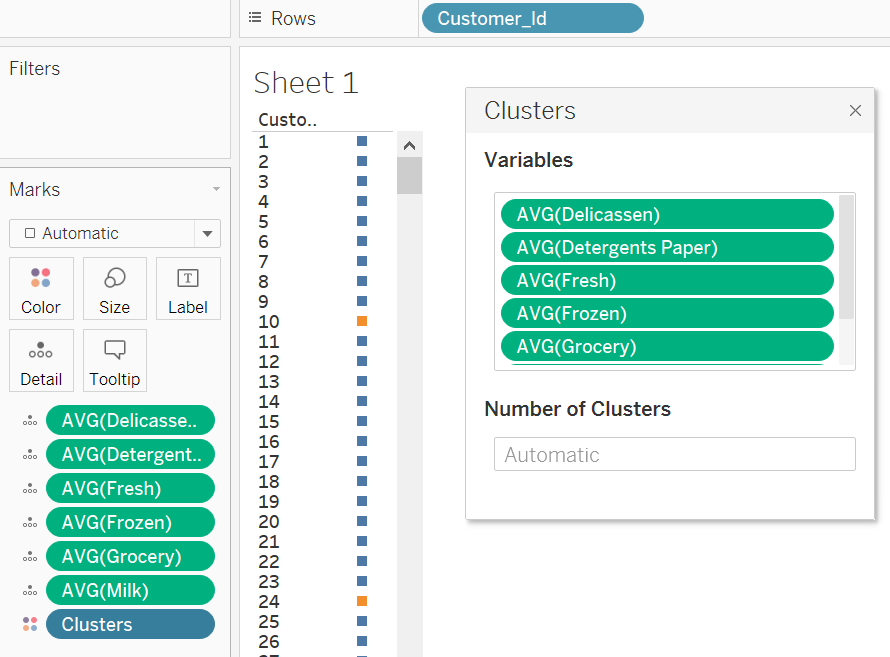
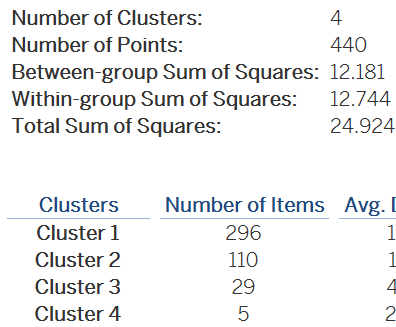
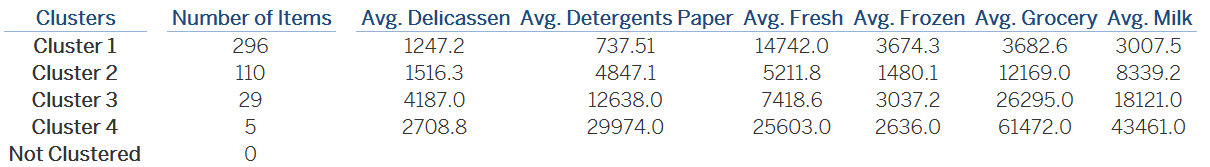
**Segmentation and Profiling**

**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.



1. 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.
2. In each segment, the customer number is showed in the table: 296, 110, 29 and 5.



1. 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.
2. From the button left graph, there is no significant association between specific regions and specific segments of clients.
3. 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.

