IDS 572 Market Segmentation

Britney Scott, Abdullah Saka

4/12/2020

Introduction

CRISA, a well-known market research company, tracks about 60-70 brands within 30 product categories in order to best determine marketing strategies for their clients. To do this, CRISA conducts household panels in India, and has data covering about 50,000 urban and 25,000 rural Indian households. Optimal households for research are selected using stratified sampling, and in urban areas data captures information from 80% of the market.

CRISA uses this data to provide market research services to their clients. These clients include two main groups: * Advertising Agencies: These agencies receive monthly data from CRISA. They utilize this database to make recommendations and decisions for their own clients' marketing and advertising strategies. * Goods Manufacturers: This group of clients are able to monitor changes in their market share with CRISA's data.

For a long time now, CRISA has implemented segmentation algorithms which cluster consumers based on their demographic characteristics. There is now a demand for CRISA to segment the market further in order to better capture brand loyalty and the consumer purchasing process. Two sets of variables which CRISA wants to implement clustering on are: * Purchase Behavior: This includes how often consumers purchase, volume purchased, use of discounts, and other variables related to the purchase process * Basis of Purchase: This set includes price and selling proposition

The objective of this additional clustering is to gain insight on purchase behaviors and brand loyalty, and identify the most important attributes which help to identify this behavior. This way, CRISA's clients can better use the information provided to make decisions. The goal is to develop unique strategies targeting different segments, in order to better reach individuals in each cluster and increase brand loyalty of consumers. This is more cost-effective than implementing a general marketing strategy which may only appeal to a fraction of consumers.

Data Exploration and Cleaning

We converted several categorical variables into dummy variables by applying one hot encoding to understand difference between clusters such as Mother Tongue, Gender, Children and Education.

K-Means Clustering

Purchase Behavior Variables

Firstly, we built a clustering model by only using variables are related to 'Purchase behavior'. Purchasing behavior can be identified based on these attributes: the number of brands purchased, brand loyalty, the number of transactions, the number of runs purchasing same brand, volume of product and average price.

We built clustering models by changing some parameters such as centers, nstart and iter.max in Kmeans model. It is shown that when we changed parameters of k-means model, there is no significant difference between models. Our baseline model is created by selecting 25 random sets and using 12 iterations.

KMeans3	SumOfSquare
Cluster 1	1242.74
Cluster 2	1141.76
Cluster 3	1585.15
Total Within	3969.64
Between	2020.36
Total	5990

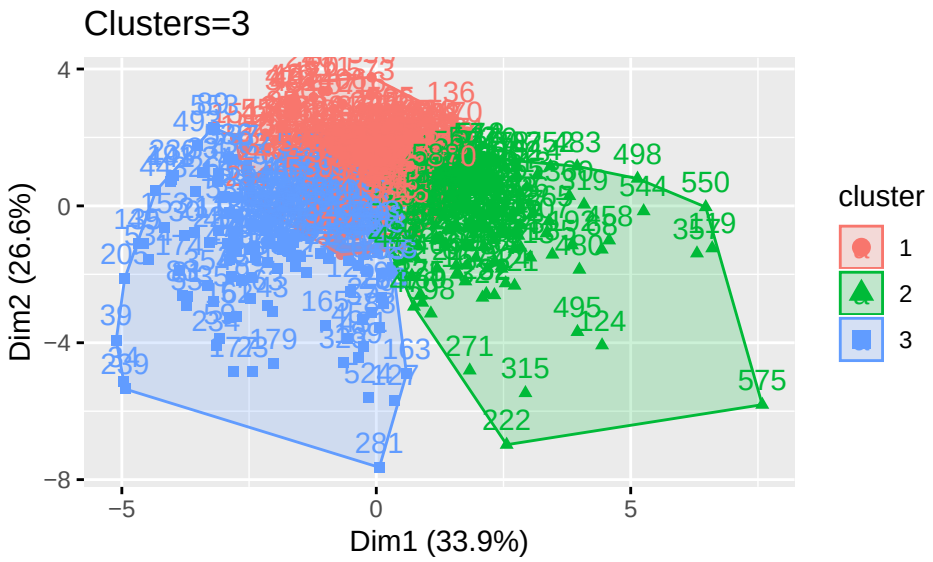
Clusters	Size
Cluster 1	259
Cluster 2	166
Cluster 3	175

Then, we checked characteristics of clusters to understand differences between clusters. Households size of cluster 1 is higher than other clusters. This increases the consumption such as number of brands, transactions and volume. Moreover, households in cluster 3 have higher brand loyalty than other clusters (generally consume products of same brand) and their affluence index is lower than others. Order of educational level is cluster 1 > cluster 2 > cluster 3 meaning people are in cluster 1 are more educated than others.

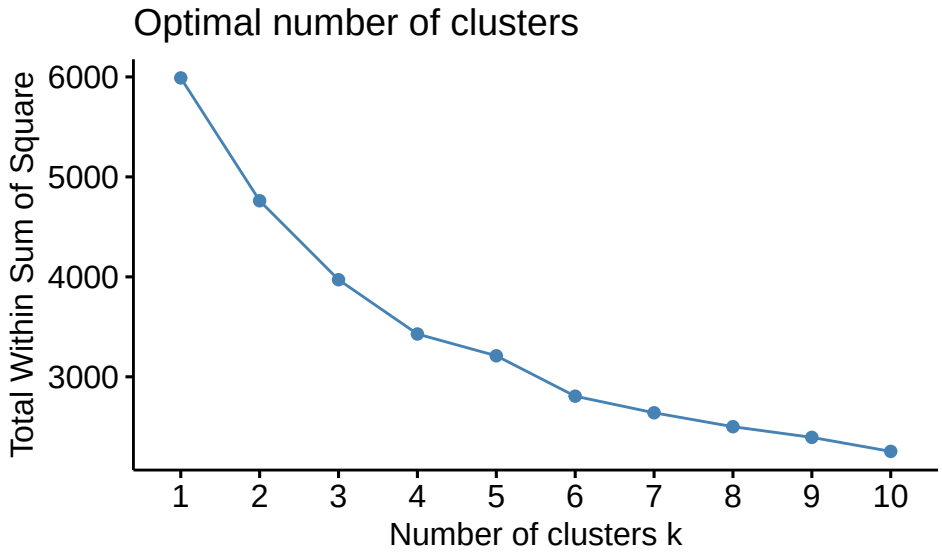
In the light of this information, marketing strategies must vary from cluster to cluster. For example, if you aim to reach people who have lower economic class and higher brand loyalty, you should consider households in cluster 3 and shape your marketing planning regarding patterns of cluster 3.

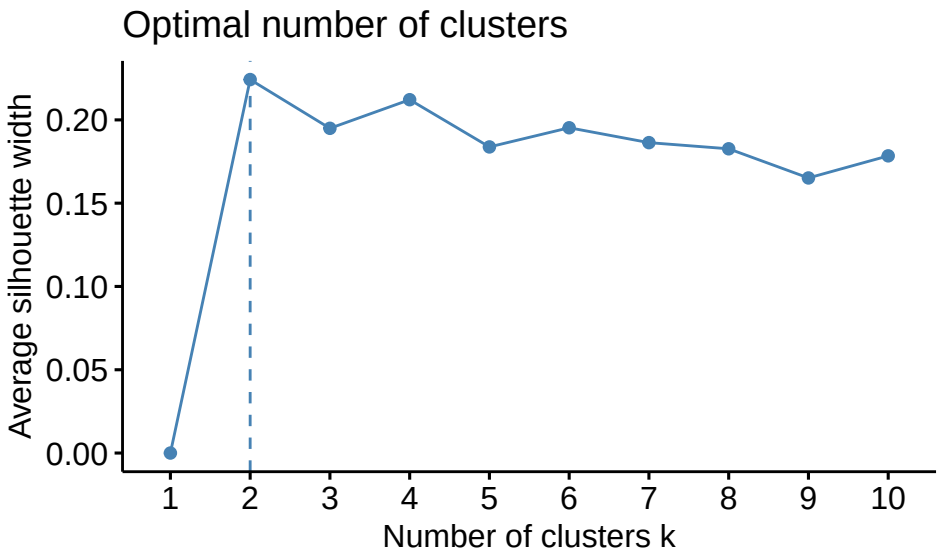
Cluster	SEC	HS	Affluence	maxBr	No of Brands	No of Trans	Brand Runs	Volume
1	2.339768	3.474903	16.60618	0.2193920	3.200772	23.91120	13.509652	7778.097
2	2.409639	5.108434	20.99398	0.2350520	5.138554	50.01205	27.180723	16856.536
3	2.822857	4.382857	13.86286	0.7250765	2.857143	23.98286	8.228571	13349.429

Next, we drew cluster plot by using fviz package. It can be clearly seen that this clustering model is not optimal since clusters are heavily overlapping. This model was not able to segment households successfully in the intersection area.



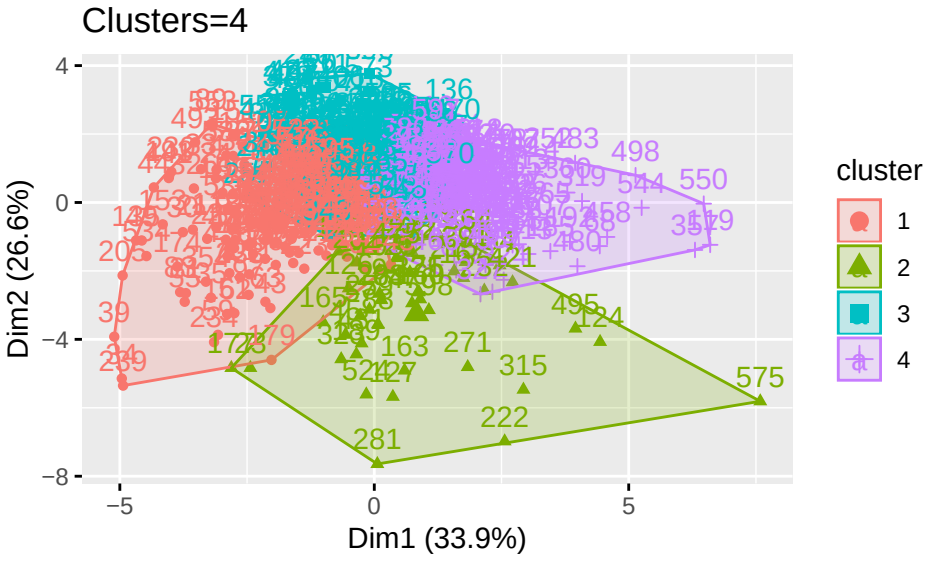
Then, we applied both the elbow and silhouette method to decide the number of clusters in the model. According to elbow method, we can say that best k value is 6. Alternatively, the silhouette method determined the number of clusters as 4. As a result, we selected the number of clusters as 4 since if we increase the number of clusters, the scope of the business cannot be easily managed by marketing teams.





