Project

Harika

04/12/2019

library(factoextra)

## Loading required package: ggplot2

## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ

library(hrbrthemes)

## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.

## Please use hrbrthemes::import\_roboto\_condensed() to install Roboto Condensed and

## if Arial Narrow is not on your system, please see http://bit.ly/arialnarrow

library(GGally)

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(viridis)

## Loading required package: viridisLite

Problem Statement: CRISA has traditionally segmented markets on the basis of purchaser demographics. They would now like to segment the market based on two key sets of variables more directly related to the purchase process and to brand loyalty: 1. Purchase behavior (volume, frequency, susceptibility to discounts, and brand loyalty) 2. Basis of purchase (price, selling proposition)Doing so would allow CRISA to gain information about what demographic attributes are associated with different purchase behaviors and degrees of brand loyalty, and thus deploy promotion budgetsmore effectively. More effective market segmentation would enable CRISA’s clients (in this case, a firm called IMRB)to design more cost-effective promotions targeted at appropriate segments. Thus, multiple promotions could be launched, each targeted at different market segments at different times of the year. This would result in a more cost-effective allocation of the promotion budget to different market segments. It would also enable IMRB to design more effective customer reward systems and thereby increase brand loyalty.

#Reading the data

data <- read.csv("BathSoap.csv")  
mydata\_norm <- as.data.frame(scale(data))

#Brand\_Loyal is calculated by the maximum value of all eight brands and comparing with other999 brand.

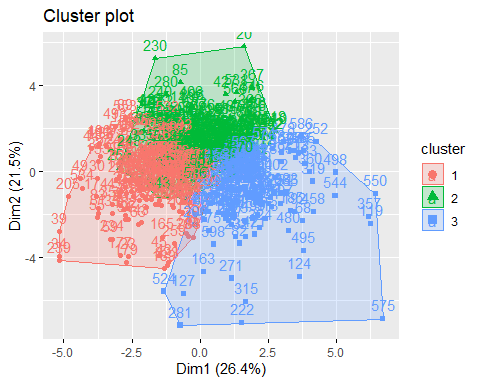
BrandLoyal<-data[,23:30]  
Max\_Brand<-apply(BrandLoyal, 1, max)  
Max\_Brand<-as.data.frame(Max\_Brand)  
BrandComp<-cbind(Max\_Brand,data[,31])  
MaxLoyal<-as.data.frame(ifelse(BrandComp[,1]>=BrandComp[,2],1,0))  
MaxLoyalComp<-cbind(BrandComp,MaxLoyal)  
BathSoapComp<-cbind(data,MaxLoyal)  
names(BathSoapComp)[47] <- "Brand\_Loyal"

#1.(a)Purchase Behavior Purchase behavior variables includes volume, frequency, susceptibility to discounts, and brand loyalty

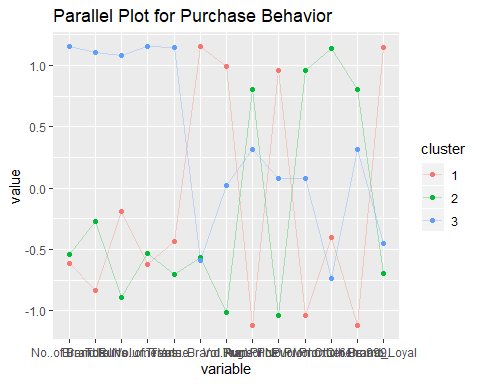
#Variable selection  
data1 <- BathSoapComp[,c(12:22,31,47)]  
data1 <- as.data.frame(scale(data1))  
  
#k-means  
set.seed(123)  
data1\_K3 <- kmeans(data1, centers =3 , nstart = 200) # k = 3  
  
# Size and Center for the Clusters  
data1\_K3$size

## [1] 200 237 163

data1\_clusters <- data.frame(data1\_K3$centers)  
cluster <- matrix(c("1","2","3"),nrow = 3)  
data1.visual <- cbind(cluster,data1\_clusters)  
  
#Cluster visiualization   
fviz\_cluster(data1\_K3, data = data1)



#Parallel Plot  
ggparcoord(data1.visual,  
 columns = 2:14, groupColumn = 1,   
 showPoints = TRUE,   
 title = "Parallel Plot for Purchase Behavior",  
 alphaLines = 0.3  
 )

 when K=2 the clusters are CLuster1:n=223 High loyal customers,favoring no.of brand,brand runs low,Others 999,promotion 6% are low.All are other parameters like Total volume,Trans.brandRuns,No.of trans are high values. Cluster2:n=377 Low Brand Loyal customers with characteristics opposite to cluster1.

