Project - Technical Project Report Phase-2
BUAN 6337.501
Shervin Shahrokhi Tehrani
Group 38

Submitted By:

Subhash Anagalabylu Ramachandra - sxa210266

Table of Contents

1. Company Overview

Conagra Brands, headquartered in Chicago, Illinois, is a North American packaged foods company. Since its inception in 1919, the company has grown to become one of the largest packaged foods companies in the United States. Conagra Brands' portfolio includes well-known brands like Chef Boyardee, Hunt's, Orville Redenbacher's, PAM, Peter Pan, and Slim Jim.

Conagra Brands manufactures a diverse range of food products such as frozen meals, canned goods, snacks, condiments, and more. Customers in the United States, Canada, and other international markets are served by the company's presence in both the retail and foodservice channels.

Conagra Brands has prioritized sustainability and innovation lately. To adapt to shifting consumer expectations, the company has created new items like plant-based meat substitutes and healthier snack options. Conagra Brands has also pledged to lessen its influence on the environment, with targets to cut greenhouse gas emissions and enhance packaging sustainability.

2. Data Set Overview

The Conagra dataset contains sales and distribution data for Conagra's food and household products in various markets across the United States. The dataset is divided into several sub-datasets, each of which contains unique information such as sales volume, revenue, product attributes, and promotional activities.

The IRI POS datasets provide point-of-sale data on Conagra's product sales volume, revenue, and distribution in various retail channels such as grocery stores and online retailers. The IRI POS datasets also show the effect of promotional activities like discounts and advertising campaigns on sales volume and revenue. This type of dataset consists of 10 excel sheets provided by Conagra for prediction analytics analysis.

The IRI Panel datasets provide consumer-level data on Conagra's products' purchasing behavior and product preferences. Household purchases, market penetration, market share, buyer behavior, demographic trends, and product preferences are all included in the datasets. Conagra provides 2 excel sheets for this type of dataset to analyze for prediction analytics.

Overall, the Conagra dataset is a valuable resource for businesses and analysts seeking insights into consumer behavior, product distribution, and sales performance for Conagra's products. This data can be used to inform business decisions and optimize business strategies.

3. Market Research of Product Categories

3.1 Table spreads

According to Grand View Research, the global table spreads market was worth USD 7.89 billion in 2020 and is predicted to increase at a compound annual growth rate (CAGR) of 4.9% between 2021 and 2028. The market is anticipated to increase during this time due to rising consumer health consciousness and rising demand for natural and organic spreads. The Grand View Research report also emphasizes how nut-based spreads, including almond and cashew butter, are becoming more and more well-liked as a nutritious and protein-rich substitute for conventional spreads. Besides boosting their demand in the market, consumers who follow vegan or plant-based diets like nut-based spreads.

According to Statista, the butter and margarine market in the United States was valued at roughly $7.7 billion in 2020. With a market value of around $4.1 billion, the margarine category in this market had a bigger share than butter, which had a market value of about $3.6 billion. Furthermore, according to the report, butter has become more and more popular as a natural alternative to margarine as well as health concerns regarding its effects on human health. In contrast, as customers become more health conscious and environmentally sensitive, the demand for organic and plant-based spreads has increased. With a rise in demand for organic and plant-based spreads, these trends are probably still present in 2023.

Below are some basic descriptions and compositions of the table spread categories.

3.1.1 **Total Butter** is a dairy product created by agitating milk or cream until the butterfat separates from the liquid. The butterfat is then washed with water, kneaded, and possibly salted to create the final product. Since it is high in saturated fat and cholesterol, it should be consumed in moderation. Butter is a versatile and delectable ingredient that can enhance the taste and texture of many dishes.

3.1.2 **Stick Butter** is a type of butter that is produced by first pressing the butterfat into a stick-like shape after milk or cream has been churned to separate the butterfat from the liquid. Stick butter is a versatile component in cooking and baking due to its rich, creamy texture and somewhat sweet, nutty flavor, just like other varieties of butter. It is used in moderation for a healthy and balanced diet.

3.1.3 **Tub Butter** is a type of butter that comes in a plastic container or tub, rather than the traditional rectangular stick shape. The production process of tub butter is, whereby milk or cream is churned to separate the butterfat from the liquid. The butterfat is then washed, shaped, and packaged in a tub with a lid to maintain freshness. Tub butter is commonly available in a variety of flavors such as garlic, herb, or honey, as well as low-fat or plant-based alternatives.

3.1.4 **Total Margarine** is a non-dairy, vegetable-based spread that was invented in the 19th century as a cheaper substitute for butter. It is produced by mixing water, emulsifiers, salt, vitamins, and artificial colors or tastes with vegetable oils such soybean, canola, or palm oil. Margarine is promoted as a healthier alternative to butter as it contains less saturated fat and cholesterol. It comes in a variety of flavors, including plain, herb, and garlic, as well as reduced-fat and plant-based options.

3.1.5 Stick Margarine is a type of margarine that is sold in the shape of rectangular sticks, close to butter. It is made from a combination of vegetable oils, water, emulsifiers, and other ingredients and is frequently marketed as a healthier alternative to butter due to its lower levels of saturated fat and cholesterol. It is made by combining vegetable oils with water and emulsifiers, then heating, cooling, and churning the mixture until it becomes a solid form that can be shaped into sticks.

3.1.6 Tub Margarine is margarine sold in a plastic tub or container rather than the traditional rectangular stick shape. It is frequently marketed as a healthier alternative to butter as it has lower levels of saturated fat and cholesterol. It is made by combining vegetable oils, water, and emulsifiers, then heating, cooling, and churning the mixture until it is soft and spreadable enough to be packaged in a plastic tub. It is mostly available in a variety of flavors and in both regular and reduced-fat versions.

3.2 Cooking Spray

According to MarketsandMarkets, the global cooking spray market was worth USD 1.14 billion in 2020, with the United States being one of the leading markets. The cooking spray market in the United States was valued at USD 149.3 million in 2020, with a 5.5% compound annual growth rate (CAGR) expected from 2021 to 2028.

One of the primary drivers of the cooking spray industry in the United States is the growing popularity of home cooking, which has been accelerated by the COVID-19 pandemic. The market is being driven by consumers' increased interest in healthier cooking methods and their growing health consciousness. Cooking spray is frequently touted as a healthier alternative to typical cooking oils since it can minimize the amount of oil used in cooking while also being lower in calories and fat.

Another factor driving the market is consumer awareness of the potential health dangers linked with various chemical additives generally used in cooking oils. Overall, the cooking spray market in the United States is likely to expand in the coming years, owing to rising demand for convenient and nutritious cooking solutions.

3.2.1 Cooking spray is a substance used in the kitchen to add a thin layer of oil to cooking surfaces and prevent food from sticking, making it a helpful tool for both home cooks and professional chefs to use it conveniently on cooking surfaces such as pans, grills, or baking sheets. It normally comes in an aerosol can and contains a mixture of emulsifiers and oils that enable a thin and even application. Depending on the brand and type, cooking spray's contents might change, but they often consist of a mixture of vegetable oil, canola oil, soybean oil, or other types of oils like an emulsifier.

3.3 Cooking & Salad Oil

The cooking and salad oil business in the United States is mature, with intense competition. The retail sales value of cooking oil in the United States was around USD 9.1 billion in 2020, according to a market research report by Statista. The market is predicted to grow at a CAGR of 3.9% between 2023 and 2030, spurred by the rise in consumer demand for natural and healthier cooking oils.

