**BUAN/MKT 6337.001 – Predictive Analytics using SAS**

**Group 3 – Project report**

**A picture containing table, food, cup, ground

Description automatically generated**

**PEANUT BUTTER**

*Scanner Data Analysis*

*Ayushi Paliwal*

*Keerthi Bisaramanempalli Reddiwandla*

*Minakshi Kesarwani*

*Priya Selarka*

*Vithika Agrawal*

**Introduction:**

Detailed data on sales of consumer goods obtained by 'scanning' the bar codes for individual products at electronic points of sale in retail outlets. The data can provide detailed information about quantities, characteristics and values of goods sold as well as their prices.

**Objective:**

Our objective is to analyze the various brands, regions, market share for the product – Peanut Butter and draw interpretations from running various models on the panel data set.

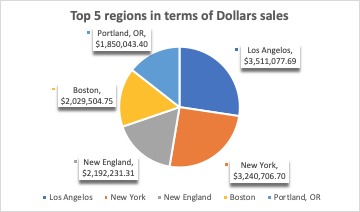
**Data:**

Product data: The details of the different varieties of peanut butter for various brands

Panel data: Details of customers across who purchased various products across different retail stores

Store data: Details of peanut butter purchases made by customers across different grocery stores

**Exploratory Data Analysis:**



**Top 7 brands**

The percentage of market share for the top 7 brands are displayed in the pie chart.  
  
We observe that 31% of market is occupied by JIF products and the second leading brand is Skippy with 26% market share.

Plotted besides are the average price per unit ounce in USD of the brands.

Here we observe that products of Smuckers have highest average price per unit ounce in USD of 0.169 and the second highest price is observed in Skippy products followed by JIF and Peter Pan.

**Problem Statement I**

Find out the best group of customers for the Brand JIF

**Analysis Method Implemented: RFM Clustering**

In this project, we studied the several attributes of the customers to unfold their buying patterns and clustered them into 5 segments on the basis of recency, frequency and monetary value of their purchase behaviour. After customer segmentation was done, further studies from the segmented data helped us to derive critical insights and make meaningful recommendations.

In this case, we identified the customer’s purchase patterns for JIF only, the leading Peanut Butter brand in the given dataset.

**Insights:**

A screenshot of a cell phone

Description automatically generated

1. We calculated the R, F and M scores and mapped the combined scores to form 5 clusters. On drilling down to see Cluster vs RFM Score, we figured out that Cluster 2 and Cluster 5 has high values of R, F and M in the range of 3-5 each.
2. Cluster 2 accounts for 64.71% of the total number of customers purchasing a JIF product and is our area of analysis as these are the customers who have transacted most recently and visit the store most frequently and spend large amounts to contribute towards the sales of JIF. These are the most loyal customers towards JIF.

We studied the demographics and other properties of these customers to make meaningful recommendations.

**Sales Analysis across Cluster:**

A screenshot of a cell phone

Description automatically generated

1. Following are the Total Sales in each of the 5 clusters:-

* Cluster 1 : $6.68\*349=$2331.32
* Cluster 2 : $23.64\*1471900=$34,795,716.2
* Cluster 3 : $22.69\*801952=$18,196.291
* Cluster 4 : $16.5\*288=$4752
* Cluster 5 : $7.37\*258=$1901.46

Here, we can see that Cluster 2 accounts for the most sales in dollars in JIF products, along with Recency, Frequency as well as Monetary.

**Distribution of Age across Clusters :**

A screenshot of a cell phone

Description automatically generated

1. Most of the customers from Cluster 2 belong to Age Band 4 which basically comprises people having age between 45 - 54’ years. Hence, we can say that people within this age band has a greater tendency to choose Jif Peanut Butter.

