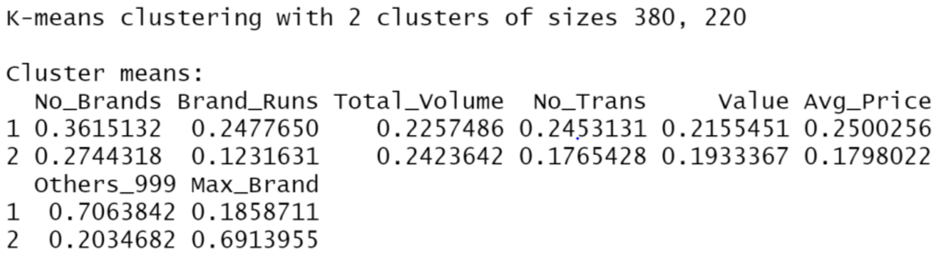
Bath Soap Case

Q1] Use k-means clustering to identify clusters of households based on (a) The variables that describe purchase behavior (including brand loyalty).
[Variables: # brands, brand runs, total volume, # transactions, value, Avg. price, share to other brands, max to one brand].

1. Clustering based on customer purchase behavior
    - Cluster size = 2. Here, clusters sizes are not same. Looks like cluster 1 has more customers which basically means that customers who are not brand loyal are more in comparison to brand loyal customers

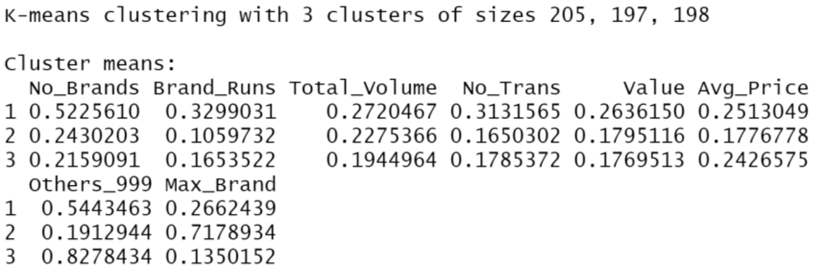


```
K-means clustering with 2 clusters of sizes 380, 220

Cluster means:
  No_Brands Brand_Runs Total_Volume  No_Trans     Value Avg_Price
1 0.3615132  0.2477650    0.2257486 0.2453131 0.2155451 0.2500256
2 0.2744318  0.1231631    0.2423642 0.1765428 0.1933367 0.1798022
  Others_999 Max_Brand
1  0.7063842 0.1858711
2  0.2034682 0.6913955
```

Here, clustering is basically one on brand loyalty, considering important variables as Others_9999 & Max_Brand

Cluster 1: Cluster 1 has greater value of Others_999 & less value of Max_Brand. This cluster basically represents customers who are particularly not loyal to any brand.
Cluster 2: Cluster 2 has greater value of Max_Brand & less value of Others_999. This cluster basically represents customers who are brand loyal.

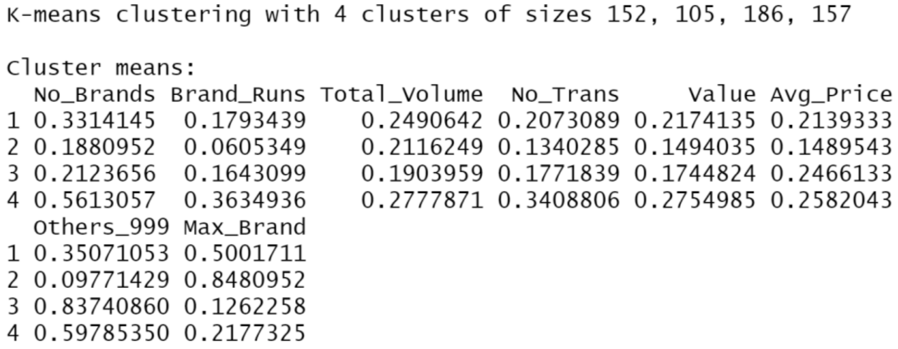
- Cluster size = 3. Here, sizes of clusters are evenly distributed.



```
K-means clustering with 3 clusters of sizes 205, 197, 198

Cluster means:
  No_Brands Brand_Runs Total_Volume  No_Trans     Value Avg_Price
1 0.5225610  0.3299031    0.2720467 0.3131565 0.2636150 0.2513049
2 0.2430203  0.1059732    0.2275366 0.1650302 0.1795116 0.1776778
3 0.2159091  0.1653522    0.1944964 0.1785372 0.1769513 0.2426575
  Others_999 Max_Brand
1  0.5443463 0.2662439
2  0.1912944 0.7178934
3  0.8278434 0.1350152
```

Cluster 1: This cluster has the largest value of No_Brands, Brand_Runs & No_Trans value. This group has moderate value of Others & Max_Brand as compared to other groups.
Cluster 2: Cluster 2 has greatest value of Max_Brand & least value of Others_999. This cluster basically represents customers who are brand loyal. Looks like customers who are brand loyal contribute to the lowest value of Brand_runs, No_Trans & Avg_Price.
Cluster 3: Customers in this group are not at all brand loyal with highest value of Others_999 & least value of Max_Brand. They have least number brands & Volume of Transactions.

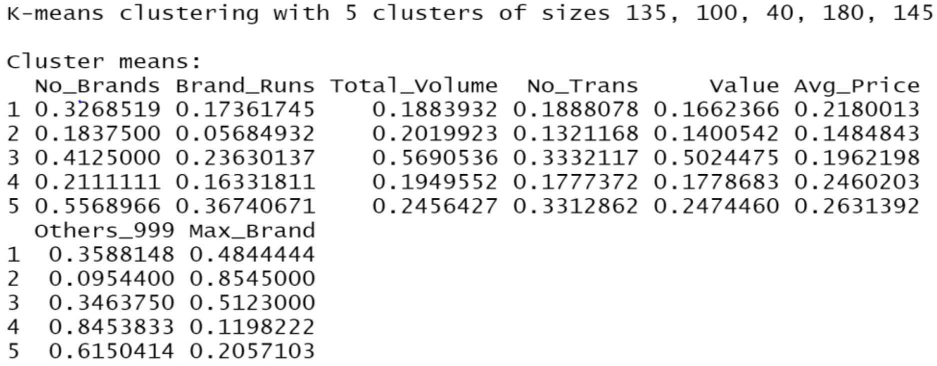
- Cluster size = 4. Looks like all the clusters are nearly equally distributed except for cluster 2.



```
K-means clustering with 4 clusters of sizes 152, 105, 186, 157

Cluster means:
  No_Brands Brand_Runs Total_Volume  No_Trans     Value Avg_Price
1 0.3314145  0.1793439    0.2490642 0.2073089 0.2174135 0.2139333
2 0.1880952  0.0605349    0.2116249 0.1340285 0.1494035 0.1489543
3 0.2123656  0.1643099    0.1903959 0.1771839 0.1744824 0.2466133
4 0.5613057  0.3634936    0.2777871 0.3408806 0.2754985 0.2582043
  Others_999 Max_Brand
1 0.35071053 0.5001711
2 0.09771429 0.8480952
3 0.83740860 0.1262258
4 0.59785350 0.2177325
```

Cluster 1: Cluster 1 seems to be moderately brand loyal and has a considerably less value for Others_999.
Cluster 2: This group is highly brand loyal. Customers in this group have the lowest value for No_Brands, Brand_runs, Value & Avg_Price.
Cluster 3: Customers in this group are not at all brand loyal with highest value of Others_999 & least value of Max_Brand. They have the least number of Total_Volume.
Cluster 4: Customers in this group are less brand loyal and more tending towards other brands. This group has highest value for all the other variables though.

