**Name: Avirneni vinay kumar rao**

**ID: 700734549**

**Course : Machine learning**

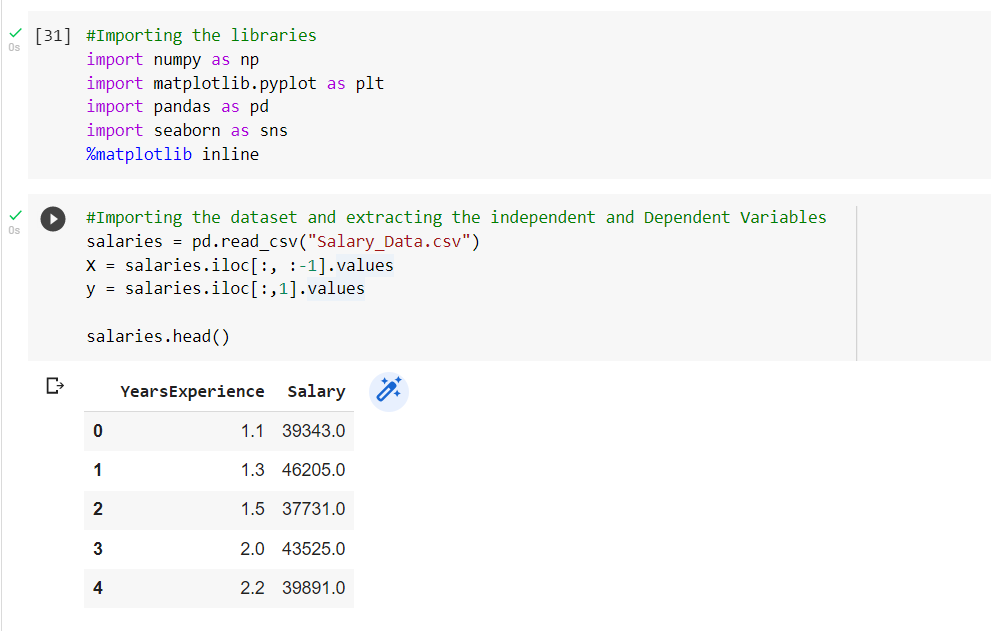
**Github link:** [**https://github.com/vinay779/assignment.4**](https://github.com/vinay779/assignment.4)

