

Executive Summary – Airbnb NYC Data Analysis

This project explores the **Airbnb New York City dataset (2019)**, analyzing 48,895 listings across five boroughs. The analysis focused on data cleaning, handling missing values, outlier treatment, and visualization of key insights.

Key Insights

- **Data Preparation**
 - Missing values in reviews were imputed, and unrealistic price/outlier values were removed.
 - Extreme minimum nights (>30) were capped to ensure meaningful analysis.
- **Pricing Trends**
 - Most listings are priced under **\$200 per night**, with a heavy concentration below **\$500**.
 - A small fraction of luxury listings (>\$1000) skew the distribution.
- **Room Types**
 - **Entire home/apartment** dominates in Manhattan.
 - **Private rooms** are more common in Brooklyn and Queens, catering to budget travelers.
 - **Shared rooms** are the least frequent across all boroughs.
- **Geographic Insights**
 - **Manhattan** has the highest concentration of listings and highest average prices.
 - **Brooklyn** offers a strong balance between affordability and availability.
 - **Bronx and Staten Island** have limited listings and lower demand.
- **Reviews & Popularity**
 - Higher review counts are concentrated in affordable private rooms.

- Neighborhoods with vibrant tourism (e.g., Manhattan & Brooklyn) dominate review volumes.

Conclusion

The analysis highlights **Manhattan as the premium Airbnb market**, while **Brooklyn provides affordability with strong demand**. Outlier handling ensured robust insights, and the visualizations clearly mapped how price, location, and room type influence Airbnb listings across NYC.

This project demonstrates strong application of **data cleaning, exploratory analysis, and visualization skills** using Python, pandas, and matplotlib/seaborn.