BUAN 6337 – Predictive Analytics using SAS

Group -11 Project

CLASSICO MARKET ANALYSIS REPORT

TEAM

Anjana Sebastian

Arjitkumar Sureshkumar

Harkirat Kaur

Prithvi Shivashankar

Soham Bhalerao

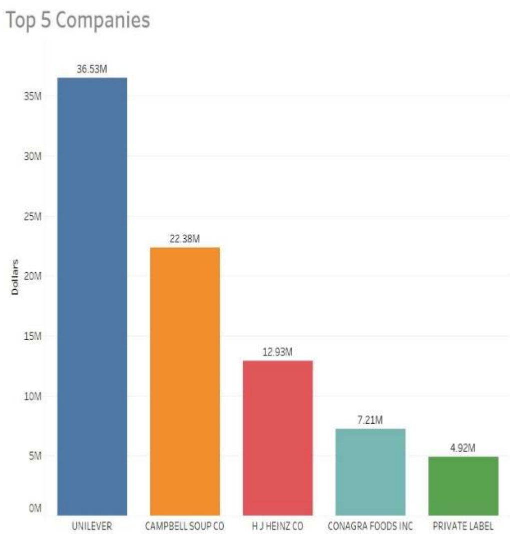
Sai Pratheek Banda

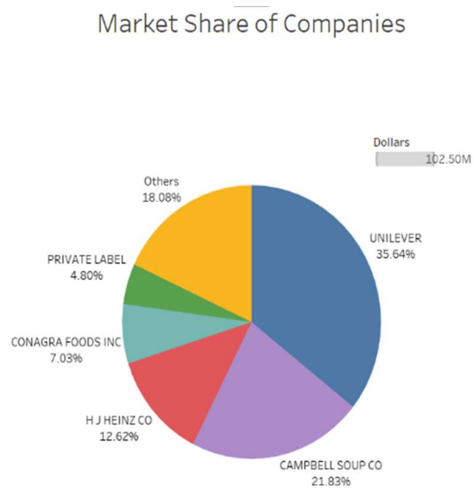
OBJECTIVE:

To provide insights as the brand manager of the brand CLASSICO to enhance market share in the sauce industry.

1. Analyse the effect of change in sales to the change in price for the brand itself (self-price elasticity) and based on the competitor's price (cross-price elasticity)
2. Analyse the effect of non-brand specific characteristics on the choice of brand.
3. Analyse the customer segmentation based on recency, frequency and monetary value of purchase using RFM method.

Exploratory Data Analysis - Company Analysis:

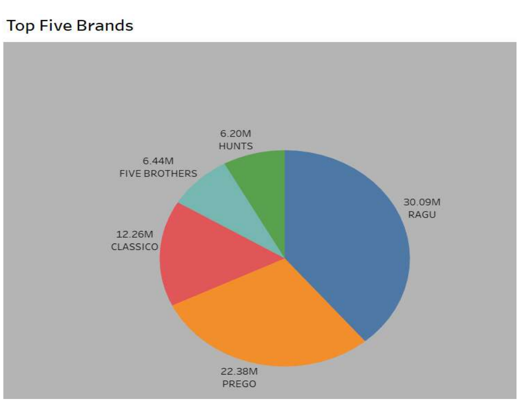


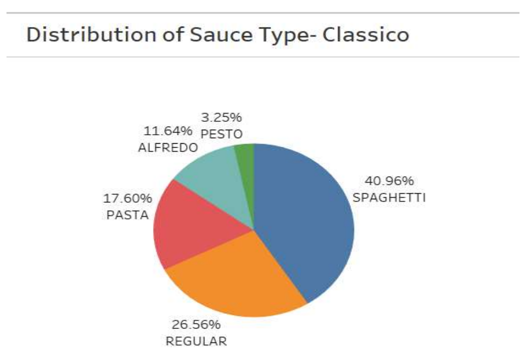


- The Spaghetti Sauce Market is led by the company UNILIVER followed by in the respective order CAMPBELL SOUP CO, HJ HEINZ, CONAGRA FOODS, PRIVATE LABEL.
- The Market share is predominantly shared by these top 5 Companies while the rest of the 234 Companies are together are shown as others, sharing 18% of the market.