This plot shows the clusters when we apply 4 different clusters regarding elbow and silhoutte model. This model is slightly better than previous model, but still not best (clusters are overlapping).

KMeans3	SumOfSquare
Cluster 1	1170.81
Cluster 2	502.28
Cluster 3	879.76
Cluster 4	875.11
Total Within	3427.96
Between	2562.04
Total	5990



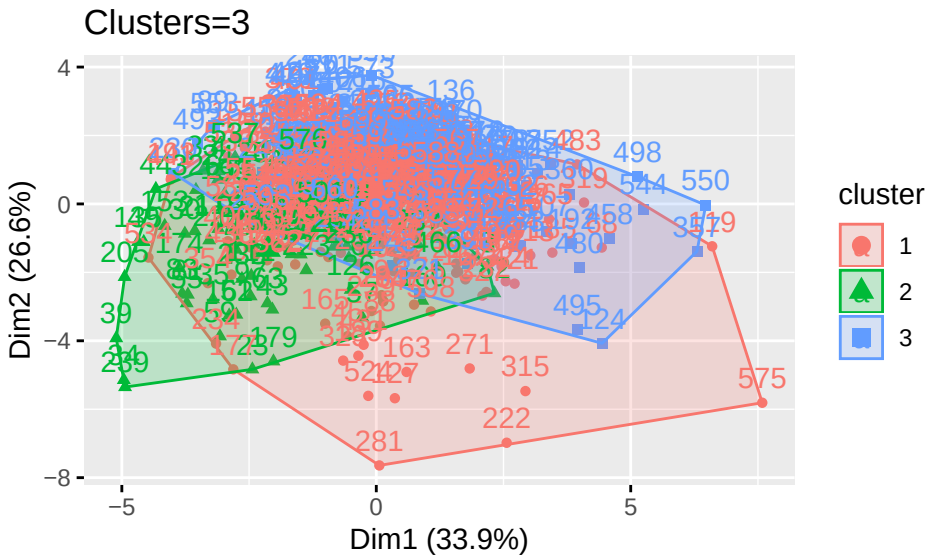
Basis for Purchase Variables

Secondly, we applied k-means clustering by using different variables. Basis of purchase variables obtains percent of volume purchased not on promotion, on promo code 6 and other than 6, proposition of beauty, health and baby products.

KMeans3	SumOfSquare
Cluster 1	1693
Cluster 2	229.32
Cluster 3	2500.69
Total Within	4423.01
Between	1566.99
Total	5990

Clusters	Size
Cluster 1	335
Cluster 2	73
Cluster 3	192

The graph below indicates that basis for purchase variables are not sufficient to segment households in consumption of consumer goods using 3 clusters. Clusters overlapped and so variance between clusters is small.

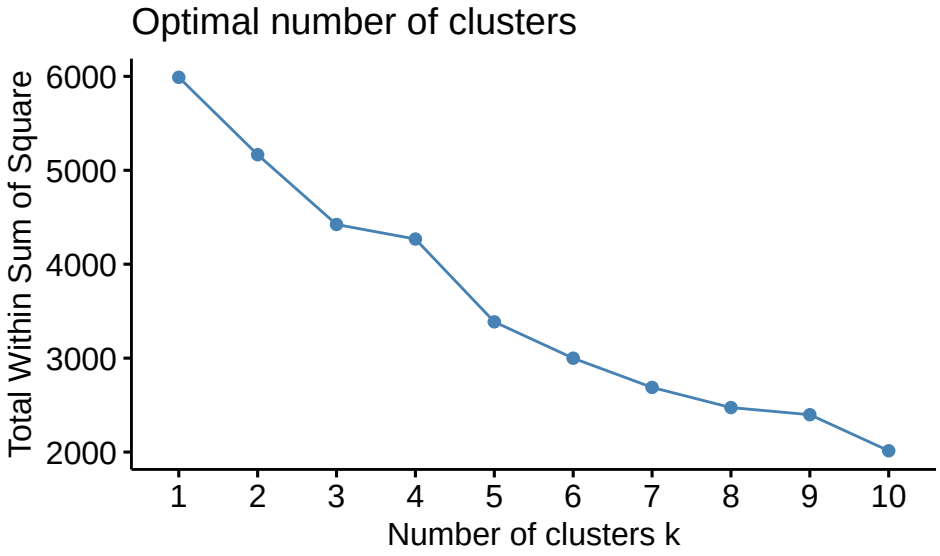


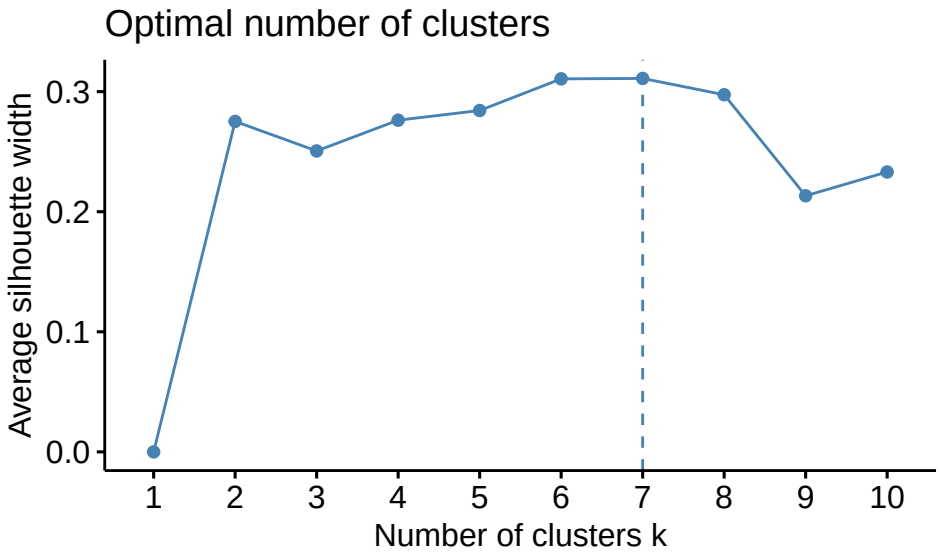
Then, we tried to learn behavior differences between clusters. Households in cluster 2 have higher brand loyalty (77%) than other clusters (meaning they generally consume products of same brand) and their affluence index is lowest among all others. Also, social economic status of cluster 3 is the lower than others (almost 3.4). People in cluster 3 are more educated than others. Overall, households in cluster 1 and 3 shows similar patterns in consumption.

Based on these insights, marketing strategies must vary from cluster to cluster. For instance, if you work on launching products which have medium price, you should target households in cluster 2 and build your marketing strategies matching with characteristics of cluster 2.

Cluster	SEC	HS	Affluence	maxBr	No of Brands	No of Trans	Brand Runs	Volume
1	2.591045	4.483582	17.14030	0.3670430	3.800000	31.60597	15.853731	13008.752
2	3.356164	4.150685	8.90411	0.7642848	2.904110	25.46575	8.506849	13279.315
3	2.015625	3.697917	19.89583	0.2290487	3.630208	32.52604	18.328125	9487.188

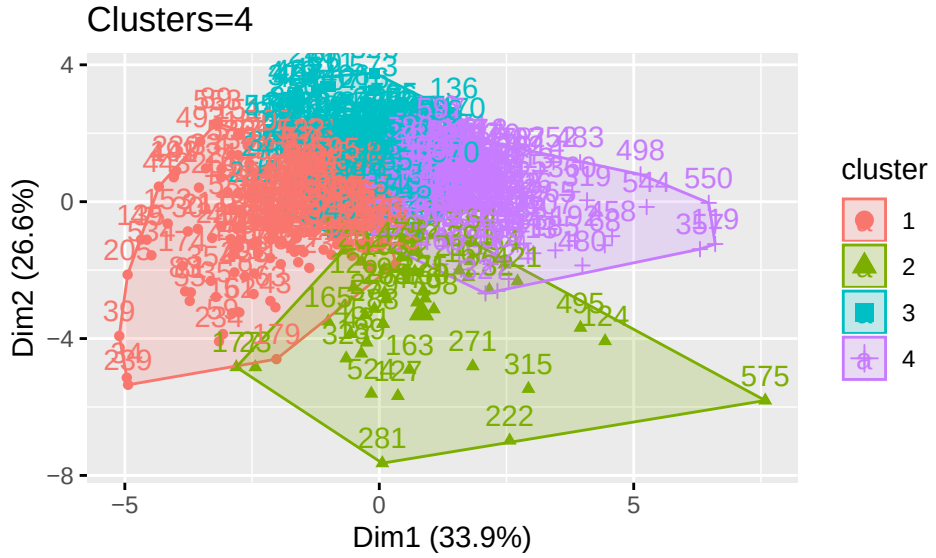
Then, we applied both elbow and silhouette methods to decide the number of clusters in the model. According to the elbow method, we can say that the best k value is 7. Besides of this, silhouette method determined the number of clusters as 8. As a result, instead of using 7 or 8 for the number of clusters, we determined the number of clusters as 4 since this help to manage marketing plans systematically.





This plot shows the clustering model when we ran 4 different clusters considering elbow and silhoutte above. This model is slightly better than previous model using only 3 clusters, but still not best because the clusters are overlapping.

KMeans3	SumOfSquare
Cluster 1	1170.81
Cluster 2	502.28
Cluster 3	879.76
Cluster 4	875.11
Total Within	3427.96
Between	2562.04
Total	5990



Combined Variables

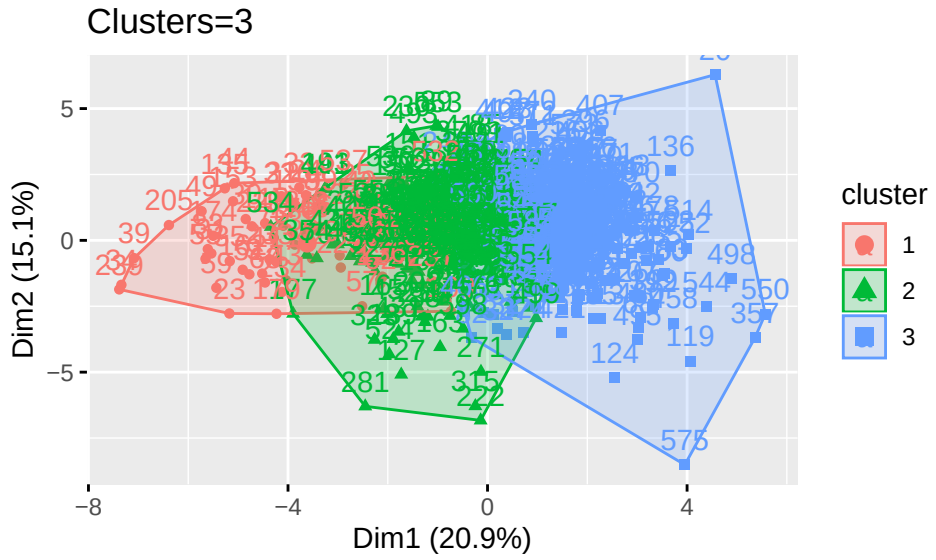
Lastly, we applied k-means clustering by using combined variables in part a and part b. Combined variables includes both purchase behavior and basis for purchase variables. The tables below give information about clustering model when we selected k as 3.

