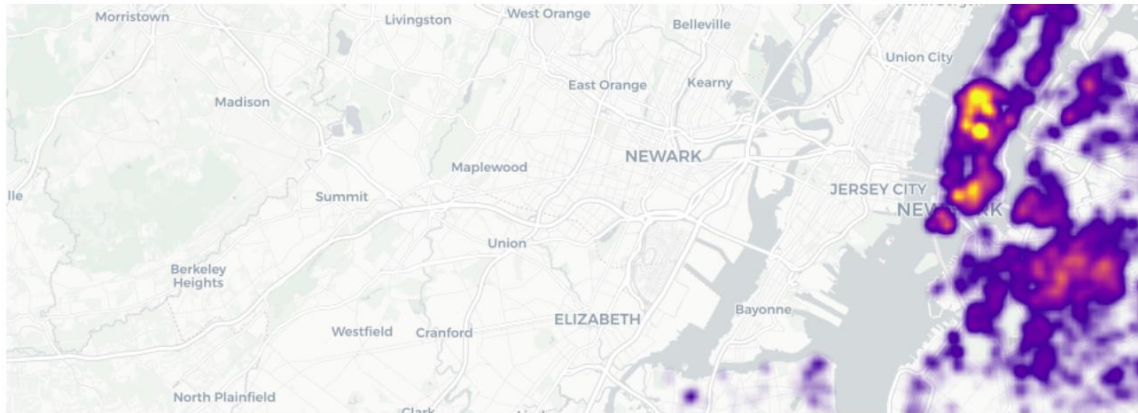
- Manhattan dominates with the highest average price of \$196.88, followed by Brooklyn (\$124.38).
- Bronx (\$87.50) and Staten Island (\$114.81) are budget-friendly alternatives.
- Boxplots revealed Manhattan's wide price distribution, with several high-end outliers.



2. Room Type Distribution:

- Entire homes command higher prices but exhibit greater variability.
- Shared rooms are the least common and least expensive, consistently clustering at lower price points.

High-Demand Areas for Airbnb Listings in NYC (Based on Reviews Per Month)



3. Geospatial Analysis:

- Listings are densely clustered in Manhattan and Brooklyn.

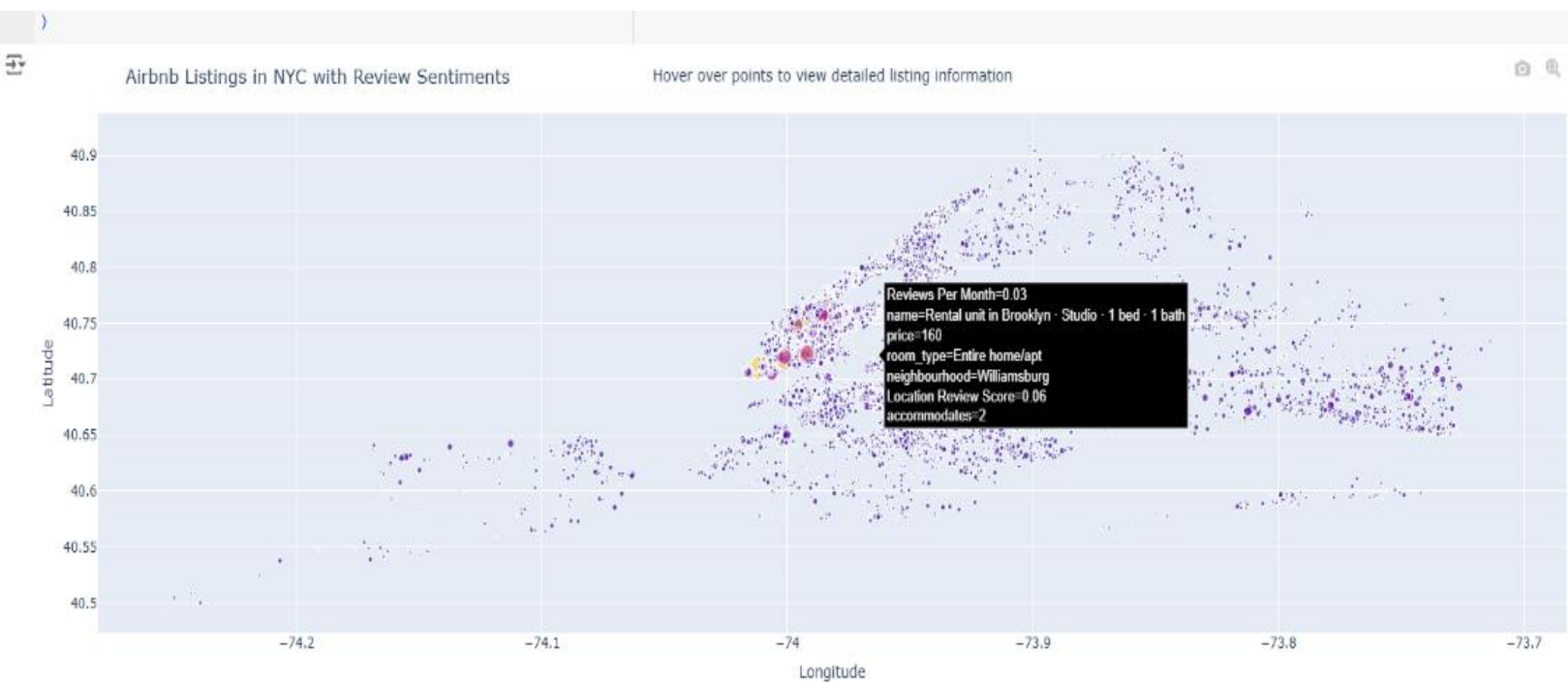
- High prices correlate with central locations near tourist hotspots, such as Times Square and the Financial District.