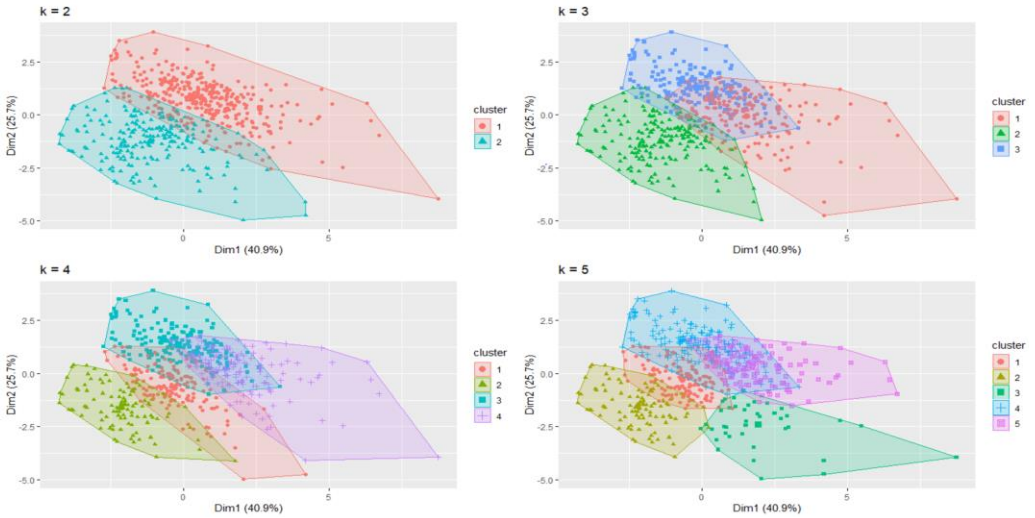
- Cluster = 5. As we can see, clusters doesn't seem to be evenly distributed. Cluster 3 has the lowest number of such instances.



```
K-means clustering with 5 clusters of sizes 135, 100, 40, 180, 145

Cluster means:
  No_Brands Brand_Runs Total_Volume  No_Trans     Value Avg_Price
1 0.3268519 0.17361745    0.1883932 0.1888078 0.1662366 0.2180013
2 0.1837500 0.05684932    0.2019923 0.1321168 0.1400542 0.1484843
3 0.4125000 0.23630137    0.5690536 0.3332117 0.5024475 0.1962198
4 0.2111111 0.16331811    0.1949552 0.1777372 0.1778683 0.2460203
5 0.5568966 0.36740671    0.2456427 0.3312862 0.2474460 0.2631392
  Others_999 Max_Brand
1  0.3588148 0.4844444
2  0.0954400 0.8545000
3  0.3463750 0.5123000
4  0.8453833 0.1198222
5  0.6150414 0.2057103
```

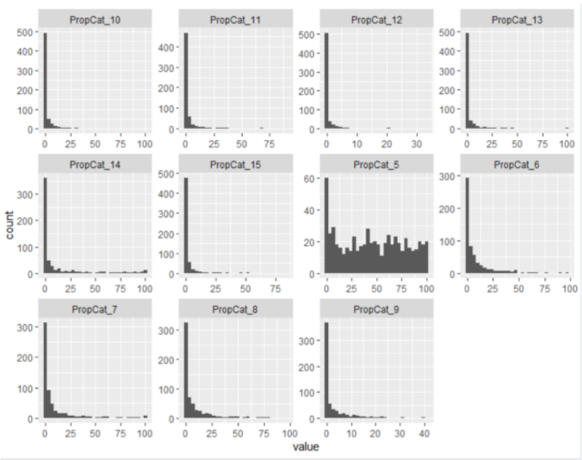
Cluster 1: Cluster 1 group seem to be a little brand loyal and has the least value for Total_Volume.
Cluster 2: This group is highly brand loyal. Customers in this group have the lowest value for No_Brands, Brand_runs, Value & Avg_Price.
Cluster 3: Customers in this group are moderately brand loyal with least value of Max_Brand. These are the customers have highest number of Total_Volume, No_Trans & Value.
Cluster 4: Customers in this group are the least brand loyal and more tending towards other brands. They are neither least or highest in other categories when compared to other clusters.
Cluster 5 : This group customers are less brand loyal. Customers in this group have the highest value for No_Brands, Brand_runs & Avg_Price.

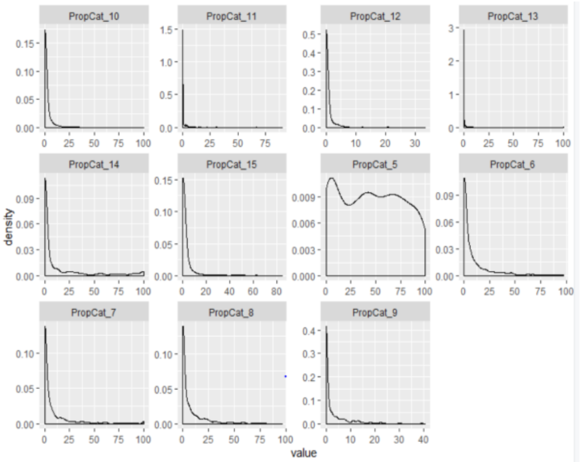
Clusters plot of Customer Purchase Behavior.



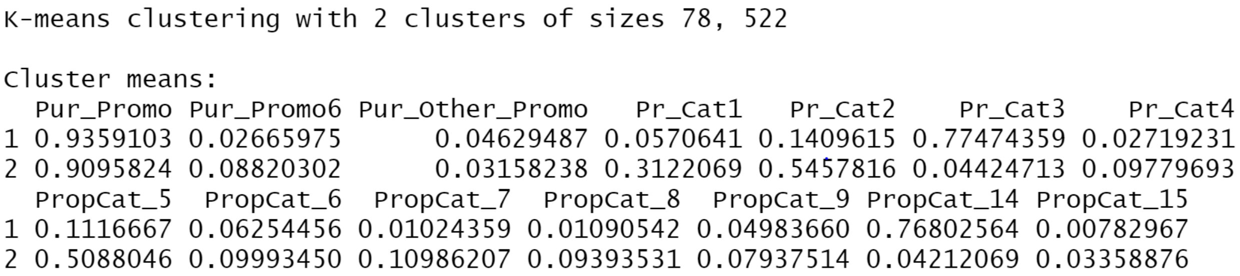
(b) The variables that describe basis-for-purchase.
[Variables: Pur-vol-no-promo, Pur-vol-promo-6, Pur-vol-other, all price categories, selling propositions]
[Note: would you use all selling-propositions? Explore the data.]

From the below graph, it's very clear that propositions category 10-13 have very less values as compared to other proposition categories. Thus, considering Selling propositions category from 5-9, 14 & 15.





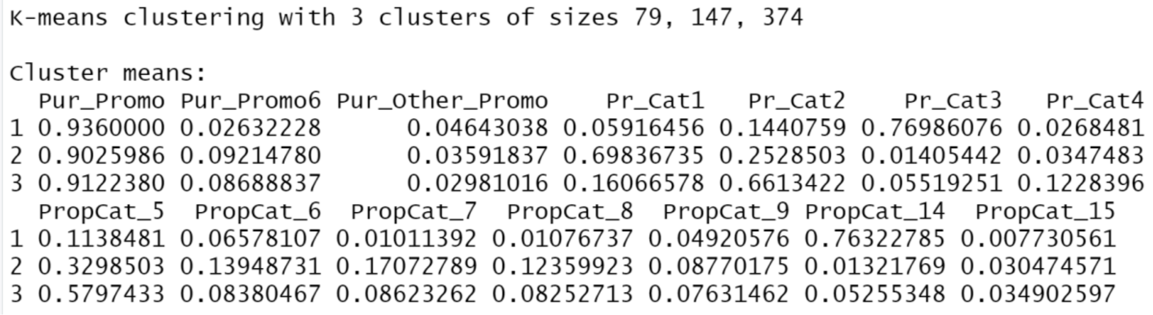
2. Clustering based on basis-for-purchase
    - Cluster size = 2. Here, clusters size are not evenly distributed. Cluster 1 has very few customers as compared to cluster 2



```
K-means clustering with 2 clusters of sizes 78, 522

Cluster means:
  Pur_Promo Pur_Promo6 Pur_Other_Promo   Pr_Cat1   Pr_Cat2    Pr_Cat3    Pr_Cat4
1 0.9359103 0.02665975      0.04629487 0.0570641 0.1409615 0.77474359 0.02719231
2 0.9095824 0.08820302      0.03158238 0.3122069 0.5457816 0.04424713 0.09779693
  PropCat_5  PropCat_6  PropCat_7  PropCat_8  PropCat_9 PropCat_14 PropCat_15
1 0.1116667 0.06254456 0.01024359 0.01090542 0.04983660 0.76802564 0.00782967
2 0.5088046 0.09993450 0.10986207 0.09393531 0.07937514 0.04212069 0.03358876
```