The growing demand for healthy cooking options is one of the major factors driving the US cooking and salad oil industry. Olive oil, avocado oil, and coconut oil are among the healthier and more natural cooking oils that consumers are choosing as they become more health conscious. The demand for healthier cooking oils is projected to increase as the trend toward healthy eating habits and lifestyles persists. Also fueling the market's expansion is the usage of salad oils in catering and foodservice.

The rising demand for premium and organic cooking and salad oils is another market-driving trend. Customers are prepared to spend extra for high-end, organic goods because they believe they are better-made and healthier. Yet, the market is also confronted with issues such as raw material price instability, which can have an influence on the profitability of producers and retailers. Furthermore, rising competition from alternative cooking goods such as cooking sprays is posing a challenge to the cooking and salad oil sector.

Overall, the cooking and salad oil market in the United States is likely to expand in the future years, driven by rising demand for healthier and more natural cooking options. The market is predicted to become increasingly competitive, and producers ought to focus on innovation and differentiation to stay competitive.

3.3.1 **Cooking and salad oil** is a product used in a range of cooking procedures, including sautéing, frying, baking, and roasting, as well as salad dressings and marinades. It's usually prepared from a blend of vegetable oils like soybean oil, canola oil, or sunflower oil. The term "vegetable oil" or "cooking oil" may be used to describe cooking and salad oil, which is frequently marketed in liquid form in bottles or jugs. It is also a good source of healthy fats, such as polyunsaturated and monounsaturated fats, which are beneficial for heart health.

4. Predictive Analytics for Conagra's Data Set

To recommend and inform decisions using predictive analytics for Conagra's data set in phase-2 of the project report, we identified the below three hypotheses or the business problem to be answered. Further, we applied the following steps to apply predictive analytics:

- **Data Preparation**
- **Feature Selection**
- **Model Selection**
- **Model Training**
- **Model Evaluation**

4.1 Hypothesis-1

4.1.1 What product attributes (form, flavor, size, etc.) have the greatest opportunity for CAG? What is the optimal assortment for Conagra's products in the Tablespreads category? Are there items where we have inefficiencies in product offering?

Business Statement to Prove:

- ✓ Hypothesis (Null): Irrespective of form or form value, sales are same across all forms for tablespreads.
- ✓ Hypothesis (Alternate): Specific combination of forms or form value contributes differently towards sales for tablespreads.

1. Case-1: Comparing within the "Conagra Brands" for all geographies.

- **Analysis-1**: For ANY/ No merchandise and CAG Form Value, ALL OTHER FORM is contributing more towards the Sales. While Conagra's total sales/market share is lesser compared to several other brands in comparison, it is also evident that Conagra is not making a lot of sales in the Sticks/Tubs forms which are the materially contributing to the sales of other competitor products. This can be considered as a scope to develop in the Sales/Tub form. The t-test analysis result is shown in Appendix Reference Section 6.1 under Exhibit 1.
- **Analysis-2**: For ANY/NO merchandise and Sub-Category Name, Margarine/Spreads is doing better than RFG Butter Blends. However, Conagra is not making RFG Butter at all, and RFG Butter Blends is the only category in Butters and comes as Tubs. The rest of the other products are made as Margarine/Spreads with a combination of Spray/Squeeze/Sticks/Tubs/Other Forms. The t-test analysis result is shown in Appendix Reference Section 6.1 under Exhibit 2.

2. ***Case-2: Comparing all the brands without "Conagra Brands" and "Private Labels" for all geographies.***

- **Analysis-1**: For ANY merchandise and CAG Form Value, STICKS' sales are significantly higher than SPRAY/SQUEEZE forms. TUBS's sales are significant but it is not a good idea to sell in this combination as it's not producing many sales. The t-test analysis result is shown in Appendix Reference Section 6.1 under Exhibit 3.
- **Analysis-2**: For NO merchandise and CAG Form Value, STICKS' sales are significantly on the higher side. TUBS's sales are doing significantly better in this combination and can be thought of selling in this combination than in recommendation1. The t-test analysis result is shown in Appendix Reference Section 6.1 under Exhibit 4.
- **Analysis-3**: For ANY merchandise and Sub-Category Name, RFG BUTTER sales are significantly higher than any other subcategory. The t-test analysis result is shown in Appendix Reference Section 6.1 under Exhibit 5.
- **Analysis-4**: In areas where there is NO merchandise in place, the RFG BUTTER BLENDS Sub-Category' sales are significantly higher than the rest of the subcategories, followed by the RFG Butter subcategory. The t-test analysis result is shown in Appendix Reference Section 6.1 under Exhibit 6.
- **Analysis-5**: For ANY merchandise and considering the interaction between the CAG Form Value and the Sub-Category Name, STICKS with RFG BUTTER is significantly more profitable to sell together. The presence of RGF BUTTER in the form of Sticks is helping the sales significantly. Another Form: 'Tubs' sales are better and significant where there are other forms/categories in place. Given that we have RFG Butter and only Butter Blend in Tubs form, the sales of Butter in Tub Form are causing decremental sales and are significant. Therefore, would be ideal to produce Tubs with Margarine/spreads. The t-test analysis result is shown in Appendix Reference Section 6.1 under Exhibit 7.
- **Analysis-6**: For NO merchandise and considering the interaction between the CAG Form Value and the Sub-Category Name, Sticks with Margarine/Spreads and Butter have material and statistical significance. However, it also occurs that Tubs with Butter are causing decremental sales, as was observed in the Any Merchandise section. The t-test analysis result is shown in Appendix Reference Section 6.1 under Exhibit 8.
- **Analysis-7**: For No Merch, RFG Butter and Margarine sold as a 16 OZ item materially and statistically outperforms the sales of any other category and size combination. The regression analysis result is shown in Appendix Reference Section 6.1 under Exhibit 9.

```
lm_model <- lm(`Dollar Sales No Merch` ~  `Sub-Category Name` * `CAG Ounces Value OZ`, data = conagra_final); summary(lm_model)

Call:
lm(formula = `Dollar Sales No Merch` ~ `Sub-Category Name` *
    `CAG Ounces Value OZ`, data = conagra_final)

Residuals:
    Min      1Q  Median      3Q     Max
 -96561  -29002  -13087    1000 1922387
`Sub-Category Name`MARGARINE/SPREADS:`CAG Ounces Value OZ`16 OZ     2.034e+04  6.367e+03   3.194 0.001405 **
`Sub-Category Name`RFG BUTTER:`CAG Ounces Value OZ`16 OZ            3.484e+04  1.831e+03  19.028  < 2e-16 ***
```

- **Analysis-8**: For Any Merch, RFG Butter sold as a 16 OZ item materially and statistically outperforms the sales of any other category and size combination. The regression analysis result is shown in Appendix Reference Section 6.1 under Exhibit 9.

```
lm_model <- lm(`Dollar Sales Any Merch` ~  `Sub-Category Name` * `CAG Ounces Value OZ`, data = conagra_final); summary(lm_model)

Call:
lm(formula = `Dollar Sales Any Merch` ~ `Sub-Category Name` *
    `CAG Ounces Value OZ`, data = conagra_final)

Residuals:
    Min      1Q  Median      3Q     Max
 -28159   -7184   -4489    -237 4377702
`Sub-Category Name`RFG BUTTER:`CAG Ounces Value OZ`16 OZ            21254.261   1533.988  13.856  < 2e-16 ***
```

Final Recommendation

Conagra, as a brand, has not been making a lot of sales in the Sticks/Tubs category, while most of the sales are from forms other than Sticks/Tubs. Also, only Margarine/Spreads and RFG Butter Blends is the most significant selling category by them.

