**Summary**

|  |  |
| --- | --- |
| PREDICTIVE ANALYTICS WITH SAS | Marketing Insights for Prego Spaghetti Sauce  **GROUP 5**  Sajal Singh  Aditi Prakash  Amanjit Singh  Nitasha Bansal  Varadharajan  Dilip Merala |

Recommendations for Prego:

Target Promotions:

* Emphasize more on Major displays rather than medium advertisement. It might be a cheaper option compared to medium size advertisement and provides better returns in terms of sales.
* Price rebates and coupon distributions should be done more often since it will increase our customer base and retain our existing customers.
* Spend more on large scale advertisement rather than small or medium since it increases sales by more than two times as compared to small scale advertisement.

Targeted Households:

* Household with Female heads of age 25-34 and 45-54 years. They are the most loyal customers for Prego.
* Target household with teen age children say 13-17 years as they have more affinity towards Prego brand of sauce.
* Household with Male head who has completed some college, Technical school education and also High school.
* Target family, whose combine d pre-tax income more than 100K as they are more loyal towards a brand mainly Prego.
* Crete a retention strategy for retaining more customers especially household with 5+ customers.

The loyalty of the customer was scored using R&M model where R+M>8 and R&M score > 4 are considered as Loyal Customers.

1. **Effect of Advertising on Sales of Prego:**

* When the **price** **increase** sale per **16 ounce of Prego** will decrease by **850%** approx.
* When the **Price of RAGU increase by 1$** sales per 16 Ounce increase by **26.72% approx**. for PREGO.
* If the PREGO is on large ads compared to no adds then **sales per 16 ounce** of spaghetti sauce will increase **by 475%** approx.
* If **PREGO is on price reduction** compared to no reduction, then sales per 16 ounce of spaghetti sauce will **increase by 12.62%** approx.
* If **PREGO is put on major display** compared to no display the sales per 16 ounce will go **up by 300%.**
* If **PREGO is put on minor display** compared to no display the sales per 16 ounce will go up by **234**%
* If the **Brand is on medium** ads compared to no adds then sales per 16 ounce of spaghetti sauce will increase by **250% approx**.
* If **the Brand is on small ads** compared to no adds then sales per 16 ounce of spaghetti sauce will increase **by 155% approximately**.

1. **Customer Behaviour of our Brand Prego**

* Households with **female** heads between the age of **25-34 years** are **1.533** times more likely to be loyal customers as compared to households with no female heads
* Households with **female heads** between the age of **45-54** years are **1.487** times more likely to be loyal customers as compared to households with no female heads
* The household with Children of age **group 0-5** are **0.467** times less likely to be loyal towards a brand when compared to a household with no children.
* Household with **male** **head** education level of **Technical** **school** are **1.411** times more likely to be loyal towards Prego when compared to households with no Male heads.

Household with male **head** education level of **Some** **College** are **1.383** times more likely to be loyal towards Prego when compared to households with no Male heads

1. **When are customers likely to churn out from our Brand Prego?**

When compared to cumulative brands, Prego has a lesser survival probability based on the family size of the customers. Prego needs to device a new retention strategy to reduce the churn and concentrate on Large families as they form more percentage of loyal customers.

**Overview**

* We selected Prego as our brand
* Linear regression model was used to determine the effect of advertisement on Sales.
* To find the characteristic of loyal customers we used RFM modelling and logistic regression.
* To predict how many customers will churn or switch from Prego, we performed Survival Analysis.

**Recommendations for Prego:**

**Target Promotions:**

* Emphasize more on Major displays rather than medium advertisement. It might be a cheaper option compared to medium size advertisement and provides better returns in terms of sales.
* Price rebates and coupon distributions should be done more often since it will increase our customer base and retain our existing customers.
* Spend more on large scale advertisement rather than small or medium since it increases sales by more than two times as compared to small scale advertisement.
* When product on both display and price rebate, the focus should be on Minor display and Price Rebate than Major Display and Price Rebate.

