PREDICTIVE ANALYTICS WITH SAS

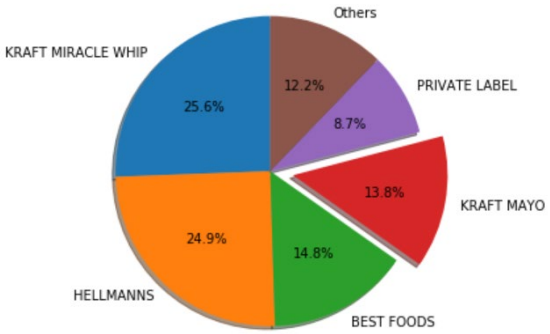
Marketing Insights for Kraft Mayo:

Sai Rahul Dhulipalla- sxd170033
Maruthi Nandan Pulletikurthi - mxp180010
Madhusudan Rudresh - mxr170019

DECEMBER 2

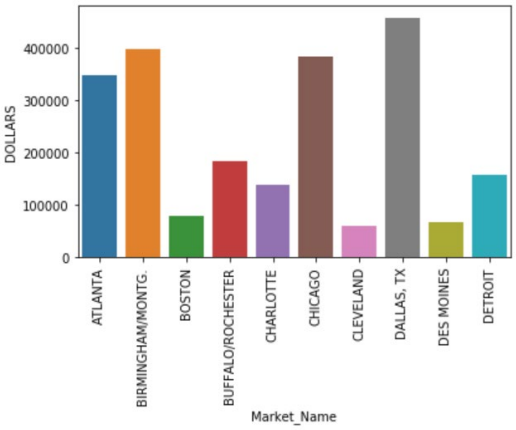
Group 6

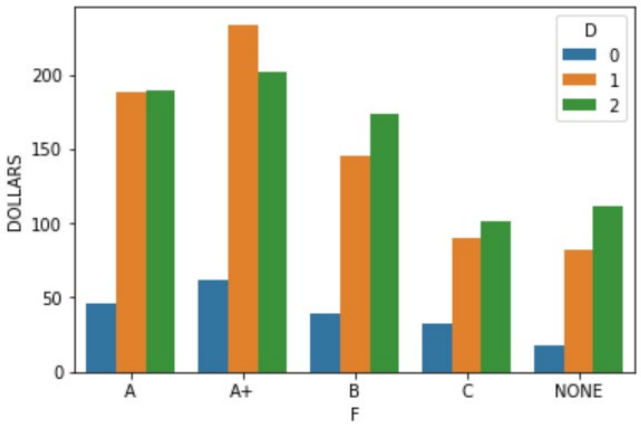
Exploratory Data Analysis of Kraft Mayo:



The Pie chart represents market share of top 6 brands. From the pie chart its evident that Kraft miracle whip (25.6%) has the highest market share followed by Hellmann's with 24.9%. We have chosen Kraft Mayo as our brand to suggest some recommendations which will improve sales for the brand.

