What Drives Airbnb Prices in NYC?

Data Visualization Project

Florence De León

1. Introduction

1.1 Dataset Overview

For this project, I reviewed a dataset of my choice to visualise how different variables can affect potential outcomes. This project uses the "New York City Airbnb Open Data" dataset, which contains information on 48895 Airbnb locations listed in NYC in 2019. This dataset is an open dataset available on Kaggle on the following link: https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data, and it has been accessed on the 1st of April 2025.

In the table below, you can see a detailed explanation of each of the 16 variables in the dataset.

Table 1

Table of Variables NYC Airbnb Dataset

Variable Names	Type	Description
ID	Integer	Identification of Airbnb
`Name	Text	Name of the accommodation
Host ID	Integer	Identification of host
Host Name	Text	Name of the host
Neighborhood Group	Text	Main neighborhood group
Neighborhood	Text	Detailed neighborhood
Latitute	Integer	Latitude
Longitude	Integer	Longitude
Room Type	Categorical	Whether it's a private or shared apartment
Price	Integer	Airbnb price per night in USD
Minimum Nights	Integer	Minimum nights of stay
Number of Reviews	Integer	Total number of reviews of the accommodation
Last Review Date	Date	When was the last review published
Review per Month	Integer	How many reviews per month it received
Calculated Host Listing Counts	Integer	How many other listings does the host have
365 Days Availability	Integer	For how many days is it available during the year

1.2 Motivation

The motivation for this study is to identify the key variables that most significantly influence Airbnb prices across New York City. To achieve this, a series of exploratory analyses were conducted, where Airbnb prices were plotted against various variables to uncover potential insights.

Given the large volume of data in the dataset, it is worth noting that, at times, the data was filtered to focus on the most prominent neighbourhood groups in NYC. This approach helped reduce visual clutter and allowed for clearer identification of the main trends.

2. Methodology

This project employed an exploratory data analysis (EDA) approach to investigate the factors influencing Airbnb prices in New York City. The dataset used was the New York City Airbnb Open Data from 2019, publicly available on Kaggle. It includes 48,895 listings and a range of variables such as location, price, room type, minimum stay requirements, and number of reviews.

The analysis was conducted using Python 3.13.2. In addition, visualisations were created using a combination of Python (Pandas, Matplotlib, and Seaborn) and Tableau to uncover patterns and enhance interpretability. Logarithmic transformation was applied to price data in select plots to reduce skewness and better visualise distribution. Each variable was analysed independently to highlight its potential influence on nightly price, without implementing predictive modeling or multivariable statistical testing.

The analysis began with an overview of listing distribution across New York City's five boroughs. Based on the concentration of listings and price variability, the focus was narrowed to Brooklyn and Manhattan. Further analyses were performed within Manhattan to explore how room type, neighborhood, and minimum stay duration relate to nightly price. Prices were log-transformed in several visualisations to normalise distribution and improve interpretability.

Each analysis was conducted independently, with visual outputs used to identify patterns and trends. No predictive models were developed, as the aim was to uncover descriptive insights rather than infer causal relationships.

