Project\_ML

Harika

12/7/2019

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))

#Finding Brand Loyalty

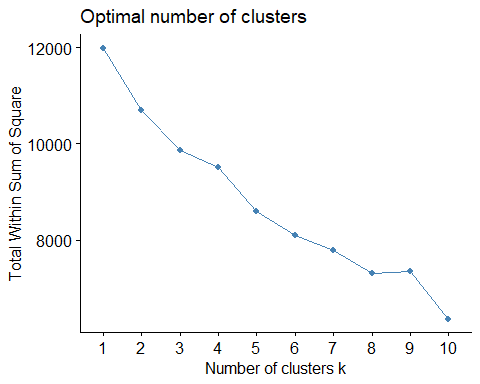
BrandLoyal<-data[,23:30]  
Max\_Brand<-apply(BrandLoyal, 1, max)  
Max\_Brand<-as.data.frame(Max\_Brand)  
BrandComp<-cbind(Max\_Brand,data[,31])  
MaxLoyal<-as.data.frame(ifelse(BrandComp[,1]>=BrandComp[,2],1,0))  
MaxLoyalComp<-cbind(BrandComp,MaxLoyal)  
BathSoapComp<-cbind(data,MaxLoyal)  
data 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

#Data preparation

data$SEC<-as.factor(data$SEC)  
data$SEX<-as.factor(data$SEX)  
data$AGE<-as.factor(data$AGE)  
data$EDU<-as.factor(data$EDU)  
data$HS<-as.factor(data$HS)  
data$CHILD<-as.factor(data$CHILD)  
data$CS<-as.factor(data$CS)  
data$Affluence.Index<-as.factor((data$Affluence.Index))

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

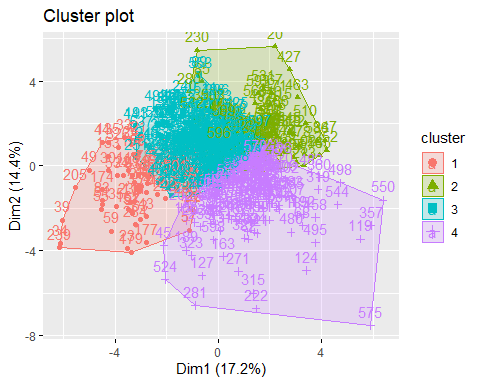
#variables needed to analyse purchase behavior  
data\_pb <- data[,c(12:31)]  
data\_pb <- as.data.frame(scale(data\_pb))  
fviz\_nbclust(data\_pb, kmeans, method = "wss")



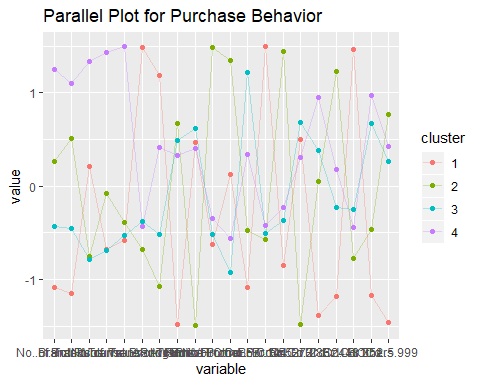
#k-means  
set.seed(123)  
k1 <- kmeans(data\_pb, centers =4 , nstart = 200) # k = 3  
  
# Size and Center for the Clusters  
k1$size

## [1] 65 95 284 156

k1\_clusters <- data.frame(k1$centers)  
cluster <- matrix(c("1","2","3","4"),nrow = 4)  
k1.visual <- cbind(cluster,k1\_clusters)  
  
#Cluster visiualization   
fviz\_cluster(k1, data = data\_pb)



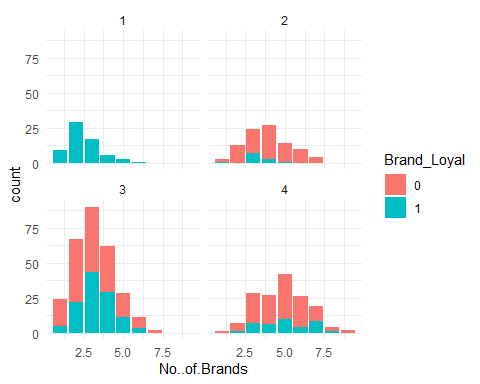
#Parallel Plot  
ggparcoord(k1.visual,  
 columns = 2:21, groupColumn = 1,   
 showPoints = TRUE,   
 title = "Parallel Plot for Purchase Behavior",  
 alphaLines = 0.3  
 )