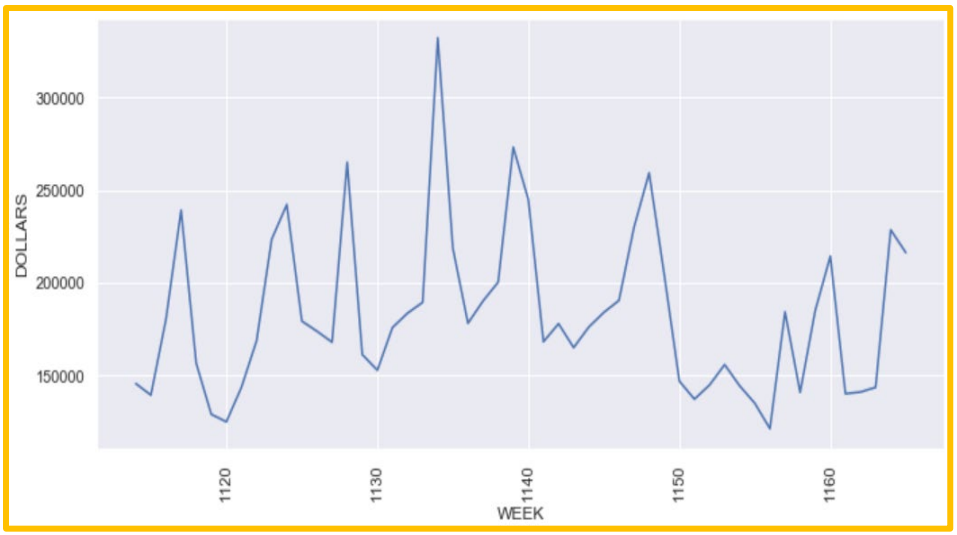
Dallas is the top Market for Kraft Mayo in terms of revenue followed by Birmingham and Chicago. It has good presence in all regions of US Market which could be put into use to improve sales uniformly throughout the States.





From the bar graph for Kraft Mayo, we can see the dollar sales value when we have both display and Feature. The maximum sales occurred when we have a rebate and retailer coupon feature along with Minor Display. In the overall feature and display interaction having a major display along with large ad featuring is giving out more dollar sales value compared to any other possible interaction.

The Line graph shows the weekly dollar sales of Kraft Mayo over the observed period. The maximum revenue was generated between the Weeks 1134 and the lowest in week number 1156.



1. Analysis on effect of Price, Featuring & Display on sales of overall Mayo and Kraft Mayo:

About the data: We have data on weekly sales of different mayonnaise UPC's at different stores for a year. The data is basically panel data with each store as different panel.

Panel Data Regression on Overall Mayo Sales:

To figure out the effect of price, featuring and display on dollar sales of Mayonnaise as a whole, we ran Panel data Fixed effects & random effects regression with each store as a different panel and select fixed effects after conducting Hausman test.

Since we have sales data at UPC level for every week at different stores, we had to calculate the average price per unit volume, average featuring & average display for the whole mayo sales in every week at different store. To find these average values we used the weighted average method with each UPC weekly sale weighted at the weight of each UPC yearly sales on all the UPC's yearly sales in specific stores.

The Model output can be found below.

Model Log_Total_Dollars = Avg_Price_Per_Unit_Vol Avg_F_A Avg_F_A_Plus Avg_F_B Avg_F_C Avg_D_1 Avg_D_2 Price_Inter_F_A Price_Inter_F_A_Plus Price_Inter_F_B Price_Inter_F_C

The SAS System

The PANEL Procedure
Fixed One-Way Estimates

Dependent Variable: Log_Total_Dollars

Model Description	
Estimation Method	FixOne
Number of Cross Sections	1450
Time Series Length	52

Fit Statistics			
SSE	3572.1901	DFE	66654
MSE	0.0536	Root MSE	0.2315
R-Square	0.8505		

F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
1449	66654	236.69	<.0001

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	6.891323	0.0475	145.11	<.0001	Intercept
Avg_Price_Per_Unit_Vol	1	-0.04384	0.0139	-3.14	0.0017	
Avg_F_A	1	2.169626	0.0632	34.31	<.0001	
Avg_F_A_Plus	1	1.416813	0.0920	15.40	<.0001	
Avg_F_B	1	2.030286	0.0860	23.60	<.0001	
Avg_F_C	1	0.773467	0.2052	3.77	0.0002	
Avg_D_1	1	0.426013	0.0109	39.10	<.0001	
Avg_D_2	1	0.649611	0.00887	73.26	<.0001	
Price_Inter_F_A	1	-0.945	0.0418	-22.61	<.0001	
Price_Inter_F_A_Plus	1	-0.49994	0.0543	-9.22	<.0001	
Price_Inter_F_B	1	-0.94332	0.0536	-17.58	<.0001	
Price_Inter_F_C	1	-0.28498	0.1300	-2.19	0.0284	

In this output,

Avg_F_A = Avg large size ad

Avg_F_B = Avg medium size ad

Avg_F_C = Avg small ad, usually 1 line of text

Avg_F_A_Plus = Avg retailer coupon or rebate

Avg_D1 = Display – MINOR

Avg_D2 = Display – MAJOR

Price_Inter_F_A, Price_Inter_F_A_Plus , Price_Inter_F_B , Price_Inter_F_C = price interacted with different levels of advertising.