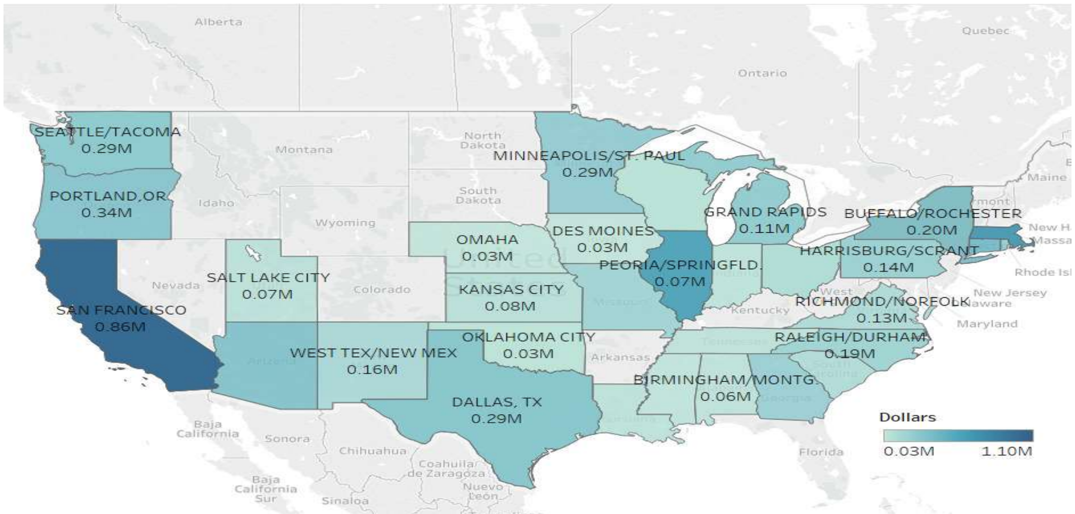
Brand Analysis

- Within and across each Company, there are 302 Brands competing in the sauce industry. Brand RAGU leads with the highest sales of $30.09 M followed by PREGO and CLASSICO.





Geo Market Dispersion- CLASSICO



PRICE ELASTICITY ANALYSIS:

Data Preparation:

We consider the top 2 brands in terms of market share for analysing the self-price elasticity and cross price elasticity of CLASSICO. The weighted price is calculated for each of the brand so that one observation is recorded for each brand for each store for that specific week it was open.

The price per ounce is calculated as:

- Price/ounce = ((DOLLARS/UNITS)/VOL_EQ)

The price per ounce is calculated as such because each of the bottled sauces may vary in terms of size and hence the price per ounce is a more reliable estimate in this case.

The weighted price calculation is as follows:

- Weighted Price$_i$ = $\sum$ Price per ounce$_i$ * $\frac{\textit{Sales of sauce Brand}}{\textit{Total Sales of Brand}}$

Similarly, the weighted value is also calculated for the feature, display and price reduction score.

Regression equation through SAS:

We use a PROC PANEL regression model with total sales taken as the dependent variable and the price for each brand which has been grouped separately alongside the interaction terms that is the weighted price of each of the brands, weighted display of brands, weighted feature of brands, weighted price reduction score of brands, interaction between price and feature, interaction between price and price reduction score and the interaction between feature and price reduction score.

Sales$_1$ = β_0 + β_1 * wt_price_brand1 + β_2 * wt_price_brand2 + β_3 * wt_price_brand3 + β_4 * disp_wt_brand1 + β_5 * disp_wt_brand2 + β_6 * disp_wt_brand3 + β_7 * Feature_wt_brand1 + β_8 * Feature_wt_brand2 + β_9 * Feature_wt_brand3 + β_{10} * PR_wt_brand1 + β_{11} * PR_wt_brand2 + β_{12} * PR_wt_brand3 + β_{13} * price_PR1 + β_{14} * price_PR2 + β_{15} * price_PR3 + β_{16} * price_F1 + β_{17} * price_F2 + β_{18} * price_F3 + β_{19} * PR_F1 + β_{20} * PR_F2 + β_{21} * PR_F3

To calculate Self-Price Elasticity:

Self-Price Elasticity = $\frac{\%change\ in\ Sale\ \ \ Brand1}{\%chang\ \ \ in\ Price\ Bran}$

Therefore, Self-Price Elasticity = $\frac{\triangle Sales\ Brand1}{Sales\ Bran} * \frac{Price\ Bran}{\triangle Price\ Bran}$