We believe the market share of Conagra is not amongst the top competitors due to the underlying shortfall in the product design/category combinations manufactured and would recommend addressing the below shortfalls to match the market demands. What we have observed to be successful in sales for rest of the brands across United States is as follows:

- During merchandising, having RFG Butter in the form of Sticks.
- During merchandising, having Margarine/Spreads in the form of Tubs.
- During No merchandising, Sticks as such is selling more. So, having RFG Butter and Margarine/Spreads in the form of Sticks will boost sales.
- The ideal size to reach customer demands is 16 OZ of RFG Butter and Margarine/Spreads during No merchandising and similar 16 OZ of RFG Butter during merchandising.

Therefore, we believe that Conagra should focus on short term goal of increasing the marketing for margarine/spreads and RFG Butter Blends and try to optimize it with 16 OZ and Tub or Stick assortment more than “ALL OTHER FORMS” to increase its market share. This strategy should help Conagra to elevate its ranking amongst other competitors.

Additionally, we believe that Conagra should focus on long term goal of producing RFG Butter and selling it both with and without merchandising and combine it with the 16 OZ and Stick category to increase its market share.

4.2 Hypothesis-2

4.3 Hypothesis-3

4.3.1 Should Conagra have varying merchandising strategies by brand or market? Any segments that respond better to merchandising activity?

Business Statement to Prove:

The entry of a competitor brand product (COUNTRY CROCK RFG VEGETABLE OIL SPREAD INDIVDL WRAP IN BOX STICK 16 OZ – 0027400000221) in the Northeast - IRI Standard - Multi Outlet + Conv Region has had a significant impact on the sales of Conagra brand product (BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222). This impact needs to be analyzed to determine if it is statistically significant, and if so, what steps can be taken to mitigate the effects and maintain Conagra's market share.

- ✓ Hypothesis (Null): The launch of the competitor (Upfield) product on a particular date in the Northeast - IRI Standard - Multi Outlet + Conv Region did not have a significant impact on the total sales of Conagra Brands product.
- ✓ Hypothesis (Alternate): The launch of the competitor (Upfield) product on a particular date in the Northeast - IRI Standard - Multi Outlet + Conv Region (3/31/2019) had a significant impact on the sales of Conagra Brands product, leading to a difference in sales before and after the launch.

1. ***Key Insights and Findings from Prior Report submitted in HW1***

- Northeast region holds significant potential for Table Spread category across all brands, with a 20.27% contribution to No Merch dollar sales and a 21.73% contribution to Any Merch dollar sales.
- Incremental dollar sales to incremental unit sales and incremental volume sales ratios for Northeast are highest, indicating for every incremental unit or volume sold, there is a gain of $3.28 and $3.36 as incremental dollar sales, respectively.
- Merchandising activities in the Northeast region can prove to be highly beneficial for all brands in the Table Spread category, using strategies like buy one get one offer or King Size pack offer based on the ratio of dollar sales to unit or volume.
- While Northeast ranks second for Conagra Brands, the BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ product dominates the sales charts, mostly coming from the Southeast region.
- A focused approach to tailoring merchandising strategies for **BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ** product in the **Northeast region** can help Conagra Brands capture a larger market share, increase brand awareness, and boost overall sales.

2. ***Case-1: Competitor Product Entries Similar to Conagra Brands' BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ -0029000008222 Between 2019-2020.***

- **Analysis-1**: We analyzed sales data for Northeast - IRI Standard - Multi Outlet + Conv Region across years to identify new product entries launched between 2019-2020. Filtered out products launched post-2019 and before 2020, identified incremental sales for these products and sorted them in decreasing order.

UPC 13 dig	Time	Incremental Dolla
104787564	2019-12-01 00:00:00	$ 447,549.25
104787564	2019-12-01 00:00:00	$ 427,704.96
193340037	2019-12-01 00:00:00	$ 319,819.44
98864533	2019-04-21 00:00:00	$ 305,600.62
104787564	2019-11-24 00:00:00	$ 278,838.48
98864533	2019-07-21 00:00:00	$ 228,077.97
193340037	2019-12-01 00:00:00	$ 217,869.56
104787564	2019-12-29 00:00:00	$ 190,482.69
104787564	2019-08-25 00:00:00	$ 166,543.74
104787564	2019-11-03 00:00:00	$ 162,550.68
104787564	2019-12-22 00:00:00	$ 158,531.26
98864533	2019-06-23 00:00:00	$ 154,052.68
98864533	2019-03-10 00:00:00	$ 144,270.29
193340037	2019-12-29 00:00:00	$ 132,181.61
104787564	2019-09-01 00:00:00	$ 131,752.15
98864533	2019-05-12 00:00:00	$ 124,875.54
104787564	2019-11-24 00:00:00	$ 123,155.59
193340037	2019-12-22 00:00:00	$ 118,205.38
193340037	2019-11-24 00:00:00	$ 115,591.63
98864533	2019-01-27 00:00:00	$ 114,195.27
104787564	2019-11-10 00:00:00	$ 103,903.52

- **Analysis-2**: We filtered out private label products and only included those with similar CAG Tier and CAG Ounces values as our product, BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222. Identified **COUNTRY CROCK RFG VEGETABLE OIL SPREAD INDIVDL WRAP IN BOX STICK 16 OZ – 0027400000221** as the competitor product launched during 2019 - 2020 with the same CAG Tier and CAG Ounces values as our product BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222. This product was launched on **3/31/2019**.

3. ***Case-2: Paired T-Test on BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222 Product, Total Sales Data.***

- **Analysis-1**: Base Sales and Incremental Sales were added together to calculate Total Sales for the BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222 Product of Conagra Brand. We split the data of Total sales of Conagra Brands' product (29000008222) into two parts. Sales before 3/31/2019 and Sales After 3/31/2019. We calculated summary statistics for both these data. We saw that the number of entries after 3/31/2019 (Week of launch of the competitor brand - 27400000221) is 197, whereas before that date, there were only 63 entries.

	Total Sales
count	63.000000
mean	121807.315899
std	28092.413232
min	92526.173171
25%	105357.239745
50%	115319.079789
75%	128126.411785
max	229157.343908

	Total Sales
count	197.000000
mean	110666.462464
std	47508.842346
min	61303.812128
25%	81999.789444
50%	97788.219119
75%	119764.905619
max	421601.811122

- **Analysis-2**: After observing that the number of entries after 3/31/2019 for our brand BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222 was 197 and before it was just 63, we restricted our analysis to the first 63 entries after 3/31/2019. We then conducted a paired t-test on the total sales data for our brand before and after 3/31/2019. It was necessary to restrict our analysis to only the first 63 entries before the launch of the competitor brand, as there were no further entries available to test the hypothesis.

	Total Sales
count	63.000000
mean	121851.817794
std	28590.309903
min	92586.927963
25%	102311.512079
50%	111727.515514
75%	131279.588159
max	205548.431587

```
stats.ttest_rel(df_1_before_date_sales, df_1_after_date_sales[0:63])

TtestResult(statistic=-0.008582089177142475, pvalue=0.9931801227820116, df=62)
```

- After conducting a paired t-test on the total sales data of Conagra Brands' product BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222, before and after the launch of Upfield's product COUNTRY CROCK RFG VEGETABLE OIL SPREAD INDIVDL WRAP IN BOX STICK 16 OZ – 0027400000221, we observed a significantly higher p-value with a 90% confidence interval. This indicates that we reject the null hypothesis, which states that the launch of Upfield's product did not have a significant impact on Conagra Brands' total sales. Instead, we accept the alternative hypothesis, which suggests that the **launch of Upfield's product on 3/31/2019 in the Northeast - IRI Standard -**

Multi Outlet + Conv Region had a significant impact on Conagra Brands' total sales of BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222, leading to a difference in sales before and after the launch. This analysis was conducted after filtering the sales data of BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222 based on specific criteria, including the product's CAG Tier Value, CAG Ounces Value, and launch date, which allowed us to make a more accurate comparison of the sales data.

