Final Exam

SABIHA MHATARNAIK

5/7/2021

####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 budgets more 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.

QUESTIONS:

1. Use k-means clustering to identify clusters of households based on:
2. The variables that describe purchase behavior (including brand loyalty)
3. The variables that describe the basis for purchase
4. The variables that describe both purchase behavior and basis of purchase Note 1: How should k be chosen? Think about how the clusters would be used. It is likely that the marketing efforts would support two to five different promotional approaches. Note 2: How should the percentages of total purchases comprised by various brands be treated? Isn’t a customer who buys all brand A just as loyal as a customer who buys all brand B? What will be the effect on any distance measure of using the brand share variables as is? Consider using a single derived variable.
5. Select what you think is the best segmentation and comment on the characteristics (demographic, brand loyalty, and basis for purchase) of these clusters. (This information would be used to guide the development of advertising and promotional campaigns.)
6. 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.

####Libraries Used

library(dplyr) #for conversion of %columns to numeric variables for calculations  
library(factoextra) #for clustering algorithms & visualization  
library(ggplot2) #visualization  
library(GGally) #graph  
library(mcclust) #modeling clusters

#Importing dataset in R

data <- read.csv("BathSoap.csv")  
sum(is.na(data)) #To check missing values

## [1] 0

#Data Transformation

#Duplicating the dataset and then using it further for transformation and analysis   
my\_data <- data  
  
#Converting the % data to decimal form  
my\_data[ ,20:46] <- apply(my\_data[ ,20:46],2,function(x) {as.numeric(sub("%", "",x,fixed=TRUE))/100})  
  
  
#Converting variables to factors  
my\_data[,c(1:10,12)] <- lapply(my\_data[,c(1:10,12)], factor)  
  
#Giving row number as the name of " Member Id " and then eliminating the variable name from the dataset to avoid redundency   
rownames(my\_data) <- my\_data$Member.id  
my\_data$Member.id = NULL

#1. Kmeans Clustering to identify clusters of households based on:

##A. Purchase Behavior (including brand loyalty) ###I. Variables that describes purchase behavior: #Brand Runs, Total Volume, No. of Trans, Value, Trans/ Brand Runs, Vol/Trans, Avg. Price,Pur Vol No Promo

###II. Brand Loyalty: #Br. Cd. 57, 144,Br. Cd. 55,Br. Cd. 272,Br. Cd. 286,Br. Cd. 24,Br. Cd. 481,Br. Cd. 352,Br. Cd. 5,Others 999

PurchaseBehavior\_Data <- my\_data[,c("Brand.Runs","Total.Volume", "No..of..Trans","Value","Trans...Brand.Runs","Vol.Tran","Avg..Price", "Pur.Vol.No.Promo....","Br..Cd..57..144","Br..Cd..55","Br..Cd..272","Br..Cd..286", "Br..Cd..24","Br..Cd..481","Br..Cd..352","Br..Cd..5", "Others.999")]  
#Normalizing the dataset   
PurchaseBehavior\_Data <- as.data.frame(scale(PurchaseBehavior\_Data))  
sum(is.na(PurchaseBehavior\_Data))

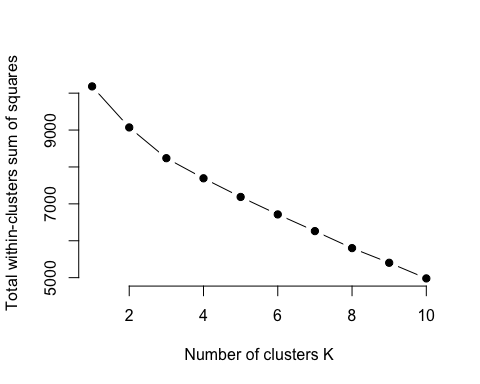
## [1] 0

##Determining K value

#Elbow Method for finding the optimal number of clusters  
set.seed(123)  
k.max <- 10  
DATA1 <- PurchaseBehavior\_Data  
wss <- sapply(1:k.max,   
 function(k){kmeans(DATA1, k, nstart=25,iter.max = 15 )$tot.withinss})  
wss

## [1] 10183.000 9071.059 8237.850 7692.864 7187.065 6713.718 6262.083  
## [8] 5801.142 5403.302 4979.912

plot(1:k.max, wss,  
 type="b", pch = 19, frame = FALSE,   
 xlab="Number of clusters K",  
 ylab="Total within-clusters sum of squares")