The Insights we get from this model output are:

1) As we reduce the Price, sales will increase.
2) Featuring has a significant boost on Sales and different levels of featuring has different effect.
   - On an average, Dollar sale with Large Ad are 63% higher compared with no Ad.
   - On an average, Dollar sale with Medium Ad are 49% higher compared with no Ad.
   - On an average, Dollar sale with Small Ad are 30% higher compared with no Ad.
   - On an average sale with retailer coupon or rebate are 60% higher compared with no Ad.

3) Display's also boost the sales significantly.
   a) On an average, Dollar sale with Minor Display are 42% higher compared with no Display.
   b) On an average, Dollar sale with Major Display are 64.9% higher compared with no Display.
4) There are interaction effects between price and different levels of advertisements.
   a) On an average Dollar sale increases by 4% when the price per unit volume is decreased by $1 and there is no Advertising.
   b) On an average Dollar sale increases by 98% when the price per unit volume is decreased by $1 and there is a Large Ad.
   c) On an average Dollar sale increases by 98% when the price per unit volume is decreased by $1 and there is a Medium Ad.
   d) On an average Dollar sale increases by 32% when the price per unit volume is decreased by $1 and there is a Small Ad.
   e) On an average Dollar sale increases by 54% when the price per unit volume is decreased by $1 and there is a retailer coupon or rebate.

From the above Insights we can conclude that having Featuring and Display helps in increasing sales and if there is planning for price reduction, then giving medium size ad helps in getting the maximum benefit for low price of advertising.

Panel Data Regression on Kraft Mayo Sales:

To figure out the effect of price, featuring and display on dollar sales of Kraft Mayo we ran Panel data Fixed effects regression with each store as a different panel.

To calculate the average price per unit volume, Average Featuring & Average Display for the Kraft mayo sales in every week at different store we used the weighted average method used while we calculated for overall mayo sales.

We ran both the Fixed effects and random effects model and from Hausman test results we concluded that Fixed effects model is better.

The SAS System

The PANEL Procedure
Fixed One-Way Estimates

Dependent Variable: Log_Total_Dollars

Model Description	
Estimation Method	FixOne
Number of Cross Sections	1450
Time Series Length	52

Fit Statistics			
SSE	12824.9138	DFE	66419
MSE	0.1931	Root MSE	0.4394
R-Square	0.7964		

F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
1449	66419	150.46	<.0001

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	6.466214	0.0799	80.96	<.0001	Intercept
Avg_Price_Per_Unit_Vol	1	-0.71631	0.0134	-53.65	<.0001	
Avg_F_A	1	1.010839	0.0523	19.33	<.0001	
Avg_F_A_Plus	1	-0.05216	0.0852	-0.61	0.5404	
Avg_F_B	1	0.876763	0.0616	14.24	<.0001	
Avg_F_C	1	0.91776	0.1408	6.52	<.0001	
Avg_D_1	1	0.500123	0.0150	33.32	<.0001	
Avg_D_2	1	0.732442	0.0114	64.19	<.0001	
Price_Inter_F_A	1	-0.37522	0.0395	-9.50	<.0001	
Price_Inter_F_A_Plus	1	0.587742	0.0578	10.16	<.0001	
Price_Inter_F_B	1	-0.30329	0.0461	-6.58	<.0001	
Price_Inter_F_C	1	-0.31038	0.1039	-2.99	0.0028	

