Machine learning-Assignment4

Student id:700674089

Student Name: Jahnavi Danda

CRN: CS571013466

**GitHub Repository**

https://github.com/jxd40890/Machine-Learning-Assignment4.git

**Assignment4 video URL**

https://drive.google.com/drive/folders/1mL8MAGIcK8H\_MrGYz8xt7ub\_Wz70mvJD?usp=sharing

Question1

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.

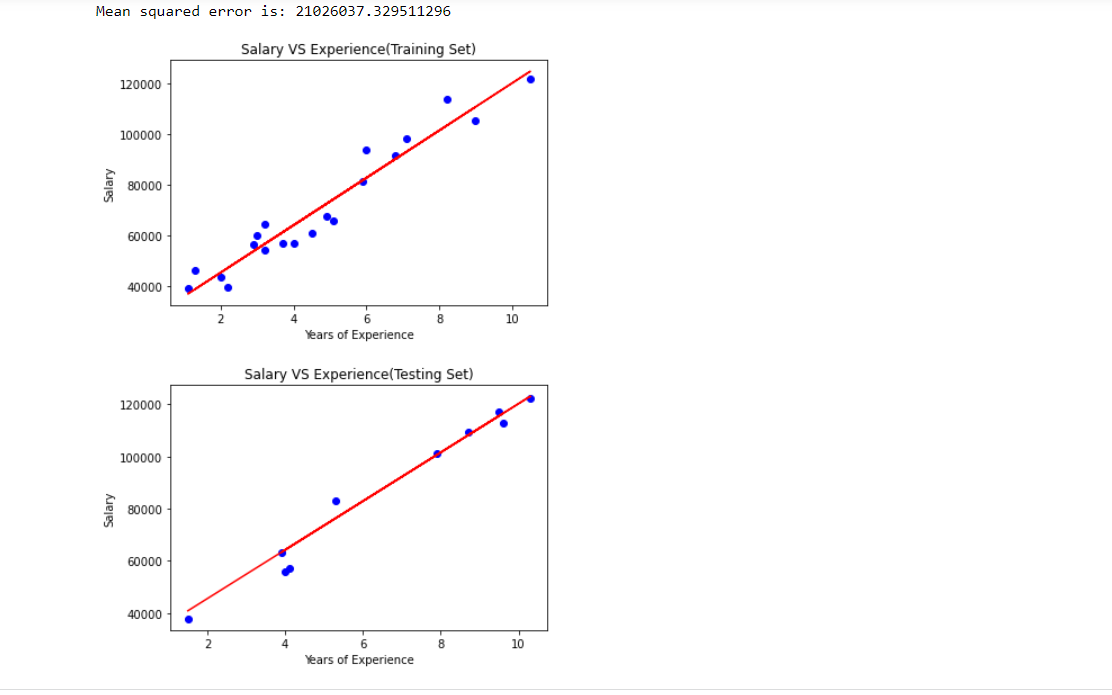
d. Calculate the mean\_squared error

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





Output:



Question2: -

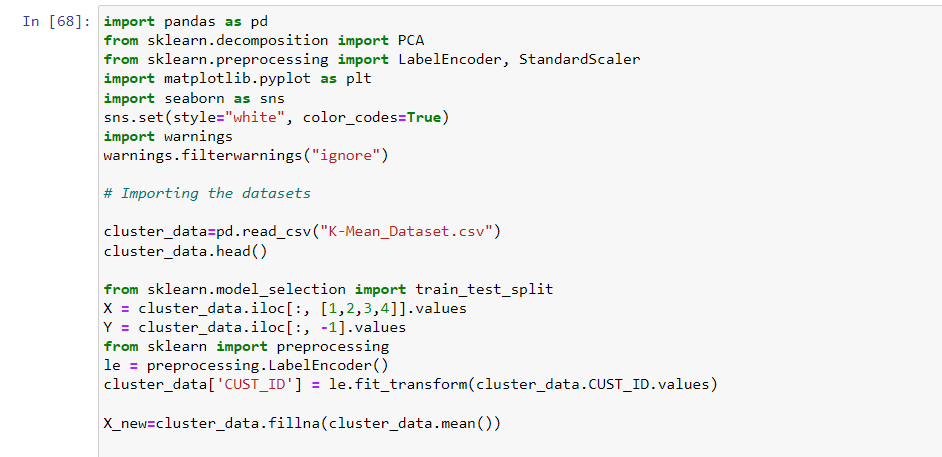
1. Apply K means clustering in the dataset provided:

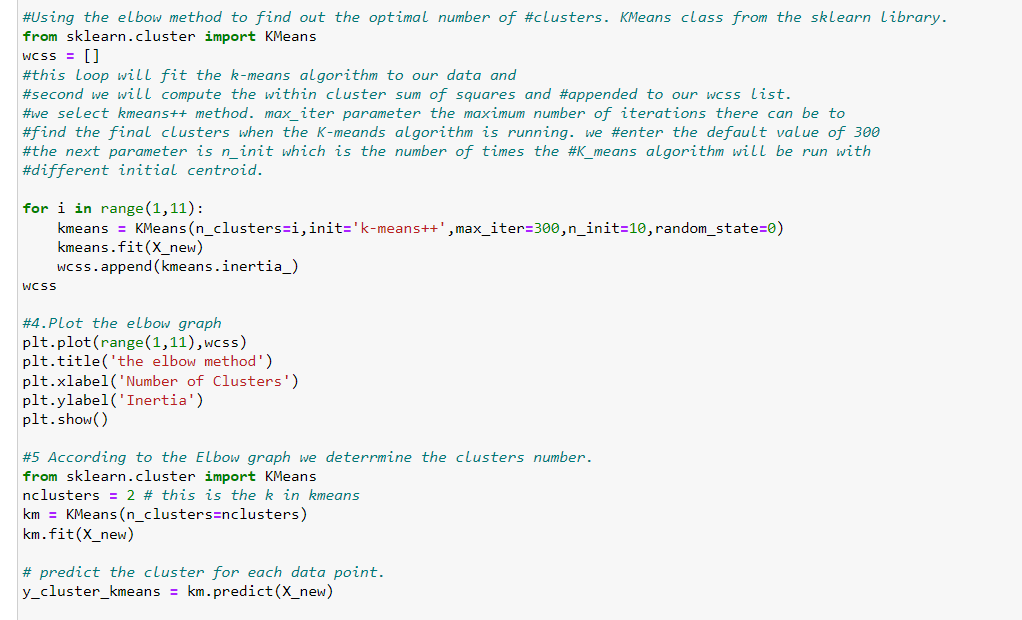
• Remove any null values by the mean.

