# **Github Link:** [**https://github.com/SrikanthMajhi/Assignment---4**](https://github.com/SrikanthMajhi/Assignment---4)

# **Question – 1**

1. Apply Linear Regression to the provided dataset using underlying steps.

a. Import the given “Salary\_Data.csv”

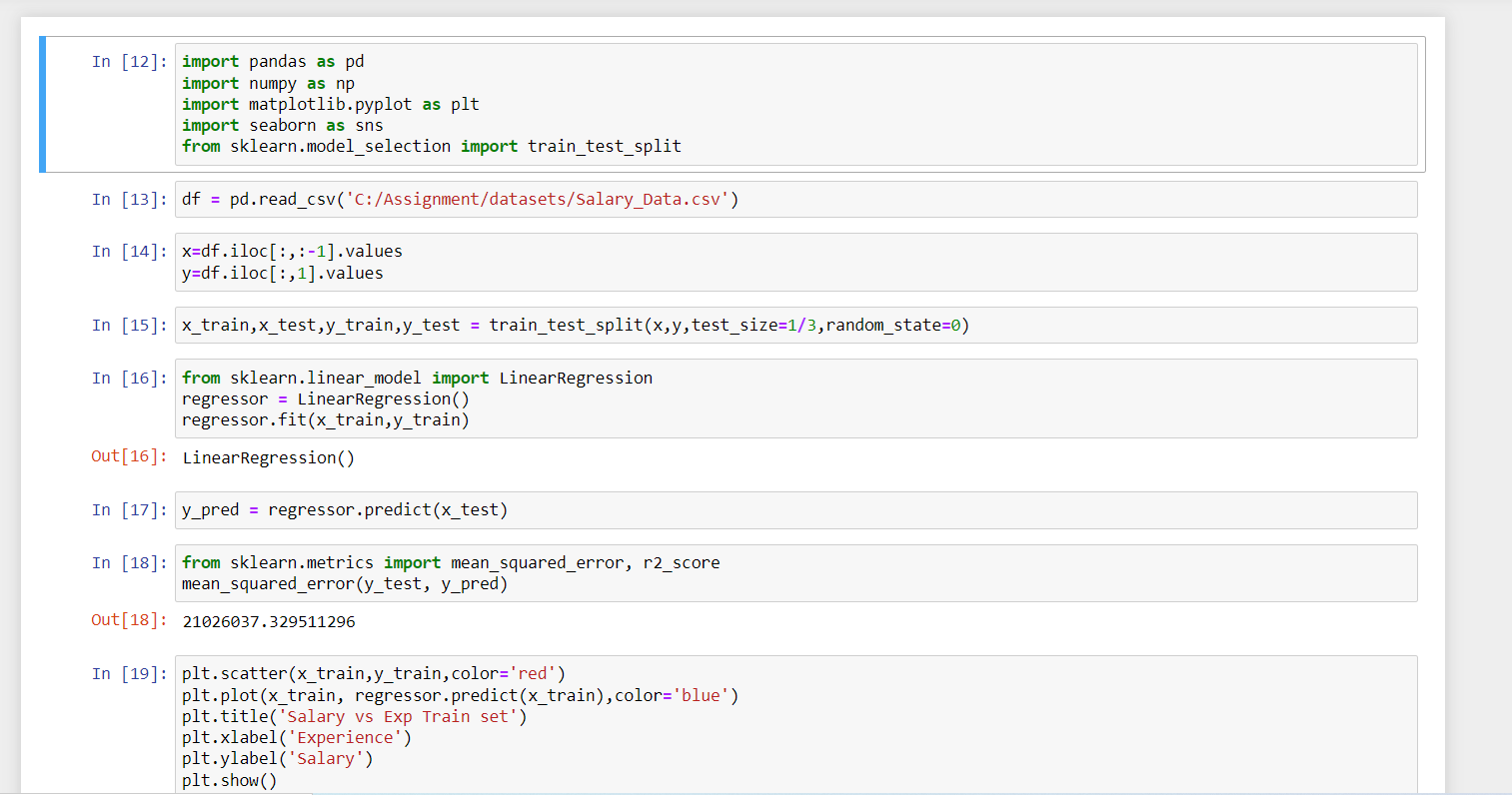
b. Split the data in train\_test partitions, such that 1/3 of the data is reserved as test subset.

c. Train and predict the model.

d. Calculate the mean\_squared error

e. Visualize both train and test data using scatter plot.

**Source Code:**



Chart, scatter chart

Description automatically generated

Chart

Description automatically generated

# **Question – 2 &3**

2. Apply K means clustering in the dataset provided:

• Remove any null values by the mean.

• Use the elbow method to find a good number of clusters with the K-Means algorithm

• Calculate the silhouette score for the above clustering.

**Source Code:**

Table

Description automatically generated with low confidence

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated

**NO of Clusters = 2**

Graphical user interface

Description automatically generated with medium confidence

Graphical user interface, text, application

Description automatically generated

Graphical user interface

Description automatically generated

**No of Clusters = 3**

**Yes, the Silhouette score is improved by 1. The Reason is K-Means requires Feature scaling to give the accurate and good clusters.**