Name: Shawon Barman
Reg. ID: 18201043

# UNIVERSITY OF ASIA PACIFIC

Department of Computer Science and Engineering

**Course Title :**

**Artificial Intelligence and Expert Systems Lab**

**Course Code : CSE 404**

**Project : 02**

**Submitted By:**
**Shawon Barman**
Section: A
ID: 18201043

**Submitted to:**
**Dr. Nasima Begum**
Assistant Professor
Department of CSE, University Of Asia Pacific

Name: Shawon Barman
Reg. ID: 18201043

# Project Name: Implementation of Multivariable Linear Regression

## Introduction:

Multiple linear regression is simply the extension of simple linear regression, that predicts the value of a dependent variable (sometimes it is called as the outcome or target variable) on the basis of two or more independent variables (or sometimes, the predictor or explanatory variables). It is a method of statistical analysis that provides the statistical significance to explanatory variables, or which potential explanatory variables are crucial predictors for a given response (target) variable.

## Problem statement:

Implement the Multivariable Linear Regression Using Open-Source Dataset with and without SK-Learn.

## Dataset:

Dataset name : USA Housing

Dataset URL : https://github.com/huzaifsayed/Linear-Regression-Model-for-House-Price-Prediction/blob/master/USA_Housing.csv

The dataset has 7 columns:

- Avg. Area Income,
- Avg. Area House Age,
- Avg. Area Number of Rooms,
- Avg. Area Number of Bedrooms,
- Area Population,
- Price,
- Address

The variables are as follows Response:

- Variable (x1) : Avg. Area Income,
- Variable (x2) : Avg. Area House Age,

Name: Shawon Barman
Reg. ID: 18201043

- Variable (x3) : Avg. Area Number of Rooms,
- Variable (x4) : Avg. Area Number of Bedrooms,
- Variable (x5) : Area Population,
- Variable (y) : Price

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Avg. Area Income | Avg. Area House Age | Avg. Area Number of Rooms | Avg. Area Number of Bedrooms | Area Population | Price | Address |
| 2 | 79545.45857 | 5.682861322 | 7.009188143 | 4.09 | 23086.8005 | 1059033.558 | 208 Michael Ferry Apt. 674Laurabu |
| 3 | 79248.64245 | 6.002899808 | 6.730821019 | 3.09 | 40173.07217 | 1505890.915 | 188 Johnson Views Suite 079Lake K |
| 4 | 61287.06718 | 5.86588984 | 8.51272743 | 5.13 | 36882.1594 | 1058987.988 | 9127 Elizabeth StravenueDanielto |
| 5 | 63345.24005 | 7.188236095 | 5.586728665 | 3.26 | 34310.24283 | 1260616.807 | USS BarnettFPO AP 4 |
| 6 | 59982.19723 | 5.040554523 | 7.839387785 | 4.23 | 26354.10947 | 630943.4893 | USNS RaymondFPO AE |
| 7 | 80175.75416 | 4.988407758 | 6.104512439 | 4.04 | 26748.42842 | 1068138.074 | 06039 Jennifer Islands Apt. 443Tr |
| 8 | 64698.46343 | 6.025335907 | 8.147759585 | 3.41 | 60828.24909 | 1502055.817 | 4759 Daniel Shoals Suite 442Nguye |
| 9 | 78394.33928 | 6.989779748 | 6.620477995 | 2.42 | 36516.35897 | 1573936.564 | 972 Joyce ViaductLake William, |
| 10 | 59927.66081 | 5.36212557 | 6.393120981 | 2.3 | 29387.396 | 798869.5328 | USS GilbertFPO AA 2 |
| 11 | 81885.92718 | 4.42367179 | 8.167688003 | 6.1 | 40149.96575 | 1545154.813 | Unit 9446 Box 0958DPO A |
| 12 | 80527.47208 | 8.093512681 | 5.0427468 | 4.1 | 47224.35984 | 1707045.722 | 6368 John Motorway Suite 700Jan |
| 13 | 50593.6955 | 4.496512793 | 7.467627404 | 4.49 | 34343.99189 | 663732.3969 | 911 Castillo Park Apt. 717Davisbo |
| 14 | 39033.80924 | 7.671755373 | 7.250029317 | 3.1 | 39220.36147 | 1042814.098 | 209 Natasha Stream Suite 961Huffr |
| 15 | 73163.66344 | 6.919534825 | 5.993187901 | 2.27 | 32326.12314 | 1291331.518 | 829 Welch Track Apt. 992North Joh |
| 16 | 69391.38018 | 5.344776177 | 8.406417715 | 4.37 | 35521.29403 | 1402818.21 | PSC 5330, Box 4420APO A |
| 17 | 73091.86675 | 5.443156467 | 8.517512711 | 4.01 | 23929.52405 | 1306674.66 | 2278 Shannon ViewNorth Carrier |
| 18 | 79706.96306 | 5.067889591 | 8.219771123 | 3.12 | 39717.81358 | 1556786.6 | 064 Hayley UnionsNicholsboroug |

