Data Analysis Project: Airbnb Listings

Introduction

This document outlines the methodology used in the analysis of Airbnb listings data, focusing on Python for data processing and analysis, and Tableau for visualization.

Tools Used

- Python:
  - Libraries: NumPy, Pandas, Matplotlib, Seaborn
- Tableau

Data Cleaning

1. Null Values

Checked for null values in the dataset to ensure data integrity.

2. Column Selection

Removed inappropriate columns:

- Latitude and Longitude: Not relevant for the analysis.
- Last Review and Reviews Per Month: Dropped due to 25% null values and perceived less importance for the analysis.

Explanatory Data Analysis

1. Most Popular Neighbourhood Group

Identified the neighbourhood group with the highest listing frequency.

2. Top 10 Host Names

Listed the top 10 hosts based on the number of listings.

3. Host with Most Bookings in Manhattan

Identified the host with the highest number of bookings in the Manhattan neighbourhood.

4. Host with Highest Number of Listings

Determined the host with the largest number of listings.

5. Type of Rooms Preferred and Avg Price in Different Neighbourhood Groups

Explored the distribution of room types across different neighbourhood groups and calculated the average prices.

6. Most Popular and Least Popular Properties by Ratings

Analyzed the properties based on ratings to identify the most and least popular ones.

7. Top 10 Expensive and Least Expensive Properties

Identified the top 10 most expensive and least expensive properties based on pricing.

Visualization

Utilized Tableau for creating visualizations to supplement the analysis, providing a more intuitive understanding of the insights.

Conclusion

Summarized key findings and insights from the analysis.

Next Steps

Outlined potential next steps or areas for further exploration.