

Final Exam - MIS64060

Tim Pagliari tpagliar@kent.edu (mailto:tpagliar@kent.edu)

12/17/2020

Below is an analysis of market data intended to provide novel insights into customer activity related to purchasing behaviors and the basis of those purchases.

First, we'll clean and order our data, as well as look for any trends, outliers, or missing data before analysis.

	Member.id <fctr>	SEC <int>	FEH <int>	MT <int>	SEX <int>	AGE <int>	EDU <int>	HS <int>	CHILD <int>
1	1010010	4	3	10	1	4	4	2	4
2	1010020	3	2	10	2	2	4	4	2
3	1014020	2	3	10	2	4	5	6	4
4	1014030	4	0	0	0	4	0	0	5
5	1014190	4	1	10	2	3	4	4	3
6	1017020	4	3	10	2	3	4	5	2

6 rows | 1-10 of 47 columns

Since we don't see any major outliers or missing data, we can proceed to the main goal of the project, which is to cluster purchases based on the following:

- The variables that describe purchase behavior (volume, frequency, susceptibility to discounts, and brand loyalty)
 - Volume: Total.Volume
 - Frequency: No..Of..Transactions
 - Susceptibility:

```
Soap$Susceptibility <- (1-Soap$Pur.Vol.No.Promo....)
```

- Brand Loyalty: One of three types:
 - Loyalty 1: Based on the number of different brands purchased. May be defined by **1 / No.of.Brands**
 - Loyalty 2: Based on how often customers switch from one brand to another. May be defined by the number of brand runs divided by total transactions, or **(Brand Runs /**

No.Transactions)

3. Loyalty 3: Based on the proportion of purchases that go to different brands. Definition:
highest of the 9 categories = proportional loyalty
We'll define a variable "Loyalty" which is the maximum of the three scaled loyalty categories.

```
Soap$Loyalty1 <- (1/Soap$No..of.Brands)
Soap$Loyalty2 <- (Soap$Brand.Runs/Soap$No..of..Trans)
Soap <- Soap %>% mutate(Loyalty3=pmax(Br..Cd..57..144, Br..Cd..55, Br..Cd..272,
Br..Cd..286, Br..Cd..24, Br..Cd..481, Br..Cd..352, Br..Cd..5, Others.999))
Soap <- Soap %>% mutate(Loyalty=pmax(Loyalty1,Loyalty2,Loyalty3))
```

- The variables that describe the basis for purchase (price, and selling proposition) Price: Value, Average Price of Purchase, and Price Cat Selling Proposition: Proposition Category 5:15
- The variables that describe both purchase behavior and basis of purchase (volume, frequency, susceptibility to discounts, brand loyalty, price, and selling proposition)

We'll next divide these variables and normalize them as the three types defined above:

```
Type1 <- select(Soap, Total.Volume, No..of..Trans, Susceptibility, Loyalty,)
Type1.norm <- scale(Type1, scale=TRUE, center=TRUE)
Type2 <- select(Soap, Value, Avg..Price, Pr.Cat.1, Pr.Cat.2, Pr.Cat.3, Pr.Cat.4, Prop
Cat.5, PropCat.6, PropCat.7, PropCat.8, PropCat.9, PropCat.10, PropCat.11, PropCat.12
, PropCat.13, PropCat.14, PropCat.15)
Type2.norm <- scale(Type2, scale=TRUE, center=TRUE)
Type3 <- select(Soap, Total.Volume, No..of..Trans, Susceptibility, Loyalty,Value, Avg
..Price, Pr.Cat.1, Pr.Cat.2, Pr.Cat.3, Pr.Cat.4, PropCat.5, PropCat.6, PropCat.7, Pro
pCat.8, PropCat.9, PropCat.10, PropCat.11, PropCat.12, PropCat.13, PropCat.14, PropCa
t.15)
Type3.norm <- scale(Type3, scale=TRUE, center=TRUE)
```

Now we'll use k-means to cluster the households:

```

set.seed(111)
#fviz_nbclust(Type1.norm, kmeans, method = "wss") # find optimal k, in addition to the
#fviz_nbclust(Type1.norm, kmeans, method = "silhouette")
kmeans.type1<-kmeans(Type1.norm, centers=4)
Soap$Type1<- kmeans.type1$cluster

kmeans.type2<-kmeans(Type2.norm, centers=5)
Soap$Type2<- kmeans.type2$cluster

kmeans.type3<-kmeans(Type3.norm, centers=5)
Soap$Type3<- kmeans.type3$cluster

```

To focus more on purchasing behavior and develop a targeted promotion for these customers, I recommend we look at the customers segmented by Type 1 characteristics.

```

MarketSegment <- summarize_at(group_by(Soap, Soap$Type1), vars(Susceptibility, Loyalty1,
Loyalty2, Loyalty3, SEC, FEH, SEX, AGE, EDU, HS, CHILD, CS, Affluence.Index, Pr.Cat.1, Pr.Cat.2,
Pr.Cat.3, Pr.Cat.4, Avg..Price, Value, Pr.Cat.1, Pr.Cat.2, Pr.Cat.3, Pr.Cat.4, PropCat.5,
PropCat.6, PropCat.7, PropCat.8, PropCat.9, PropCat.10, PropCat.11, PropCat.12, PropCat.13,
PropCat.14, PropCat.15), mean)
MarketSegment

```

Soap\$Typ...	Susceptibility	Loyalty1	Loyalty2	Loyalty3	SEC	FEH	SEX
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.04247788	0.4524336	0.5627510	0.8089912	2.424779	1.774336	1.517699
2	0.07225490	0.2709111	0.4806144	0.7040588	2.686275	2.568627	1.960784
3	0.04807292	0.2720424	0.5080650	0.5582135	2.546875	2.203125	1.859375
4	0.32462500	0.3259821	0.6291972	0.7419750	2.362500	1.787500	1.787500

4 rows | 1-9 of 31 columns

Group 1 (n=226) - Has the lowest aggregate education of all segments as well as the fewest household members and the lowest on the Affluence Index. May try other brands but spends a large proportion in one brand category. They spend mostly in price category 2 or less and have the lowest in average price and total value of all the customer groups. They are not susceptible to promotions.

Group 2 (n=102) - Is almost entirely female, and is the oldest of the customer groups. They have an average of 6.5 people in the household, and are well above average on the Affluence Index. They switch brands often, are not very susceptible to promotions, and shop mostly in price category 2 and 1. They have the lowest average price but by far the greatest total value – more than twice that of the nearest segment.

Group 3 (n=192) - Is average in many areas, but is the most educated and among the youngest customer segments. Customers average 4.17 people in the household and have an average rating on the Affluence Index. This segment the lowest brand loyalty of the 4 groups, and buys mostly from price category 2 or 1.

Group 4 (n=80) - Is older and has an average of 3.77 members in the household with a high likelihood of children in the home. The group is high on the Affluence Index and is by far the most susceptible to promotions. Will try multiple brands but maintains long stretches of brand loyalty, and spreads their purchases across price categories 1, 2 and 4. They have the highest average price and an average total value.

—

Conclusion: Due to their high susceptibility to promotions and a wide preference for price categories as well as a willingness to try new products, Group 4 would be the ideal to train a classification model intended for a direct mail campaign highlighting promotional offers.
