**NAME: AQSA TAUHEED (2303.KHI.DEG.011)**

**PAIRING** WITH **: MOHAMMAD HUSSAM(2033.KHI.DEG.020)**

**&**

**MAVIA ALAM KHAN (2303.KHI.DEG.017)**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

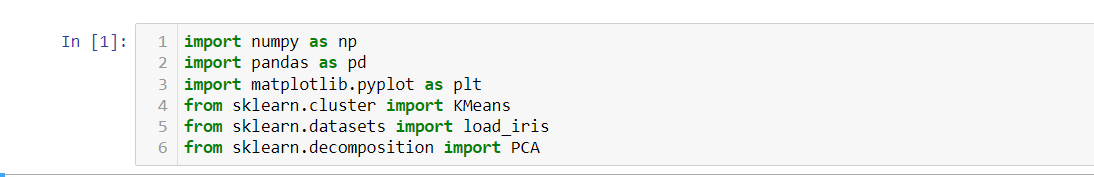
**ASSIGNMENT NO : 3.3**

Perform k-means clusterization on the Iris  
dataset. Repeat the procedure on the  
dataset reduced with PCA, and then  
compare the results.

**Solution:**

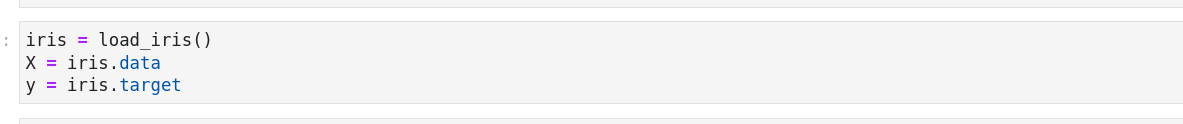
**Step#1:**

First of all we have to import the necessary libraries.



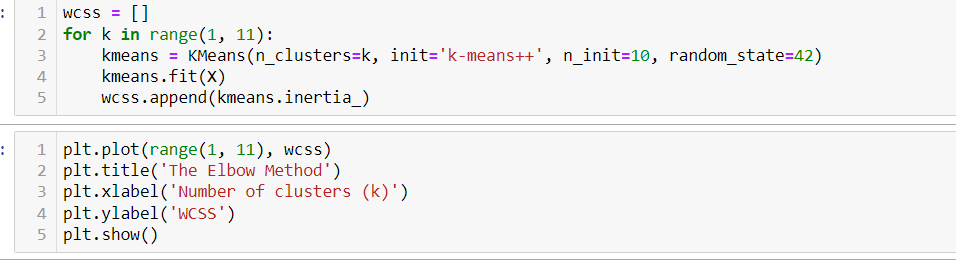
**Step#2:**

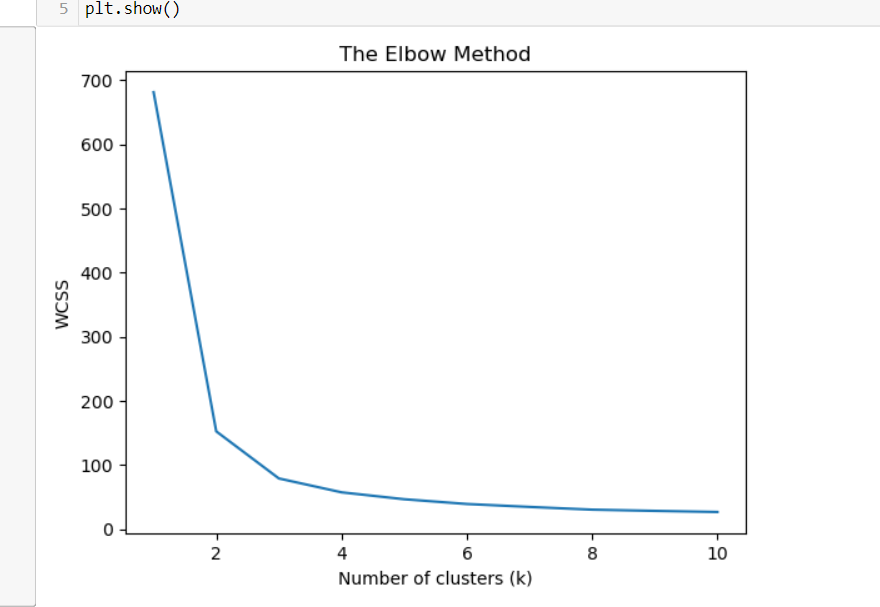
After that we load the iris data set.