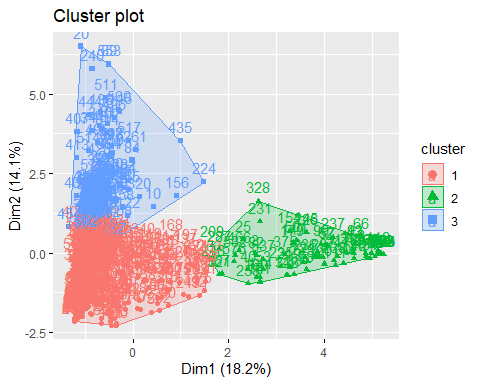
when K=3 the clusters are Cluster1:n=200 High Loyal customers,favoring low no.of brands,brand runs,avr.price,no promotion,others999.Mid range of ohter promotions.High Volume transaction. Cluster2:n=237 Low brand loyal customers,favoring towards different promotion types and others 999 brands more. Cluter3:n=163:It is having customers with high No of brands,brand runs,total volume.Middling all other factors.

#(b)Basis of purchase Basis of purchase (price, selling proposition)

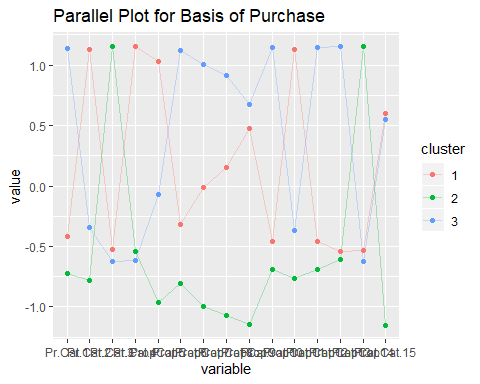
data2 <- BathSoapComp[,c(32:46)]  
data2 <- data.frame(scale(data2))  
set.seed(123)  
data2\_k3 <- kmeans(data2, centers =3 , nstart = 500)  
  
  
# Size and Center for the Clusters  
data2\_k3$size

## [1] 376 79 145

data2\_clusters <- data.frame(data2\_k3$centers)  
cluster <- matrix(c("1","2","3"),nrow = 3)  
data2.visual <- cbind(cluster,data2\_clusters)  
  
#cluster visualization  
fviz\_cluster(data2\_k3, data = data2)



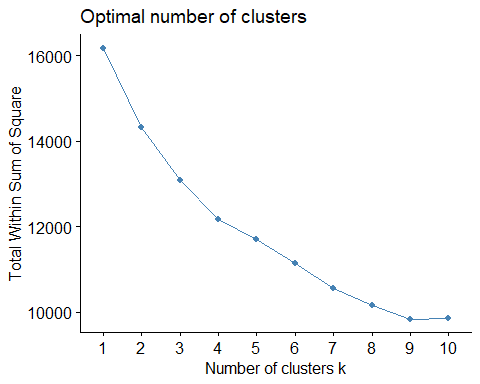
#Parallel Plot  
ggparcoord(data2.visual,  
 columns = 2:16, groupColumn = 1,   
 showPoints = TRUE,   
 title = "Parallel Plot for Basis of Purchase",  
 alphaLines = 0.3  
 )

 When k=2 Cluster1:n=78 which has Price Cat:1,2,4 and prop cat 4 to 13 ,15 are low and Price catagory 3,prop cat.14 are high Cluster2 :n=522 which shows excatly opposite to cluster1 In this k=2 analysis it is unclear so about customers trend in most of the parameters.

when k=3 Cluster1:n=376 which has Pr.cat.2 ,4 high.Middling all other values Cluster2:n= 79 which has characteristics price cat 3 and Prop.Cat14 high.All other are low. Cluster3:n=145 which has Pr.cat.1,Prop.Cat.6,10,12,13 high volume.

#(c)The variables that describe both purchase behavior and basis of purchase

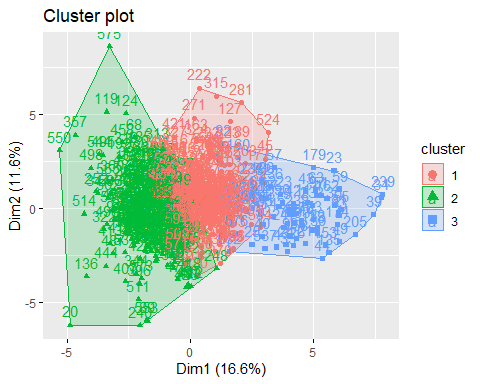
data3 <- BathSoapComp[,c(12:22,32:46,47)]  
data3 <- data.frame(scale(data3))  
  
#k-means  
set.seed(123)  
fviz\_nbclust(data3, kmeans, method = "wss")



