R Notebook

Code

Hide

```
library(tidyverse)
library(factoextra)
library(ISLR)
library(dplyr)     # for data wrangling
library(ggplot2)# for plotting
```

Hide

```
Mdata <- read.csv("BathSoap.csv")
Mdata1 <-scale(Mdata[,-c(1:11)]) ##scale the dataset
head(Mdata1)
```

```
     No..of.Brands Brand.Runs Total.Volume No..of..Trans
[1,]    -0.4030277  0.1200727   -0.5005898    -0.4104681
[2,]     0.8630280  0.8895639    0.2651391     0.5076339
[3,]     0.8630280  2.0438006    1.4394712     1.8274054
[4,]    -1.0360556 -1.1303505   -1.3403176    -1.5580955
[5,]    -0.4030277 -0.9379777   -0.4651989    -1.0416632
[6,]    -0.4030277  0.9857502    0.8056536     0.5650152
          Value Trans...Brand.Runs   Vol.Tran  Avg..Price
[1,] -0.5881031         -0.4636969 -0.3242918 -0.43944366
[2,]  0.3896410         -0.3907514 -0.2639930  0.05217678
[3,]  0.6936645         -0.3523590 -0.1944886 -0.90701745
[4,] -1.3852447         -0.6211057 -0.1610026 -1.13145287
[5,] -0.8451360         -0.1719147  0.8980852 -1.25970168
[6,]  0.4168163         -0.3984298  0.1135176 -0.65586353
     Pur.Vol.No.Promo.... Pur.Vol.Promo.6..
[1,]            0.7269843        -0.5756626
[2,]           -0.1927198         0.4986688
[3,]            0.2253275        -0.3607963
[4,]            0.7269843        -0.5756626
[5,]           -2.5337850         0.9284014
[6,]            0.7269843        -0.5756626
     Pur.Vol.Other.Promo.. Br..Cd..57..144   Br..Cd..55
[1,]           -0.46438366       0.8278127  0.002180775
[2,]           -0.18644830      -0.6938241 -0.190240508
[3,]            0.09148706      -0.6515564  1.618519544
[4,]           -0.46438366       0.9123481  1.810940826
[5,]            2.87084066      -0.5670211  0.040665031
[6,]           -0.46438366      -0.4402180 -0.228724764
     Br..Cd..272 Br..Cd..286 Br..Cd..24 Br..Cd..481
[1,]  -0.3648639 -0.30081170 -0.2426136  -0.2900917
[2,]  -0.3648639 -0.30081170 -0.2426136   0.3806406
[3,]  -0.3648639 -0.03512914 -0.2426136  -0.2900917
[4,]  -0.3648639 -0.30081170 -0.2426136  -0.2900917
[5,]  -0.3648639 -0.30081170 -0.2426136  -0.2900917
[6,]  -0.3648639 -0.30081170 -0.2426136  -0.2900917
     Br..Cd..352   Br..Cd..5 Others.999    Pr.Cat.1
[1,]  -0.2814842 -0.26749813 -0.1008403 -0.17442555
[2,]  -0.2814842  1.79584830  0.5953831  0.03922943
[3,]  -0.2814842  0.02726565 -0.4809043 -0.56612636
[4,]  -0.2814842 -0.26749813 -1.7556322 -0.99343633
[5,]  -0.2814842 -0.26749813  0.9586301 -0.99343633
[6,]  -0.2814842 -0.26749813  1.1268001 -0.21003472
```

/

```
       Pr.Cat.2     Pr.Cat.3    Pr.Cat.4    PropCat.5
[1,]  0.2143002 -0.034449105 -0.09717115  0.135304773
[2,]  0.1822114 -0.183687103 -0.14932024  0.008900937
[3,] -0.5558329  1.569859381 -0.46221482 -0.686320160
[4,] -0.2991219  1.719097380 -0.46221482 -0.180704817
[5,] -1.4222327  0.002860395  3.76186195  1.114934500
[6,] -0.1386774 -0.258306103  0.94581077  0.103703814
       PropCat.6  PropCat.7  PropCat.8  PropCat.9
[1,] -0.55572458 -0.4950510 -0.5260532 -0.4903393
[2,]  1.54967253 -0.3417581 -0.3948406 -0.3313963
[3,]  0.16612586 -0.3417581 -0.4604469 -0.3313963
[4,] -0.55572458 -0.4950510 -0.5260532 -0.4903393
[5,] -0.55572458 -0.4950510 -0.1980217 -0.4903393
[6,]  0.04581745 -0.4950510 -0.4604469  0.6222620
     PropCat.10 PropCat.11 PropCat.12 PropCat.13
[1,] -0.2654362 -0.2984808  0.9071200  -0.262012
[2,] -0.2654362  0.3103186 -0.2363086  -0.262012
[3,] -0.2654362 -0.2984808  0.5259771  -0.262012
[4,] -0.2654362 -0.2984808 -0.2363086  -0.262012
[5,] -0.2654362 -0.2984808 -0.2363086  -0.262012
[6,] -0.2654362 -0.2984808 -0.2363086  -0.262012
      PropCat.14 PropCat.15    maxbrand
[1,] -0.02455314  3.5964577  0.03013272
[2,] -0.21245979 -0.2897512 -0.81078040
[3,]  1.59144411 -0.2897512  0.62577951
[4,]  1.74176944 -0.2897512  0.80096974
[5,]  0.01302819 -0.2897512 -0.81078040
[6,] -0.25004112  2.7963559 -1.02100867
```

QS 1 Use k-means clustering to identify clusters of households based on: a. The variables that describe purchase behavior (including brand loyalty)

Hide

```
Mdata2<-(Mdata1[,c(1:8,20,36)])
##Max brand loyalty is obtained by taking maximum values out of the variables - Br. Cd. 57,144, Br. Cd. 55, Br. Cd. 272Cd.28
6, Br. Cd.24, Br. Cd.481, Br. Cd.352, Br. Cd.5
##Others999 gives the share of transactions towards other brands which indicates that a customer is not brand loyal

head(Mdata2)
```