**Targeted Households:**

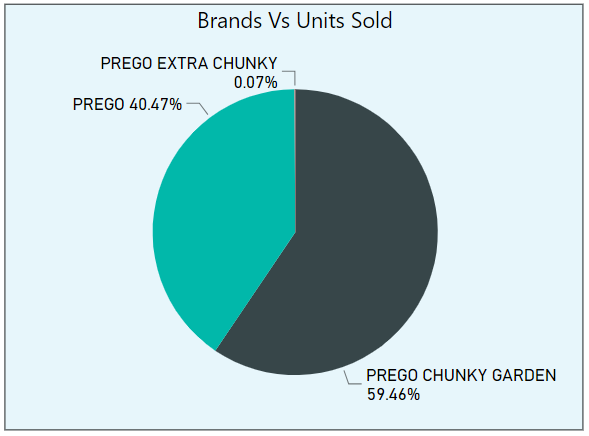
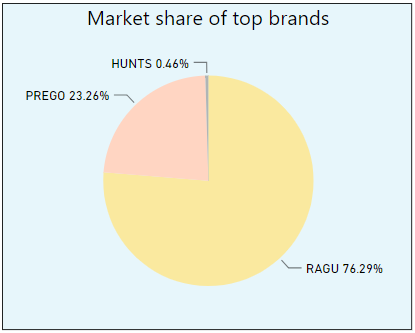
* Household with Female heads of age 25-34 and 45-54 years. They are the most loyal customers for Prego.
* Target household with teen age children say 13-17 years as they have more affinity towards Prego brand of sauce.
* Household with Male head who has completed some college, Technical school education and also High school.
* Target family, whose combine d pre-tax income more than 100K as they are more loyal towards a brand mainly Prego.
* Crete a retention strategy for retaining more customers especially household with 5+ customers.

**Retain Customers:**

* The one recommendation to Prego is try concentrating more on larger families as they are backbone for overall market. Using RFM and logistic we could see that they form a higher percentage of loyal customers. Prego must device a retention program to retain such customers who are potential to give a huge revenue.

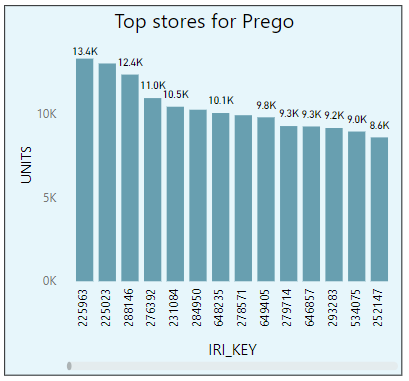
**Exploratory Data Analysis**

**Market Share and Top Brand for Prego:**



* Prego is second most popular brand in terms of market share, constituting to 23.26% of the total market share.
* Within the Prego brand, Prego Chunky Garden is the most popular brand in terms of units sold.

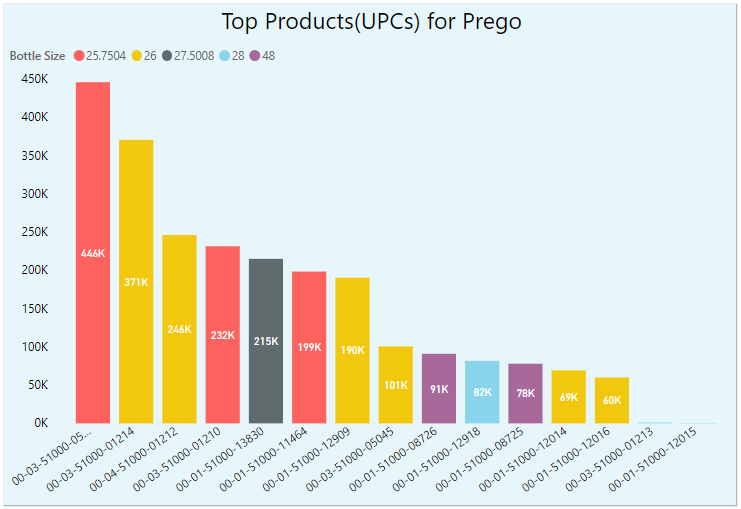
**Top Stores and Market for Prego:**





* Store 225963 sold the highest number of units of Prego products, 13.4K units.
* New York, Los Angeles, New England and Chicago are the top 4 market places for Prego.