KMeans3	SumOfSquare
Cluster 1	878.75
Cluster 2	3440.63
Cluster 3	5068.85
Total Within	9388.23
Between	2591.77
Total	11980

Clusters
Cluster 1
Cluster 2
Cluster 3

Clusters

The graph bel ow indicates that combined variables are slightly better than previos models, yet it is still not sufficient to

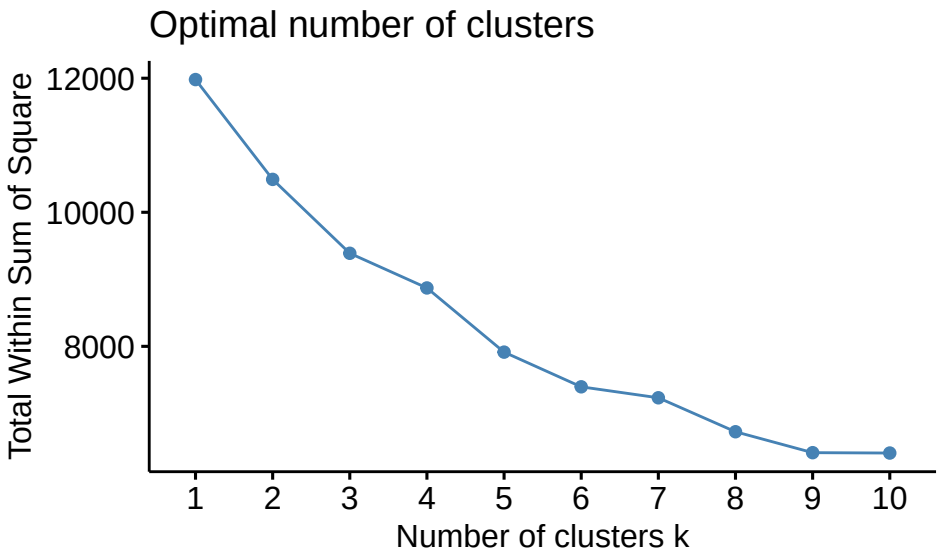


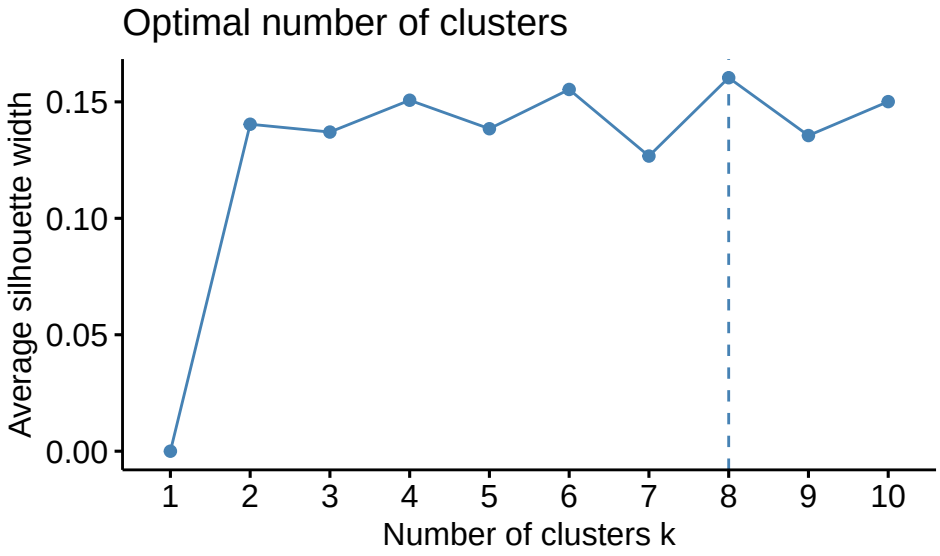
Then, we checked characteristics of clusters whether we see vital differences between clusters. Household sizes of clusters are similar to each other. Moreover, households in cluster 3 have higher brand loyalty than other clusters (hence transaction of brand runs is also highest) and their affluence index is lower than others. Order of educational level is cluster 1>cluster 2>cluster 3. People are in cluster 1 are more educated than others. Overall, households in cluster 1 and 3 shows similar pattern in consumption.

Based on this information, marketing strategies must vary from cluster to cluster. For example, if you launch premium products into the market, you have to reach people who have higher affluence index and education level which corresponds to cluster 1. For cluster 1, purchasing power is higher and not price oriented, but rather quality oriented.

Cluster	SEC	HS	Affluence	maxBr	No of Brands	No of Trans	Brand Runs	Volume
1	3.450704	4.140845	8.098591	0.7808103	2.732394	23.91549	7.492958	13055.14
2	2.588477	4.242798	15.395062	0.4660238	3.193416	24.25103	11.251029	12173.96
3	2.188811	4.160839	20.615385	0.1889798	4.237762	38.81469	21.625874	11411.45

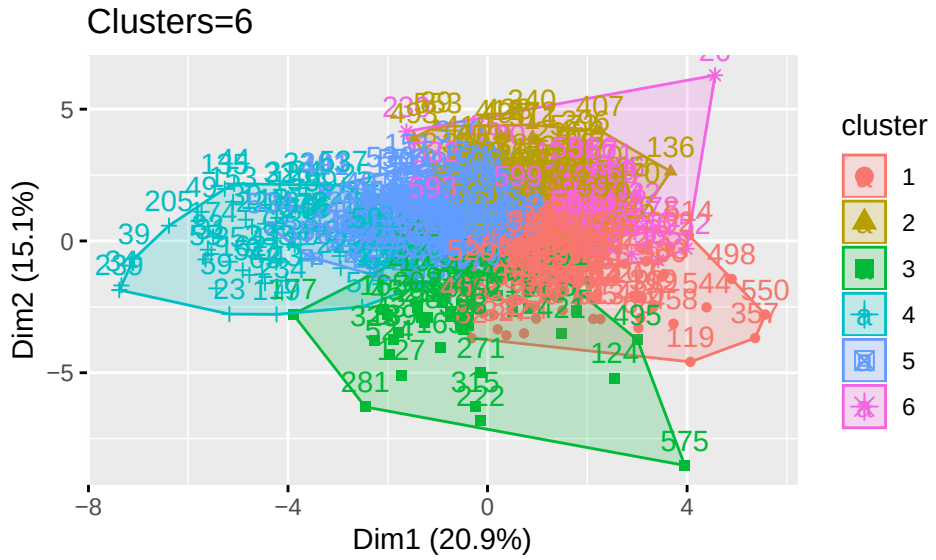
Then, we applied both elbow and silhouette method to do benchmarking among clusters. According to the elbow method, we can say that best k value is 6. Alternatively, silhouette method determined the number of clusters as 9. As a result, we determined the number of clusters as 6 due to the business implications of having too many clusters.





This plot shows the clustering model when we apply 6 different clusters regarding benchmarking analysis above. Within clusters SSQ is higher and between clusters SSQ is lower than the previous model. This model is slightly better than previous models, but still not best because clusters are overlapping.

KMeans3	SumOfSquare
Cluster 1	1506.44
Cluster 2	1526.87
Cluster 3	733.63
Cluster 4	790.09
Cluster 5	1743.87
Cluster 6	1095.17
Total Within	7396.07
Between	4583.93
Total	11980



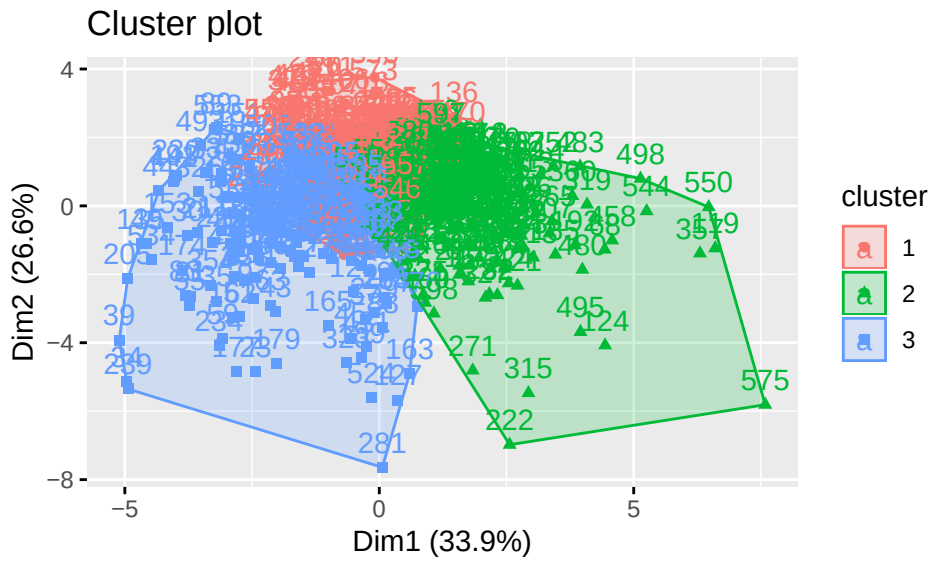
K-Medoids Clustering

Purchase Behavior Variables

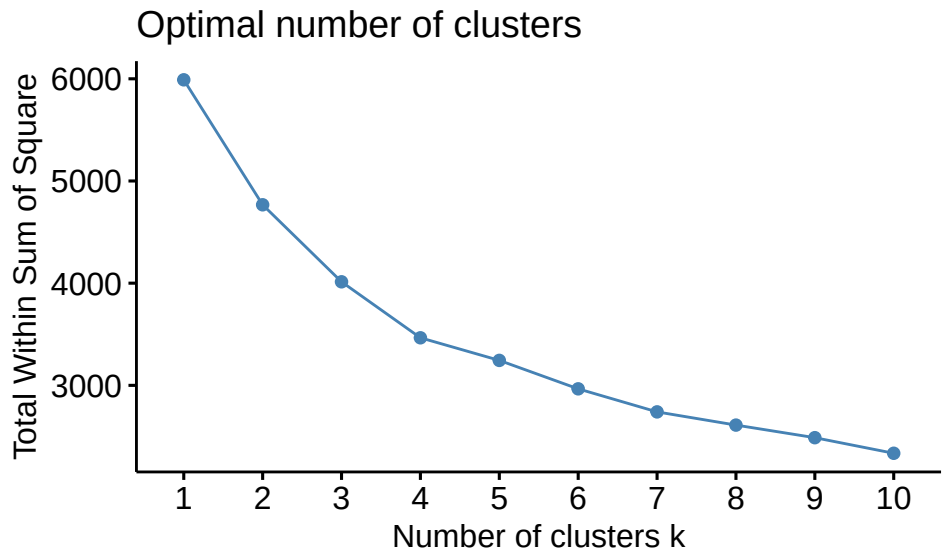
Several issues occur when using the k-means clustering. For example, the k-means algorithm is very sensitive to outliers and noise since the mean statistic is sensitive to these occurrences. In order to address these issues, we chose to explore k-medoids clustering as a second clustering technique.

