

University of Essex Online

Machine Learning

Group Project Report

## **Airbnb Business Analysis**

Presented by (Group 2) :

Elias Medig, Radha Murugan, Pavlos Papachristos and Sadiq Lawal

Submission Date

10.03.2024

# Table of Contents

## Contents

1. Introduction .....	2
1.1 Background .....	2
1.2 Business Analytics Question.....	2
2. Data Analysis & Insights .....	4
2.1 Room Type Distribution.....	4
2.2 Distribution of Airbnb Listings by Borough.....	4
Key Insights.....	5
2.3 Median Airbnb Prices by Borough .....	5
Key Insights.....	6
2.4 Correlation Analysis: Price vs. Location & Availability .....	6
Key Insights.....	7
2.5 Top 10 Most Expensive Neighbourhoods .....	7
Key Insights.....	8
2.6 Heatmap of Price and availability relation.....	8
3. ML: K-Means Clustering for NYC Airbnb Market Segmentation.....	10
3.1 Determining Optimal Clusters Using the Elbow Method .....	10
3.2 NYC Airbnb Clusters Map .....	11
Cluster Interpretations .....	12
4. Final Business Strategy for Airbnb NYC .....	13
5. Conclusion .....	14
Appendix .....	16

# **1. Introduction**

## *1.1 Background*

This report aims to provide Airbnb with analytical insights into their New York City operations by examining historical data from 2019 using the AB\_NYC\_2019 dataset (Kaggle, 2019). It outlines the formulation of a business analytics question, the data analysis process, and the resulting analytical findings.

Airbnb is an online platform that connects travelers with hosts, offering a variety of stays, experiences, and adventures. From charming homes to guided activities by local experts, it caters to both leisure and business travelers. With secure payments, verified profiles, 24/7 support, and AirCover protection for hosts, Airbnb prioritizes safety and convenience for its community (AirBnB, 2025a). Airbnb deducts a service fee from each booking, which is automatically taken from the host's payout. Most hosts pay a 3% split fee, where the cost is shared with guests, who usually contribute less than 14.2% of the booking subtotal. Alternatively, some hosts choose a host-only fee of 14% to 16%, which applies to hotels, serviced apartments, and those using property management software (AirBnB, 2025b).

## *1.2 Business Analytics Question*

For Airbnb, identifying the most promising areas for apartment rentals is crucial, as each booking contributes to their revenue. The more hosts list and successfully rent out their properties, the more Airbnb benefits financially. Understanding variations in booking demand and pricing across different neighborhoods and the factors

influencing these differences can help hosts optimize their listings, ultimately increasing their earnings while also boosting Airbnb's revenue.

To explore this, we will address the following analytics question:

**Which areas in New York City attract the most Airbnb bookings, what are the mean prices for each area, and what factors contribute to their popularity?**

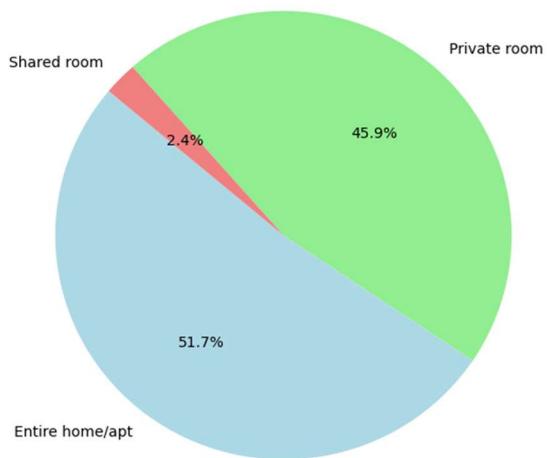
To answer this question, we will explore:

- Which boroughs and neighborhoods attract the most Airbnb guests?
- How does pricing vary across different locations?
- What factors influence booking demand (availability, reviews, pricing)?
- Applying machine learning techniques (K-Means clustering) which identifies the segment in the NYC's Airbnb market that currently fulfills the company's business objectives.

