**NAME:** Priyanka Nimmagadda

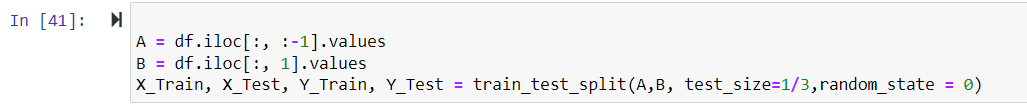
**StudentId:** 700734089

**GitHub Link: https://github.com/pnimmagadda0404/ML\_Assignment4**

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

a. Import the given “Salary\_Data.csv”  
Graphical user interface, text, application

Description automatically generated

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.   
Graphical user interface, text, application, email

Description automatically generated

d. Calculate the mean\_squared error  
Graphical user interface, text, application, Word

Description automatically generated with medium confidence

e. Visualize both train and test data using scatter plot.  
Chart, scatter chart

Description automatically generated  
Chart, scatter chart

Description automatically generated

2. Apply K means clustering in the dataset provided:

• Remove any null values by the mean.  
Table

Description automatically generated

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

Description automatically generated with medium confidence

• Calculate the silhouette score for the above clustering  
Graphical user interface, text, application, email

Description automatically generated

3. Try feature scaling and then apply K-Means on the scaled features. Did that improve the Silhouette score? If Yes, can you justify why

Graphical user interface, text, application, email

Description automatically generated

**3)Try feature scaling and then apply K-Means on the scaled features. Did that improve the Silhouette score? If yes, can you justify why**

A.) Accuracy not got improved, scaling the feature is required based on the use case and features what we are dealing, if two features’ scales in different in nature then scaling is required to put in same range.