Cluster 1: Customers in this group buy Pr_Cat3 which is any economy or carbolic products. Hence, PropCat_14 is higher which belongs to carbolic products too. Customers in this cluster doesn't usually buy any popular or premium soap as compared to cluster 2. This group usually buys product when they are not on promotion.
Cluster 2: Customers in this group tend to buy soaps/beauty products. However, more of popular soaps than premium soaps. Even this group usually buys product when they are not on promotion.

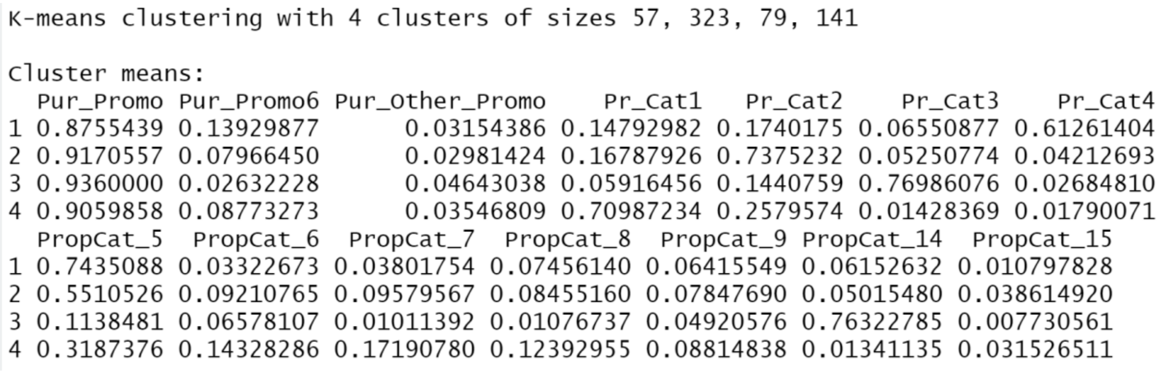
- Cluster size = 3. Here, sizes of clusters are not evenly distributed.



```
K-means clustering with 3 clusters of sizes 79, 147, 374

Cluster means:
  Pur_Promo Pur_Promo6 Pur_Other_Promo    Pr_Cat1   Pr_Cat2    Pr_Cat3   Pr_Cat4
1 0.9360000 0.02632228      0.04643038 0.05916456 0.1440759 0.76986076 0.0268481
2 0.9025986 0.09214780      0.03591837 0.69836735 0.2528503 0.01405442 0.0347483
3 0.9122380 0.08688837      0.02981016 0.16066578 0.6613422 0.05519251 0.1228396
  PropCat_5  PropCat_6  PropCat_7  PropCat_8  PropCat_9 PropCat_14  PropCat_15
1 0.1138481 0.06578107 0.01011392 0.01076737 0.04920576 0.76322785 0.007730561
2 0.3298503 0.13948731 0.17072789 0.12359923 0.08770175 0.01321769 0.030474571
3 0.5797433 0.08380467 0.08623262 0.08252713 0.07631462 0.05255348 0.034902597
```

Cluster 1: Customers in this cluster buy any economy/carbolic products that are not on promotion.
Cluster 2: Customers in this cluster buy any premium soaps that are not on promotion.
Cluster 3: Customers in this group tend to buy beauty products & are inclined towards popular soaps. They mostly buy products when there is no promotion going on.

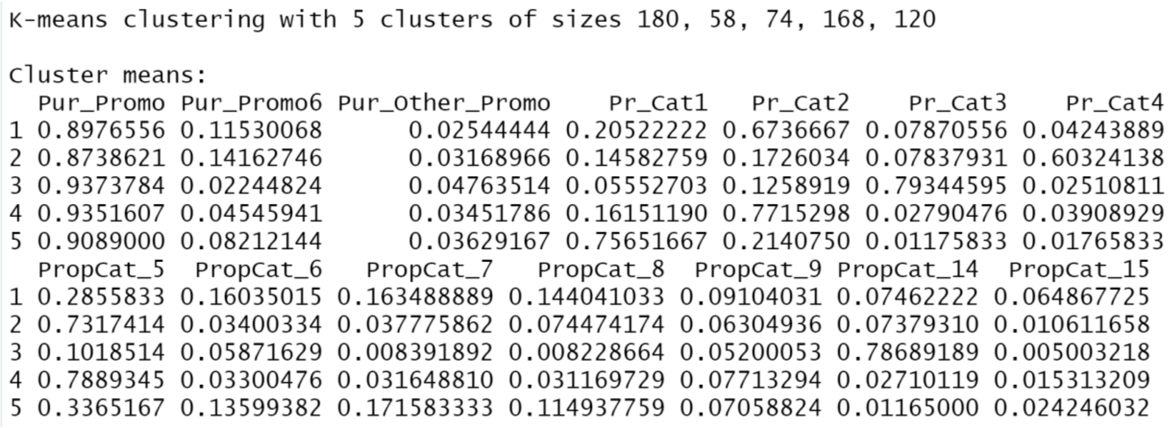
- Cluster size = 4. Custer sizes again vary.



```
K-means clustering with 4 clusters of sizes 57, 323, 79, 141

Cluster means:
  Pur_Promo Pur_Promo6 Pur_Other_Promo    Pr_Cat1   Pr_Cat2    Pr_Cat3    Pr_Cat4
1 0.8755439 0.13929877      0.03154386 0.14792982 0.1740175 0.06550877 0.61261404
2 0.9170557 0.07966450      0.02981424 0.16787926 0.7375232 0.05250774 0.04212693
3 0.9360000 0.02632228      0.04643038 0.05916456 0.1440759 0.76986076 0.02684810
4 0.9059858 0.08773273      0.03546809 0.70987234 0.2579574 0.01428369 0.01790071
  PropCat_5  PropCat_6  PropCat_7  PropCat_8  PropCat_9 PropCat_14  PropCat_15
1 0.7435088 0.03322673 0.03801754 0.07456140 0.06415549 0.06152632 0.010797828
2 0.5510526 0.09210765 0.09579567 0.08455160 0.07847690 0.05015480 0.038614920
3 0.1138481 0.06578107 0.01011392 0.01076737 0.04920576 0.76322785 0.007730561
4 0.3187376 0.14328286 0.17190780 0.12392955 0.08814838 0.01341135 0.031526511
```

Cluster 1: Customers in this category tend to buy beauty products from sub-popular price codelist under no promotion.
Cluster 2: Customers in this category tend to buy beauty products from popular soap price codelist under no promotion.
Cluster 3: : Customers in this cluster buy any economy/carbolic products that are not on promotion.
Cluster 4: Customers in this cluster buy any premium soaps that are not on promotion.

