**PRINCIPAL COMPONENT ANALYSIS**

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

**Analysis**

#Wine dataexamples:

library(gdata)

PCA<-read.csv("F:/Data Scientist/PCA 4.3/wine.xlsx.csv")

pca<-princomp(PCA[,2:7], cor=TRUE,scores=TRUE,covmat=NULL)

Summary(pca)

Result;

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6

Standard deviation 1.3680647 1.2694218 1.0037308 0.8528145 0.69675380 0.54473240

Proportion of Variance 0.3119335 0.2685720 0.1679126 0.1212154 0.08091098 0.04945556

Cumulative Proportion 0.3119335 0.5805055 0.7484180 0.8696335 0.95054444 1.00000000

Pca$scores

pca$scores

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6

[1,] 2.51084586 0.99724090 0.393273670 -0.918907445 0.24619486 0.055786410

[2,] 1.96879267 -1.45239434 0.758359215 -0.195994869 -0.20371658 -0.939703186

[3,] 0.63206208 0.85547850 -0.332309617 0.588327081 -0.21313122 -0.609105692

[4,] 2.90502038 1.17222012 0.033704421 0.798237493 -0.58477591 0.393652147

[5,] 0.56136922 2.22187083 -0.715775908 -0.089531316 -0.30895040 -0.401167712

.

.

. [165,] -1.17529521 0.09992953 1.064067431 0.141125651 1.24599663 0.601413285

[166,] -1.91249995 0.40435983 1.989702211 0.239301758 0.44582318 0.605585652

**After applying K-means clustering we have obtained different number of clusters.**