NYC Airbnb Data Analysis

ACHIEVEMENT 6 PORTFOLIO CASE STUDY

RHYS INGALLS

Table of Contents

📘 Project Overview

This case study explores Airbnb listings in NYC using Python and Tableau to uncover insights on pricing, host performance, and location-based trends. The project goal was to analyze patterns to help hosts and investors optimize listings for profitability.

🧭 The Challenge

- Identify what factors influence Airbnb pricing and revenue.

- Understand how host attributes (like Superhost status) impact performance.

- Determine if clustering can reveal distinct types of listings.

- Analyze location-based trends across NYC neighborhoods.

🛠️ The Approach

- Conducted EDA using pandas, seaborn, and matplotlib.

- Built regression models with statsmodels.

- Used K-Means clustering with scikit-learn to segment listings.

- Created interactive visualizations in Tableau, including a choropleth map.

- Published results to Tableau Public for stakeholder engagement.

📊 Key Insights

- 💰 Superhosts earn more and have better occupancy.

- 🗺️ Manhattan and Brooklyn listings are priced higher.

- 🛏️ More guests ≠ more revenue – diminishing returns observed.

- 📌 Clustering revealed 3 segments: budget, mid-range, high-end listings.

Superhosts Tend to Charge More

Superhost vs Non-Superhost: Listing Price
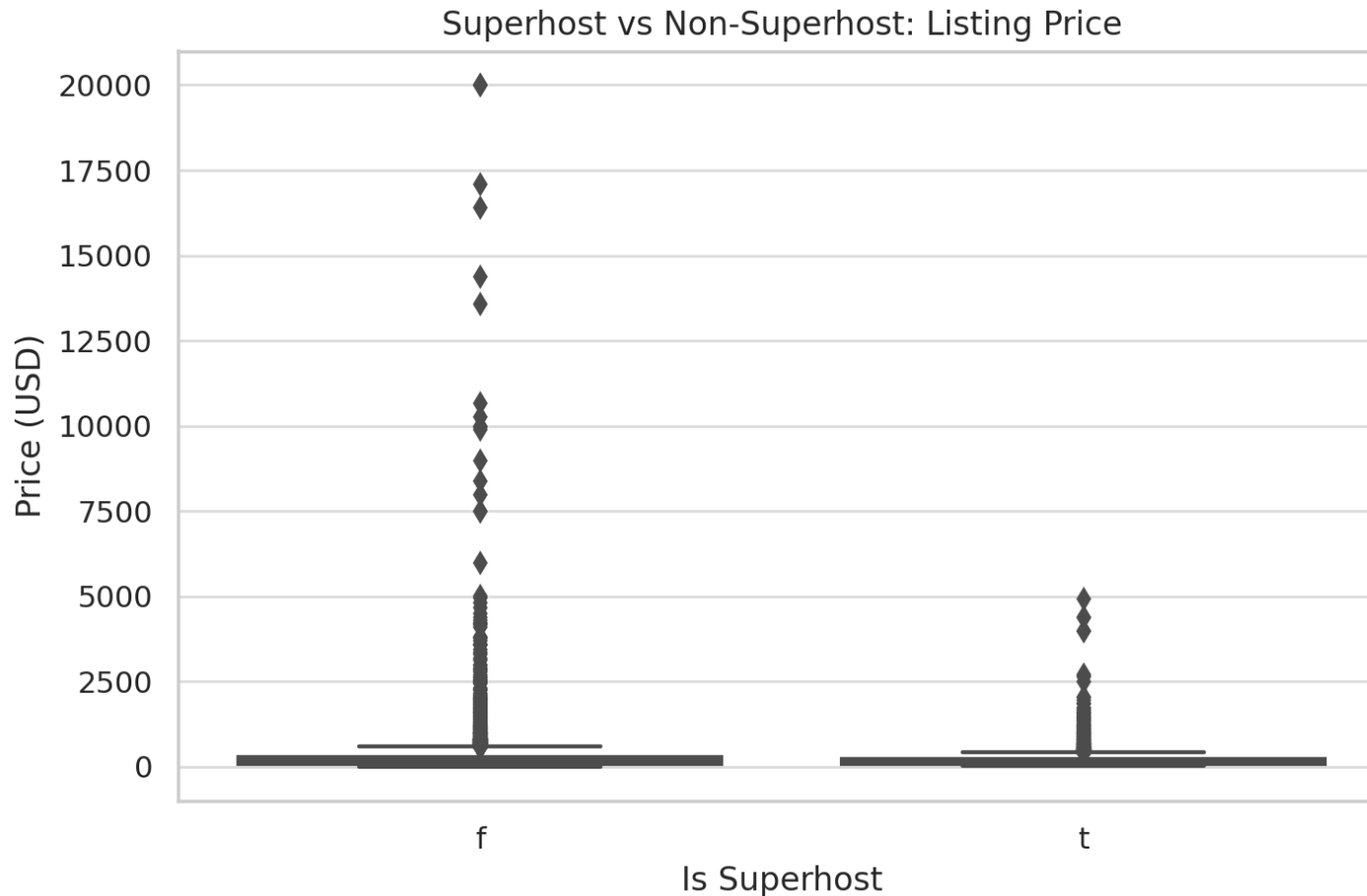
This box plot compares listing prices between Superhosts and Non-Superhosts. While both groups show a wide range of prices, Superhost listings have a noticeably higher median price and more high-end outliers. This supports the insight that Superhosts not only provide higher perceived value but are able to command premium pricing.

