

Spring 2019

BUAN 6337.001: Predictive Analytics using SAS

A Report On

Frozen Pizza Brand Marketing



**THE UNIVERSITY
OF TEXAS AT DALLAS**

Submitted by:

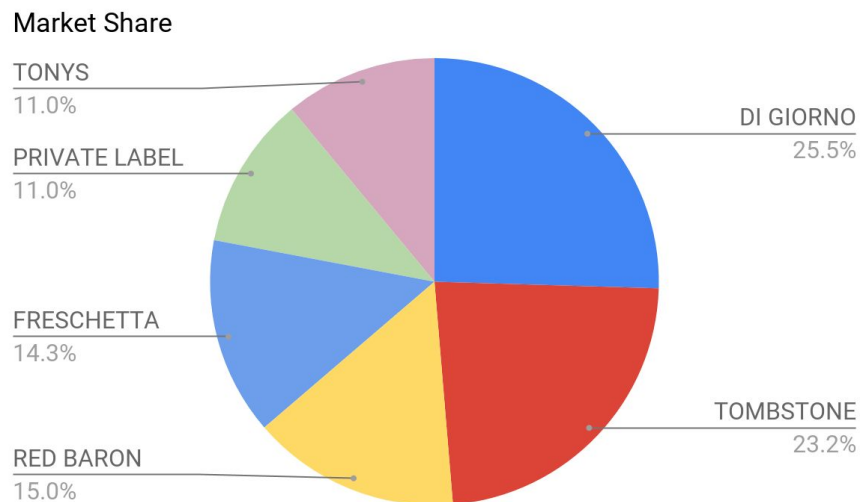
Arvinth Arul- AXA170033

Under The Guidance of

Prof. B. Murthi

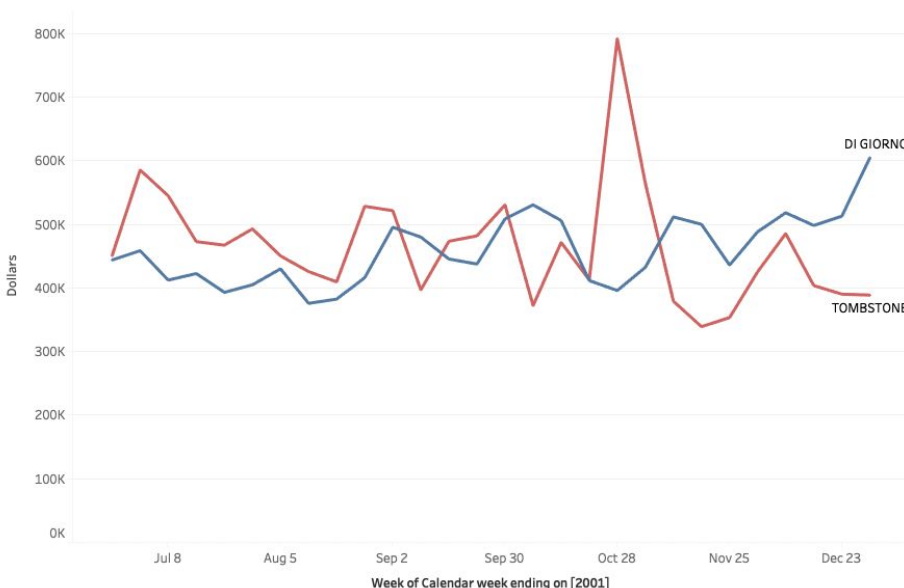
Introduction:

Pizza is undoubtedly one of the most relished food items in the United States. The Tombstone Pizza Company, a freestanding division of Kraft Foods, is one among the most popular frozen pizza brands in the country. In this report, we have captured several insights that help determine the standing that Tombstone establishes in the market against various other pizza brands.