## 2. Data Analysis & Insights

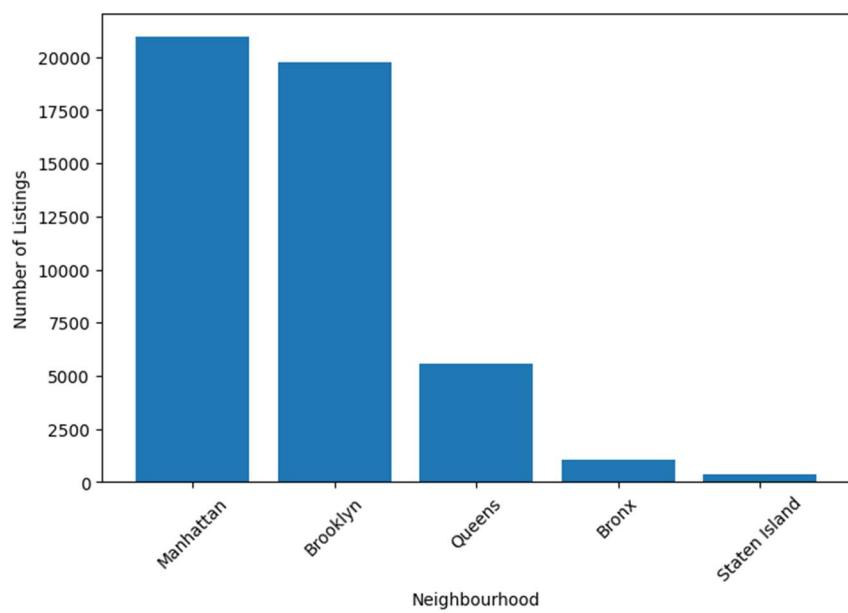
### 2.1 Room Type Distribution

The pie chart below shows the distribution of room types in NYC's Airbnb listings:



### 2.2 Distribution of Airbnb Listings by Borough

The bar chart below shows the number of listings across NYC boroughs:



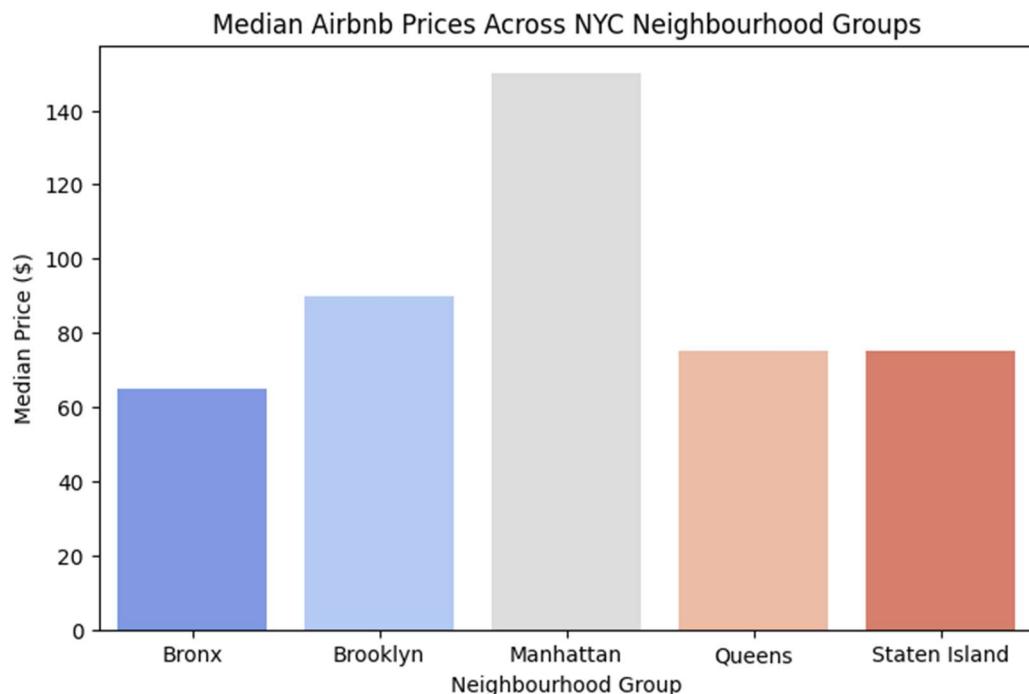
## Key Insights

- Manhattan and Brooklyn dominate Airbnb listings, each exceeding 20,000 properties.
- Queens has moderate listings (~5,000), while The Bronx and Staten Island have significantly fewer.

