

BUAN 6337 - F19 – PREDICTIVE ANALYTICS USING SAS PROJECT REPORT Group 10

Sandeep Pradeep Kumar
Deepa Subramanian
Chinmay Sunil Karandikar
Rohit Gurjar
Vidhya Venkatraman

OVERVIEW

Based on the total sales of each brand (in dollars), we have the following top 10 brands

Obs	Brand_Name	Total_Sales
1	KRAFT MIRACLE WHIP	17598441.77
2	HELLMANNS	17138996.19
3	BEST FOODS	10154714.01
4	KRAFT MAYO	9506461.18
5	PRIVATE LABEL	5991590.34
6	DUKES	1425456.00
7	HELLMANNS LIGHT	1281270.03
8	CAINS	1151269.02
9	HELLMANNS JUST 2 GOOD	1075791.55
10	BLUE PLATE	954136.52

We choose to do analysis for the brand Hellmann's and come up with findings which could help the brand achieve better results in terms of sales, market share or customer preference.

ANALYSIS 1

<u>RFM ANALYSIS – CLUSTER ANALYSIS</u>

Our objective in this section is to identify customer purchasing patterns that drive the sales of mayonnaise. This can help the company to decide which segments of customers they have to target to boost sales. This can be done with the help of a technique called RFM analysis. This technique can help us analyze the frequency with which a customer buys the product, the recency of the purchase and the average money the customer spends on his purchases.

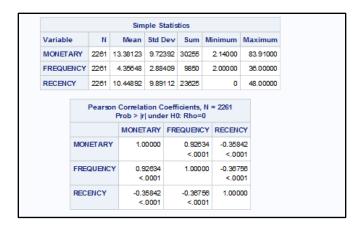
RFM analysis is a customer segmentation technique that uses past **purchase behavior** to divide customers into groups.

RECENCY (R): Time since last purchase
FREQUENCY (F): Total number of purchases
MONETARY VALUE (M): Total monetary value

Benefits of RFM Analysis

- Increased customer **retention**
- Increased **response** rate
- Increased conversion rate
- Increased revenue

To understand the customer patterns, initially we check for the RFM statistics. From the below figure, we can see that an average customer spends around \$13.4 a year and visits around 4 times in a year. Also, the customer tends to buy the products once in 10 weeks on an average.



We can see a huge correlation between frequency and monetary. This is as expected as more frequency means more money spent. Recency does not have a strong correlation with both frequency and monetary.

PANID	279	1368174.69	496638.96	1100180.00	3843060.00
FAM_INCOME_L	279	0.1505376	0.3582402	0	1.0000000
FAM_INCOME_M	279	0.5863082	0.4964743	0	1.0000000
FAM_INCOME_H	279	0.2831541	0.4513402	0	1.0000000
FAM_SIZE_S	279	0.6308244	0.4834488	0	1.0000000
FAM_SIZE_L	279	0.3891756	0.4834488	0	1.0000000
AGE_MALE_Y	279	0.0035842	0.0598684	0	1.0000000
AGE_MALE_M	279	0.4910394	0.5008180	0	1.0000000
AGE_MALE_O	279	0.5053763	0.5008695	0	1.0000000
AGE_FEMALE_Y	279	0	0	0	0
AGE_FEMALE_M	279	0.5519713	0.4981853	0	1.0000000
AGE_FEMALE_O	279	0.4480287	0.4981853	0	1.0000000
Edu_Female_S	279	0.6379928	0.4814445	0	1.0000000
Edu_Female_C	279	0.3620072	0.4814445	0	1.0000000
Edu_Male_S	279	0.6487455	0.4782201	0	1.0000000
Edu_Male_C	279	0.3512545	0.4782201	0	1.0000000
OCC_Male_WH	279	0.2903226	0.4547269	0	1.0000000
OCC_Male_WL	279	0.0501792	0.2187089	0	1.0000000
OCC_Male_B	279	0.3440860	0.4759230	0	1.0000000
OCC_Male_NW	279	0.3154122	0.4655148	0	1.0000000
OCC_Male_O	279	0	0	0	0
OCC_Female_WH	279	0.3835125	0.4871151	0	1.0000000
OCC_Female_WL	279	0.1182796	0.3235192	0	1.0000000
OCC_Female_B	279	0.1326165	0.3397692	0	1.0000000
OCC_Female_NW	279	0.3855914	0.4824610	0	1.0000000
OCC_Female_O	279	0	0	0	0
Marital_Status_S	279	0.0250896	0.1566783	0	1.0000000
Marital_Status_M	279	0.9569892	0.2032459	0	1.0000000
Marital_Status_DW	279	0.0179211	0.1329033	0	1.0000000
HH_RACE_H	279	0.0071685	0.0845143	0	1.0000000
HH_RACE_NH	279	0.9928315	0.0845143	0	1.0000000
HH_Race3_W	279	0.9858831	0.1190891	0	1.0000000
HH_Race3_B	279	0.0071685	0.0845143	0	1.0000000
HH_Race3_H	279	0	0	0	0
HH_Race3_A	279	0	0	0	0
HH_Race3_O	279	0.0035842	0.0598684	0	1.0000000
Total_Pets_0	279	0.3942652	0.4895704	0	1.0000000
Total_Pets_1	279	0.6057348	0.4895704	0	1.0000000
Child_1	279	0.0143389	0.1190891	0	1.0000000
Child_2	279	0.0845161	0.2481116	0	1.0000000
Child 3	279	0.1541219	0.3817143	0	1.0000000