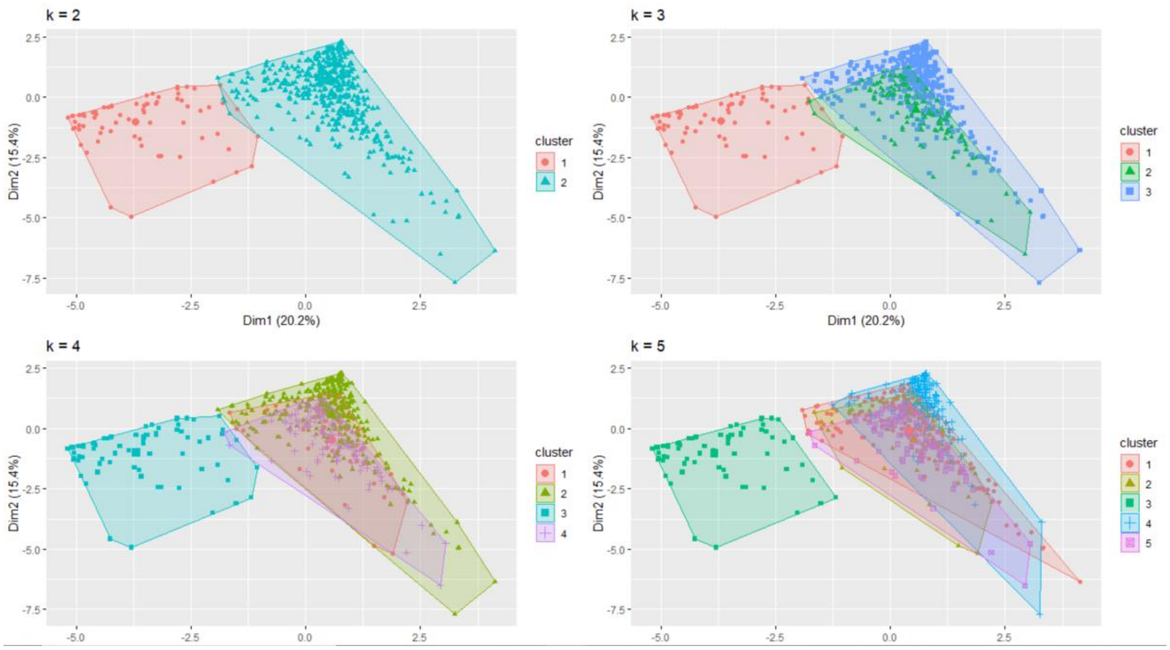
- Cluster = 5. AS we can see, clusters doesn't seem to be evenly distributed.



```
K-means clustering with 5 clusters of sizes 180, 58, 74, 168, 120

Cluster means:
  Pur_Promo Pur_Promo6 Pur_Other_Promo    Pr_Cat1   Pr_Cat2    Pr_Cat3    Pr_Cat4
1 0.8976556 0.11530068      0.02544444 0.20522222 0.6736667 0.07870556 0.04243889
2 0.8738621 0.14162746      0.03168966 0.14582759 0.1726034 0.07837931 0.60324138
3 0.9373784 0.02244824      0.04763514 0.05552703 0.1258919 0.79344595 0.02510811
4 0.9351607 0.04545941      0.03451786 0.16151190 0.7715298 0.02790476 0.03908929
5 0.9089000 0.08212144      0.03629167 0.75651667 0.2140750 0.01175833 0.01765833
  PropCat_5  PropCat_6   PropCat_7   PropCat_8  PropCat_9 PropCat_14  PropCat_15
1 0.2855833 0.16035015 0.163488889 0.144041033 0.09104031 0.07462222 0.064867725
2 0.7317414 0.03400334 0.037775862 0.074474174 0.06304936 0.07379310 0.010611658
3 0.1018514 0.05871629 0.008391892 0.008228664 0.05200053 0.78689189 0.005003218
4 0.7889345 0.03300476 0.031648810 0.031169729 0.07713294 0.02710119 0.015313209
5 0.3365167 0.13599382 0.171583333 0.114937759 0.07058824 0.01165000 0.024246032
```

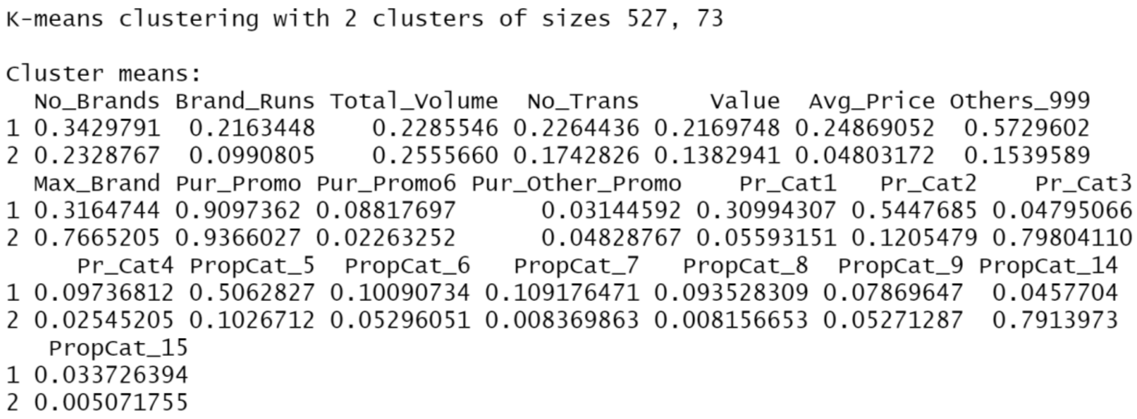
Clustering with k=5 seems similar to that of k=4 with little differences. K=5 does not explain any new customer segment different from 4.

Clusters plot of Customer Customer basis-for-purchase



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

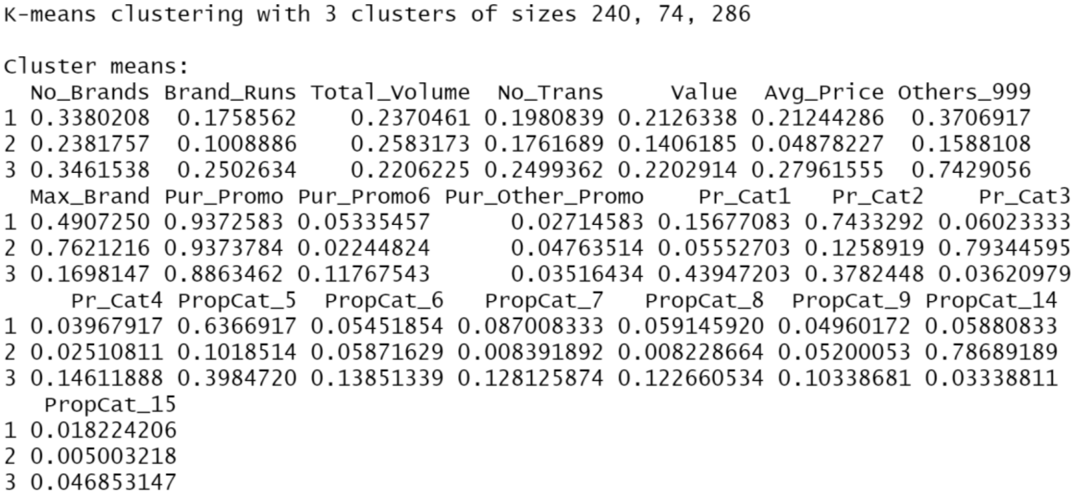
3. Clustering based on both basis-for-purchase and purchase behavior.
   - Cluster size = 2. Here, clusters sizes are not at all evenly distributed. Cluster 2 has very few customers as compared to cluster 1


```
K-means clustering with 2 clusters of sizes 527, 73

Cluster means:
  No_Brands Brand_Runs Total_Volume  No_Trans     Value  Avg_Price Others_999
1 0.3429791  0.2163448    0.2285546 0.2264436 0.2169748 0.24869052  0.5729602
2 0.2328767  0.0990805    0.2555660 0.1742826 0.1382941 0.04803172  0.1539589
  Max_Brand Pur_Promo Pur_Promo6 Pur_Other_Promo    Pr_Cat1   Pr_Cat2    Pr_Cat3
1 0.3164744 0.9097362 0.08817697      0.03144592 0.30994307 0.5447685 0.04795066
2 0.7665205 0.9366027 0.02263252      0.04828767 0.05593151 0.1205479 0.79804110
     Pr_Cat4 PropCat_5  PropCat_6   PropCat_7   PropCat_8  PropCat_9 PropCat_14
1 0.09736812 0.5062827 0.10090734 0.109176471 0.093528309 0.07869647  0.0457704
2 0.02545205 0.1026712 0.05296051 0.008369863 0.008156653 0.05271287  0.7913973
   PropCat_15
1 0.033726394
2 0.005071755
```

Cluster 1: Customers in this category tend to buy beauty products from any popular price codelist under no promotion. These customers are not that brand loyal are more likely to buy products from other brands. The average price & brand runs of the customers in this cluster are higher than that of cluster 2.