Here in the dataset the Price column is a dependent feature and the rest of the columns are independent features plotting the independent variable vs department variable to see how independent variables affect the independent variable.

## Tools And Languages:

- **Programing Language:** Python
- **IDE**: Jupyter Notebook

## Parameter Initialization:

For the weights of the features and bias we use 1,1.5, 3, 8, and 9 values.

## Hypothesis Function:

For Multiple Linear Regression the hypothesis function will be:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

Name: Shawon Barman
Reg. ID: 18201043

## Cost Function:

For cost function, we are using Means Squared Error. The formula for the MSE

$$Cost = \frac{1}{m}\sum_{i=0}^{m}(y_{predict}(i) - y_{true}(i))^2$$

## Gradient Function:

Updating the weights and bias value we are using Gradient Descent(GD)

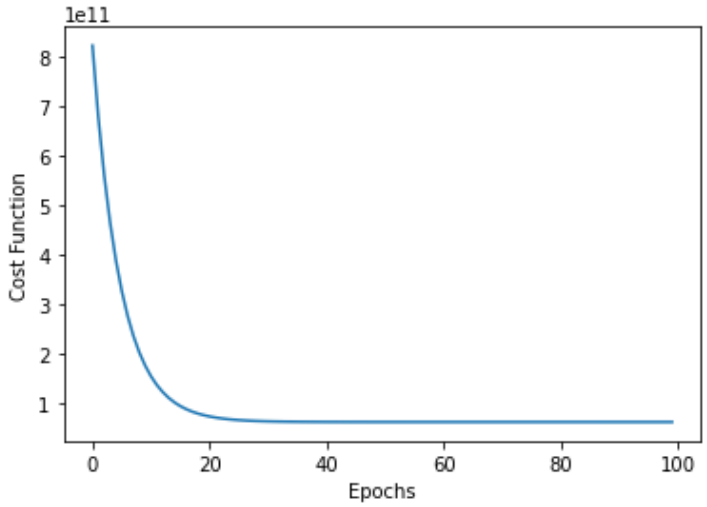
$$weights_i = weights_i - \frac{\alpha}{m} * \sum_{i=0}^{m}[(y_{predict}(i) - y_{true}(i)) * X(i)]$$