As a result, the between ss/total ss ratio for k=3 continues to shift slowly and stay stable relative to other k’s. As a result, k=3 should be a reasonable number of clusters for this results.

set.seed(123)  
#K-means   
KMeans1 <- kmeans(PurchaseBehavior\_Data, centers = 3, nstart = 25)  
KMeans1$centers #centers of clusters

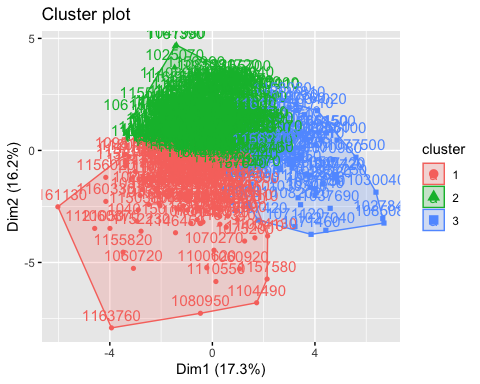
## Brand.Runs Total.Volume No..of..Trans Value Trans...Brand.Runs  
## 1 0.8670970 1.2164267 1.0299300 1.2328699 -0.1328217  
## 2 -0.1308592 -0.4618909 -0.2774206 -0.3480481 -0.2008284  
## 3 -0.8867615 0.1511614 -0.4417670 -0.4487871 1.2430693  
## Vol.Tran Avg..Price Pur.Vol.No.Promo.... Br..Cd..57..144 Br..Cd..55  
## 1 0.4697760 -0.06836852 -0.03038319 -0.1440356 -0.2017577  
## 2 -0.2846836 0.26897735 -0.04716137 0.1637939 -0.3493619  
## 3 0.5912798 -1.22673730 0.29047776 -0.5648253 2.1103184  
## Br..Cd..272 Br..Cd..286 Br..Cd..24 Br..Cd..481 Br..Cd..352 Br..Cd..5  
## 1 -0.07337963 0.004949496 -0.03560124 0.08852116 -0.0525913 0.03049299  
## 2 0.08799788 -0.003696919 0.05648124 0.01920970 -0.0927790 0.03413838  
## 3 -0.31057365 0.009726359 -0.21979726 -0.25379669 0.5586715 -0.22538902  
## Others.999  
## 1 0.3099117  
## 2 0.1583766  
## 3 -1.3453413

KMeans1$size #size of each clusters

## [1] 137 386 77

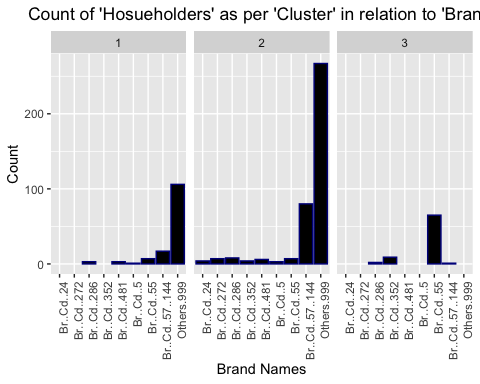
#Store the cluster assigned to the dataset  
PurchaseBehavior\_Data <- as.data.frame(cbind(KMeans1$cluster, PurchaseBehavior\_Data))  
colnames(PurchaseBehavior\_Data)[1] <- "cluster"  
#Converting the cluster to a factor   
PurchaseBehavior\_Data$cluster <- as.factor(PurchaseBehavior\_Data$cluster)

# Plot  
fviz\_cluster(KMeans1, data = PurchaseBehavior\_Data[ , -c(1)])



#Creating a new column "High Brand Value Name" that finds the highest value of all the rows and prints the name of the column it refers to in the new variable.  
DataSet <- data.frame(my\_data[,c(22:30)])  
PurchaseBehavior\_Data$High\_Brand\_Value\_Name <- colnames(DataSet)[apply(DataSet,1,which.max)]  
# Ploting graph by Brand Name  
ggplot(data = PurchaseBehavior\_Data) +  
 geom\_bar(mapping = aes(PurchaseBehavior\_Data$High\_Brand\_Value\_Name),   
 col = "blue4",  
 fill = "black") +  
 facet\_wrap(vars(PurchaseBehavior\_Data$cluster)) +  
 labs(title = "Count of 'Hosueholders' as per 'Cluster' in relation to 'Brand'") +  
 labs(x = "Brand Names", y = "Count") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1),plot.title = element\_text(hjust = 0.5))

## Warning: Use of `PurchaseBehavior\_Data$High\_Brand\_Value\_Name` is discouraged.  
## Use `High\_Brand\_Value\_Name` instead.