	Time	Product Description	Total Sales
0	2023-01-01	COUNTRY CROCK RFG VEGETABLE OIL SPREAD INDIVDL...	52536.631993
1	2023-01-01	BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREA...	187790.678180
2	2022-12-25	COUNTRY CROCK RFG VEGETABLE OIL SPREAD INDIVDL...	110383.880858
3	2022-12-25	BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREA...	421601.811122
4	2022-12-18	COUNTRY CROCK RFG VEGETABLE OIL SPREAD INDIVDL...	82420.153060

```
[ ] Total_Sales_A = df2[df2['Product Description'] == df2['Product Description'].unique()[0]]['Total Sales']
```

```
[ ] Total_Sales_B = df2[df2['Product Description'] == df2['Product Description'].unique()[1]]['Total Sales']
```

```
[ ] from scipy.stats import pearsonr
```

```
[ ] corr, pval = pearsonr(Total_Sales_A, Total_Sales_B)
```

```
[ ] print("Pearson correlation coefficient:", corr)
    print("p-value:", pval)
```

```
Pearson correlation coefficient: 0.6829773518542332
p-value: 2.081651225096214e-28
```

```
import pandas as pd
import statsmodels.api as sm

# Load the excel file into a pandas dataframe
df2 = pd.read_excel("/content/Correlation Analysis Data.xlsx")

# Create separate dataframes for each product's sales
sales_a = df2[df2['Product Description'] == df2['Product Description'].unique()[0]][['Time', 'Total Sales']]
sales_b = df2[df2['Product Description'] == df2['Product Description'].unique()[1]][['Time', 'Total Sales']]

# Merge the two dataframes on the 'Time' column
merged_sales = pd.merge(sales_a, sales_b, on='Time')

# Rename the columns to be more descriptive
merged_sales.columns = ['Time', 'Total_Sales_A', 'Total_Sales_B']

# Create a new column for the sales of product A minus the mean of product A sales
merged_sales['A_diff'] = merged_sales['Total_Sales_A'] - merged_sales['Total_Sales_A'].mean()

# Run the regression analysis with product B sales as the dependent variable and product A sales as the independent variable
X = merged_sales['A_diff']
y = merged_sales['Total_Sales_B']
X = sm.add_constant(X)
model = sm.OLS(y, X)
results = model.fit()
print(results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            Total_Sales_B   R-squared:                       0.466
Model:                              OLS   Adj. R-squared:                  0.464
Method:                   Least Squares   F-statistic:                     170.5
Date:                  Sun, 07 May 2023   Prob (F-statistic):           2.08e-28
Time:                          17:44:21   Log-Likelihood:                -2338.6
No. Observations:                   197   AIC:                             4681.
Df Residuals:                       195   BIC:                             4688.
Df Model:                             1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1.107e+05   2478.771     44.646      0.000    1.06e+05    1.16e+05
A_diff         1.9319      0.148     13.057      0.000       1.640       2.224
==============================================================================
Omnibus:                       28.619   Durbin-Watson:                   0.230
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               37.792
Skew:                           0.910   Prob(JB):                     6.22e-09
Kurtosis:                       4.137   Cond. No.                     1.68e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.68e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Based on the given regression output, it seems that there is only one independent variable, which is the difference in sales between two products (A_diff). Therefore, there is no variable to remove. However, as mentioned in the notes section, there might be strong multicollinearity or other numerical problems, which could affect the accuracy and reliability of the regression model. It might be worth exploring this issue further and consider alternative modeling approaches or data preparation techniques to address this problem.

```
import pandas as pd
import statsmodels.api as sm
from statsmodels.stats.outliers_influence import variance_inflation_factor

# Load the excel file into a pandas dataframe
```

```
df2 = pd.read_excel("/content/Correlation Analysis Data.xlsx")

# Create separate dataframes for each product's sales
sales_a = df2[df2['Product Description'] == df2['Product Description'].unique()[0]][[
'Time', 'Total Sales']]
sales_b = df2[df2['Product Description'] == df2['Product Description'].unique()[1]][[
'Time', 'Total Sales']]

# Merge the two dataframes on the 'Time' column
merged_sales = pd.merge(sales_a, sales_b, on='Time')

# Rename the columns to be more descriptive
merged_sales.columns = ['Time', 'Total_Sales_A', 'Total_Sales_B']

# Create a new column for the sales of product A minus the mean of product A sales
merged_sales['A_diff'] = merged_sales['Total_Sales_A'] -
 merged_sales['Total_Sales_A'].mean()

# Run the regression analysis with product B sales as the dependent variable and prod
uct A sales as the independent variable
X = merged_sales['A_diff']
y = merged_sales['Total_Sales_B']
X = sm.add_constant(X)
model = sm.OLS(y, X)
results = model.fit()

# Get the VIF scores for the independent variables
vif = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

# Print the VIF scores for each variable
for i in range(len(vif)):
    print('Variable', i, '\tVIF Score:', round(vif[i], 2))
```