4. Review Trends:

- Listings with positive reviews and high availability tend to perform better.
- Manhattan leads in the number of reviews, indicating its popularity among tourists.

Key Findings

- Manhattan dominates in both price and quantity of listings, but competition is fierce.
- Staten Island and Bronx listings cater to budget-conscious travelers.



This figure gives us a quick insight of NYC listing based on people's experience

Data Cleaning and Preprocessing

Steps Taken

1. Handling Missing Values:

- Replaced missing names with placeholders.
- reviews_per_month set to 0 for listings with no reviews.

2. Outlier Removal:

- Used IQR to eliminate listings with excessively high prices.

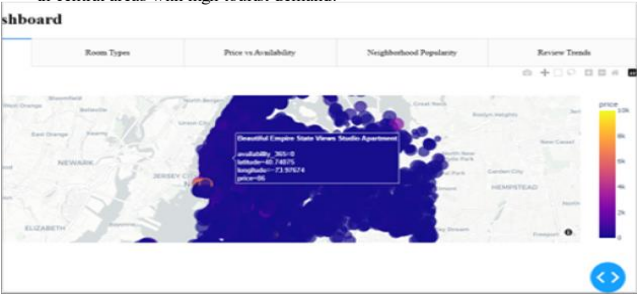
3. Feature Engineering:

- Extracted new attributes such as year and period (pre-COVID, during COVID, post-COVID) from last_review.
- Encoded categorical variables for modeling.

Dashboard no 1

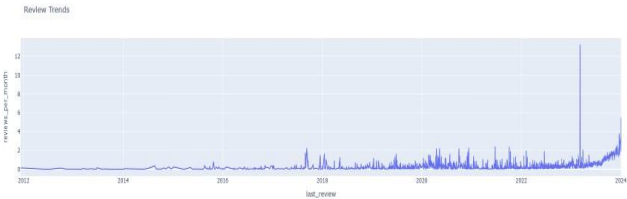
Map Overview (Geospatial Insights)

High-price clusters are likely concentrated in Manhattan, particularly in central areas with high tourist demand.

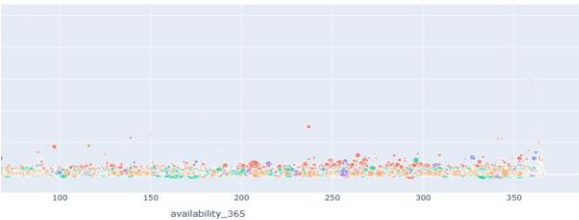


Review Trends Over Time

- Peaks in the line chart for reviews per month may coincide with travel seasons or events in NYC (e.g., holidays, festivals).
- Declining review activity could signal external factors such as travel restrictions or reduced demand.