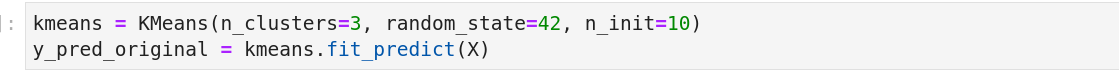
**Step#3:**

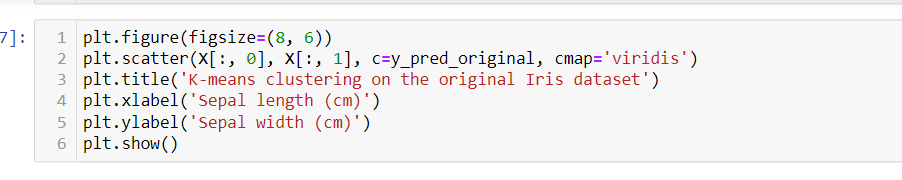
We use elbow method to find optimal number of cluster in the original data set

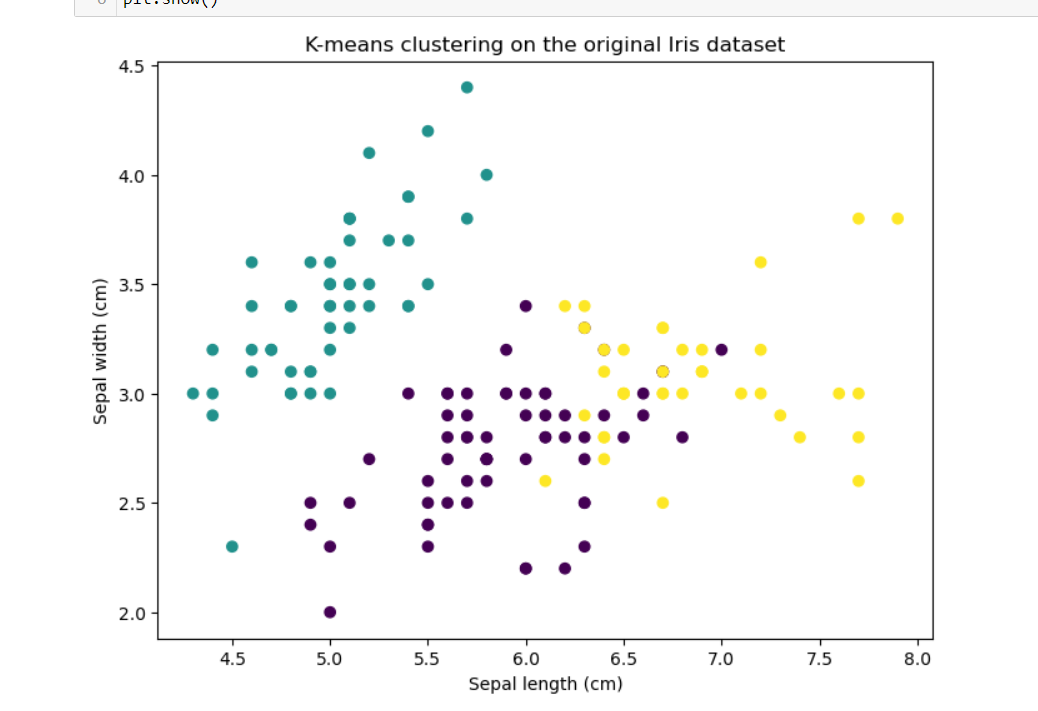




now in the above graph we see that the optimal number is 3 .so we provided, the n\_clusters parameter is set to 3, which means we are performing K-means clustering with 3 clusters. This means that the algorithm will try to group the data points into three distinct clusters based on their similarity.