```
Variable 0  VIF Score: 1.0
Variable 1  VIF Score: 1.0
```

Based on the analysis of the sales data for Product A and Product B, we conducted a variance inflation factor (VIF) analysis to quantify the severity of the multicollinearity problem between the two products. The VIF analysis indicated that there is no significant multicollinearity problem between the two products as the VIF scores were both 1.0. Therefore, we can conclude that the regression analysis results are reliable and the relationship between the sales of the two products can be interpreted with confidence.

The Pearson correlation coefficient of 0.68 indicates a strong positive correlation between the sales of Product A (competitor product) and Product B (our product). The p-value of 2.08e-28 indicates that this correlation is statistically significant, suggesting that there is a genuine relationship between the sales of the two products.

Combining this with the regression analysis, which shows that the sales of Product B are positively impacted by the sales of Product A, it can be inferred that when the sales of the competitor's product (Product A) increase, the sales of our product (Product B) also tend to increase. Therefore, it can be concluded that there is a potential opportunity for our company to leverage the existing market demand for the competitor's product by improving the marketing and advertising strategies for our product, highlighting the unique features and benefits of our product to the target audience, and offering competitive pricing and promotions.

An actionable insight from this analysis could be to focus on increasing the visibility and appeal of our product in the market by launching targeted campaigns, conducting product demos, collaborating with retailers to promote our product in-store, and offering discounts and promotions to incentivize customers to choose our product over the competitor's product. Additionally, monitoring the sales trends and customer preferences in real-time and making necessary adjustments to the product offering and marketing strategies can help the company stay ahead of the competition and capitalize on the market opportunities.

4. ***Case-3: Analyzing Competitor's (COUNTRY CROCK TABLESPREADS) Stronger Customer Base Through Buyer Distribution Index***

- **Analysis-1**: To gain insights into the strength of the customer base of **COUNTRY CROCK TABLESPREADS** compared to **BLUE BONNET TABLESPREADS**, we conducted an analysis of the buyer distribution index of both products from IRI_Panel_Buyer Distribution and Index file. This involved subtracting the buyer distribution index of BLUE BONNET TABLESPREADS from that of COUNTRY CROCK TABLESPREADS to determine the difference in their customer bases. The buyer distribution index is a measure of the share of buyers for a particular product, and by analyzing the difference in indices between the two products, we aimed to **identify any areas where COUNTRY CROCK TABLESPREADS had a stronger customer base than BLUE BONNET TABLESPREADS and vice versa**.
- **Analysis-2**: We sorted the difference column in descending order to identify which brand had a stronger customer base in the segment. A difference of zero indicates equal customer strength between the brands, while a positive difference indicates that Conagra's product has a stronger

customer base, and a **negative difference indicates that the competitor's product has a stronger customer base**.

- Using the results of this analysis, we aim to improve our marketing strategies and increase our market share. By understanding the strengths and weaknesses of our product relative to the competitor, we can develop targeted marketing campaigns and promotional offers that better resonate with our target customers. Ultimately, this analysis will allow us to position our product more effectively in the market and increase our share of the table spreads segment.

Geography	Time	Target Group	HH Demo Summary	BLUE BONNET TABLESPREADS Buyer Distribution	COUNTRY CROCK TABLESPREADS Buyer Distribution	Difference
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male Empl - ge 35 hrs/wk	39	48	-9
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	County Size A Top 25 Mkts	25	33	-8
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Child - 1+	30	38	-7
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem White Collar	35	43	-7
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Upper (>=$70k 1P + $30k all others)	21	28	-7
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Income ge $100K	19	26	-7
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Millennials-Older (Born 1981-1989)	15	21	-6
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male White Collar	21	27	-6
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem Empl - ge 35 hrs/wk	31	37	-6
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem 35-44 years old	13	19	-5
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male 35-44 years old	12	18	-5
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	LS2 - Young Families (children <12)	15	20	-5
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem 25-34 years old	14	19	-5
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Married	62	66	-4
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Acculturated Hispanic	7	12	-4
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem Graduated College	26	30	-4
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Pet Owner	61	66	-4
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male Graduated College	15	19	-4
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	5+ Person HH	12	16	-4
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	M10 - 35-44 w/Kids $70k+	4	8	-3
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Millennials-Younger (Born 1990-1996)	8	12	-3
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem Professional	17	20	-3
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	LS1 - Getting Started (no child, age <45)	10	13	-3
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male Professional	9	12	-3
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male 45-54 years old	13	16	-3
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Generation X (Born 1965-1980)	29	32	-3
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Income $70-99.9K	14	17	-3
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Race - Other / Unknown	5	8	-3
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	LS3 - Raising Teens (oldest child 12-17)	16	18	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem Prop/Managers/Officials	8	10	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	1 Dog	26	28	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Age of Children 0 to 5 only	5	7	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male Blue Collar	20	23	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	2+ Cats	17	19	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male Craftsmen/Foreman (Skilled)	10	12	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	4 Person HH	14	16	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	3 Person HH	16	18	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Age of Children 0 to 5 and 6 to 11	4	6	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Middle ($30-70k 1P + $5k per add`l Person)	39	41	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	M15 - 45-64 No Kids $100k+	7	9	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem Post Graduate School	9	11	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Income $50-69.9K	13	15	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male Prop/Managers/Officials	7	9	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male Post Graduate School	7	8	-2
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	M17 - 45-64 w/Kids $70k+	5	6	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Race - Asian	1	3	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Rent Home	26	28	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem 45-54 years old	17	18	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	M06 - Under 35 w/Kids $70k+	4	5	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male Some College	17	19	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Generation Z (Born 1997 and After)		1	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Age of Children 12 to 17 only	9	10	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	M08 - 35-44 No Kids $70k+	1	2	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Single	14	15	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem Empl - lt 35 hrs/wk	16	17	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	M05 - Under 35 w/Kids <$70k Non-City	3	4	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Age of Children 6 to 11 only	5	6	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male 25-34 years old	8	9	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem Clerical	6	7	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Age of Children 6 to 11 and 12 to 17	4	5	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	M09 - 35-44 w/Kids <$70k	6	6	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	M02 - Under 35 Single No Kids City	2	3	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	M03 - Under 35 No Kids Non-City	2	3	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Fem Sales	5	6	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	M01 - Under 35 Married No Kids City	2	2	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male Sales	3	4	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Race - White	74	75	-1
Total US - All Outlets	52 Weeks Ending Jan-01-2023	All Households	Male Other Collar	4	4	-1

Final Recommendation

- Conduct a hypothesis test: It is recommended to conduct a hypothesis test to determine if the launch of the competitor's product (COUNTRY CROCK RFG VEGETABLE OIL SPREAD INDIVDL WRAP IN BOX STICK 16 OZ –

0027400000221) in the Northeast - IRI Standard - Multi Outlet + Conv Region had a significant impact on the sales of Conagra's BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222 product. If the test confirms that the competitor's entry had a significant impact, steps can be taken to mitigate its effects and maintain Conagra's market share.

- Focus on tailoring merchandising strategies: Based on the analysis, it is recommended to tailor merchandising strategies for BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ product in the Northeast region to capture a larger market share, increase brand awareness, and boost overall sales. Using strategies like buy one get one offer or King Size pack offer based on the ratio of dollar sales to unit or volume can be highly beneficial for all brands in the Table Spread category.
- Monitor competition closely: To stay ahead of the competition, it is recommended to monitor competitor products' entries in the Northeast - IRI Standard - Multi Outlet + Conv Region closely. By analyzing sales data, filtering out private label products, and identifying products with similar CAG Tier and CAG Ounces values as Conagra's products, the impact of the competition can be determined early on, and steps can be taken to maintain market share.
- Targeting Demographic with Lower Preference: Target male employees who work 35 hours per week and reside in the top 25 markets of County Size A. This demographic has shown a lower preference for Conagra Brands' Blue Bonnet Tablespreads compared to the competitor's Country Crock brand, as indicated by the negative difference in the buyer distribution index. Therefore, this segment presents an opportunity for targeted merchandising efforts.

4.4 Hypothesis-3

4.4.1 To what extent does seasonality influence the sales performance of products in the table spread category, which products are used more across different seasons, and in which form do people prefer to buy them? and if there is an effect, what are the corresponding changes in sales trends, and can we improve our regression model using this information?

Business Statement to Prove:

- ✓ Hypothesis (Null): There are no changes in the sales of table spreads across different seasons.
- ✓ Hypothesis (Alternate): There is a significant change in sales of table spreads across different seasons.

- **Analysis-1**: To analyze the impact of seasonality on sales performance in the table spread category, a new column called "Season" was created based on the data in the "Time" column. The "Season" column was divided into four categories: winter, summer, spring, and fall. Total sales were then calculated for each season to identify any seasonal patterns in sales performance. It was observed (Exhibit 10) that fall had the highest sales with a percentage of 31.6%, followed by winter with 28%, spring with 25%, and the lowest sales were in summer with 15.4%. These findings suggest that seasonality significantly impacts sales performance in the table spread category, with customers being more likely to purchase these products during the colder months.

- **Analysis-2**: To verify our observations regarding the impact of seasonality on sales performance in the table spread category, we conducted an ANOVA test. Initially, we plotted a graph of the total sales across different seasons (Exhibit 11), but we observed that the data was not normally distributed, violating the assumptions of the ANOVA test. To address this, we took the log of the data and plotted a new graph (Exhibit 12), which showed a normal distribution. Using this new graph, we performed an ANOVA test and observed that the F-value (Exhibit 13) was highly significant, indicating that we can reject the null hypothesis and conclude that seasonality has an effect on sales performance in the table spread category. These findings provide further support for our previous observations and highlight the importance of considering seasonality when analyzing sales data in this category.
- **Analysis-3**: Based on our observations and statistical analysis, we plan to incorporate the season variable into our regression model to predict changes in sales performance across different regions and identify which product sizes are preferred during certain seasons. By taking into account the impact of seasonality on sales, we can develop more accurate models that can be used by Conagra to forecast demand and optimize their product offerings considering the impact of seasonality on sales, we can develop more accurate models that Conagra can use. These insights can help improve sales performance and provide a competitive advantage in the table spread category.

5. References

1. https://www.conagrabrands.com/
2. https://www.grandviewresearch.com/industry-analysis/butter-market-report
3. https://www.fortunebusinessinsights.com/butter-market-106457
4. https://www.statista.com/outlook/cmo/food/oils-fats/margarine/united-states#:~:text=Revenue%20in%20the%20Margarine%20segment,US%245%2C426.00m%20in%202023).
5. https://www.mordorintelligence.com/industry-reports/margarine-market
6. https://www.theinsightpartners.com/reports/cooking-spray-market
7. https://www.marketwatch.com/press-release/cooking-spray-market-booming-with-rising-demands-and-massive-opportunities-2023---2028-2023-02-21
8. https://www.fortunebusinessinsights.com/cooking-oil-market-106391
9. https://www.statista.com/outlook/cmo/food/oils-fats/edible-oils/united-states
10. https://www.factmr.com/report/670/cooking-oil-market
11. https://www.businesswire.com/news/home/20230206005341/en/Global-Vegetable-Oil-Market-Report-2023-to-2030---Increasing-Consumption-of-Plant-based-Food-Products-Drives-Growth---ResearchAndMarkets.com
12. https://www.marketwatch.com/press-release/used-cooking-oil-market-is-the-fastest-rising-industry-from-2023-to-2028-2023-02-15

6. Appendix

This section consists of all the graphical representations of the trends observed in the dataset and discussed in chapter 4 of this report.

6.1 Reference Section 1

This section relates to the references for the recommendations, linear regressing and modelling analysis mentioned in Hypothesis-1 (Section 4.1)

- **Exhibit1**