## Linear Regression Without SK-Learn: Output

### *Cost Function & Gradient Decent:*

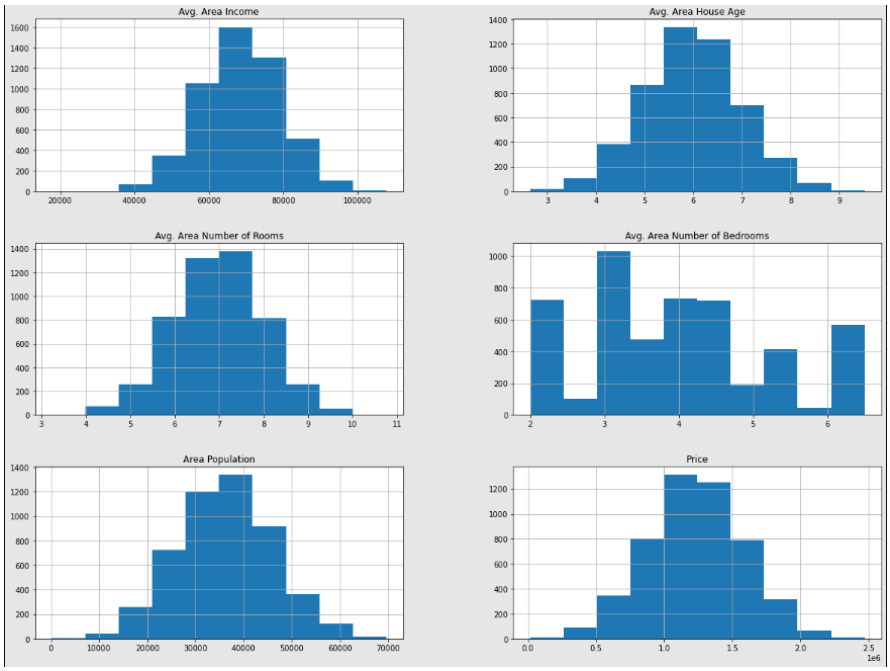
```
Iteration number:  100
Hypothesis function value is: h0(x)=theta_0+theta_1 * x
Cost function is: j(theta)=1/(2*m) * i=1_samtionSign_m (h_theta_(x)-y)**2

Cost function is: 62300201242.352455

gradient decent:
New parameter value is:  [1230689.3193730055, 22103.63785409576, 26572.07523898754, 25854.485030193617, 26240.419866977296]
result coefficient is [1230689.3193730055, 22103.63785409576, 26572.07523898754, 25854.485030193617, 26240.419866977296]
```