/

```
     No..of.Brands Brand.Runs Total.Volume No..of..Trans
[1,]    -0.4030277  0.1200727   -0.5005898    -0.4104681
[2,]     0.8630280  0.8895639    0.2651391     0.5076339
[3,]     0.8630280  2.0438006    1.4394712     1.8274054
[4,]    -1.0360556 -1.1303505   -1.3403176    -1.5580955
[5,]    -0.4030277 -0.9379777   -0.4651989    -1.0416632
[6,]    -0.4030277  0.9857502    0.8056536     0.5650152
          Value Trans...Brand.Runs   Vol.Tran  Avg..Price
[1,] -0.5881031         -0.4636969 -0.3242918 -0.43944366
[2,]  0.3896410         -0.3907514 -0.2639930  0.05217678
[3,]  0.6936645         -0.3523590 -0.1944886 -0.90701745
[4,] -1.3852447         -0.6211057 -0.1610026 -1.13145287
[5,] -0.8451360         -0.1719147  0.8980852 -1.25970168
[6,]  0.4168163         -0.3984298  0.1135176 -0.65586353
     Others.999    maxbrand
[1,] -0.1008403  0.03013272
[2,]  0.5953831 -0.81078040
[3,] -0.4809043  0.62577951
[4,] -1.7556322  0.80096974
[5,]  0.9586301 -0.81078040
[6,]  1.1268001 -1.02100867
```

Hide

```
summary(Mdata2)
```

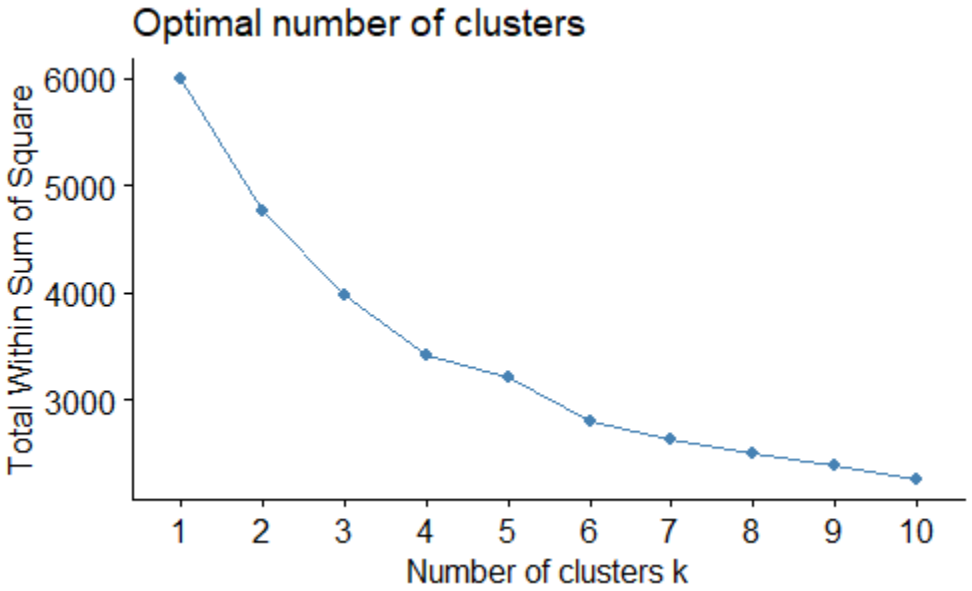
/

```
 No..of.Brands      Brand.Runs       Total.Volume
 Min.   :-1.669   Min.   :-1.4189   Min.   :-1.5141
 1st Qu.:-1.036   1st Qu.:-0.7456   1st Qu.:-0.6550
 Median :-0.403   Median :-0.0723   Median :-0.2001
 Mean   : 0.000   Mean   : 0.0000   Mean   : 0.0000
 3rd Qu.: 0.863   3rd Qu.: 0.5048   3rd Qu.: 0.4413
 Max.   : 3.395   Max.   : 5.6027   Max.   : 5.0165
 No..of..Trans         Value         Trans...Brand.Runs
 Min.   :-1.7302   Min.   :-1.4917   Min.   :-0.62111
 1st Qu.:-0.5252   1st Qu.:-0.6203   1st Qu.:-0.45986
 Median :-0.1809   Median :-0.1374   Median :-0.29669
 Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000
 3rd Qu.: 0.5076   3rd Qu.: 0.3831   3rd Qu.: 0.02773
 Max.   : 6.1310   Max.   : 5.7005   Max.   : 7.82522
    Vol.Tran         Avg..Price        Others.999
 Min.   :-1.2889   Min.   :-1.6605   Min.   :-1.75563
 1st Qu.:-0.6614   1st Qu.:-0.5543   1st Qu.:-0.81808
 Median :-0.2152   Median :-0.1562   Median : 0.01183
 Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000
 3rd Qu.: 0.3049   3rd Qu.: 0.4229   3rd Qu.: 0.86277
 Max.   : 8.4818   Max.   : 5.7432   Max.   : 1.60777
    maxbrand
 Min.   :-1.3013
 1st Qu.:-0.8458
 Median :-0.2502
 Mean   : 0.0000
 3rd Qu.: 0.6959
 Max.   : 2.2025
```

Hide

```
fviz_nbclust(Mdata2, kmeans, method = "wss")##Calculating our optimal K using Elbow chart
```

/



Hide

```
set.seed(120)##Lets run the k-means algorithm to cluster the dataset.
k3 <- kmeans(Mdata2, centers = 3, nstart = 25)
k3center <- as.data.frame(k3$centers)
k3$size
```

```
[1] 259 166 175
```

Hide

```
# Add a column for cluster number / to be also used for color
cluster <- matrix(c("1","2","3"),nrow = 3)
k3center <- cbind(cluster,k3center)
k3center
```

cluster <fctr>	No..of.Brands <dbl>	Brand.Runs <dbl>	Total.Volume <dbl>	No..of..Trans <dbl>	Value <dbl>	Trans...Brand.Runs <dbl>
1 1	-0.2759333	-0.2156513	-0.5323647	-0.4155637	-0.45499157	-0.2473209
2 2	0.9507367	1.0993197	0.6359753	1.0821389	0.76730922	-0.2548053
3 3	-0.4934603	-0.7236194	0.1846318	-0.4114518	-0.05446008	0.6077360

