



Data Analysis of AirBnb New York

By Aman Singh

Project Overview:

- This project involves analyzing a real-world Airbnb dataset using Python and popular data analysis libraries such as Pandas, Matplotlib, and Seaborn. The dataset includes various attributes like price, room type, location, number of reviews, availability, and more.
- The analysis process includes:
- Data cleaning and preprocessing (handling missing values and duplicates)
- Exploratory Data Analysis (EDA) using visualizations like histograms, scatter plots, bar plots, heatmaps, and pair plots
- Identifying trends in pricing, availability, and neighborhood-wise performance
- Correlation analysis between key numerical variables
- Geographical distribution of listings using longitude and latitude
- The insights generated help in identifying:
- The most profitable neighborhoods
- The impact of reviews on pricing
- Availability trends across different room types and regions
- Market competitiveness and customer preferences

Business Conclusions:

1. Pricing Strategy Insights

- Listings in certain neighborhoods (e.g., Manhattan) have significantly higher average prices, suggesting location is a key driver of value.
- Room type greatly impacts pricing, with entire homes/apartments commanding higher prices than shared or private rooms.

2. Customer Demand & Popularity

- Areas with higher number of reviews tend to have competitive pricing, indicating strong customer demand and host responsiveness.
- Listings with high availability throughout the year are more likely to generate consistent revenue.

3. Geographical Advantage

- Visualizing listings on a scatter plot based on latitude and longitude reveals clustering in high-demand areas.
- Hosts in central or tourist-heavy regions have better chances of booking and revenue generation.

4. Impact of Reviews

- A positive correlation exists between number of reviews and price in some regions, showing that social proof helps increase perceived value.

5. Optimization Opportunities

- Some neighborhoods have low pricing but high review counts and availability, indicating potential for price optimization.
- Hosts can optimize listings based on customer behavior trends revealed in the data.