### *Graph Plotting:*

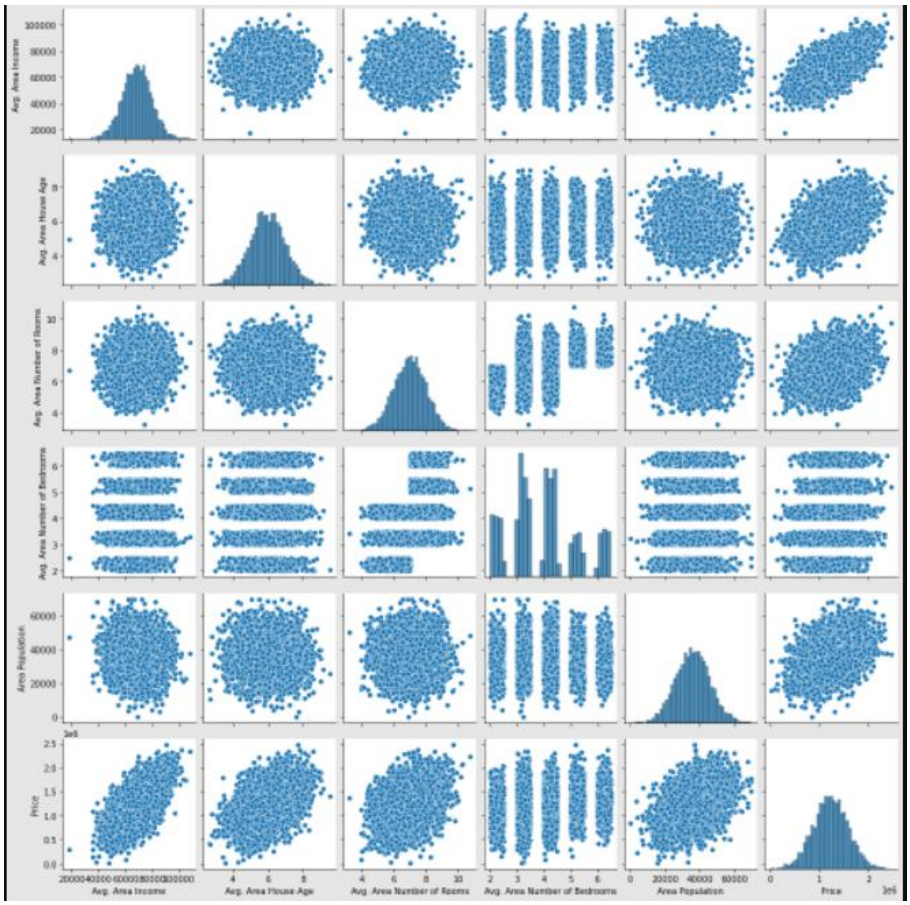


Name: Shawon Barman
Reg. ID: 18201043

## *Figure Plotting:*

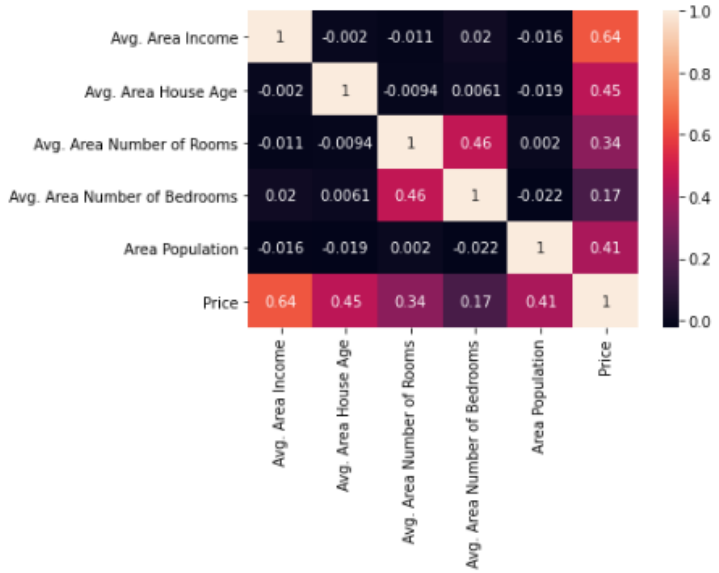



## *Pairwise Figure:*



Name: Shawon Barman
Reg. ID: 18201043

## *HeatMap:*



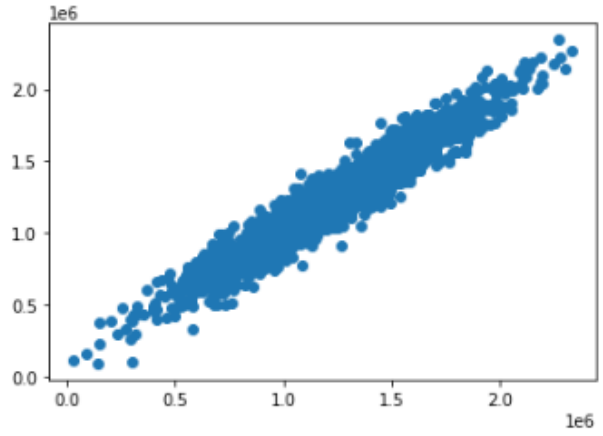

# Linear Regression With SK-Learn: Output:

## *Coefficients , Mean Error & Variance Score:*

```
In [23]: print("Coefficients: \n", lm.coef_)
         # Mean Squared Error
         print("Mean Squared Error: %.2f" %np.mean((lm.predict(X_test)-y_test)**2))
         # Explained Variance Score : 1 is perfect prediction
         print("Variance score: %.2f" %lm.score(X_test, y_test))
```

