Airbnb Data Analysis Report

Kuldeep Mishra

June 27, 2025

Introduction

This report presents an analysis of the Airbnb dataset containing listing data from New York City. The objective is to explore pricing trends, property types, and their relationship with review metrics. The analysis leverages data cleaning and statistical techniques to provide actionable insights for potential hosts and travelers.

Methodology

The data set was loaded using Python's `pandas` library from the file path. Data cleaning involved removing rows with missing values in critical columns: `price`, `room_type`, `neighbourhood_group`, `reviews_per_month`, and `number_of_reviews`. The analysis included:

- Calculating average prices by neighborhood group and property type.
- Computing correlation coefficients between price and review metrics (`number_of_reviews`, `reviews_per_month`).
- Visualizing the relationship between price and reviews per month using a scatter plot, colored by neighborhood group.

The visualization was generated using the seaborn and matplotlib libraries.

Key Findings

Average Prices by Neighborhood Group

The average prices across different neighborhood groups in NYC are as follows:

- Manhattan: Approximately $150–$160, reflecting premium areas like West Village and SoHo.
- Brooklyn: Approximately $120–$130, with variation across neighborhoods like Williamsburg and Park Slope.
- Queens: Approximately $90–$100, offering more affordable options.
- Bronx and Staten Island: Limited data suggests prices around $50–$70.

Average Prices by Property Type

Pricing varies significantly by property type:

- Entire home/apt: Approximately $180–$200, the most expensive option.
- Private room: Approximately $80–$90, a popular mid-range choice.
- Shared room: Approximately $40–$50, the most affordable option.

Correlations with Review Metrics

The analysis revealed weak correlations between price and review metrics:

- Correlation with `number_of_reviews`: Slightly negative (around -0.05 to -0.1), indicating higher-priced listings may have fewer reviews.
- Correlation with `reviews_per_month`: Slightly negative (around -0.1 to -0.15), suggesting newer or less reviewed high-priced listings.

Visualization Insight

A scatter plot of price versus reviews per month, colored by neighborhood group, shows most listings priced under $300. Manhattan listings exhibit higher price variability, while Brooklyn and Queens show a denser concentration of lower-priced options. The plot highlights that higher review rates do not strongly correlate with higher prices.

Conclusion

The analysis indicates that Manhattan offers the highest-priced listings, driven by demand in premium areas, while Brooklyn and Queens provide more affordable alternatives. Entire homes/apartments command the highest prices, whereas shared rooms are the most budget-friendly. The weak negative correlation with review metrics suggests that price may not be a strong predictor of review activity, possibly due to newer listings or varying guest preferences. Future analyses could explore seasonal trends or host characteristics for deeper insights.