

Bath Soap Consumer Segmentation

Business problem

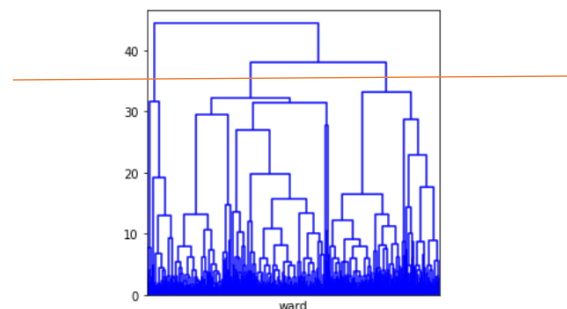
CRISA, an Asian market research agency that tracks data for purchases of consumer goods. Although demographic information for consumer records is included, demographic segmentation has already been explored. CRISA would like to pursue segmentation of the household goods market by (1) Purchase behavior (volume, frequency, susceptibility to discounts, and brand loyalty) and (2) Basis of purchase (price, selling proposition). This will help to identify purchase behaviors and brand loyalty, so targeted marketing promotions can be utilized. CRISA mainly has two types of customers (1) advertising agencies that help advise their clients on market trends and marketing strategies and (2) manufacturers of consumer goods who use the info for competitive determination and movement of goods.

Data

The data is split into three main types (1) demographic data (e.g. gender, age, education, children, affluence), (2) consumer purchase behavior (e.g. brands purchased, avg. price of purchase, volume purchase by brand), and (3) basis of purchase (purchase during a promotion, percent of volume purchased). Brand purchase category 999, price category 1 and 2, and proposition category 5 contain the fewest outliers of the brandwise, categorywise, and propositionwise categories. Thus, they may be stronger indicators of brand trends.

Purchase Behavior Clustering Analysis

Purchase Behavior includes the variables: Total Volume, Number of Transactions, Value, Volume per Transaction, Average Price, Percent Volume Purchased of each Brand, and the Brand Loyalty Variables (Number of brands purchased, Number of consecutive purchases of a given brand, and average transactions per consecutive brand). After running hierarchical clustering, it was determined that 3 clusters would be optimal (see figure below showing ward method with line showing cluster cut-off) .

