BUAN 6337

Spring 2023

Group 35

Final Project Report

Emily Le dxl170001

Turner Pikulinski Tcp170230

Dhatri Chunchu dxc220013

Jagjeevan Kaur jxk210072

Ishtiaq Omair ixo220002

Rana Udayveer Singh rxx210005

Table of Content

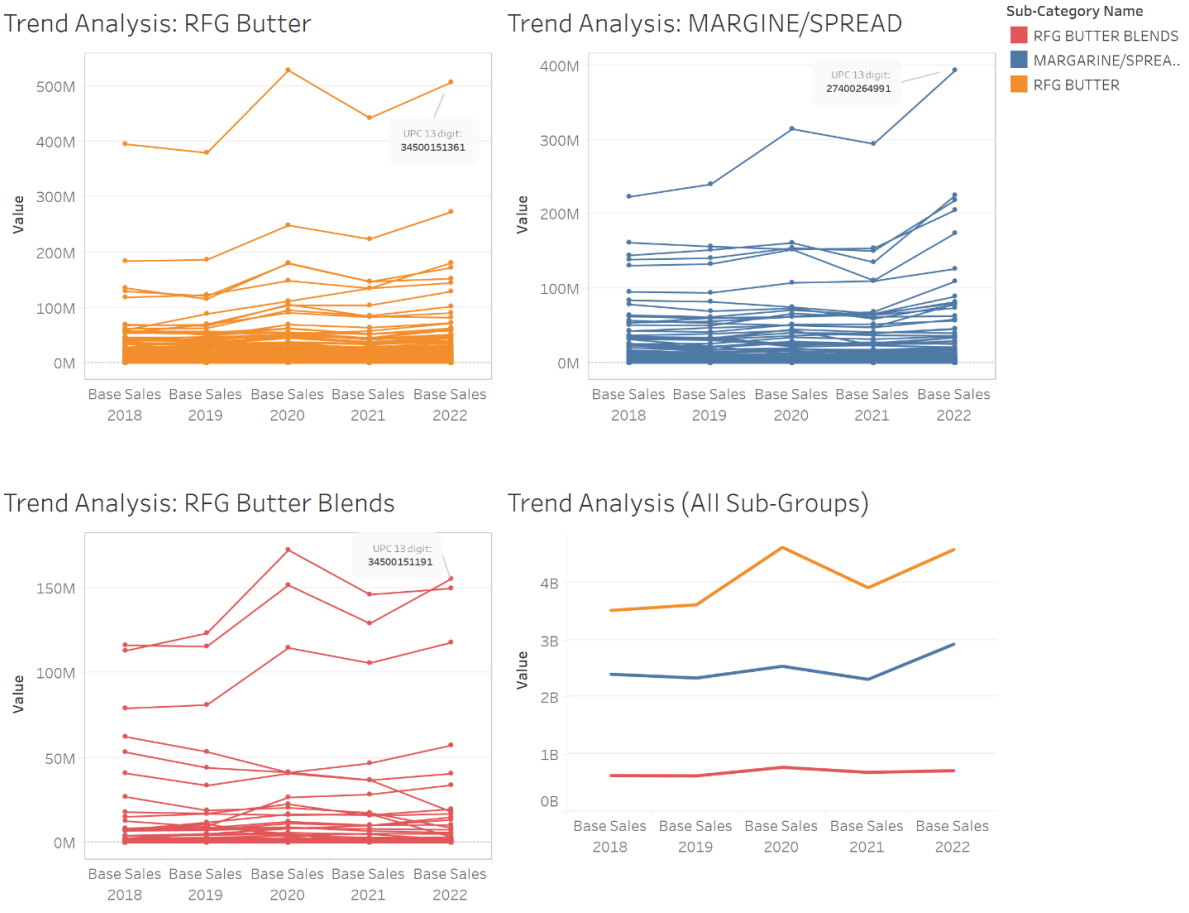
1. Market Research

Brand

Conagra Brands is a large, packaged food company with a diverse portfolio of brands, including some of the most recognized names in the industry with approximately 18,000 employees and operations in North America, South America, and Asia. The company's revenue in 2020 was $11.1 billion. The company has been in operation for over 100 years and has established a strong reputation for quality and innovation.

The company was founded in 1919 and is headquartered in Chicago, Illinois. Conagra Brands has a diverse portfolio of brands, which includes some of the most recognized names in the industry. The company's portfolio includes brands such as Banquet, Chef Boyardee, Healthy Choice, Hunt's, Orville Redenbacher's, and Slim Jim, among others. The company's brands are organized into three main segments: Grocery and Snacks, Refrigerated and Frozen, and International.

Market Size and Growth: Table 1



According to the latest data from Statista, the US butter market size was valued at $7.6 billion in 2020, and it is expected to grow at a compound annual growth rate (CAGR) of 3.8% from 2021 to 2028. On the other hand, the US margarine market size was valued at $1.7 billion in 2020, and

it is expected to grow at a CAGR of 2.6% from 2021 to 2028. Dashboard displays Market growth trends in all Sub-categories of the Table Spread Market

Consumer Trend

There are several consumer trends that are impacting the margarine and butter markets in the US. These include:

- Health concerns: Consumers are increasingly concerned about their health, and they are looking for products that are low in fat, sugar, and salt. As a result, there has been a shift towards margarine in recent years due to its lower levels of saturated fat and cholesterol.
- Natural and organic: There is a growing demand for natural and organic products in the US, and this has led to an increased interest in butter, which is seen as a more natural and wholesome product than margarine.
- Convenience: Consumers are looking for convenient and easy-to-use products, and this has led to an increased demand for spreadable butter and spray margarine.
- Plant-based diets: The demand for plant-based products is growing in the US, and this has led to an increased interest in margarine, which is a plant-based alternative to butter.
- The US margarine and butter markets can be segmented by product type, distribution channel, and application.
- Product Type: The product types in the margarine market include hard, soft, and liquid margarine, while the product types in the butter market include salted, unsalted, whipped, and spreadable butter.
- Distribution Channel: The distribution channels for margarine and butter include supermarkets/hypermarkets, convenience stores, online retailers, and others.
- Application: The applications of margarine and butter include spreads, dressings, sauces, bakery and confectionery, and others.

Price

Conagra's table spread products have prices vary depending on the brand, type of spread, and size of the container. Here is a breakdown of some of the prices for Conagra's table spread products: (via Walmart and Instacart)

- Blue Bonnet Original Spread (15 oz.): $1.79
- Blue Bonnet Light Spread (15 oz.): $1.99
- Parkay Original Spread (12 oz.): $2.98
- Parkay Vegetable Oil Spread (13 oz.): $2.48

It's worth noting that these prices are for regular-sized containers, and prices may vary by region and store. Additionally, prices may fluctuate due to sales or promotions. In comparison to other table spread brands, Conagra's prices appear to be competitive. For example, some popular brands like Country Crock and I Can't Believe It's Not Butter! have similar prices for their products.