3. Data Analysis & Key Findings

3.1. Distribution of Accommodations per Neighbourhood Group

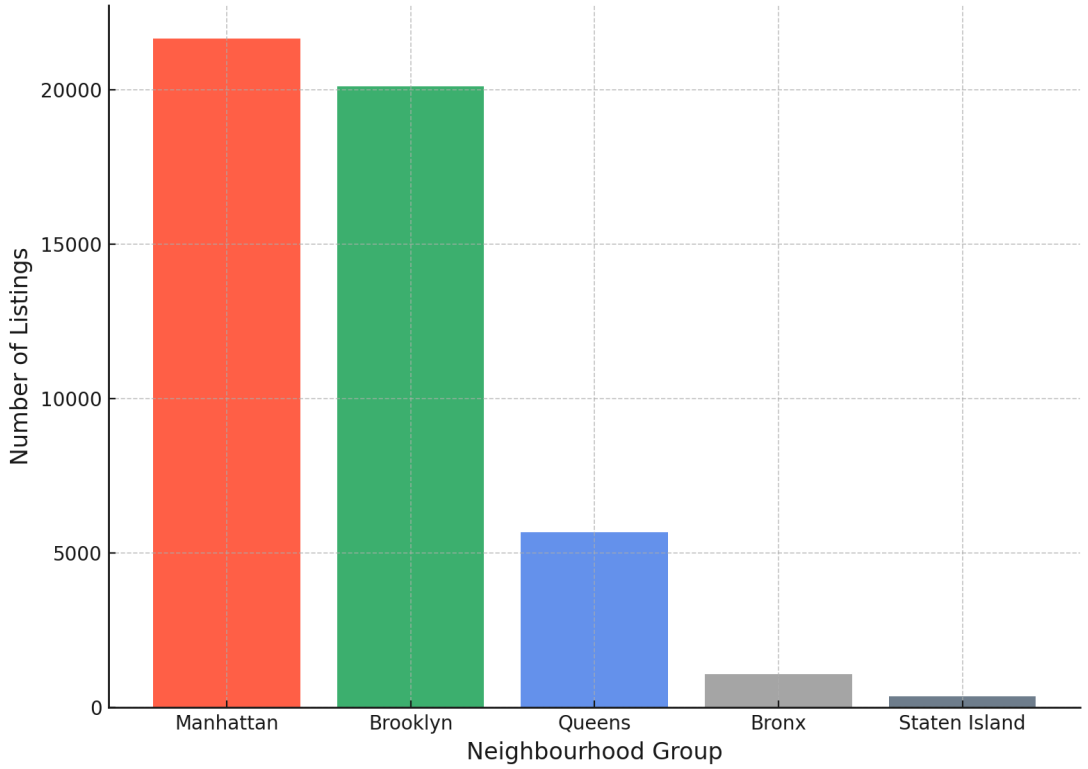
Given the high volume of data, it is useful to first examine how accommodations are distributed across different neighbourhood groups in New York City. This step provides context for understanding broader pricing trends and identifying where to focus deeper analysis.

The initial plot (Figure 1) shows that Manhattan and Brooklyn have, by far, the highest number of available Airbnb listings. Based on this observation, the analysis will focus primarily on these two neighbourhood groups in the following sections.

This focus enables a more detailed exploration of how other variables, such as the number of reviews, room type, and minimum stay requirements, may influence nightly accommodation prices.

Figure 1

Number of Listings by NYC Neighborhood Group



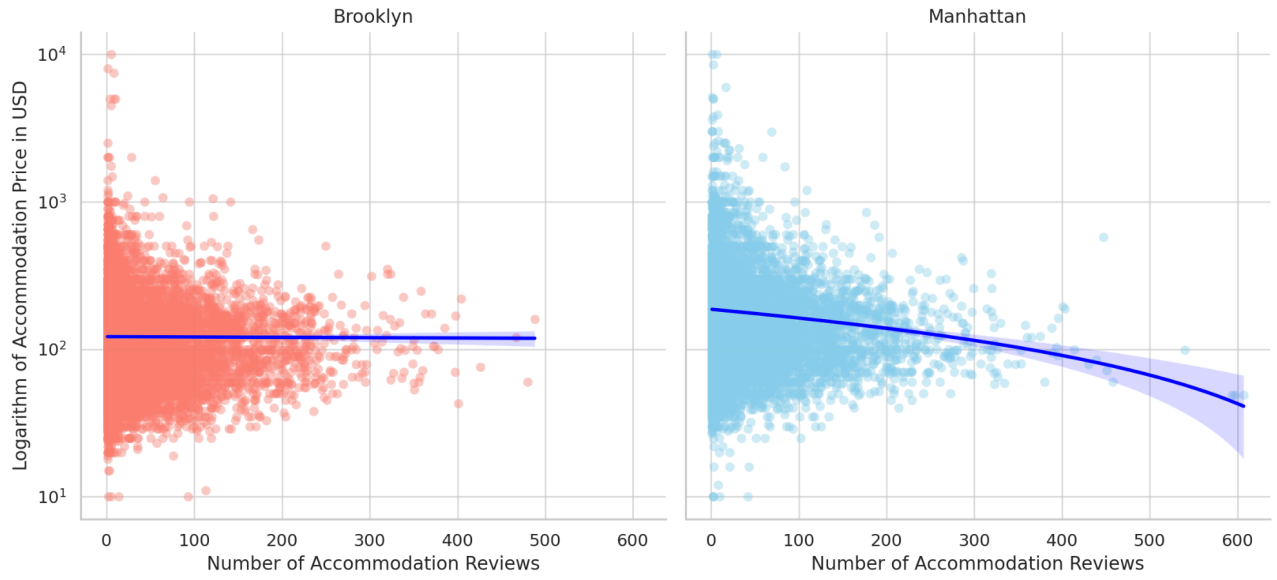
Note. This bar chart displays the number of Airbnb listings in each New York City neighborhood group as of 2019. Manhattan and Brooklyn have the highest concentration of listings, followed by Queens, the Bronx, and Staten Island.