Child_4	279	0.0143369	0.1190891	0	1.0000000
Child_5	279	0.0035842	0.0598684	0	1.0000000
Child_6	279	0.0537634	0.2259558	0	1.0000000
Child_7	279	0.0071685	0.0845143	0	1.0000000
MONETARY	279	29.0789964	10.5386284	18.5100000	78.6100000
FREQUENCY	279	8.7992832	3.1295527	3.0000000	21.0000000
RECENCY	279	4.3820072	5.3187727	0	31.0000000

PANID	193	1834616.94	977923.91	1100057.00	3842039.00
FAM_INCOME_L	193	0.1243523	0.3308413	0	1.0000000
FAM_INCOME_M	193	0.5025907	0.5012937	0	1.0000000
FAM_INCOME_H	193	0.3730570	0.4848749	0	1.0000000
FAM_SIZE_S	193	0.6735751	0.4701243	0	1.0000000
FAM_SIZE_L	193	0.3264249	0.4701243	0	1.0000000
AGE_MALE_Y	193	0	0	0	0
AGE_MALE_M	193	0.4863212	0.5001619	0	1.0000000
AGE_MALE_O	193	0.5336788	0.5001619	0	1.0000000
AGE_FEMALE_Y	193	0	0	0	0
AGE_FEMALE_M	193	0.5077720	0.5012398	0	1.0000000
AGE_FEMALE_O	193	0.4922280	0.5012398	0	1.0000000
Edu_Female_S	193	0.5233161	0.5007550	0	1.0000000
Edu_Female_C	193	0.4766839	0.5007550	0	1.0000000
Edu_Male_S	193	0.5899482	0.4963707	0	1.0000000
Edu_Male_C	193	0.4300518	0.4983707	0	1.0000000
OCC_Male_WH	193	0.3575130	0.4805141	0	1.0000000
OCC_Male_WL	193	0.0362694	0.1874480	0	1.0000000
OCC_Male_B	193	0.2901554	0.4550143	0	1.0000000
OCC_Male_NW	193	0.3160622	0.4881487	0	1.0000000
OCC_Male_O	193	0	0	0	0
OCC_Female_WH	193	0.4145078	0.4939181	0	1.0000000
OCC_Female_WL	193	0.1191710	0.3248322	0	1.0000000
OCC_Female_B	193	0.0880829	0.2841525	0	1.0000000
OCC_Female_NW	193	0.3782383	0.4882088	0	1.0000000
OCC_Female_O	193	0	0	0	0
Marital_Status_S	193	0.0466321	0.2113980	0	1.0000000
Marital_Status_M	193	0.9430052	0.2324357	0	1.0000000
Marital_Status_DW	193	0.0103627	0.1015319	0	1.0000000
HH_RACE_H	193	0	0	0	0
HH_RACE_NH	193	1.0000000	0	1.0000000	1.0000000
HH_Race3_W	193	0.9948187	0.0719816	0	1.0000000
HH_Race3_B	193	0.0051813	0.0719816	0	1.0000000
HH_Race3_H	193	0	0	0	0
HH_Race3_A	193	0	0	0	0
HH_Race3_O	193	0	0	0	0
Total_Pets_0	193	0.4196891	0.4947915	0	1.0000000
Total_Pets_1	193	0.5803109	0.4947915	0	1.0000000
Child_1	193	0.0259087	0.1592702	0	1.0000000
Child_2	193	0.0518135	0.2222267	0	1.0000000
Child 3	193	0.1088083	0.3122084	0	1.0000000