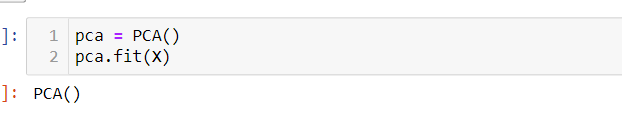


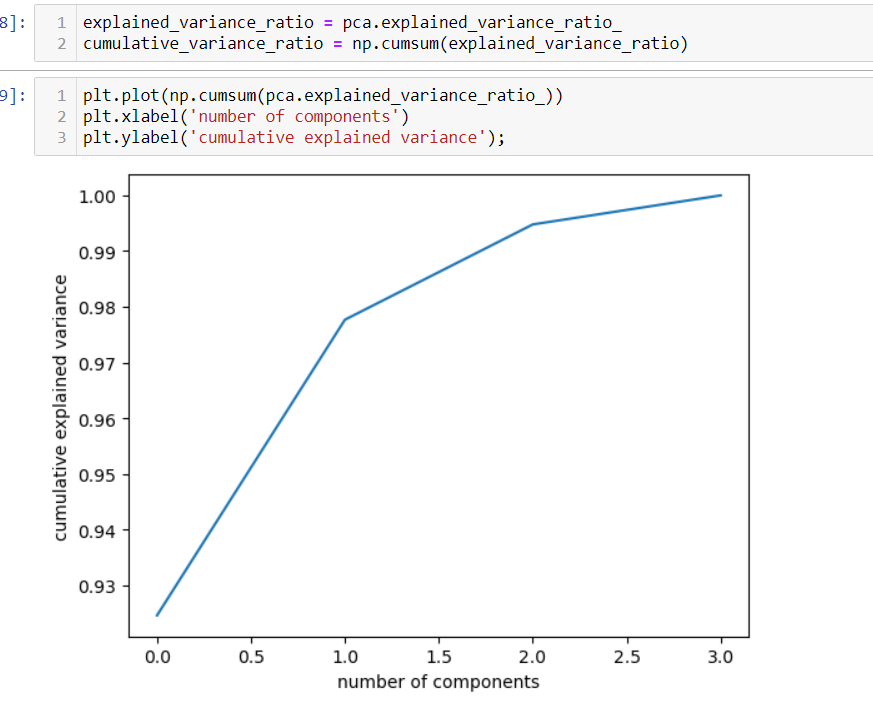
**Step # 4:**

Applying PCA to reduce the dimensionality of the dataset .

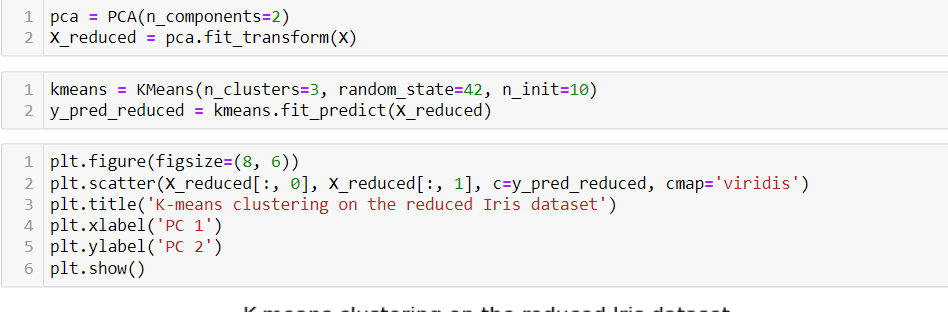
Now first we use cumulative explained variance to find out the n number of components and then we reduced it .