Cluster 2: Customers in this category are brand loyal and tend to buy any economy/carbolic products under no promotion. Their brand runs & average price are comparatively lesser to cluster 1.

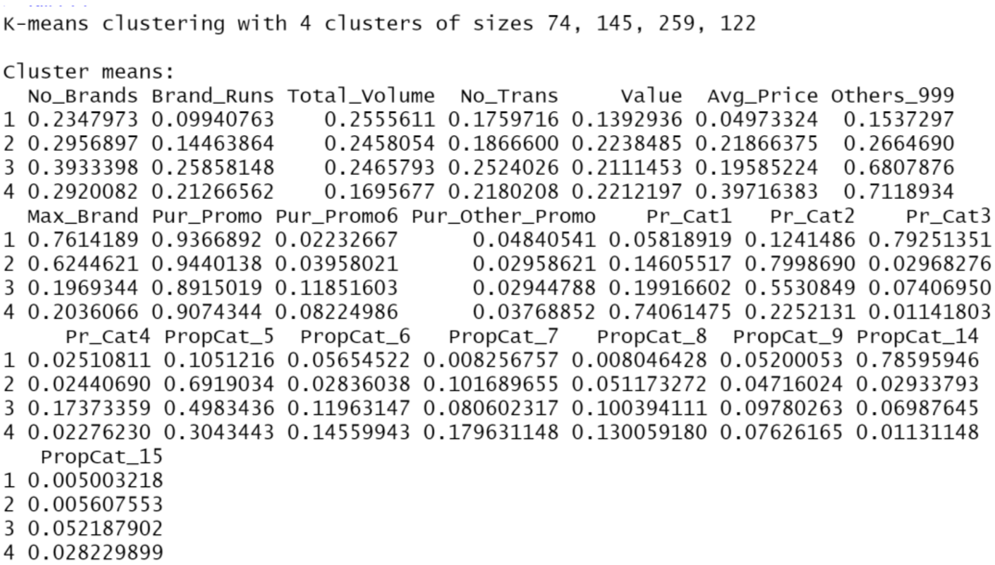
- Cluster size = 3. Here, clusters sizes are again not evenly distributed.



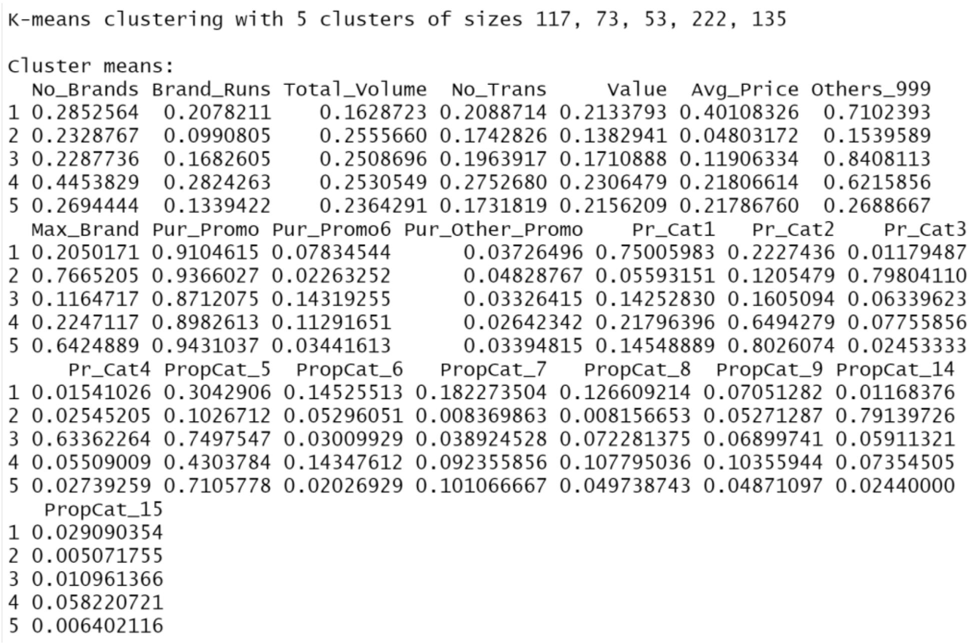
```
K-means clustering with 3 clusters of sizes 240, 74, 286

Cluster means:
  No_Brands Brand_Runs Total_Volume  No_Trans     Value  Avg_Price Others_999
1 0.3380208  0.1758562    0.2370461 0.1980839 0.2126338 0.21244286  0.3706917
2 0.2381757  0.1008886    0.2583173 0.1761689 0.1406185 0.04878227  0.1588108
3 0.3461538  0.2502634    0.2206225 0.2499362 0.2202914 0.27961555  0.7429056
  Max_Brand Pur_Promo Pur_Promo6 Pur_Other_Promo    Pr_Cat1   Pr_Cat2    Pr_Cat3
1 0.4907250 0.9372583 0.05335457      0.02714583 0.15677083 0.7433292 0.06023333
2 0.7621216 0.9373784 0.02244824      0.04763514 0.05552703 0.1258919 0.79344595
3 0.1698147 0.8863462 0.11767543      0.03516434 0.43947203 0.3782448 0.03620979
     Pr_Cat4 PropCat_5  PropCat_6   PropCat_7   PropCat_8  PropCat_9 PropCat_14
1 0.03967917 0.6366917 0.05451854 0.087008333 0.059145920 0.04960172 0.05880833
2 0.02510811 0.1018514 0.05871629 0.008391892 0.008228664 0.05200053 0.78689189
3 0.14611888 0.3984720 0.13851339 0.128125874 0.122660534 0.10338681 0.03338811
   PropCat_15
1 0.018224206
2 0.005003218
3 0.046853147
```

Cluster 1: Customers in this category are moderately brand loyal & tend to buy beauty products from any popular soap price codelist under no promotion.
Cluster 2: Customers in this category are highly brand loyal and tend to buy any economy/carbolic products under no promotion. Their brand runs, no of brands, No of transactions, value & average price are least as compared to the other 2 clusters. However, the value of Total Volume is marginally higher.
Cluster 3: Customers in this category are not at all brand loyal & mostly tending towards buying from other brands. Looks like this group buys beauty products from any premium soap price codelist under no promotion. Their brand runs, no of brands, No of transactions, value & average price are the highest among all the clusters.

- Output of clusters with size 4 & 5 are attached below.