**Business Recommendation:** Expand the Airbnb marketing in the areas of Queens, Bronx, and Staten Island to attract new hosts and balance demand. This can be done by highlighting the benefits such as good transportation, local markets, good city views etc., that these areas can offer to travellers or to people who are looking for short term lettings.

### 2.3 Median Airbnb Prices by Borough

The bar chart below illustrates price variations by neighbourhood group:



## Key Insights

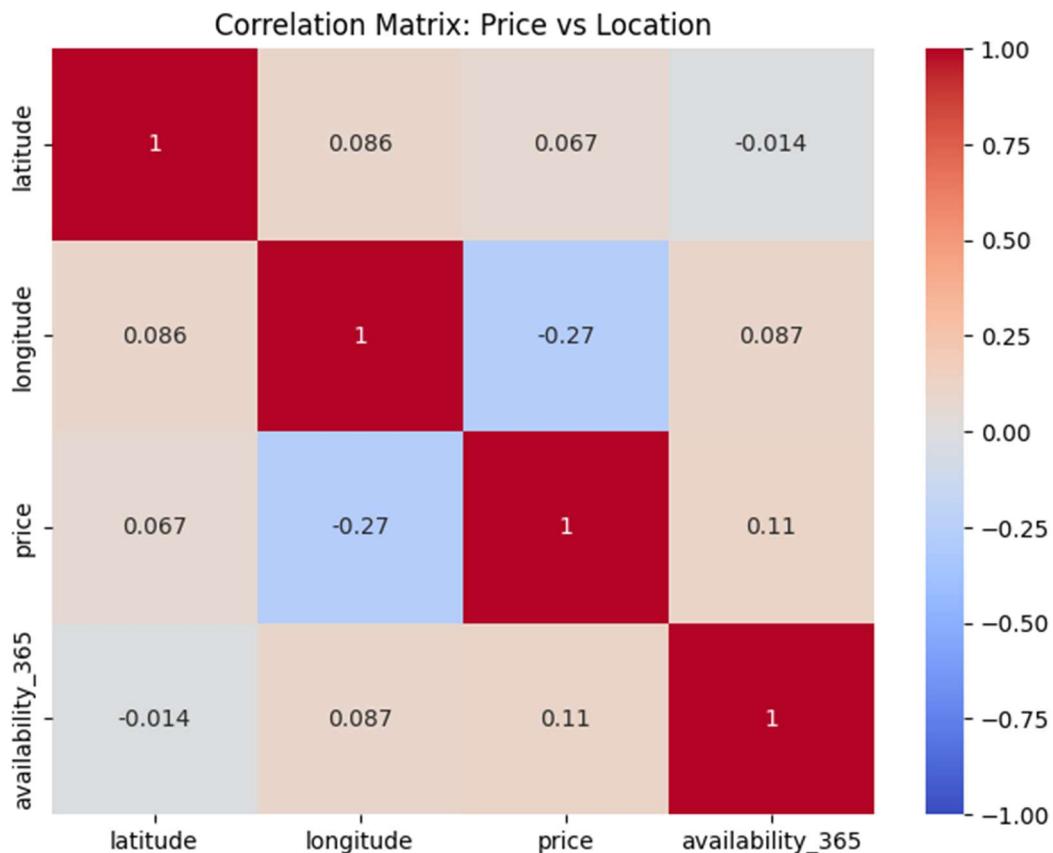
- Manhattan has the highest median price (~\$150).
- Brooklyn follows with ~\$90, making it a mid-range alternative.
- The Bronx, Queens, and Staten Island have the most affordable stays (\$75 - \$65).

### Business Recommendation:

- A. Use dynamic pricing models in Manhattan to maximize revenue during peak demand periods.
- B. Attract budget travellers to the Bronx & Queens by developing strategic promotions.