Price vs. Availability



It reveals that **high-priced listings are less frequently available** (lower availability values), while affordable options may see high occupancy rates.

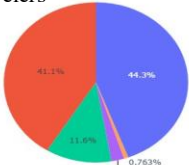
Room Types and Neighborhood Groups



Entire homes/apartments may dominate Manhattan and Brooklyn, reflecting demand for privacy in high -end or trendy areas.

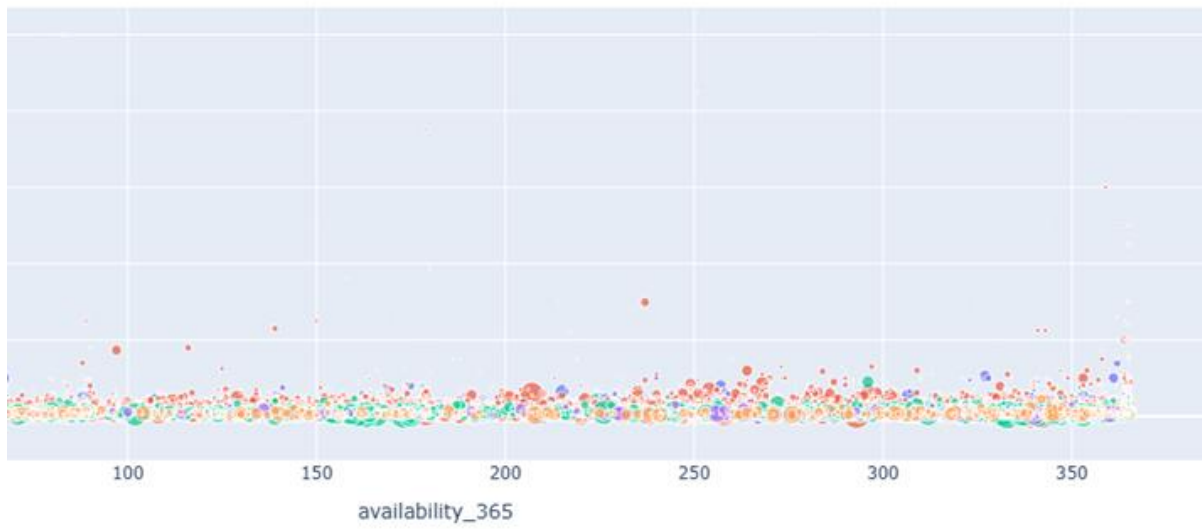
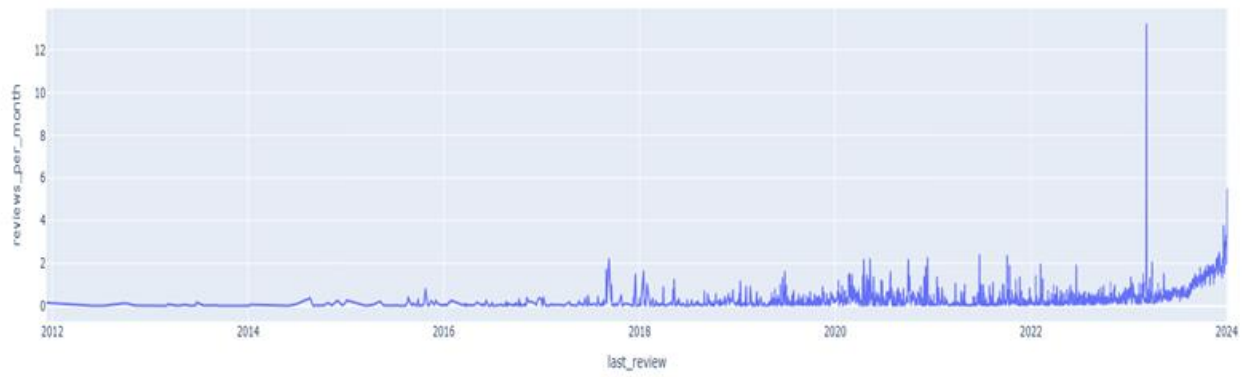
Neighborhood Popularity

