**Clustering on EastWestAirlines**

#hierarchical clustering

library(readxl)

AirLine\_DF<-read\_excel("E:\\Data Science\\Assignments\\EastWestAirlines.xlsx",sheet="data")

AirLine\_DF$cc1\_miles = ifelse(AirLine\_DF$cc1\_miles==1,2500,

ifelse(AirLine\_DF$cc1\_miles==2,7500,

ifelse(AirLine\_DF$cc1\_miles==3,17500,

ifelse(AirLine\_DF$cc1\_miles==4,32500,

ifelse(AirLine\_DF$cc1\_miles==5,50000,0)))))

AirLine\_DF$cc2\_miles = ifelse(AirLine\_DF$cc2\_miles==1,2500,

ifelse(AirLine\_DF$cc2\_miles==2,7500,

ifelse(AirLine\_DF$cc2\_miles==3,17500,

ifelse(AirLine\_DF$cc2\_miles==4,32500,

ifelse(AirLine\_DF$cc2\_miles==5,50000,0)))))

AirLine\_DF$cc3\_miles = ifelse(AirLine\_DF$cc3\_miles==1,2500,

ifelse(AirLine\_DF$cc3\_miles==2,7500,

ifelse(AirLine\_DF$cc3\_miles==3,17500,

ifelse(AirLine\_DF$cc3\_miles==4,32500,

ifelse(AirLine\_DF$cc3\_miles==5,50000,0)))))

data = scale(AirLine\_DF)

d <- dist(data[,2:11], method = "euclidean")

fit <- hclust(d, method="ward.D2")

fit <- as.dendrogram(fit)

library(dendextend)

cd = color\_branches(fit,k=3) #Coloured dendrogram branches

plot(cd)

clusters <- cutree(fit, k=3)

table(clusters)

g1 = aggregate(AirLine\_DF[,2:11],list(clusters),median)

data.frame(Cluster=g1[,1],Freq=as.vector(table(clusters)),g1[,-1])

centroid = function(i, dat, clusters)

{

ind = (clusters == i)

colMeans(dat[ind,])

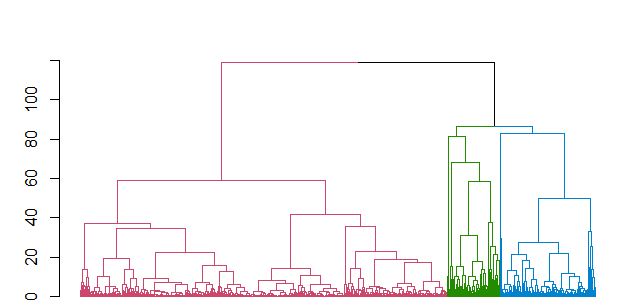
}

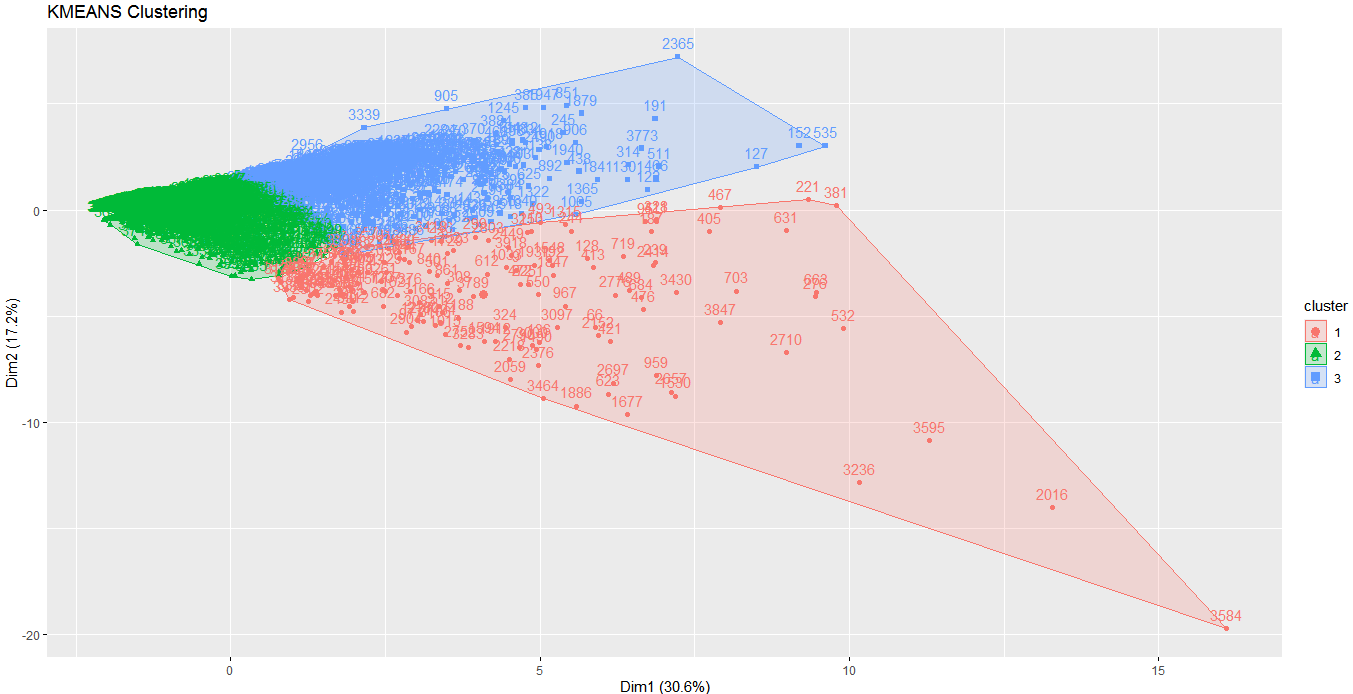
sapply(unique(clusters), centroid, AirLine\_DF[,2:11], clusters)

**#Kmeans clustering**

km.3 <- eclust(data[,2:11], "kmeans", k = 3, nstart = 25, graph = TRUE)

fviz\_silhouette(km.3)





Conclusion -

For the non-frequent flyers who are more in numbers promotions like more miles per fly, discounted air fare rates can be offered to improve the number of flyers.

These offers would help customers to fly frequently.