## 2.4 Correlation Analysis: Price vs. Location & Availability

The heatmap below shows correlations between price, location, and availability:



## Key Insights

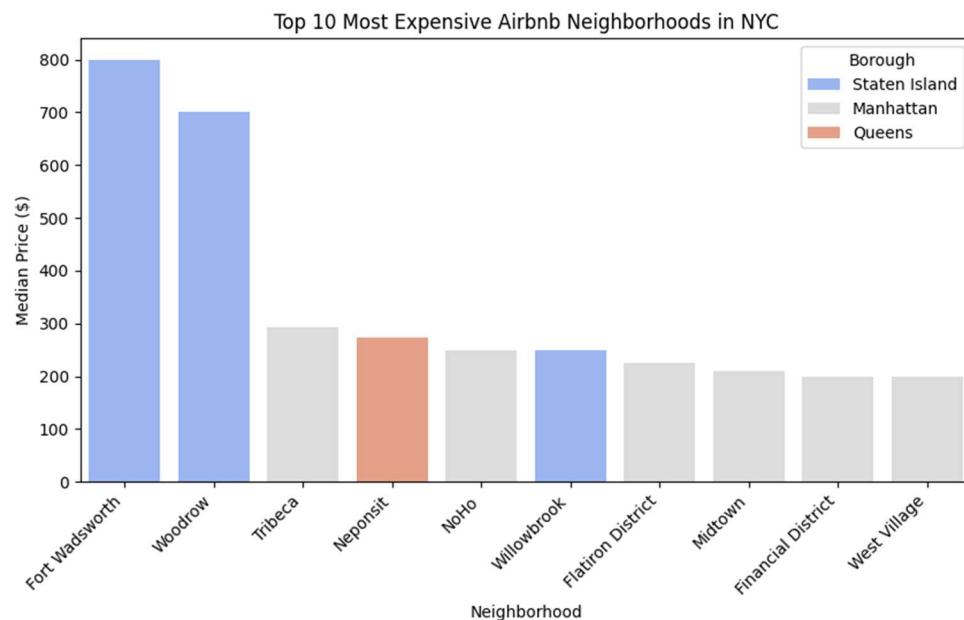
- Availability (0.11 correlation with price): Higher-priced properties are not significantly more available.
- Weak correlation between latitude/longitude and availability, suggesting demand is spread across multiple areas.

### Business Recommendation:

- A. The company needs to optimise its pricing strategies per location to maximize occupancy rates since the 11% positive correlation indicates possible further increase in the price won't necessarily lead to a drop in demand. This needs to be explored further at area/neighbourhood level.
- B. Encourage long-term stays in lower-priced listings to enhance revenue stability.

## 2.5 Top 10 Most Expensive Neighbourhoods

The bar chart below shows the top 10 most expensive neighbourhoods in NYC:



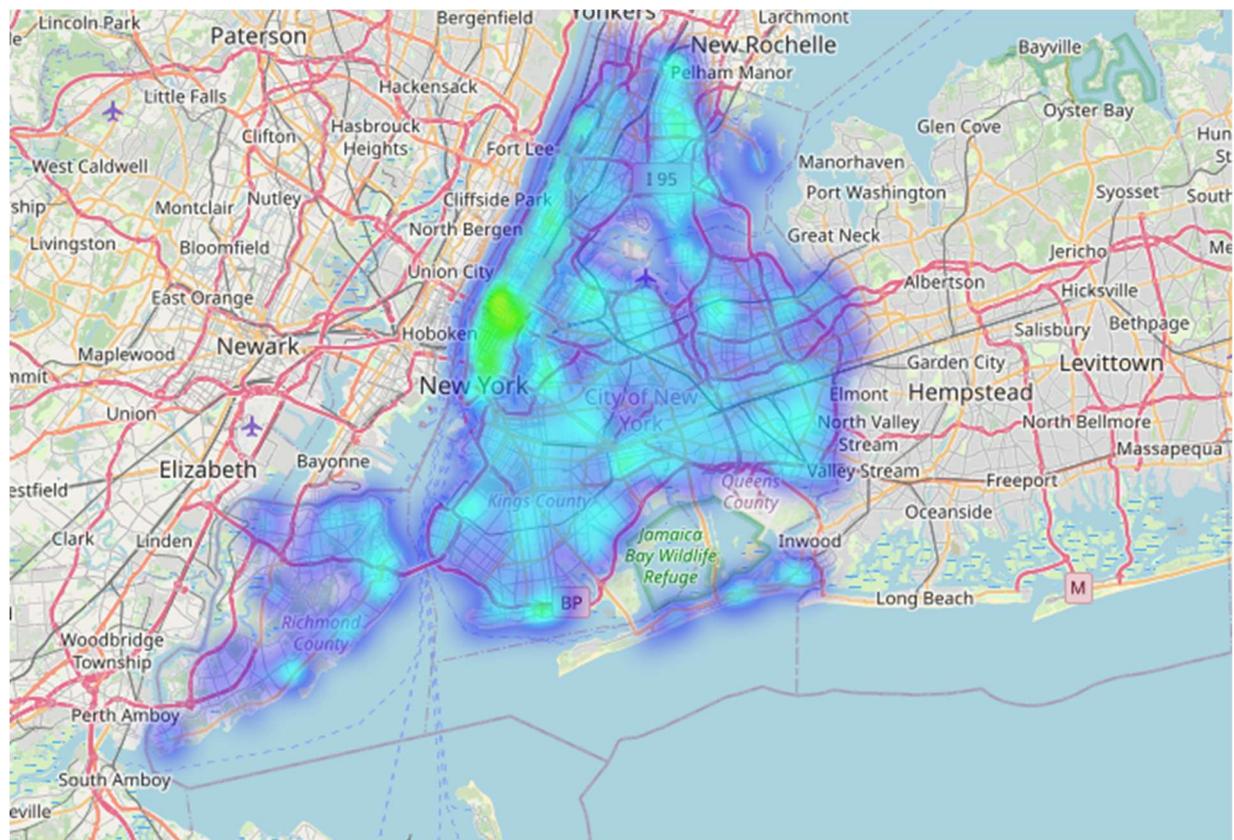
## Key Insights

- Fort Wadsworth & Woodrow (Staten Island) have the highest median prices (~\$700-\$800).
- Tribeca (Manhattan) and Neponsit (Queens) follow (~\$270-\$300).
- Staten Island's high-priced listings are likely due to premium properties with low competition.

**Business Recommendation:** The company needs to consider the promotion of traveller reviews and experiences in the high-end locations like Tribeca & Fort Wadsworth in order to increase bookings in premium locations.

## 2.6 Heatmap of Price and availability relation

The heatmap below depicts the price and availability relation on the NY City's map.



## Key Insights

- The heat higher heat shoes where price and availability are higher, the heat is grading lower when price and availability are lower.
- Manhattan and in particular properties in West Manhattan (closer to tourist spots) are more expensive and in high demand.

**Business Recommendation:** The company needs to consider expanding offerings in areas which show potential for expansion such neighbourhoods closer to transportation (e.g. JF Kennedy airport) where there are spots with positive correlation between prices and availability.