- Country CRock Original Spread 15 oz: 3.43
- Country CRock Light Vegetable Oil spread 15 oz: 3.43
- I cant believe its not butter Original Spread 15 oz: 3.98
- I cant believe its not butter Light spread 15 oz: 3.98
- I cant believe its not butter Bang stick 16 oz: 3.98

SWOT Analysis

Strength	**Weakness**
• Brand recognition • Innovation • Accessibility • Wide product Range • Quality	• Dependence on a few key brands • Economic conditions
Opportunity	**Threat**
• Growing Demand of natural and organic products • Product line expansion and Diversification • Societal increased focus on Health & Wellness • International expansion	• Price sensitivity • Competition • Health concerns • Regulatory changes

Major Questions

1) Where are the opportunities for Conagra to expand their brands in terms of product form throughout the major regions of the United States? More specifically, which combination of form factor and tier value has the best unit sales per geographical location out of top 5 Manufacturers Displayed in Table 1?

    - Dependent variable: Total Unit Sales
    - Independent variables: CAG Form, Regional Geography

2) What is Conagra's ideal product? What is the optimal combination of Sub-Category, Price, and Tier Value in terms of most Unit Sales?

    - Dependent variable: CAG Form
    - Independent variables: CAG Sub-Category, Price buckets, and Tier Value

3) What is the ideal price for Conagra to sell product form factors based on major US regions Total Unit Sales?

    - Dependent variable: Price Buckets
    - Independent variables: Form, Sub-Category, and Geography

4) Is there any variation in sales before and after Covid in terms of changes in regional sales, price per unit, and Total Dollar Sales? Where can Conagra target its operations?

    - Dependent variable: Major Brand
    - Independent variables: Geography, Years (Covid: Before and After)

Potential Hypothesis:

Q1 Hypothesis:

Conagra should promote their SPREAD/SQUEEZE specifically in the Northeast for the greatest improvement in sales. The hypothesis suggests that these products have a greater impact on sales compared to other forms of products, and that focusing on the Northeast region and California can lead to positive impacts on sales. Also, maintaining competitive pricing is crucial for Conagra, as it has a significant impact on sales across all predictor variables. By analyzing the pricing strategies of their competitors and offering competitive prices, Conagra can maintain customer loyalty and increase market share.

Q2 Hypothesis:

One potential hypothesis for Conagra to improve their marketing strategy is by focusing on the high-priced STICK form category with a VALUE tier value positioning. This is because consumers are more likely to choose a medium-priced product over a low-priced one, according to the analysis of the multinomial logistic regression. Moreover, consumers in this market segment prefer products with mainstream tier value. Therefore, Conagra can capture a significant market share and drive growth and profitability by offering high-quality margarine and spread products in this category with a mainstream positioning. However, it's important to avoid positioning the products as premium or super premium to ensure they remain affordable for the target market.

Q3 Hypothesis:

Based on the multinomial logistic regression analysis, it can be hypothesized that Conagra should focus on targeting production within the ideal price brackets for medium-sized TUB form and Subcategory Margarine/Spread factors for the Mid-South US regions. This may help Conagra capture a significant market share in this region by providing high-quality, affordable margarine and spread products that are tailored to the specific needs and preferences of consumers in the Mid-South. Moreover, it is hypothesized that highlighting the health benefits of its margarine and spread products, positioning them as a healthier alternative to traditional butter while maintaining an affordable price point, may help Conagra successfully capture a significant market share in the Mid-South US regions and drive long-term growth and profitability. Further research is needed to determine the effectiveness of these marketing solutions.

Q4 Hypothesis

Another potential hypothesis for Conagra could be which product they need to focus on what geographical area/areas based on their performance based on before and after covid period and of course the brand's geographical location. By doing so Conagra can better allocate their resources across the US locations and focus on specific brands and their specific needs. This will not only be efficient from their end but will also give them a competitive edge over other competitors in these categories.

2. Dataset Overview and Data Manipulation Process

Data was partially cleaned in R and concatenated and joined together in Python.

After partially cleaning the data in R, loaded all the years data of table spreads sale and appended all year's data in python for ease of analysis. After appending, we also joined the product attributes file to get all information of form, tier, prices in one table. We also created three new columns for identifying total sales dollars, total sales units and total sales volume by adding base sales columns to incremental sales columns. Below is the screenshot of the code for reference:

We utilized the pandas library in Python to prepare the sales data for tablespreads from five Excel files, covering the period between 2018 and 2022. A new column named "Year" was appended to each dataset to distinguish the sales data by year. We used the concat() method to merge the tablespreads data from 2018-2022 into a single consolidated dataset. A left join was executed on the consolidated sales data and product attribute data using the "UPC 13 digit" common column. We were then able to obtain further insights into the product information such as form value, tier value, and product description. Finally, the cleaned and joined data was stored in a new DataFrame called "Final_Table".

3. Summary Statistics

To understand the data better, we created a correlation matrix to identify if any metrics or features are correlated. Using this we were able update our models to get better accuracy and prediction. In the graph below, we can see there is a consistent correlation between the different Tier values of products and Unit price and Volume price. This is to be expected considering the higher quality products command higher prices from consumers. Additionally, the Categorical variables, Form Value, Sub-Category, and Tier value all share relatively strong correlations. RFG Butter regularly comes in Stick form and ‘All other Form’, while Tubs and ‘Sprays and Squeezes’ are regularly used for Margarine/Spreads and Butter Blends. In terms of Tier, most Super Premium products are in the ‘All other Form’ Category and are RFG Butter. It can also be noted that the Value Tier has a relatively heavy correlation with Margarine/Spreads.

Using these correlations alone, we can identify that new types of Spreads and Blends are usually in forms that are not the basic Stick and are generally in the Value

Table 1: Correlation Matrix Heat Map Displays relationship between Sales, Price, Volume, Form Value, Tier Value

