Home Assignment <4>: Predicting House Prices using Regression

Learning Objective:

The objective of this assignment is to understand how to build and evaluate a **Simple Linear Regression** model to predict **house prices** based on **square footage**.
Students will perform data analysis, preprocessing, model training, and performance evaluation using statistical and visualization techniques.

Dataset :

You are provided with a dataset containing details of residential properties.
Focus on the following two numerical features for your analysis:

- **Square Footage (X)**
- **Price (Y)**

Expected Completion Time:

Best Case: 30 minutes
Average Case: 45 minutes

Assignment Details:

1. **Load the Dataset**
   - Load the provided dataset using **Pandas**.
   - Retain only the columns **Square Footage** and **Price** for model building.
2. **Exploratory Data Analysis (EDA)**
   - Display the first few rows of the dataset.
   - Check for missing or null values and handle them appropriately.
   - Visualize the relationship between **Square Footage** and **Price** using a **scatter plot**.
3. **Feature and Target Selection**
   - Assign **Square Footage** as the **independent variable (X)**.
   - Assign **Price** as the **dependent variable (Y)**.
4. **Train-Test Split**
   - Split the dataset into **training** and **testing** sets using an 80-20 ratio.
5. **Model Building**
   - Create a **Linear Regression model** using `LinearRegression` from `sklearn.linear_model`.
   - Fit the model on the training data.
   - Display the **intercept (b_0)** and **coefficient (b_1)** of the regression line.
6. **Prediction and Evaluation**
   - Predict the house prices for the test set.
   - Calculate and print the following evaluation metrics:
     - Mean Squared Error (MSE)
     - Root Mean Squared Error (RMSE)

- R^2 Score (Coefficient of Determination)

7. **Visualization**
    - Plot the **regression line** along with the actual data points.
    - Visualize **actual vs predicted prices** to assess model performance.

Expected Outcome:

Upon completion of this assignment, you should be able to:

- Understand the concept of **Linear Regression** and its mathematical formulation.
- Apply **data preprocessing and EDA** on real-world datasets.
- Build and interpret a **Simple Linear Regression model**.
- Evaluate model performance using key statistical metrics.
- Visualize regression results effectively.
- Develop insights on how **square footage affects housing prices**.