**ASSIGNMENT-4**

**NAME: PAVANI REDDY GANGA SARAM.**

**ID: 700734357**

**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.**

**OUTPUTS:**

Read\_csv() is used to read the dataset and train\_test\_split() is used to split the data into training and testing sets.

**Graphical user interface, text, application, email

Description automatically generated**

Mean\_squared\_error is used to find the mean square error.

**Graphical user interface, text, application

Description automatically generated**

**Graphical user interface, text

Description automatically generated**

**Chart, scatter chart

Description automatically generated**

**Text

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.**

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

**• Calculate the silhouette score for the above clustering.**

**OUTPUTS:**

All the missing values are replaced with mean using a simple imputer and the model is fitted with the resulting data.

**Graphical user interface, text

Description automatically generated**

**Table

Description automatically generated**

**Graphical user interface, application

Description automatically generated**

The elbow graph bends at 2 so we consider this as the number of clusters.

**Graphical user interface, text, application

Description automatically generated**

**Chart, line chart

Description automatically generated**

The Silhouette score is reduced after scaling

**Graphical user interface, text, application

Description automatically generated with medium confidence**

**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.**

Silhouette is reduced after scaling

**Graphical user interface, text, application, email

Description automatically generated**