3 rows | 1-8 of 11 columns

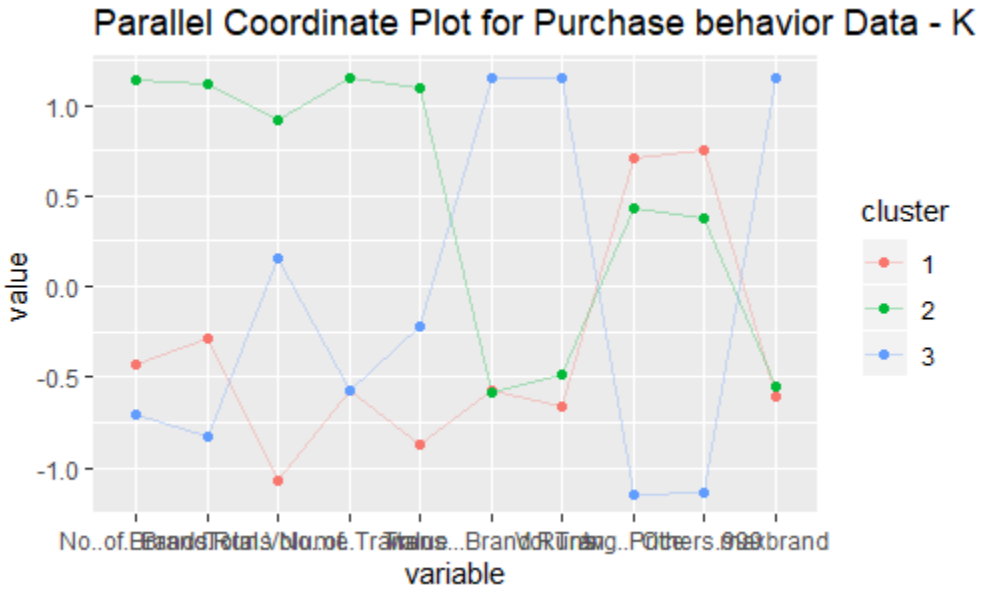
Code

Hide

```
library(hrbrthemes)
library(GGally)
library(viridis)
```

Hide

```
##visualize the data
ggparcoord(k3center,
    columns = 2:11, groupColumn = 1,
    showPoints = TRUE,
    title = "Parallel Coordinate Plot for Purchase behavior Data - K = 3",
    alphaLines = 0.3
    )
```



Cluster 1: This group has the lowest brand loyalty because they mostly buy from other999 brands and also has the lowest number of brands and they have highest average price of total transaction. it is least interest because it has lowest value

Cluster 2: This group purchase a large number of brands with highest brand runs. They have highest number and least volume of transactions whereas their brand loyalty is between cluster 1 and cluster 3 customers.

/

Cluster 3: This group has maximum brand loyality and also has the lowest number of brands and least average price of total transaction.

QS b The variables that describe the basis for purchase.

Hide

```
Mdata3 <- Mdata1[,c(9:11,21:35)]
head(Mdata3)
```

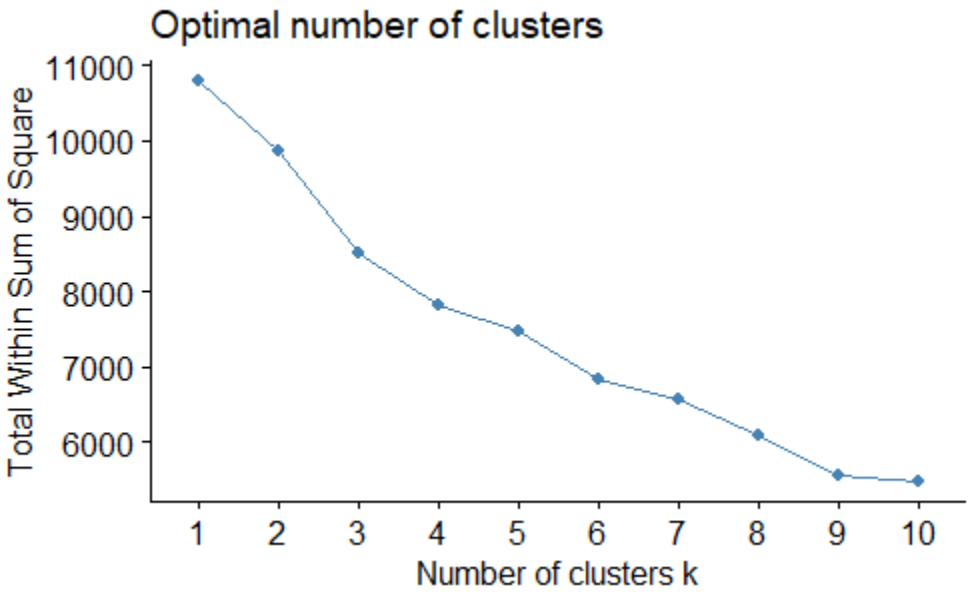
/

```
     Pur.Vol.No.Promo.... Pur.Vol.Promo.6..
[1,]            0.7269843        -0.5756626
[2,]           -0.1927198         0.4986688
[3,]            0.2253275        -0.3607963
[4,]            0.7269843        -0.5756626
[5,]           -2.5337850         0.9284014
[6,]            0.7269843        -0.5756626
     Pur.Vol.Other.Promo..    Pr.Cat.1   Pr.Cat.2
[1,]           -0.46438366 -0.17442555  0.2143002
[2,]           -0.18644830  0.03922943  0.1822114
[3,]            0.09148706 -0.56612636 -0.5558329
[4,]           -0.46438366 -0.99343633 -0.2991219
[5,]            2.87084066 -0.99343633 -1.4222327
[6,]           -0.46438366 -0.21003472 -0.1386774
         Pr.Cat.3    Pr.Cat.4    PropCat.5   PropCat.6
[1,] -0.034449105 -0.09717115  0.135304773 -0.55572458
[2,] -0.183687103 -0.14932024  0.008900937  1.54967253
[3,]  1.569859381 -0.46221482 -0.686320160  0.16612586
[4,]  1.719097380 -0.46221482 -0.180704817 -0.55572458
[5,]  0.002860395  3.76186195  1.114934500 -0.55572458
[6,] -0.258306103  0.94581077  0.103703814  0.04581745
      PropCat.7  PropCat.8  PropCat.9 PropCat.10
[1,] -0.4950510 -0.5260532 -0.4903393 -0.2654362
[2,] -0.3417581 -0.3948406 -0.3313963 -0.2654362
[3,] -0.3417581 -0.4604469 -0.3313963 -0.2654362
[4,] -0.4950510 -0.5260532 -0.4903393 -0.2654362
[5,] -0.4950510 -0.1980217 -0.4903393 -0.2654362
[6,] -0.4950510 -0.4604469  0.6222620 -0.2654362
     PropCat.11 PropCat.12 PropCat.13  PropCat.14
[1,] -0.2984808  0.9071200  -0.262012 -0.02455314
[2,]  0.3103186 -0.2363086  -0.262012 -0.21245979
[3,] -0.2984808  0.5259771  -0.262012  1.59144411
[4,] -0.2984808 -0.2363086  -0.262012  1.74176944
[5,] -0.2984808 -0.2363086  -0.262012  0.01302819
[6,] -0.2984808 -0.2363086  -0.262012 -0.25004112
     PropCat.15
[1,]  3.5964577
[2,] -0.2897512
[3,] -0.2897512
[4,] -0.2897512
[5,] -0.2897512
[6,]  2.7963559
```