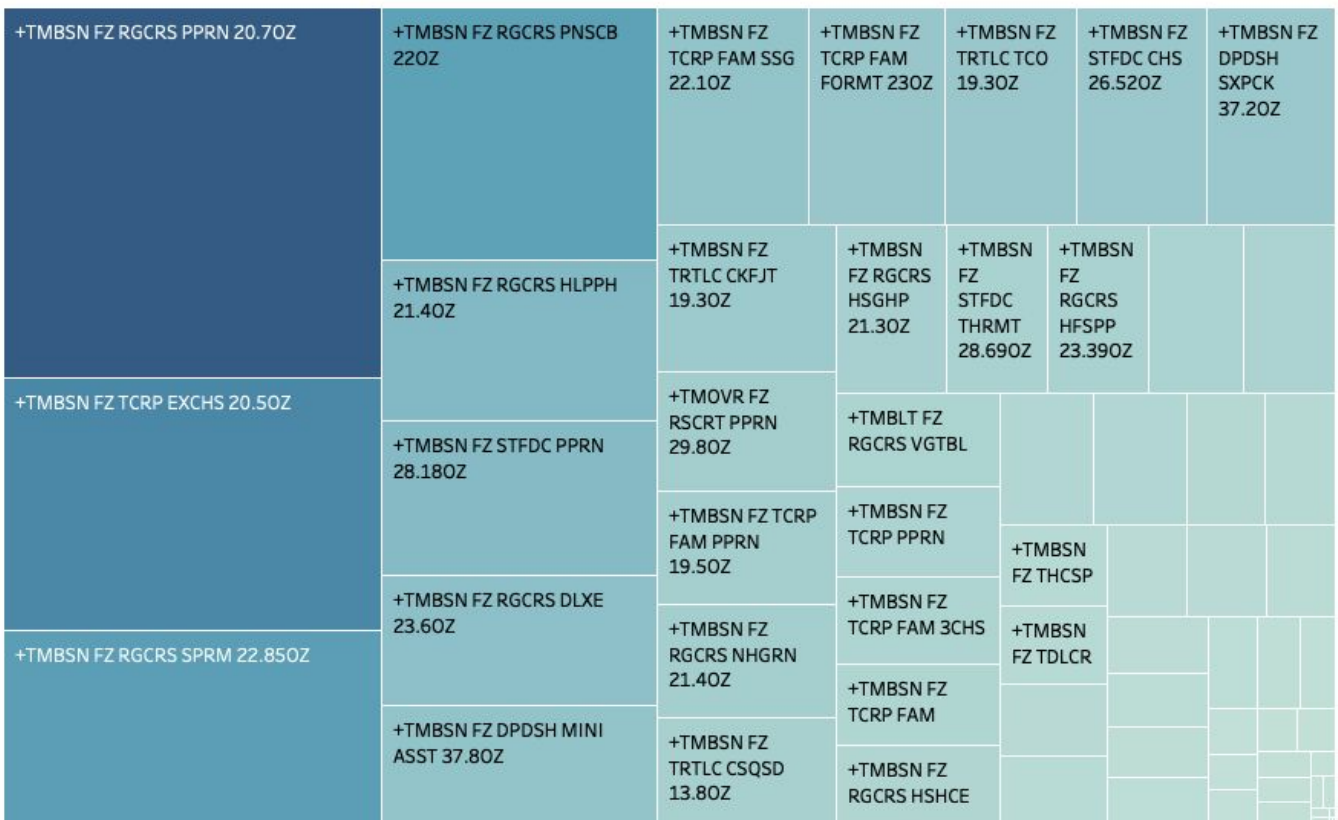


Tombstone has a market share of 23.21% and Di Giorno tops the chart with 25.5%.

Comparing Sales Trend of Tombstone and Di Giorno in the year 2001



Understanding the contribution of Tombstones Products with respect to sales



PROBLEM STATEMENT-1

Determine the effect of price, display, feature and promotion on the sales of Tombstone frozen pizza

Data Preparation:

861 stores which had data for all 52 weeks were filtered to create a balanced data set for **Panel regression**.