### 3. ML: K-Means Clustering for NYC Airbnb

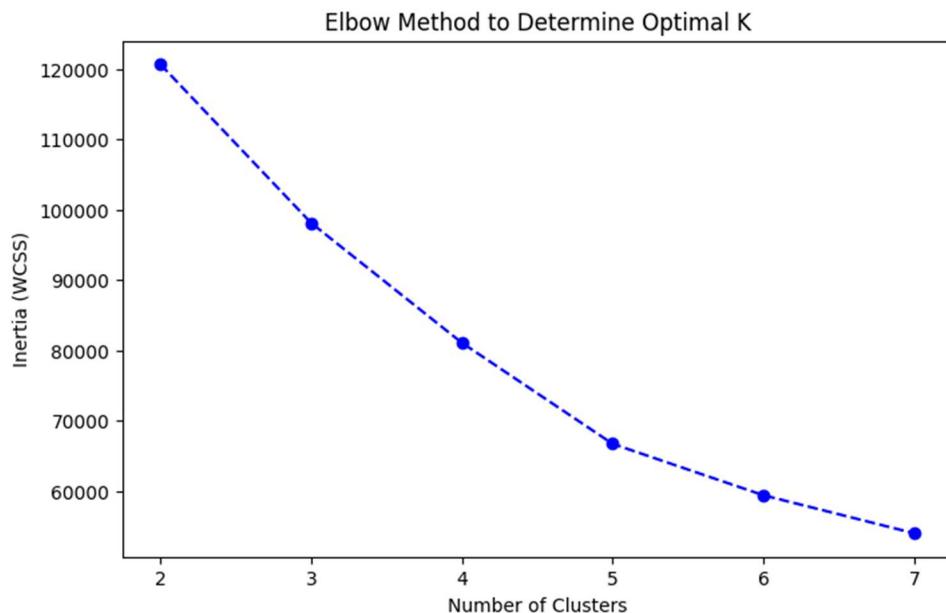
#### Market Segmentation

Using K-Means Clustering, we segment NYC Airbnb listings into four market categories based on:

- Latitude & Longitude (Geographical distribution)
- Availability (Booking frequency in days per year)
- Price (Revenue potential per listing)

##### *3.1 Determining Optimal Clusters Using the Elbow Method*

The Elbow Method graph below shows the best k value:

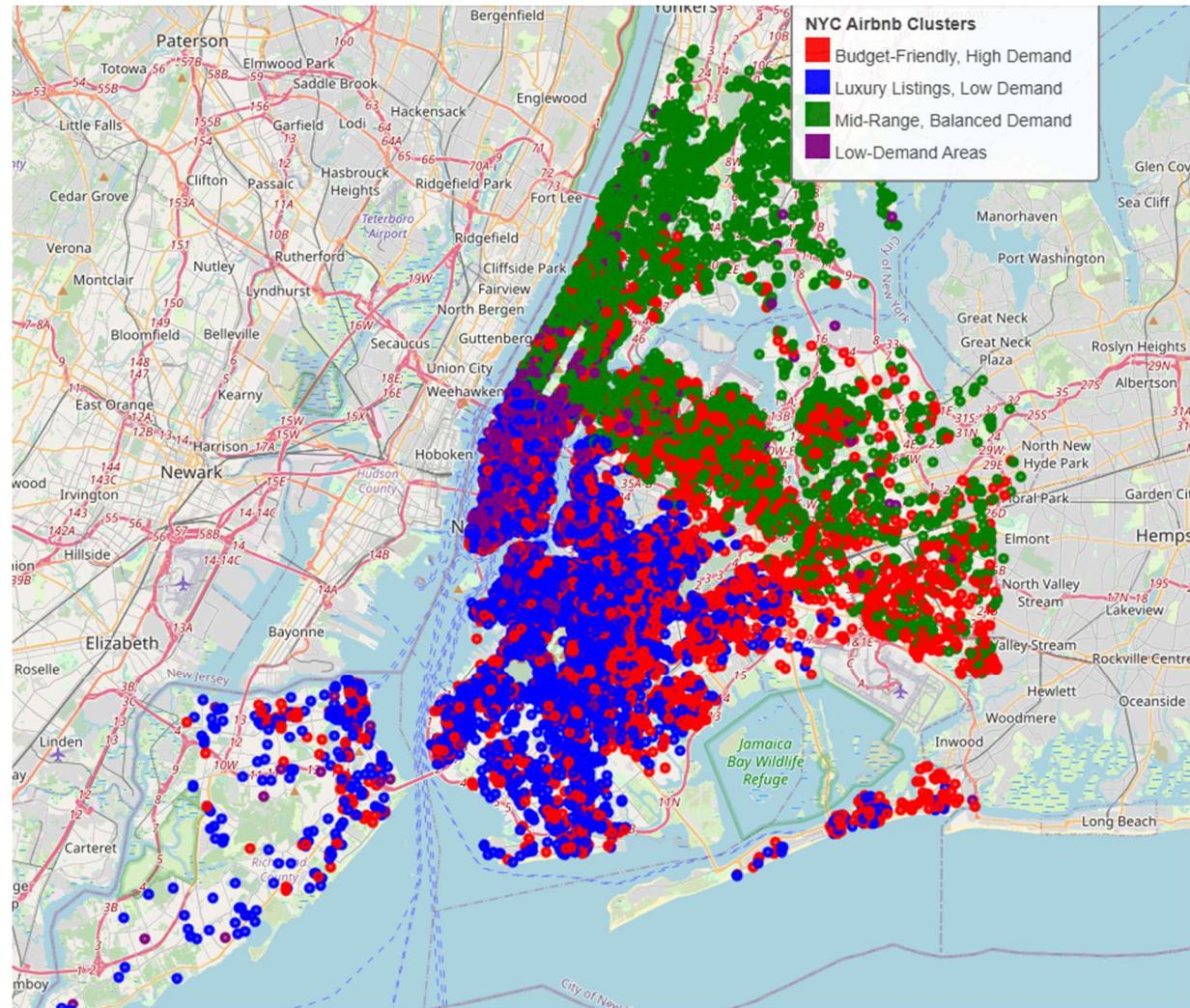


The Elbow Method is used to identify the optimal number of clusters by plotting the within-cluster sum of squares (WCSS, or inertia) against the number of clusters.

The optimal cluster count is typically at the point where the inertia reduction significantly slows down (the "elbow"). The optimal number of clusters is 4 (k=4).

### 3.2 NYC Airbnb Clusters Map

The interactive map below visualizes clustered Airbnb demand zones:



## Cluster Interpretations

Cluster	Colour	Price	Availability	Key Locations
0	Red	Low (~\$75-\$100)	High (~300 days)	Bronx, Queens
1	Blue	High (~\$250+)	Low (~50-100 days)	Manhattan (Luxury areas)
2	Green	Moderate (~\$120-\$180)	Balanced (~150-200 days)	Brooklyn, Queens
3	Purple	Very Low (~\$50-\$75)	Low (~30-100 days)	Staten Island and Manhattan, Low-Demand Areas

### Business Recommendation:

- Cluster 0 (Red) → Expand budget-friendly listings to meet high demand.
- Cluster 1 (Blue) → Target high-spending tourists with luxury experiences.
- Cluster 2 (Green) → Optimize mid-range pricing for steady revenue.
- Cluster 3 (Purple) → Promote long-term stays in low-demand areas to attract more bookings.

## 4. Final Business Strategy for Airbnb NYC

Goal	Recommended Action
<b>Maximize Bookings</b>	Expand budget-friendly listings in Cluster 0 (Red).
<b>Increase Revenue in High-Demand Areas</b>	Use dynamic pricing in Cluster 1 (Blue) (Manhattan).
<b>Improve Occupancy in Underserved Areas</b>	Promote Queens, Bronx, & Staten Island listings. Increase budget-friendly options to attract travelers.
<b>Enhance Long-Term Stay Strategy</b>	Offer discounts for monthly bookings in low-demand areas.
<b>Target High-Value Travelers</b>	Increase luxury property offerings in premium locations (Cluster1).

## 5. Conclusion

This report provides data-driven insights to help Airbnb optimize listings, pricing, and host strategies in NYC. The key takeaways from this analysis can be summarised as:

- Manhattan & Brooklyn dominate bookings but are expensive.
- The Bronx & Staten Island offer growth opportunities.
- Dynamic pricing can optimize revenue in high-demand areas.
- Clustering helps segment the market for targeted marketing.
- Airbnb should implement these strategies to maximize growth, revenue, and customer satisfaction.

# References

- Kaggle. (2019) AB\_NYC\_2019 [Dataset]. Available at: <https://www.kaggle.com/code/whyalwaysme/ab-nyc-2019> (Accessed: 15 February 2025).
- Airbnb. (2025a) About Airbnb: What it is and how it works. Available from: <https://www.airbnb.com/help/article/2503> (Accessed: 15 February 2025).
- Airbnb. (2025b) How much does Airbnb charge hosts? Available from: <https://www.airbnb.com/resources/hosting-homes/a/how-much-does-airbnb-charge-hosts-288> (Accessed: 15 February 2025).

# **Appendix**

The code can be found in the colab file handed in alongside this report.