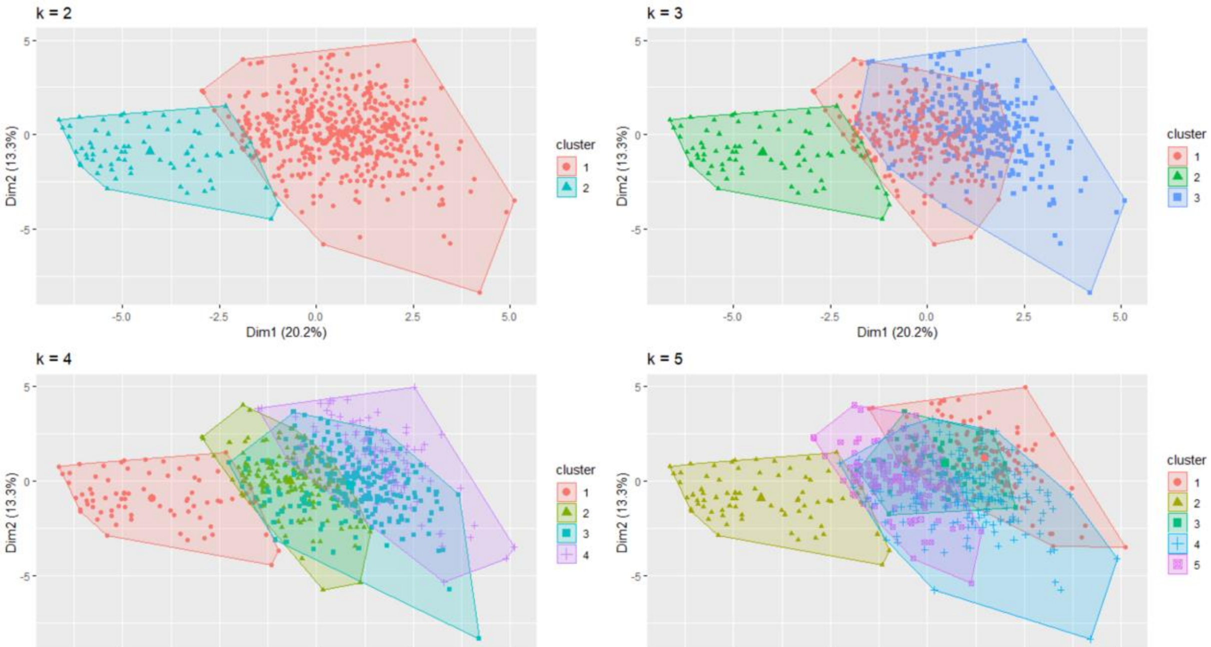
```
K-means clustering with 4 clusters of sizes 74, 145, 259, 122

Cluster means:
  No_Brands Brand_Runs Total_Volume  No_Trans     Value  Avg_Price Others_999
1 0.2347973 0.09940763    0.2555611 0.1759716 0.1392936 0.04973324  0.1537297
2 0.2956897 0.14463864    0.2458054 0.1866600 0.2238485 0.21866375  0.2664690
3 0.3933398 0.25858148    0.2465793 0.2524026 0.2111453 0.19585224  0.6807876
4 0.2920082 0.21266562    0.1695677 0.2180208 0.2212197 0.39716383  0.7118934
  Max_Brand Pur_Promo Pur_Promo6 Pur_Other_Promo    Pr_Cat1   Pr_Cat2    Pr_Cat3
1 0.7614189 0.9366892 0.02232667      0.04840541 0.05818919 0.1241486 0.79251351
2 0.6244621 0.9440138 0.03958021      0.02958621 0.14605517 0.7998690 0.02968276
3 0.1969344 0.8915019 0.11851603      0.02944788 0.19916602 0.5530849 0.07406950
4 0.2036066 0.9074344 0.08224986      0.03768852 0.74061475 0.2252131 0.01141803
     Pr_Cat4 PropCat_5  PropCat_6   PropCat_7   PropCat_8  PropCat_9 PropCat_14
1 0.02510811 0.1051216 0.05654522 0.008256757 0.008046428 0.05200053 0.78595946
2 0.02440690 0.6919034 0.02836038 0.101689655 0.051173272 0.04716024 0.02933793
3 0.17373359 0.4983436 0.11963147 0.080602317 0.100394111 0.09780263 0.06987645
4 0.02276230 0.3043443 0.14559943 0.179631148 0.130059180 0.07626165 0.01131148
   PropCat_15
1 0.005003218
2 0.005607553
3 0.052187902
4 0.028229899
```



```
K-means clustering with 5 clusters of sizes 117, 73, 53, 222, 135

Cluster means:
  No_Brands Brand_Runs Total_Volume  No_Trans     Value  Avg_Price Others_999
1 0.2852564  0.2078211    0.1628723 0.2088714 0.2133793 0.40108326  0.7102393
2 0.2328767  0.0990805    0.2555660 0.1742826 0.1382941 0.04803172  0.1539589
3 0.2287736  0.1682605    0.2508696 0.1963917 0.1710888 0.11906334  0.8408113
4 0.4453829  0.2824263    0.2530549 0.2752680 0.2306479 0.21806614  0.6215856
5 0.2694444  0.1339422    0.2364291 0.1731819 0.2156209 0.21786760  0.2688667
  Max_Brand Pur_Promo Pur_Promo6 Pur_Other_Promo    Pr_Cat1   Pr_Cat2    Pr_Cat3
1 0.2050171 0.9104615 0.07834544      0.03726496 0.75005983 0.2227436 0.01179487
2 0.7665205 0.9366027 0.02263252      0.04828767 0.05593151 0.1205479 0.79804110
3 0.1164717 0.8712075 0.14319255      0.03326415 0.14252830 0.1605094 0.06339623
4 0.2247117 0.8982613 0.11291651      0.02642342 0.21796396 0.6494279 0.07755856
5 0.6424889 0.9431037 0.03441613      0.03394815 0.14548889 0.8026074 0.02453333
     Pr_Cat4 PropCat_5  PropCat_6   PropCat_7   PropCat_8  PropCat_9 PropCat_14
1 0.01541026 0.3042906 0.14525513 0.182273504 0.126609214 0.07051282 0.01168376
2 0.02545205 0.1026712 0.05296051 0.008369863 0.008156653 0.05271287 0.79139726
3 0.63362264 0.7497547 0.03009929 0.038924528 0.072281375 0.06899741 0.05911321
4 0.05509009 0.4303784 0.14347612 0.092355856 0.107795036 0.10355944 0.07354505
5 0.02739259 0.7105778 0.02026929 0.101066667 0.049738743 0.04871097 0.02440000
   PropCat_15
1 0.029090354
2 0.005071755
3 0.010961366
4 0.058220721
5 0.006402116
```

Clustering with k=4 & k=5 seem more or less similar to clusters of k=3. K=4 & k=5 does not explain any new customer segment different from that of 3.

Clusters plot of Customer basis-for-purchase and purchase behavior

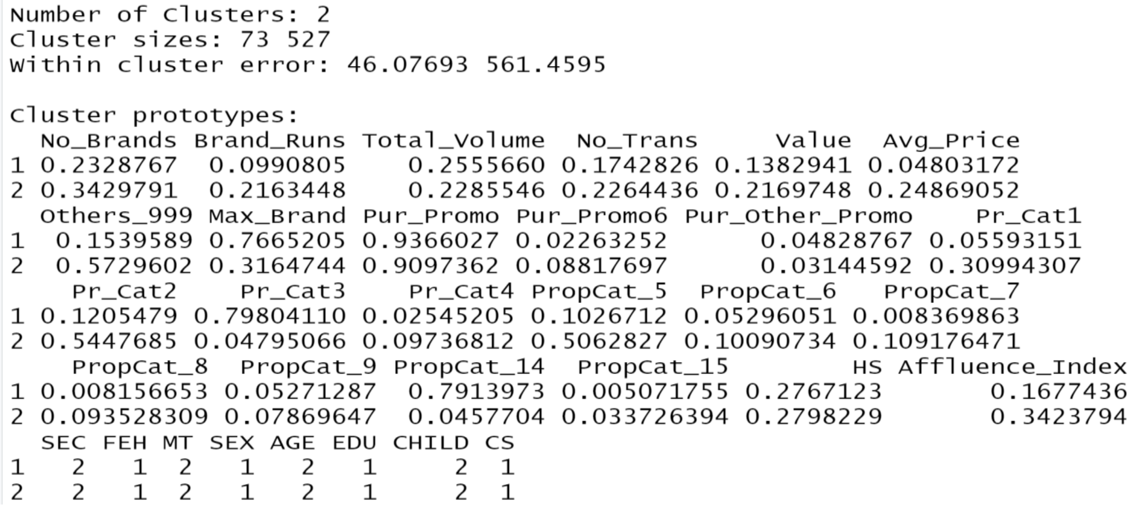


Q] How should k be chosen?
Silhouette measure could be considered in choosing the optimal value of k. Silhoutte measure basically tells us how similar the object is to its own cluster and how dissimilar or far it is from its neighbouring clusters.