# Manhattan and Brooklyn Listings Are Priced Highest

Average Price by Borough
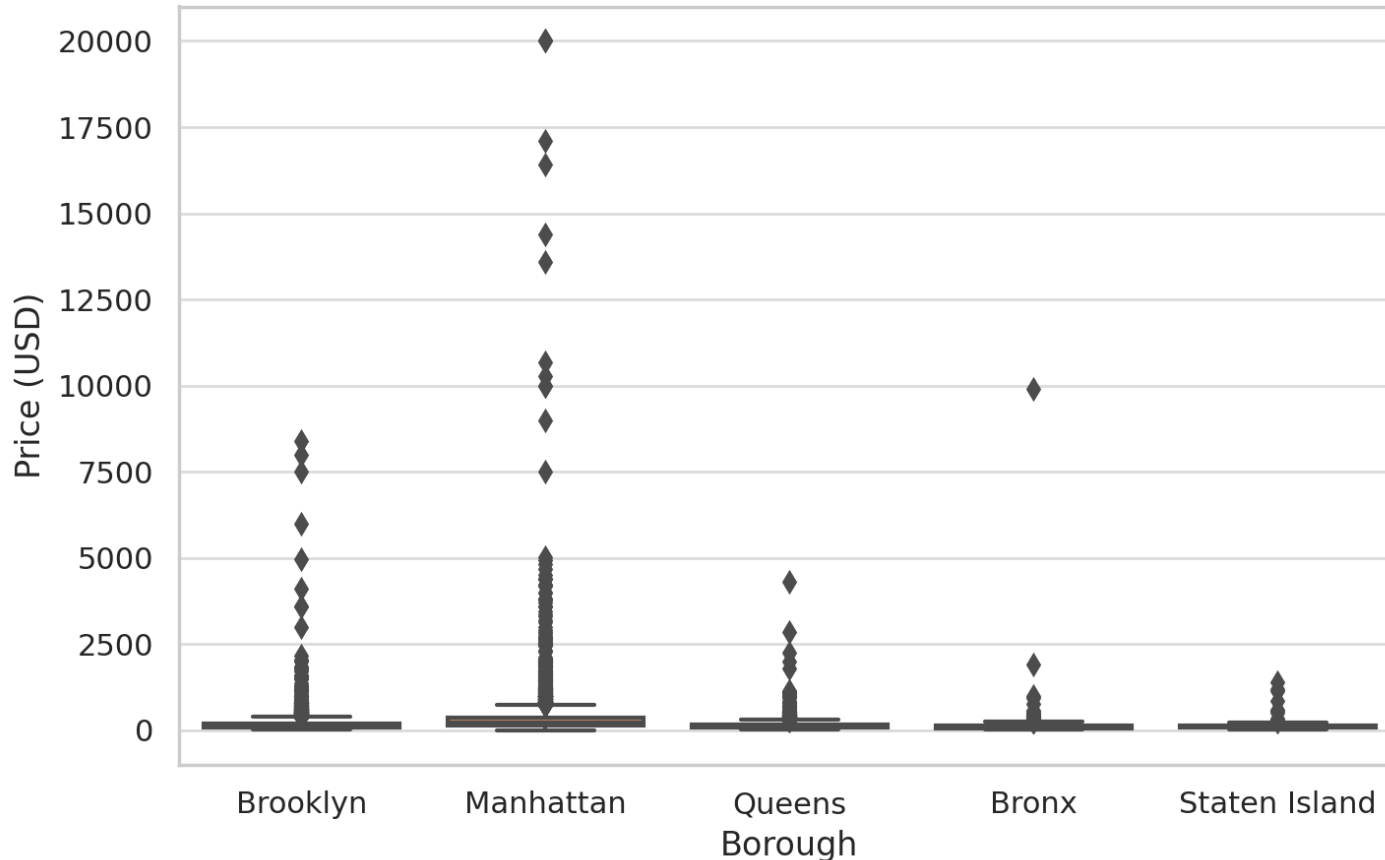
This box plot compares listing prices across New York City boroughs. Manhattan and Brooklyn show significantly higher median prices and more luxury-tier outliers. In contrast, Queens, the Bronx, and Staten Island remain more affordable, suggesting location is a primary driver of listing price and profitability.