3.2 The Relationship Between Number of Reviews and Accommodation Price

Intuitively, it seemed that people preferred Airbnbs with a higher number of reviews, as reviews could serve as a reliable indicator of the accommodation's quality. Based on that assumption, the analysis investigated whether the number of reviews for an Airbnb listing had any impact on its price and whether people were willing to pay more for listings with more reviews.

Figure 2

Logarithm of Airbnb Prices in USD by Review and Neighbourhood



Note. This figure compares Airbnb prices in Brooklyn and Manhattan based on the number of accommodation reviews. Prices are shown on a logarithmic scale.

Based on this plot, the number of accommodation reviews does not appear to be a factor that increases the price of an Airbnb, as most listings with a high number of reviews (over 100) fall within the mid-price range. The graphic also shows that prices in Manhattan exhibit greater variance than those in Brooklyn, with more outliers concentrated in the highest price range.

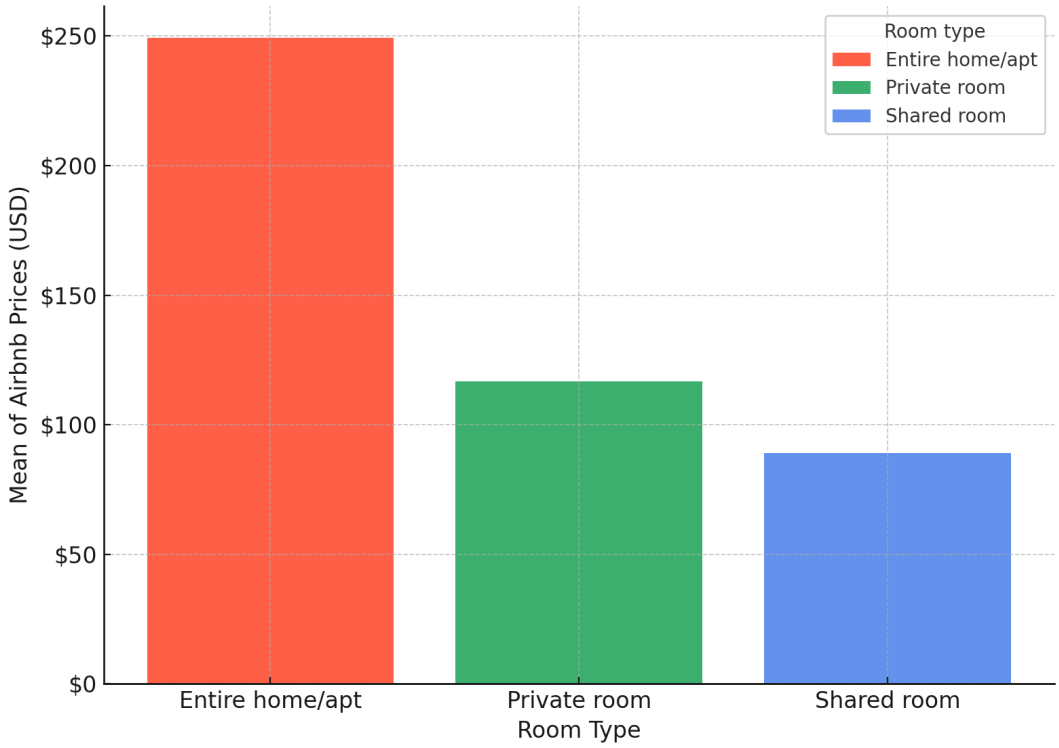
Moving forward, to continue exploring which variables may affect Airbnb pricing, the analysis will focus solely on the Manhattan neighbourhood group.