**Video link:** [**https://youtu.be/bdQTL-\_5VWM**](https://youtu.be/bdQTL-_5VWM)

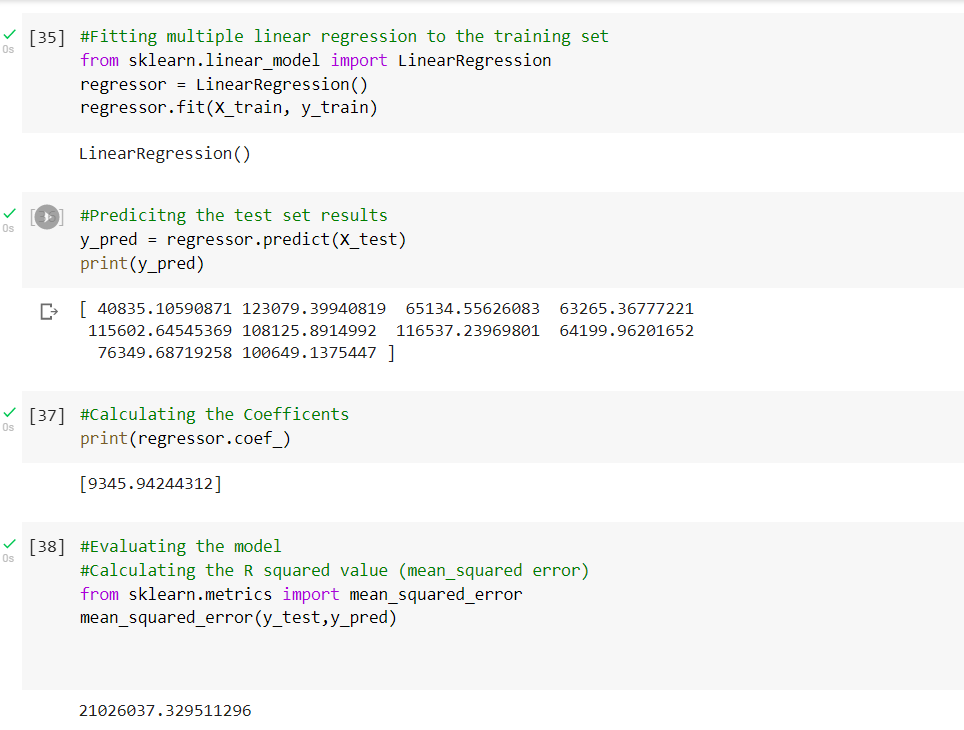
**Question 1.**

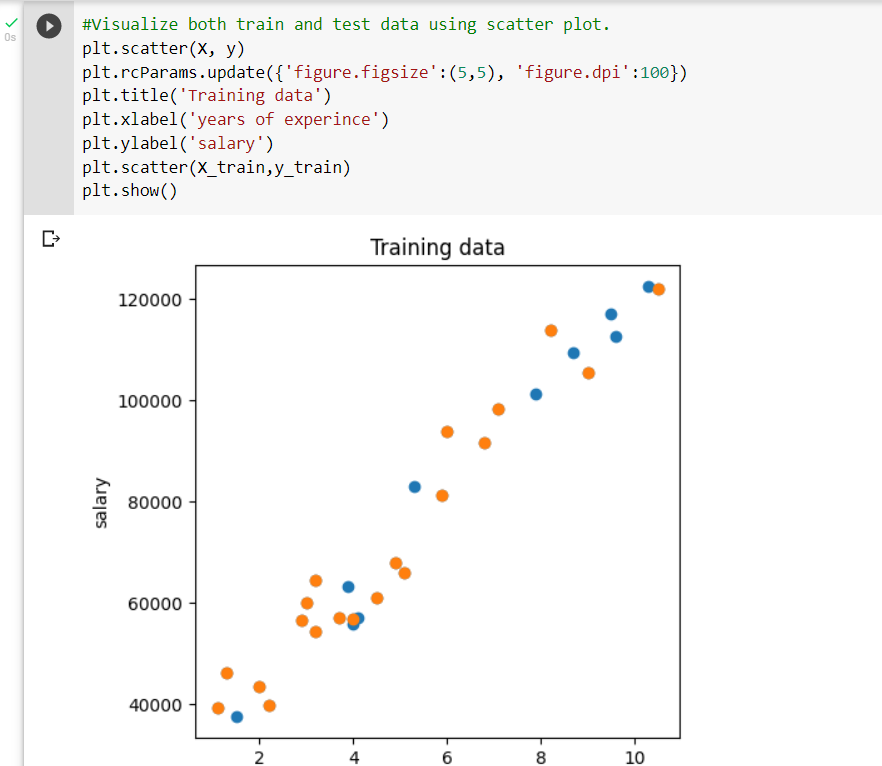
**Apply Linear Regression to the provided dataset using underlying steps.**