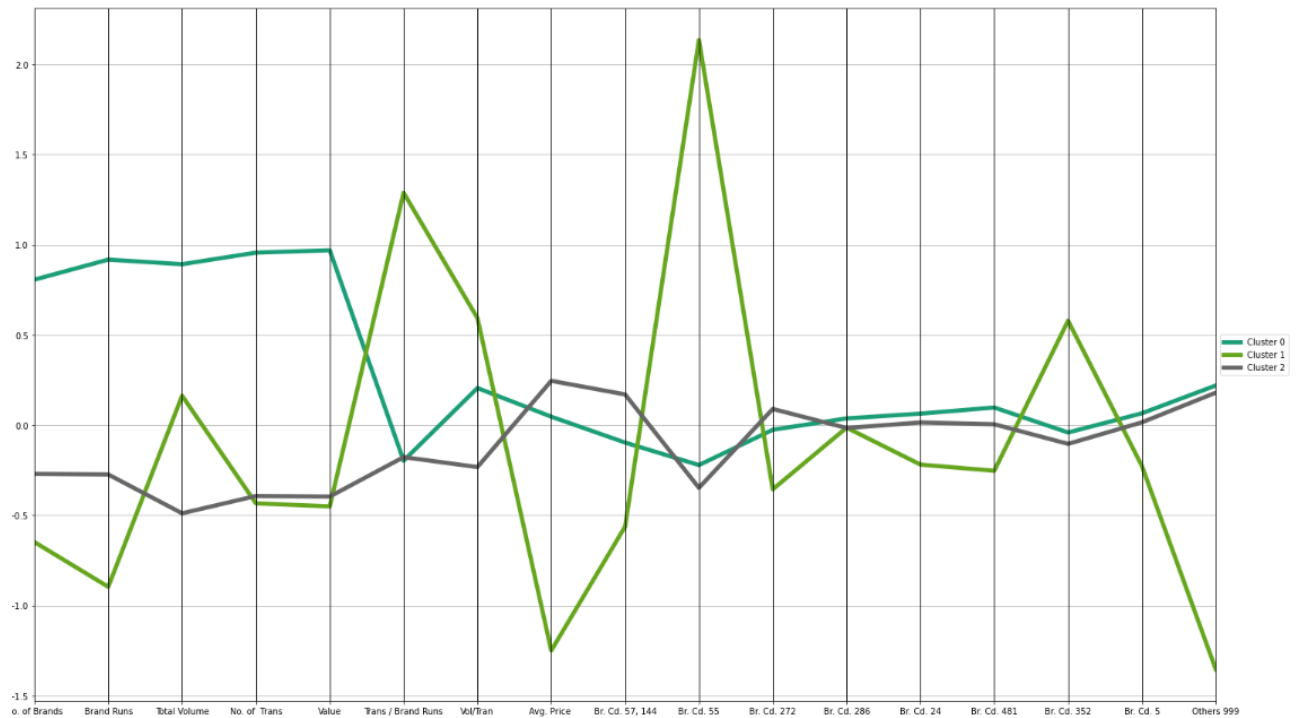


After running KMeans with 3 clusters, the following clusters appeared:

```
Cluster 0 (177 members): 2365.44 within cluster
Cluster 1 (75 members): 1261.83 within cluster
```

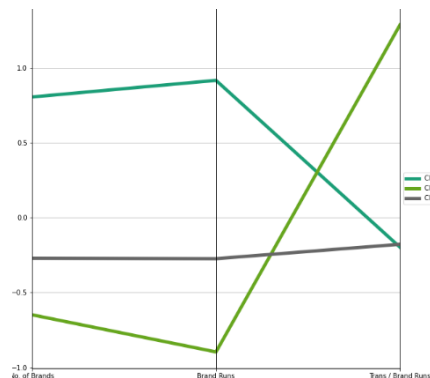
Cluster 2 (348 members): 4497.55 within cluster

The clusters appear fairly distinct and well distributed. The following behavior appeared with these three clusters:



Cluster 0 has a very high number of transactions but a moderate loyalty toward any brand. Cluster 1 customers purchase large quantities of lower cost brands with a preference for brands 55 and 352. Cluster 2 shows an average preference and purchase trend.

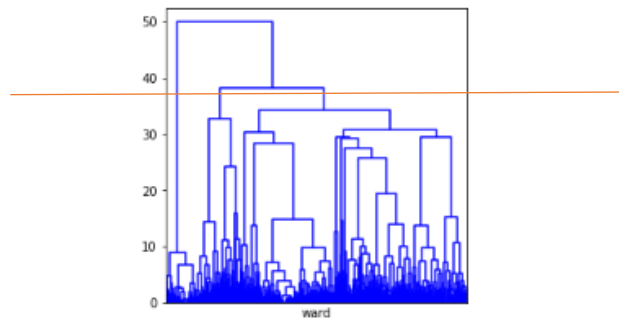
When specifically dealing with brand loyalty (as seen in the figure below), cluster 0 shows a propensity toward a large number of brands and will buy those brands consecutively, but, overall, they have fewer total transactions. Cluster 1 purchases few brands and is more apt to switch products but will purchase high quantity in each transaction (these are the cost conscious consumers). Cluster 2 has average brand loyalty characteristics.



A random forest prediction model was run to see how well cluster membership could be predicted with given data. An accuracy rate of 93.89% was achieved.

Basis of Purchase Clustering Analysis

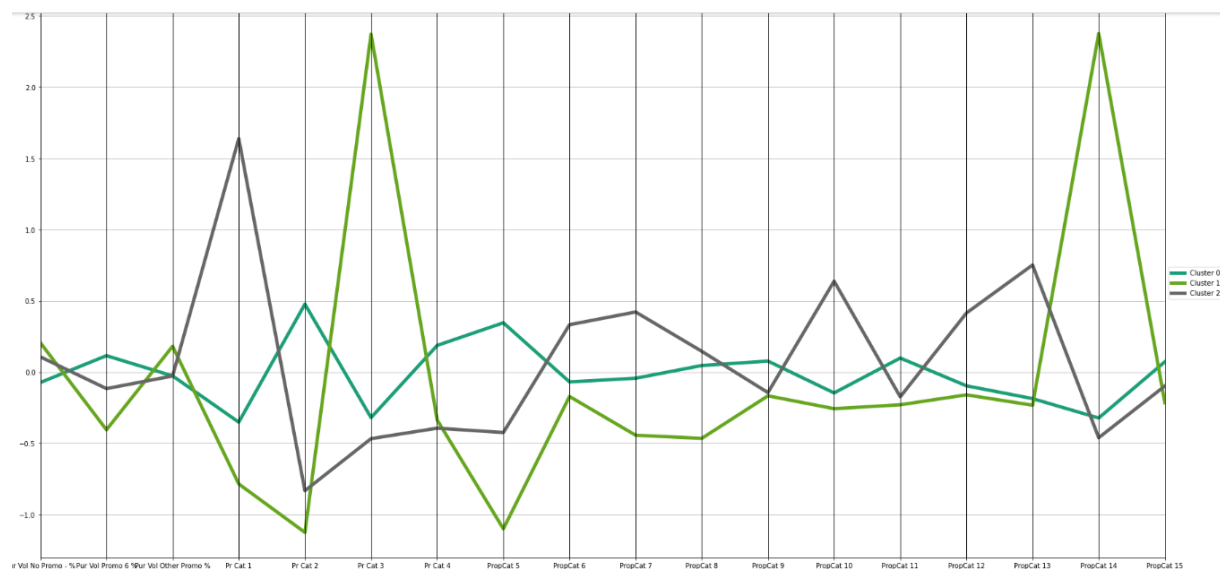
Basis of Purchase involves the variables with percent of volume purchased with no promotions, a 6% promo, and other promotions, as well as type of soap (e.g. premium or economy) purchased, and purchases from related categories. Using hierarchical clustering with ward's method, 3 clusters were determined to be optimal (see figure below).