3.3 The Relationship Between Room Type and Price

Another common assumption about Airbnb accommodations is that people are willing to pay more for a space they can have entirely to themselves. To examine whether this holds true, a plot was created comparing accommodation types and their average prices, focusing exclusively on the Manhattan area.
The plot clearly shows that listings offering an entire home or apartment in Manhattan have significantly higher average prices than those offering shared spaces, such as private or shared rooms. On average, entire accommodations charge nearly twice as much. In contrast, the price difference between private and shared rooms appears relatively small. It is worth noting that this graph presents average prices and does not account for variability within each category or other influencing factors such as location or amenities.

Figure 3

Mean Price of Airbnb Listings in Manhattan by Room Type



Note. Prices reflect average nightly rates for Airbnb listings located in Manhattan as of 2019.

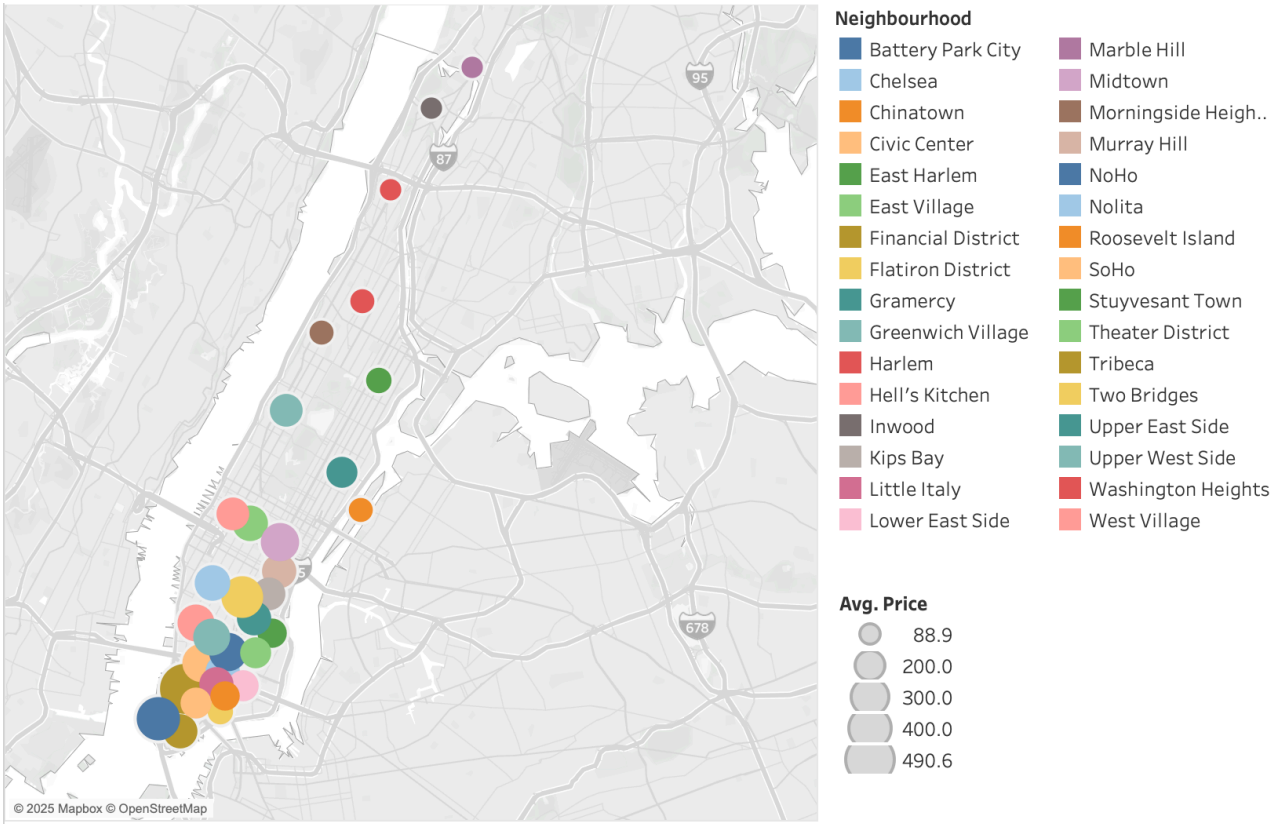
3.4 The Relationship Between Neighbourhood and Price