```
> lm_model <- lm(`Dollar Sales Any Merch` ~  `CAG Form Value`, data = conagra_final);summary(lm_model)

Call:
lm(formula = `Dollar Sales Any Merch` ~ `CAG Form Value`, data = conagra_final)

Residuals:
   Min     1Q Median     3Q    Max 
-31238  -3149  -2022   -262 252957 

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                     31244.7      493.7   63.28   <2e-16 ***
`CAG Form Value`SPRAY/SQUEEZE -28517.1      608.0  -46.91   <2e-16 ***
`CAG Form Value`STICKS        -29761.1      537.2  -55.40   <2e-16 ***
`CAG Form Value`TUBS          -27886.4      505.9  -55.12   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10070 on 11835 degrees of freedom
Multiple R-squared:  0.2145,	Adjusted R-squared:  0.2143 
F-statistic:  1077 on 3 and 11835 DF,  p-value: < 2.2e-16
```

```
> lm_model <- lm(`Dollar Sales No Merch` ~  `CAG Form Value`, data = conagra_final);summary(lm_model)

Call:
lm(formula = `Dollar Sales No Merch` ~ `CAG Form Value`, data = conagra_final)

Residuals:
    Min      1Q  Median      3Q     Max 
-187962  -21939  -10675   11223  846860 

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                     188073       2089   90.02   <2e-16 ***
`CAG Form Value`SPRAY/SQUEEZE  -149562       2572  -58.14   <2e-16 ***
`CAG Form Value`STICKS         -175804       2273  -77.33   <2e-16 ***
`CAG Form Value`TUBS           -161063       2141  -75.24   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 42610 on 11835 degrees of freedom
Multiple R-squared:  0.3433,	Adjusted R-squared:  0.3431 
F-statistic:  2062 on 3 and 11835 DF,  p-value: < 2.2e-16
```

- **Exhibit2**

```
> lm_model <- lm(`Dollar Sales Any Merch` ~  `Sub-Category Name` , data = conagra_final); summary(lm_model)

Call:
lm(formula = `Dollar Sales Any Merch` ~ `Sub-Category Name`, 
    data = conagra_final)

Residuals:
   Min     1Q Median     3Q    Max 
 -3785  -3672  -3112   -635 280417 

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)                           3784.92      37.34  101.36   <2e-16 ***
`Sub-Category Name`RFG BUTTER BLENDS -3148.50     301.52  -10.44   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9755 on 69307 degrees of freedom
Multiple R-squared:  0.001571,	Adjusted R-squared:  0.001556 
F-statistic:   109 on 1 and 69307 DF,  p-value: < 2.2e-16
```

```
> lm_model <- lm(`Dollar Sales No Merch` ~  `Sub-Category Name` , data = conagra_final); summary(lm_model)

Call:
lm(formula = `Dollar Sales No Merch` ~ `Sub-Category Name`, data = conagra_final)

Residuals:
    Min      1Q  Median      3Q     Max 
 -24355  -21795  -14117    5389 1010579 

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)                           24354.6      152.3   159.9   <2e-16 ***
`Sub-Category Name`RFG BUTTER BLENDS -22752.4     1229.9   -18.5   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 39790 on 69307 degrees of freedom
Multiple R-squared:  0.004914,	Adjusted R-squared:  0.004899 
F-statistic: 342.2 on 1 and 69307 DF,  p-value: < 2.2e-16
```

- **Exhibit3**

```
> lm_model <- lm(`Dollar Sales Any Merch` ~  `CAG Form Value`, data = conagra_final);summary(lm_model)

Call:
lm(formula = `Dollar Sales Any Merch` ~ `CAG Form Value`, data = conagra_final)

Residuals:
    Min      1Q  Median      3Q     Max 
 -19691  -13349   -5760   -2816 4386169 

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                     3797.0      224.0  16.952   <2e-16 ***
`CAG Form Value`SPRAY/SQUEEZE  11562.7     1002.5  11.533   <2e-16 ***
`CAG Form Value`STICKS         15894.5      276.1  57.564   <2e-16 ***
`CAG Form Value`TUBS            2537.3      263.6   9.625   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 55050 on 336697 degrees of freedom
Multiple R-squared:  0.0148,	Adjusted R-squared:  0.01479 
F-statistic:  1686 on 3 and 336697 DF,  p-value: < 2.2e-16
```

- **Exhibit 4**

```
> lm_model <- lm(`Dollar Sales No Merch` ~  `CAG Form Value`, data = conagra_final);summary(lm_model)

Call:
lm(formula = `Dollar Sales No Merch` ~ `CAG Form Value`, data = conagra_final)

Residuals:
    Min      1Q  Median      3Q     Max 
 -39149  -31651  -19764   -1738 1928864 

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                    10731.2      275.7   38.92   <2e-16 ***
`CAG Form Value`SPRAY/SQUEEZE  19121.0     1234.0   15.49   <2e-16 ***
`CAG Form Value`STICKS         28418.0      339.9   83.61   <2e-16 ***
`CAG Form Value`TUBS           21752.7      324.5   67.04   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 67770 on 336697 degrees of freedom
Multiple R-squared:  0.02082,	Adjusted R-squared:  0.02082 
F-statistic:  2387 on 3 and 336697 DF,  p-value: < 2.2e-16
```

- **Exhibit 5**

```
> lm_model <- lm(`Dollar Sales Any Merch` ~  `Sub-Category Name` , data = conagra_final); summary(lm_model)

Call:
lm(formula = `Dollar Sales Any Merch` ~ `Sub-Category Name`, 
    data = conagra_final)

Residuals:
    Min      1Q  Median      3Q     Max 
 -14083  -13600   -7081   -3522 4391778 