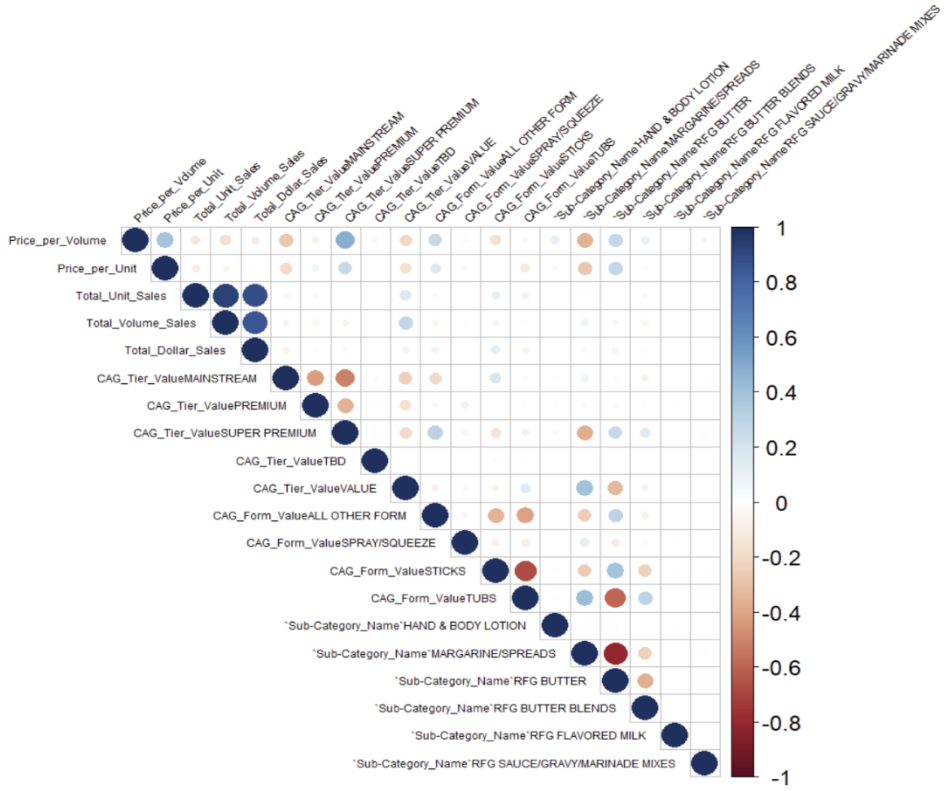


Table 41: Large Outliers in Base Unit Sales, Volume, and Dollar Sales

```
> summary(my_data[, c(5:9)])
 Price per Unit    Price per Volume    Base Unit Sales    Base Volume Sales  Base Dollar Sales
 Min.   :  0.050   Min.   :  0.02133   Min.   :     0.0   Min.   :     0.0   Min.   :      0.0
 1st Qu.:  2.683   1st Qu.:  2.91675   1st Qu.:    68.1   1st Qu.:    47.0   1st Qu.:    255.3
 Median :  3.538   Median :  4.39091   Median :   528.8   Median :   455.2   Median :   1833.4
 Mean   :  4.062   Mean   :  5.70958   Mean   :  5561.6   Mean   :  6309.8   Mean   :  18741.4
 3rd Qu.:  4.922   3rd Qu.:  6.71302   3rd Qu.:  3154.6   3rd Qu.:  3037.6   3rd Qu.:  11420.2
 Max.   :199.990   Max.   :103.92000   Max.   :734069.0   Max.   :734069.0   Max.   :2521574.3
>
```

4. Questions and Analysis

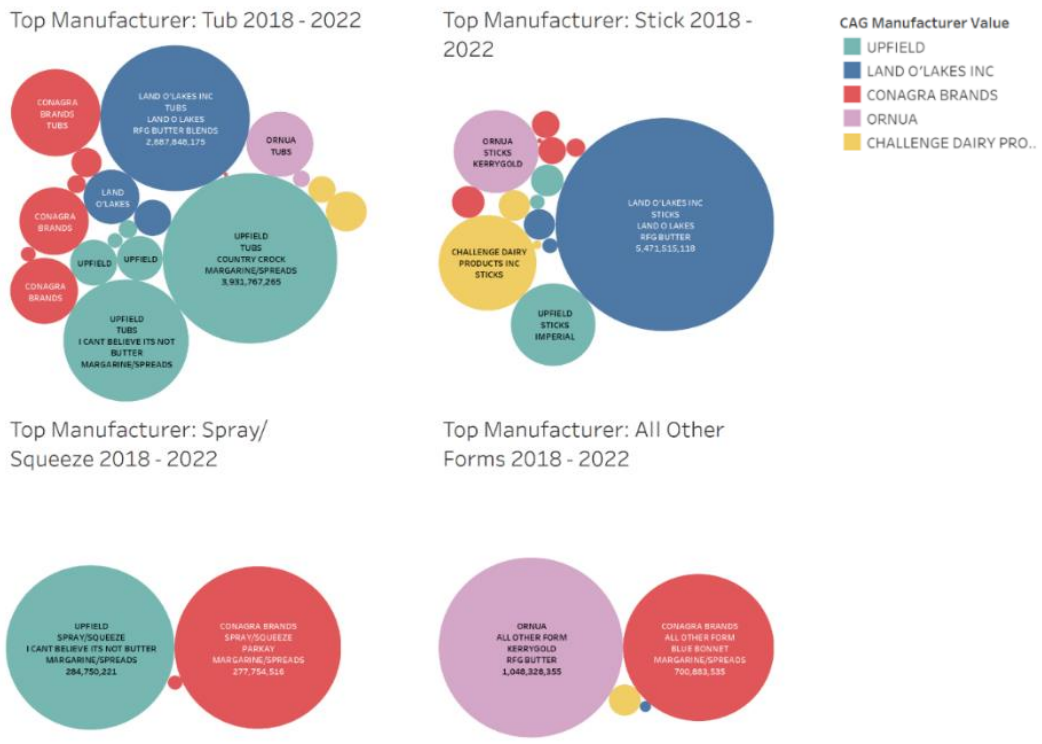
1) Where are opportunities for Conagra to expand their brands in terms of product form throughout the major regions of the United States? More specifically, which combination of form factor and tier value has the best unit sales per geographical location out of top 5 Manufacturers Displayed in Table 1 (Below) excluding Private Label?

Our group will first run a general regression to identify where a majority of ONLY Conagra's sales are located throughout the regions of the US and their top performing forms based on Total unit Sales. We will then run an additional general regression for the TOP 5 Manufacturers (Including Conagra) to identify potential regions and forms that are more increasingly controlled in terms of Total Unit sales by top competitors. The regressions will provide context for Conagra in terms of what form factors to offer in what regions based on major market players offerings.

The coefficients for Conagra will represent the impact of the independent variables on Conagra's sales only, while the coefficients for the other manufacturers will represent the impact of the independent variables on the total sales of all manufacturers included in the analysis. Therefore, you can interpret the coefficients for Conagra as the unique effect of the independent variables on Conagra's sales, while the coefficients for the other manufacturers represent the combined effect of the independent variables on all manufacturers' sales.

Additionally, you could compare the magnitude and direction of the coefficients for Conagra and the other manufacturers to identify any differences in the relationship between the independent variables and Total Unit Sales. This could potentially provide insights into areas where Conagra may need to adjust its strategies to better compete with the top 5 manufacturers.

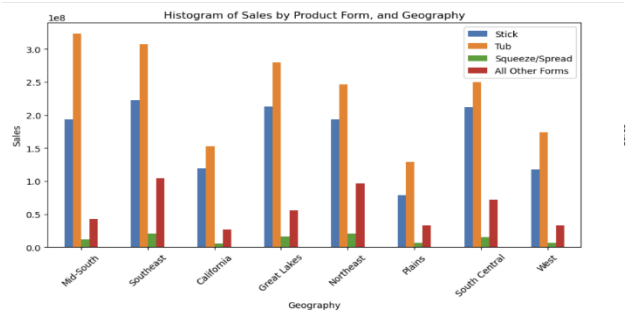
Table 12: Sales split by Form Value and categorized by CAG Manufacturer displaying Top 5 Manufacturers in 2018-2022 (Excluding Private Label)



- Conagra has major market share in *Spray/Squeeze* and *All Other Forms* Markets, however they are 3rd largest in the *Tub* Market

- Conagra is the 4th largest player in the *Stick* Market which is the most competitive among Manufacturers.
- Lack of Market Share in the *Stick* Market is most likely due to Conagra being a Margarine/Spread Manu which is generally not as popular in Stick form.

Figure 3: (TOP 5 BRANDS ONLY) In the below Histograms Stick Sales in the Mid-South are most significant and are predominantly sold with Mainstream brands. Tubs are also very significant sellers in the Mid-south so predominantly with Value brands.