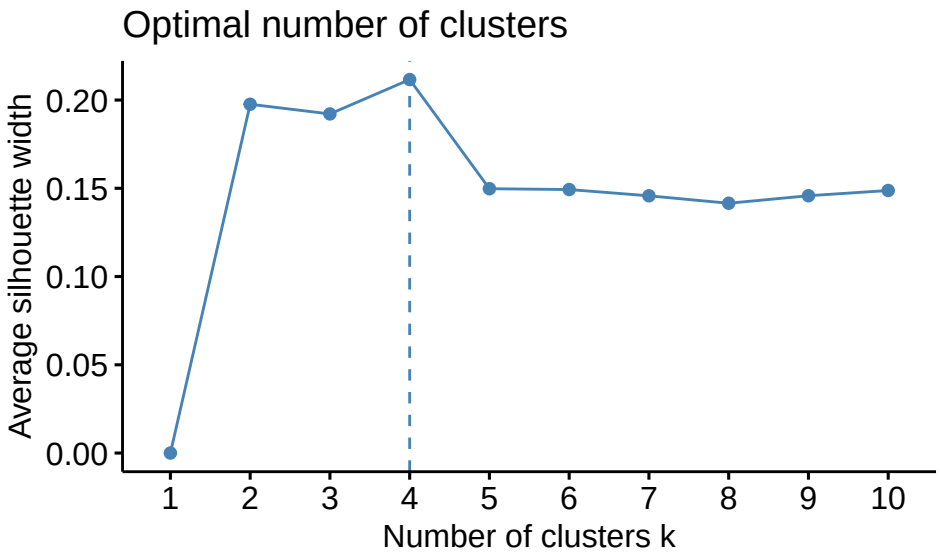
First, we will apply this clustering using the purchasing behavior. We started by using 3 clusters again. Here is a plot of the clusters using this choice.

size	max_diss	av_diss	diameter	separation
160	6.474649	2.044774	9.059644	0.4692172
242	8.784529	2.255817	11.178010	0.6147253
198	10.019243	2.553799	13.270996	0.4692172



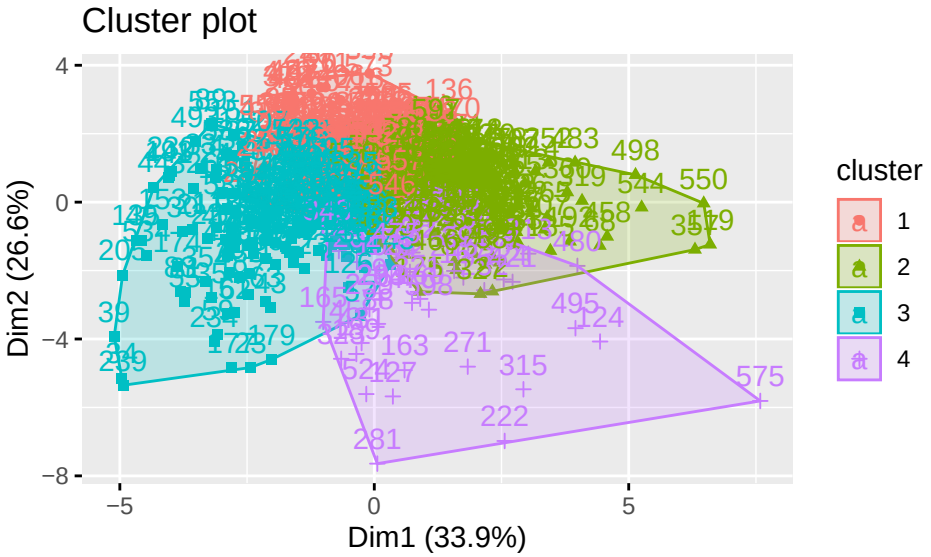
This plot does not vary much visually from the k-means algorithm. Once again, though, we wanted to verify what the optimal number of clusters actually is. Therefore, we chose to use both the elbow and silhouette methods to check for the optimal number of clusters. The elbow method does not demonstrate a clear elbow, but the silhouette method suggests 4 clusters to be optimal. Because of this, we will run the k-medoids again using 4 clusters instead of 3.





The graph below indicates that k medoids model works slightly better than k mean regarding basis for purchase variables, yet there is a still overlapping so distance between clusters is small.

size	max_diss	av_diss	diameter	separation
156	6.474649	2.005804	9.059644	0.4692172
212	7.349266	2.017348	9.071292	0.6147253
180	8.425903	2.324993	11.441277	0.4692172
52	7.794822	3.004423	12.126794	1.0068283



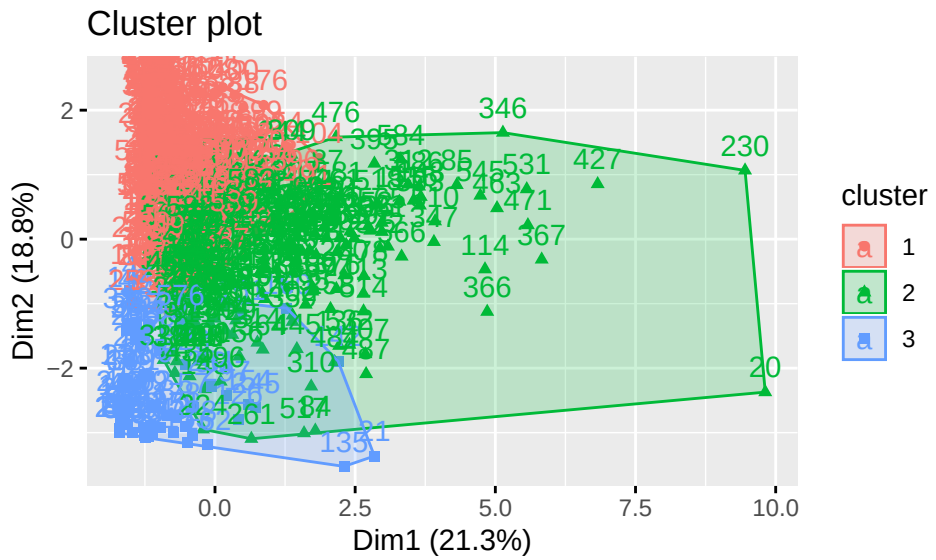
According to table, cluster 2 has higher affluence index and lower brand loyalty than other clusters. On the other hand, households in cluster 3 is the most loyal customers in this market and brand runs metric is the lowest among all households. Cluster 1 and Cluster 4 have similar consumption patterns with slight differences. Household size of cluster 4 is the highest; hence total consumption vary significantly from other clusters.

Cluster	SEC	HS	Affluence	maxBr	No of Brands	No of Trans	Brand Runs	Volume
1	2.391026	3.141026	13.98077	0.1591571	2.634615	20.59615	11.185897	6841.955
2	2.316038	4.551887	21.52830	0.2478214	5.014151	44.36792	25.268868	12592.934
3	2.727778	4.005556	13.92222	0.7224212	2.866667	22.88889	8.205556	10747.722
4	2.788461	6.519231	18.48077	0.2947516	3.692308	37.55769	16.769231	28408.173

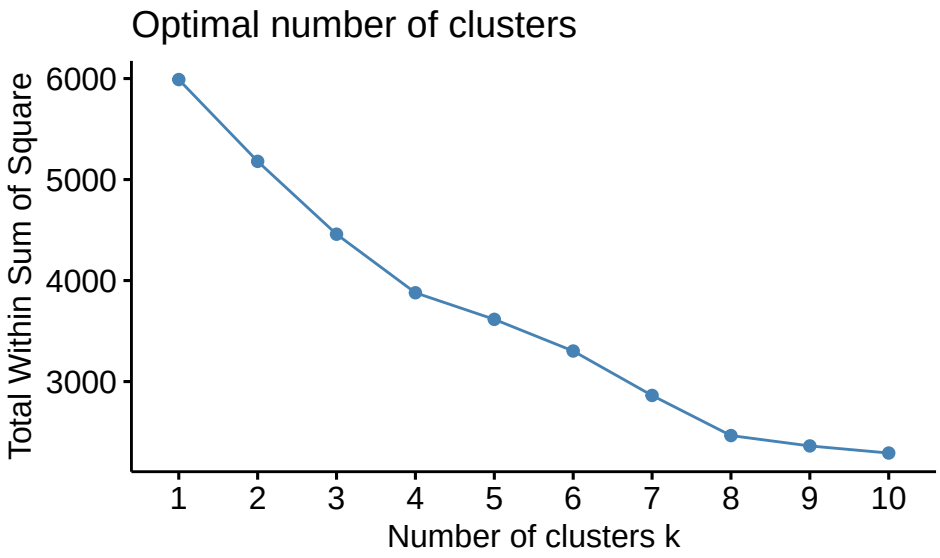
Basis for Purchase Variables

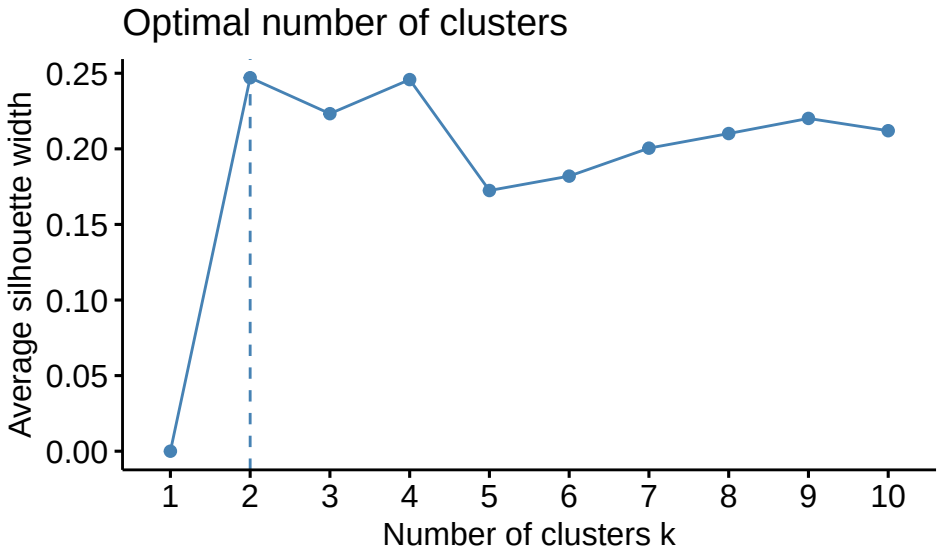
Now, just like with k means, we will cluster on the basis for purchase variables. Once again, three clusters will be used as a baseline. Here is the result of the three clusters:

size	max_diss	av_diss	diameter	separation
290	9.565754	1.967946	12.687231	0.4508119
237	15.534372	2.951697	21.109567	0.4508119
73	6.857316	1.564375	7.758894	0.6283877