/

Hide

```
fviz_nbclust(Mdata3, kmeans, method = "wss")
```



Hide

```
set.seed(120)
k3a <- kmeans(Mdata3, centers = 3, nstart = 25)
```

Hide

```
k3acenter <- as.data.frame(k3a$centers)
```

Hide

```
# Add a column for cluster number / to be also used for color
acluster <- matrix(c("1","2","3"),nrow = 3)
k3acenter <- cbind(acluster,k3acenter)
k3acenter
```

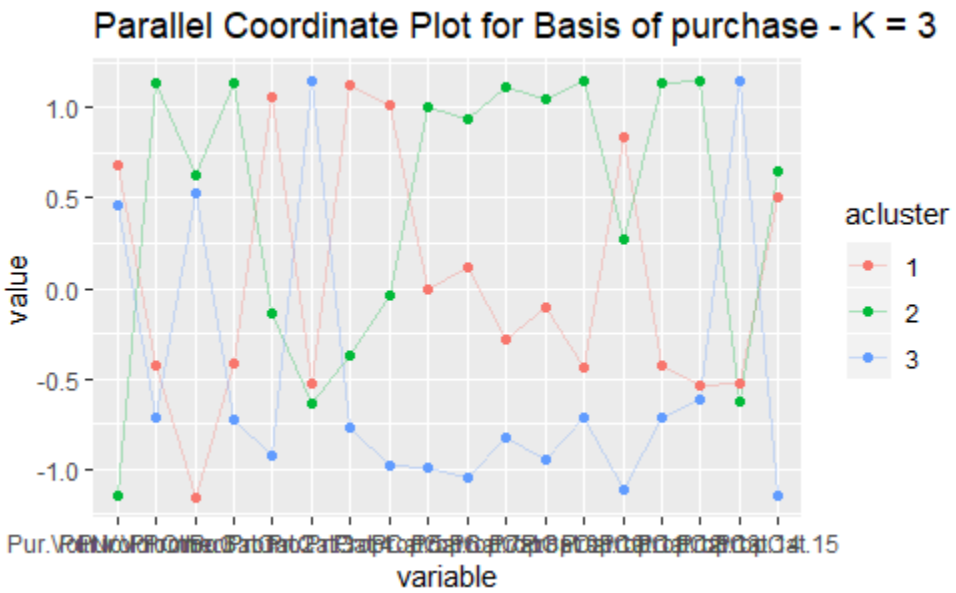
acluster <fctr>	Pur.Vol.No.Promo.... <dbl>	Pur.Vol.Promo.6.. <dbl>	Pur.Vol.Other.Promo.. <dbl>	Pr.Cat.1 <dbl>	Pr.Cat.2 <dbl>
1 1	0.2860651	-0.2360563	-0.1703973	-0.4651435	0.5449425

acluster <fctr>	Pur.Vol.No.Promo.... <dbl>	Pur.Vol.Promo.6.. <dbl>	Pur.Vol.Other.Promo.. <dbl>	Pr.Cat.1 <dbl>	Pr.Cat.2 <dbl>
2 2	-0.5626809	0.5576736	0.2131738	1.1091649	-0.4708722
3 3	0.1856666	-0.3842112	0.1912587	-0.7825205	-1.1334328

3 rows | 1-7 of 19 columns

Hide

```
ggparcoord(k3acenter,
    columns = 2:19, groupColumn = 1,
    showPoints = TRUE,
    title = "Parallel Coordinate Plot for Basis of purchase - K = 3",
    alphaLines = 0.3
    )
```



Cluster1 customer in this cluster are responsive to pricing category 2,4 and 5

Cluster 2 customer in this cluster are highly responsive to promotion and pricing category 1

Cluster3 they buy products irespective of promotion, they are highely responsive to selling proposition and pricing category 3.