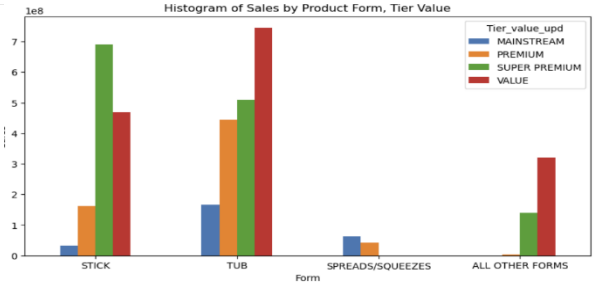
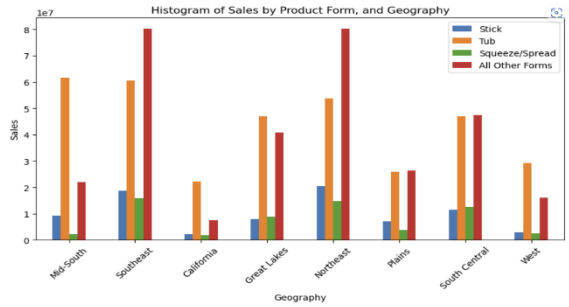
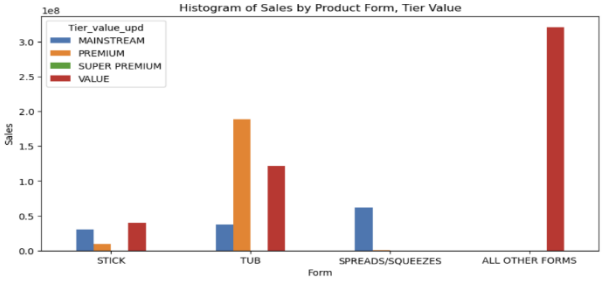


Figure 3: (CONGARA ONLY) In the below Histograms Stick Sales in the Mid-South are most significant and are predominantly sold with Mainstream brands. Tubs are also very significant sellers in the Mid-south so predominantly with Value brands.





- Based on the Comparison Histograms, Conagra has opportunities to release a Super Premium version in its already Popular forms such as TUB, All OTHER FORMS, and STICK. Additionally, the Stick form has opportunities for Conagra to add additional sales in all Regions of the United States.

a) Define the Dummy variables used for regression model and run GLM

In [303]:
```
Dummydata.reset_index()
Dummy_upd = Dummydata.reset_index()



Dummy_upd['Form_upd'] = np.select([Dummy_upd['CAG Form Value'] =='STICKS',
                                   Dummy_upd['CAG Form Value'] == 'TUBS',
                                   Dummy_upd['CAG Form Value'] == 'SPRAY/SQUEEZE',
                                   Dummy_upd['CAG Form Value'] == 'ALL OTHER FORM'],
                                  [ 0, 1, 2, 3 ], default=4)

Dummy_upd['SubCat_upd'] = np.select([Dummy_upd['Sub-Category Name'] == 'RFG BUTTER',
                                     Dummy_upd['Sub-Category Name'] == 'MARGARINE/SPREADS',
                                     Dummy_upd['Sub-Category Name'] == 'RFG BUTTER BLENDS',
                                     Dummy_upd['Sub-Category Name'] == 'RFG FLAVORED MILK'],
                                    [ 0, 1, 2, 3 ], default=4)

Dummy_upd['Geo_upd'] = np.select([Dummy_upd['Geography'] =='Northeast - IRI Standard - Multi Outlet + Conv',
                                  Dummy_upd['Geography'] == 'Southeast - IRI Standard - Multi Outlet + Conv',
                                  Dummy_upd['Geography'] == 'California - IRI Standard - Multi Outlet + Conv',
                                   Dummy_upd['Geography'] == 'Great Lakes - IRI Standard - Multi Outlet + Conv',
                                   Dummy_upd['Geography']=='Mid-South - IRI Standard - Multi Outlet + Conv',
                                   Dummy_upd['Geography']=='Plains - IRI Standard - Multi Outlet + Conv',
                                   Dummy_upd['Geography']=='South Central - IRI Standard - Multi Outlet + Conv',
                                   Dummy_upd['Geography']=='West - IRI Standard - Multi Outlet + Conv'],
                                   [ 0, 1, 2, 3, 4, 5, 6, 7], default=8)

Dummy_upd['Tier_value_upd'] = np.select([Dummy_upd['CAG Tier Value'] =='MAINSTREAM',
                                  Dummy_upd['CAG Tier Value'] == 'PREMIUM',
                                  Dummy_upd['CAG Tier Value'] == 'SUPER PREMIUM',
                                   Dummy_upd['CAG Tier Value'] == 'VALUE'],
                                   [ 0, 1, 2, 3 ], default=4)
```

a. Conagra-only General Linear Regression – Model 1

In [327]:
```
#Subset Data for only Conagra
Dummy_upd_Conagra = Dummy_upd.loc[Dummy_upd['CAG Manufacturer Value'] == 'CONAGRA BRANDS']
```