#B. Basis of Purchase

#\*Variables used: Pur Vol No Promo-%, Vol Promo 6%, Pur Vol Other Promo %, Price Cat 1 to 4, Proposition Cat 5 to 15

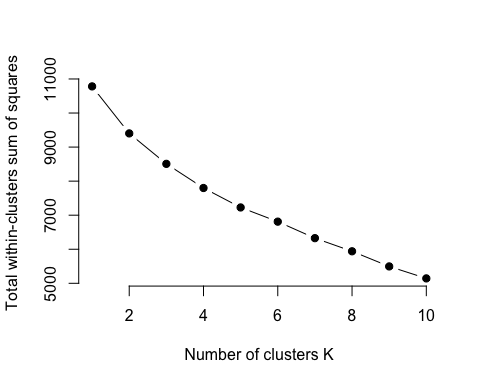
Purchase\_Basis\_data <- my\_data[,c("Pur.Vol.No.Promo....","Pur.Vol.Promo.6..", "Pur.Vol.Other.Promo..","Pr.Cat.1" ,"Pr.Cat.2","Pr.Cat.3","Pr.Cat.4", "PropCat.5","PropCat.6" ,"PropCat.7","PropCat.8","PropCat.9","PropCat.10","PropCat.11","PropCat.12","PropCat.13","PropCat.14","PropCat.15"   
)]  
#Data Normalization  
Purchase\_Basis\_data <- as.data.frame(scale(Purchase\_Basis\_data))

#####Determining the value of K

#Elbow Method for finding the optimal number of clusters  
set.seed(123)  
k.max <- 10  
DATA2 <- Purchase\_Basis\_data  
wss <- sapply(1:k.max,   
 function(k){kmeans(DATA2, k, nstart=25,iter.max = 15 )$tot.withinss})  
wss

## [1] 10782.000 9401.887 8507.874 7798.887 7227.351 6810.742 6325.761  
## [8] 5939.965 5495.652 5143.901

plot(1:k.max, wss,  
 type="b", pch = 19, frame = FALSE,   
 xlab="Number of clusters K",  
 ylab="Total within-clusters sum of squares")



#As a result, the between ss/total ss ratio for k=3 continues to shift slowly and stay stable relative to other k’s. As a result, k=3 should be a reasonable number of clusters for this results.

set.seed(123)  
#k-means   
KMeans2 <- kmeans(Purchase\_Basis\_data, centers = 3, nstart = 25)  
#size of each cluster  
KMeans2$centers

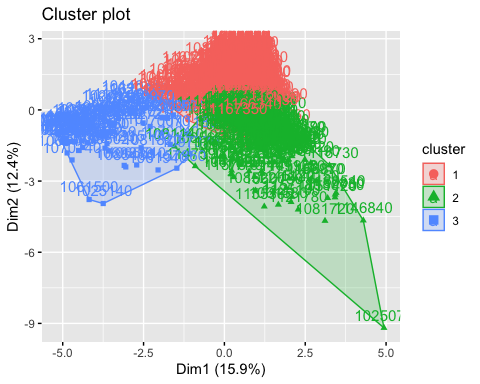
## Pur.Vol.No.Promo.... Pur.Vol.Promo.6.. Pur.Vol.Other.Promo.. Pr.Cat.1  
## 1 0.2860651 -0.2360563 -0.1703973 -0.4651435  
## 2 -0.5626809 0.5576736 0.2131738 1.1091649  
## 3 0.1856666 -0.3842112 0.1912587 -0.7825205  
## Pr.Cat.2 Pr.Cat.3 Pr.Cat.4 PropCat.5 PropCat.6 PropCat.7  
## 1 0.5449425 -0.2889248 0.1995556 0.4650497 -0.02823256 -0.03967626  
## 2 -0.4708722 -0.4653448 -0.2106562 -0.3516447 0.11719213 0.24917427  
## 3 -1.1334328 2.3701003 -0.3204763 -1.0914607 -0.17089192 -0.44919415  
## PropCat.8 PropCat.9 PropCat.10 PropCat.11 PropCat.12 PropCat.13  
## 1 -0.1912417 -0.03863186 -0.1612525 0.06575023 -0.09960688 -0.2035148  
## 2 0.5131099 0.13143273 0.3787795 -0.01931633 0.23567662 0.4408922  
## 3 -0.4629703 -0.16226455 -0.2570818 -0.22953559 -0.16301187 -0.2325107  
## PropCat.14 PropCat.15  
## 1 -0.2913920 0.02639850  
## 2 -0.4620933 0.04781956  
## 3 2.3724613 -0.22967026