**Top products for Prego (in terms of total units sold):**



* Product UPC 00-03-51000-05044, which translates to “PCHG1 GARCB SPGHT PSSC 25.75OZ”, is the most popular product for Prego.
* It is followed by UPC 00-03-51000-01214, which translates to “PCHG1 TMGON SPGHT PSSC 26OZ”, is the second most popular product for Prego.
* Both top 2 products are of same brand in Prego, that is, Prego Chunky Garden.

**Advertising Effect on Sales of PREGO:**

**Weighted components:**

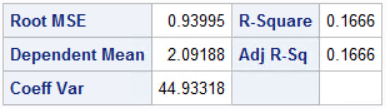
In order to identify the effect of brand on Sales, Weighted price is calculated using the market share of each brand.

**Market share** = Total Sales of Brand Total Sales

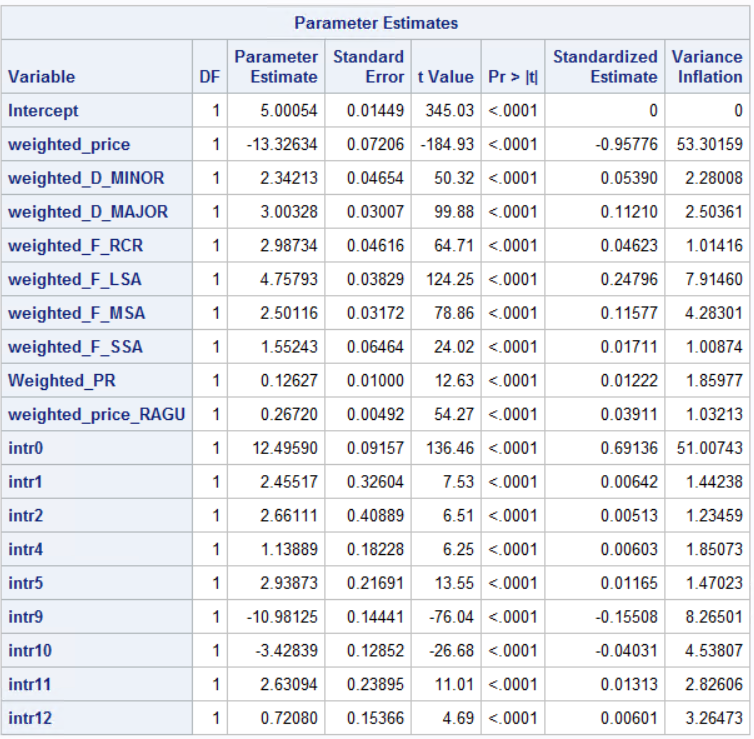
**Weighted price** = Dollars \* Market Share Units\*Volume Equivalent

**Regression on log(sales) without interaction and diminishing Variables:**

* The regression R2 is 0.1666 with all the coefficient are significant at 5% level
* Since all the vairables are within the VIF <10 and Collin Index <100. There is no multicollinearity.