**Distribution of Occupation across Clusters:**

A screenshot of a cell phone

Description automatically generated

1. The most loyal customers (Cluster 2) are those who are mostly working in Professional or Technical Area (HH\_OCC=1)

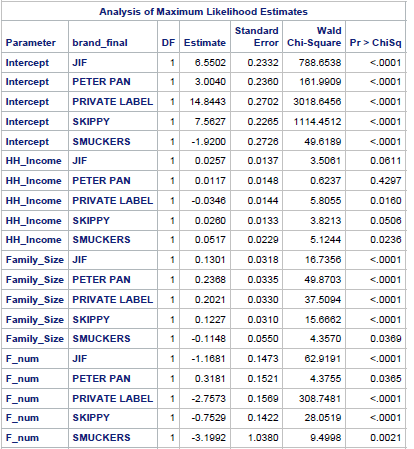
**Problem Statement II**

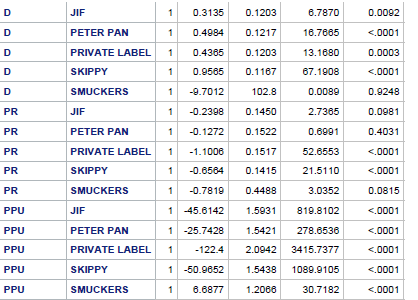
Find out the attributes that affect the brand of choice of the household panelists.

**Analysis Method Implemented: Multinomial Logistic Regression**

We have used the top 5 brands in terms of market share i.e., JIF,SKIPPY,PRIVATE LABEL,PETER PAN,SMUCKERS and clubbed all the other brands as “OTHER”.We used multinomial regression to study how the attributes like the income of the household,family size of the household,the price per unit ounce of the brand,whether the brand was on display or featured or if it had price reduction will affect the choice of the customer.

**Insights:**





1.The preference of the customers in the brands is PRIVATE LABEL>SKIPPY>JIF>PETER PAN>OTHER>SMUCKERS

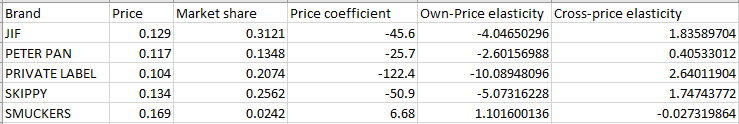
2. As the income of the household increases by 10k USD,the odss of choosing JIF compared to other brands will increase by 2.6%.Income is significant for JIF brand at 90% confidence.

3. As the number of family memebers increases by 1,the odss of choosing JIF compared to other brands will increase by 13.9%

4.The odds of choosing JIF when it is on display increases by 36.8% compared to choosing JIF when it is not on display.

5.As the price per unit ounce of a brand increases, the odds of choosing that brand decreases.

6.Price elasticities:



The own-price elasticity of brand JIF is -4.04 implyng that if JIF increases it’s price per unit ounce by 1 USD the odds of choosing that barnd decreases by 304%.

The cross-price elasticity of brand JIF is 1.836 implying that if the competitive brands increase their price per unit by 1 USD the odds of choosing brand JIF goes up by 83.6%

**Problem Statement III**

Find the most effective method of Marketing Communication from Types of Features and Display and Price Reduction.

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Analysis Method Implemented: Panel Regression**

**We have the data that portrays the different purchases by consumers along with the information of types of promotions (features and display) going on at that time. The panel regression thus is an apt method to determine which of the promotion methods and marketing communications were most effective to influence the consumer buying pattern.**

**Output:**

**1)We ran Hausman Test as P value was <0.05 so we rejected the null hypothesis and ran fixed effect model.**

**2)Here we can see that amongst the features and display promotions, Coupons are the most effective mode of marketing communications affecting the total sales in the most positive manner followed by display, large advertisements, medium advertisements and small advertisements. We can also assess that an increase in the price reduction is also significant and will increase total sales.**

**Insights and Interpretations:**

1. The regression output shows that attributes like types of advertisements for the products, and whether a Product is displayed- all these factors play a significant role in determining the sales of the product.
2. Compared to no coupons, A product advertised through coupons in a week in a store is expected to generate $ 371.24 more sales.
3. Compared to no advertisement, a product having a large advertisement in a week in a store is expected to increase total sales by $259.31.
4. Compared to no advertisement, a product having a medium advertisement in a week in a store is expected to increase total sales by $192.87.
5. Compared to no advertisement, a product having a small advertisement in a week in a store is expected to increase total sales by $47.56.
6. If a product is displayed at the store, the total sales in a week is expected to increase by $348.96 compared to the case when it is not displayed in that week in that store.
7. Compared to no Price reduction, If there is price reduction in a week in a store is expected to increase total sales by $30.2.

**Recommendations:**

1. The Jif brand should focus on coupon features to be marketed to the best customers (age group 45-54, working in Professional or Technical) to get highest increase in sales.
2. The households with higher income are preferring the Jif brand.Hence the management should come up with products that target this section of customers.