NYC Airbnb Analysis

Ms. Manasi Patil

Ms. Neha Kedar

Anudip Foundation

Under The Guidance of

Ajay Swarnkar

Data Analytics Using Python

TABLE OF CONTENTS

I. INTRODUCTION

Airbnb has revolutionized the travel and hospitality industry by enabling individuals to rent out their homes or spare rooms. New York City, being one of the most popular tourist destinations in the world, has a large number of Airbnb listings. This analysis provides critical insights for hosts to optimize their listing strategies and for guests to make informed accommodation choices. City planners and regulators can also benefit from understanding the distribution and concentration of listings.

This report provides an exploratory data analysis (EDA) of Airbnb listings in New York City. The objective is to understand the distribution, pricing, and trends of listings across different boroughs. The dataset includes various features such as location, price, room type, number of reviews, and availability. This analysis is crucial for potential hosts, This project employs powerful Python libraries such as Pandas, Numpy, Matplotlib, and Seaborn to conduct data cleaning, visual exploration, and insightful analysis. The data provides a rich view into the NYC rental landscape, allowing us to:

- Understand the distribution of listings across boroughs.
- Investigate pricing anomalies and outliers.
- Explore availability patterns and review behaviors.
- Analyze host performance and identify professional hosting trends.

The primary objective is twofold:

1. **For Guests –** Help renters discover affordable and well-reviewed listings by understanding which neighborhoods offer the best value and availability.

2. **For Hosts –** Provide data-driven recommendations to optimize pricing, improve availability, and boost guest satisfaction.

This project not only helps to make more informed decisions for short-term rentals but also lays the groundwork for future enhancements, such as price prediction, sentiment analysis of reviews, and interactive dashboards for real-time insights.

II. DATA UNDERSTANDING

The dataset used in this analysis was sourced from the Inside Airbnb website. It includes over 48,000 listings with detailed attributes. Key features such as price, number of reviews, location coordinates, and availability provide a comprehensive view of the Airbnb landscape in NYC. Understanding the data is essential before drawing conclusions or making recommendations.

The dataset contains several columns, including:

-ID,Name,HostID, Host Name
- Neighborhood Group, Neighborhood
- Latitude, Longitude
- Room Type, Price, Minimum Nights
- Number of Reviews, Availability 365

Initial data inspection reveals some missing values and outliers that need to be addressed. Categorical and numerical data types are identified for further analysis.

III. DATA CLEANING

Prices over $1,000 were considered outliers and were filtered out to prevent skewing the analysis. Missing host names and review counts were handled using median imputation or row removal depending on the severity. The room type column had consistent categories and was one of the main predictors in pricing analysis.

The cleaning process included:

- **Handle missing data:** \`price\`, \`neighborhood\`, and \`beds\` columns had null values.
- **Fix data types Converted :**\`last_review\` to a datetime object.
- **Remove outliers:** Listings with prices > $1,000 were capped to avoid skewed visualizations.

Ensuring data types were appropriate for analysis Post-cleaning, the dataset was more consistent and ready for meaningful visualizations and insights.

IV. UNIVARIATE ANALYSIS

A frequency distribution of room types revealed that private rooms make up a significant portion of listings, especially in Brooklyn and Queens. The distribution of price values shows a long tail, suggesting a few listings with extremely high prices, which is typical in real estate data. Review count distribution also showed that a small number of listings account for a large number of reviews.

This section examines individual features:

- **Room Type:** Entire home/apt, Private room, Shared room
- **Neighborhood Group:** Manhattan, Brooklyn, Queens, Bronx, Staten Island
- **Price:** Distribution is right-skewed with some extreme values
- **Number of Reviews:** Reveals listinhttps://meet.google.com/osq-msfe-snxg popularity and engagement

V. GEOSPATIAL ANALYSIS

Neighborhood clustering is visible in popular areas like Midtown Manhattan and Williamsburg. These regions have both a high density of listings and elevated price points. Geospatial analysis was further enhanced using interactive Folium maps and Seaborn heatmaps to visualize availability and pricing variations.Listings were plotted on a scatter map using latitude and longitude. Color-coding was used to indicate price and room type.

Key findings:

- Manhattan has the highest concentration of listings
- Certain neighborhoods show clustering of high or low prices
- Heatmaps and interactive maps offer deeper location insights

VI. BIVARIATE ANALYSIS

Listings in Manhattan consistently show higher average prices across all room types. Shared rooms, although least expensive, are also the least common. Availability trends showed that listings with extremely high availability are often part of professional rental businesses rather than casual hosts.

Relationships between pairs of variables were analyzed:

- **Price vs. Neighborhood Group:** Manhattan is the most expensive
- **Price vs. Room Type:** Entire homes are pricier than private rooms
- **Reviews vs. Price:** No strong correlation observed
- **Availability vs. Price:** Seasonal trends may exist

Boxplots and scatter plots were used for visualization.

VII. INSIGHTS & OBSERVATIONS

Analysis also found that super hosts (not explicitly listed but inferred from high review counts and consistent availability) tend to have better pricing strategies. Areas such as Harlem and the Bronx offer hidden gems with good value for money. Some neighborhoods with many listings had surprisingly low average review scores, which may point to quality issues or inflated expectations.

1. Price Trends:

- Manhattan has the most expensive listings, followed by Brooklyn.

- Entire homes/apartments cost significantly more than private or shared rooms.

2. Room Type Distribution:

- Entire homes/apartments are the most common, but private rooms offer budget-friendly options.

3. Outliers in Price:

- Few listings priced at $10,000+ were detected, indicating the need to filter such extreme values.

4. Availability Patterns:

- Listings with high availability tend to have lower prices and more reviews, likely due to better guest experience.

5. Host Behavior:

- Some hosts manage multiple listings, indicating a trend toward professional hosting.

VIII. RECOMMENDATION

Dynamic pricing tools could help hosts remain competitive while maximizing revenue. Guests are encouraged to book early in high-demand areas and explore beyond Manhattan for better rates. The city could use this data to balance tourism growth with residential needs by regulating heavily saturated neighborhoods.

For Hosts:

- ➢ Price competitively in Manhattan
- ➢ Consider under-served neighborhoods for new listings

For Guests:

- ➢ Explore Brooklyn and Queens for better deals

For Policy Makers:

- ➢ Monitor neighborhoods with dense listings to manage impact on housing

IX. CONCLUSION

Overall, the EDA highlights the complex dynamics of short-term rentals in New York City. The data points to patterns that are useful for multiple stakeholders. Future analysis could incorporate seasonal trends and customer sentiment extracted from textual reviews for deeper insights. This analysis reveals spatial and pricing trends across NYC's Airbnb market. The data shows how location and room type impact price and availability. Further analysis could include time series, review sentiment, and host behavior patterns.