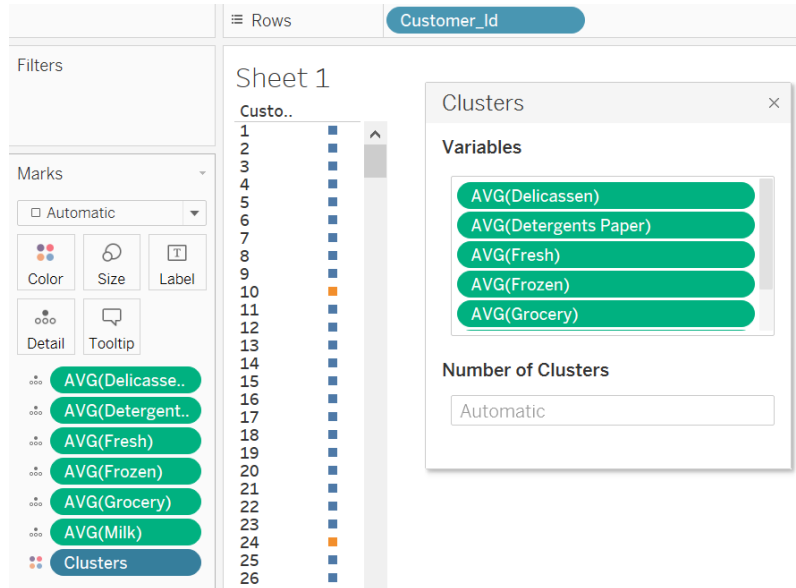


Segmentation and Profiling

By Fangling Zhang

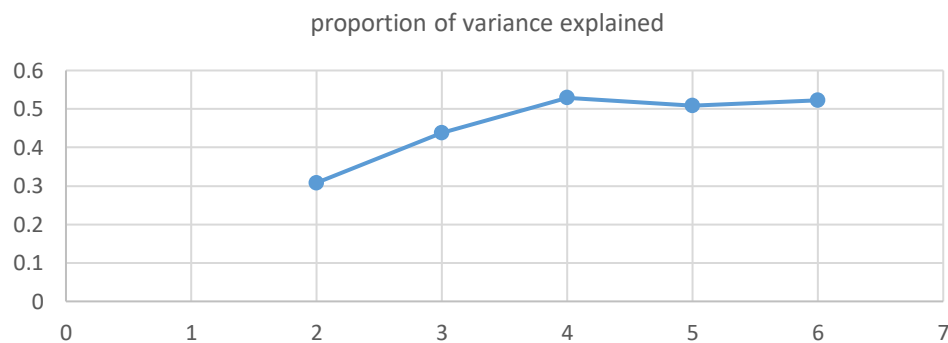
1. I created a Customer Id for each customer and put it in 'Rows'. Then I put average value of columns c – h in 'Detail' of 'Marks'. I applied cluster analysis with the automatically determined number of cluster.



Number of Clusters: 4
 Number of Points: 440
 Between-group Sum of Squares: 12.181
 Within-group Sum of Squares: 12.744
 Total Sum of Squares: 24.924

Clusters	Number of Items	Avg. I
Cluster 1	296	1
Cluster 2	110	1
Cluster 3	29	4
Cluster 4	5	2

- a. I also tried 3, 4, 5 and 6 number of cluster, the proportion of variance explained by the model (the ratio (between-group sum of squares)/(total sum of squares)) is showed in the following graph and we can see that 4 clusters is the best choice.



- b. In each segment, the customer number is showed in the table: 296, 110, 29 and 5.

Clusters	Number of Items	Avg. Delicassen	Avg. Detergents Paper	Avg. Fresh	Avg. Frozen	Avg. Grocery	Avg. Milk
Cluster 1	296	1247.2	737.51	14742.0	3674.3	3682.6	3007.5
Cluster 2	110	1516.3	4847.1	5211.8	1480.1	12169.0	8339.2
Cluster 3	29	4187.0	12638.0	7418.6	3037.2	26295.0	18121.0
Cluster 4	5	2708.8	29974.0	25603.0	2636.0	61472.0	43461.0
Not Clustered	0						

2.

- a. We can see from the above table that the cluster 1 purchases most Frozen, but least Delicassen, Detergents paper, Grocery and milk. The cluster 2 purchases most Milk, but least Fresh and Frozen. The cluster 3 purchases most Delicassen, but least Milk. The cluster 4 bought most Detergents paper, Fresh, Grocery and Milk.
- b. From the button left graph, there is no significant association between specific regions and specific segments of clients.
- c. From the button right graph, Channel 1 (Horeca (Hotel/Restaurant/CafÃ©))'s customers are mainly from Cluster 1. Comparatively, Channel 2 (or Retail channel (Nominal))'s customers are mainly from Cluster 2.

Region	Customer_Id (clusters)			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	68.83%	22.08%	9.09%	
2	68.09%	21.28%	8.51%	2.13%
3	66.77%	26.27%	5.70%	1.27%

Channel	Customer_Id (clusters)			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	89.93%	9.73%	0.34%	
2	19.72%	57.04%	19.72%	3.52%