**Regression on log(Sales) with interaction and diminishing variables:**

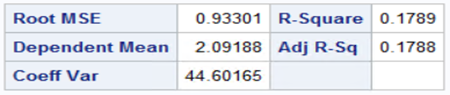


* After running regression on log sales with interaction

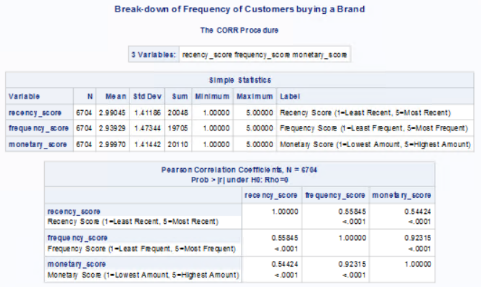
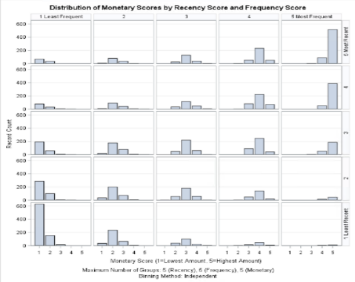
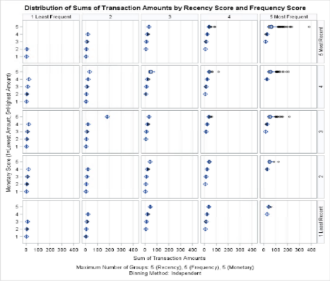
and diminishing variables the R2 increase to 0.1788

* All the parameters are significant and even running the F-test shows the interaction variables are significant
* Intr0 -weighted\_price^2
* Intr1-Minor Display \* Large Size Ads
* Intr2-Minor Display \* Medium Size Ads
* Intr4-Major Display \* Large Size Ads
* Intr5-Major Display \* Medium Size Ads
* Intr9-Price Rebate \* Large Size Ads
* Intr10-Price Rebate \* Medium Size Ads
* Intr11-Price Rebate \* Minor Display
* Intr12-Price Rebate \* Major Display

**Interpreting Significant Coefficients :**

* When the price increase sale per 16 ounce of Prego will decrease by 850% approx.
* When the Price of RAGU increase by 1$ sales per 16 Ounce increase by 26.72% approx. for PREGO.
* If the PREGO is on large ads compared to no adds then sales per 16 ounce of spaghetti sauce will increase by 475% approx.
* If PREGO is on price reduction compared to no reduction, then sales per 16 ounce of spaghetti sauce will increase by 12.62% approx.
* If PREGO is put on major display compared to no display the sales per 16 ounce will go up by 300%.
* If PREGO is put on minor display compared to no display the sales per 16 ounce will go up by 234%
* If the Brand is on medium ads compared to no adds then sales per 16 ounce of spaghetti sauce will increase by 250% approx.
* If the Brand is on small ads compared to no adds then sales per 16 ounce of spaghetti sauce will increase by 155% approx.

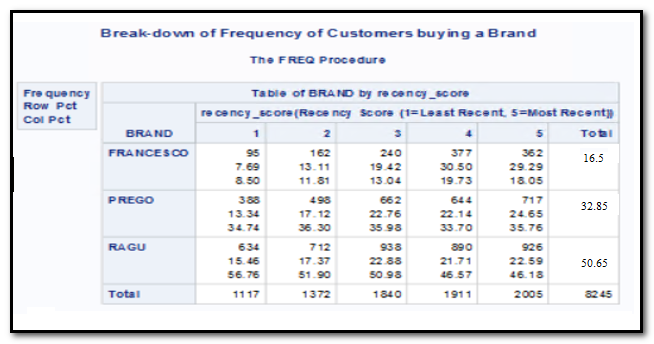
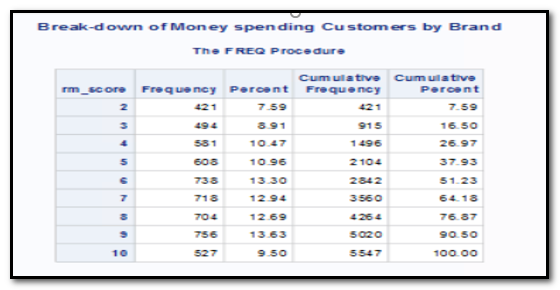
**Performing RFM to identify Loyal Customers**

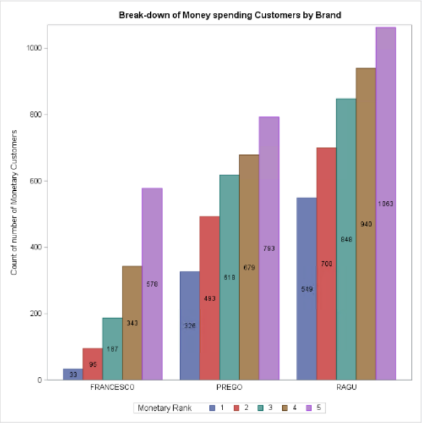
We are trying to find Loyal customers for the top 3 brands Ragu Prego and Francesco using RFM analysis.