Coefficients:
                                                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                        125.42    3520.50   0.036   0.9716    
`Sub-Category Name`MARGARINE/SPREADS              7058.37    3524.67   2.003   0.0452 *  
`Sub-Category Name`RFG BUTTER                    13957.25    3522.88   3.962 7.44e-05 ***
`Sub-Category Name`RFG BUTTER BLENDS              4691.19    3529.32   1.329   0.1838    
`Sub-Category Name`RFG FLAVORED MILK              1537.90    9234.82   0.167   0.8677    
`Sub-Category Name`RFG SAUCE/GRAVY/MARINADE MIXES  -78.15    6897.92  -0.011   0.9910    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 55330 on 336695 degrees of freedom
Multiple R-squared:  0.004944,	Adjusted R-squared:  0.004929 
F-statistic: 334.6 on 5 and 336695 DF,  p-value: < 2.2e-16
```

- **Exhibit 6**

```
> lm_model <- lm(`Dollar Sales No Merch` ~  `Sub-Category Name` , data = conagra_final); summary(lm_model)

Call:
lm(formula = `Dollar Sales No Merch` ~ `Sub-Category Name`, data = conagra_final)

Residuals:
    Min      1Q  Median      3Q     Max 
 -35965  -27978  -23804   -2642 1939871 

Coefficients:
                                                Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                        826.3     4351.4   0.190    0.849    
`Sub-Category Name`MARGARINE/SPREADS             35138.8     4356.5   8.066 7.30e-16 ***
`Sub-Category Name`RFG BUTTER                    27316.1     4354.3   6.273 3.54e-10 ***
`Sub-Category Name`RFG BUTTER BLENDS             29515.5     4362.3   6.766 1.33e-11 ***
`Sub-Category Name`RFG FLAVORED MILK              5988.3    11414.3   0.525    0.600    
`Sub-Category Name`RFG SAUCE/GRAVY/MARINADE MIXES -779.2     8525.9  -0.091    0.927    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 68390 on 336695 degrees of freedom
Multiple R-squared:  0.002793,	Adjusted R-squared:  0.002779 
F-statistic: 188.6 on 5 and 336695 DF,  p-value: < 2.2e-16
```

- **Exhibit 7**

```
> lm_model <- lm(`Dollar Sales Any Merch` ~  `CAG Form Value` * `Sub-Category Name` , data = conagra_final); summary(lm_model)

Call:
lm(formula = `Dollar Sales Any Merch` ~ `CAG Form Value` * `Sub-Category Name`,
    data = conagra_final)

Residuals:
    Min      1Q  Median      3Q     Max
 -22414   -7506   -4679   -1965 4383447

Coefficients: (9 not defined because of singularities)
                                                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                        -4824.59    3590.98  -1.344   0.1791
`CAG Form Value`SPRAY/SQUEEZE                                       -348.06   14697.90  -0.024   0.9811
`CAG Form Value`STICKS                                              4353.94    1855.33   2.347   0.0189 *
`CAG Form Value`TUBS                                                4950.02     827.98   5.978 2.26e-09 ***
`Sub-Category Name`MARGARINE/SPREADS                                5393.99    3738.93   1.443   0.1491
`Sub-Category Name`RFG BUTTER                                       9109.70    3598.94   2.531   0.0114 *
`Sub-Category Name`RFG BUTTER BLENDS                                5197.95    3504.19   1.483   0.1380
`Sub-Category Name`RFG FLAVORED MILK                                2133.98    9322.55   0.229   0.8189
`Sub-Category Name`RFG SAUCE/GRAVY/MARINADE MIXES                    -78.15    6846.44  -0.011   0.9909
`CAG Form Value`SPRAY/SQUEEZE:`Sub-Category Name`MARGARINE/SPREADS 17657.15   14772.29   1.195   0.2320
`CAG Form Value`STICKS:`Sub-Category Name`MARGARINE/SPREADS           29.83    2168.90   0.014   0.9890
`CAG Form Value`TUBS:`Sub-Category Name`MARGARINE/SPREADS           1996.86    1344.19   1.486   0.1374
`CAG Form Value`SPRAY/SQUEEZE:`Sub-Category Name`RFG BUTTER        -3740.79   14932.21  -0.251   0.8022
`CAG Form Value`STICKS:`Sub-Category Name`RFG BUTTER               13775.20    1878.87   7.332 2.28e-13 ***
`CAG Form Value`TUBS:`Sub-Category Name`RFG BUTTER                 -4516.26     915.71  -4.932 8.14e-07 ***
`CAG Form Value`SPRAY/SQUEEZE:`Sub-Category Name`RFG BUTTER BLENDS       NA         NA      NA       NA
`CAG Form Value`STICKS:`Sub-Category Name`RFG BUTTER BLENDS              NA         NA      NA       NA
`CAG Form Value`TUBS:`Sub-Category Name`RFG BUTTER BLENDS                NA         NA      NA       NA
`CAG Form Value`SPRAY/SQUEEZE:`Sub-Category Name`RFG FLAVORED MILK       NA         NA      NA       NA
`CAG Form Value`STICKS:`Sub-Category Name`RFG FLAVORED MILK              NA         NA      NA       NA
`CAG Form Value`TUBS:`Sub-Category Name`RFG FLAVORED MILK                NA         NA      NA       NA
`CAG Form Value`SPRAY/SQUEEZE:`Sub-Category Name`RFG SAUCE/GRAVY/MARINADE MIXES NA  NA      NA       NA
`CAG Form Value`STICKS:`Sub-Category Name`RFG SAUCE/GRAVY/MARINADE MIXES NA         NA      NA       NA
`CAG Form Value`TUBS:`Sub-Category Name`RFG SAUCE/GRAVY/MARINADE MIXES   NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 54920 on 336686 degrees of freedom
Multiple R-squared:  0.01977,    Adjusted R-squared:  0.01973
F-statistic:   485 on 14 and 336686 DF,  p-value: < 2.2e-16
```

- **Exhibit 8**

```
> lm_model <- lm(`Dollar Sales No Merch` ~  `CAG Form Value` * `Sub-Category Name`, data = conagra_final); summary(lm_model)

Call:
lm(formula = `Dollar Sales No Merch` ~ `CAG Form Value` * `Sub-Category Name`,
    data = conagra_final)

Residuals:
    Min      1Q  Median      3Q     Max
 -40435  -33176  -16262    -403 1927579