Up to this point, the analysis has focused exclusively on neighbourhood groups, which represent broader areas within New York City. However, each group is made up of several smaller, more specific neighbourhoods. This section explores whether the specific city area in which an accommodation is located has an effect on its final price.

In this Manhattan area map, it is evident that the majority of neighbourhoods with the highest mean price per night are located in the southern part of the borough. This trend may be influenced by the concentration of accommodations in that area and their proximity to popular tourist attractions. In contrast, the northern part of Manhattan shows significantly lower Airbnb prices per night, suggesting there may be less demand for stays in that region.

Figure 4

Airbnb Listing Distribution and Average Prices by Manhattan Neighborhood in 2019



Note. This map visualises the geographic spread and average nightly prices of Airbnb listings in Manhattan. Each circle represents a neighborhood, with size indicating average price and color corresponding to neighborhood name.

3.5 The Relationship Between Minimum Stay Duration and Airbnb Price in NYC

Lastly, this section examines whether the minimum number of nights per stay has an impact on the price per night for Airbnb listings in Manhattan. This variable represents a restriction that hosts can set when listing their property on the platform.

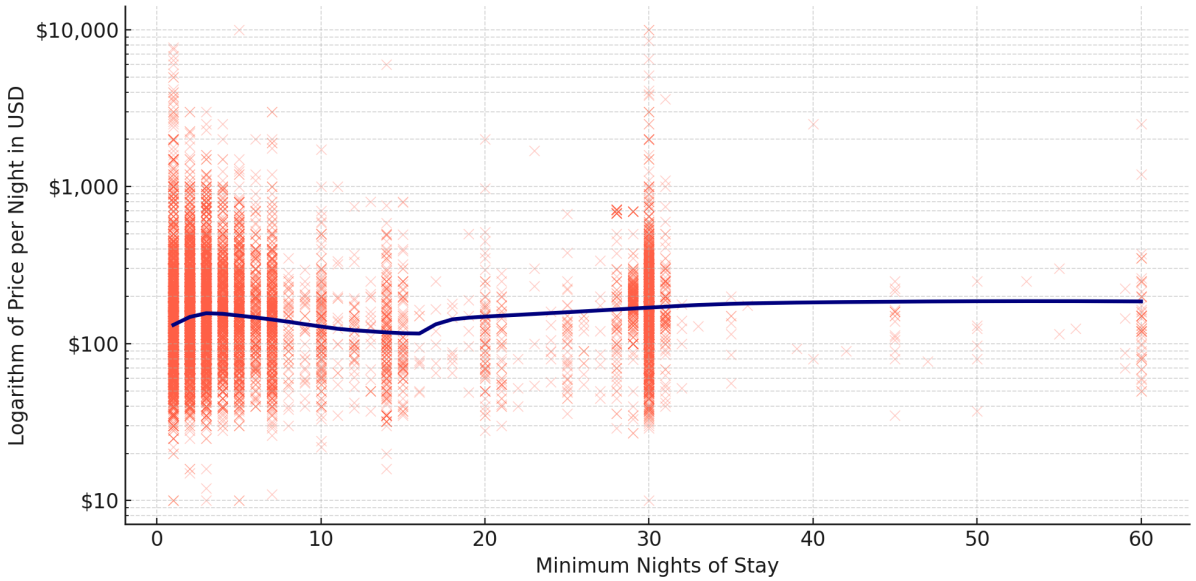
To improve data clarity, listings with a minimum stay requirement exceeding 60 nights were excluded, as these are considered outliers. This adjustment resulted in a filtered dataset of 21,414 listings. The upper limit was set because it is unlikely that hosts are using Airbnb to offer long-term rental contracts (over two months), given the platform's primary focus on short-term, tourist-oriented stays.