From the Correlation between R, F & M we found that F and M are highly correlated, and we are selecting only R and M as our important components in identifying customer loyalty.

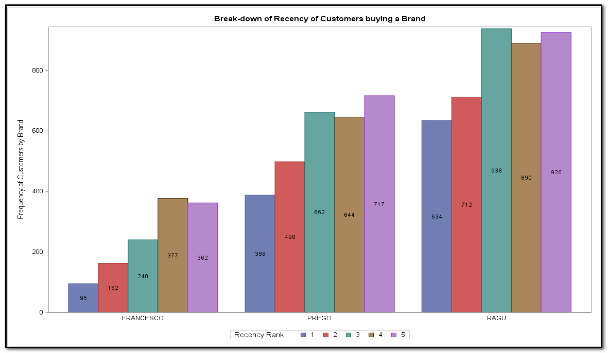
Looking at the distribution of monetary score by recency and frequency score, individuals are grouped together towards left and right extreme ends. The distribution of the Sum of Transactions amount by Recency & Frequency is accumulated towards the right. (frequency is considered only for analytical purpose and not included in our further analysis) Assigning the highest value of 5 to each group, we ranked customers based on the sum of individual scores for the two factors (R & M). Those who scored 8 or above on with a minimum of 4 for each - recency and monetary - were considered loyal.

The below Frequency comparison on brand shows that Raghu has the highest loyal customer with a share of 50%. Our brand Prego is popular among 33% of loyal customers.



From the break down of the RM scores, we could see that a total of 1987 customer is loyal out of 6704 customers which accounts for 30% of total customers.

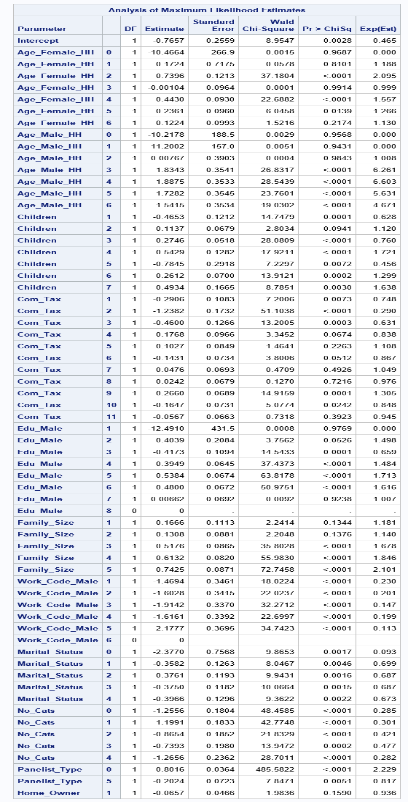
From the break down of Money spending customer by brand, we could see that Ragu has the highest number of 5 ratings followed by Prego and

From the Recency of customers buying a Brand, also, we could see Ragu has the highest number. Interesting Ragu has a lot of customers with Recency score of 3, which tolls a little higher than the score of 5.