KMeans2$size

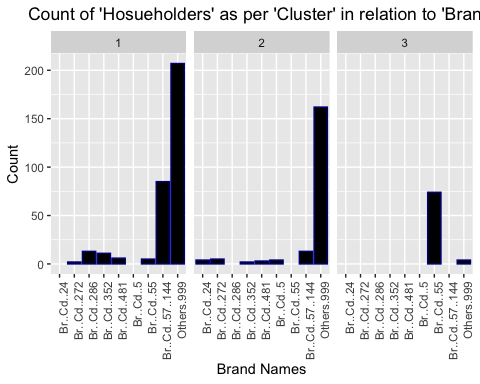
## [1] 329 193 78

#Store the cluster assigned to the dataset  
Purchase\_Basis\_data <- as.data.frame(cbind(KMeans2$cluster, Purchase\_Basis\_data))  
colnames(Purchase\_Basis\_data)[1] <- "cluster"  
#Converting the cluster to a factor   
Purchase\_Basis\_data$cluster <- as.factor(Purchase\_Basis\_data$cluster)

# Plots of the variables  
fviz\_cluster(KMeans2, data = Purchase\_Basis\_data[ , -c(1)])



# Plot by Brand Name  
ggplot(data = Purchase\_Basis\_data) +  
 geom\_bar(mapping = aes(PurchaseBehavior\_Data$High\_Brand\_Value\_Name),   
 col = "blue4",  
 fill = "black") +  
 facet\_wrap(vars(Purchase\_Basis\_data$cluster)) +  
 labs(title = "Count of 'Hosueholders' as per 'Cluster' in relation to 'Brand'") +  
 labs(x = "Brand Names", y = "Count") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1),plot.title = element\_text(hjust = 0.5))



####c. Purchase Behaviour & Basis of Purchase

#I. Variables that describes “Purchase behavior”: # Brand Runs, Total Volume, No. of Trans, Value, Trans/ Brand Runs, Vol/Trans, Avg. Price,Pur #Vol,Br.Cd.(57,144),55,272,286,24,481,352,5,and 999(others)

#II. Variables that describe “Basis of”Purchase": #Pur Vol No Promo-%, Vol Promo 6%, Pur Vol Other Promo %, Price Cat 1 to 4, Proposition #Cat 5 to 15

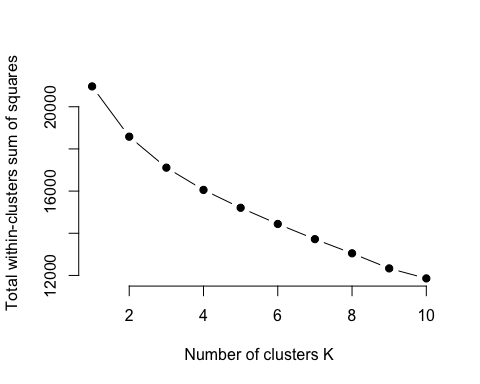
PurchaseBehaviour\_n\_PurchaseBasis<- my\_data[,c("Brand.Runs","Total.Volume", "No..of..Trans","Value","Trans...Brand.Runs","Vol.Tran","Avg..Price", "Pur.Vol.No.Promo....", "Br..Cd..57..144","Br..Cd..55","Br..Cd..272","Br..Cd..286", "Br..Cd..24","Br..Cd..481", "Br..Cd..352","Br..Cd..5","Others.999","Pur.Vol.No.Promo....","Pur.Vol.Promo.6..", "Pur.Vol.Other.Promo..","Pr.Cat.1" ,"Pr.Cat.2","Pr.Cat.3","Pr.Cat.4", "PropCat.5","PropCat.6" ,"PropCat.7","PropCat.8","PropCat.9","PropCat.10","PropCat.11","PropCat.12","PropCat.13","PropCat.14","PropCat.15")]  
  