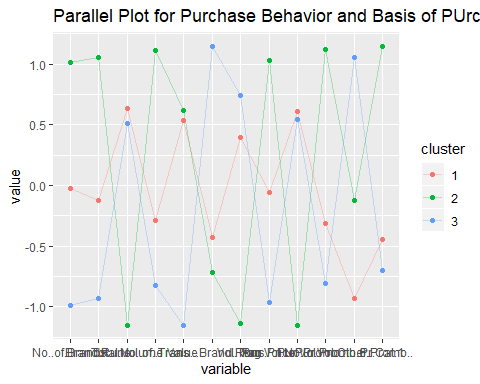
set.seed(123)  
data3\_k3 <- kmeans(data3, centers =3 , nstart = 500)   
  
# Size and Center for the Clusters  
data3\_k3$size

## [1] 280 247 73

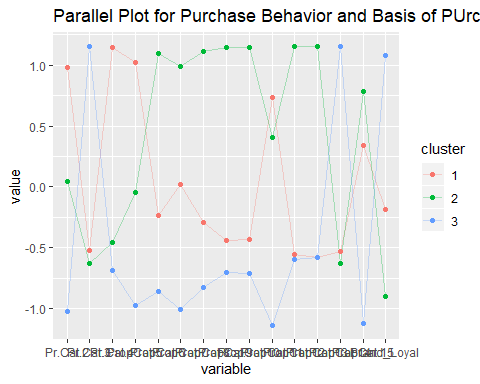
data3\_clusters <- data.frame(data3\_k3$centers)  
cluster <- matrix(c("1","2","3"),nrow = 3)  
data3.visual <- cbind(cluster,data3\_clusters)  
  
#cluster visualization  
fviz\_cluster(data3\_k3, data = data3)



#Parallel Plot  
ggparcoord(data3.visual,  
 columns = 2:13, groupColumn = 1,   
 showPoints = TRUE,   
 title = "Parallel Plot for Purchase Behavior and Basis of PUrchase",  
 alphaLines = 0.3  
 )



ggparcoord(data3.visual,  
 columns = 14:28, groupColumn = 1,   
 showPoints = TRUE,   
 title = "Parallel Plot for Purchase Behavior and Basis of PUrchase",  
 alphaLines = 0.3  
 )

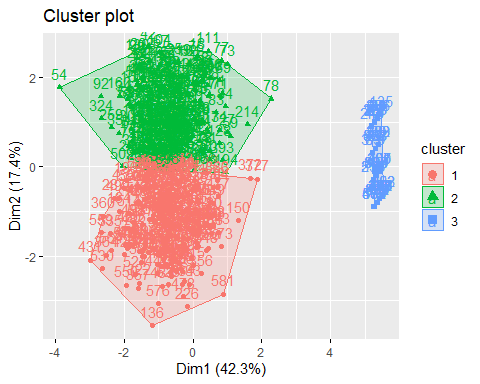
 whenk=3 Cluster3:shows high brand loyal customers and their characteristics Cluster2:Low brand loyal customers and their characteristics.

# Including Demographics

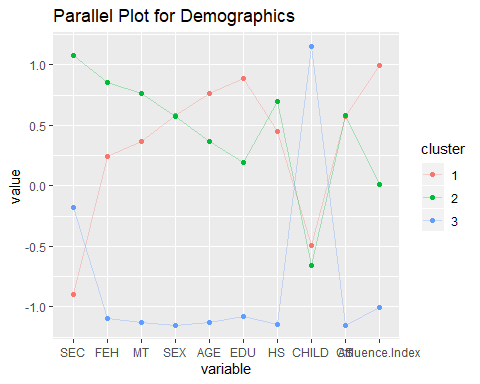
data4 <- BathSoapComp[,c(2:11)]  
data4 <- data.frame(scale(data4))  
  
#k-means  
  
set.seed(123)  
data4\_k3 <- kmeans(data4, centers =3 , nstart = 500)   
  
# Size and Center for the Clusters  
data4\_k3$size

## [1] 237 295 68

data4\_clusters <- data.frame(data4\_k3$centers)  
cluster <- matrix(c("1","2","3"),nrow = 3)  
data4.visual <- cbind(cluster,data4\_clusters)  
  
#cluster visualization  
fviz\_cluster(data4\_k3, data = data4)



#Parallel Plot  
ggparcoord(data4.visual,  
 columns = 2:11, groupColumn = 1,   
 showPoints = TRUE,   
 title = "Parallel Plot for Demographics",  
 alphaLines = 0.3  
 )



Each segmentation has it own significance depending upon the companies marketing approaches.Since demographics has been traditional segmentation,Purchase behavior can be helpful for promotion campaigns if company is looking for customers to have trusting relationship that is brand loyalty.