Self-Price elasticity for Brand1 = ($\beta_1 + \beta_{13} * PR_1 + \beta_{16} * F1$) * $\frac{Pric}{Sales1}$

To calculate Cross-Price Elasticity:

Cross-Price Elasticity = $\frac{\%chang\ \ \ in\ Sales\ Brand1}{\%change\ in\ Price\ Bran}$

Therefore, Cross-Price Elasticity = $\frac{\triangle Sales\ Bran}{Sale\ \ \ Brand1} * \frac{Price\ Bran}{\triangle Price\ Brand2}$

Cross-Price elasticity for Brand1 = ($\beta_2 + \beta_{14} * PR_2 + \beta_{17} * F2$) * $\frac{Pric}{Sales1}$

Hausman test:

To check for correlation between the error term U_i and X variable.

H_0: Correlation exists between U_i and X variable.

H_1: No correlation exists between U_i and X variable.

For each brand, the panel model is run and based on each case, the p-value being highly significant. Hence, Fixed effects model is the ideal model for the dataset.

The SAS System

The PANEL Procedure
Fixed Two-Way Estimates

Dependent Variable: tot_units

Model Description	
Estimation Method	FixTwo
Number of Cross Sections	1047
Time Series Length	52

Fit Statistics			
SSE	125002282.4	DFE	48158
MSE	2595.6701	Root MSE	50.9477
R-Square	0.7642		

F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
1097	48158	80.40	<.0001

Brand Classico

The SAS System

The PANEL Procedure
Fixed Two-Way Estimates

Dependent Variable: tot_units

Model Description	
Estimation Method	FixTwo
Number of Cross Sections	1047
Time Series Length	52

Fit Statistics			
SSE	1834493738	DFE	48158
MSE	38093.2293	Root MSE	195.1749
R-Square	0.5024		

F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
1097	48158	22.24	<.0001

Brand Prego

The SAS System

The PANEL Procedure
Fixed Two-Way Estimates

Dependent Variable: tot_units

Model Description	
Estimation Method	FixTwo
Number of Cross Sections	1047
Time Series Length	52

Fit Statistics			
SSE	1632105303	DFE	48158
MSE	33890.6371	Root MSE	184.0941
R-Square	0.6011		

F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
1097	48158	23.98	<.0001

Brand Ragu

Estimates for PROC PANEL regression:

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	126.8719	10.8065	11.74	<.0001	Intercept
wt_price_brand1	1	-3.15419	1.1849	-2.66	0.0078	
wt_price_brand2	1	-9.72041	2.4051	-4.04	<.0001	
wt_price_brand3	1	-17.6962	3.5411	-5.00	<.0001	
disp_wt_brand1	1	157.1401	2.2692	69.25	<.0001	
disp_wt_brand2	1	-5.76311	1.5729	-3.66	0.0002	
disp_wt_brand3	1	-1.0629	1.9203	-0.55	0.5799	
Feature_wt_brand1	1	350.0068	7.7477	45.18	<.0001	
Feature_wt_brand2	1	11.87001	5.9682	1.99	0.0467	
Feature_wt_brand3	1	11.21203	9.2198	1.22	0.2240	
PR_wt_brand1	1	45.55713	6.1473	7.41	<.0001	
PR_wt_brand2	1	-3.09599	1.2248	-2.53	0.0115	
PR_wt_brand3	1	-4.71992	1.5342	-3.08	0.0021	
price_PR1	1	-57.6302	2.9737	-19.38	<.0001	
price_PR2	1	12.51246	2.5378	4.93	<.0001	
price_PR3	1	54.60591	3.2610	16.75	<.0001	
price_F1	1	-158.56	3.6711	-43.19	<.0001	
price_F2	1	-14.567	3.9052	-3.73	0.0002	
price_F3	1	-10.3988	6.5590	-1.59	0.1129	
PR_F1	1	-13.2358	3.6483	-3.63	0.0003	
PR_F2	1	-1.29577	3.5637	-0.36	0.7162	
PR_F3	1	-3.61155	4.8674	-0.74	0.4581	

Mean Total Sales	
Brand	**Tot Units**
Classico	78.52
Ragu	198.09
Prego	170.89

Price Elasticity	
No Feature & No PR	-0.0802
if PR only	-1.54
if Feature only	-4.11
Feature and PR	-5.58

Cross Price Elasticity w.r.t Ragu	
if PR only	0.045

Cross Price Elasticity w.r.t Prego	
if PR only	0.5922
Feature and PR	0.4255

Insights:

From the above regression estimates and the price elasticity calculations, we understand that if CLASSICO reduces price by 1% then, the sales will increase by 0.08%. If CLASSICO reduces price and offer a discounted price alongside a featured advertisement then, the sales will increase by 5.58%. If CLASSICO reduces price by 1% and adds a featured advertisement, the sales will increase by 4.11%.