QS c The variables that describe both purchase behavior and basis of purchase

/

Hide

```
Mdata4 <-Mdata1[,c(1:11,20:36)]
head(Mdata4)
```

/

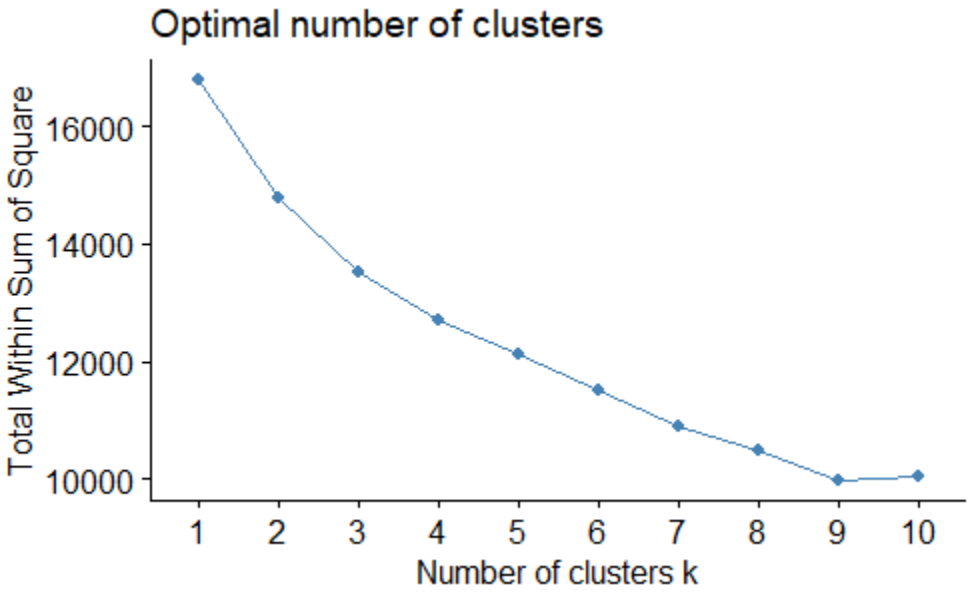
```
     No..of.Brands Brand.Runs Total.Volume No..of..Trans
[1,]    -0.4030277  0.1200727   -0.5005898    -0.4104681
[2,]     0.8630280  0.8895639    0.2651391     0.5076339
[3,]     0.8630280  2.0438006    1.4394712     1.8274054
[4,]    -1.0360556 -1.1303505   -1.3403176    -1.5580955
[5,]    -0.4030277 -0.9379777   -0.4651989    -1.0416632
[6,]    -0.4030277  0.9857502    0.8056536     0.5650152
          Value Trans...Brand.Runs   Vol.Tran  Avg..Price
[1,] -0.5881031         -0.4636969 -0.3242918 -0.43944366
[2,]  0.3896410         -0.3907514 -0.2639930  0.05217678
[3,]  0.6936645         -0.3523590 -0.1944886 -0.90701745
[4,] -1.3852447         -0.6211057 -0.1610026 -1.13145287
[5,] -0.8451360         -0.1719147  0.8980852 -1.25970168
[6,]  0.4168163         -0.3984298  0.1135176 -0.65586353
     Pur.Vol.No.Promo.... Pur.Vol.Promo.6..
[1,]            0.7269843        -0.5756626
[2,]           -0.1927198         0.4986688
[3,]            0.2253275        -0.3607963
[4,]            0.7269843        -0.5756626
[5,]           -2.5337850         0.9284014
[6,]            0.7269843        -0.5756626
     Pur.Vol.Other.Promo.. Others.999    Pr.Cat.1
[1,]           -0.46438366 -0.1008403 -0.17442555
[2,]           -0.18644830  0.5953831  0.03922943
[3,]            0.09148706 -0.4809043 -0.56612636
[4,]           -0.46438366 -1.7556322 -0.99343633
[5,]            2.87084066  0.9586301 -0.99343633
[6,]           -0.46438366  1.1268001 -0.21003472
       Pr.Cat.2     Pr.Cat.3    Pr.Cat.4    PropCat.5
[1,]  0.2143002 -0.034449105 -0.09717115  0.135304773
[2,]  0.1822114 -0.183687103 -0.14932024  0.008900937
[3,] -0.5558329  1.569859381 -0.46221482 -0.686320160
[4,] -0.2991219  1.719097380 -0.46221482 -0.180704817
[5,] -1.4222327  0.002860395  3.76186195  1.114934500
[6,] -0.1386774 -0.258306103  0.94581077  0.103703814
       PropCat.6  PropCat.7  PropCat.8  PropCat.9
[1,] -0.55572458 -0.4950510 -0.5260532 -0.4903393
[2,]  1.54967253 -0.3417581 -0.3948406 -0.3313963
[3,]  0.16612586 -0.3417581 -0.4604469 -0.3313963
[4,] -0.55572458 -0.4950510 -0.5260532 -0.4903393
[5,] -0.55572458 -0.4950510 -0.1980217 -0.4903393
[6,]  0.04581745 -0.4950510 -0.4604469  0.6222620
```

/

```
     PropCat.10 PropCat.11 PropCat.12 PropCat.13
[1,] -0.2654362 -0.2984808  0.9071200  -0.262012
[2,] -0.2654362  0.3103186 -0.2363086  -0.262012
[3,] -0.2654362 -0.2984808  0.5259771  -0.262012
[4,] -0.2654362 -0.2984808 -0.2363086  -0.262012
[5,] -0.2654362 -0.2984808 -0.2363086  -0.262012
[6,] -0.2654362 -0.2984808 -0.2363086  -0.262012
      PropCat.14 PropCat.15    maxbrand
[1,] -0.02455314  3.5964577  0.03013272
[2,] -0.21245979 -0.2897512 -0.81078040
[3,]  1.59144411 -0.2897512  0.62577951
[4,]  1.74176944 -0.2897512  0.80096974
[5,]  0.01302819 -0.2897512 -0.81078040
[6,] -0.25004112  2.7963559 -1.02100867
```

Hide

```
fviz_nbclust(Mdata4, kmeans, method = "wss")
```



Hide

```
set.seed(120)
k3b <- kmeans(Mdata4, centers = 3, nstart = 25)
k3bcenter <- as.data.frame(k3b$centers)
```

/

Hide

```
# Add a column for cluster number / to be also used for color
bcluster <- matrix(c("1","2","3"),nrow = 3)
k3bcenter <- cbind(bcluster,k3bcenter)
k3bcenter
```