Child_4	193	0.0207254	0.1428340	0	1.0000000
Child_5	193	0	0	0	0
Child_6	193	0.0829016	0.2764505	0	1.0000000
Child_7	193	0	0	0	0
MONETARY	193	5.1584974	0.8888701	2.1400000	5.8800000
FREQUENCY	193	2.0000000	0	2.0000000	2.0000000
RECENCY	193	17.5544041	11.0622478	0	48.0000000

We divide the customers into 2 clusters:

- 1) These are high revenue generating customers in the top 20%. Also these customers are frequently visiting and have high recency on comparison with other segments. These are the best customers.
- 2) The other segment are those whose monetary value is in the bottom 20% and the frequency of visits is in the bottom 40%.

We want to find interesting patterns that are common and distinct across the two groups. Upon analyzing, we can see that the attributes demonstrate almost similar characteristics for both the clusters.

For the customers choosing **HELLMANNS** as their preferred brand, we can see that:

- The average monetary value of the high revenue customers is \$29 compared to low revenue customers which is only 5\$. The high revenue group also have a high frequency of around 8 and recency of 4 weeks. On the other hand, low revenue group has frequency of 2 and recency of around 17 weeks. Clearly there is a huge difference between the clusters.
- In both clusters, we can see that medium and high-income households contribute around 85-90% of the revenue. Low Income customers certainly do not prefer this brand indicating the price per unit is high.
- Another interesting insight is that the revenue of the company is mostly from small family households which consist of only 1-3 members around 63-67%. Large family sizes which have more than 3 members do not prefer our brand much owing to the high cost of the product. Buying for a big family means spending more for the product which these families do not prefer.
- Young males and females in the age group of 18-24 who are heading the household do not seem to like mayonnaise. Middle aged and older males and females are the ones who are contributing to the revenue.
- Another interesting thing is that married people are the significant contributors of revenue above 90%. Single, unmarried and divorced people do not buy our product. This maybe because married people cook food usually at home and hence are required to have mayonnaise as an ingredient while single and unmarried are free to eat outside often and do not have the necessity to buy mayonnaise.
- Among the employed people higher paid white-collar employees buy our product. This also suggests that the people are price sensitive as lower white collars and blue collars contribute less to the revenue. Also retired and unemployed people buy our product. Retired people are those who have a good brand sense and would have experience in choosing a good brand like ours in this case.
- Children in the age group 12-17 are the highest contributors in the children groups. Toddlers and small children apparently do like the taste of the product until they grow a bit older.

RECOMMENDATIONS:

- Our findings are not able to differentiate the exact reason for the purchase patterns of high and low revenue customers. Clearly income plays a major part. Our brand being a bit high priced is not able to attract the low-income families much. The low-income families must be targeted by giving specialized discounts to make them accustomed to our taste. Once they become addicted to our brand in a few months, they will be willing to pay the same price as everyone else. This can boost revenue to a huge extent.
- Another target area is the **age group of 18-24**. This group has a very high potential to boost revenue. Since this group uses lot of social media, **lot of targeted ads in Facebook and Instagram as well as emails with promotions can be sent to promote our brand** and attract this segment. This is an easily manipulatable group as they have wavering mindset and diverse opinions.
- Children in the age group 12-17 do not seem to like the taste of mayonnaise. New flavours with ingredients liked by the children can be introduced in the market. Samples can be initially tested out before a market release. Once the children start liking the product, the parents need to continuously keep buying in bulks to satisfy the adamant children and hence drive the sales.

ANALYSIS 2

Determine the effect of price, display, feature and promotion on the sales of Hellman's mayonnaise.

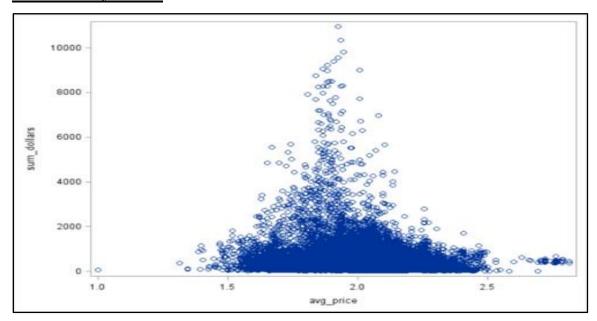
Data Preparation:

Standardized price per unit for a unit volume equivalent was calculated so that they can be compared with all other brands.