PurchaseBehaviour\_n\_PurchaseBasis <- as.data.frame(scale(PurchaseBehaviour\_n\_PurchaseBasis))

#####Determining the value of K

#Elbow Method for finding the optimal number of clusters  
set.seed(123)  
k.max <- 10  
DATA3 <- PurchaseBehaviour\_n\_PurchaseBasis  
wss <- sapply(1:k.max,   
 function(k){kmeans(DATA3, k, nstart=25,iter.max = 15 )$tot.withinss})  
wss

## [1] 20965.00 18580.65 17111.77 16055.75 15208.13 14441.16 13721.32 13049.80  
## [9] 12334.14 11859.59

plot(1:k.max, wss,  
 type="b", pch = 19, frame = FALSE,   
 xlab="Number of clusters K",  
 ylab="Total within-clusters sum of squares")



#As a result, the between ss/total ss ratio for k=3 continues to shift slowly and stay stable relative to other k’s. As a result, k=3 should be a reasonable number of clusters for this results.

set.seed(123)  
#K-means   
KMeans3 <- kmeans(PurchaseBehaviour\_n\_PurchaseBasis, centers = 3, nstart = 25)  
#size of each cluster  
KMeans3$centers

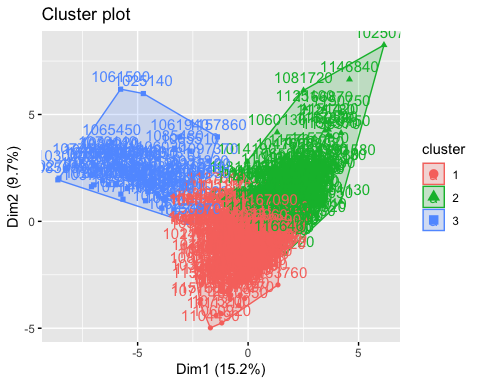
## Brand.Runs Total.Volume No..of..Trans Value Trans...Brand.Runs  
## 1 -0.1134779 0.1857282 -0.05192827 0.110054759 -0.05754053  
## 2 0.3968272 -0.3164902 0.18838478 0.008486536 -0.24686715  
## 3 -0.7105098 0.1729055 -0.34533247 -0.480321544 0.97186725  
## Vol.Tran Avg..Price Pur.Vol.No.Promo.... Br..Cd..57..144 Br..Cd..55  
## 1 0.2225872 -0.2332921 0.3570261 0.4340049 -0.2767672  
## 2 -0.4812561 0.7615287 -0.5651620 -0.4102464 -0.4346155  
## 3 0.5103331 -1.2993101 0.2038602 -0.5750176 2.4365700  
## Br..Cd..272 Br..Cd..286 Br..Cd..24 Br..Cd..481 Br..Cd..352 Br..Cd..5  
## 1 -0.1930194 0.1608045 -0.1897110 0.07376960 0.1424151 -0.1283041  
## 2 0.3706977 -0.1414022 0.3329282 -0.03145328 -0.1071466 0.2289109  
## 3 -0.3039129 -0.2445636 -0.2053058 -0.21153749 -0.2703619 -0.1499909  
## Others.999 Pur.Vol.No.Promo.....1 Pur.Vol.Promo.6.. Pur.Vol.Other.Promo..  
## 1 -0.1056985 0.3570261 -0.2968982 -0.2091554  
## 2 0.5578813 -0.5651620 0.5523854 0.2241380  
## 3 -1.2214882 0.2038602 -0.4145129 0.1985297  
## Pr.Cat.1 Pr.Cat.2 Pr.Cat.3 Pr.Cat.4 PropCat.5 PropCat.6  
## 1 -0.4306719 0.5423199 -0.2740346 0.13187586 0.5089396 -0.08333716  
## 2 0.8660095 -0.3583404 -0.4399355 -0.07204477 -0.3300912 0.18745326  
## 3 -0.7937363 -1.1772298 2.4410866 -0.33113736 -1.1231821 -0.21268304  
## PropCat.7 PropCat.8 PropCat.9 PropCat.10 PropCat.11 PropCat.12  
## 1 -0.02381718 -0.2730616 -0.0503891 -0.1219042 0.07621598 -0.1117521  
## 2 0.18547671 0.5394528 0.1194240 0.2558788 -0.03005563 0.2124005  
## 3 -0.45292998 -0.4746321 -0.1466787 -0.2566303 -0.22580882 -0.1693510  
## PropCat.13 PropCat.14 PropCat.15  
## 1 -0.2117651 -0.2737444 -0.0002652065  
## 2 0.3722178 -0.4415351 0.0817246866  
## 3 -0.2309160 2.4446419 -0.2418686199

