BUAN 6337 – Predictive Analytics using SAS

Group – 7 Project
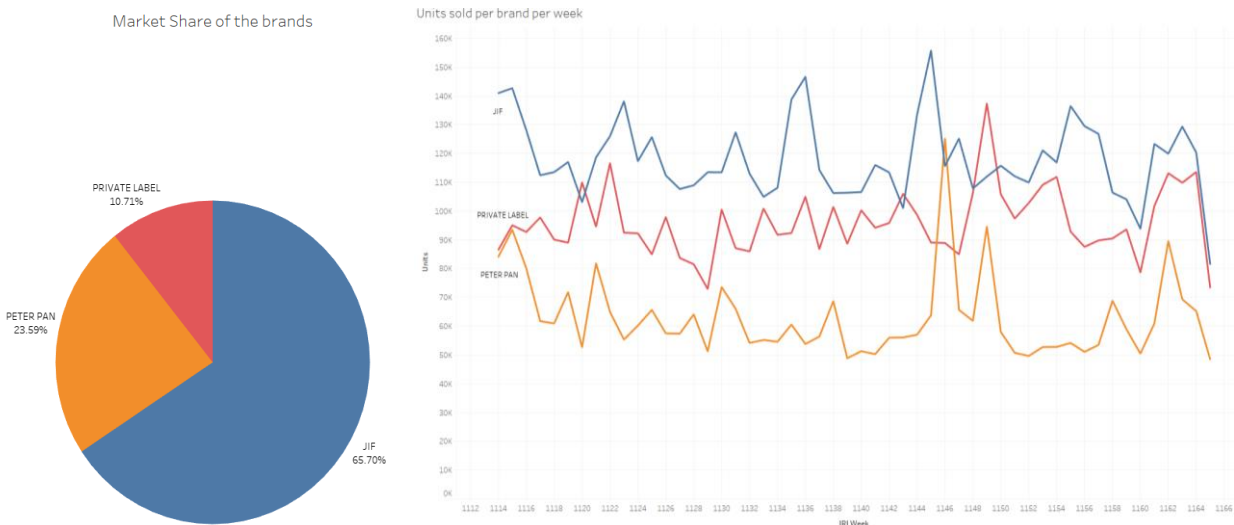
PEANUT BUTTER MARKET ANALYSIS