The Insights we get from this model output are:

1) As we reduce the Price, sales will increase.
2) Featuring helps boost Sales.
   - On an average Dollar sale with Large Ad are 40% higher compared with no Ad.
   - On an average Dollar sale with Medium Ad are 38% higher compared with no Ad.
   - On an average Dollar sale with Small Ad are 41% higher compared with no Ad.
   - On an average Dollar sale with retailer coupon or rebate are 95% higher compared with no Ad sales.
3) Display's also boost the sales significantly.
   - On an average Dollar sale with Minor Display are 50% higher compared with no Display.

- On an average Dollar sale with Major Display are 73.2% higher compared with no Display.

4) There are interaction effects between price and different levels of advertisements.
   - On an average Dollar sale increases by 71% when the price per unit volume is decreased by $1 and there is no Advertising.
   - On an average Dollar sale increases by 109% when the price per unit volume is decreased by $1 and there is a Large Ad.
   - On an average Dollar sale increases by 102% when the price per unit volume is decreased by $1 and there is a Medium Ad.
   - On an average Dollar sale increases by 102% when the price per unit volume is decreased by $1 and there is a Small Ad.
   - On an average Dollar sale increases by 12% when the price per unit volume is decreased by $1 and there is a retailer coupon or rebate.

From the above Insights we can conclude that

1) Retailer coupon or rebates provide higher boost of Dollar sales compared with all other advertising when there is no price reduction, so this method of advertising is best.
2) Large Ad, Medium Ad & Small Ad boost of Dollars sales are similar when there is no price reduction so we should go with Small Ad as it is more cost effective.
3) Major Displays gives higher boost in sales.
4) When we are planning to decrease the price to boost sales then placing Large Advertisements help getting maximum boost of Dollar Sales. We should not go with retailer coupon or rebate when decreasing the price because this way of advertising nullifies the effect of price reduction giving least benefit.

Effect of other competing brand Price, Featuring & Display on Kraft Mayo Sales:

To check the effect of competing brands price, featuring and Display on our brand, we filtered out the retail store sales of top 4 selling UPC from Kraft Mayo, HELLMANNS, KRAFT MIRACLE WHIP & PRIVATE LABEL. These UPC's account for almost 50% of sales in their respective brands. We ran panel data models on these data to figure out the required effects.

Panel Data Regression to Know the effect of competitors price:

model Log_KM_DOLLARS = KM_Price_Per_Unit_Vol KM_F KM_D KMW_Price_Per_Unit_Vol HELL_Price_Per_Unit_Vol PL_Price_Per_Unit_Vol

The abbreviations are given below for reference:
KM = Kraft Mayo
KMW = Kraft Whip Mayo
PL = Private Label
Hell = Hellmanns

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	4.583292	0.1069	42.88	<.0001	Intercept
KM_Price_Per_Unit_Vol	1	-0.31341	0.0230	-13.62	<.0001	
KM_F	1	0.622567	0.0120	51.76	<.0001	
KM_D	1	0.487086	0.00899	54.21	<.0001	
KMW_Price_Per_Unit_Vol	1	-0.28847	0.0172	-16.75	<.0001	
HELL_Price_Per_Unit_Vol	1	0.349977	0.0162	21.63	<.0001	
PL_Price_Per_Unit_Vol	1	-0.01903	0.0201	-0.95	0.3427	

The Insights we get from this model output are:

1) When Hellmann's reduce their price by $1 our sales decreases by 35% approximately. This effect is very high. To reduce our loss, we should reduce our price too during that time.
2) There is no effect on Kraft Mayo due to price changes from Private Label.

Panel Data Regression to Know the effect of competitors Featuring and Display:

model Log_KM_DOLLARS = KM_PR KM_F KM_D Competitors_F Competitors_D Competitors_PR KM_F_Inter_Competitors_F KM_D_Inter_Competitors_D KM_PR_Inter_Competitors_PR

Competitors Display, featuring = 1 if any of Hellmann's, Private label has display or any featuring.