KMeans3$size

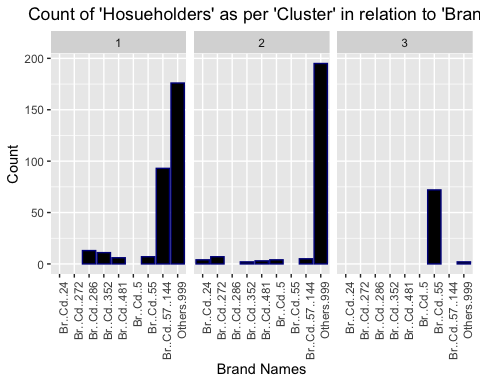
## [1] 306 220 74

#Store the cluster assigned to the dataset  
PurchaseBehaviour\_n\_PurchaseBasis <- as.data.frame(cbind(KMeans3$cluster, PurchaseBehaviour\_n\_PurchaseBasis))  
colnames(PurchaseBehaviour\_n\_PurchaseBasis)[1] <- "cluster"  
#Converting the cluster to a factor   
PurchaseBehaviour\_n\_PurchaseBasis$cluster <- as.factor(PurchaseBehaviour\_n\_PurchaseBasis$cluster)

# Ploting variables  
fviz\_cluster(KMeans3, data = PurchaseBehaviour\_n\_PurchaseBasis[ , -c(1)])



# Ploting graph by Brand Name  
ggplot(data = PurchaseBehaviour\_n\_PurchaseBasis) +  
 geom\_bar(mapping = aes(PurchaseBehavior\_Data$High\_Brand\_Value\_Name),   
 col = "blue4",  
 fill = "black") +  
 facet\_wrap(vars(PurchaseBehaviour\_n\_PurchaseBasis$cluster)) +  
 labs(title = "Count of 'Hosueholders' as per 'Cluster' in relation to 'Brand'") +  
 labs(x = "Brand Names", y = "Count") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1),plot.title = element\_text(hjust = 0.5))

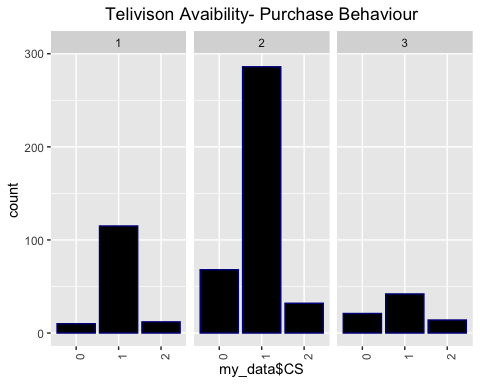


###2. Market Segmentation

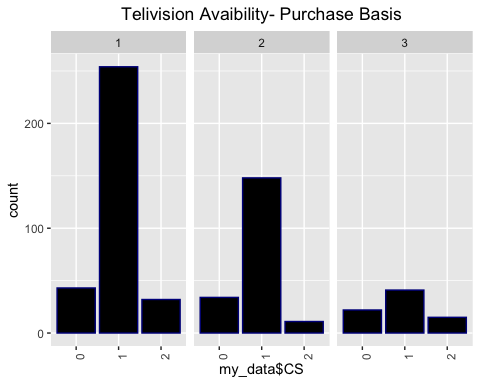
#\*As per the analysis of characteristics (demographic, brand loyalty and basis of purchase) the best segmentation is the third one where we consider both the aspects i.e purchase behaviour (brand loyalty) and basis of purchase.

#\*To get the result/s of the other variable/s we need to replace the CS variable name to the respective variable we want to do visualization on.