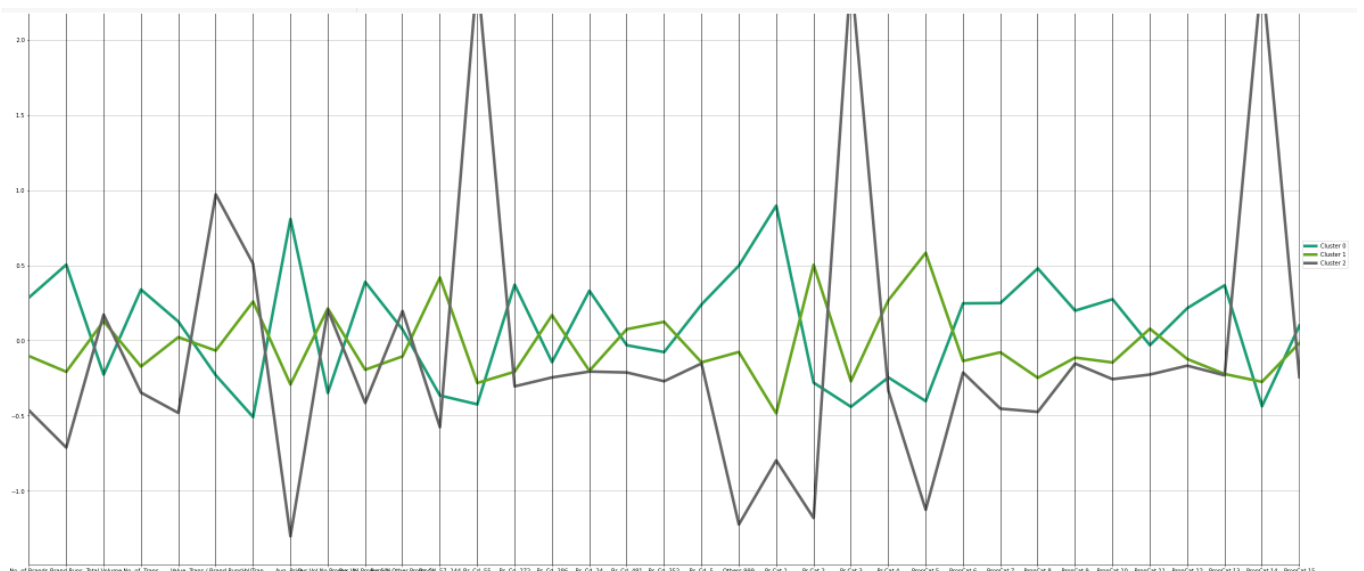


Running KMeans with 3 clusters provided the following three clusters:

```
Cluster 0 (399 members): 5492.44 within cluster
Cluster 1 (78 members): 402.51 within cluster
Cluster 2 (123 members): 2674.74 within cluster
```

The clusters appear tightly formed and distinct. The clusters have the following characteristics (see figure below): Cluster 0 is defined by high purchases of popular soap and beauty products, cluster 1 is defined by low purchases of popular soap, very high purchases of economy soap, low purchases of beauty products and very high purchases of carbohc products, cluster 2 is defined by very high purchases of premium soaps, low purchases of popular soaps, low purchases of beauty products, and high purchases of skin care and glycerine products.





A random forest prediction model was run to see how well combined cluster membership could be predicted with given data. An accuracy rate of 95.00% was achieved.

Market segmentation analysis

The combined categories provided the most useful customer segmentation by dividing the clusters into cluster 0, the luxury buyers with brand loyalty, cluster 1, the average market with a preference for popular products, and cluster 2, the economy buyers who prefer low-cost products. Advertising agencies who use CRISA data would benefit from targeting cluster 0 with luxury ads, cluster 1 with popular ads, and cluster 2 with ads targeted at low-cost alternatives. Manufacturers who use CRISA could use consumer groups to target R&D and packaging/ingredients/costs appealing to each cluster.