The Insights we get from this model output are:

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	4.282534	0.0976	43.89	<.0001	Intercept
KM_F	1	0.849533	0.00935	90.87	<.0001	
KM_D	1	0.542602	0.0101	53.66	<.0001	
Competitors_F	1	-0.12087	0.00798	-15.16	<.0001	
Competitors_D	1	-0.06191	0.00914	-6.77	<.0001	
KM_F_Inter_Competitors_F	1	0.021301	0.0241	0.88	0.3777	
KM_D_Inter_Competitors_D	1	-0.06664	0.0190	-3.50	0.0005	

1) When Competitors Features any advertisement Kraft sales go down by 12% on average compared to when there is no competition.
2) When Competitors Has Display, Kraft sales go down by 6% on average compared to when there is no competition.
3) There is interaction effect between Kraft Mayo display and Competitors display. When Both Kraft Mayo and competitors both have a display, the sales of Kraft go down by 6% more than when there is only competitors feature. From these results we can conclude that Kraft Mayo to get full benefit from featuring and display, they should do it when competitors are not doing featuring and display. To combat the reduction in sales due to competitors Featuring and Display we should be doing featuring and display.

Recommendation for Kraft Mayo from the Analysis of Price, Display & Featuring:

- *Kraft Mayo should be doing Featuring and Display's frequently to boost their sales. While doing Featuring and Display they should follow the below guide lines.*
  1) *Retailer Coupon or rebates provide higher boost of Dollar sales compared with all other advertising when there is no price reduction, so this method of featuring is best when there is no price reduction.*
  2) *Larger Ad, Medium Ad & Small Ad effect have same magnitude of effect on Dollars sales when there is no price reduction so we should go with Small Ad as it is more cost effective.*
  3) *Major Displays should be used as they give higher boost in sales.*
  4) *When we are planning to decrease the price to boost sales then placing Large Advertisements help getting maximum boost on Dollar Sales. We should not go with retailer coupon or rebate when decreasing the price because this way of advertising nullifies the effect of price reduction giving least benefit.*
  5) *Kraft Mayo to get full benefit from featuring and Display we should do it when competitors are not doing featuring and display.*
  6) *To Combat the reduction in sales due to competitors Featuring and Display we should be doing featuring and display at the same time.*
- *When Competitor brand Hellmann's reduce their price Kraft sales get effected by 34% approximately. To reduce our loss, we should reduce our price too during that time and do featuring to combat that effect.*

2. Analysis on effect of customer demographics on Brand Choice:

To check how demographic variables such as household income, Age, education, number of kids, number of cats and dogs effect brand choice we ran Conditional logit on all the household purchase history data of top 4 brands (Kraft Mayo, Kraft Miracle Whip, Hellmann's, Private Label).

The output of Conditional Logit with Kraft Mayo as Base brand(Brand-1), Brand-2 as Kraft Miracle Whip, Brand – 3 as Hellmanns, Brand – 4 as Private Label can be found below.