Vivekpandian Veerapandian

VXV180020

To provide insights as PETER PAN's brand manager to enhance market share in the Peanut butter industry.
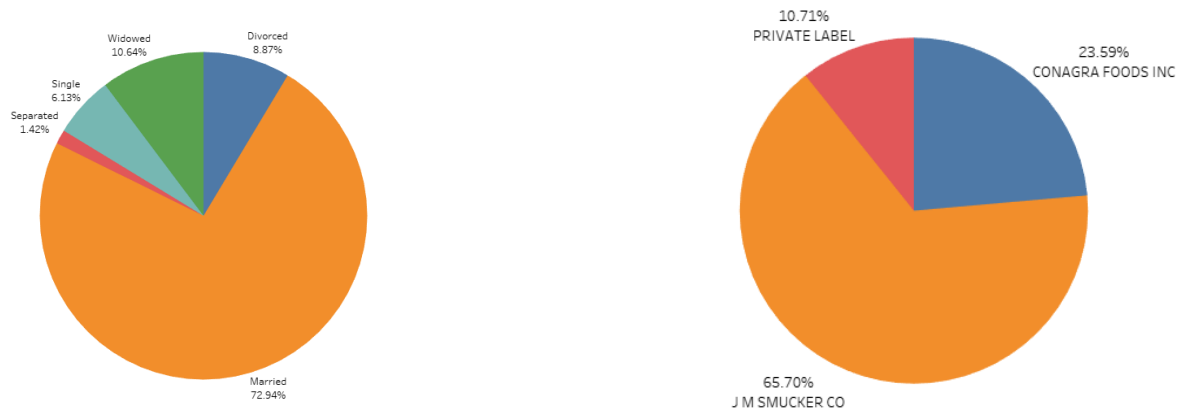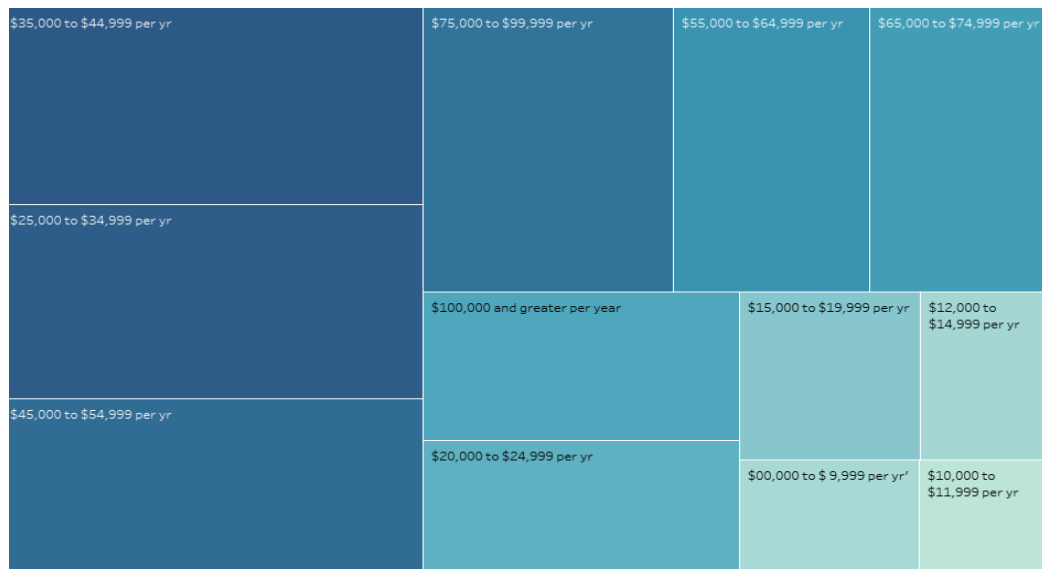
Exploratory Data Analysis:

Brand Analysis:

- JIF has the major market share in the peanut butter industry.
- Eventhough, JIF is the leading brand, there is an instance where we can notice that Peter Pan sold more units than the leaders in the market. It indicates that there is scope for improving our brand.

Customer Base Analysis:

- Customer prefer the brand produced by J M SMUCKER CO as their product is sold around 65% of total market.
- Most of the customer base consists of the customers who are married and are with a family.

$35,000 to $44,999 per yr	$75,000 to $99,999 per yr	$55,000 to $64,999 per yr	$65,000 to $74,999 per yr
$25,000 to $34,999 per yr			
	$100,000 and greater per year	$15,000 to $19,999 per yr	$12,000 to $14,999 per yr
$45,000 to $54,999 per yr			
	$20,000 to $24,999 per yr	$00,000 to $ 9,999 per yr'	$10,000 to $11,999 per yr

- Maximum sales of the penaut butter is contributed by the customers who have their income range from $35000 to $44,999 per year