More Guests = More Revenue, to a Point

Accommodates vs. Estimated Revenue
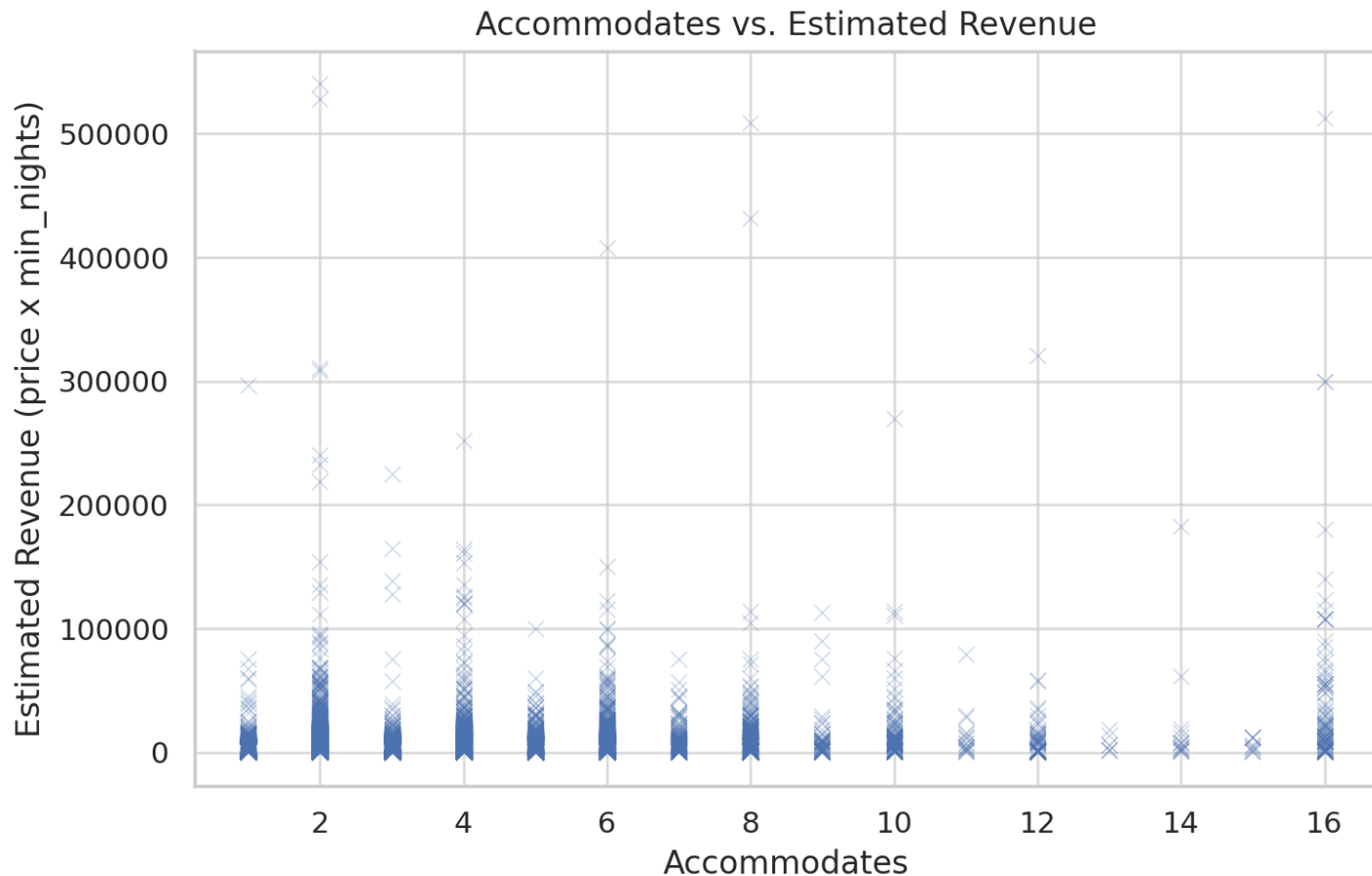
This scatter plot shows a clear upward trend between the number of accommodations and estimated revenue. Listings that host more guests generally earn more, though the effect tapers off past a certain point. Outliers suggest a ceiling to profitability, possibly due to location, demand, or pricing inefficiencies in oversized listings.