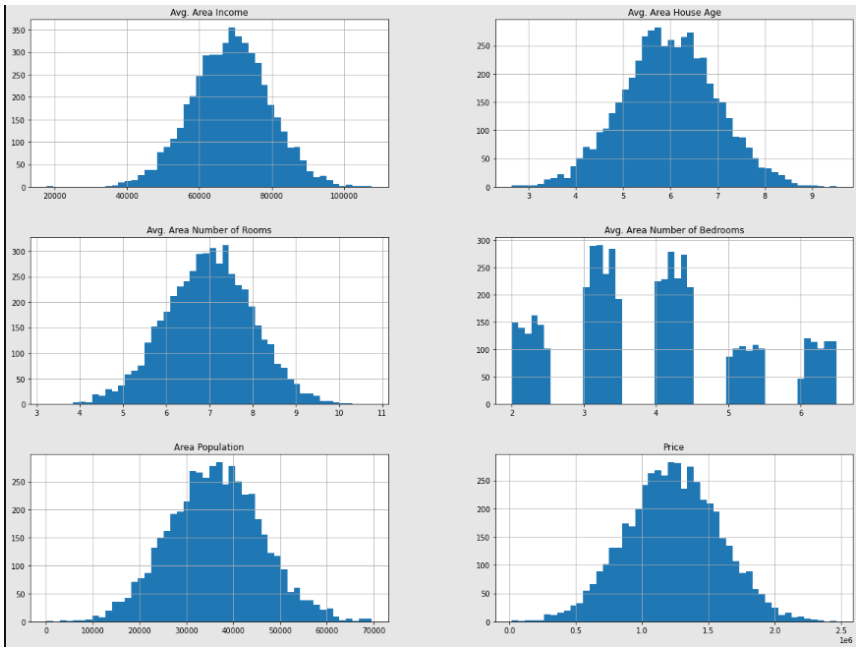
```
Coefficients:
 [2.15282755e+01 1.64883282e+05 1.22368678e+05 2.23380186e+03
 1.51504200e+01]
Mean Squared Error: 10460958907.21
Variance score: 0.92
```