Geo Market Distribution

Peanut butter sales across the United States
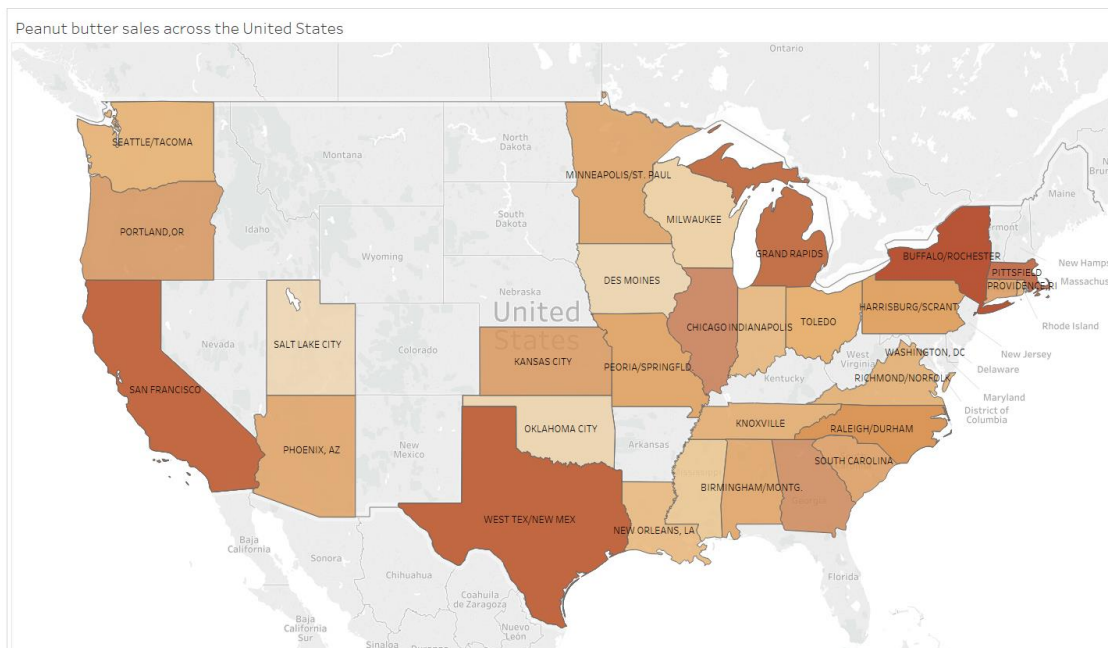
ANALYSIS OF PRICE ELASTICITY:

1. To analyze 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)

We consider the top 3 brands in terms of market share for analyzing the self-price elasticity and cross price elasticity of the brand 'PETER PAN'.

The price per ounce is calculated as:

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

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

The weighted price calculation is as follows:

- Weighted Price = Σ Price per ounce *(sales of peanut butter brand/total sales of peanut butter)

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

UNDERSTANDING THE IMPACT OF PRICE ELASTICITY on PETER PAN:

We use a PROC regression model with total ounces sold 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 and the interaction between feature and price reduction.

$Total\_ounces\_sold\_Peter\_Pan = \beta_0 + \beta_1 * wt\_price\_brand1 + \beta_2 * wt\_price\_brand2 + \beta_3 * wt\_price\_brand3 + \beta_4 * disp\_wt\_brand1 + \beta_5 * disp\_wt\_brand2 + \beta_6 * disp\_wt\_brand3 + \beta_7 * Feature\_wt\_brand1 + \beta_8 * Feature\_wt\_brand2 + \beta_9 * Feature\_wt\_brand3 + \beta_{10} * PR\_wt\_brand1 + \beta_{11} * PR\_wt\_brand2 + \beta_{12} * PR\_wt\_brand3 + \beta_{13} * price\_PR1 + \beta_{14} * price\_PR2 + \beta_{15} * price\_PR3 + \beta_{16} * price\_F1 + \beta_{17} * price\_F2 + \beta_{18} * price\_F3 + \beta_{19} * PR\_F1 + \beta_{20} * PR\_F2 + \beta_{21} * PR\_F3$

Self-Price elasticity for Brand1 = (β1 + β13 * PR1 + β16 * F1) * (price/sales1)

Cross-Price elasticity for Brand1 = (β2 + β14 * PR2 + β17 * F2) * (price/sales)

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: total_ounces_sold

Number of Observations Read	1465722
Number of Observations Used	1465722

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	21	40296778911	1918894234	9802.06	<.0001
Error	1.47E6	2.869317E11	195764		
Corrected Total	1.47E6	3.272285E11			