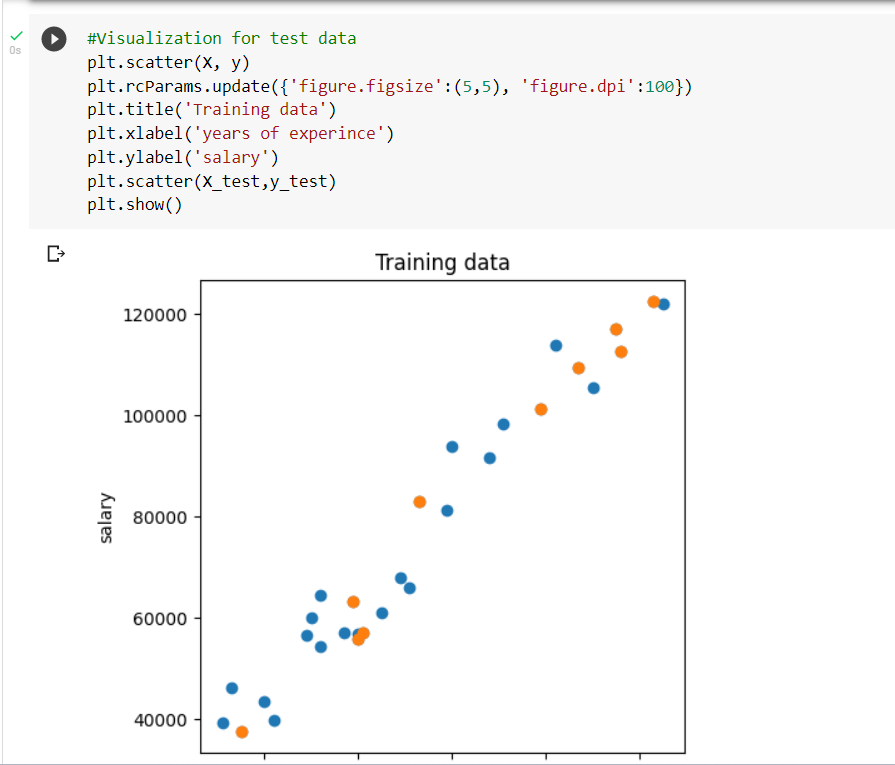
* **Importing the given “Salary\_Data.csv”**
* **Splitting the data in train test partitions, such that 1/3 of the data is reserved as test subset.**
* **Training and predicting the model.**
* **Calculating the mean squared error**
* **Visualizing both train and test data using scatter plot.**