If RAGU brand offers discounted price implying that for every 1% decrease in RAGU price, there will be 0.045% decrease in sales of CLASSICO.

If PREGO brand offers discounted price implying that for every 1% decrease in PREGO price, there will be 0.59% decrease in sales of CLASSICO.

Customer Demographic Brand Preference Study:

Data Preparation:

The data is structured to the right format and then only the top 3 brands in terms of market share as well as overall sales are retained in the dataset. The data containing the customer details is merged with the product sauce details. Based on this, each of the brands are coded categorically and then a multinomial regression is performed on the top 3 brands (RAGU, PREGO, CLASSICO).

Multinomial regression equation through SAS:

We use a multinomial regression model on the dataset. The Brand variable containing each of the brands as a categorical number is taken as the dependent variable to check how each of the variables affect the choice of the brand.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	43855.170	43309.683
SC	43871.288	43889.953
-2 Log L	43851.170	43165.683

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	685.4864	70	<.0001
Score	708.7465	70	<.0001
Wald	650.5163	70	<.0001

Fam_Big_2 =	1 * Intercept_2 - Fam_Small_2 - Fam_Med_2
Fam_Big_3 =	1 * Intercept_3 - Fam_Small_3 - Fam_Med_3

Analysis of Maximum Likelihood Estimates						
Parameter	brand	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	2	1	0.365	0.5452	0.4481	0.5032
Intercept	3	1	-0.1654	0.7192	0.0529	0.8181
low_income	2	1	1.6363	0.5998	7.4417	0.0064
low_income	3	1	1.6338	0.7789	4.3992	0.036
med_income	2	1	1.7863	0.6014	8.8217	0.003
med_income	3	1	1.7916	0.7803	5.2726	0.0217
rent	2	1	0.418	0.0721	33.5947	<.0001
rent	3	1	0.2219	0.0759	8.5395	0.0035
gen_y	2	1	2.5604	0.8412	9.2643	0.0023
gen_y	3	1	0.6462	0.954	0.4589	0.4981
gen_x	2	1	2.2721	0.8315	7.4672	0.0063
gen_x	3	1	0.1354	0.945	0.0205	0.886
baby_boom	2	1	2.156	0.8287	6.7695	0.0093
baby_boom	3	1	-0.0728	0.9423	0.006	0.9384
grade_school	2	1	-1.3408	0.9219	2.1152	0.1458
grade_school	3	1	0.2094	1.0407	0.0405	0.8405
high_school	2	1	-1.4425	0.8823	2.6727	0.1021
high_school	3	1	0.2132	0.998	0.0456	0.8309
tech_school	2	1	-1.4379	0.8847	2.6417	0.1041
tech_school	3	1	0.4219	1.0004	0.1779	0.6732
college	2	1	-1.4793	0.8833	2.8045	0.094
college	3	1	0.2868	0.9991	0.0824	0.7741
grad	2	1	-1.9504	0.8853	4.8537	0.0276
grad	3	1	-0.0888	1.0011	0.0079	0.9294
emp	2	1	0.3326	0.3541	0.8823	0.3476
emp	3	1	0.1366	0.3657	0.1396	0.7087
male_emp	2	1	-0.1874	0.0608	9.4942	0.0021