Total Volume= Volume Equivalent of a Product * Number of Units Sold

Standardized Price= Total Cost of a Purchase / Total Volume

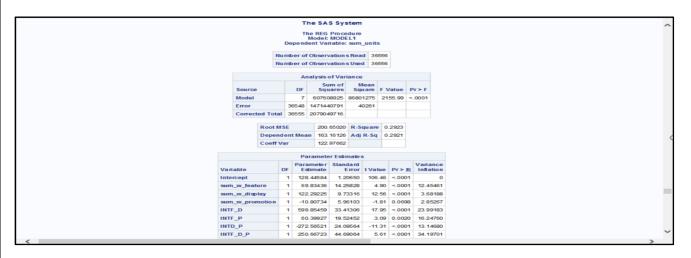
Non-Linearity Check:



The scatterplot above clearly shows the Non-Linear characteristics of Price. For this reason, we have also added a squared price term in our model.

Multicollinearity Check:

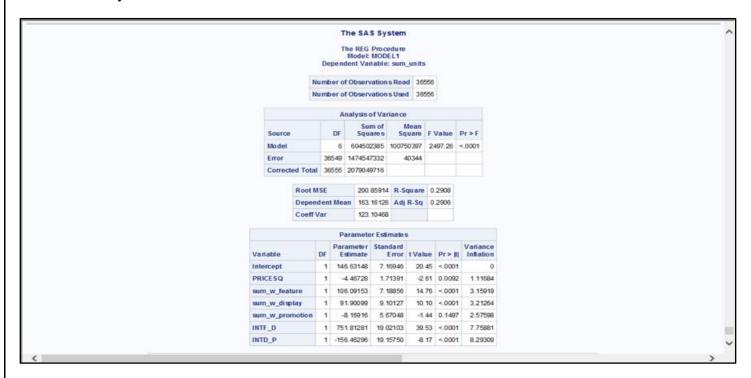
We observe that the VIF values are well above 10. Therefore, we cannot rule out multicollinearity in this model.



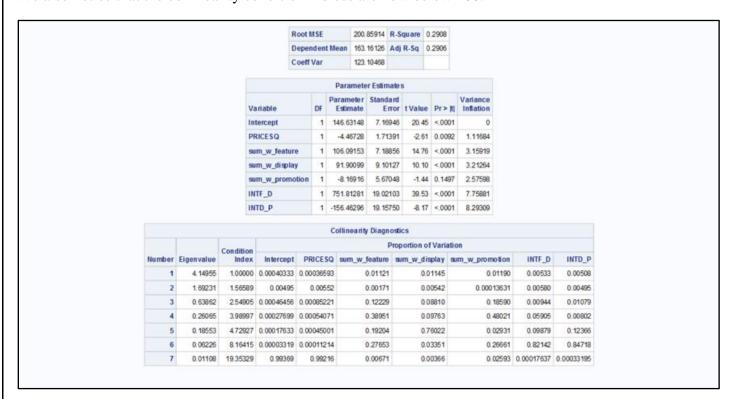
To correct this, we decided to remove two factors, as there is collinearity between the below-mentioned factors.

- 1. Interaction factor between feature, display and promotions.
- 2. Interaction factor between feature and display.

We can now see that the variation inflation factors drop below 10 and thus remove the problem of multicollinearity.



We also notice that the collinearity condition indices are now below 100.



Panel Regression:

We consider the Fixed effects model as according to the Hausman test, we go on to reject the null hypothesis and hence the fixed effects model holds true.

Dependent Variable: Total Units Sold/Week In a Store

Independent Variables: PRICESQ, Weighted Display, Weighted Feature, Weighted Promotion, Interaction between Display and Feature, Interaction between Display and Promotion.

Parameter Estimates							
Variable	DF	Estimate	Standard Error	t Value	Pr > t	Label	
Intercept	1	97.08544	23.3319	4.16	<.0001	Intercept	
PRICESQ	1	9.248387	2.9869	3.10	0.0020		
sum_w_feature	1	90.62983	5.3890	16.82	<.0001		
sum_w_display	1	59.62616	6.8826	8.66	<.0001		
sum_w_promotion	1	32.65932	4.5708	7.15	<.0001		
INTF_D	1	609.4995	13.8606	43.97	<.0001		
INTD_P	1	-85.4545	14.0035	-6.10	<.0001		

Effects of Price, Display, Feature and Price Reduction on Sales

PRICESQ: Nonlinear relationship between average price and number of units sold is seen from the estimates. Hence, we have considered PRICESQ which is the square of average price.