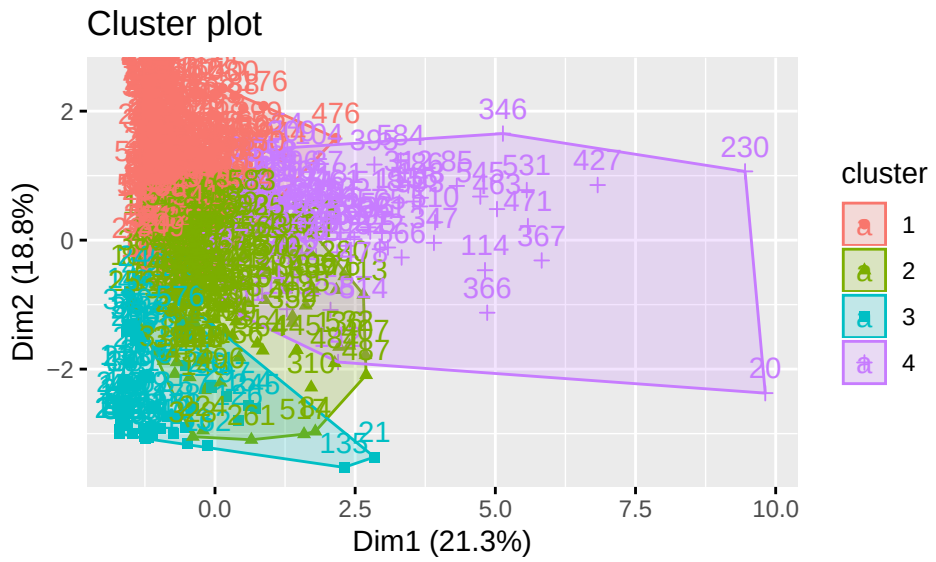
The silhouette method clearly indicates that 2 clusters would be optimal in the case of clustering on basis for purchase variables.





Running the pam algorithm with only two clusters yields clusters of very different sizes, as is visible below.

size	max_diss	av_diss	diameter	separation
254	9.825695	1.717100	12.687231	0.4625144
179	11.236392	2.662583	13.730433	0.4625144
70	6.857316	1.466505	7.758894	0.6932850
97	15.234251	2.775529	21.109567	0.6949584



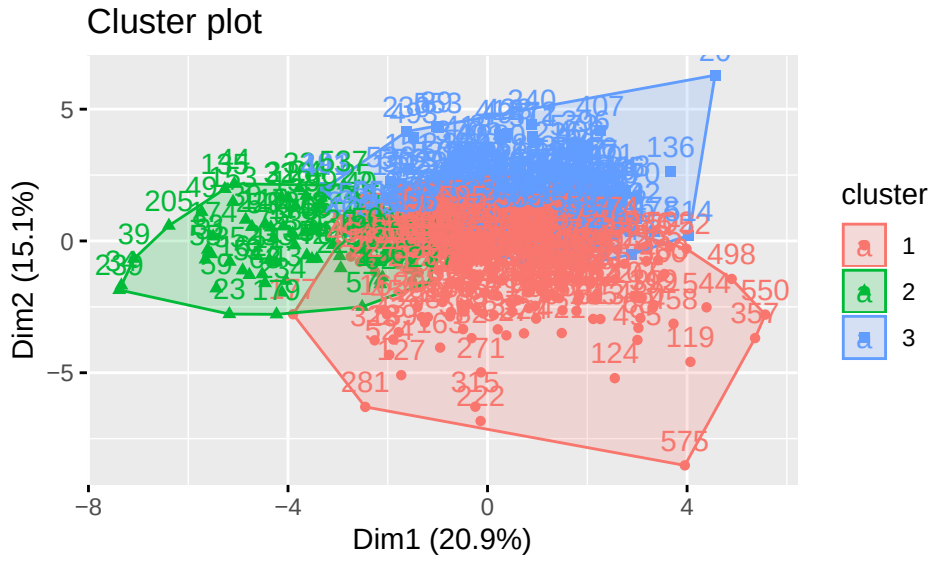
This table shows differences between clusters. For example, if we want to launch premium soap, we should try to understand cluster 1 and cluster 2. Nonetheless, when we want to increase sales of low priced products, we have to focus cluster 2 and cluster 4.

Cluster	SEC	HS	Affluence	maxBr	No of Brands	No of Trans	Brand Runs	Volume
1	2.622047	4.433071	16.374016	0.4110557	3.610236	30.13780	14.606299	13184.57
2	2.039106	3.905028	20.061453	0.2254913	3.765363	34.11173	17.944134	10513.85
3	3.385714	3.928571	8.471429	0.7709884	2.857143	24.92857	8.114286	12940.21
4	2.391753	4.278351	19.268041	0.2473120	4.030928	32.84536	20.216495	10434.90

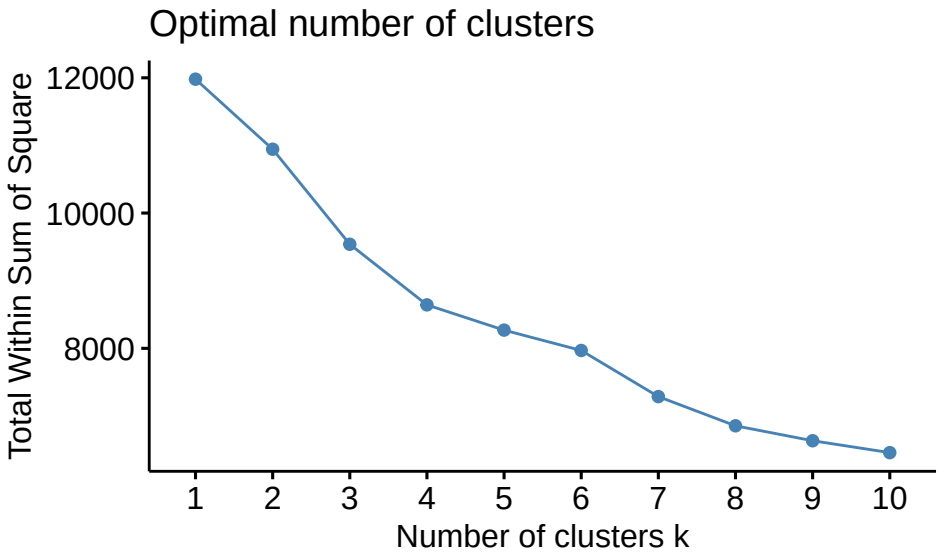
Combined Variables

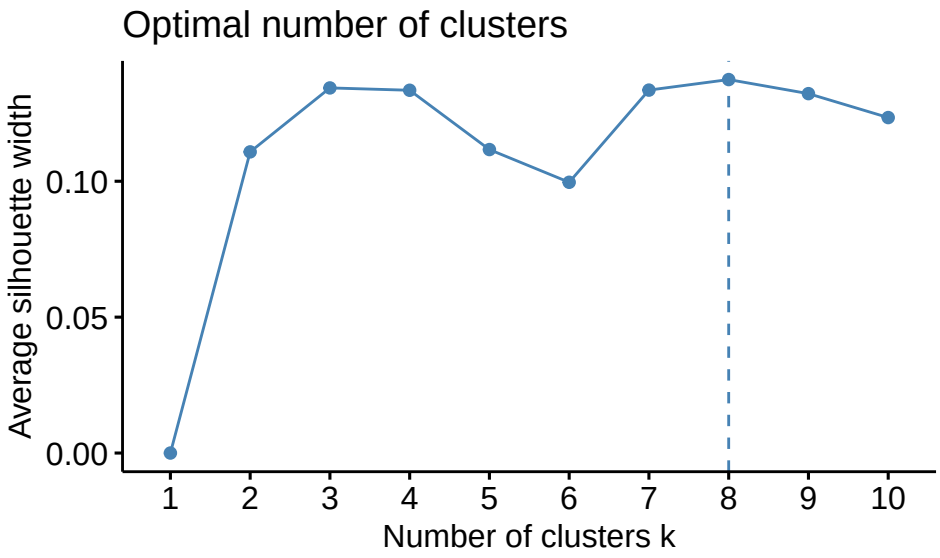
Lastly, we will impliment the k medoids algorithm on the combined variables. First, we will start with the baseline of three clusters.

size	max_diss	av_diss	diameter	separation
308	10.502413	3.628277	14.88709	1.296149
72	8.068223	3.358211	11.94597	1.656396
220	16.028957	4.176661	21.67919	1.296149



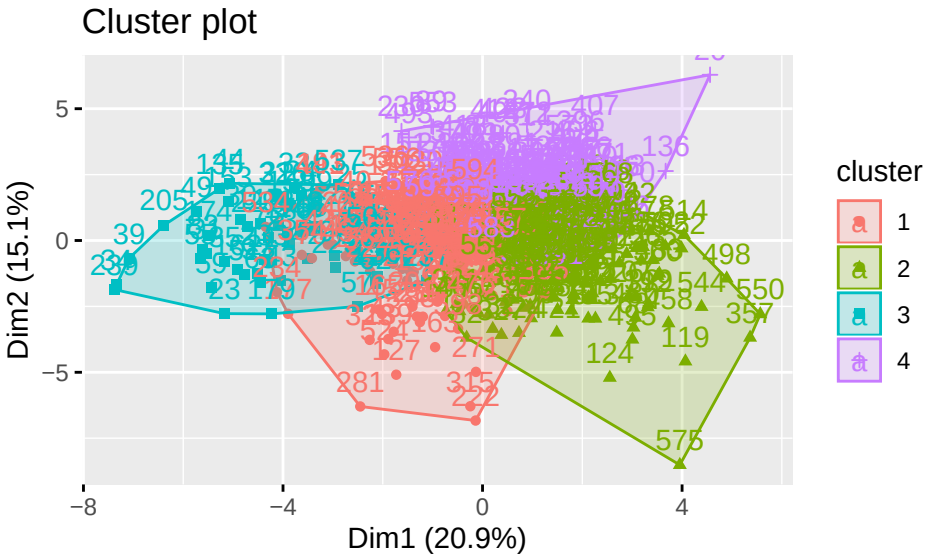
The three clusters are visualized below. The above table shows that the size differences in clusters are large. Instead, we will search for the optimal cluster size in an attempt to balance this.