Silhouette coefficient is calculated using the mean intra cluster distance and the mean nearest cluster difference. Higher Silhouette measure means that the object is close to its own cluster and far from its neighbouring clusters and vice versa.

Q2. (a) Select what you think is the best segmentation - explain why you think this is the \best".
(b) 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.)

- Clustering based on demographic, brand loyalty & basis for purchase – I did consider cluster sizes from 2 to 5. However, cluster 2 had the maximum Silhouette measure of 0.46 as compared to other clusters. Thus cluster 2 seems to be the best cluster segmentation.



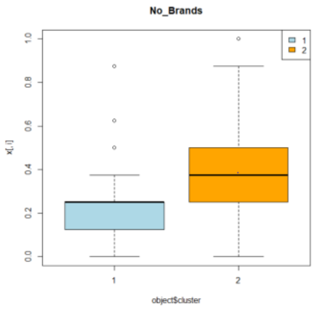
```
Number of Clusters: 2
Cluster sizes: 73 527
Within cluster error: 46.07693 561.4595

Cluster prototypes:
  No_Brands Brand_Runs Total_Volume  No_Trans     Value  Avg_Price
1 0.2328767  0.0990805    0.2555660 0.1742826 0.1382941 0.04803172
2 0.3429791  0.2163448    0.2285546 0.2264436 0.2169748 0.24869052
  Others_999 Max_Brand Pur_Promo Pur_Promo6 Pur_Other_Promo    Pr_Cat1
1  0.1539589 0.7665205 0.9366027 0.02263252      0.04828767 0.05593151
2  0.5729602 0.3164744 0.9097362 0.08817697      0.03144592 0.30994307
    Pr_Cat2    Pr_Cat3    Pr_Cat4 PropCat_5  PropCat_6   PropCat_7
1 0.1205479 0.79804110 0.02545205 0.1026712 0.05296051 0.008369863
2 0.5447685 0.04795066 0.09736812 0.5062827 0.10090734 0.109176471
    PropCat_8  PropCat_9 PropCat_14  PropCat_15        HS Affluence_Index
1 0.008156653 0.05271287  0.7913973 0.005071755 0.2767123       0.1677436
2 0.093528309 0.07869647  0.0457704 0.033726394 0.2798229       0.3423794
  SEC FEH MT SEX AGE EDU CHILD CS
1   2   1  2   1   2   1     2  1
2   2   1  2   1   2   1     2  1
```

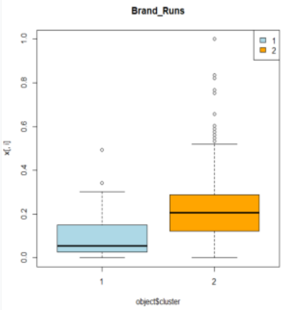
The clusters are segmented mainly based on Price categories & proposition categories. Another important variable is brand loyalty. Both the clusters seem to have same values of demographics which can be seen in below graphs. Demographics does not seem to play a vital role while grouping the customers.

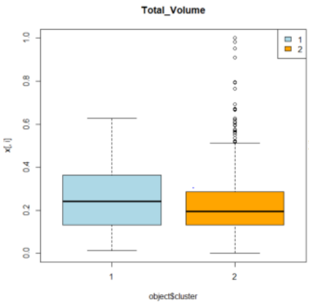
Cluster 1: Customers in this group are highly brand loyal and tend buy any carbolic products from price code list of any economy/carbolic under no promotion. Affluence index of customers in this group are less than that of cluster 2.

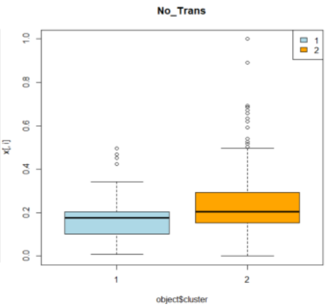
Cluster 2: Customers in this category are not that brand loyal & tend to buy products from other brands. Customers in this cluster seem to buy any beauty products from price code list of any premium soaps & mostly popular soaps under no promotion. People in this group have better affluence index as compared to cluster 1.

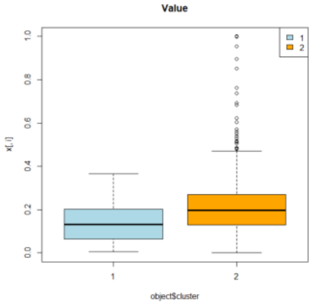
<u>Plots of demographic, brand loyalty and basis-for-purchase characteristics of cluster size = 2</u>