Weighted Display: As the display increases by 1% number of units sold per week in a shop increases by 59 units.

Weighted Promotion: As the promotion increases by 1% number of units sold per week in a shop increases by 32 units.

Interaction between Display and Feature: This produces a very strong synergistic effect in number of units sold per week in a shop.

Interaction between Display and Promotion: This produces a cancellation effect in number of units sold per week in a shop.

RECOMMENDATIONS:

- We can clearly see that the promotions and display together leads to a cancellation effect.
- This could be because, customers tend to not like the idea of being constantly pestered for the same product again and again.
- As a result, the customers start doubting the product and woo away from it.
- To stop this the display or the promotions one of the two needs to be subtle so that the apparent idea about the product is not harmed in the eyes of the potential customers.

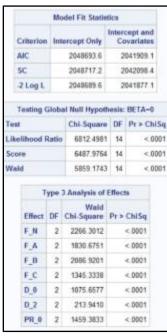
ANALYSIS 3

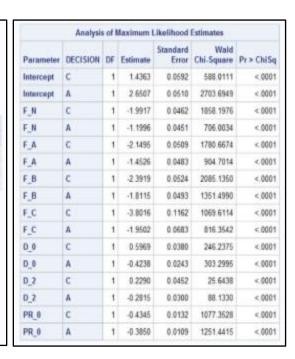
Our objective in this analysis to find the impact of factors such as Display (D), Feature (D) and Price Reduction (PR) on brand selection. We wish to know if these factors could make a considerable impact on the customer's purchasing decision. We choose to analyze customer brand preference of **Hellmann's** over Kraft Miracle Whip and Best Foods.

In order to achieve this, we carry out a multinomial logistic regression. We create a column called DECISION where, Kraft Miracle Whip products are assigned the variable **A**, Hellmann's as **B** and Best Foods as **C**. The dummy variables in the model are: Decision=B (Hellmann's), Feature= A+, Display=1 (Minor Display) and PR=1.

By running the multinomial logistic regression, we get the following output:









Interpretation:

When comparing Hellmanns to Kraft Miracle Wip and Best Foods, keeping all the other factors constant:

With Hellmans having Price Reduction grater then 5% and has a minor advertisement displayed with A+ level feature:

FEATURE:

The relative odds of selecting Best Foods when there is **no** feature is 0.136 and that of Kraft Miracle Wip is 0.301.

The relative odds of selecting Best Foods when there is **A** level feature is 0.11 and that of Kraft Miracle Wip is 0.23.

The relative odds of selecting Best Foods when there is **B** level feature is 0.09 and that of Kraft Miracle Wip is 0.16.3.

The relative odds of selecting Best Foods when there is C level feature is 0.022 and that of Kraft Miracle Wip is 0.14.

We find that the relative odds of not choosing Kraft brand and Best Food brand is more when Hellmanns has a price reduction with an A+ feature displayed. Hence having a A+ feature display has a considerable impact on Hellmans when the price is reduced.

DISPLAY:

The relative odds of selecting Best Foods when there is **no** display is 1.816 and that of Kraft Miracle Wip is 0.65.

The relative odds of selecting Best Foods when there is **major** display is 1.25 and that of Kraft Miracle Wip is 0.75.

We find that the effect of a minor display of Hellmanns when there is a price reduction does not increase the odds of selection when compared to Best Foods but has a significant impact when compared to Kradt Miracle Wip.

PRICE REDUCTION:

The relative odds of selecting Best Foods when there is **Price Reduction** is 0.64 and that of Kraft Miracle Wip is 0.68.

We can see that Price Reduction has a considerable impact on the odds of selecting Hellmans compared to the other two brands.

RECOMMENDATIONS:

When price reduction is introduced, it is advisable to go for a major display since a minor display does not significantly increase the odds of buying Hellmann's. An A+ level feature definitely increases the chances of a customer preferring the Hellman's brand. We also find that the effect of price reduction impacts the customers of Kraft Miracle Whip to switch to Hellmann's and does not have a huge impact on the customers of Best Foods. Hence it would be reasonable to introduce Price Reduction with a major display in stores where customers buy more Kraft Miracle Whip products.