****

****

****

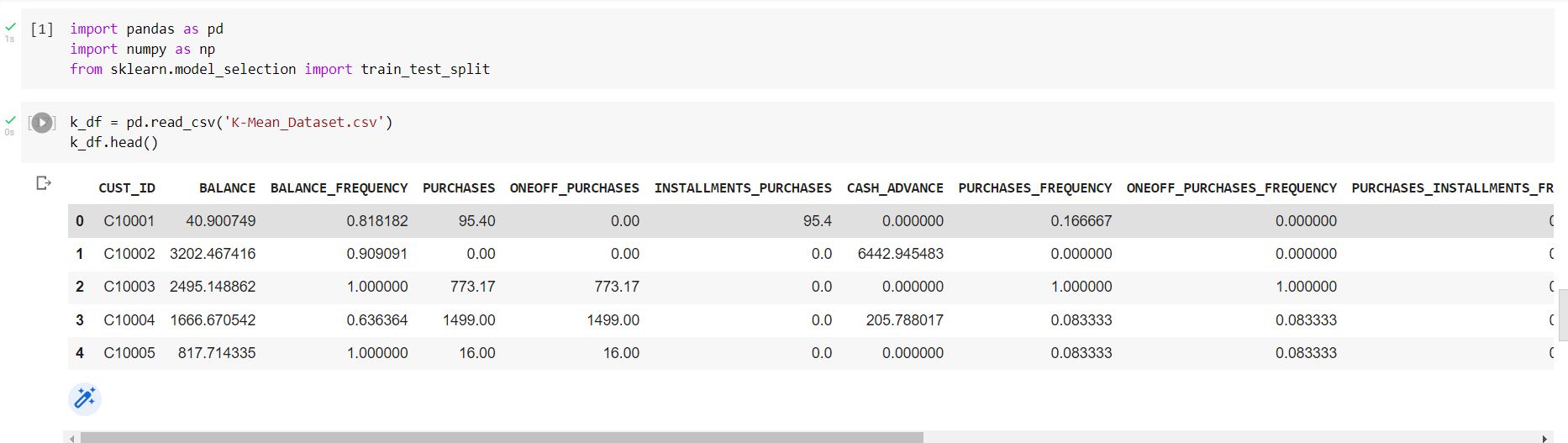
****

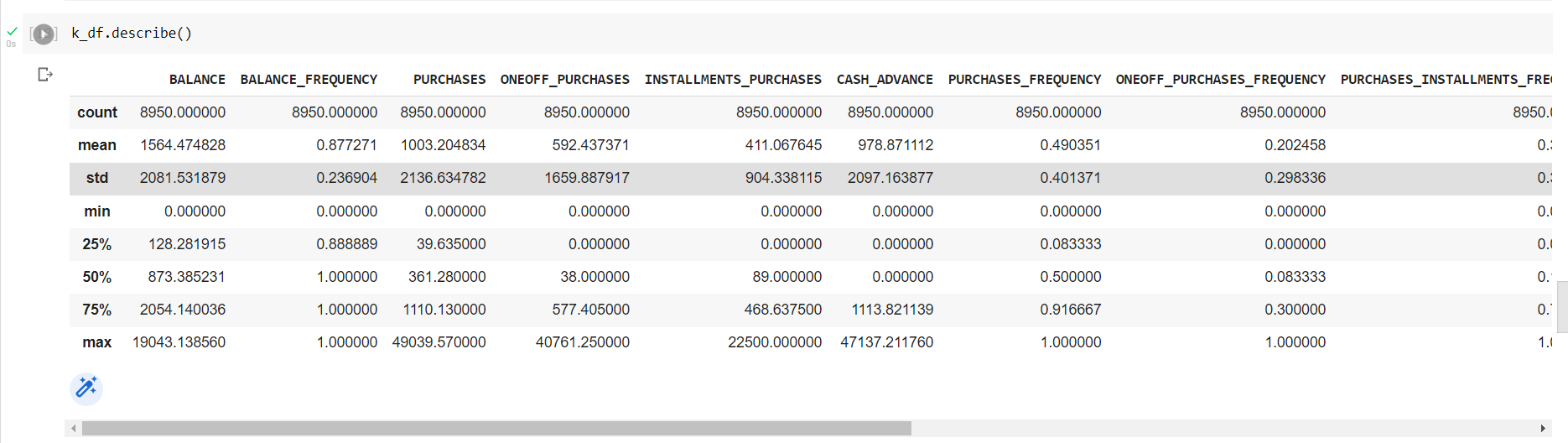
****

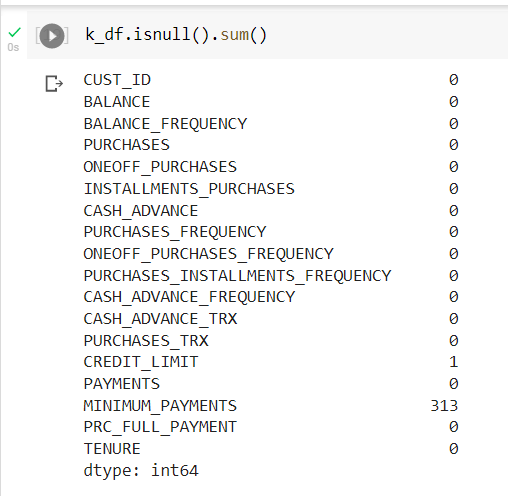
**Question 2:**

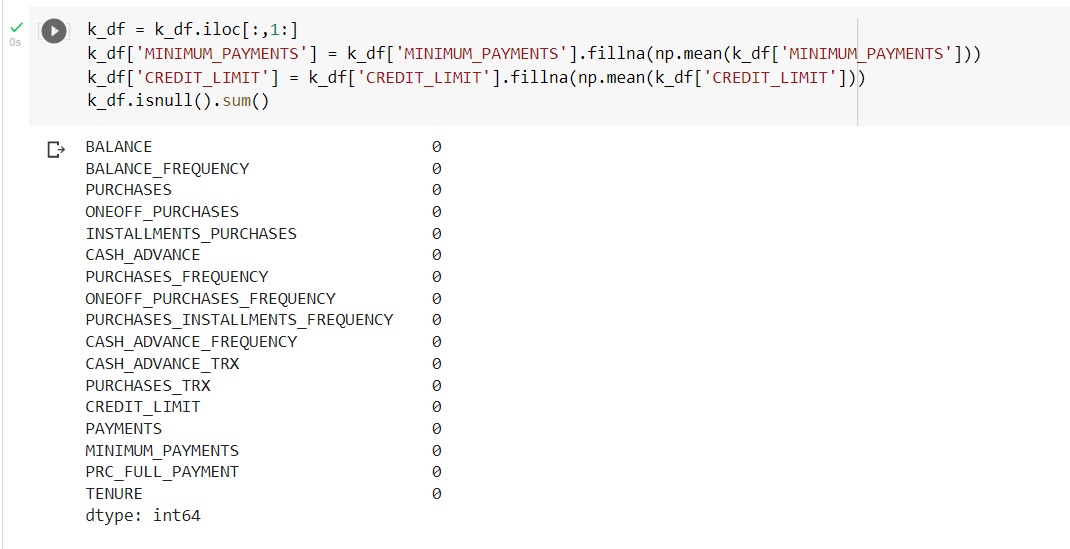
**Apply K means clustering in the dataset