male_emp	**3**	1	-0.2138	0.0638	11.2368	0.0008
female_emp	**2**	1	0.1415	0.0549	6.6381	0.01
female_emp	**3**	1	0.069	0.0575	1.4392	0.2303
single	**2**	1	0.4178	0.2816	2.2012	0.1379
single	**3**	1	0.5783	0.3245	3.1771	0.0747
married	**2**	1	0.6733	0.2645	6.4815	0.0109
married	**3**	1	0.9568	0.3079	9.6535	0.0019
divorced	**2**	1	0.6652	0.2745	5.8718	0.0154
divorced	**3**	1	0.817	0.3178	6.608	0.0102
widowed	**2**	1	0.9211	0.2759	11.1431	0.0008
widowed	**3**	1	1.1271	0.3189	12.4907	0.0004
seperated	**2**	1	0.8157	0.3255	6.2796	0.0122
seperated	**3**	1	0.4745	0.3735	1.6136	0.204
White	**2**	1	-1.5565	0.2537	37.6298	<.0001
White	**3**	1	-0.8273	0.2603	10.1013	0.0015
Afro	**2**	1	-0.2287	0.5247	0.19	0.6629
Afro	**3**	1	0.6159	0.5338	1.3311	0.2486
Asian	**2**	1	0.4953	1.0434	0.2253	0.635
Asian	**3**	1	0.9431	1.0565	0.7968	0.3721
Other_race	**2**	1	-2.5603	0.3242	62.3535	<.0001
Other_race	**3**	1	-2.4668	0.3633	46.1003	<.0001
native_race	**2**	1	-1.1304	0.5414	4.3592	0.0368
native_race	**3**	1	-0.5261	0.5584	0.8879	0.346
Hawaiian	**2**	1	-4.4911	0.5537	65.7858	<.0001
Hawaiian	**3**	1	-4.265	0.8192	27.103	<.0001
kidg6_11	**2**	1	-0.1505	0.2192	0.4715	0.4923
kidg6_11	**3**	1	-0.5624	0.2237	6.3216	0.0119
kidg12_17	**2**	1	-0.1962	0.2082	0.8884	0.3459
kidg12_17	**3**	1	-0.5427	0.2115	6.5811	0.0103

kidg0_11	**2**	1	-0.4448	0.2705	2.7046	0.1001
kidg0_11	**3**	1	-0.4524	0.2743	2.7187	0.0992
kidg05_1217	**2**	1	-0.7876	0.3455	5.1956	0.0226
kidg05_1217	**3**	1	-1.5374	0.3779	16.5467	<.0001
kidg06_17	**2**	1	-0.1115	0.2231	0.2498	0.6172
kidg06_17	**3**	1	-0.1132	0.2259	0.2511	0.6163
kidg0_17	**2**	1	0.0747	0.4305	0.0301	0.8622
kidg0_17	**3**	1	0.1134	0.4329	0.0687	0.7933
kidg0	**2**	1	-0.5754	0.2026	8.0669	0.0045
kidg0	**3**	1	-0.5373	0.2053	6.8476	0.0089
Fam_Small	**2**	1	-0.2801	0.0976	8.2429	0.0041
Fam_Small	**3**	1	-0.5375	0.1009	28.3828	<.0001
Fam_Med	**2**	1	-0.1358	0.0842	2.6016	0.1068
Fam_Med	**3**	1	-0.2715	0.0868	9.7908	0.0018
Fam_Big	**2**	0	0	.	.	.
Fam_Big	**3**	0	0	.	.	.
dogs	**2**	1	-0.133	0.0488	7.4357	0.0064
dogs	**3**	1	-0.1338	0.051	6.8904	0.0087
cats	**2**	1	0.171	0.0517	10.9426	0.0009
cats	**3**	1	0.2629	0.0537	23.9901	<.000

Interpreting results and deriving actionable insights:

Several variables are insignificant in the model run above. The reference variable used is the Classico brand and based on that a few actionable insights have been derived which would be of big use in created a set of targeted customers.

Interpretation:

- A person who lives in a rented house is more likely to pick Ragu or Prego over Classico
- In comparison to Classico and Prego, a person who is Generation Y(gen_y) will most likely prefer Ragu
- A person who is Generation X will more likely prefer Ragu over Classico and Prego
- A person who is in a college will more likely pick Classico in comparison to Ragu or Prego
- A person enrolled in grad school is likely to pick Classico or Prego over Ragu
- A male employee is more like to pick Classico in comparison to Prego and Ragu

- A person of every marital status is more likely to pick Prego or Ragu over Classico and this shows that the status is a factor to be investigated to understand the reason for the preference
- A person who is White will more likely prefer Classico over Ragu or Prego
- A person who is of a native race will more likely prefer Classico or Prego over Ragu
- A Hawaiian will more likely prefer Classico over Ragu and Prego
- A family with kids in the age group of 6-11 are more likely to pick Classico or Ragu over Prego and a similar case is observed where, families with kids in the age group 12-17 prefer Classico or Prego over Ragu
- A very interesting observation is where families with children in the age group 0-5 and 12-17 are more likely to pick Classico over Ragu and Prego
- In comparison to Ragu and Prego, a small family will more likely pick Classico
- As compared to a large family, a small family and a medium sized family are more likely to pick Classico
- People with dogs and cats are more likely to prefer Classico over Prego and Ragu