The borough with the largest pie segment (likely Manhattan or Brooklyn) indicates the highest density of listings. This reflects the draw of these areas for travelers



Including the images from dashboad below for better visiability

Review Trends



The dashboard is designed to provide a comprehensive analysis of Airbnb listings in New York City. It incorporates five interactive tabs, each focusing on a unique aspect of the dataset:

1. **Map Overview (Geospatial Insights)**

This tab visualizes the geographic distribution of Airbnb listings using a scatter map. Key attributes like price and availability (visualized through color and size) provide a quick overview of areas with high or low-priced listings and their availability patterns.

2. **Room Types and Neighborhood Groups**

This bar chart breaks down room types (e.g., entire home, private room) across neighborhood groups. It allows users to understand which room types dominate each part of NYC.

3. **Price vs. Availability**

This scatter plot explores the relationship between price and availability, color-coded by neighborhood groups. It helps identify patterns such as whether higher prices correlate with lower availability.

4. **Neighborhood Popularity**

A pie chart highlights the proportion of listings in each neighborhood group, offering a quick look at the popularity and market share of different boroughs.

5. **Review Trends Over Time**

This line chart tracks the average number of reviews per month over time. It is particularly useful for analyzing how user engagement with Airbnb listings changes seasonally or with external factors.

1: Map Overview

- **High-price clusters** are likely concentrated in Manhattan, particularly in central areas with high tourist demand.
- Listings with high availability (size of points) could indicate less popular or underutilized listings.

Tab 2: Room Types and Neighborhood Groups

- Entire homes/apartments may dominate Manhattan and Brooklyn, reflecting demand for privacy in high-end or trendy areas.
- Private and shared rooms could be more prevalent in Queens or the Bronx, catering to budget travelers.

Tab 3: Price vs. Availability

- A scatter plot might reveal that **high-priced listings are less frequently available** (lower availability values), while affordable options may see high occupancy rates.
- Outliers (e.g., extremely expensive listings with high availability) could suggest niche or luxury offerings.

Tab 4: Neighborhood Popularity

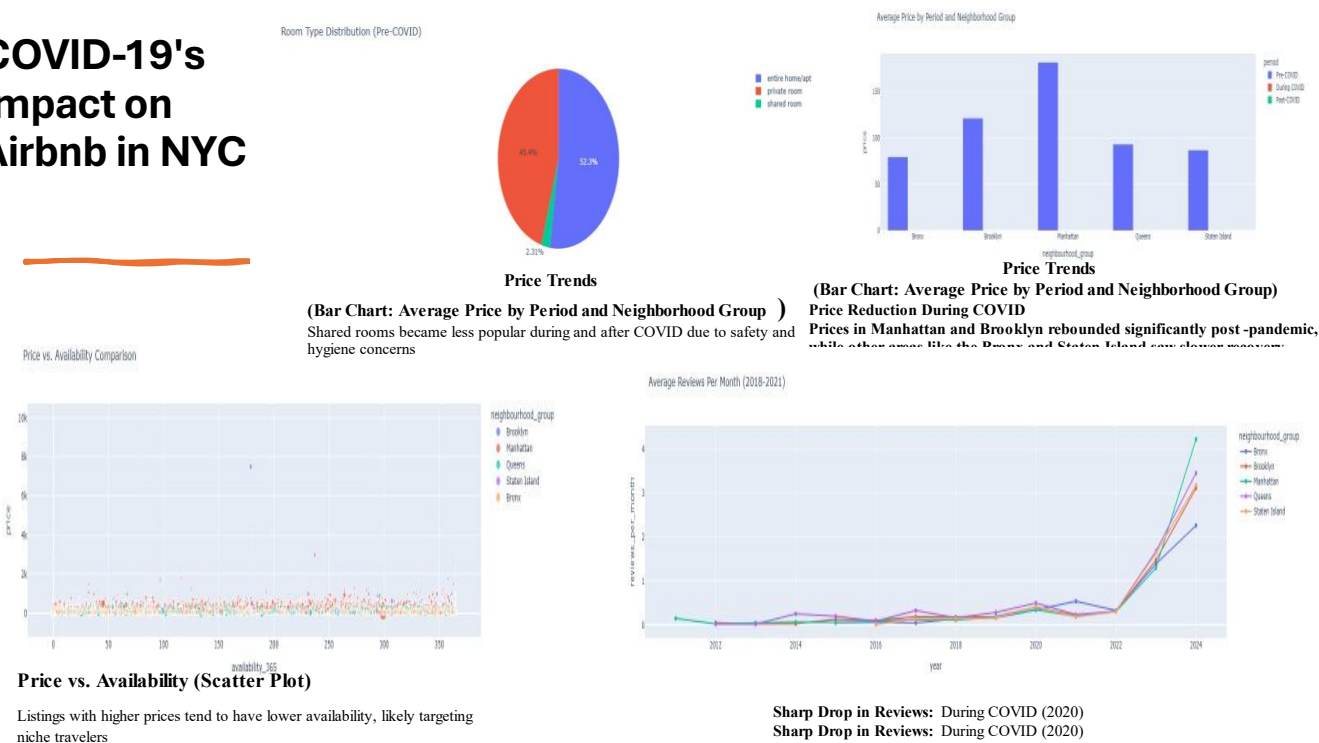
- The borough with the largest pie segment (likely Manhattan or Brooklyn) indicates the highest density of listings. This reflects the draw of these areas for travelers.