The silhouette method indicates the optimal number of clusters as 8. This is a large number of clusters, though, given the business implications - 8 separate marketing plans would be very intensive to develop and implement. Therefore, we will select 4 as the best number of clusters given the result of the elbow method. 4 clusters also do not look too bad in the silhouette graph.

size	max_diss	av_diss	diameter	separation
177	10.327528	3.317120	13.87253	0.8590833
196	9.397880	3.490261	14.05870	1.0664848
65	8.068223	3.176974	11.94597	1.2785782
162	16.028957	4.023063	21.67919	0.8590833



The clusters are somewhat closer in size than just using 3 groups, though cluster 3 is still a lot smaller than the others. In the table below, we can see the differences between the 4 clusters on some of the key variables. Cluster three has higher brand loyalty than the other clusters. The other three, larger clusters vary on some

important variables such as brand runs, volume, and household size. Cluster 4 has the lowest household size, and purchases the smallest volume.

Cluster	SEC	HS	Affluence	maxBr	No of Brands	No of Trans	Brand Runs	Volume
1	2.728814	4.711864	16.050847	0.4815375	3.112994	25.58757	10.949153	14096.186
2	2.403061	4.821429	21.066326	0.2102490	5.000000	46.08673	26.132653	14036.668
3	3.446154	3.938461	7.953846	0.7977172	2.646154	23.92308	7.123077	13018.692
4	1.987654	2.962963	16.820988	0.2743024	2.956790	22.06790	11.901235	6521.204

Hierarchical Clustering

As a third clustering algorithm, we chose to use hierarchical clustering.

Purchase Behavior Variables

We implemented agglomerative hierarchical clustering by using different distance techniques such as weighted, complete and Ward's method.

We chose the clustering based on Ward's method rather than complete method. The sizes of clusters based on complete and weighted measures vary significantly. Most of households are in cluster 1 (86%). However, clustering with Ward's method creates more balanced clusters. We can check the agglomerative coefficient, which measures the amount of clustering structure found (values closer to 1 suggest strong clustering structure). We can clearly say that Ward's method is better than others on the basis of this dataset.

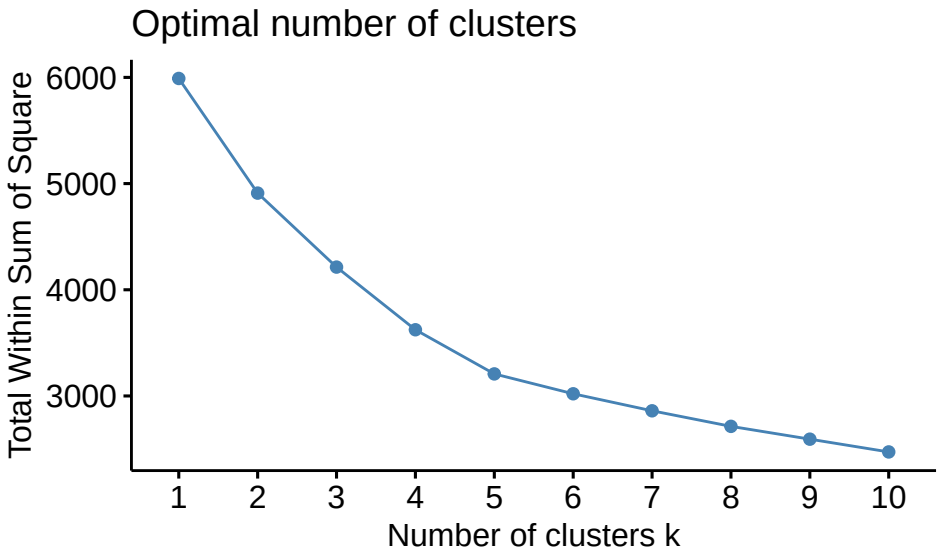
Method	Value
Agg. Coef of Weighted	0.89
Agg. Coef of Complete	0.93
Agg. Coef of Ward	0.98

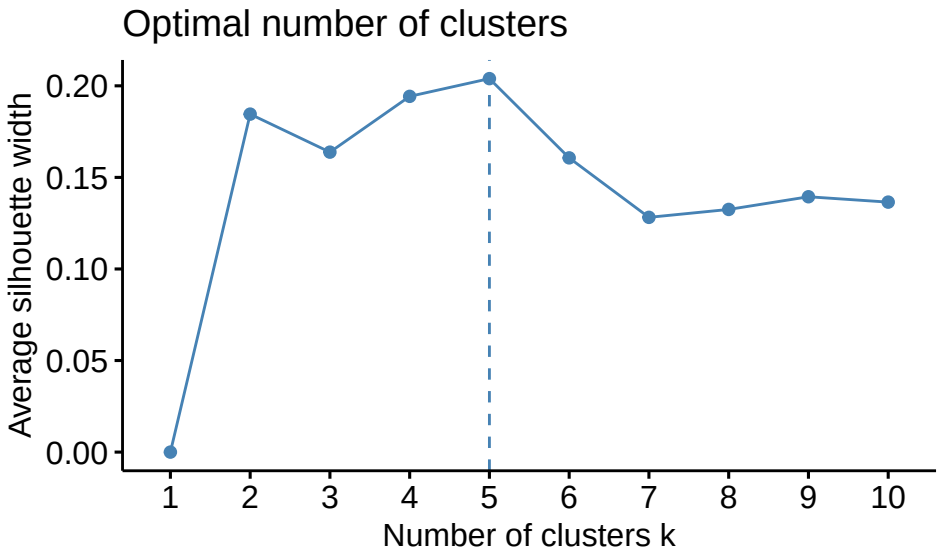
Cluster label based on weighted method	Size
1	328
2	269
3	3

Cluster label based on Complete method	Size
1	533
2	53
3	14

Cluster label based on Ward's method	Size
1	217
2	241
3	142

We checked for the optimal number of clusters using the hierarchical method.





This determines 5 clusters to be good based on both the elbow and silhouette methods. However, we determined the number of clusters as 4 since when we increase the number of clusters, size of several clusters is smaller. This does not add any significant value to business, because adding value of focusing some households in small clusters is not worth it.

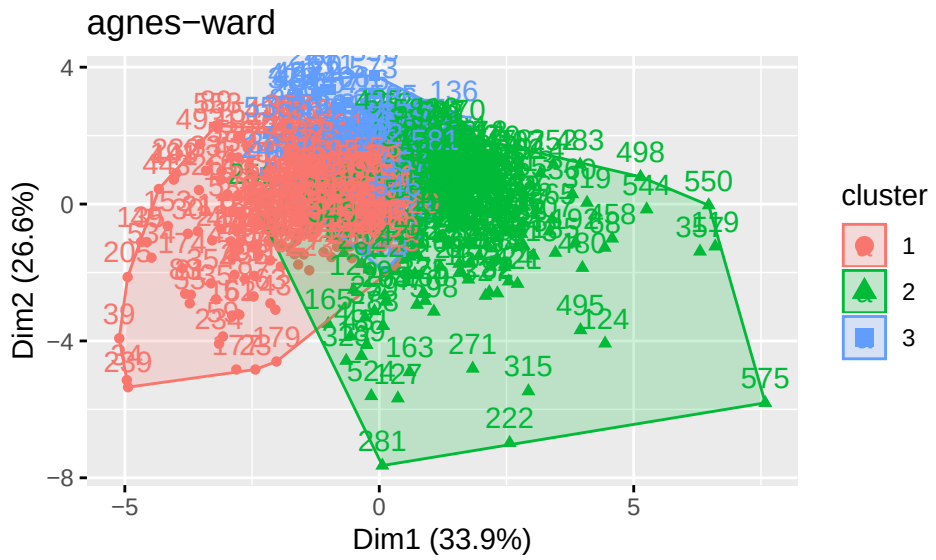
Cluster label k=4	Size
1	217
2	202
3	142
4	39

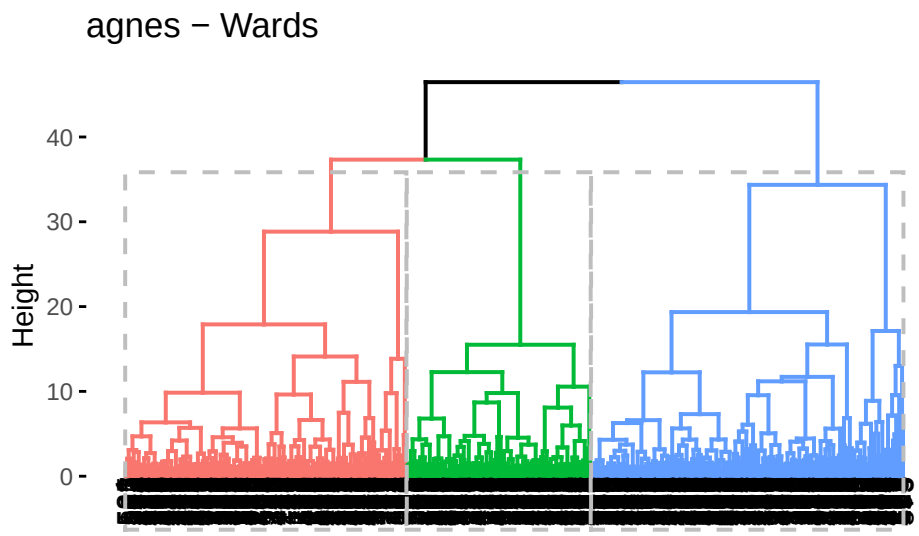
Cluster label k=5	Size
1	193
2	202
3	142
4	39
5	24

When we check features of the households, we can say that cluster 2 has much higher affluence index but less brand loyalty, meaning households in cluster 2 purchase a wider variety of brands in the market. On the other hand, cluster 1 has higher brand loyalty in this market. The average household size of cluster 4 is much higher than others, which is why volume of consumption is also higher.

Cluster	SEC	HS	Affluence	maxBr	No of Brands	No of Trans	Brand Runs	Volume
1	2.645161	3.903226	14.12903	0.6618342	3.082949	22.84332	9.336405	10251.530
2	2.326733	4.678218	21.40594	0.2316532	5.039604	45.64356	25.950495	13423.292
3	2.500000	3.436620	14.50000	0.1445449	2.598591	22.75352	11.450704	7752.676
4	2.589744	6.025641	19.56410	0.3023641	3.230769	32.92308	14.282051	28510.128

The first table shows cluster distribution of observations visually and the second one indicates the dendogram of the hierarchical clustering method which was built by using Ward's method with 4 clusters.