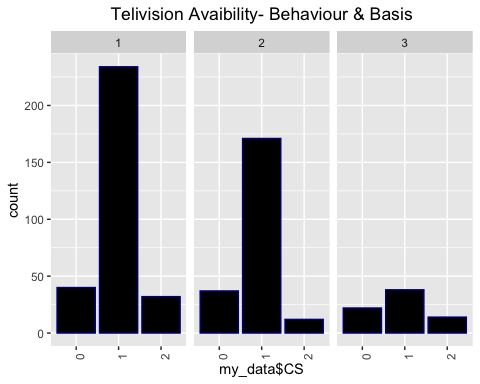
#Ploting graph by "CS":- Telivision Avaibiity  
# Purchase Behaviour Data  
ggplot(data = PurchaseBehavior\_Data) +  
 geom\_bar(mapping = aes(my\_data$CS),   
 col = "blue4",  
 fill = "black") +  
 facet\_wrap(vars(PurchaseBehavior\_Data$cluster)) +  
 labs(title = "Telivison Avaibility- Purchase Behaviour") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1),plot.title = element\_text(hjust = 0.5))



# Purchase Basis Data  
ggplot(data = Purchase\_Basis\_data) +  
 geom\_bar(mapping = aes(my\_data$CS),   
 col = "blue4",  
 fill = "black") +  
 facet\_wrap(vars(Purchase\_Basis\_data$cluster)) +  
 labs(title = "Telivision Avaibility- Purchase Basis") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1),plot.title = element\_text(hjust = 0.5))



#Analysis:  
#Ploting graph by "CS":- Telivision Avaibiity  
#Behaviour & Basis Data  
ggplot(data = PurchaseBehaviour\_n\_PurchaseBasis) +  
 geom\_bar(mapping = aes(my\_data$CS),   
 col = "blue4",  
 fill = "black") +  
 facet\_wrap(vars(PurchaseBehaviour\_n\_PurchaseBasis$cluster)) +  
 labs(title = "Telivision Avaibility- Behaviour & Basis") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1), plot.title = element\_text(hjust = 0.5))



##Insights on Demographics:  
Demographics\_data = cbind(data, Cluster =PurchaseBehaviour\_n\_PurchaseBasis$cluster)  
aggregate\_data = aggregate(cbind(SEC, FEH, MT, SEX, AGE, EDU, HS, CHILD, CS,  
Affluence.Index) ~ Cluster, data = Demographics\_data, mean, na.rm = FALSE)

Analysis: 1. Based on the chart’s observations/visuals, we can deduce that “Others 999” is the most common brand name in Clusters 1 and 2.

2.More extensive data collection is needed to determine if the “Other 999” brand has any secret patterns. Whether it has a larger number of customers within the same brand than the competition.

3.Clusters 1 and 2 have the largest percentages of households with televisions. This may be one of the reasons that more people in the same cluster have a lot of “Other 999” brand items. They must have seen a commercial on television that enticed them to try a different product than normal.

4.The ‘CS’ variable, which refers to Telivision availability, has zero in it. As a result, many homeowners have not responded to the survey. As a result, there’s a chance that I’ll miss out on crucial details. As a result, detailed questions relating to variables should be made obligatory to answer.

5.Cluster 3 has the highest level of socioeconomic status, and it has a high level of brand loyalty to “Br..Cd..55.” There’s a chance that the food is pricey and serves as a status statement in society.

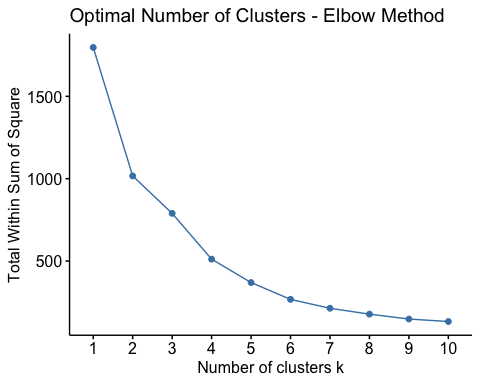
6.Cluster 3 has the lowest Affluence Index, perhaps because the householders have children and their earnings will be spent on tuition, health, insurance, schooling, and other expenses, implying that they do not have any income or savings.

