



Airbnb Data Analysis Project Report

Introduction

This project aims to analyse Airbnb listings to uncover insights related to room types, pricing, host information, neighbourhood distribution, and guest reviews. Using data visualization and statistical analysis, we explore patterns that can assist hosts, guests, and stakeholders in making informed decisions.

Abstract

The dataset contains over **100,000** Airbnb listings, primarily in New York City. The project includes data cleaning, transformation, and visualization. **Python (Pandas, Matplotlib, Seaborn)** was used for data preprocessing and analysis, while **Power BI** was used to create an interactive dashboard. Key findings show that Manhattan and Brooklyn have the highest number of listings, and 'Entire home/apt' is the most common room type. The project also highlights pricing trends, availability, and review patterns.

Tools Used

- **Python:** For data preprocessing and analysis (libraries: Pandas, NumPy, Matplotlib, Seaborn)
 - **Power BI:** For dashboard creation and visual exploration
 - **Jupyter Notebook:** For coding and documenting Python scripts
 - **CSV File:** Dataset in .csv format
-

Steps Involved in Building the Project

1. **Data Loading**
Imported the dataset using Pandas from a .csv file containing 102,599 rows and 26 columns.
2. **Data Cleaning**
 - Removed null or missing values in key columns like host name and NAME.
 - Converted price and service fee to numeric values by stripping currency symbols.

- Dropped unnecessary columns such as license and house_rules.
- Removed duplicates to ensure accuracy.

3. Exploratory Data Analysis (EDA)

- Descriptive statistics were used to understand the distribution of prices, reviews, availability, etc.
- Identified data types and anomalies (e.g., extreme price values).

4. Visualizations in Python

- Histograms for price distribution
- Count plots for room types and neighbourhoods
- Box plots to compare price vs. room types
- Line plot showing reviews over time

5. Dashboard Creation in Power BI

- Developed a dashboard to visualize:
 - Room type distribution
 - Neighbourhood group listings
 - Price vs. room type
 - Geographic distribution of listings
 - Review trends and availability by region

Conclusion

The Airbnb data analysis provides crucial insights into listing dynamics, price trends, and customer preferences. Manhattan and Brooklyn dominate in listing volume. Most listings are Entire home/apartment types. The dashboard helps users visualize the data clearly, enabling strategic decision-making for both hosts and travellers. This project also demonstrates the power of combining Python and Power BI for a complete data analysis workflow.