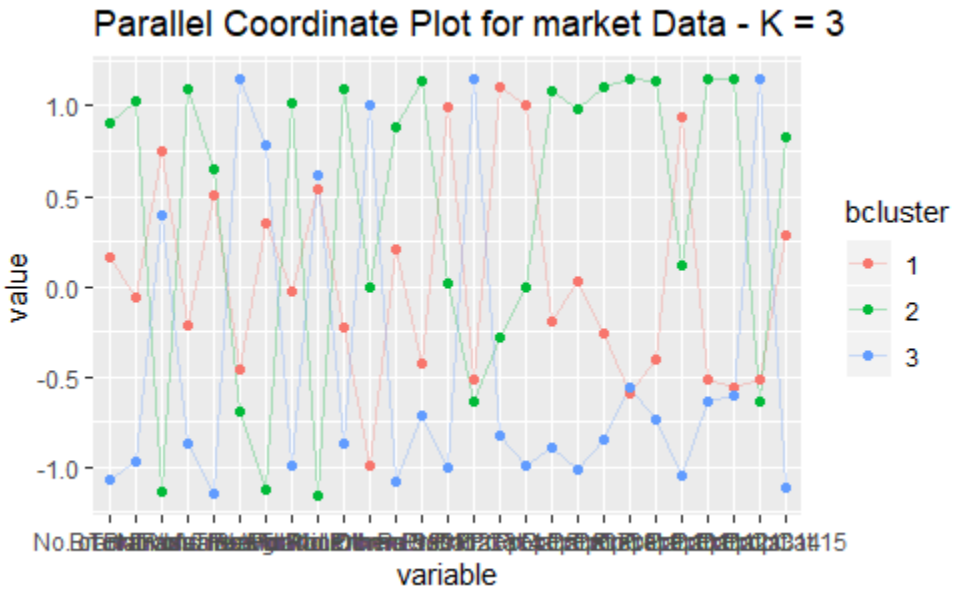
bcluster <fctr>	No..of.Brands <dbl>	Brand.Runs <dbl>	Total.Volume <dbl>	No..of..Trans <dbl>	Value <dbl>	Trans...Brand.Runs <dbl>
1 1	-0.06413404	-0.2128553	0.16901166	-0.1714756	0.04853479	-0.06387311
2 2	0.25424375	0.5039997	-0.24019541	0.3408616	0.09947329	-0.23090832
3 3	-0.59852165	-0.8120867	0.09190381	-0.4290326	-0.55575082	1.07696718

3 rows | 1-8 of 29 columns

Hide

```
ggparcoord(k3bcenter,
    columns = 2:28, groupColumn = 1,
    showPoints = TRUE,
    title = "Parallel Coordinate Plot for market Data - K = 3",
    alphaLines = 0.3
    )
```

/



Cluster1 Customers in this cluster have very low brand loyalty and they buy products from other brands very often. customer are highly responsive towards pricing category 2.

Cluster2 Customers in this cluster have low brand loyalty and buy products from other brands very often.

Cluster3 Customers in this cluster are highly brand loyal and very much responsive to promotion. and the volume of transaction is also highest. the customes in this cluster are very responsive to price category 3.

Hide

```
k3b$size
```

```
[1] 297 235  68
```

We can add demographic information:

Hide

```
Mdata5<-Mdata[,c(2:11)]
head(Mdata5)
```

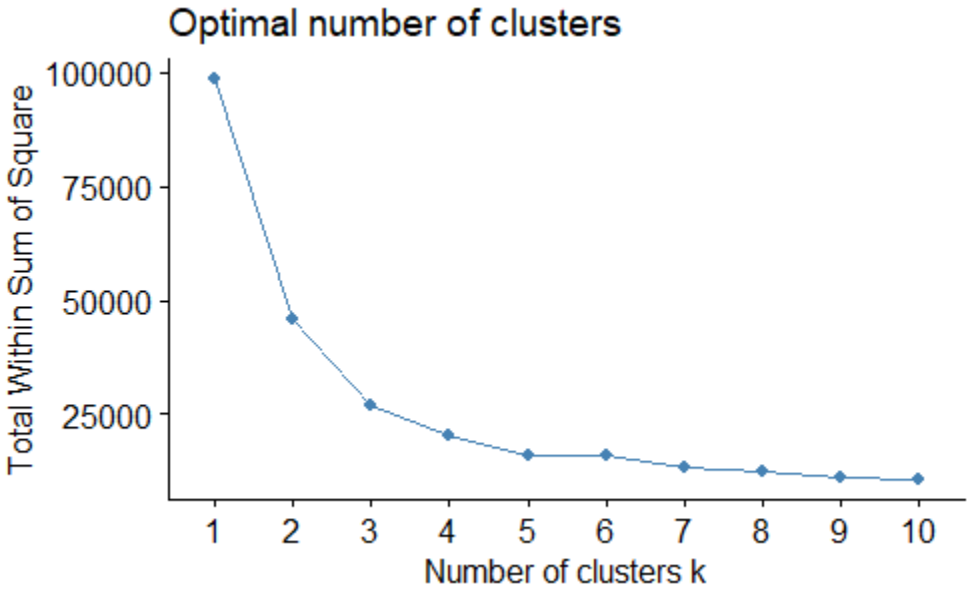
	SEC <int>	FEH <int>	MT <int>	SEX <int>	AGE <int>	EDU <int>	HS <int>	CHILD <int>	CS <int>
1	4	3	10	1	4	4	2	4	1

	SEC <int>	FEH <int>	MT <int>	SEX <int>	AGE <int>	EDU <int>	HS <int>	CHILD <int>	CS <int>
2	3	2	10	2	2	4	4	2	1
3	2	3	10	2	4	5	6	4	1
4	4	0	0	0	4	0	0	5	0
5	4	1	10	2	3	4	4	3	1
6	4	3	10	2	3	4	5	2	1

6 rows | 1-10 of 10 columns

Hide

```
fviz_nbclust(Mdata5, kmeans, method = "wss")
```



Hide

```
set.seed(120)
k3c <- kmeans(Mdata5, centers = 3, nstart = 25)
k3ccenter <- as.data.frame(k3c$centers)
```

