**Question ) Perform Principal component analysis and perform clustering using first 3 principal component scores (both heirarchial and k mean clustering(scree plot or elbow curve) and obtain optimum number of clusters and check whether we have obtained same number of clusters with the original data (class column we have ignored at the begining who shows it has 3 clusters)df**

**Soln. :**

**Objective :**

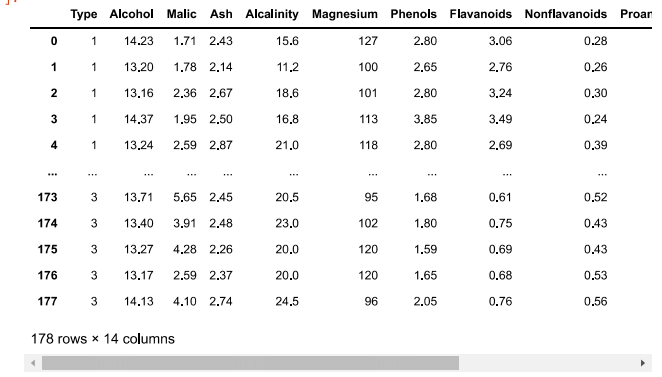
Use Scree plot or elbow curve and obtain optimum number of clusters and check whether we have obtained same number of clusters with the original data

STEPS TO BE FOLLOWED TO OBTAIN SOLUTION:

STEP1:

Loading of wine data on which PCA method are applied. wine data set is of 8 rows and 15

columns. The details of the data is



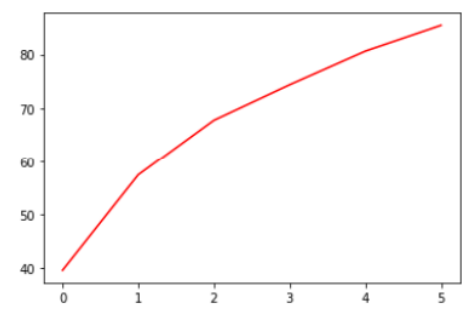
STEP2:

AFTER describing data ,to obtain PCA s PCA module is imported from sklearn.decomposition . PCA technique is applied only on normalized data. For normalization Scale module is imported from sklearn preprocessing .

STEP3:

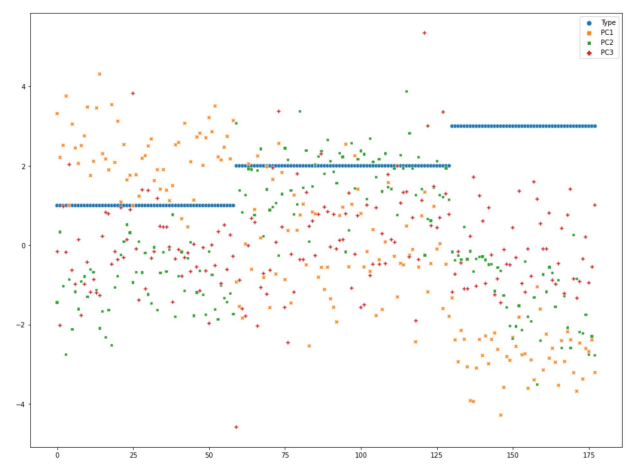
For n\_components = 13 ;PCA is obtained and its fitted to normalized wine data. Variance plot for PCA

components obtained



STEP 4 :

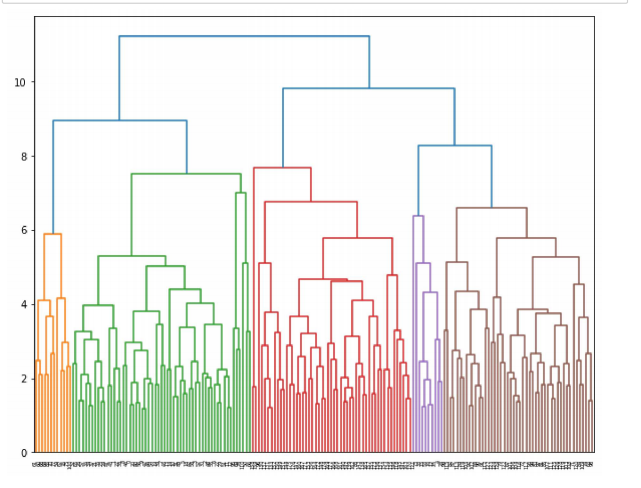
Later pca data is concated to main data index column and Scatter diagram is plotted



STEP 5:

check whether we have obtained same number of clusters with the original data Clustring techniques is applied on original wine data.

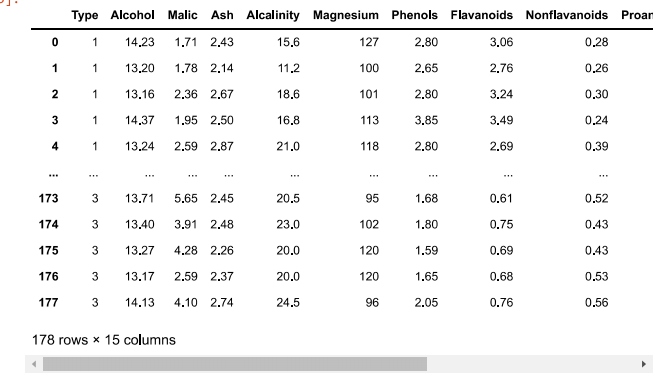
Dendogram is drawn for original wine data



STEP7:

After plotting Dendogram for both the data cluster groups were obtained from

Agglomerative Clustering which is imported from sklearn.cluster by complete method cluster grouping for main data



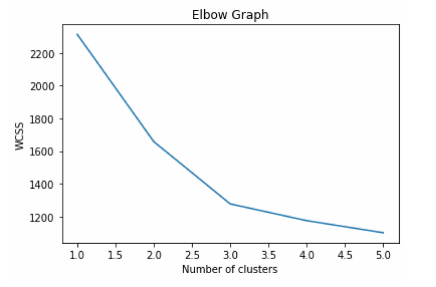
STEP 8:

Same way kmeans clustering is applied on both the data sets of given and after PCA formed. to do kmeans ,the regarding module kmeans is imported from sklearn.cluster.

STEP 9:

scree plot or elbow curve is plotted between no.of clusters picked and total data points with in sum of squares.

scree plot is plotted for PCA applied data



**Summary :**

Using Scree plot or elbow curve we can determine the obtain optimum number of clusters obtained same number of clusters with the original data