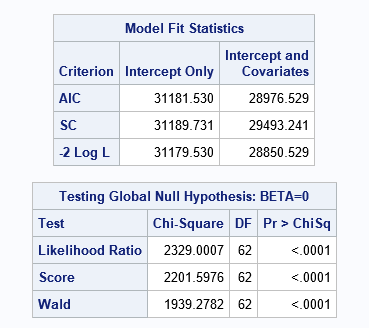
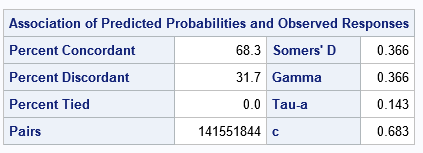
Insights: Ragu has the highest percentage of loyal customer and they are the top selling brand. This analysis gives only the customer loyalty and to do a quick check on their behavioural trend, we need to analyse the demographics of the customers

**Customer loyalty traits using logistic regression**

Based on the results obtained from RFM, we are determining the loyalty traits of customers using a logistic regression on customer demographics. In this, we are taking the loyalty as a categorical variable and running a logit model against customer demographics. Also, we are running a separate model for Prego to identify the customer who are loyal towards Prego.

We applied economic theory, subject knowledge and utilized the assistance of statistical selection models like stepwise that could predict whether a customer would be loyal or not.

**Meaning of Coefficients: Considering ceteris paribus**

* For the households with female family heads in age group 25-34, the odds of a customer being loyal increases by **109**% compared to households with no female heads.
* For the households with female family heads in age group 45-54, the odds of a customer being loyal increases by **57**% compared to households with no female heads.
* For the households with male family heads in age group 35-44, the odds of a customer being loyal increases by **526**% compared to households with no male heads.
* For the households with male family heads in age group 45-54, the odds of a customer being loyal increases by **560**% compared to households with no male heads.
* If the household has kids in age group 0-5, the odds of being loyal towards a brand decreases by **37.2**% compared to the family who do not have any kids.
* Household having a combined salary < $9,999 per year the odds of being loyal towards a brand decreases by **25**% when compared to households whose salary is >$100K
* Household having a combined salary < $11,999 per year the odds of being loyal towards a brand decreases by **71**% when compared to households whose salary is >$100K
* The household with a family size of five people, the odds of being loyal towards a brand increases by **110.1**% compared household with six or more people, keeping all other variables constant
* The household with a family size of three people, the odds of being loyal towards a brand increases by **67.8**% compared household with six or more people, keeping all other variables constant

**Interpreting fit of the model using -2logL, AIC and SC**

Model with the smallest AIC, SC and -2 Log L is considered the best model. Logistic Regression with all possible demographics which we obtained after removing few redundant and irrelevant demographics.

Intercept value of 2 Log L= 31179.530 (without X variables)

Intercept and covariates value of 2 Log L = 28850.529 (with X variables)

So here the improvement of model = McFadden’s R-square

= diff. in (-2LogL)/Null model’s (-2logL)

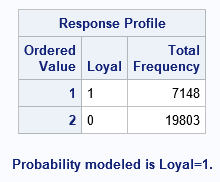
= (31179.530 - 28850.529)/ 31179.530 = 7.46%

= Therefore, R-square value is 7.46% which signifies improvement in the model

Compared to Intercept only (Null model i.e no X variable) the AIC and BIC Decreased for intercept and Covariates. The Best Model is with lower AIC and BIC. Compared to the Null model, the model with explanatory variables have reduced the AIC and SC(BIC) which shows there is an improvement in the model.

% of Concordance: Concordance tells how the model is when paired with events than with that a no events and how good it is predicting within the model

The value of predicting the random pair of loyal customers than not loyal customer is 68.3%. This means when running the model with new data set we can at least expect accuracy of 68.3% in prediction.

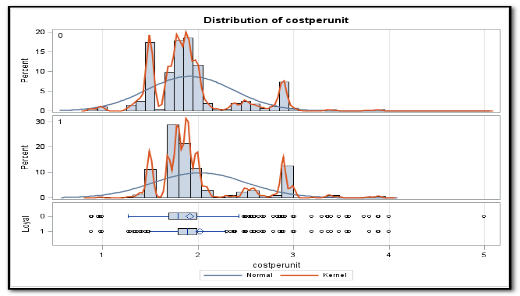
**Naive Ratio**: 7148/ (7148 + 19803) = 26.52%

Number of Correct Prediction= 20128

Hit Ratio is No. of correct prediction/ (Total predictions)

**Hit ratio** =20128/26951= 74.68%

There is an improvement of 48.16% from the naive ratio.