Generalized Linear Model Regression Results

Dep. Variable:	Total_Unit_Sales	No. Observations:	70825
Model:	GLM	Df Residuals:	70791
Model Family:	Gaussian	Df Model:	33
Link Function:	identity	Scale:	1.8151e+08
Method:	IRLS	Log-Likelihood:	-7.7391e+05
Date:	Mon, 08 May 2023	Deviance:	1.2850e+13
Time:	23:58:03	Pearson chi2:	1.28e+13
No. Iterations:	3	Pseudo R-squ. (CS):	0.9969
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	1.013e+04	373.878	27.096	0.000	9397.751	1.09e+04
Form_upd[T.TUB]	4171.0905	357.736	11.660	0.000	3469.941	4872.240
Form_upd[T.SPRAY/SQUEEZE]	-1421.2161	642.865	-2.211	0.027	-2681.208	-161.225
Form_upd[T.ALL OTHER FORMS]	8.78e+04	899.165	97.648	0.000	8.6e+04	8.96e+04
Geo_upd[T.Southeast]	5438.3874	462.342	11.763	0.000	4532.213	6344.562
Geo_upd[T.California]	-1631.2010	530.710	-3.074	0.002	-2671.374	-591.028
Geo_upd[T.Great lakes]	-574.0290	447.818	-1.282	0.200	-1451.736	303.678
Geo_upd[T.Mid-South]	5092.3182	443.327	11.487	0.000	4223.414	5961.223
Geo_upd[T.Plains]	-964.0266	448.981	-2.147	0.032	-1844.013	-84.040
Geo_upd[T.South Central]	1666.3978	469.249	3.551	0.000	746.687	2586.108
Geo_upd[T.West]	-3021.1277	461.107	-6.552	0.000	-3924.882	-2117.374
Geo_upd[T.Southeast]:Form_upd[T.TUB]	-3158.8411	512.734	-6.161	0.000	-4163.781	-2153.901
Geo_upd[T.California]:Form_upd[T.TUB]	102.7846	588.449	0.175	0.861	-1050.555	1256.124
Geo_upd[T.Great lakes]:Form_upd[T.TUB]	-530.9722	497.241	-1.068	0.286	-1505.546	443.602
Geo_upd[T.Mid-South]:Form_upd[T.TUB]	-5623.9693	492.348	-11.423	0.000	-6588.954	-4658.985
Geo_upd[T.Plains]:Form_upd[T.TUB]	-2640.7954	501.409	-5.267	0.000	-3623.539	-1658.052
Geo_upd[T.South Central]:Form_upd[T.TUB]	-1809.6886	521.680	-3.469	0.001	-2832.163	-787.214
Geo_upd[T.West]:Form_upd[T.TUB]	930.4877	519.483	1.791	0.073	-87.681	1948.656
Geo_upd[T.Southeast]:Form_upd[T.SPRAY/SQUEEZE]	1.47e+04	889.555	16.520	0.000	1.3e+04	1.64e+04
Geo_upd[T.California]:Form_upd[T.SPRAY/SQUEEZE]	2230.0008	1063.347	2.097	0.036	145.879	4314.123
Geo_upd[T.Great lakes]:Form_upd[T.SPRAY/SQUEEZE]	1.165e+04	899.032	12.955	0.000	9884.797	1.34e+04
Geo_upd[T.Mid-South]:Form_upd[T.SPRAY/SQUEEZE]	1.474e+04	886.830	16.615	0.000	1.3e+04	1.65e+04
Geo_upd[T.Plains]:Form_upd[T.SPRAY/SQUEEZE]	3932.1917	919.737	4.275	0.000	2129.540	5734.844
Geo_upd[T.South Central]:Form_upd[T.SPRAY/SQUEEZE]	1.731e+04	925.024	18.715	0.000	1.55e+04	1.91e+04
Geo_upd[T.West]:Form_upd[T.SPRAY/SQUEEZE]	3228.7484	922.741	3.499	0.000	1420.209	5037.288
Geo_upd[T.Southeast]:Form_upd[T.ALL OTHER FORMS]	2.171e+05	1261.483	172.106	0.000	2.15e+05	2.2e+05
Geo_upd[T.California]:Form_upd[T.ALL OTHER FORMS]	-6.423e+04	1297.647	-49.494	0.000	-6.68e+04	-6.17e+04
Geo_upd[T.Great lakes]:Form_upd[T.ALL OTHER FORMS]	7.594e+04	1262.537	60.149	0.000	7.35e+04	7.84e+04
Geo_upd[T.Mid-South]:Form_upd[T.ALL OTHER FORMS]	2.277e+05	1260.896	180.610	0.000	2.25e+05	2.3e+05
Geo_upd[T.Plains]:Form_upd[T.ALL OTHER FORMS]	1.939e+04	1262.985	15.353	0.000	1.69e+04	2.19e+04
Geo_upd[T.South Central]:Form_upd[T.ALL OTHER FORMS]	9.847e+04	1270.165	77.521	0.000	9.6e+04	1.01e+05
Geo_upd[T.West]:Form_upd[T.ALL OTHER FORMS]	-2.873e+04	1267.428	-22.666	0.000	-3.12e+04	-2.62e+04
Price_per_unit	-2181.1568	90.671	-24.056	0.000	-2358.868	-2003.446
np.power(Price_per_unit, 2)	109.6131	8.344	13.136	0.000	93.258	125.968

a. Top 5 Manufacturers General Linear Regression – Model 2 *(Upfield, Land-O-Lakes, Conagra, Ornua, Challenge Dairy) -*

Generalized Linear Model Regression Results

Dep. Variable:	Total_Unit_Sales	No. Observations:	255748
Model:	GLM	Df Residuals:	255714
Model Family:	Gaussian	Df Model:	33
Link Function:	identity	Scale:	1.4034e+09
Method:	IRLS	Log-Likelihood:	-3.0562e+06
Date:	Tue, 09 May 2023	Deviance:	3.5886e+14
Time:	03:46:48	Pearson chi2:	3.59e+14
No. Iterations:	3	Pseudo R-squ. (CS):	0.08676
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	5.079e+04	536.022	94.750	0.000	4.97e+04	5.18e+04
Form_upd[T.TUB]	-1.764e+04	497.023	-35.482	0.000	-1.86e+04	-1.67e+04
Form_upd[T.SPRAY/SQUEEZE]	-2.176e+04	1195.730	-18.199	0.000	-2.41e+04	-1.94e+04
Form_upd[T.ALL OTHER FORMS]	1048.4167	1128.450	0.929	0.353	-1163.304	3260.137
Geo_upd[T.Southeast]	-964.7556	591.559	-1.631	0.103	-2124.189	194.678
Geo_upd[T.California]	-1.138e+04	610.427	-18.639	0.000	-1.26e+04	-1.02e+04
Geo_upd[T.Great lakes]	-2014.0649	586.065	-3.437	0.001	-3162.731	-865.399
Geo_upd[T.Mid-South]	-8018.8483	583.753	-13.737	0.000	-9162.984	-6874.713
Geo_upd[T.Plains]	-2.179e+04	594.511	-36.645	0.000	-2.3e+04	-2.06e+04
Geo_upd[T.South Central]	-7361.5095	587.923	-12.521	0.000	-8513.818	-6209.201
Geo_upd[T.West]	-1.494e+04	583.264	-25.617	0.000	-1.61e+04	-1.38e+04
Geo_upd[T.Southeast]:Form_upd[T.TUB]	1208.3293	690.731	1.749	0.080	-145.479	2562.138
Geo_upd[T.California]:Form_upd[T.TUB]	6335.1013	716.041	8.847	0.000	4931.688	7738.515
Geo_upd[T.Great lakes]:Form_upd[T.TUB]	-161.9832	682.869	-0.237	0.812	-1500.381	1176.414
Geo_upd[T.Mid-South]:Form_upd[T.TUB]	3746.5803	680.898	5.502	0.000	2412.044	5081.116
Geo_upd[T.Plains]:Form_upd[T.TUB]	1.276e+04	694.269	18.385	0.000	1.14e+04	1.41e+04
Geo_upd[T.South Central]:Form_upd[T.TUB]	3485.7233	688.987	5.059	0.000	2135.333	4836.113
Geo_upd[T.West]:Form_upd[T.TUB]	9294.8054	689.679	13.477	0.000	7943.060	1.06e+04
Geo_upd[T.Southeast]:Form_upd[T.SPRAY/SQUEEZE]	7114.3417	1654.233	4.301	0.000	3872.105	1.04e+04
Geo_upd[T.California]:Form_upd[T.SPRAY/SQUEEZE]	6648.9795	1791.764	3.711	0.000	3137.186	1.02e+04
Geo_upd[T.Great lakes]:Form_upd[T.SPRAY/SQUEEZE]	5623.2693	1667.541	3.372	0.001	2354.950	8891.589
Geo_upd[T.Mid-South]:Form_upd[T.SPRAY/SQUEEZE]	1.371e+04	1658.049	8.266	0.000	1.05e+04	1.7e+04
Geo_upd[T.Plains]:Form_upd[T.SPRAY/SQUEEZE]	1.559e+04	1692.382	9.211	0.000	1.23e+04	1.89e+04
Geo_upd[T.South Central]:Form_upd[T.SPRAY/SQUEEZE]	1.035e+04	1688.362	6.128	0.000	7037.692	1.37e+04
Geo_upd[T.West]:Form_upd[T.SPRAY/SQUEEZE]	9017.4995	1685.083	5.351	0.000	5714.797	1.23e+04
Geo_upd[T.Southeast]:Form_upd[T.ALL OTHER FORMS]	4.348e+04	1558.326	27.900	0.000	4.04e+04	4.65e+04
Geo_upd[T.California]:Form_upd[T.ALL OTHER FORMS]	-6999.3358	1504.610	-4.652	0.000	-9948.317	-4050.355
Geo_upd[T.Great lakes]:Form_upd[T.ALL OTHER FORMS]	1886.0136	1497.433	1.259	0.208	-1048.901	4820.928
Geo_upd[T.Mid-South]:Form_upd[T.ALL OTHER FORMS]	4.008e+04	1535.547	26.102	0.000	3.71e+04	4.31e+04
Geo_upd[T.Plains]:Form_upd[T.ALL OTHER FORMS]	1.039e+04	1552.305	6.694	0.000	7348.292	1.34e+04
Geo_upd[T.South Central]:Form_upd[T.ALL OTHER FORMS]	2.326e+04	1553.782	14.972	0.000	2.02e+04	2.63e+04
Geo_upd[T.West]:Form_upd[T.ALL OTHER FORMS]	227.8216	1499.044	0.152	0.879	-2710.251	3165.894
Price_per_unit	-6460.1622	141.874	-45.535	0.000	-6738.229	-6182.095
np.power(Price_per_unit, 2)	385.4035	13.430	28.698	0.000	359.082	411.725

