Perform clustering (Both hierarchical and K means clustering) for the airlines data to obtain optimum number of clusters.

Draw the inferences from the clusters obtained.

**Ans :**

**R Code :**

## Hierarchal Clustering

########## Crime Data Set #########

# Data Load

# Data Load

mydata1 <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Clustering\\crime\_data.csv')

mydata <- scale(mydata1[,2:5])

d <- dist(mydata, method = "euclidean") # Computing the distance matrix

fit <- hclust(d, method = "average") #Computing the algorithm # try with centroid

plot(fit) #display dendrogram

groups <- cutree(fit, k=4) #cut tree into 4 clusters

#draw dendrogram with red borders around the 5 clusters

rect.hclust(fit, k=4, border = "red")

#attach the cluster number to universities

clusters=data.frame('ID'=mydata1[,1],'Cluster'=groups)

fit1 <- hclust(d, method = "centroid") #Computing the algorithm # trying with centroid

plot(fit1) #display dendrogram

groups <- cutree(fit1, k=4) #cut tree into 4 clusters

#draw dendrogram with red borders around the 5 clusters

rect.hclust(fit1, k=4, border = "red")

#attach the cluster number to universities

clusters1=data.frame('ID'=mydata1[,1],'Cluster'=groups)

write.csv(as.data.frame(clusters), file = "clusters-average.csv")

write.csv(as.data.frame(clusters1), file = "clusters-centroid.csv")

## K-means Clustering ##

crime\_data <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Clustering\\crime\_data.csv')

data <- scale(crime\_data[2:5])

plot(data)

#Elbow Chart

wss <- c()

for (i in 2:15) wss[i] <- sum(kmeans(data, centers = i)$withinss)

plot(1:15,wss,type = "b", xlab = "No of Clusters", ylab = "Avg Distance")

### Cluster algorithm building

km <- kmeans(data, 8)

km$centers

km$cluster

##Animation

install.packages("animation")

library(animation)

windows()

km <- kmeans.ani(data, 8)

**Results :**

Clusters enclosed as different CSV files.

**Plots :**