Insights:

Collectively, Classico is the preferred brand for people who are white, Native Americans and Hawaiians. This shows that traditional Americans prefer Classico as a brand over the others. They should be a focus as target customers to extract higher business value. Families with children between the age group between 6-11 and age group 12-17 are more likely to pick Classico and this shows the favourability of the brand with children. Interestingly, small and medium families prefer Classico and this should our target customers to build upon a larger market share value. Finally, people with dogs and cats are more likely to prefer Classico over the other brands and this could be partly attributed to the fact that small families are the ones more likely to have pets and this is an observation that should be noted for our targeted customer base to gain more loyal customers.

CUSTOMER SEGMENTATION (RECENCY, FREQUENCY, MONETARY AND LOYALTY):

The data set containing the purchase information of each panel ID across week. It was filtered for specific brands. The metrics evaluated are:

Recency by Week - How recently a customer purchased? – Difference between last week of purchase and each week a customer bought

Frequency - How often a customer purchased? – Count of the number of weeks of purchase available for each customer in the dataset.

Monetary - How much a customer spent for the product? – Sum of all the dollar amount paid by a customer made during purchase across weeks in the dataset.
Loyalty - How many times a specific brand was chosen over the other.

We have looked at people who have purchased CLASSICO and PREGO at least once. On that data set of semi-loyal customers, we have created RFM segments to better understand our present customers, our target customers and their attributes and demographics.

On Various combinations we segmented people who were buying CLASSICO and compared to PREGO:

The SAS System

The MEANS Procedure

Cluster	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
1	35	loyality_classico	35	0.7867528	0.2251634	0.2266667	1.0000000
		recency	35	2.3428571	2.5081380	0	9.0000000
		frequency	35	17.8285714	6.1762625	11.0000000	35.0000000
		monetary	35	2.9591314	0.6661475	1.9288889	5.8561765
2	110	loyality_classico	110	0.2582289	0.2563160	0.0222222	1.0000000
		recency	110	18.8000000	17.3764578	0	50.0000000
		frequency	110	2.0000000	1.8523479	1.0000000	12.0000000
		monetary	110	5.0461855	0.8832039	3.7850000	9.5600000
3	291	loyality_classico	291	0.6666630	0.2661933	0.1891892	1.0000000
		recency	291	12.2542955	12.4170603	0	49.0000000
		frequency	291	3.7731959	2.5127475	1.0000000	11.0000000
		monetary	291	2.7315834	0.4635313	1.8900000	4.6500000
4	681	loyality_classico	681	0.1806671	0.1208049	0.0222222	0.6666667
		recency	681	23.9441997	15.1695263	0	51.0000000
		frequency	681	1.8252570	1.1969358	1.0000000	9.0000000
		monetary	681	2.5910493	0.2360524	1.8900000	3.8850000

PREGO

The MEANS Procedure

Cluster	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
1	892	loyality_PREGO	892	0.8710200	0.1713599	0.3928571	1.0000000
		monetary	892	2.2406487	0.5916581	0.9900000	4.2466667
		frequency	892	4.1289238	2.8811422	1.0000000	12.0000000
		recency	892	13.0033632	12.9085965	0	51.0000000
2	132	loyality_PREGO	132	0.8349725	0.1901329	0.1612903	1.0000000
		monetary	132	2.5639774	0.6542117	1.4450000	5.1566667
		frequency	132	19.7803030	10.2609572	12.0000000	76.0000000
		recency	132	2.6818182	3.8726250	0	26.0000000
3	260	loyality_PREGO	260	0.3347726	0.3036538	0.0217391	1.0000000
		monetary	260	4.6111795	1.3379311	3.3800000	14.4500000
		frequency	260	2.1846154	1.9758567	1.0000000	11.0000000
		recency	260	20.2192308	15.2642748	0	51.0000000
4	1625	loyality_PREGO	1625	0.2451462	0.1465681	0.0227273	0.6666667
		monetary	1625	2.2534584	0.4816152	0.9900000	3.4680000
		frequency	1625	2.2775385	1.7535731	1.0000000	12.0000000
		recency	1625	17.8966154	14.2444059	0	51.0000000

1.Valuable customers:

These group of people give highest value with high loyalty, monetary and frequency scores:

CLASSICO: We checked for people whose loyalty score was high and corresponding monetary and frequency value. We can see 5%(35/1117) customers were loyal to CLASSICO followed by more 42%(291/1117) people who are recent but provide good loyalty score.