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

$$\text{Total Volume} = \text{Volume Equivalent of a Product} * \text{Number of Units Sold}$$

$$\text{Standardised Price} = \text{Total Cost of a Purchase} / \text{Total Volume}$$

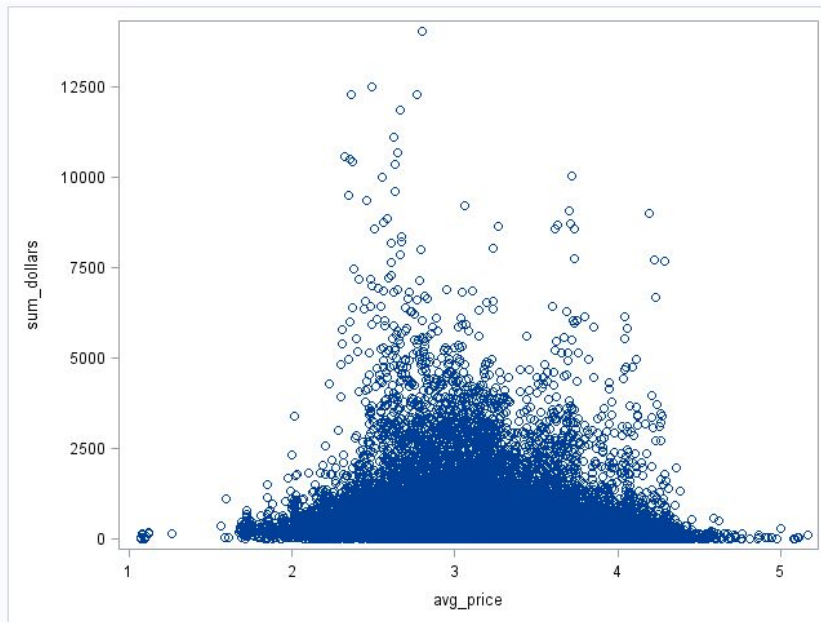
Multicollinearity Check:

We observe that the VIF values are well within 10 and the Collinearity Condition indices are much lesser than 100. Therefore, we can rule out multicollinearity in this model.

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	-11.09529	9.16668	-1.21	0.2261	0
avg_price	1	13.34401	2.53267	5.27	<.0001	2.02872
sum_w_feature	1	136.76509	3.66840	37.28	<.0001	1.86582
sum_w_display	1	278.05012	4.52171	61.49	<.0001	1.16159
sum_w_promotion	1	54.46818	3.87247	14.07	<.0001	2.59061

Collinearity Diagnostics							
Number	Eigenvalue	Condition Index	Proportion of Variation				
			Intercept	avg_price	sum_w_feature	sum_w_display	sum_w_promotion
1	3.29316	1.00000	0.00068720	0.00073793	0.02200	0.02721	0.01474
2	0.90263	1.91008	0.00208	0.00346	0.10049	0.19775	0.01007
3	0.58550	2.37161	0.00006138	0.00029127	0.16444	0.77500	0.04158
4	0.21372	3.92543	0.00022967	0.00154	0.69811	0.00004582	0.58671
5	0.00499	25.68017	0.99694	0.99397	0.01496	0.00000190	0.34690

Non-Linearity Check :



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

Panel Regression:

Based on the Hausman test, we go with Fixed effects model.

Dependent Variable: Total Units Sold/Week In a Store

Independent Variables: Weighted Price, Weighted Display, Weighted Feature, Weighted Promotion, Interaction between Display and Feature, Interaction between Display and Promotion, Interaction between Feature and Promotion, Interaction between Display, Feature and Promotion

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Pr > t	Label
Intercept	1	876.4733	30.8510	28.41	<.0001	Intercept
avg_price	1	-83.7802	15.0070	-5.58	<.0001	
pricesq	1	9.303238	2.3335	3.99	<.0001	
sum_w_feature	1	-5.50804	7.8806	-0.70	0.4846	
sum_w_display	1	87.46746	6.6998	13.06	<.0001	
sum_w_promotion	1	15.55397	3.5379	4.40	<.0001	
INF_D	1	154.308	26.3535	5.86	<.0001	
INF_P	1	89.38004	9.9525	8.98	<.0001	
IND_P	1	-7.43432	12.4133	-0.60	0.5492	
INF_D_P	1	102.0227	31.6742	3.22	0.0013	

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

Weighted Price: Nonlinear relationship between price and number of sheets sold is seen from the estimates. As price increases the amount spent increases to a certain limit and then decreases.

Weighted Display: As the display increases by 1% number of units sold per week in a shop increases by 87 units.

Weighted Promotion: As the promotion increases by 1% number of units sold per week in a shop increases by 15 units.

