Logo

Description automatically generated

**Module 5: Final Project**

**By**

**Smit Pareshbhai Ranpariya & Vishvesh Gaurangbhai Vyas**

**Project: Airbnb**

**College of Professional Studies: Northeastern University**

**Prof.**

**Subject: ALY6110: Big Data and Data Management**

**Date: May 30, 2024**

**Project Proposal: Data-Driven Insights for Airbnb**

**Introduction**

This project aims to analyze Airbnb using a comprehensive dataset. By leveraging this data, we aim to uncover trends, patterns, and insights that can help hosts optimize their listings, improve guest experiences, and potentially increase revenue. The dataset contains detailed information on 48,895 listings, including attributes such as price, location, room type, number of reviews, availability, and more.

**Objectives**

1. Price Optimization: Identify factors that influence the pricing of Airbnb listings to help hosts set competitive and profitable rates.

2. Occupancy Insights: Analyze the availability data to understand booking trends and optimize occupancy rates.

3. Review Analysis: Examine reviews and ratings to identify key drivers of guest satisfaction.

4. Geographical Trends: Explore geographical patterns to determine the most popular neighborhoods and their characteristics.

5. Host Performance: Evaluate the performance of hosts based on their listing counts and review scores.

**Data Overview**

The dataset includes the following columns:

- id: Unique identifier for each listing

- name: Name of the listing

- host\_id: Unique identifier for each host

- host\_name: Name of the host

- neighbourhood\_group: Borough of the listing

- neighbourhood: Specific neighborhood of the listing

- latitude and longitude: Geographical coordinates

- room\_type: Type of room offered

- price: Price per night in USD

- minimum\_nights: Minimum number of nights required for booking

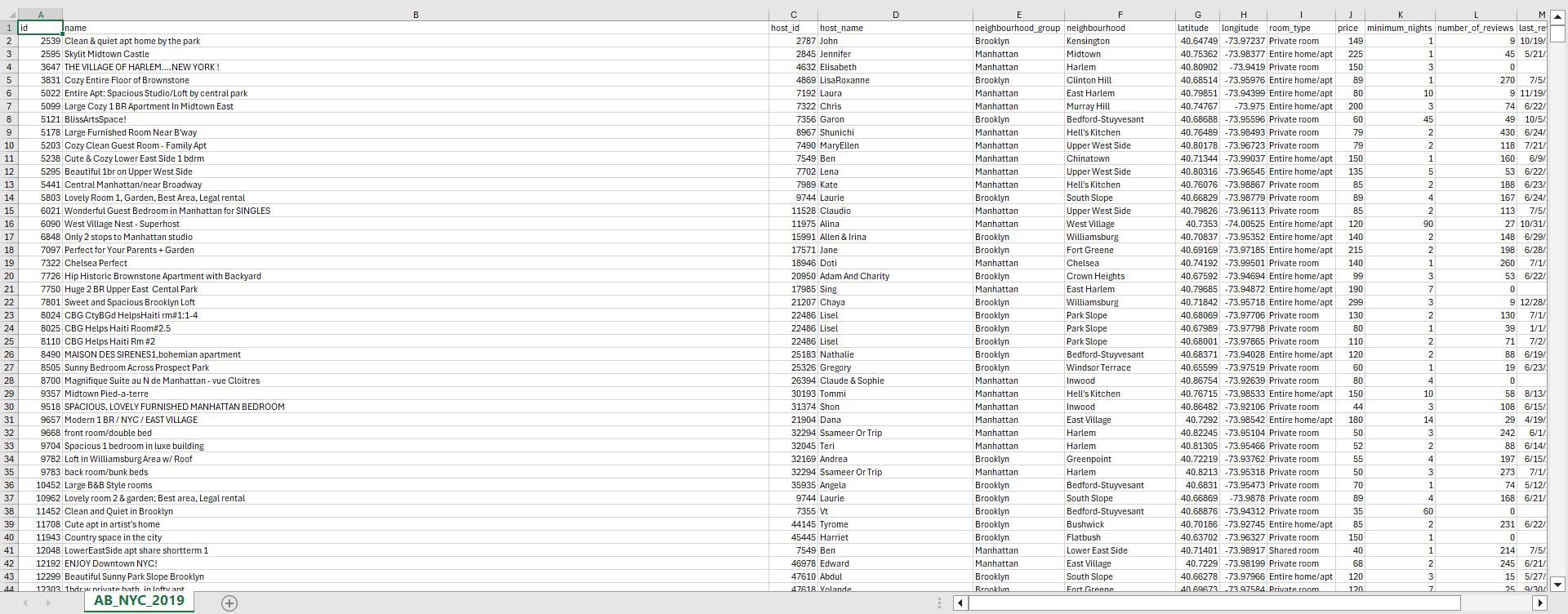
- number\_of\_reviews: Total number of reviews received

- last\_review: Date of the last review

- reviews\_per\_month: Average number of reviews per month

- calculated\_host\_listings\_count: Total listings managed by the host

- availability\_365: Number of available days in a year



**Methodology**

1. Data Cleaning: Handle missing values, correct data types, and ensure data consistency.

2. Exploratory Data Analysis (EDA): Perform descriptive statistic and visualization to understand the distribution and relationship between variable.

3. Price Analysis: Use regression models to identify factors influencing price variations across different neighborhoods and room types.

4. Occupancy Analysis: Analyze availability data to identify high and low demand periods.

5. Sentiment Analysis: Conduct text analysis on review comments to extract sentiments and key themes.

6. Geospatial Analysis: Map listings and analyze spatial trends using GIS tools.

7. Host Analysis: Compare performance metrics across different hosts and identify characteristics of successful hosts.

**Expected Outcomes**

1. Pricing Strategy: A dynamic pricing model that considers various factors such as location, room type, and seasonality.

2. Occupancy Optimization: Insights into booking patterns that help hosts maximize occupancy rates.

3. Enhanced Guest Experience: Key drivers of positive reviews to guide hosts in improving their service.

4. Neighborhood Insights: Identification of high-demand neighborhoods and their unique features.

5. Host Best Practices: Best practices for hosts to increase their ratings and manage multiple listings effectively.

**Resources Required**

- Software: Python or R, Jupyter Notebook, GIS tools, Data visualization and dashbord (Tableau or Power BI)

- Hardware: Standard computational resources for data analysis

- Data: Provided dataset, additional data from public sources if necessary

**Conclusion**

By conducting a thorough analysis of Airbnb , this project will provide valuable insights and actionable recommendations for hosts. These insights will help in enhancing guest satisfaction, optimizing pricing, and improving overall performance, thereby contributing to the success of Airbnb hosts in a competitive market.

**Reference:**

[**https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data**](https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data)