Case Study 2 Bath Soap Customer by Joshua Troup

1. **k-means clustering - to identify clusters of households based on • the variables that describe the purchase behavior including brand loyalty**--#brands, brand runs, total volume, #transactions, Value, Avg. price, share to other brands, max to one brand.

**• the variables that describe the basis for purchase**—Pur\_vol\_no\_promo, Pur\_vol\_promo\_6, Pur\_vol\_other, all price categories (4 variables), selling propositions 5 and 14 (2 variables) (most popular of these propositions)

• **the variables that describe both purchase behavior and basis of purchase**--#brand, brand runs, total volume, #transactions, Value, Avg. price, share to other brands, max to one brand, Pur\_vol\_no\_promo, Pur\_vol\_promo\_6, Pur\_vol\_other, all price categories (4), selling propositions 5 and 14 (2).

1. **Select the best segmentation and comment on the characteristics of these clusters.**

We ran 8 clusters schemes.

Clustering scheme 1 (K= 2, 3 and 4 clusters) on the basis of brand loyalty and other purchase behavior variables.

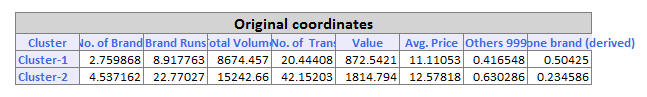
Clustering scheme 2 (K = 2, 3, and 4 clusters) on the basis of purchase variables (price and selling proposition).

Clustering scheme 3 (K = 2, 3 clusters) on the basis of both sets of variables used in schemes 1 and 2.

Option 1 which includes only brand loyalty variables is simple and achieves separation focusing an area of prime interest to client.

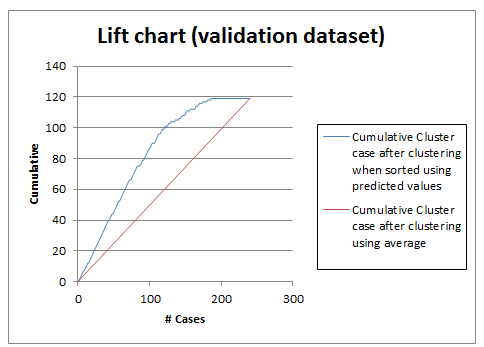
**Cluster 1** is brand loyal, relatively low volume, responsive to middling-high price, and relatively less affluent and with bigger households.

**Cluster 2** is less brand loyal, higher volume, responsive to a slightly higher price, relatively more affluent and with smaller households and more access to television.

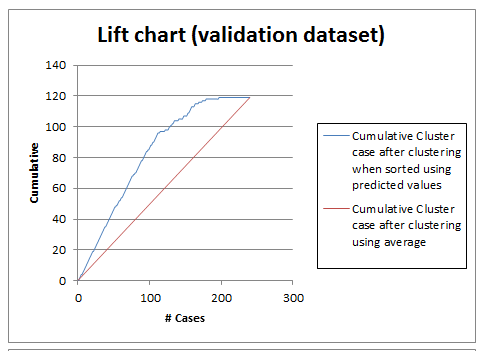


Cluster 2 is the targeted area to run sales or targeted campaign. The numbers above show the difference between Cluster 1 & 2. Much higher volume and number of transactions in Cluster 2.

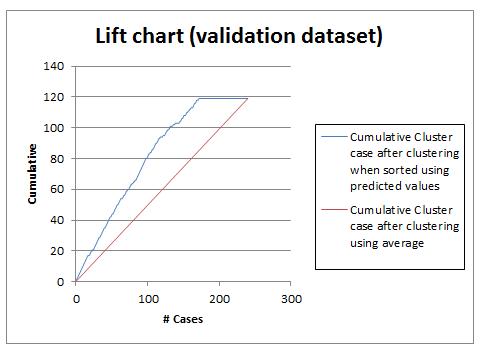
**Develop a model that classifies the data into these segments. As this will be used in direct mail promotions to target customers, select a market segment that would be defined as a success in the classification model.**

****

Logistic Regression: From the Lift Chart, if we assigned 100 cases to success class 2, about 100 class 2s would be included. If 100 cases were selected at random, we could expect about 50 only

****

Neural Network: From the Lift Chart, if we assigned 100 cases to success class 2, about 95 class 2s would be included. If 100 cases were selected at random, it would be 50.

****

Classification Tree: From the Lift Chart, if we assigned 100 cases to success class 2, about 80 class 2s would be included. If 100 cases were selected at random, we could expect about 50 only.

Conclusion: Logistic Regression & Neural Network seem to be the best classification model based on the lift chart results.