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

Interaction between Promotion and Feature: This produces a synergistic effect in number of units sold per week in a shop

Interaction between Display, Promotion and Feature: This produces a synergistic effect in number of units sold per week in a shop

Recommendation: Featuring alone does not have a significant effect on sales. But when combined with Display or Promotion it has a significant effect than having those separately. This insight can save money when spent only for featuring.

PROBLEM STATEMENT-2

Analyze the demographic features of the most valuable customers of the Tombstone frozen pizza brand

Data Preparation & Analysis:

Panel Data has purchase information for all frozen pizza products purchased in all stores. We filter out for Tombstone and perform **RFM analysis** to segment customers who are most valuable with respect to the following 3 metrics:

- Recency: Difference between last purchase date and the latest week of year to find out how recently a purchase has been made.
- Frequency: Number of weeks a customer has made a purchase
- Monetary: Total Amount spent by customers

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
MONETARY	1255	24.32140	19.95297	30523	4.38000	172.08000
FREQUENCY	1255	4.77849	3.82058	5997	2.00000	31.00000
RECENCY	1255	8.18247	6.82331	10269	0	27.00000

Pearson Correlation Coefficients, N = 1255 Prob > r under H0: Rho=0			
	MONETARY	FREQUENCY	RECENCY
MONETARY	1.00000	0.87009 <.0001	-0.32846 <.0001
FREQUENCY	0.87009 <.0001	1.00000	-0.34583 <.0001
RECENCY	-0.32846 <.0001	-0.34583 <.0001	1.00000

Since Monetary and Frequency are highly correlated we will use only Monetary and Recency to split our customers into 3 segments.

Variable	Minimum	20th Pctl	40th Pctl	60th Pctl	80th Pctl	Maximum
MONETARY	4.3800000	10.0000000	14.9700000	21.9600000	34.9300000	172.0800000
REGENCY	0	2.0000000	4.0000000	9.0000000	13.0000000	27.0000000

Based on the Pareto principle, we assume that 20% of the customers contribute more revenue to the company than the rest. So we split the customers into 3 segments.

Segment 1 - Monetary > \$34.39

Segment 2 - (Monetary > \$21.96 AND Monetary<\$34.39) AND (Recency>13)

Segment 0 - Rest

For Segment 1 we mapped it to the demographics and found the average for each characteristics of a customer using dummy variables.

Customer Demographics	Levels	Average
Large Family Size	4,5,6	53%
Normal Family Size	1,2,3	46%
Low Income	1,2,3,4	4%
Medium Income	5,6,7,8	40%
High Income	9,10,11,12	55%
Age of Male - Young	1	0%
Age of Male - Mid	2,3,4	69%
Age of Male - Old	5,6	31%
Age of Female - Young	1	1%
Age of Female - Mid	2,3,4	74%
Age of Female - Old	5,6	24%
Education - Male -School	1,2,3	1%
Education - Male - College	4,5,6	70%
Education - Male - Graduate	7,8	28%
Education - Female -School	1,2,3	1%
Education - Female - College	4,5,6	70%
Education - Female - Graduate	7,8	28%
One Child	1,2,3	32%
Two Children	4,5,6	13%

Three Children	7	0%
No Children	8	53%
Male - White Collar High Post	1,2,3	46%
Male - White Collar Low Post	4,5	7%
Male Blue Collar	6,7,8,9	33%
Male No Occupation	10,13	13%
Female - White Collar High Post	1,2,3	44%
Female - White Collar Low Post	4,5	22%
Female Blue Collar	6,7,8,9	11%
Female No Occupation	10,13	22%
Pets in House (Cats+Dogs)		68%

The following characteristics of top 20% Customers are as follows:

- Family size ranging from 4 to 6
- High income from \$55K to \$100K
- Age group 25-54
- Education Level - graduated High School, Technical School or studying in some College
- Houses with no children and have at least 1 dog or cat.
- People having a high position white collar job

Recommendation : Concentrating on the Top 20% and understanding their purchase history enables us to attract similar customers in the future. New customers with similar demographic features can be targeted to improve sales in long run.

PROBLEM STATEMENT-3

Evaluating the Brand preferences among different customers

Analysis:

We build a **Multinomial Logistic Regression** for our analysis. The outcome of this analysis measures Brand preference among – Tombstone, Di Giorno or Red Baron- from which we are going to see what relationships exists with Family Size (Fam_Size), Family Income (Fam_Income) and number of children (Childcount).