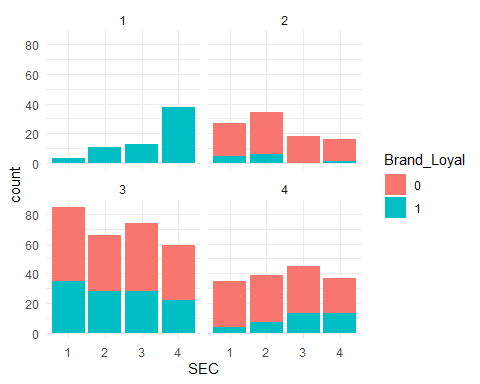
Analysis: Cluster1:Customers who are brand loyal to certain particular brand. Cluster2:Customer who are not particular brand loyal people. Cluster3 and Cluster4 :middling all factors with differences in no.of brands,brand runs.

After doing kmeans clustering for purhase behavior alone.Now add demographics and loyalty to purchase behavior

data\_pb1 <- data[,c(1:31,47)]  
data\_pb1[,33]<-data.frame(k1$cluster)  
  
library(ggplot2)  
data\_pb1$Brand\_Loyal<-as.factor(data\_pb1$Brand\_Loyal)  
ggplot(data\_pb1) +  
 aes(x =No..of.Brands,fill = Brand\_Loyal ) +  
 geom\_bar() +  
 scale\_fill\_hue() +  
 theme\_minimal() +  
 facet\_wrap(vars(k1.cluster))

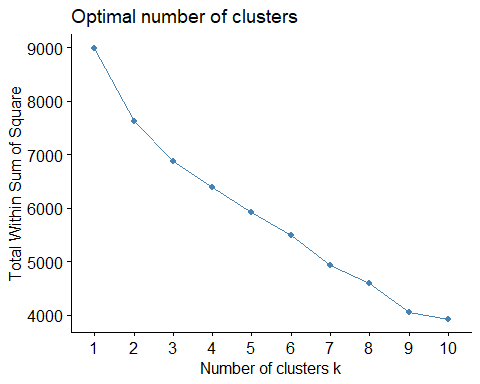


#lets consider SEC,Brand\_Loyal and in four clusters.  
library(ggplot2)  
data\_pb1$Brand\_Loyal<-as.factor(data\_pb1$Brand\_Loyal)  
ggplot(data\_pb1) +  
 aes(x = SEC,fill = Brand\_Loyal ) +  
 geom\_bar() +  
 scale\_fill\_hue() +  
 theme\_minimal() +  
 facet\_wrap(vars(k1.cluster))



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

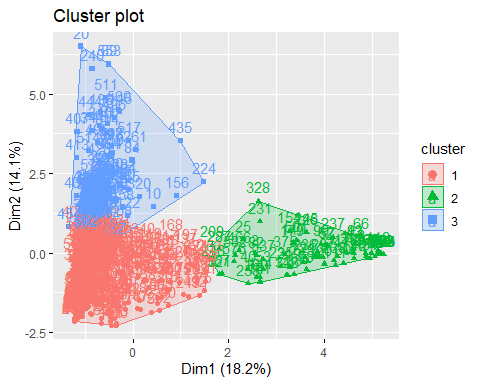
data\_bp <- BathSoapComp[,c(32:46)]  
data\_bp <- data.frame(scale(data\_bp))  
  
fviz\_nbclust(data\_bp, kmeans, method = "wss")



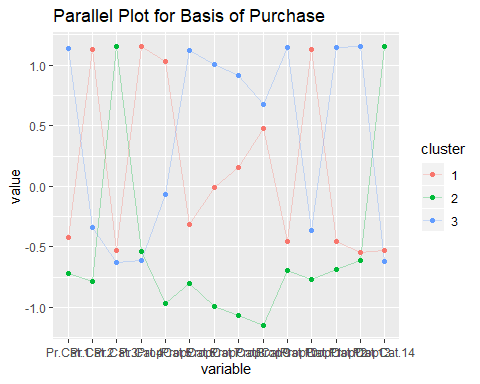
set.seed(123)  
k2 <- kmeans(data\_bp, centers =3 , nstart = 500)  
  
  
# Size and Center for the Clusters  
k2$size

## [1] 376 79 145

k2\_clusters <- data.frame(k2$centers)  
cluster <- matrix(c("1","2","3"),nrow = 3)  
k2.visual <- cbind(cluster,k2\_clusters)  
  