## *Graph Plotting:*

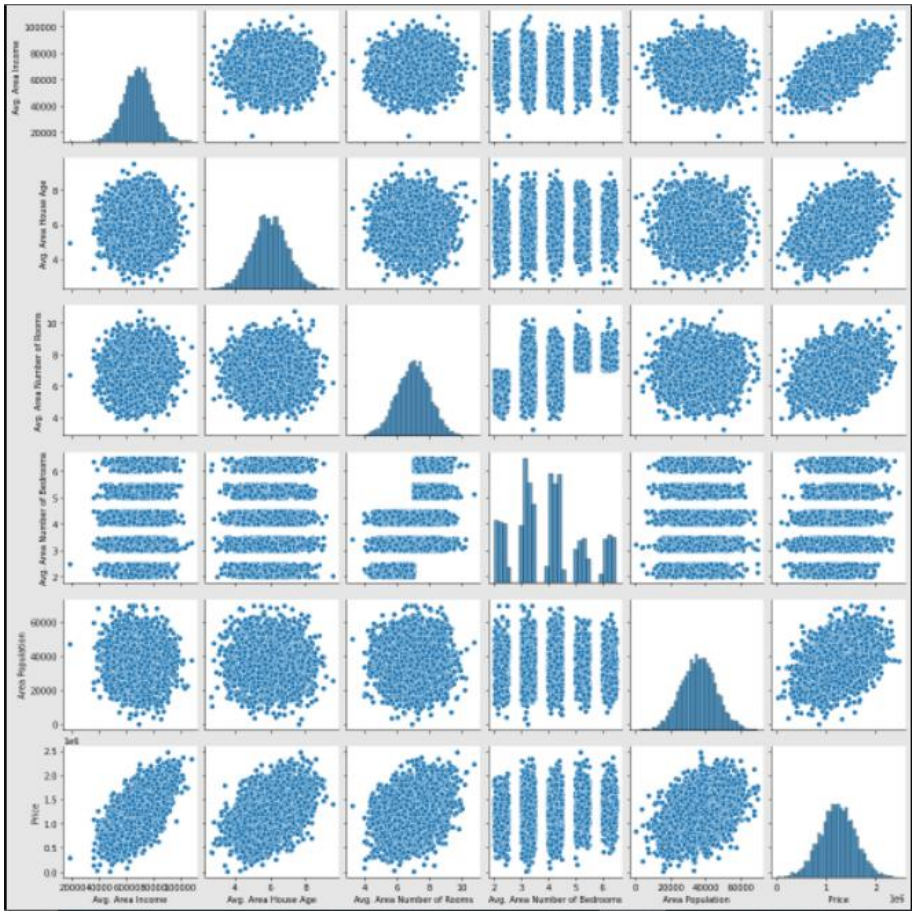


Name: Shawon Barman
Reg. ID: 18201043

## *Figure Plotting:*

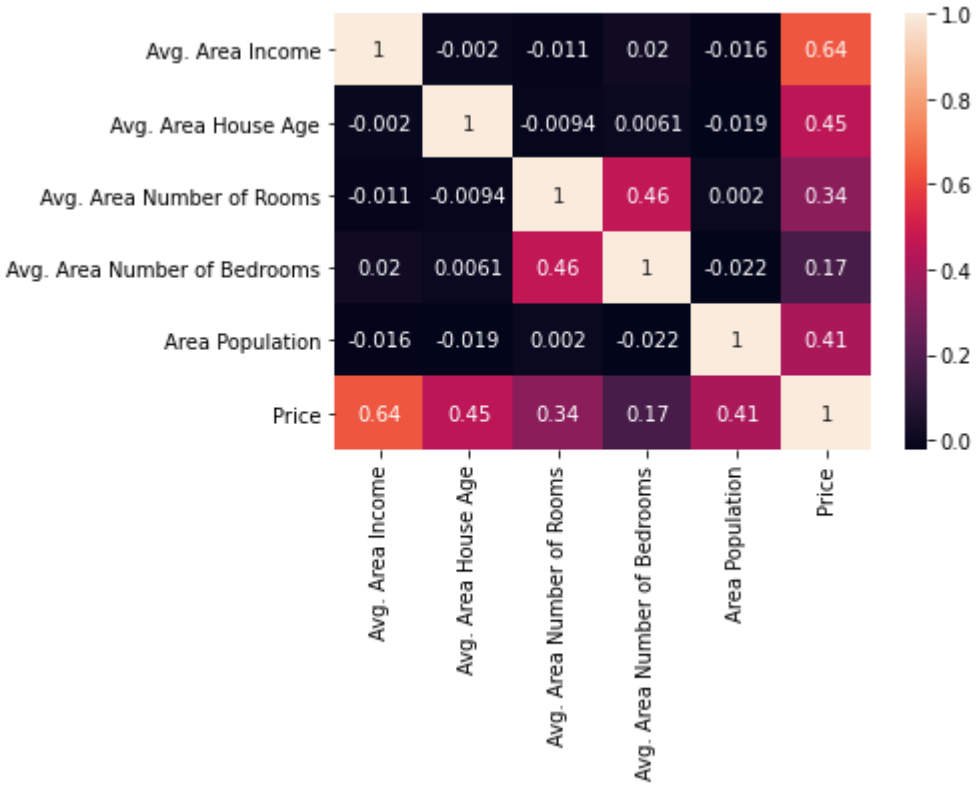



## *Pairwise Figure:*



Name: Shawon Barman
Reg. ID: 18201043

## ***HeatMap:***




## Conclusion:

Multiple regression is an extension of simple linear regression. In this project, after successful implementation, I've a brief knowledge about linear regression. Multiple linear regression models are useful in helping an enterprise to consider the impact of multiple independent predictors and variables on a dependent variable and can be beneficial for forecasting and predicting results. So, the outcome knowledge of this project will help me a lot to real-life problem predicting and solving on various field.