Tab 5: Review Trends

- Peaks in the line chart for reviews per month may coincide with travel seasons or events in NYC (e.g., holidays, festivals).
- Declining review activity could signal external factors such as travel restrictions or reduced demand.

Dasboard no 2: Covid-19’s imapct on Airbnb makrket in NYC

COVID-19's Impact on Airbnb in NYC



1. Booking Trends (Line Chart: Average Reviews Per Month)

What it shows:

- The line chart highlights the **average reviews per month** for each neighborhood group (Manhattan, Brooklyn, Queens, Bronx, and Staten Island) over time, segmented into Pre-COVID, During COVID, and Post-COVID periods.

Insights:

- **Sharp Drop in Reviews:** During COVID (2020), all neighborhoods show a noticeable decline in reviews per month, indicating a significant reduction in booking activity due to travel restrictions and decreased demand.
 - **Recovery Post-COVID:** Post-COVID (2021-2022), reviews start to recover, though the rate varies by neighborhood.
 - **Manhattan and Brooklyn Lead:** These two neighborhoods consistently show higher engagement compared to others, suggesting higher demand and popularity pre- and post-pandemic.
-

2. Price Trends (Bar Chart: Average Price by Period and Neighborhood Group)

What it shows:

- The bar chart compares **average prices** across neighborhoods for Pre-COVID, During COVID, and Post-COVID periods.

Insights:

- **Price Reduction During COVID:** Prices dropped across all neighborhoods during COVID, reflecting lower demand and attempts to attract limited travelers.
 - **Post-COVID Recovery:** Prices in Manhattan and Brooklyn rebounded significantly post-pandemic, while other areas like the Bronx and Staten Island saw slower recovery.
 - **Manhattan is Most Expensive:** Consistently, Manhattan has the highest average prices across all periods, indicating its premium market status.
-

3. Room Type Distribution (Pie Charts: Pre-COVID vs Post-COVID)

What it shows:

- Two pie charts display the **distribution of room types** (Entire Home, Private Room, Shared Room) before and after COVID.

Insights:

- **Shift in Room Type Preference:**
 - Pre-COVID: Entire homes/apartments dominated, followed by private rooms.
 - Post-COVID: The share of **entire homes/apartments increased**, reflecting travelers' preference for private, isolated spaces post-pandemic.