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Income2	1	-0.002637	0.0124	-0.21	0.8313
Income3	1	-0.007455	0.0123	-0.60	0.5456
Income4	1	-0.1208	0.0144	-8.39	<.0001
Family_Size2	1	-0.1319	0.0319	-4.13	<.0001
Family_Size3	1	-0.0911	0.0317	-2.88	0.0040
Family_Size4	1	-0.0864	0.0375	-2.30	0.0214
HH_AGE2	1	0.009839	0.0338	0.29	0.7708
HH_AGE3	1	0.007475	0.0337	0.22	0.8243
HH_AGE4	1	-0.2295	0.0389	-5.90	<.0001
HH_EDU2	1	-0.0584	0.0233	-2.51	0.0122
HH_EDU3	1	-0.0871	0.0233	-3.74	0.0002
HH_EDU4	1	-0.1050	0.0282	-3.73	0.0002
No_Dogs2	1	-0.0308	0.0429	-0.72	0.4727
No_Dogs3	1	0.0423	0.0425	1.00	0.3192
No_Dogs4	1	0.1559	0.0492	3.17	0.0015
No_Cats2	1	-0.0176	0.0356	-0.50	0.6205
No_Cats3	1	0.0475	0.0352	1.35	0.1767
No_Cats4	1	0.0349	0.0414	0.84	0.3984
No_TVs2	1	-0.0330	0.0219	-1.51	0.1309
No_TVs3	1	-0.0646	0.0218	-2.96	0.0031
No_TVs4	1	-0.0320	0.0262	-1.22	0.2212
Kids_below_6_2	1	0.1129	0.1666	0.68	0.4979
Kids_below_6_3	1	-0.0911	0.1667	-0.55	0.5848
Kids_below_6_4	1	0.0347	0.1906	0.18	0.8554
childeren_6_11_2	1	0.3531	0.1169	3.02	0.0025
childeren_6_11_3	1	0.5036	0.1155	4.36	<.0001
childeren_6_11_4	1	0.3319	0.1325	2.50	0.0123
childeren_12_17_2	1	-0.004111	0.0964	-0.04	0.9660
childeren_12_17_3	1	0.0528	0.0955	0.55	0.5801
childeren_12_17_4	1	0.1279	0.1116	1.15	0.2520
intercept2	1	2.4942	0.2384	10.46	<.0001
intercept3	1	2.6234	0.2376	11.04	<.0001
intercept4	1	3.0224	0.2741	11.03	<.0001

From the above output, we can conclude that

- With increase in income of household there is no much difference in brand choice of 1,2 or 1,3 but between brand 4 and brand 1, with increase in income people tend to choose brand 1(kraft Mayo) over brand 4(Private Label).
- With Family size seeing all coefficients negative we can say with increase in family size there is a pattern of choosing our Kraft Mayo over other brands.
- Age also follows the same pattern as of income, no significant difference in brand choice between 1,2 or 1,3 with age but increase in age decreases the loyalty towards brand 4 with respect to brand 1(Kraft Mayo)
- With increase in Education of household, we can say the choice of household choosing our Kraft Mayo over their brand is increasing.
- Number of dogs is not significant between brand choice of 1,2 or 1,3 but in case of brand choice between 1,4 with increase in number of dogs the probability of them choosing brand 4 is increases.
- Number of cats in a household has no effect over the brand decision.
- Households with kids below 6 and households between 12-17 have no effect on brand choice.

- But For Number of children between 6-11, increase in the number increases the probability of household choosing that brand with respect to our Kraft mayo.

Recommendations from Analysis of Household Data on brand choice:

- *For the households between 6-11, they are choosing other brands compared to ours, we must spend more targeted ads to those households. And households with kids age between 12-17 are not caring about the brands, so we can reach out to those households and can turn them into our loyal customers.*

3. Analysis of how purchasing frequency of mayonaisse varies based on customer demographics using survival analysis (Lifetest and Phreg):