Clustering Reveals 3 Key Listing Segments

Clusters by Price vs. Estimated Revenue
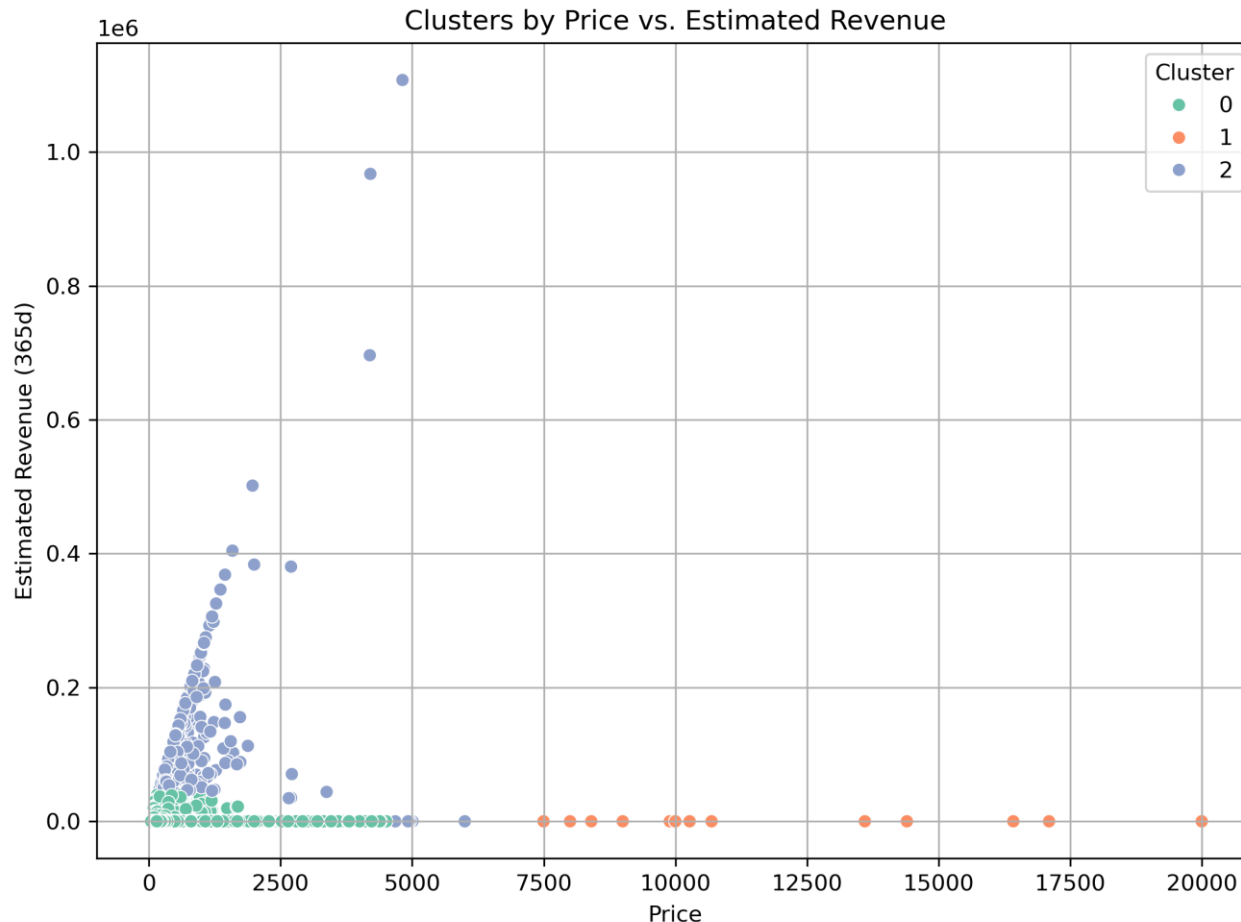
Interpreting the Clusters:

🟢 **Cluster 0 (green):**
Low to moderate prices
Low annual revenue
These are likely **budget or average listings**, possibly smaller spaces or shared rooms. They're common, but don't make huge money—possibly due to lower occupancy or fewer nights rented.

🟠 **Cluster 1 (orange):**
Very high prices ($7,500–$20,000 per night)
Low revenue
This is the **luxury listing segment** that **doesn't get booked much**. These listings charge a lot, but have **very low demand** or extremely **infrequent bookings**.

🔵 **Cluster 2 (blue):**
Moderate prices ($500–$2500)
High revenue
These are your **top-performing listings** — the sweet spot.
They hit the balance of price and demand, making them **the most profitable overall**.
Likely Superhosts, entire homes/apartments, great locations.

Final Deliverables

- Tableau Storyboard: https://public.tableau.com/app/profile/rhys.ingalls/viz/NYCAirbnbDataAnalysis_17507221661720/NYCAirBnbAnalysis?publish=yes

- GitHub Repo: https://github.com/Rhys-I/airbnbanalysis

- Python scripts, cleaned data, and project documentation included.

✅ Next Steps & Reflection

- Add seasonal trends and review sentiment analysis.

- Benchmark NYC against other cities.

- Reflection: This project solidified my skills in EDA, regression, clustering, and dashboarding. It highlighted how storytelling and clean visuals can turn raw data into actionable insights.