- **Shared Rooms Decline:** Shared rooms became less popular during and after COVID due to safety and hygiene concerns.
-

4. Price vs. Availability (Scatter Plot)

What it shows:

- This scatter plot explores the relationship between **price** and **availability** for different neighborhoods.

Insights:

- **Lower Availability, Higher Price:** Listings with higher prices tend to have lower availability, likely targeting niche travelers.
 - **High Availability Clusters:** Many listings (particularly in the Bronx and Staten Island) show high availability but low prices, indicating underutilization or lower demand in those areas.
 - **Outliers:** Some high-priced listings with unusually high availability may indicate luxury or long-term stays.
-

5. Geographical Density (Heatmap)

What it shows:

- The density map visualizes **geographic clusters of listings** and their prices, particularly during the COVID period.

Insights:

- **Price Clusters in Manhattan:** Listings in central Manhattan show the highest price density, even during COVID, highlighting its resilience and premium nature.
 - **Brooklyn's Growing Importance:** Brooklyn emerges as another significant cluster, reflecting its role as a more affordable alternative to Manhattan.
 - **Sparse Listings in Outer Boroughs:** Staten Island and parts of the Bronx show sparse density and lower prices, indicating lower demand.
-

Summary of Insights

1. **Demand Shock During COVID:** A significant drop in bookings (reviews) occurred across all neighborhoods during the pandemic.

2. **Room Preference Shift:** Travelers favored entire homes/apartments post-COVID, reducing demand for shared accommodations.
3. **Price Recovery:** Prices rebounded post-COVID, particularly in Manhattan and Brooklyn.
4. **Geographic Trends:** Central Manhattan and Brooklyn remained strong markets, while outer boroughs struggled with lower demand and prices.
5. **Availability vs Price:** Higher-priced listings showed lower availability, while more affordable listings often had high availability, especially in less popular areas.

Dashboard no 3: Insights from Host Activity and Listing Analysis Dashboard



This dashboard provides valuable insights into **host behavior** and their listing patterns on Airbnb in NYC. Below are key takeaways from each visualization:

1. Average Price: Single vs Multi-Listing Hosts (Top Chart)

What it shows:

- A comparison of the **average price** of listings managed by **single-listing hosts** versus **multi-listing hosts**.

Insights:

- **Price Distribution Similarity:** Both single and multi-listing hosts have a significant number of listings priced below \$500.
- **Outliers:** There are a few listings managed by **multi-listing hosts** priced significantly higher (above \$5,000), indicating that luxury properties are more common with hosts managing multiple listings.

- **Central Tendency:** While both groups share similar price distributions for the majority of listings, **multi-listing hosts** may slightly skew toward higher prices.
-

2. Top 10 Hosts with the Most Listings (Middle Chart)

What it shows:

- This bar chart highlights the **top 10 hosts** with the **largest number of listings** and their respective counts.

Insights:

- **Highly Active Hosts Dominate:** Some hosts have over **300 listings**, which indicates they operate as commercial entities rather than individual hosts.
 - **Uneven Distribution:** The listings are highly concentrated among a small number of hosts, suggesting a disparity between small-scale hosts and large operators.
 - **Hosts with 50+ Listings:** Several hosts manage between **50 to 150 listings**, highlighting their significant influence on the Airbnb market in NYC.
-

3. Distribution of Total Listings per Host (Bottom Chart)

What it shows:

- This histogram illustrates the **distribution** of the number of listings managed per host.

Insights:

- **Majority of Hosts Have Few Listings:** Most hosts manage **only 1 to 2 listings**, representing individuals or small-scale hosts.
 - **Skewed Distribution:** There is a sharp drop-off as the number of listings increases, indicating fewer hosts manage a large portfolio.
 - **Long Tail:** A small number of hosts manage **over 100 listings**, confirming the presence of commercial-scale operators in the market.
-

Summary of Insights

1. Host Landscape:

- The majority of Airbnb hosts in NYC are **small-scale operators** with one or two listings.
- However, a small number of commercial hosts dominate the market, managing **50+ listings** and some with over **300 listings**.

2. Pricing Patterns:

- Listings managed by **multi-listing hosts** tend to include luxury or high-priced properties, particularly outliers above \$5,000.
- Despite this, most listings fall into a similar price range regardless of host type.