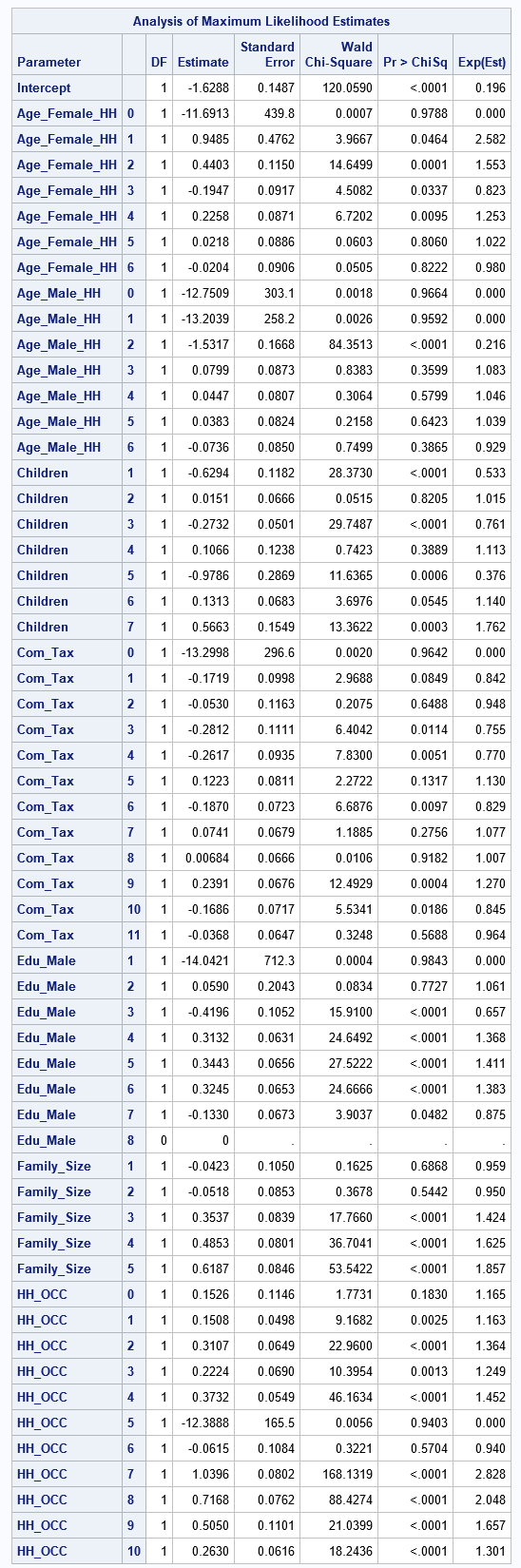
The **Area under Curve** also shows that the model correct prediction 68.3%

To understand the difference between the monetary benefit we received from loyal and non-loyal customers. As, we expected the loyal customers bring in more revenue per sale with a approximate mean of $2.1 and non-loyal customer with an approximate mean of $1.90. We performed a T-test to identify the statically significance and found the difference is highly significant. With respect to unequal variances, we found that the mean was different from each other. Hence, for any brand to make the customer to become loyal, it has a leverage of spending $.20 per unit sold. If the brand is spending more that this amount, they are less likely to get more gains out of it.

**Analysing the Customer Behaviour of Prego Brand:**

To find the what factors are affecting the loyalty of a customer towards Prego brand, we ran a simple logistic regression model on our data with brand Prego alone. The overall output was like that of Loyalty behaviour of households, but in this case, we were able to get a deeper insight for Prego to adopt a better marketing strategy to reach out right customer segment to get maximum benefit.