#cluster visualization  
fviz\_cluster(k2, data = data\_bp)



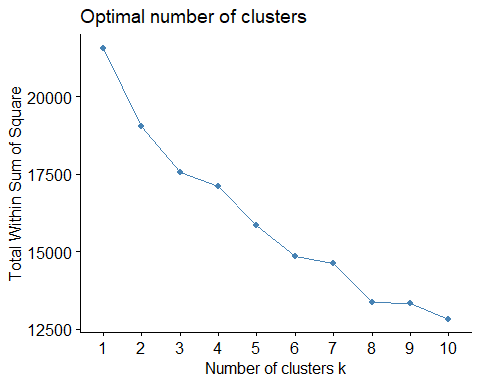
#Parallel Plot  
ggparcoord(k2.visual,  
 columns = 2:15, groupColumn = 1,   
 showPoints = TRUE,   
 title = "Parallel Plot for Basis of Purchase",  
 alphaLines = 0.3  
 )

 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

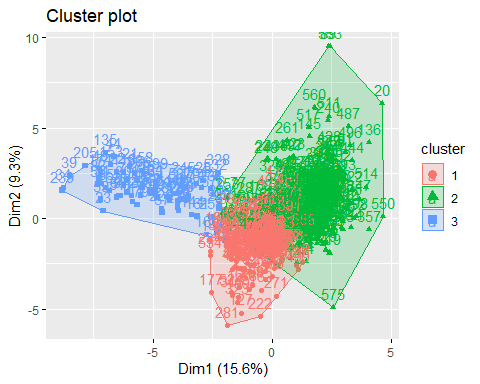
data\_both <- BathSoapComp[,c(12:47)]  
data\_both <- data.frame(scale(data\_both))  
  
#k-means  
set.seed(123)  
fviz\_nbclust(data\_both, kmeans, method = "wss")



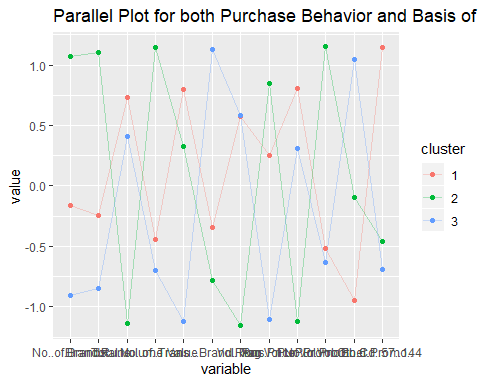
set.seed(123)  
k3 <- kmeans(data\_both, centers =3 , nstart = 500)   
  
# Size and Center for the Clusters  
k3$size

## [1] 180 344 76

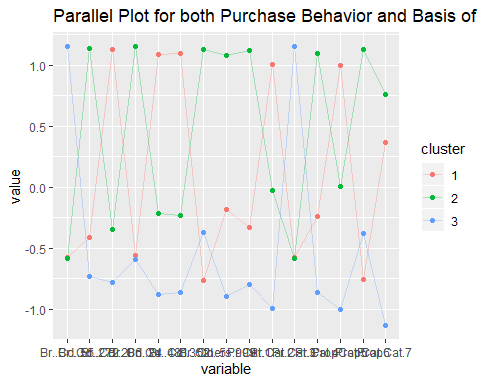
k3\_clusters <- data.frame(k3$centers)  
cluster <- matrix(c("1","2","3"),nrow = 3)  
k3.visual <- cbind(cluster,k3\_clusters)  
  
#cluster visualization  
fviz\_cluster(k3, data = data\_both)



#Parallel Plot  
ggparcoord(k3.visual,  
 columns = 2:13, groupColumn = 1,   
 showPoints = TRUE,   
 title = "Parallel Plot for both Purchase Behavior and Basis of PUrchase",  
 alphaLines = 0.3  
 )

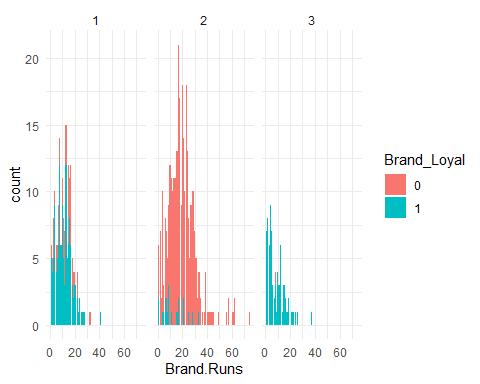


ggparcoord(k3.visual,  
 columns = 14:28, groupColumn = 1,   
 showPoints = TRUE,   
 title = "Parallel Plot for both Purchase Behavior and Basis of PUrchase",  
 alphaLines = 0.3  
 )