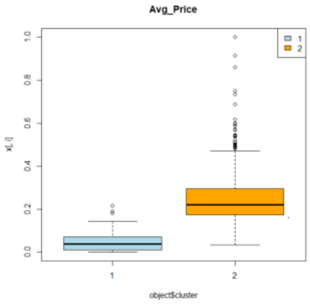

No_Brands

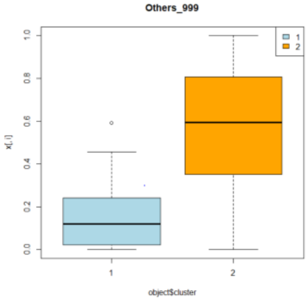

Brand_Runs

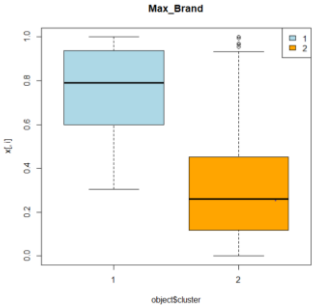

Total_Volume

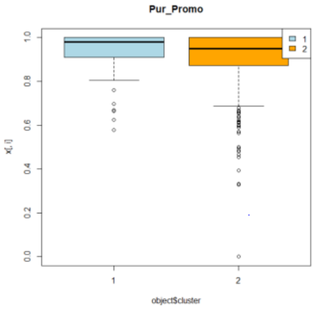

No_Trans

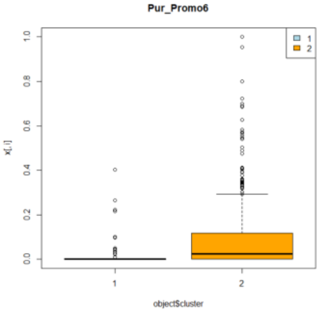

Value

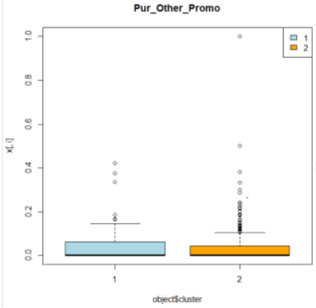

Avg_Price

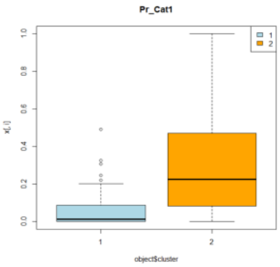

Others_999

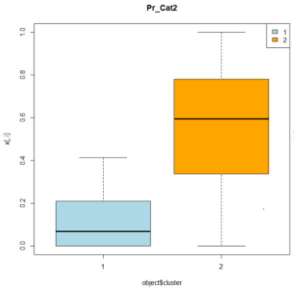

Max_Brand

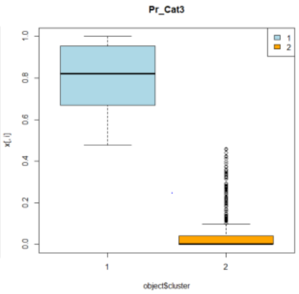

Pur_Promo

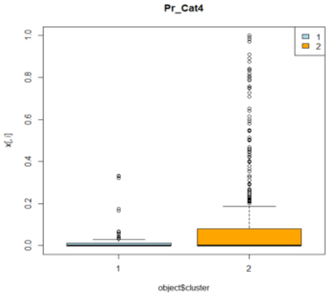

Pur_Promo6

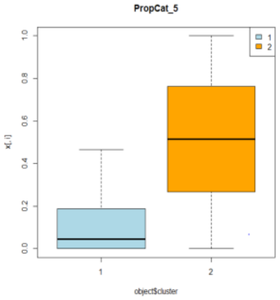

Pur_Other_Promo

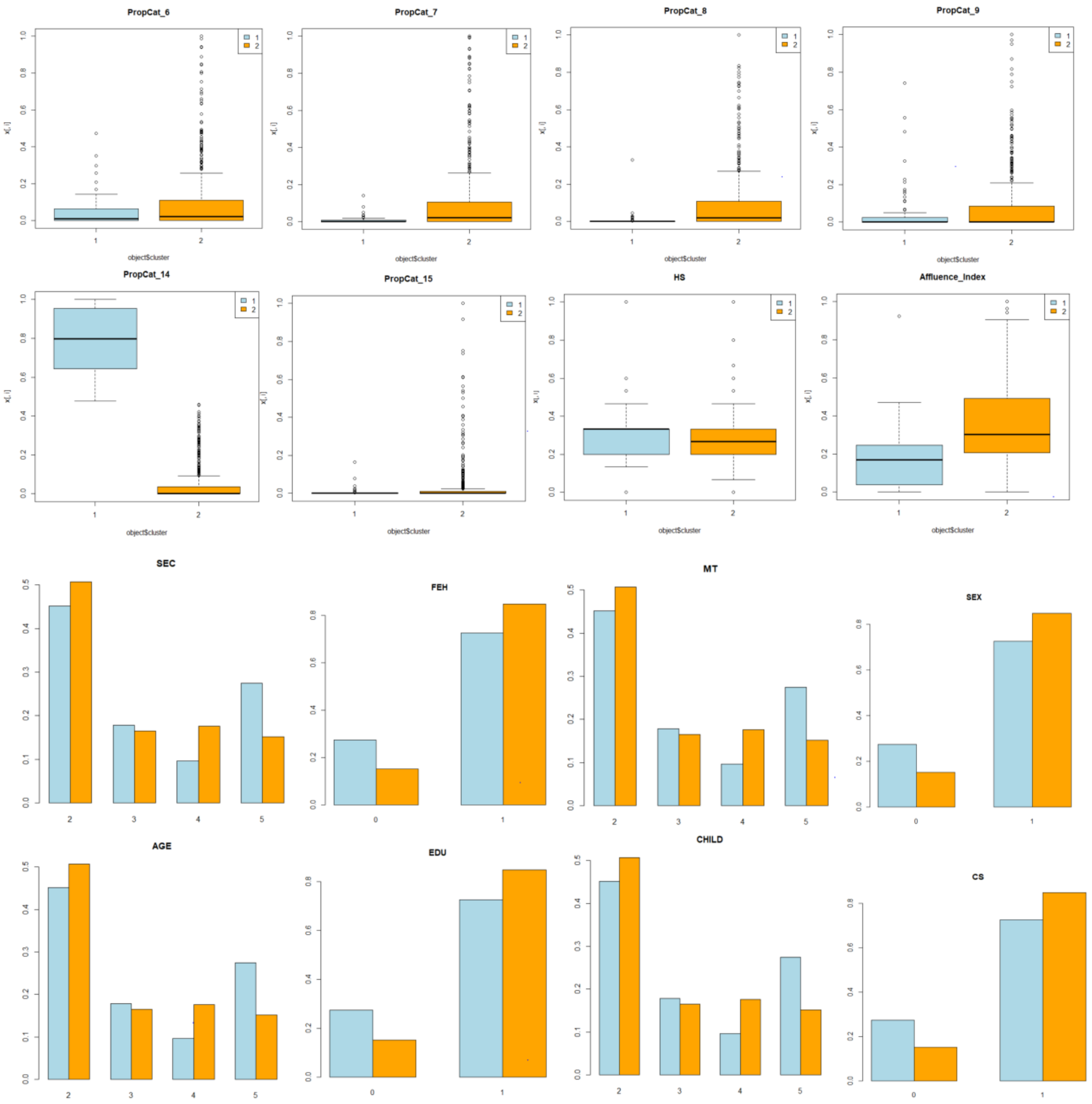

Pr_Cat1


Pr_Cat2


Pr_Cat3

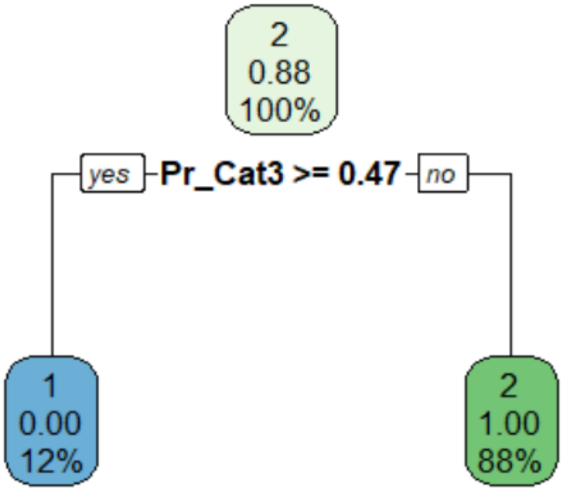

Pr_Cat4

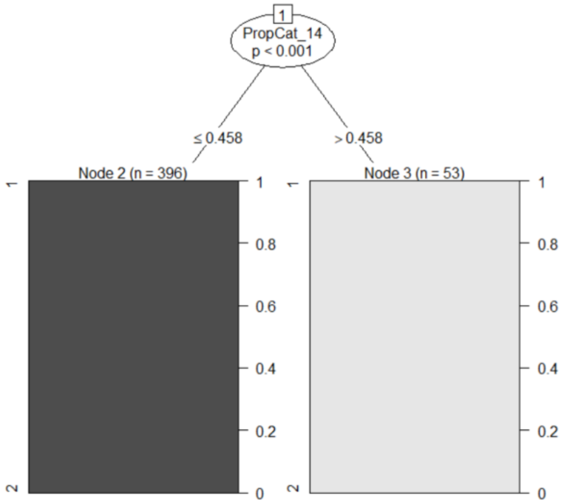

PropCat_5



Q3. For the best segmentation, obtain a description of the clusters using a decision tree how effective is the tree in identifying the different clusters? Does the tree help in explaining/interpreting different clusters? (explain why/why not)

Below are the snapshots of two decision trees plotted using packages party & rpart in R. Below decision trees do given an idea that customers with Proposition category of any carbolic<=0.458 & Price Codelist of any economy/carbolic <= 0.47 fall under one cluster group and vice versa for the other cluster. However, the decision trees do not explain about other variables like brand loyalty. This decision tree model gives an accuracy of 100%.


2
0.88
100%
yes
Pr_Cat3 >= 0.47
no
1
0.00
12%
2
1.00
88%


1
PropCat_14
p < 0.001
≤ 0.458
> 0.458
Node 2 (n = 396)
Node 3 (n = 53)
1
2
1
0.8
0.6
0.4
0.2
0
1
2
1
0.8
0.6
0.4
0.2
0