**Project- Part A: Airbnb Price Prediction and Insights**

(Video link:- https://drive.google.com/file/d/1A2QOWxQScFxAk3DRQjqA0v6uePF51EWm/view?usp=sharing)

**1.Data Exploration and Preprocessing:**

**Analyze the dataset for trends, missing values, and outliers:**

In this task to find trend by plotting a boxplot to analyse the trend for price by city and I find out that the city SF has the highest Average log price and to find outliers by plotting boxplot for column like ‘log\_price', 'bathrooms', 'latitude', 'longitude', 'number\_of\_reviews', 'review\_scores\_rating', 'bedrooms', 'beds'. Then I created a missing\_values(df) function to print the missing values.

**Perform data cleaning, feature engineering, and transformations:**

In this first I handle the missing values by putting meadian values in numeric columns and put the ‘Unknown’ in non numeric columnand also dropped unnecessary columns like 'id', 'description', and 'amenities' etc which are not going to use further. Then I use LabelEncoder to convert the non-numeric data into numerical lables.

1. **Model Development:**

**Build a regression model to predict listing prices:**

In this first the target variable X will contain all columns except ‘log\_price’ by using drop() function and then target variable y will contain log\_price. After I split the data into training and testing phase and then I created a Linear Regression model by using LinearRegression() function using the training data (X\_train, y\_train).

1. **Model Evaluation:**

**Evaluate the model’s performance using appropriate metrics like RMSE, MAE, and R²:**

In this I created a variable y\_pred that will be the predicted values for log\_price based on the testing set X\_test after that I calculated the RMSE, MAE, and R² and in the result the value of MAE is 0.3686 which indicate that the difference between the actual and predicted value is around 0.3686.

RMSE is 0.2394 which Indicate that average magnitude of error is about 0.2394 unit.

R^2 is 0.5339 which Indicate that 53.39% variance in log\_price which shows that the model is moderately effective.

Explanation Video Link:

https://drive.google.com/file/d/1A2QOWxQScFxAk3DRQjqA0v6uePF51EWm/view?usp=sharing