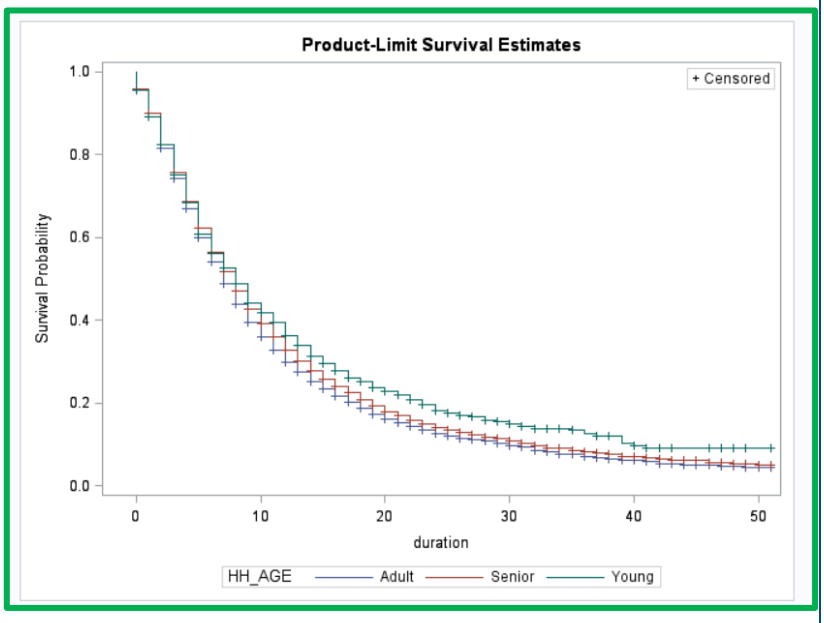
Survival Analysis – Lifetest:

- Additional variables were created such as start (week of purchase), end (week of next purchase or that week itself if it's last transaction), duration (end-start), censor (1 = the last transaction of every panid, else 0).
- The following categories were regrouped
  - Income: Low=0-44,999, Medium=45,000-74,999, High=75,000+
  - Family_size: Small=1-2, Medium=3-4, Large=5+
  - HH_Age: Young=18-34, Adult=35-54 , Senior= 55+
- Shortlisted 12 variables on the dataset and they are as follows: PANID, L5, censor, Income, Family_size, HH_Race, HH_Age, HH_EDU, HH_OCC, F, D, PR, duration, total_ounces

Model Interpretation

1. Life test was conducted on the dataset with HH_Age as strata, time variable as duration and censoring variable as censor. There are 3 groups in this HH_Age variable; Young=18-34 , Adult=35-54 , Senior= 55+

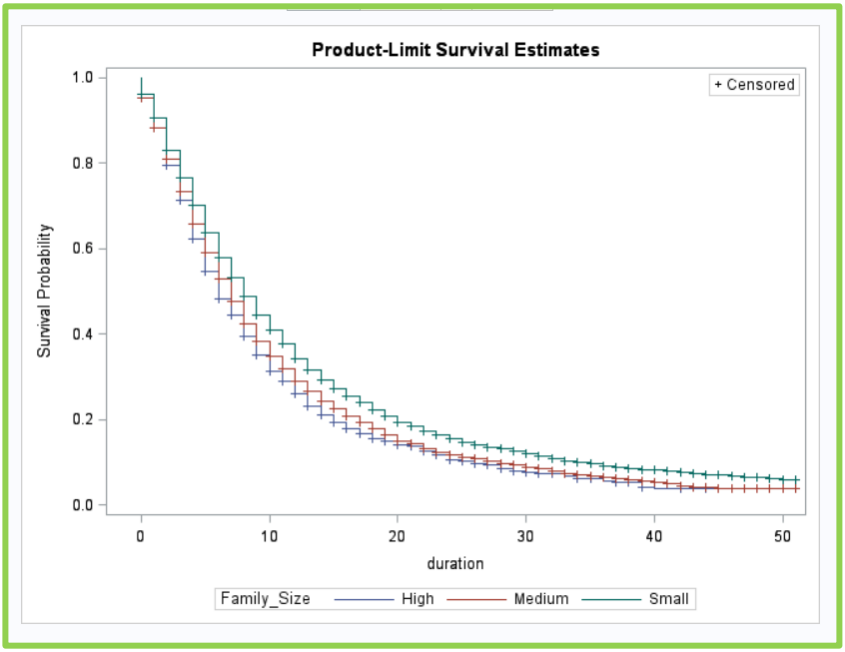
Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	29.0695	2	<.0001
Wilcoxon	19.3992	2	<.0001
-2Log(LR)	32.9992	2	<.0001



Both Log-Rank (p-value < 0.0001) and Wilcoxon (p-value < 0.0001) confirms that there is significant difference between the strata groups considered for the analysis. We see that survival probability for the Adult and Senior population is less indicating that their frequency of buying mayonaisse is higher compared to the young population.