Analysis

The Analysis will consist of the comparison of each Categorical variable between Only Conagra and then the Top 5 brands. The Intercept Variables are Northeast Region and Stick. We will interpret the largest coefficients as variables with the most impact on Total Unit Sales.

In terms of Form, Conagra's 'All other Forms' located in the Mid-South, Southeast, and South Central has the largest Impact on Total Unit Sales compared to the intercept variables Northeast and Stick.

- Geo_upd[T.Mid-South]:Form_upd[T.ALL OTHER FORMS]: Coefficient = 2.277e+05
- Geo_upd[T.Southeast]:Form_upd[T.ALL OTHER FORMS]: Coefficient = 2.171e+05
- Geo_upd[T.South Central]:Form_upd[T.ALL OTHER FORMS]: Coefficient = 9.847e+04

The 'All other Forms' category of Conagra's products plays a significant role in influencing Total Unit Sales, with implications varying across different regions. The coefficient for the 'All other Forms' category (8.78e+04) suggests that, on average, products falling under this category tend to have higher Total Unit Sales compared to the reference category 'Stick', while controlling for other variables in the model.

Additionally, the interactions between the 'All other Forms' category and different regions provide further insights. Specifically, in the Mid-South region, the interaction coefficient (2.277e+05) indicates that the effect of the 'All other Forms' category on Total Unit Sales is even more pronounced. This means that Conagra's 'All other Forms' products have a particularly strong positive impact on Total Unit Sales in the Mid-South region specifically.

On the other hand, when considering the interactions with the Southeast and South Central regions, the coefficients (2.171e+05 and 9.847e+04, respectively) suggest a less substantial effect compared to the Mid-South region. While still positive, the influence of the 'All other Forms' category on Total Unit Sales is relatively smaller in these regions compared to the Mid-South.

In terms of Top 5 Brands, Similarly the form type ALL OTHER FORMS in the Southeast, Mid-South, and South Central had the greatest impact on Total Unit Sales

- Geo_upd[T.Southeast]:Form_upd[T.ALL OTHER FORMS] - Coefficient: 4.348e+04
- Geo_upd[T.Mid-South]:Form_upd[T.ALL OTHER FORMS] - Coefficient: 4.008e+04
- Geo_upd[T.South Central]:Form_upd[T.ALL OTHER FORMS] - Coefficient: 2.326e+04

As for why "ALL OTHER FORMS" has such a large coefficient compared to other forms, when the average price per unit for the "ALL OTHER FORMS" category is lower, a larger coefficient is needed to account for the same impact on total unit sales as the other form categories with higher average prices per unit. The larger coefficient indicates that the "ALL OTHER FORMS" category has a stronger influence on total unit sales despite its lower average price per unit.

When comparing the difference between coefficients in each Model to identify differences in form offering in specific regions, our group Identified the top greatest differences between Conagra’s Total Unit Sales and the Top 5 brand’s Unit Sales.

Here are the variables with the largest differences in coefficients:

- **Form_upd[T.ALL OTHER FORMS]**: The coefficient difference is approximately 1.05e+04.
- **Geo_upd[T.Great lakes]:Form_upd[T.ALL OTHER FORMS]**: The coefficient difference is approximately 1.86e+03.
- **Geo_upd[T.California]:Form_upd[T.ALL OTHER FORMS]**: The coefficient difference is approximately 1.47e+04.
- **Geo_upd[T.Great lakes]**: The coefficient difference is approximately 1.81e+03.

- **Geo_upd[T.Southeast]:Form_upd[T.ALL OTHER FORMS]**: The coefficient difference is approximately 4.31e+04.

These major differences it form offering in specific locations implies that Conagra relies heavily on its 'All Other Form' products compared to the overall Top 5 Manufacturers which are more equally distributed in product offering in the above 5 variables.

2. What is the optimal combination of Sub-Category, Price, and Tier Value in terms of most Unit Sales?

Our group will run a Multinomial Logistic Regression function with independent variables Sub-Category, Price, and Tier Value to find the optimal combination of variables for each Form factor. The reference Category is All_other_Forms, price_low, Tier_value, and Rfg_Butter_Blends.

MNLogit Regression Results

Dep. Variable:	CAG_Form_Value	No. Observations:	100256
Model:	MNLogit	Df Residuals:	100229
Method:	MLE	Df Model:	24
Date:	Fri, 28 Apr 2023	Pseudo R-squ.:	0.2508
Time:	01:04:39	Log-Likelihood:	-82190.
converged:	False	LL-Null:	-1.0970e+05
Covariance Type:	nonrobust	LLR p-value:	0.000

CAG_Form_Value=SPRAY/SQUEEZE	coef	std err	z	P>|z|	[0.025	0.975]
const	-24.7599	1160.539	-0.021	0.983	-2299.375	2249.856
price_medium	0.2047	0.073	2.807	0.005	0.062	0.348
price_high	-2.3384	0.145	-16.095	0.000	-2.623	-2.054
Tier_MAINSTREAM	6.0982	0.710	8.587	0.000	4.706	7.490
Tier_PREMIUM	6.7764	0.711	9.537	0.000	5.384	8.169
Tier_SUPER_PREMIUM	5.1503	0.720	7.156	0.000	3.740	6.561
Sub-Cat_MARGARINE/SPREADS	18.8871	1160.539	0.016	0.987	-2255.728	2293.502
Sub-Cat_RFG_BUTTER	14.3298	1160.539	0.012	0.990	-2260.285	2288.945
Sub-Cat_RFG_BUTTER_BLENDS	15.4940	1160.539	0.013	0.989	-2259.121	2290.109