Root MSE	442.45259	R-Square	0.1231
Dependent Mean	214.19086	Adj R-Sq	0.1231
Coeff Var	206.56931		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	190.85399	0.41674	457.97	<.0001
wt_price_brand1	1	-56.34014	0.65246	-86.35	<.0001
wt_price_brand2	1	-243.82279	3.13248	-77.84	<.0001
wt_price_brand3	1	-103.88424	3.84683	-27.01	<.0001
disp_wt_brand1	1	1239.95091	11.42617	108.52	<.0001
disp_wt_brand2	1	4450.05489	26.94642	165.14	<.0001
disp_wt_brand3	1	1973.55502	10.26924	192.18	<.0001
Feature_wt_brand1	1	721.19787	19.92132	36.20	<.0001
Feature_wt_brand2	1	910.91494	38.38380	23.73	<.0001
Feature_wt_brand3	1	1055.05837	24.04462	43.88	<.0001
PR_wt_brand1	1	223.25095	5.90191	37.83	<.0001
PR_wt_brand2	1	76.16057	12.80230	5.95	<.0001
PR_wt_brand3	1	158.28619	6.09642	25.96	<.0001
price_PR1	1	-126.40903	6.11407	-20.68	<.0001
price_PR2	1	-566.94414	49.74387	-11.40	<.0001

price_PR3	1	-155.01766	26.33507	-5.89	<.0001
price_F1	1	-350.90940	8.49876	-41.29	<.0001
price_F2	1	-3812.34709	83.29602	-45.77	<.0001
price_F3	1	-1159.60320	59.61928	-19.45	<.0001
PR_F1	1	449.42952	49.65586	9.05	<.0001
PR_F2	1	6609.05451	186.34784	35.47	<.0001
PR_F3	1	777.38240	74.93427	10.37	<.0001

PRICE ELASTICITY OF PETER PAN:

When there's No Feature & No PR: - 0.664

When there's PR only: - 2.20

When there's Feature only: - 11.05

When there's Feature and PR: 17.34

Cross Price Elasticity of PETER PAN w.r.t JIF:

When there's PR only: - 0.797

When there's Feature and PR: 1.71

Cross Price Elasticity of PETER PAN w.r.t Private Label:

When there's PR only: - 0.227

When there's Feature and PR: 0.59

PRICE ELASTICITY INSIGHTS:

From the above regression estimates and the price elasticity calculations, we understand that if PETER PAN reduces price by 1% then, the sales will increase by 0.66%. If PETER PAN reduces price by 1% and offer a discounted price then, the sales will increase by 2.20%. If PETER PAN reduces price by 1% and adds a featured advertisement, the sales will increase by 11.05%. If PETER PAN reduces price and offer a discounted price alongside a featured advertisement then, the sales will increase by 17.34%.

If JIF brand offers discounted price alongside a featured advertisement implying that for every 1% decrease in JIF price, there will be 1.71% decrease in sales of PETER PAN.

If PRIVATE LABEL brand offers discounted price alongside a featured advertisement implying that for every 1% decrease in PRIVATE LABEL price, there will be 0.59% decrease in sales of PETER PAN.

PRODUCT CHARACTERISTICS ANALYSIS:

2. To determine whether there is an effect of size, texture, process, sugar content, and salt/sodium content on the total ounces of Peter Pan brand sold.

For the brand PETER PAN, Size is missing for more than 95% of the records, Process is missing for more than 99% of the records, salt/sodium content is missing for more than 80% of the records. So, considering only Flavor/scent, Texture, Sugar content and Product type for further analysis.

a. Flavor/Scent

The SAS System

The ANOVA Procedure

Dependent Variable: total_ounces_sold

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4133298991	1377766330	5739.25	<.0001
Error	353172	84782637666	240060		
Corrected Total	353175	88915936657			

R-Square	Coeff Var	Root MSE	total_ounces_sold Mean
0.046485	285.9897	489.9597	171.3207

Source	DF	Anova SS	Mean Square	F Value	Pr > F
FLAVOR_SCENT	3	4133298991	1377766330	5739.25	<.0001

b. Sugar Content

The SAS System

The ANOVA Procedure

Dependent Variable: total_ounces_sold

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	926887007	231721752	930.09	<.0001
Error	353171	87989049650	249140		
Corrected Total	353175	88915936657			

R-Square	Coeff Var	Root MSE	total_ounces_sold Mean
0.010424	291.3479	499.1393	171.3207

Source	DF	Anova SS	Mean Square	F Value	Pr > F
SUGAR_CONTENT	4	926887007.4	231721751.9	930.09	<.0001

c. Texture

The SAS System

The ANOVA Procedure

Dependent Variable: total_ounces_sold

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	618000318	206000106	823.95	<.0001
Error	353172	88297936339	250014		
Corrected Total	353175	88915936657			