Hide

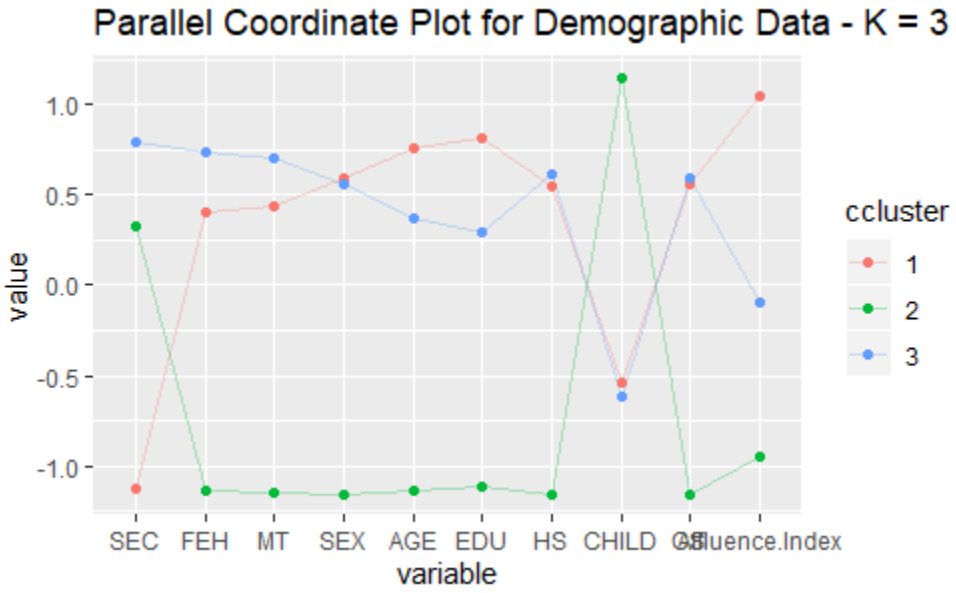
```
# Add a column for cluster number / to be also used for color
ccluster <- matrix(c("1","2","3"),nrow = 3)
k3ccenter <- cbind(ccluster,k3ccenter)
k3ccenter
```

ccluster <fctr>	SEC <dbl>	FEH <dbl>	MT <dbl>	SEX <dbl>	AGE <dbl>	EDU <dbl>	HS <dbl>	CHILD <dbl>
1 1	1.687898	2.0955414	8.547771	1.9808917	3.343949	5.490446	4.681529	3.044586
2 2	2.562500	0.7083333	2.447917	0.5520833	2.875000	0.937500	1.114583	4.489583
3 3	2.850144	2.3976945	9.596542	1.9567723	3.247839	4.247839	4.821326	2.971182

3 rows | 1-10 of 11 columns

Hide

```
ggparcoord(k3ccenter,
    columns = 2:11, groupColumn = 1,
    showPoints = TRUE,
    title = "Parallel Coordinate Plot for Demographic Data - K = 3",
    alphaLines = 0.3
    )
```



Cluster1 customer are high socio economic class and maximum household member.

Cluster2 customer have average socio economic class and low affluence index. customer have minimum household member, and they are least educated.

Cluster3 customer are having high affluence index and they are educated class.

2. 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.)

Cluster 1 is distinguished mostly by the purchase behavior variables - it has lowest brand loyalty together with low value, volume and frequency. The customers having lowest number of brand as well - this group is not particularly responsive to promotions, pricing or selling propositions. Demographically it has high socio economic class, and maximum household member.

Cluster 2 stands out in both groups of variables - it has average loyalty, highest number of brand runs,low value and price per purchase, and very differential response to price (highly responsive to categories 2,4 and 5, unresponsive to category 3), and selling proposition Demographically it has low affluence index and education and it has relatively small family size.

Cluster 3 is characterized by highest brand loyalty,with highest volume of transaction.low volume and sensitivity to promotions and price (responsive to pricing cat. 3), and highly responsive to selling proposition. Demographically, it has high affluence index, and highest educated class of high socio-economic status.

The three clusters are well separated in almost all variables.
Cluster 3 has more loyal customer with lower socioeconomic status. Volume of transaction is also more in this cluster, the customers are very much responsive to promotion and price category 3.

So our success category is cluster 3, the more affluent group, lower socioeconomic group, which also turns out to be highly loyal. we can target these customer by running promotions and offers.

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.

Hide

```
Mdata8<-Mdata[,c(23:31)] ##to build a model based on brand loyalty
```

Hide

```
Kbrand <- kmeans(Mdata8,centers = 2,nstart = 25)
Kbrand
```

/

```
K-means clustering with 2 clusters of sizes 281, 319

Cluster means:
  Br..Cd..57..144 Br..Cd..55 Br..Cd..272 Br..Cd..286
1       0.3023843  0.2440925  0.02932384  0.04565836
2       0.0800000  0.0284326  0.03655172  0.02366771
  Br..Cd..24 Br..Cd..481 Br..Cd..352  Br..Cd..5 Others.999
1 0.01932384  0.03455516  0.05259786 0.02163701  0.2510320
2 0.01934169  0.01836991  0.01799373 0.01507837  0.7606552

Clustering vector:
  [1] 1 2 1 1 2 2 1 1 2 1 2 1 2 1 2 1 2 1 2 2 1 1 1 1 1 2 1
 [28] 1 2 1 1 2 1 1 1 2 2 2 1 1 2 1 1 1 2 1 2 1 1 2 1 1 1 2
 [55] 1 1 1 2 1 1 2 1 1 1 2 1 1 2 2 1 1 2 1 2 1 2 2 1 2 2 2
 [82] 1 1 2 2 1 2 1 1 1 2 1 1 2 2 2 1 1 1 1 1 1 1 2 2 2 2 2
[109] 2 2 1 1 2 2 2 2 2 1 1 2 2 2 1 2 1 1 1 1 1 2 2 2 1 1 1
[136] 2 2 1 1 2 2 1 2 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 2 1 1 1
[163] 1 2 1 1 2 2 1 2 2 2 2 1 1 1 1 1 1 2 2 2 2 2 1 1 2 2 1
[190] 2 2 2 1 1 1 2 1 1 2 1 1 2 2 2 1 2 2 2 1 1 1 2 2 1 2 2
[217] 1 1 1 1 2 1 1 2 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 2 2
[244] 1 1 1 2 2 2 1 2 2 1 2 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 2
[271] 2 2 1 2 2 2 2 2 1 2 1 2 2 1 2 2 2 1 1 2 2 2 1 2 2 2 1
[298] 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2
[325] 1 1 1 1 1 2 1 1 2 2 2 2 2 2 1 2 1 1 2 2 1 2 2 2 1 2 1
[352] 1 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2
[379] 1 1 2 2 2 2 2 1 1 2 1 2 2 2 2 1 1 2 2 2 2 1 2 2 1 1 2
[406] 2 2 2 1 1 2 2 2 1 2 2 2 1 1 1 2 2 2 1 2 2 2 2 2 1 1 2
[433] 2 2 2 1 2 1 1 2 1 2 1 2 2 1 1 2 1 1 2 1 1 2 1 1 2 2 2
[460] 2 2 2 1 1 2 1 2 1 1 2 2 2 2 2 2 1 2 2 1 2 1 2 2 2 2 1
[487] 2 1 2 2 1 2 1 1 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 1
[514] 2 1 1 1 2 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1
[541] 2 2 1 2 2 2 2 1 1 2 1 2 1 1 1 2 2 1 2 1 2 1 2 2 2 2 2
[568] 2 1 2 1 1 2 2 2 1 2 2 2 1 2 2 1 2 1 2 1 2 2 2 1 2 1 2
[595] 1 1 2 2 2 1

Within cluster sum of squares by cluster:
[1] 87.83885 18.42597
 (between_SS / total_SS =  33.5 %)

Available components:

[1] "cluster"      "centers"      "totss"
```