CAG_Form_Value=STICKS	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.1318	0.109	-10.389	0.000	-1.345	-0.918
price_medium	0.3518	0.027	12.901	0.000	0.298	0.405
price_high	0.2127	0.027	7.989	0.000	0.160	0.265
Tier_MAINSTREAM	1.2070	0.053	22.620	0.000	1.102	1.312
Tier_PREMIUM	0.2901	0.054	5.346	0.000	0.184	0.396
Tier_SUPER_PREMIUM	-0.8981	0.055	-16.342	0.000	-1.006	-0.790
Sub-Cat_MARGARINE/SPREADS	1.9476	0.110	17.640	0.000	1.731	2.164
Sub-Cat_RFG_BUTTER	1.4898	0.111	13.365	0.000	1.271	1.708
Sub-Cat_RFG_BUTTER_BLENDS	-0.8001	0.143	-5.599	0.000	-1.080	-0.520

CAG_Form_Value=TUBS	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.1406	0.105	-10.898	0.000	-1.346	-0.935
price_medium	0.2217	0.029	7.766	0.000	0.166	0.278
price_high	-0.0898	0.028	-3.166	0.002	-0.145	-0.034
Tier_MAINSTREAM	1.2750	0.050	25.718	0.000	1.178	1.372
Tier_PREMIUM	0.8848	0.051	17.476	0.000	0.786	0.984
Tier_SUPER_PREMIUM	0.5111	0.051	9.985	0.000	0.411	0.611
Sub-Cat_MARGARINE/SPREADS	3.1439	0.104	30.243	0.000	2.940	3.348
Sub-Cat_RFG_BUTTER	-0.1367	0.104	-1.315	0.188	-0.340	0.067
Sub-Cat_RFG_BUTTER_BLENDS	2.7165	0.110	24.795	0.000	2.502	2.931

Analysis

This is a multinomial logistic regression with CAG_Form_Value as the dependent variable and 24 independent variables. The model estimates the probability of each level of CAG_Form_Value (SPRAY/SQUEEZE, STICKS, TUBS) for a given combination of values of independent variables. The coefficients indicate the change in the log-odds of being in a particular level of CAG_Form_Value associated with a one-unit increase in the independent variable, holding other independent variables constant. Here is a summary of the coefficients for each independent variable:

const:

The intercept term for each level of CAG_Form_Value. It represents the log-odds of being in the reference level (SPRAY/SQUEEZE) when all other independent variables are zero. The coefficients for STICKS and TUBS are significantly different from zero, indicating that the log-odds of being in these levels are different from the log-odds of being in SPRAY/SQUEEZE. However, the intercept coefficient for each level is not informative on its own, as it depends on the values of other independent variables.

price_medium:

The coefficient indicates the change in the log-odds of being in a particular level of CAG_Form_Value associated with a one-unit increase in the variable price_medium, holding other independent variables constant. For example, for STICKS, a one-unit increase in price_medium is associated with an increase of 0.3518 in the log-odds of being in that level, compared to SPRAY/SQUEEZE. The coefficient is statistically significant for all levels of CAG_Form_Value, except for SPRAY/SQUEEZE.

Tier_MAINSTREAM:

The coefficient indicates the change in the log-odds of being in a particular level of CAG_Form_Value associated with a one-unit increase in the variable Tier_MAINSTREAM, holding other independent variables constant. For example, for STICKS, a one-unit increase in Tier_MAINSTREAM is associated with an increase of 1.2070 in the log-odds of being in that level, compared to SPRAY/SQUEEZE. The coefficient is statistically significant for all levels of CAG_Form_Value.

Based on the provided output, the best combination of price, tier, and sub-category for each CAG_Form_Value can be determined by examining the coefficients with the highest values within each category. Here are the best combinations:

For CAG_Form_Value = SPRAY/SQUEEZE:

Price: None of the price coefficients (price_medium and price_high) have significant positive effects on SPRAY/SQUEEZE.

Tier: The best tier combination is Tier_SUPER_PREMIUM, with a coefficient of 5.1020.

Sub-Cat: None of the sub-category coefficients (Sub-Cat_MARGARINE/SPREADS, Sub-Cat_RFG_BUTTER, and Sub-Cat_RFG_BUTTER_BLENDS) have significant positive effects on SPRAY/SQUEEZE.

For CAG_Form_Value = STICKS:

Price: The best price combination is price_medium with coefficients of 0.3494

Tier: The best tier combination is Tier_MAINSTREAM, with a coefficient of 1.2506.

Sub-Cat: The best sub-category combination is Sub-Cat_MARGARINE/SPREADS, with a coefficient of 1.9217.

For CAG_Form_Value = TUBS:

Price: The best price combination is price_medium, with a coefficient of 0.2214.

Tier: The best tier combination is Tier_MAINSTREAM, with a coefficient of 1.3552.

Sub-Cat: The best sub-category combination is Sub-Cat_MARGARINE/SPREADS, with a coefficient of 3.1131.

3, What are the ideal price brackets for Conagra to sell different product form factors based on different major US regions?

Our group will run a Multinomial Logistic Regression function with independent variables Form, Sub-Category, Geography to find the optimal combination of variables for each Form factor. The reference Category is Stick, price_low, Northeast, and Rfg_Butter_Blends

```
Optimization terminated successfully.
         Current function value: 1.028396
         Iterations 7
                          MNLogit Regression Results
==============================================================================
Dep. Variable:          price_buckets   No. Observations:               701796
Model:                        MNLogit   Df Residuals:                   701768
Method:                           MLE   Df Model:                           26
Date:                Fri, 28 Apr 2023   Pseudo R-squ.:                0.06369
Time:                        15:19:17   Log-Likelihood:           -7.2172e+05
converged:                       True   LL-Null:                  -7.7082e+05
Covariance Type:            nonrobust   LLR p-value:                    0.000
=================================================================================================================
                                  price_buckets=2       coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------------------
const                                                -2.1453      0.065    -32.896      0.000      -2.273      -2.018
Form_TUBS                                            -0.0541      0.008     -6.434      0.000      -0.071      -0.038
Form_ALL OTHER FORM                                  -0.2955      0.009    -32.295      0.000      -0.313      -0.278
Form_SPRAY/SQUEEZE                                    1.5060      0.051     29.585      0.000       1.406       1.606
Sub-Cat_MARGARINE/SPREADS                             3.4237      0.065     52.722      0.000       3.296       3.551
Sub-Cat_RFG_BUTTER                                    1.7917      0.065     27.664      0.000       1.665       1.919
Sub-Cat_RFG_BUTTER_BLENDS                             2.9885      0.066     45.581      0.000       2.860       3.117
Geo_California - IRI Standard - Multi Outlet + Conv  -1.0591      0.014    -73.164      0.000      -1.088      -1.031
Geo_Great Lakes - IRI Standard - Multi Outlet + Conv -0.3227      0.013    -25.227      0.000      -0.348      -0.298
Geo_Mid-South - IRI Standard - Multi Outlet + Conv    0.0837      0.013      6.492      0.000       0.058       0.109
Geo_Northeast - IRI Standard - Multi Outlet + Conv   -0.3139      0.012    -25.600      0.000      -0.338      -0.290
Geo_Plains - IRI Standard - Multi Outlet + Conv      -0.3319      0.014    -24.429      0.000      -0.359      -0.305
Geo_Southeast - IRI Standard - Multi Outlet + Conv   -0.1423      0.013    -10.636      0.000      -0.168      -0.116
Geo_West - IRI Standard - Multi Outlet + Conv        -0.6203      0.014    -45.232      0.000      -0.647      -0.593
-----------------------------------------------------------------------------------------------------------------
                                  price_buckets=3       coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------------------
const                                                -1.0315      0.039    -26.404      0.000      -1.108      -0.955
Form_TUBS                                             0.0498      0.008      6.141      0.000       0.034       0.066
Form_ALL OTHER FORM                                  -0.2872      0.008    -35.757      0.000      -0.303      -0.271
Form_SPRAY/SQUEEZE                                    2.0843      0.051     41.015      0.000       1.985       2.184
Sub-Cat_MARGARINE/SPREADS                             1.3679      0.039     35.357      0.000       1.292       1.444
Sub-Cat_RFG_BUTTER                                    1.1273      0.038     29.495      0.000       1.052       1.202
Sub-Cat_RFG_BUTTER_BLENDS                             1.5437      0.040     38.923      0.000       1.466       1.621
Geo_California - IRI Standard - Multi Outlet + Conv  -0.5796      0.013    -43.320      0.000      -0.606      -0.553
Geo_Great Lakes - IRI Standard - Multi Outlet + Conv -0.1717      0.012    -13.795      0.000      -0.196      -0.147
Geo_Mid-South - IRI Standard - Multi Outlet + Conv   -0.0117      0.013     -0.914      0.361      -0.037       0.013
Geo_Northeast - IRI Standard - Multi Outlet + Conv   -0.2824      0.012    -23.433      0.000      -0.306      -0.259
Geo_Plains - IRI Standard - Multi Outlet + Conv      -0.1501      0.013    -11.430      0.000      -0.176      -0.124
Geo_Southeast - IRI Standard - Multi Outlet + Conv   -0.1112      0.013     -8.391      0.000      -0.137      -0.085
Geo_West - IRI Standard - Multi Outlet + Conv        -0.3154      0.013    -24.183      0.000      -0.341      -0.290
=================================================================================================================
```