2. Life test was conducted on the dataset with Family_size as strata, time variable as duration and censoring variable as churn. There are three groups in this Family_size variable; Small: 1-2 people, Medium: 3-4 people, Large: 5+ people

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	118.4778	2	<.0001
Wilcoxon	107.9061	2	<.0001
-2Log(LR)	131.7447	2	<.0001



Both Log-Rank (p-value < .0001) and Wilcoxon (p-value < 0.0001) confirms that there is significant difference between the strata groups considered for the analysis. We see that Medium and High family size population have a lower survival probability indicating that their frequency of buying mayonaise is higher than the small family size population.

Survival Analysis – PHREG:

We have used Proportional Hazard as part of Survival Analysis to determine the importance of covariates and the impact of each of the features on hazard of coming back to buy frequently. We have modeled for the censor variable (1) for each other panelists.

We have obtained the following results:

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	280838.76	280606.76
AIC	280838.76	280634.76
SBC	280838.76	280741.74

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Income	2	9.6540	0.0080
Family_Size	2	101.2388	<.0001
HH_RACE	1	2.0451	0.1527
HH_AGE	2	20.9850	<.0001
F	4	68.2725	<.0001
D	2	0.5383	0.7640
PR	1	3.8644	0.0493

The model statistics provides us with information to compare the fit of two different models. The AIC and BIC of the model with Covariates is 280634.76 and 280741.74. The null Log likelihood of the model is 280606.76. Type 3 test below states the significance of overall feature and determine the contributing demographics of panelists in determining the churn of the model.

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Income	High	1	-0.07663	0.02492	9.4527	0.0021	0.926	Income High
Income	Low	1	-0.01874	0.01904	0.9690	0.3249	0.981	Income Low
Family_Size	High	1	0.25462	0.02995	72.2778	<.0001	1.290	Family_Size High
Family_Size	Medium	1	0.17209	0.02047	70.6908	<.0001	1.188	Family_Size Medium
HH_RACE	1	1	-0.10963	0.07666	2.0451	0.1527	0.896	HH_RACE 1
HH_AGE	Adult	1	0.15699	0.04331	13.1401	0.0003	1.170	HH_AGE Adult
HH_AGE	Senior	1	0.19706	0.04381	20.2283	<.0001	1.218	HH_AGE Senior

The variables income, Family size, HH_age, F, PR suggest some differences among their respective classes which might be key in evaluating strategies for that panelists representing that group. Whereas age does not seem to be significant in determining with Kraft mayo.

The significant parameters in determining the hazard of buying of the customer staying with the credit card company are Income, Family Size, HH_race, HH_Age, Feature and Price Reduction.

Insights and Recommendations:

- *As compared to Medium Income category which we have created for salaries ranging from ($45,000-$75,000) the hazard of repurchasing is lower in case of High-income households by 7.4% ((1-0.926) ∗100) and in case of Low-Income category its 1.9% ((1-0.981) ∗100) lower. Thus, we should target middle income category people.*
- *As compared to Small Family Size having a Large family, the hazard of purchasing again increases by 29%, and in case of medium sized family it increases by 18.8%. Thus, we should target more on middle and large sized family.*
- *As compared to Hispanic Community, the hazard of repurchasing is 10.4% lower for non-Hispanic. Thus, we should target Hispanic community more.*
- *The hazard of repurchasing is 17% higher for Adults and 21.8% higher for seniors compared with young populations. Therefore, we should create a retention strategy for older population as they are more frequent mayonnaise buyers.*