/

```
[4] "withinss"     "tot.withinss" "betweenss"
[7] "size"         "iter"         "ifault"
```

Hide

```
set.seed(120)
Kbrand1<-as.data.frame(Kbrand$centers)
```

Hide

```
dcluster <- matrix(c("1","2"),nrow = 2)
Kbrand2 <- cbind(dcluster,Kbrand1)
Kbrand2
```

dcluster <fctr>	Br..Cd..57..144 <dbl>	Br..Cd..55 <dbl>	Br..Cd..272 <dbl>	Br..Cd..286 <dbl>	Br..Cd..24 <dbl>	Br..Cd..481 <dbl>	Br..Cd..352 <dbl>
1 1	0.3023843	0.2440925	0.02932384	0.04565836	0.01932384	0.03455516	0.05259786
2 2	0.0800000	0.0284326	0.03655172	0.02366771	0.01934169	0.01836991	0.01799373

2 rows | 1-9 of 10 columns

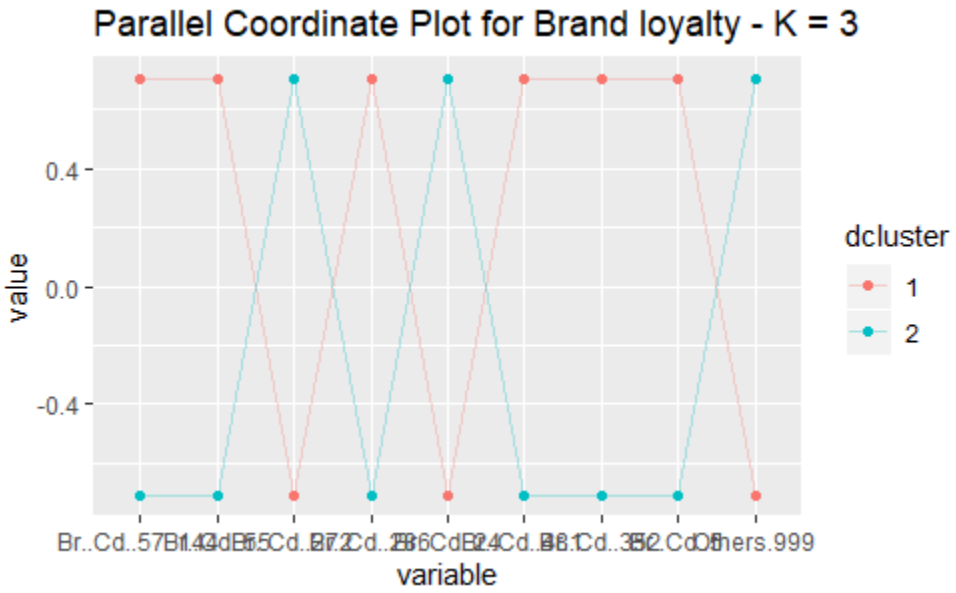
Hide

```
NA
```

Hide

```
ggparcoord(Kbrand2,
    columns = 2:10, groupColumn = 1,
    showPoints = TRUE,
    title = "Parallel Coordinate Plot for Brand loyalty - K = 3",
    alphaLines = 0.3
    )
```

/



Hide

```
Mdata6 <- cbind(Mdata[,2:11],Mdata1[,-(12:19)])
```

Hide

```
Mdata9 <- cbind(Mdata6, clusterNum = Kbrand$cluster)

Mdata9$clusterNum <- as.factor(Mdata9$clusterNum)
head(Mdata9)
```

PropCat.10 <dbl>	PropCat.11 <dbl>	PropCat.12 <dbl>	PropCat.13 <dbl>	PropCat.14 <dbl>	PropCat.15 <dbl>	maxbrand <dbl>	clusterNum <fctr>
-0.2654362	-0.2984808	0.9071200	-0.262012	-0.02455314	3.5964577	0.03013272	1
-0.2654362	0.3103186	-0.2363086	-0.262012	-0.21245979	-0.2897512	-0.81078040	2
-0.2654362	-0.2984808	0.5259771	-0.262012	1.59144411	-0.2897512	0.62577951	1
-0.2654362	-0.2984808	-0.2363086	-0.262012	1.74176944	-0.2897512	0.80096974	1
-0.2654362	-0.2984808	-0.2363086	-0.262012	0.01302819	-0.2897512	-0.81078040	2
-0.2654362	-0.2984808	-0.2363086	-0.262012	-0.25004112	2.7963559	-1.02100867	2

6 rows | 33-40 of 39 columns

Code

`

The customer who are in cluster 1 are brand loyal customer and customers in cluster 2 are non brand loyal, they freqently do shopping from other999 brands.