provided**

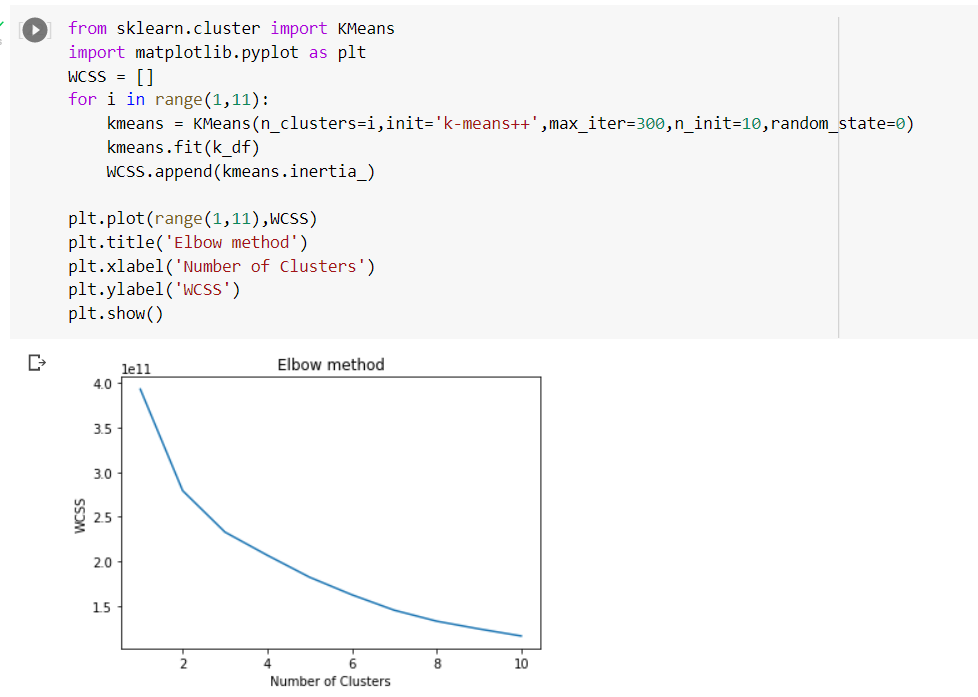
* + **Removing any null values by the mean.**
  + **By using the elbow method to find a good number of clusters with the K-Means algorithm**
  + **Calculating the silhouette score for the above clustering**