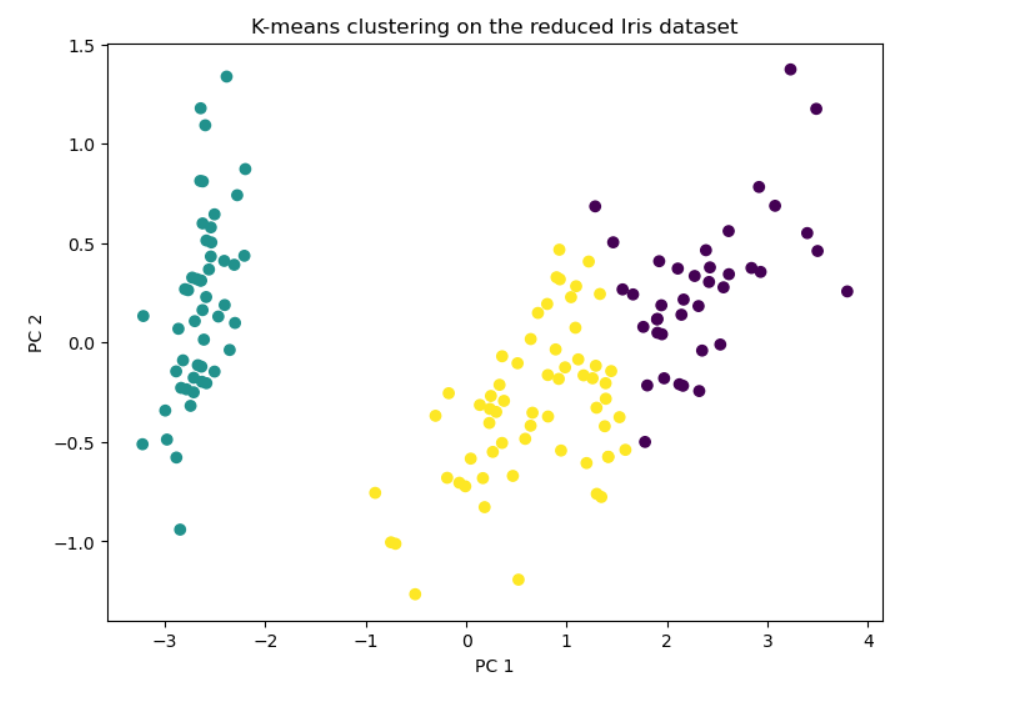
We reduce our data because beneficial when dealing with high-dimensional data, as it helps to simplify the dataset and reduce the computational complexity of subsequent analysis or modeling tasks.





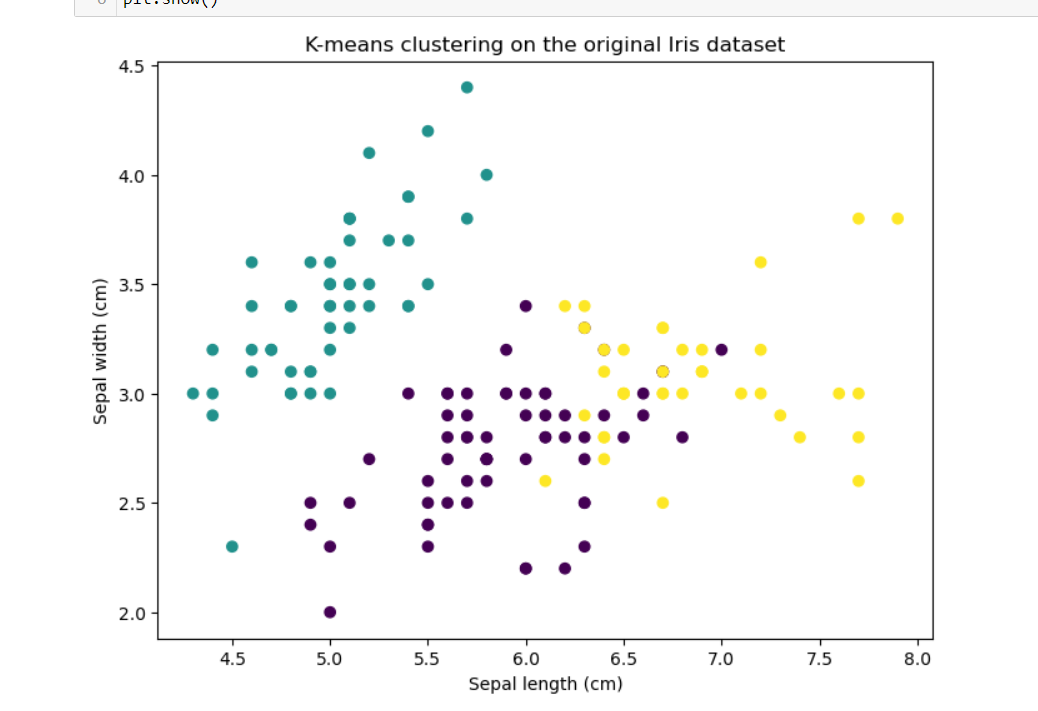
So in the above graph we see that number of components that is 2 It determines the number of components or dimensions that reduce our data. n\_component=2 means you are reducing the dimensionality to 2 components.





NOW CAMPARISION :

BEFORE



AFTER