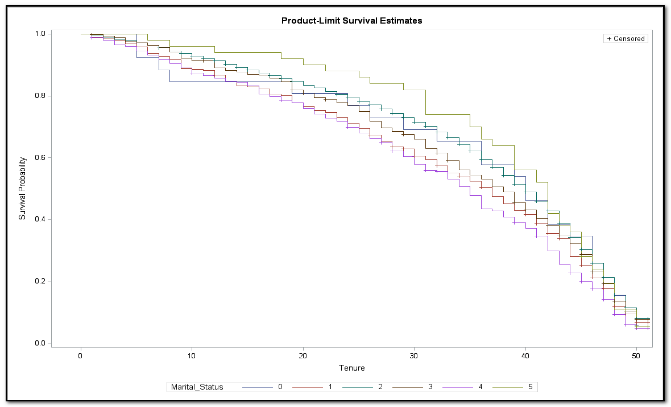
**Meaning of Coefficients: Considering ceteris paribus**

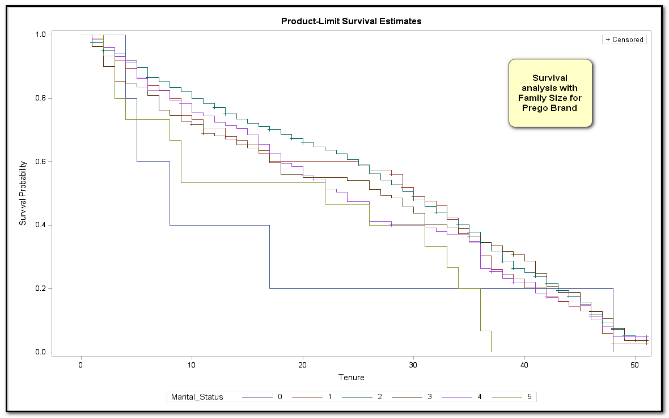
* Households with female heads between the age of 25-34 years are 1.533 times more likely to be loyal customers as compared to households with no female heads
* Households with female heads between the age of 45-54 years are 1.487 times more likely to be loyal customers as compared to households with no female heads
* The household with children of age group 0-5 are 0.467 times less likely to be loyal towards a brand when compared to a household with no children.
* The household with children age group 6-17 are 1.140 times more likely to be loyal toward a brand when compared to a house with no children.
* Household with a combined salary of 55K – 64.9K are 1.270 times more loyal than household having combined salary more than 100K
* Household with male head education level of Graduated high school are 1.368 times more likely to be loyal towards Prego when compared to households with no Male heads.
* Household with male head education level of Technical school are 1.411 times more likely to be loyal towards Prego when compared to households with no Male heads.
* Household with male head education level of Some College are 1.383 times more likely to be loyal towards Prego when compared to households with no Male heads.
* The Households with family size 3,4,5 are 1,424, 1.625, 1.857 times more likely to be loyal towards Prego when compared to households with family size more than 6 people.

**SURVIVAL ANALYSIS OF CUSTOMERS:**

The Survival analysis is best model to determine when the customer will switch brand and from that we could work on devising different retention strategies in addition to the results which we derived from the RFM analysis. To determine after how many weeks, the customer will be going to switch or will not be using the product. The Survival Analysis tell the time when the event of interest will occur.

We have performed survival analysis on tenure, which we calculated by taking the maximum week and minimum week the customer had bought and subtracted the same. We have censored the data based on week, by taking Churn = 0, when week >= 1165. We have the life test for all brands and our brand Prego with Family size as a strata. From the result, we could confirm that the survival probability across different groups are different. The chi-sq test confirms the same.

From the life test, we could see that the survival probability of the customer with different family size greatly varies. The first figure showing the survival probability for all brands, where we could see a nice trend. The survival probability decreases with time, here we cannot comment on whether the customer switch or did not buy sauce after 51 weeks. But we can compare the plot with Prego brand and we can try to get some interesting insights.

When comparing to the survival probability of Prego, the Prego has a lesser survival probability when compared to overall market standard. The Households with family five are dominating the overall data with highest survival probability where as in Prego, the same category seems to struggle. Almost most of the customer from that segment get churned out or switched to other brand at around 40 weeks. The households with family size 2 and 3 have a better survival probability and forms a loyal group to Prego