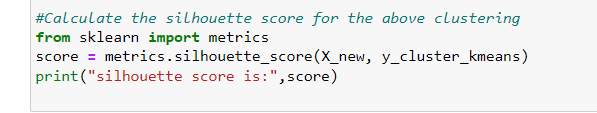
• Use the elbow method to find a good number of clusters with the K-Means algorithm

• Calculate the silhouette score for the above clustering

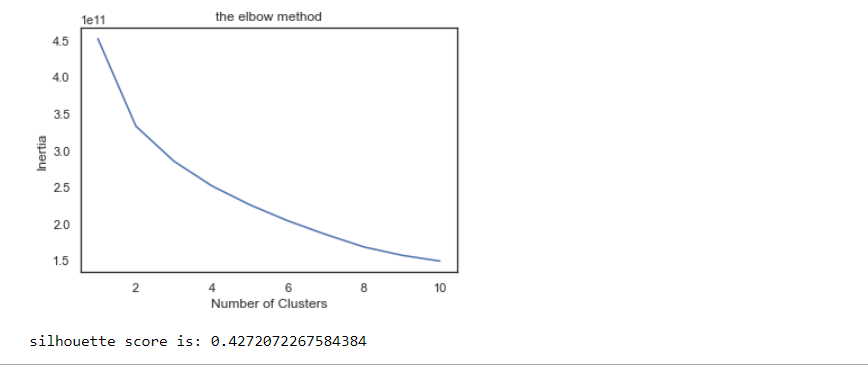
**Answer**

****

****



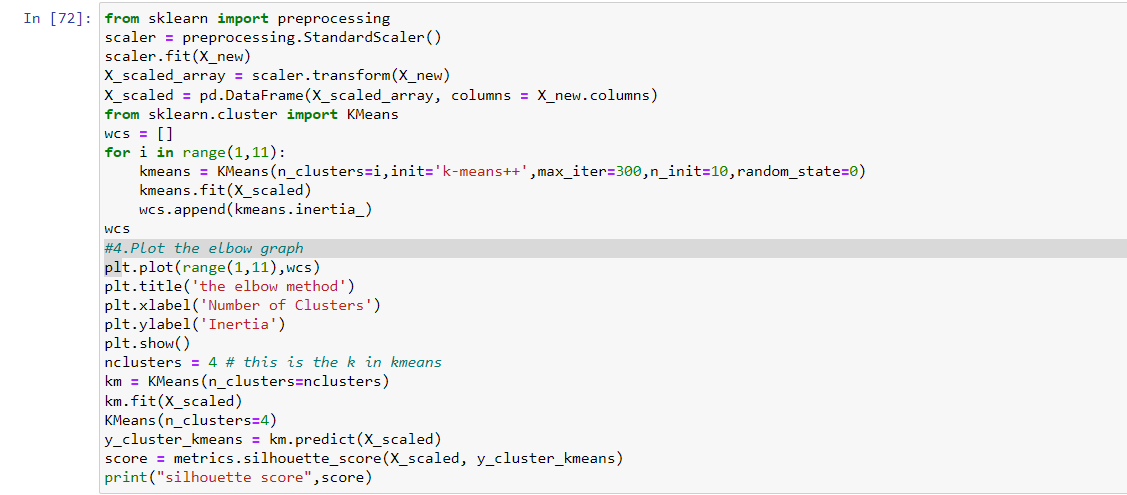
Output



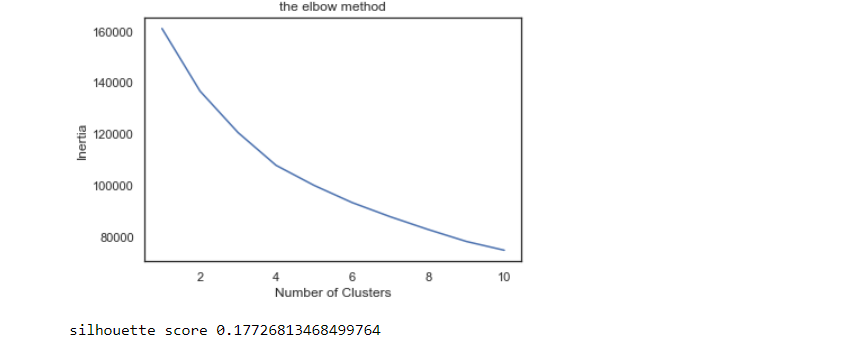
Question3: -

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

**Code Screenshot**



**Output Screenshot:**



**The silhouette score is not improved after feature scaling.**

The silhouette score near to 0 means clusters are indifferent, or we can say that the distance between clusters is not significant.