###3. Classification Model

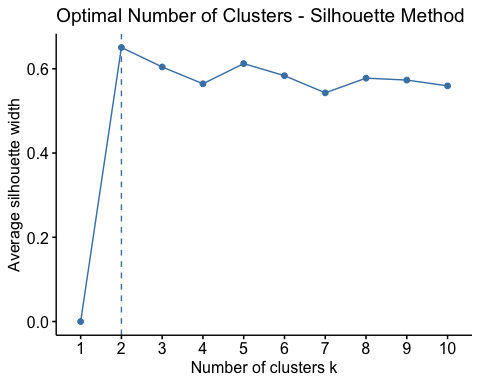
#\*I. Method: k means clustering for classification model

#\*II. Variables used:“Pur.Vol.No.Promo….”,“Pur.Vol.Promo.6..”,“Pur.Vol.Other.Promo..”

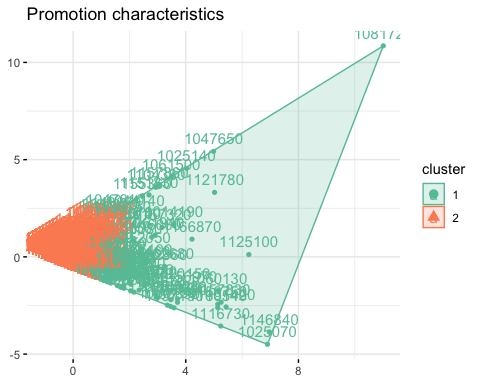
#Building model based on 3 variables.  
Classification\_Model<-my\_data[,c(19:21)]  
Classification\_Model\_Scaled<-as.data.frame(scale(Classification\_Model))  
set.seed(123)  
fviz\_nbclust(Classification\_Model\_Scaled, kmeans, method = "wss") +  
 labs(title = "Optimal Number of Clusters - Elbow Method")



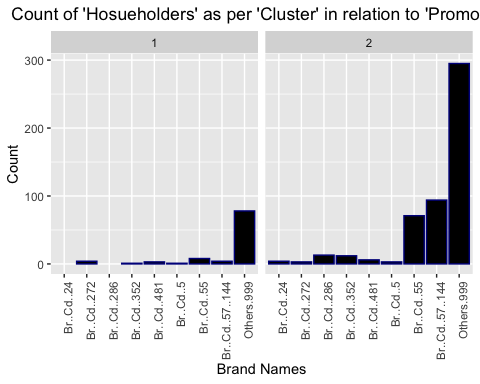
fviz\_nbclust(Classification\_Model\_Scaled, kmeans, method = "silhouette") +  
 labs(title = "Optimal Number of Clusters - Silhouette Method")



#K-means for classification model  
set.seed(123)  
km\_class <- kmeans(Classification\_Model\_Scaled, centers =2 , nstart = 100) #k=2  
#plotting k-means model  
fviz\_cluster(km\_class, data = Classification\_Model\_Scaled, main="Promotion characteristics",   
 xlab = FALSE,ylab = FALSE,palette = "Set2", ggtheme = theme\_minimal())



#Store the cluster assigned to the dataset  
Classification\_Model\_Scaled <- as.data.frame(cbind(km\_class$cluster, Classification\_Model\_Scaled))  
colnames(Classification\_Model\_Scaled)[1] <- "cluster"  
#Converting the cluster to a factor   
Classification\_Model\_Scaled$cluster <- as.factor(Classification\_Model\_Scaled$cluster)  
# Plot on basis of Promotion  
ggplot(data = Classification\_Model\_Scaled) +  
 geom\_bar(mapping = aes(PurchaseBehavior\_Data$High\_Brand\_Value\_Name),   
 col = "blue4",  
 fill = "black") +  
 facet\_wrap(vars(Classification\_Model\_Scaled$cluster)) +  
 labs(title = "Count of 'Hosueholders' as per 'Cluster' in relation to 'Promotion'") +  
 labs(x = "Brand Names", y = "Count") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1),plot.title = element\_text(hjust = 0.5))

 Conclusion:

–>Brand Loyal residents can be found in Cluster 2. If we have discounts or deals, cluster 1 householders should be our target demographic because they are price sensitive and could be future customers.

–>As a result, instead of spending time on cluster 2, it is recommended to aim direct mail deals to consumers from other clusters.

–>Other clusters are at differences with each other. They react well to promotions, and if given the opportunity, some potential customers can switch to brandloyal.