Tombstone is chosen as the reference category and coded as 1. Di Giorno is coded as 2 and Red Baron as 3.

Response Profile		
Ordered Value	DECISION	Total Frequency
1	1	9158
2	2	1910
3	3	3476

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	26181.339	26051.483
SC	26196.509	26112.163
-2 Log L	26177.339	26035.483

DECISION: Two models were defined in this multinomial regression: one relating DI GIORNO to the referent category, TOMBSTONE, and another model relating RED BARON to TOMBSTONE. The DECISION number indicates to which model an estimate, standard error, chi-square, and p-value refer. Our DECISION categories 2 and 3 are DI GIORNO AND RED BARON.

Analysis of Maximum Likelihood Estimates						
Parameter	DECISION	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	2	1	-0.9692	0.1273	57.9967	<.0001
Intercept	3	1	-0.3606	0.0988	13.3130	0.0003
FAM_INCOME	2	1	0.0347	0.0418	0.6898	0.4062
FAM_INCOME	3	1	-0.1954	0.0323	36.5507	<.0001
CHILDCOUNT	2	1	0.00674	0.0467	0.0208	0.8852
CHILDCOUNT	3	1	0.1352	0.0343	15.5434	<.0001
FAM_SIZE	2	1	-0.4772	0.0658	52.5861	<.0001
FAM_SIZE	3	1	-0.1452	0.0508	8.1617	0.0043

Odds Ratio Estimates		
Effect	DECISION	Point Estimate
FAM_INCOME	2	1.035
FAM_INCOME	3	0.823
CHILDCOUNT	2	1.007
CHILDCOUNT	3	1.145
FAM_SIZE	2	0.621
FAM_SIZE	3	0.865

MODEL 1 : DI GIORNO relative to TOMBSTONE

Intercept – This is the multinomial logit estimate for DI GIORNO relative to TOMBSTONE when the predictor variables in the model are evaluated at zero. The logit for preferring DI GIORNO to TOMBSTONE is -0.96.

FAM_SIZE: This is the multinomial logit estimate for when FAMILY SIZE increases from regular (1,2,3) to large family size(4,5,6) for DI GIORNO relative to TOMBSTONE, given the other variables in the model are held constant.

If a customer were to move from regular (1,2,3) to large family size(4,5,6) the multinomial log-odds for preferring DI GIORNO to TOMBSTONE would be expected to decrease by -0.47 unit while holding all other variables in the model constant.

MODEL 2: RED BARON relative to TOMBSTONE

Intercept: This is the multinomial logit estimate for DI GIORNO relative to TOMBSTONE when the predictor variables in the model are evaluated at zero. The logit for preferring RED BARON to TOMBSTONE is -0.36.

FAM_INCOME: This is the multinomial logit estimate for when FAMILY INCOME bracket increases from low to medium or medium to high for RED BARON relative to TOMBSTONE, given the other variables in the model are held constant.

If a customer were to move from lower income bracket to higher income bracket the multinomial log-odds for preferring RED BARON to TOMBSTONE would be expected to decrease by 0.19 unit while holding all other variables in the model constant.

CHILDCOUNT: If the number of children in a house increase by one unit the multinomial log-odds for preferring RED BARON to TOMBSTONE would be expected to increase by 0.13 unit while holding all other variables in the model constant.

FAM_SIZE: This is the multinomial logit estimate for when FAMILY SIZE increases from regular (1,2,3) to large family size(4,5,6) for RED BARON to TOMBSTONE, given the other variables in the model are held constant.

If a customer were to move from regular (1,2,3) to large family size(4,5,6) the multinomial log-odds for preferring RED BARON to TOMBSTONE would be expected to decrease by -0.14 unit while holding all other variables in the model constant.

Recommendation:

With an increase in income and family size, Tombstone is preferred. But to sustain increase in consumption from Families, Tombstone needs to target children as it is clearly seen that Red Baron is preferred in the houses with children. Just like how McDonald's introduced Happy Meals with Action Comic characters, Tombstone can also experiment with children-friendly product lines.