Machine Learning Assignment – Multiple Linear Regression

**Assignment: Predicting House Prices using Multiple Linear Regression**

**Objective**: Implement a Multiple Linear Regression model to predict house prices based on square footage, number of bedrooms, and age of the house.

**Dataset**: A dataset with features such as square footage, number of bedrooms, house age, and the target variable being house prices.

**Steps:**

1. **Load the Data**:
   * Load the data into a pandas DataFrame.
2. **Explore the Data**:
   * Display the first few rows of the dataset.
   * Check for any missing values.
   * Understand the basic statistics of the features and the target variable.
3. **Visualize the data through Charts**
   1. **Pairplot**:
   * Show the pairwise relationships between all variables. It helps in understanding the distribution and interactions between features and the target variable.
   1. **Distribution Plot**:

* Display the distribution of house prices.
  1. **Scatter Plot (Square Footage vs House Price)**:
* Show the relationship between square footage and house price.
  1. **Box Plot (Number of Bedrooms vs House Price)**:
* Show the distribution of house prices for different numbers of bedrooms.
  1. **Scatter Plot (House Age vs House Price)**:
* Show the relationship between house age and house price.
  1. **Correlation Heatmap**:
* Display the correlation coefficients between features.

1. **Split the Data**:
   * Split the data into training and testing sets (80% training, 20% testing).
2. **Train the Model**:
   * Implement a Multiple Linear Regression model using scikit-learn.
   * Train the model on the training data.
3. **Evaluate the Model**:
   * Predict the target variable for the test data.
   * Calculate and display the Mean Squared Error (MSE) of the predictions and r2 score of the model.
4. **Visualize the Results**:
   * Create a scatter plot comparing the actual and predicted values of the target variable for the test set.
5. **Create a Prediction Function**:
   * Implement a function that takes input features and returns the predicted house price.

Hint:

# Function to predict house prices

def predict\_house\_price(square\_footage, num\_bedrooms, house\_age):

input\_features = pd.DataFrame({

'Square\_Footage': [square\_footage],

'Num\_Bedrooms': [num\_bedrooms],

'House\_Age': [house\_age]

})

input\_features\_scaled = scaler.transform(input\_features)

predicted\_price = model.predict(input\_features\_scaled)

return predicted\_price[0]

1. **Determine feature importance and visualize it:**

* This will determine which features contribute the most to the model's predictions. In the context of linear regression, the importance of a feature can be gauged by looking at the absolute value of its coefficients. Higher absolute values indicate greater importance.

1. **Normalize the data using StandardScaler from scikit-learn and check its impact on the model performance:**

* This can be particularly useful for models like linear regression where different scales of features can affect the coefficients and model performance.