According to Figure 5, a vast majority of listings require fewer than 15 nights as a minimum stay. Interestingly, higher prices are often observed among listings with no minimum stay requirement. As the minimum stay increases—for example, to 10 nights—average prices appear to decrease slightly. Another notable trend is the significant concentration of listings

requiring a minimum of 30 nights. In this range, prices tend to be higher than those for listings with no minimum stay requirement.

Figure 5

The Relationship Between Logarithmic Price per Night and Minimum Stay Requirement for Airbnbs in Manhattan



Note. This scatter plot displays the relationship between the logarithm of the nightly price (in USD) and the minimum number of nights required for a stay. Each point represents a listing, and the blue line shows a smoothed trend across the data. Listings with lower minimum night requirements appear to show greater price variability.

4. Final Remarks

The objective of this project was to investigate which variables may influence Airbnb prices in New York City, USA. One of the primary factors identified is the neighbourhood group in which the listing is located. A higher concentration of accommodations was observed in Brooklyn and Manhattan, with the highest prices found in Manhattan.

The number of reviews does not appear to significantly impact pricing, as listings with the most reviews were typically in the mid-price range. Further analysis within the Manhattan neighbourhood group revealed that entire home/apartment listings command substantially higher prices than shared accommodations, including private and shared rooms. Among these, shared rooms represent the lowest price category.

Still focusing on Manhattan, higher Airbnb prices tend to cluster in the southern part of the borough. Although beyond the main scope of this analysis, this pattern may be linked to the proximity of major tourist attractions in that area.

Lastly, the relationship between the minimum number of nights required for a stay and the nightly price was examined. While no strong trend emerged, some insights can be drawn.

The highest individual prices were found among listings with no minimum stay requirement. However, the highest average prices occurred in listings that required a minimum stay of 30 nights. This may indicate that hosts offering 30-night minimums target longer-stay guests, such as business travelers or temporary residents, who typically have higher budgets. Overall, location and accommodation type appear to be the most significant drivers of Airbnb prices in New York City.

5. Limitations

While this analysis provides insights into factors that may influence Airbnb pricing in New York City, several limitations should be noted:

- Temporal scope: The dataset reflects listings from 2019 only. Market dynamics may have changed since then, particularly due to the COVID-19 pandemic and its impact on travel behavior.
- Seasonality: The data does not account for seasonal price fluctuations, which can significantly affect nightly rates depending on the time of year.
- Missing variables: Other influential factors such as listing amenities, cleanliness ratings, guest satisfaction scores, or proximity to landmarks were not included in the analysis.
- Aggregated data: Some analyses are based on average values and do not capture the full variability within each category (e.g., room type or neighborhood).
- Causality: The analysis is exploratory and descriptive. No causal relationships can be inferred from the observed correlations or trends.

These limitations should be considered when interpreting the findings. Future analyses could incorporate additional variables and updated datasets to offer a more comprehensive understanding of Airbnb pricing dynamics.