PREGO: Around 30%(892/2909) are loyal to PREGO followed by more 4%(132/2909)

2.Opportunistic Customers:

These customers are Recent with less frequency and so can be targeted to improve retention.

CLASSICO: Around 61% (681/1117) people have recently bought from CLASSICO who can be targeted and retain to increase our customer base.

PREGO: Around 10%(260/2909) people recently bought from PREGO.

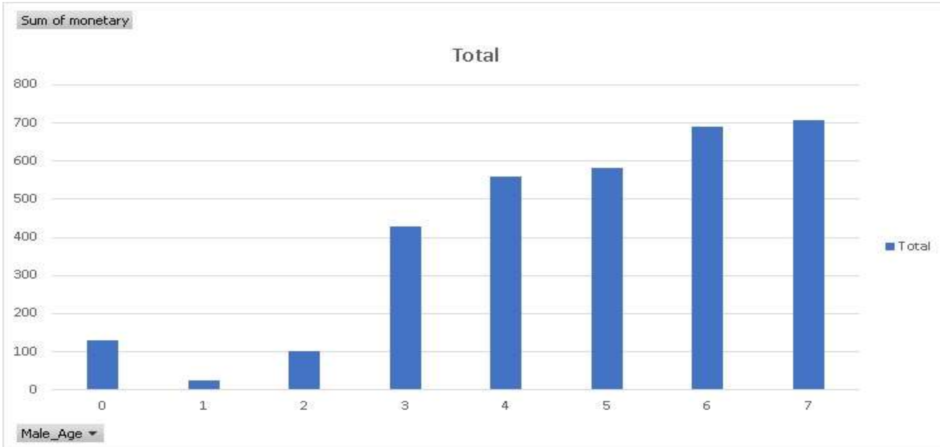
Clear advantage for CLASSICO.

3.Marginal Customers:

These are erratic customers and so difficult to identify their behaviour. There around 10% of such people in CLASSICO and 55% of PREGO customers. One interpretation can be, they change brands based on current state of market namely discounts, advertisements and marketing.

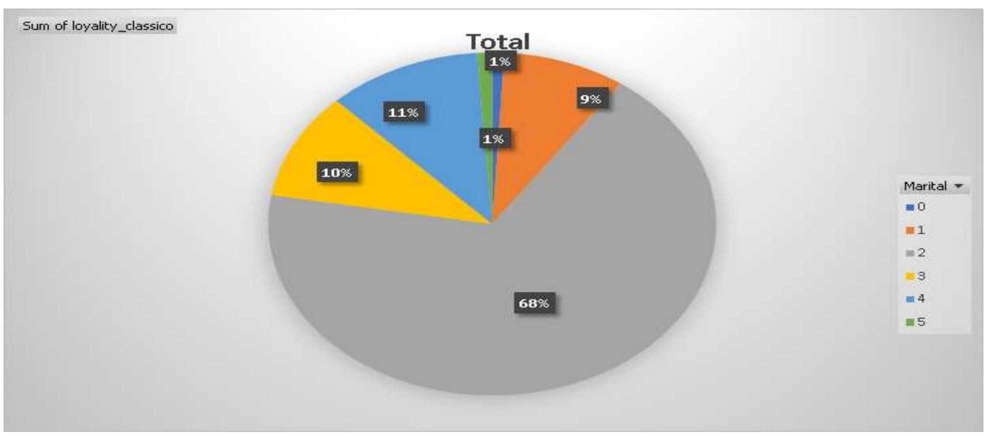
INSIGHTS:

Monetary Value (M): This is the amount of money these customers have spent on purchases. Customers who have spent higher contribute more value to the business as compared to those who have spent less. Here you can see with increase in age the customers are spending more. Hence the company can target older age group for more business.



Loyalty:

Married status contributes to 68% of the loyalty distribution. Second contribution is made by widowed followed by Divorced. Business should focus on making customised promotional strategies and loyalty schemes for these customers in order to retain this valuable customer base.



From the below output, we can see that small-sized family household prefer Classico brand