R-Square	Coeff Var	Root MSE	total_ounces_sold Mean
0.006950	291.8584	500.0140	171.3207

Source	DF	Anova SS	Mean Square	F Value	Pr > F
TEXTURE	3	618000318.0	206000106.0	823.95	<.0001

d. Product Type

The SAS System

The ANOVA Procedure

Dependent Variable: total_ounces_sold

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	587546669	293773335	1174.63	<.0001
Error	353173	88328389988	250099		
Corrected Total	353175	88915936657			

R-Square	Coeff Var	Root MSE	total_ounces_sold Mean
0.006608	291.9084	500.0995	171.3207

Source	DF	Anova SS	Mean Square	F Value	Pr > F
PRODUCT_TYPE	2	587546669.0	293773334.5	1174.63	<.0001

INSIGHTS:

From the results of ANOVA hypotheses tests, it can be understood that the effect of Product type, flavor/scent, texture and sugar content is significant on the total ounces sold of the brand Peter Pan. Thus, the total ounces of Peter Pan sold varies based on different product types, flavors/scent, textures and sugar content.

CUSTOMER BRAND DEMOGRAPHIC STUDY:

Log (total units ounces sold) = β_0 + β_1 combined_pre_tax_income_of_hh + β_2 family_size + β_3 hh_age + β_4 hh_edu + β_5 hh_occ + β_6 male_working_hour_code + β_7 marital_status + β_8 children_group_code + β_9 number_of_dogs

Interpretation:

Several variables are insignificant in the model run above. The reference variable used is the total ounces sold of PETER PAN brand and based on that a few actionable insights have been derived which would be helpful in creating a set of targeted customers.

1. The percentage difference of total unit ounces sold for an income range between '$10,000 to $11,999 per yr' and '$25,000 to $34,999 per yr' is 14%.

2. The percentage difference of total unit ounces sold for an income range between '$12,000 to $14,999 per yr' and '$25,000 to $34,999 per yr' is 12.1%.

3. The percentage difference of total unit ounces sold for an income range between '$35,000 to $44,999 per yr' and '$25,000 to $34,999 per yr' is 12.8%.

4. The percentage difference of total unit ounces sold for an income range between '$55,000 to $64,999 per yr' and '$25,000 to $34,999 per yr' is 14.4%.

5. The percentage difference of total unit ounces sold for an age range between '25 - 34' and above 65 is 36.9%.

6. The percentage difference of total unit ounces sold for an age range between '35 - 44' and above 65 is 31.4%.

7. The percentage difference of total unit ounces sold for an age range between '45 - 54' and above 65 is 23.5%.

8. The percentage difference of total unit ounces sold for an age range between '55 - 64' and above 65 is 12.4%.

9. The percentage difference of total unit ounces sold for an education level between 'Some high school' and 'Graduated high school' is 17%.

10. The percentage of total unit ounces sold for an education level 'Technical school' is 17% less than 'Graduated high school'.

11. The percentage of total unit ounces sold for an education level 'Graduated from college' is 20% less than 'Graduated high school'.

12. The percentage of total unit ounces sold for an education level 'Post graduate work' is 13.8% less than 'Graduated high school'.

13. The percentage of total unit ounces sold for an occupation level between 'Other' and 'Retired' is 38.1%.

14. The percentage difference of total unit ounces sold for a child group 'Child in [12-17)' and 'Family size>0 yet no children' is 18.6%.

15. The percentage difference of total unit ounces sold for a child group 'Children in [0-5) is 42.9% less than 'Family size>0 yet no children'.

INSIGHTS:

The percentage of units sold is more when the customer's education level is less than a high school graduate and less when his education level is greater than a high school graduate. The marital status of a customer does not have any effect on the number of units sold. As the average age of a customer's household increases, the number of units sold decreases. Intuitively, family size should influence the number of units sold but results show that family size doesn't have any effect.

Note: PANEL regression results are captured in a separate document.