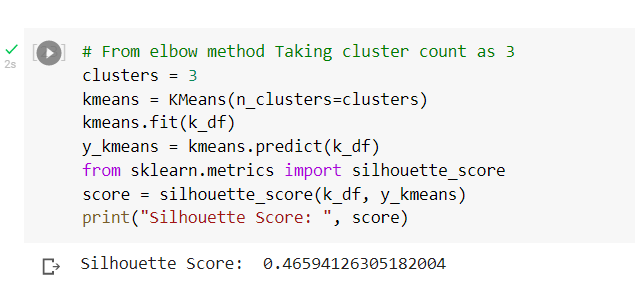
****

****

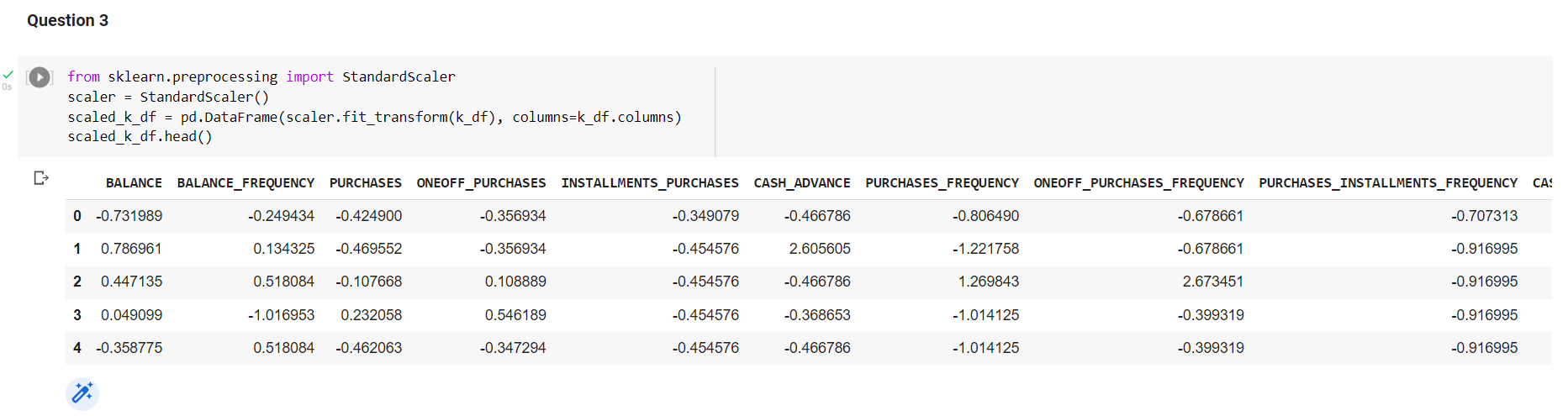
****

****

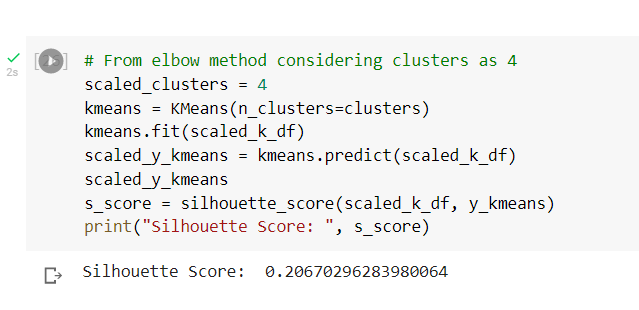
****

****

**Question 3:**  **By using feature scaling and then applying K-Means on the scaled features.**

****

****

****

**By observing the silhouette score we can say that On applying the feature scale, the value comes under the same range therefore it is difficult to predict the values.**