Basis for Purchase Variables

Secondly, we applied hierarchical clustering by using different variables. Basis of purchase variables obtains percent of volume purchased not on promotion, on promo code 6 and other than 6, proposition of beauty, health and baby products.

We chose the clustering based on Ward's method rather than the complete method. The sizes of the clusters based on complete and weighted measure vary significantly. Most of households are in cluster 1 (99%). However, clustering with Ward's method creates more balanced clusters. We can clearly say that Ward's method is better than the others on the basis of this dataset.

Method	Value
Agg. Coef of Weighted	0.95
Agg. Coef of Complete	0.96
Agg. Coef of Ward	0.98

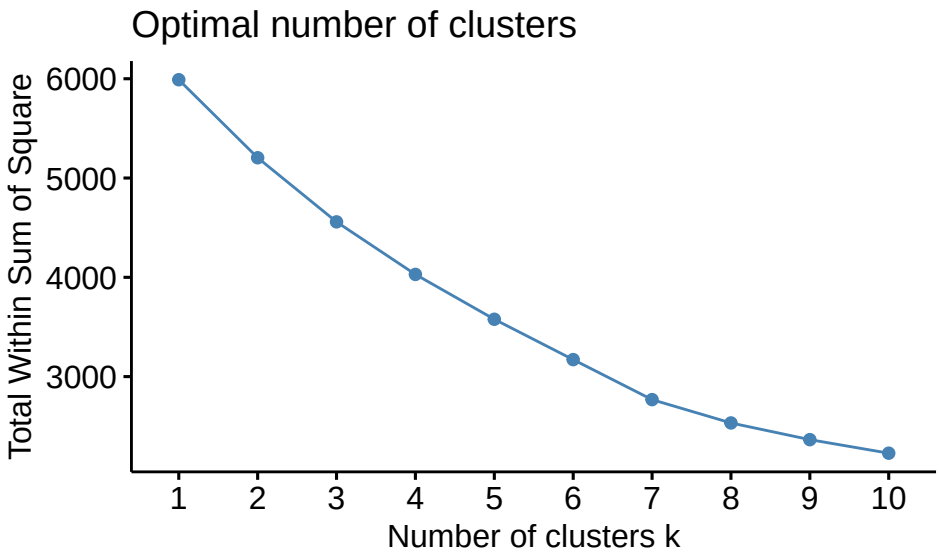
Cluster label based on weighted method	Size
1	592
2	7
3	1

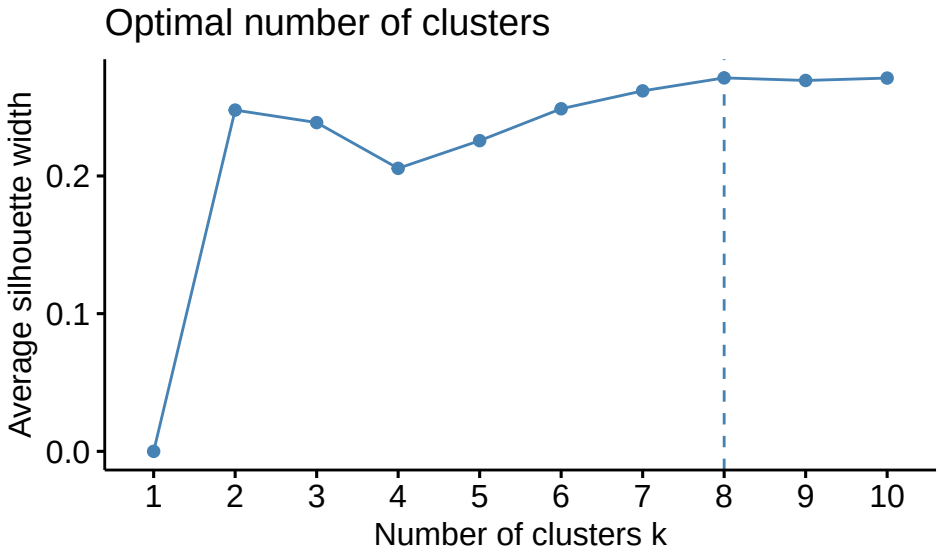
Cluster label based on Complete method	Size
1	594
2	5
3	1

Cluster label based on Ward's method	Size
1	419
2	66

Cluster label based on Ward's method	Size
3	115

Using these clustering variables, we again checked for the optimal clusters using both silhouette and elbow methods. The elbow and silhouette methods indicate 7-8 clusters as optimal, but this is a large amount from a business perspective.





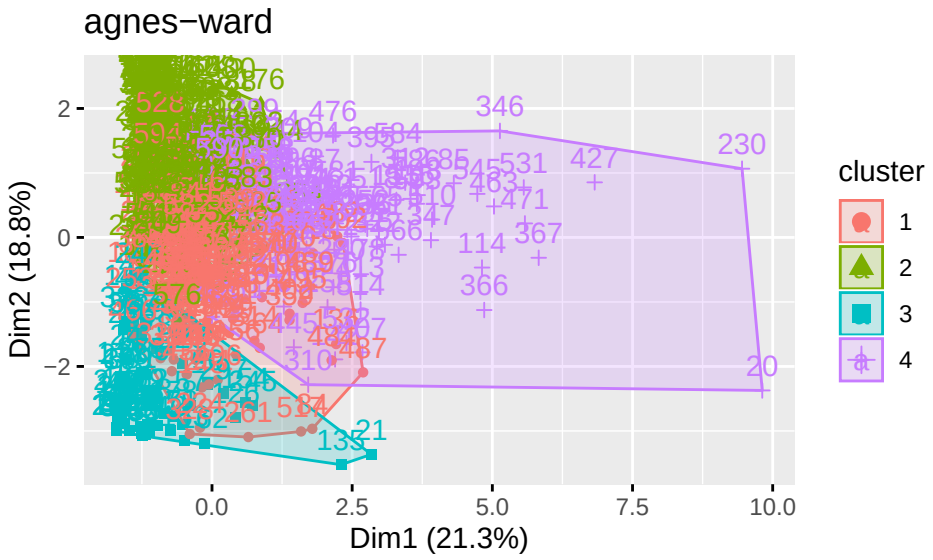
Instead, we decided to try smaller numbers of clusters and compare. In the first table, cluster 1 dominates other clusters, thus, 4 clusters are better than others. On the other hand, when we increase the number of clusters, some clusters are smaller and that is not sufficient for market segmentation. As a result, we determined the number of clusters as 4 since the clusters are more balanced than others. Concentrating on feasible customer segments is much better.

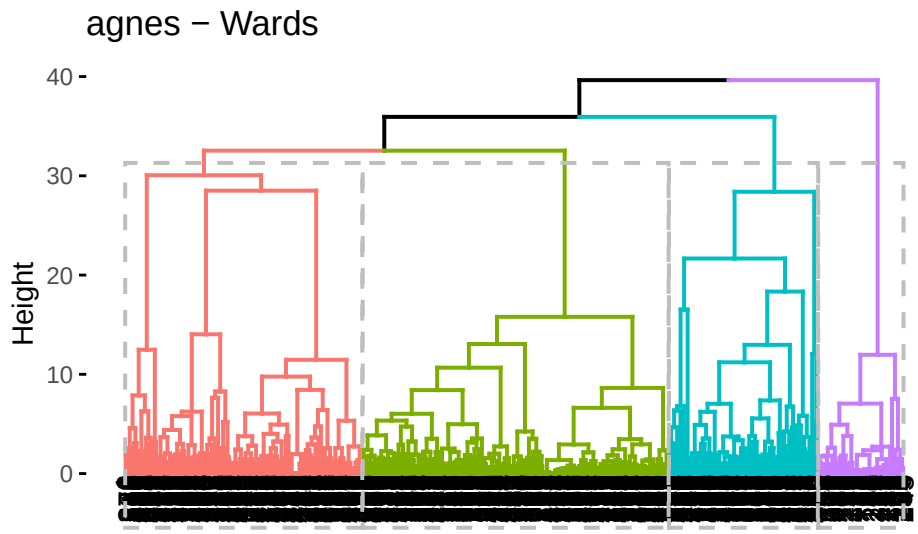
Cluster label k=3	Size
1	419
2	66
3	115

Cluster label k=4	Size
1	183
2	236
3	66
4	115

Cluster label k=5	Size
1	25
2	236
3	66
4	115
5	158

The first table shows the cluster distribution of observations, and the clusters are overlapping. The features were not able to seperate observations properly. The second one indicates the dendogram of the hierarchical clustering method which was built by using Ward's method with 4 clusters.





The following table indicates the differences between the clusters. Cluster 2 shows significant differences from the other clusters (higher brand loyalty and less affluence index). Nonetheless, clusters 3 and 4 show similar consumption behaviors.

Cluster	SEC	HS	Affluence	maxBr	No of Brands	No of Trans	Brand Runs	Volume
1	2.065574	3.841530	19.224044	0.2375024	3.677596	32.89071	16.912568	10649.81
2	2.635593	4.559322	17.072034	0.4206826	3.758475	31.03814	15.211864	13402.30
3	3.439394	3.848485	7.636364	0.7899970	2.818182	24.54545	7.757576	12936.59
4	2.373913	4.191304	18.791304	0.2421365	3.791304	32.41739	19.600000	10288.61

Combined Variables

Lastly, we applied hierarchical clustering by using combined variables in part a and part b. Combined variables includes both purchase behavior and basis for purchase variables.

We used Ward's method to measure distance between points (Ward performs well) and cut the tree by checking dendrogram. Overall, we can use 5 clusters to segment households and then concentrate on characteristics of these clusters.

