# Clustering activity

Name: Hemanth sai kumar

RegNO: 12220407

Code:-

data1<-read.csv("snsdata.csv")

sum(is.na(data1))

# Replace NAs in each column with the column's mean

for (i in 1:ncol(data1)) {

data1[is.na(data1[, i]), i] <- mean(data1[, i], na.rm = TRUE)

}

data<-data.frame(data1[,-2])

d=dist(data,method="euclidean")

hfit=hclust(d,method = "average")

plot(hfit)

grps<-cutree(hfit,k=3)

grps

rect.hclust(hfit,k=3,border = "red")

sum(is.na(data))

# data<-na.omit(data)

set.seed(123)

kmeans.re<-kmeans(data,centers=2,nstart = 1000)

kmeans.re

kmeans.re$cluster

kmeans.re$centers

cm<-table(data1$gender,kmeans.re$cluster)

cm

library(cluster)

clusplot(data[,c("gradyear","age")],

kmeans.re$cluster,

lines = 0,

shade = TRUE,

color = TRUE,

labels = 2,

plotchar = FALSE,

span=TRUE,

main=paste("SNS cluster"),

xlab="gradyear",

ylab="age"

)