3. Market Disparity:

- There is a clear divide between individual hosts (1-2 listings) and commercial operators, which could impact market dynamics and competition.

Data Modeling

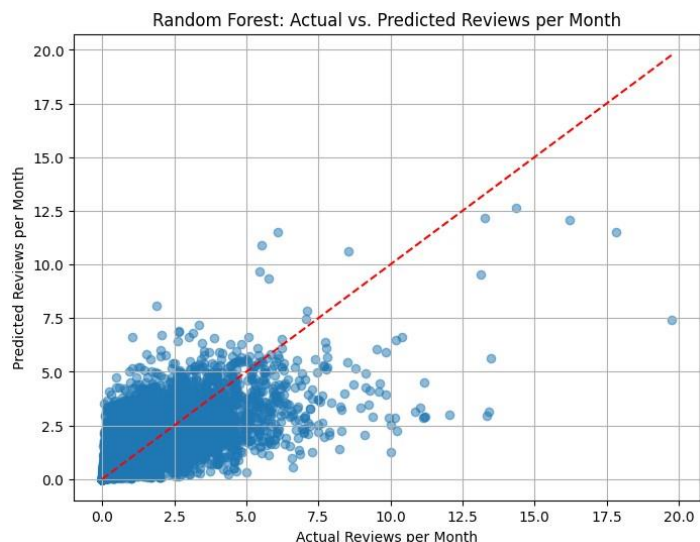
Modeling helps us predict the future of Airbnb listings. Random Forest stood out for predicting review trends, while Logistic Regression excelled at identifying ‘fully booked’ listings. Linear Regression struggled with NYC’s pricing variability, but it offered valuable baseline insights. The lesson? Models help uncover hidden patterns, but they’re only as good as the data we feed them.

Models Used

1. Random Forest Regression:

Random Forest

- RMSE: 1.0632857965330489
- A lower RMSE generally indicates better model performance.
- **Scatter Plot Analysis**
- **X-axis:** Actual Reviews per Month
- **Y-axis:** Predicted Reviews per Month
- The majority of blue dots (predictions) are close to the red dashed line, which means the Random Forest model is predicting values reasonably well



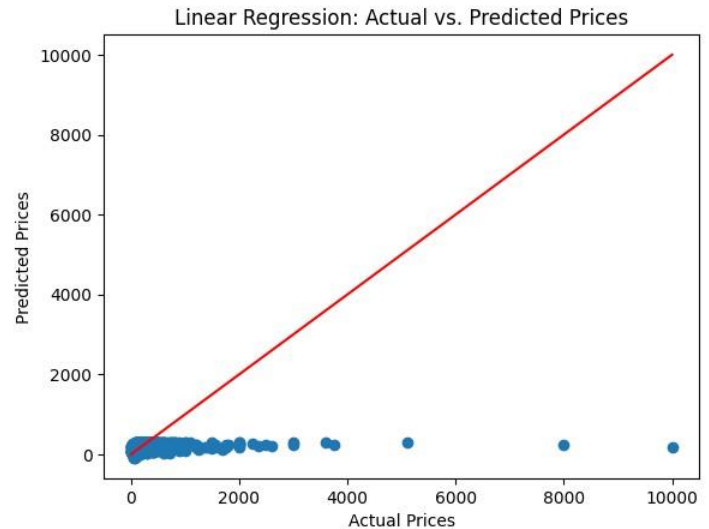
- The model performed pretty well with less RMSE.

2. Linear Regression:

- Predicted pricing but struggled with high variability in the dataset.
- R^2 : 0.116, suggesting limited explanatory power.

Models Results

- RMSE: 197.72503141630483
 R^2 : 0.11625801812234327
- The **scatter plot** compares **actual prices** (x-axis) with the **predicted prices** (y-axis)
- The model performs poorly on this dataset, as shown by the high RMSE and low R^2 score.