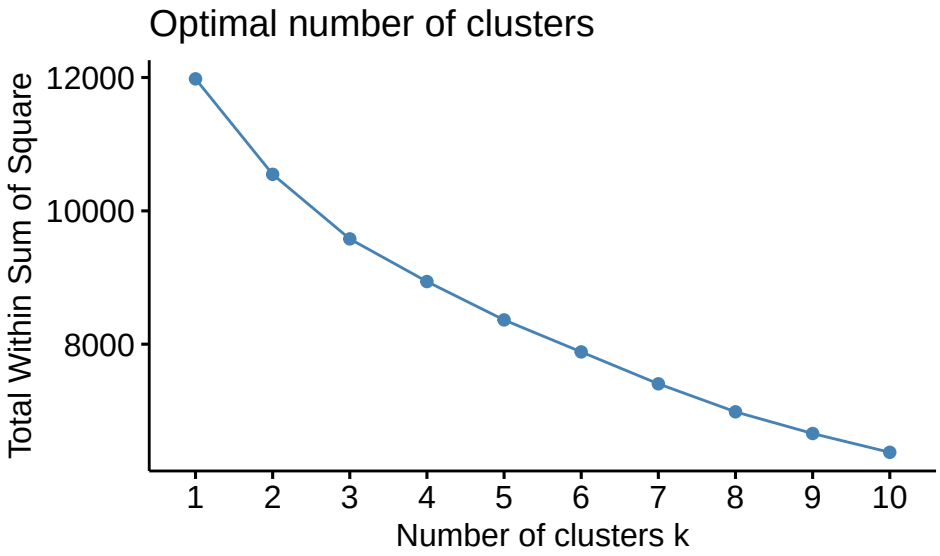
Method	Value
Agg. Coef of Ward	0.96

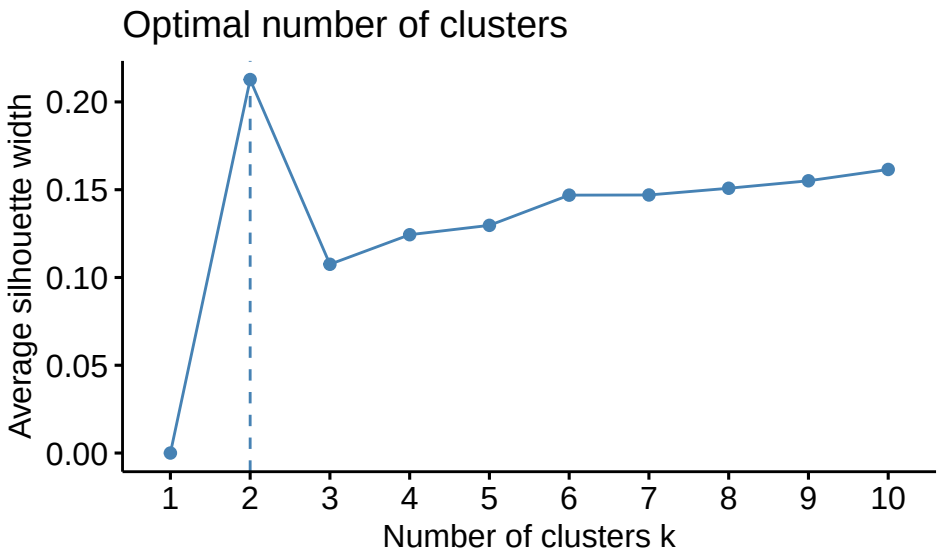
Cluster label k=3	Size
1	340
2	68
3	192

Cluster label k=4	Size
1	290
2	68
3	192
4	50

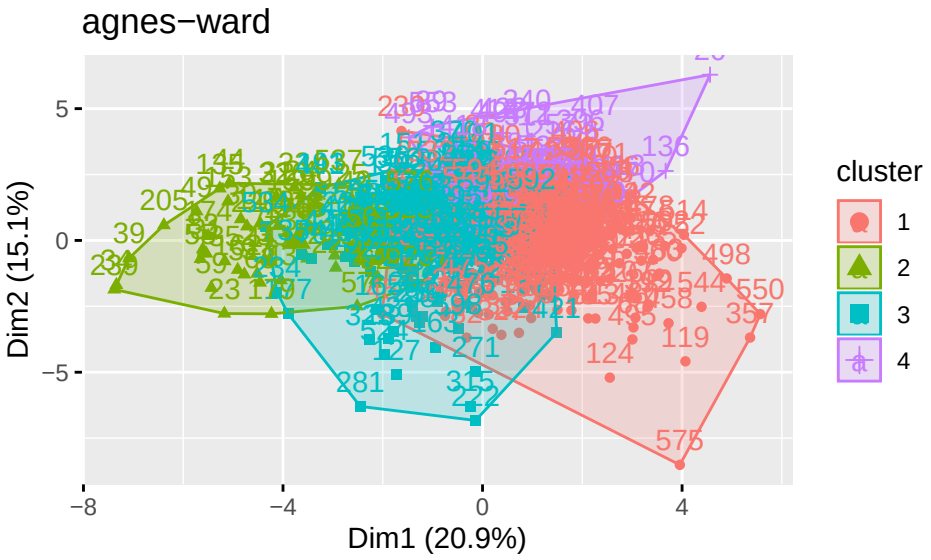
Cluster label k=5	Size
1	109
2	181
3	68
4	192
5	50

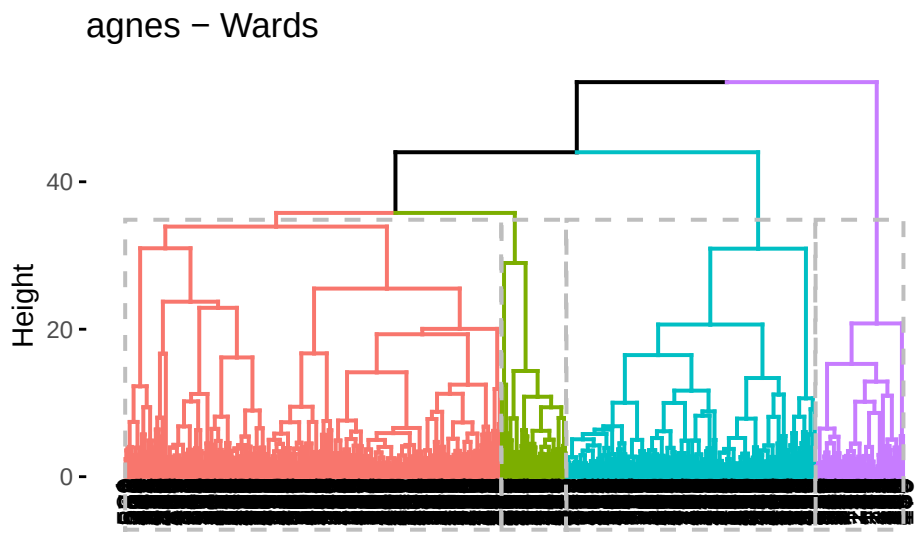
We also checked the elbow and silhouette method for the combined clustering variables. Elbow doesn't give a super distinctive result in this case. Silhouette indicates that we should use 2 clusters, but this is very few, so we still feel a few more are better.





First table shows cluster distribution of observations, and it is evident that clusters are overlapping, meaning the features were not able to seperate observations properly. The second one indicates dendrogram of hierarchical clustering method which was built by using Ward's method with 4 clusters.





This table shows differences between clusters based on combined variables. It can be clearly seen that behavior of cluster 1 and 2 are totally different. Cluster 2 has the lowest affluence index and the highest brand loyalty.(focus this group for low priced products). For premium products, you have to consider households in cluster 1 and cluster 3.

Cluster	SEC	HS	Affluence	maxBr	No of Brands	No of Trans	Brand Runs	Volume
1	2.400000	4.465517	19.720690	0.2172261	4.434483	38.97241	21.748276	12500.06
2	3.382353	3.852941	8.397059	0.7820743	2.720588	23.45588	7.308823	12401.69
3	2.578125	4.281250	15.708333	0.4877121	3.088542	25.20833	11.229167	12566.09
4	1.580000	2.720000	18.120000	0.2582438	2.360000	19.10000	9.820000	5356.80

Best Segmentation and Clustering Model

We implemented 3 different clustering methods: k-means, k-medoids and hierarchical clustering. The clusters obtained from these procedures are slightly different since the modeling approach varies from model to model. For example, similarity of k-means is based on means. However, k-means is sensitive to outliers and hence k-medoids calculates distances using the median. Additionally, agglomerative hierarchical clustering begins with the maximum number of clusters (the number of observations) and then continues until one cluster is left. We have a chance to cut the dendrogram at the proper level. The advantage of hierarchical clustering is that this model gives an analysis more in depth than other models.

We decided to use a low number of clusters such as 3 and 4 to segment households, because we did not find any benefit from using higher number of clusters in terms of business interpretation. Besides, the total number of households is in order of 100, a considerably small dataset, which also justifies using small number of clusters.

Overall, k-medoids performs slightly better than k-means. Hierarchical clustering and k-medoids shows similar performance by considering distribution of observations. We selected k-medoids model with 4 clusters for best segmentation and clustering. This shows the segmentation of the households in the soap market.

Cluster	SEC	HS	Affluence	maxBr	No of Brands	No of Trans	Brand Runs	Volume
1	2.728814	4.711864	16.050847	0.4815375	3.112994	25.58757	10.949153	14096.186
2	2.403061	4.821429	21.066326	0.2102490	5.000000	46.08673	26.132653	14036.668
3	3.446154	3.938461	7.953846	0.7977172	2.646154	23.92308	7.123077	13018.692
4	1.987654	2.962963	16.820988	0.2743024	2.956790	22.06790	11.901235	6521.204

The table below states that cluster 3 has the lowest affluence index and higher brand loyalty. Cluster 1 and 2 have similar households size, yet their brand loyalties are different. Cluster 4 purchases less than other clusters because the household size is lower. This model explains consumer behaviors more and adds vital value to marketing planning.

For this one 'best' segmentation, we built a decision tree by predicting labels of observations. We used information gain as splitting criteria and determined minsplit and complexity parameters as 40 and 0.001, respectively. Then, we checked variable importance as this table shows importance score of each feature:

```
##            Avg__Price             Pr_Cat_3            Brand_Runs
##            154.719836           140.866784            118.814076
##             Pr_Cat_1         No__of__Trans         No__of_Brands
##             98.126803            45.781570             42.372578
##            Others_999  Pur_Vol_No_Promo____              Pr_Cat_2
##             39.129662            36.143951             36.056754
##              Vol_Tran                 maxBr     Pur_Vol_Promo_6__
##             32.375785            31.303730             25.681229
##    Trans___Brand_Runs            PropCat_15 Pur_Vol_Other_Promo__
##             13.043221            10.900374              8.560410
##            PropCat_12          Total_Volume             PropCat_5
##             7.630262             6.009459              5.007882
```

According to this table,'Avg___Price','Pr__Cat_3','Brand_Runs','Pr_Cat_1' and 'No___of___Trans' are the most 5 important variable to predict clustering label accurately. This shows that the most important variables are combination of purchase behavior and basis for purchase.

This table demonstrates the confusion matrix of training data and train accuracy:

```
##    predTrn
##       1   2   3   4
##   1 116   7   1   3
##   2  11 120   3   3
##   3   0   0  50   0
##   4   8  20   0  78
```

Metric	Result
Training Accuracy	0.8667

This table shows the confusion matrix of testing data and testing accuracy:

```
##    predTst
##      1  2  3  4
##   1 44  4  1  1
##   2 12 44  0  3
```

```
##    3  0  0 15  0
##    4  7 12  0 37
```

Metric	Result
Testing Accuracy	0.7778