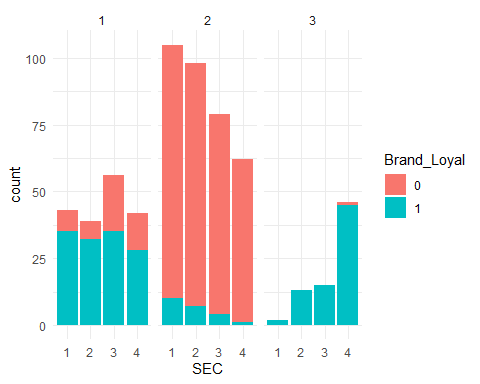


Now add demographics and analyse the data

data\_all <- data[,c(1:47)]  
data\_all[,48]<-data.frame(k3$cluster)  
  
library(ggplot2)  
data\_all$Brand\_Loyal<-as.factor(data\_all$Brand\_Loyal)  
ggplot(data\_all) +  
 aes(x = Brand.Runs,fill = Brand\_Loyal ) +  
 geom\_bar() +  
 scale\_fill\_hue() +  
 theme\_minimal() +  
 facet\_wrap(vars(k3.cluster))



#lets consider SEC,Brand\_Loyal and in three clusters.  
library(ggplot2)  
data\_all$Brand\_Loyal<-as.factor(data\_all$Brand\_Loyal)  
ggplot(data\_all) +  
 aes(x = SEC,fill = Brand\_Loyal ) +  
 geom\_bar() +  
 scale\_fill\_hue() +  
 theme\_minimal() +  
 facet\_wrap(vars(k3.cluster))



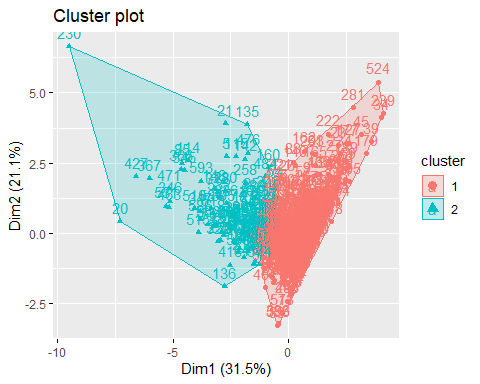
whenk=3 Cluster3:shows high brand loyal customers and their characteristics Cluster2:Low brand loyal customers and their characteristics.

#3.Develop a model that classifies the data into these segments. Since this information would most likely be used in targeting direct-mail promotions, it would be useful to select a market segment that would be defined as a success in the classification model. In this model I’m considering some variables like value,trans brand runs,vol.trans,avg price and all three promotions categories.

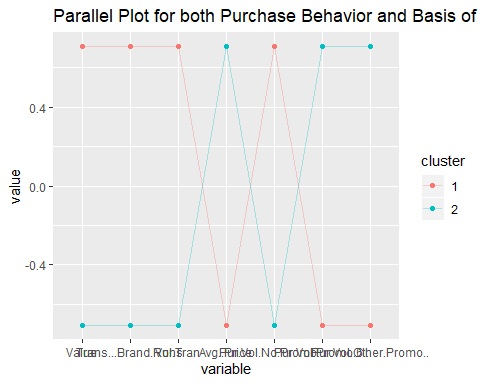
data\_classification <- data[,c(16:22)]  
data\_classnorm <- as.data.frame(scale(data\_classification))  
  
  
set.seed(123)  
k4 <- kmeans(data\_classnorm, centers =2 , nstart = 500)   
  
# Size and Center for the Clusters  
k4$size

## [1] 491 109

k4\_clusters <- data.frame(k4$centers)  
cluster <- matrix(c("1","2"),nrow = 2)  
k4.visual <- cbind(cluster,k4\_clusters)  
  
#cluster visualization  
fviz\_cluster(k4, data = data\_classnorm)



#Parallel Plot  
ggparcoord(k4.visual,  
 columns = 2:8, groupColumn = 1,   
 showPoints = TRUE,   
 title = "Parallel Plot for both Purchase Behavior and Basis of PUrchase",  
 alphaLines = 0.3  
 )



Cluster2:Shows customers behavior when there is some kind of promotions. Cluster1:Shows behavior when there is no promotions. In this case cluster/group two is success because they respond to the promotions.