Coefficients: (9 not defined because of singularities)
                                                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                        -32740.4     4412.9  -7.419 1.18e-13 ***
`CAG Form Value`SPRAY/SQUEEZE                                        -557.3    18062.1  -0.031  0.97538
`CAG Form Value`STICKS                                               9441.8     2280.0   4.141 3.46e-05 ***
`CAG Form Value`TUBS                                                33566.7     1017.5  32.989  < 2e-16 ***
`Sub-Category Name`MARGARINE/SPREADS                                36146.3     4594.7   7.867 3.65e-15 ***
`Sub-Category Name`RFG BUTTER                                       44794.3     4422.7  10.128  < 2e-16 ***
`Sub-Category Name`RFG BUTTER BLENDS                                33386.6     4306.3   7.753 9.00e-15 ***
`Sub-Category Name`RFG FLAVORED MILK                                30113.2    11456.4   2.629  0.00858 **
`Sub-Category Name`RFG SAUCE/GRAVY/MARINADE MIXES                    -779.2     8413.5  -0.093  0.92621
`CAG Form Value`SPRAY/SQUEEZE:`Sub-Category Name`MARGARINE/SPREADS 31559.4    18153.5   1.738  0.08213 .
`CAG Form Value`STICKS:`Sub-Category Name`MARGARINE/SPREADS         20783.1     2665.3   7.798 6.33e-15 ***
`CAG Form Value`TUBS:`Sub-Category Name`MARGARINE/SPREADS             636.5     1651.9   0.385  0.70000
`CAG Form Value`SPRAY/SQUEEZE:`Sub-Category Name`RFG BUTTER         -9004.9    18350.0  -0.491  0.62362
`CAG Form Value`STICKS:`Sub-Category Name`RFG BUTTER                18939.0     2308.9   8.203 2.36e-16 ***
`CAG Form Value`TUBS:`Sub-Category Name`RFG BUTTER                 -28486.3     1125.3 -25.314  < 2e-16 ***
```

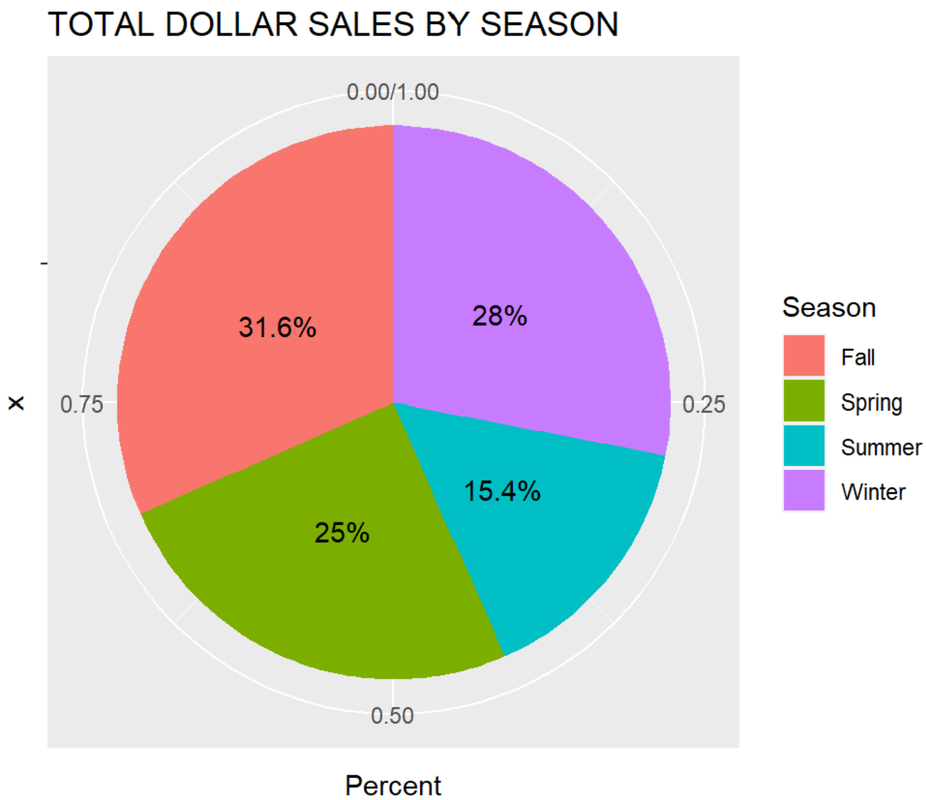
```
`CAG Form Value`SPRAY/SQUEEZE:`Sub-Category Name`RFG BUTTER BLENDS                 NA         NA       NA        NA
`CAG Form Value`STICKS:`Sub-Category Name`RFG BUTTER BLENDS                        NA         NA       NA        NA
`CAG Form Value`TUBS:`Sub-Category Name`RFG BUTTER BLENDS                          NA         NA       NA        NA
`CAG Form Value`SPRAY/SQUEEZE:`Sub-Category Name`RFG FLAVORED MILK                 NA         NA       NA        NA
`CAG Form Value`STICKS:`Sub-Category Name`RFG FLAVORED MILK                        NA         NA       NA        NA
`CAG Form Value`TUBS:`Sub-Category Name`RFG FLAVORED MILK                          NA         NA       NA        NA
`CAG Form Value`SPRAY/SQUEEZE:`Sub-Category Name`RFG SAUCE/GRAVY/MARINADE MIXES     NA         NA       NA        NA
`CAG Form Value`STICKS:`Sub-Category Name`RFG SAUCE/GRAVY/MARINADE MIXES            NA         NA       NA        NA
`CAG Form Value`TUBS:`Sub-Category Name`RFG SAUCE/GRAVY/MARINADE MIXES              NA         NA       NA        NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 67490 on 336686 degrees of freedom
Multiple R-squared:  0.02893,	Adjusted R-squared:  0.02889 
F-statistic: 716.4 on 14 and 336686 DF,  p-value: < 2.2e-16
```

- Exhibit 9

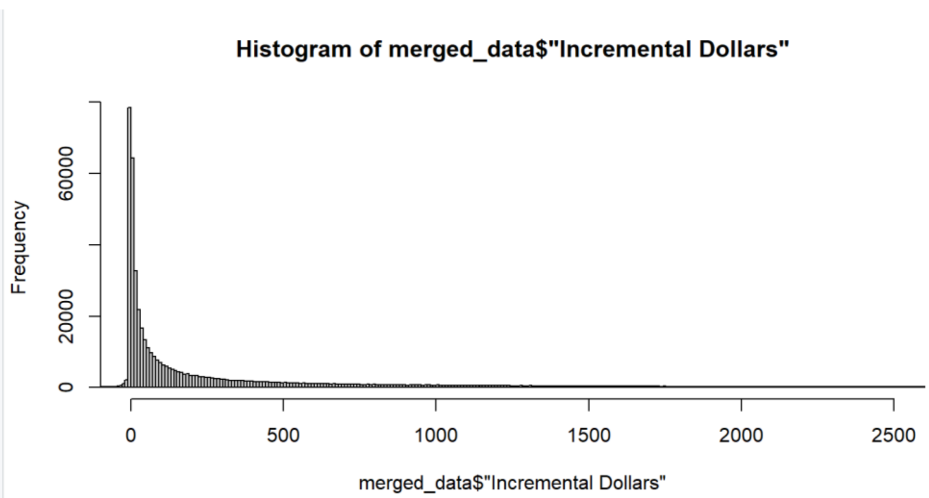
Case2-Analysis7.txt   Case2-Analysis8.txt

This section relates to the Hypothesis-3 (Section 4.4)

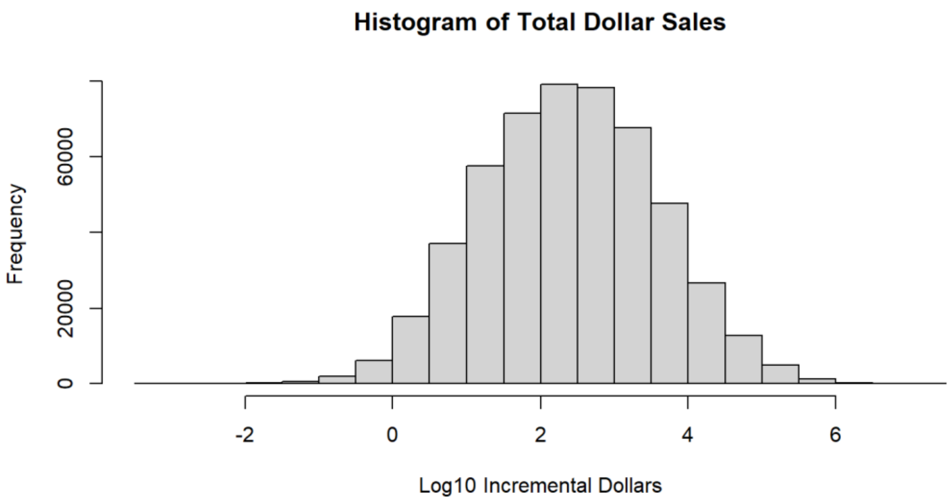
- Exhibit 10



- Exhibit 11



- Exhibit 12



- Exhibit 13

```
> summary(anova_result)
                Df  Sum Sq Mean Sq F value Pr(>F)    
Season           3    8065  2688.4   358.7 <2e-16 ***
Residuals   512223 3839457     7.5                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```