○

3. Logistic Regression:

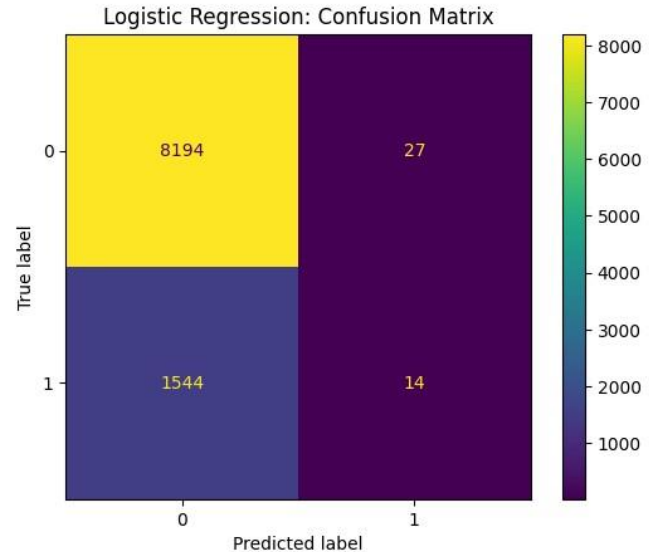
- Classified listings as "fully booked" (availability > 300 days).
- Accuracy: 84%, highlighting the role of availability and pricing in booking likelihood.

Model Insights

- Random Forest outperformed other models in prediction accuracy.
- Listings with high availability, competitive pricing, and positive reviews are more likely to be fully booked.

Logistic regression

- **Feature Engineering** :price, minimum_nights , number_of_reviews
- With fully booked room_type fully_booked
- While the Logistic Regression model performs well in identifying the negative class (**True 0**), it performs poorly for the positive class (**True 1**). Adjusting for class imbalance and exploring other methods could improve the overall performance
- Can perform better



•

Evaluation

Metrics

1. RMSE for regression models to assess prediction accuracy.
2. Precision, recall, and accuracy for classification models.

Model Comparison

- Random Forest: Best for predicting review frequency.
- Logistic Regression: Effective for identifying fully booked listings.
- Linear Regression: Underperformed due to high variance in price data.

Discussion

Key Learnings

Here's what hosts should know: Manhattan listings earn the most but face stiff competition. Bronx and Staten Island listings cater to budget-conscious travelers, offering untapped potential. To maximize profits, focus on competitive pricing, flexibility, and delivering top-notch guest experiences. Remember, reviews are gold—happy guests mean more booking

- Manhattan offers significant earning potential but requires competitive strategies due to high competition.
- Flexible minimum stays and positive reviews can boost profitability across all boroughs.
- Budget-conscious travelers gravitate toward Staten Island and Bronx, presenting opportunities for unique value propositions.

Limitations

Every project has its hurdles. Sparse data made analysis tricky for Staten Island, and we didn't include real-time factors like events or economic changes. But the future is bright next steps could include analyzing other cities or even using guest reviews to understand satisfaction trends. Airbnb hosting is a journey, and data helps light the way.

- Sparse data in low-demand areas (e.g., Staten Island) limited model accuracy.
- External factors like seasonal demand or local events were not included.

Conclusion

Summary

This project provides actionable insights for prospective Airbnb hosts in NYC. Key findings include:

1. Manhattan listings dominate but require high-quality listings to compete.
2. Budget-friendly options in Bronx and Staten Island cater to untapped markets.
3. Interactive dashboards offer users the ability to explore profitability trends.

Future Work

1. Incorporate real-time data, such as local events or economic conditions.
2. Expand the analysis to other major cities to identify universal patterns in Airbnb profitability.
3. Integrate sentiment analysis from reviews to assess customer satisfaction.

References

NYC Airbnb Dataset: [Kaggle](https://www.kaggle.com/datasets/rhonarosecortez/new-york-airbnb-open-data)

<https://www.kaggle.com/datasets/rhonarosecortez/new-york-airbnb-open-data>

https://github.com/otavio-s-s/data_science/blob/master/Analysing%20New%20York%20City%20Airbnb%20Data.ipynb

<https://medium.com/almabetter/exploratory-data-analysis-on-nyc-airbnb-2019-dataset-fe908c2accaa>