2247101 Chandana Pacha

**Project Proposal**

**“New York Airbnb Data Analysis”**

**Introduction:**

The sharing economy has revolutionized the way people travel, with Airbnb being a forefront platform offering short-term lodging in private homes. The pricing of listings on Airbnb can be affected by various factors including location, amenities, and availability. This project aims to utilize linear regression analysis to understand the influences on Airbnb pricing in New York City and potentially aid hosts in optimizing their rental strategy while providing insights for travelers.

**Objectives:**

1. Determine the impact of neighborhood group and type of room on Airbnb pricing.
2. Explore the relationship between the number of reviews, review frequency, and price, assuming that customer feedback may reflect on pricing dynamics.
3. Analyze how minimum nights, availability, and host listing count influence pricing.
4. Develop a predictive model for Airbnb listing prices based on relevant factors.
5. Examine the interaction between room type and other variables to understand complex dynamics in pricing.

**Methodology:**

**Data Collection:**

Data is collected from Kaggle datasets.

<https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data>

**Data Preprocessing:**

* Inspect the dataset for completeness and cleanliness.
* Handle missing values through imputation or removal.
* Identify and treat outliers that could skew analysis.
* Convert categorical variables such as “neighbourhood\_group” and “room\_type” into dummy variables.

**Exploratory Data Analysis (EDA):** Conduct a thorough EDA to uncover patterns and relationships within the data using visualization and correlation analysis.

**Linear Regression Analysis:**

1. **Simple Linear Regression Analysis:** Execute simple linear regression to assess the direct impact of individual factors on the pricing.
2. **Multiple Linear Regression Analysis:** Implement multiple linear regression to evaluate the combined effect of several predictors on the price.

**Statistical Validation:** Validate the models' assumptions through residual diagnostics and identify potential model improvements.

**Model Selection and Evaluation:** Employ model selection techniques like stepwise regression to refine the models and use metrics such as R-squared and RMSE for evaluation.

**Expected Outcomes:**

* Insights into the significance of various factors on Airbnb pricing in NYC.
* A comprehensive model for predicting Airbnb prices, helping hosts in pricing decisions.
* Recommendations for travelers seeking value-for-money listings based on model findings.