**5. Implementation of Multiple Linear Regression for House Price Prediction using sklearn**

**Aim:** Write a program to implement Multiple Linear Regression.

**Description:**

**Multiple Linear Regression:** Multiple linear regression is a regression model that estimates the relationship between a quantitative dependent variable and two or more independent variables using a straight line.

**Mean Squared Error**

Mean squared error (MSE) is a metric used to measure the average squared difference between the predicted values and the actual values in the dataset. It is calculated by taking the average of the squared residuals, where the residual is the difference between predicted value and the actual value for each data point. The MSE value provides a way to analyse the accuracy of the model.

**Mean Absolute Error**

Mean Absolute Error calculates the average difference between the calculated values and actual values.

**R2 Score**

R-Squared (R² or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable.

**Program:**

import pandas as pd

import numpy as np

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error,mean\_absolute\_error, r2\_score

import matplotlib.pyplot as plt

df=pd.read\_csv(r"C:\Users\Tabeen Fatima\Downloads\Real-estate.csv")

print(df)

df=df.drop(columns=['No'])

print(df)

x=df.drop(columns=['Y house price of unit area'])

y=df['Y house price of unit area']

x\_train, x\_test, y\_train, y\_test= train\_test\_split(x,y,test\_size= 0.3, random\_state=42)

model=LinearRegression()

model.fit(x\_train,y\_train)

y\_predict=model.predict(x\_test)

print(y\_predict)

print('mean\_squared\_error : ', mean\_squared\_error(y\_test, y\_predict))

print('mean\_absolute\_error : ', mean\_absolute\_error(y\_test, y\_predict))

print('R-Square :',r2\_score(y\_test, y\_predict))

m, b = np.polyfit(y\_test, y\_predict, 1) # Fit a line (y = mx + b)

plt.plot(y\_test, m\*y\_test + b, color='red', label="Best Fit Line")

plt.scatter(y\_test,y\_predict, color='blue', label="Predictions")

plt.xlabel("Actual House Prices")

plt.ylabel("Predicted House Prices")

plt.title("Actual vs Predicted House Prices with Best Fit Line")

plt.legend()

plt.show()

Output:

| **No** | **X1 transaction date** | **X2 house age** | **X3 distance to the nearest MRT station** | **X4 number of convenience stores** | **X5 latitude** | **X6 longitude** | **Y house price of unit area** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1 | 2012.917 | 32.0 | 84.87882 | 10 | 24.98298 | 121.54024 | 37.9 |
| **1** | 2 | 2012.917 | 19.5 | 306.59470 | 9 | 24.98034 | 121.53951 | 42.2 |
| **2** | 3 | 2013.583 | 13.3 | 561.98450 | 5 | 24.98746 | 121.54391 | 47.3 |
| **3** | 4 | 2013.500 | 13.3 | 561.98450 | 5 | 24.98746 | 121.54391 | 54.8 |
| **4** | 5 | 2012.833 | 5.0 | 390.56840 | 5 | 24.97937 | 121.54245 | 43.1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |
| **409** | 410 | 2013.000 | 13.7 | 4082.01500 | 0 | 24.94155 | 121.50381 | 15.4 |
| **410** | 411 | 2012.667 | 5.6 | 90.45606 | 9 | 24.97433 | 121.54310 | 50.0 |
| **411** | 412 | 2013.250 | 18.8 | 390.96960 | 7 | 24.97923 | 121.53986 | 40.6 |
| **412** | 413 | 2013.000 | 8.1 | 104.81010 | 5 | 24.96674 | 121.54067 | 52.5 |
| **413** | 414 | 2013.500 | 6.5 | 90.45606 | 9 | 24.97433 | 121.54310 | 63.9 |

| **X1 transaction date** | **X2 house age** | **X3 distance to the nearest MRT station** | **X4 number of convenience stores** | **X5 latitude** | **X6 longitude** | **Y house price of unit area** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 2012.917 | 32.0 | 84.87882 | 10 | 24.98298 | 121.54024 | 37.9 |
| **1** | 2012.917 | 19.5 | 306.59470 | 9 | 24.98034 | 121.53951 | 42.2 |
| **2** | 2013.583 | 13.3 | 561.98450 | 5 | 24.98746 | 121.54391 | 47.3 |
| **3** | 2013.500 | 13.3 | 561.98450 | 5 | 24.98746 | 121.54391 | 54.8 |
| **4** | 2012.833 | 5.0 | 390.56840 | 5 | 24.97937 | 121.54245 | 43.1 |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **409** | 2013.000 | 13.7 | 4082.01500 | 0 | 24.94155 | 121.50381 | 15.4 |
| **410** | 2012.667 | 5.6 | 90.45606 | 9 | 24.97433 | 121.54310 | 50.0 |
| **411** | 2013.250 | 18.8 | 390.96960 | 7 | 24.97923 | 121.53986 | 40.6 |
| **412** | 2013.000 | 8.1 | 104.81010 | 5 | 24.96674 | 121.54067 | 52.5 |
| **413** | 2013.500 | 6.5 | 90.45606 | 9 | 24.97433 | 121.54310 | 63.9 |

array([47.55430212, 41.08372744, 44.25551663, 40.51685112, 27.43467608,

45.85435624, 44.70296876, 46.57752144, 23.86072119, 54.35199522,

34.63011363, 36.34958785, 37.34518438, 23.49428018, 36.02735916,

34.56828629, 44.01823263, 45.04246804, 29.9554971 , 43.96338099,

0.09450675, 35.291724 , 48.09455663, 45.54503941, 10.6006418 ,

39.9786752 , 11.9116313 , 42.79356737, 35.67204264, 38.09573519,

12.98819108, 40.13164278, 35.44632378, 27.09263458, 46.57230583,

32.08134705, 50.62714445, 13.2533591 , 49.8007962 , 40.83574406,

39.02898306, 40.92260381, 47.61743515, 37.68719274, 39.67028734,

46.64468542, 43.57375428, 19.40851794, 46.95629589, 45.93056809,

48.03966928, 48.79469044, 43.06419598, 43.00021216, 36.56049741,

14.20568662, 37.31495695, 34.30916207, 28.33264665, 46.09215428,

32.92580137, 32.83028385, 14.20568662, 8.3906898 , 9.79703308,

34.33939647, 27.76485421, 46.75860125, 33.77722462, 30.85241149,

45.70501009, 39.07276122, 45.50942841, 45.00472078, 35.59161518,

37.74197487, 38.450681 , 37.63852526, 40.76300959, 33.88758966,

42.52501106, 43.31378324, 27.56441871, 39.40936415, 33.8977811 ,

30.61439071, 39.93460642, 45.36129734, 34.86088945, 39.96760778,

12.55982 , 41.4248276 , 41.87470921, 48.61756325, 46.67493517,

40.34138938, 34.2628639 , 45.3400411 , 32.29706728, 47.46416107,

29.69148895, 34.11146083, 43.36677103, 30.63583283, 48.77316661,

49.23795893, 30.34082584, 26.11105207, 40.9879696 , 38.89338679,

37.24238803, 40.02563617, 48.48697441, 45.26509346, 18.10869352,

41.23164604, 45.60093933, 43.63953623, 31.35568965, 31.75622985,

46.20909815, 14.58093806, 50.40747066, 46.60197459, 37.99631926])

mean\_squared\_error : 73.56837932850034

mean\_absolute\_error : 6.184836340097152

R-Square : 0.5600638168604377