Price_bucket 2:

The model is predicting the probability of a product falling into different price buckets based on various independent variables such as form, sub-category, and geographic region.

The "price buckets=2" suggests that this model is predicting the probability of a product falling into range of 34%-66% price bucket. The coefficients represent the estimated effect of each independent variable on the log-odds of a product falling into price bucket 2. The standard errors of the coefficients, and the "P>|z|" column gives the p-values for the coefficients, which indicate whether each variable is statistically significant in predicting the price bucket.

Const: Form_STICKS

- If log-odds of Product in STICKs form in the range of 34%-66% of the Price_bucket 2 increase by 1, the probability of customer buying will drop by 2.1453
- If log-odds of Product in SPRAY/SQUEEZE form in the range of 34%-66% of the Price_bucket 2 increase by 1, the probability of customer buying will drop by 0.2955

Sub-Category:

- If log-odds of Product with Sub-Cat MARGARINE/SPREADS in the range of 34%-66% of the Price_bucket 2 increase by 1, the probability of customer buying will increase by 3.4237

- If log-odds of Product with Sub-Cat RFG BUTTER BLEND in the range of 34%-66% of the Price_bucket 2 increase by 1, the probability of customer buying will increase by 2.9885

Geography:

- If log-odds of Product from California in the range of 34%-66% of the Price_bucket 2 increase by 1, the probability of customer buying will drop by 1.0591
- If log-odds of Product from Mid-South in the range of 34%-66% of the Price_bucket 2 increase by 1, the probability of customer buying will increase by 0.0837

Price_bucket 3:

Const: Form_STICKS:

- If log-odds of Product in STICKs form in the range of 67%-100% of the Price_bucket 3 increases by 1, the probability of customer buying decreases by 1.0315
- If log-odds of Product in SPRAY/SQUEEZE form in the range of 67%-100% of the Price_bucket 3 increase by 1, the probability of customer buying will increase by 2.0843

Sub-Category:

- If log-odds of Product with Sub-Cat RFG BUTTER BLEND in the range of 67%-100% of the Price_bucket 3 increase by 1, the probability of customer buying will increase by 1.5437
- If log-odds of Product with Sub-Cat MARGARINE/SPREADS in the range of 67%-100% of the Price_bucket 3 increase by 1, the probability of customer buying will increase by 1.3679

Geography:

- If log-odds of Product from California in the range of 67%-100% of the Price_bucket 3 increase by 1, the probability of customer buying will drop by 0.5796
- If log-odds of Product from West in the range of 67%-100% of the Price_bucket 3 increase by 1, the probability of customer buying will drop by 0.3154

4, Is there any variation in sales before and after covid in terms of changes in regional sales, price per unit, and Total Dollar Sales?

Defined covid_segree as pre covid = 0 (years 2018 and 2019) & post covid = 1 (years 2020, 2021, 2022)

Major_Brands_upd=EARTH BALANCE	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.2495	0.032	-7.721	0.000	-0.313	-0.186
covid_segree	0.2573	0.020	12.613	0.000	0.217	0.297
Geo_California - IRI Standard - Multi Outlet + Conv	0.6593	0.042	15.518	0.000	0.576	0.743
Geo_Great Lakes - IRI Standard - Multi Outlet + Conv	0.4509	0.040	11.220	0.000	0.372	0.530
Geo_Mid-South - IRI Standard - Multi Outlet + Conv	0.4294	0.040	10.718	0.000	0.351	0.508
Geo_Northeast - IRI Standard - Multi Outlet + Conv	0.5389	0.040	13.510	0.000	0.461	0.617
Geo_Plains - IRI Standard - Multi Outlet + Conv	0.3175	0.042	7.643	0.000	0.236	0.399
Geo_Southeast - IRI Standard - Multi Outlet + Conv	0.2638	0.041	6.491	0.000	0.184	0.343
Geo_West - IRI Standard - Multi Outlet + Conv	0.4142	0.042	9.969	0.000	0.333	0.496

Major_Brands_upd=SMART BALANCE	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3461	0.030	11.669	0.000	0.288	0.404
covid_segree	-0.2025	0.019	-10.502	0.000	-0.240	-0.165
Geo_California - IRI Standard - Multi Outlet + Conv	0.1174	0.042	2.799	0.005	0.035	0.200
Geo_Great Lakes - IRI Standard - Multi Outlet + Conv	0.3886	0.038	10.318	0.000	0.315	0.462
Geo_Mid-South - IRI Standard - Multi Outlet + Conv	0.3335	0.038	8.860	0.000	0.260	0.407
Geo_Northeast - IRI Standard - Multi Outlet + Conv	0.4335	0.038	11.548	0.000	0.360	0.507
Geo_Plains - IRI Standard - Multi Outlet + Conv	0.1592	0.039	4.055	0.000	0.082	0.236
Geo_Southeast - IRI Standard - Multi Outlet + Conv	0.3279	0.038	8.727	0.000	0.254	0.402
Geo_West - IRI Standard - Multi Outlet + Conv	0.1787	0.040	4.512	0.000	0.101	0.256

Major_Brands_upd=PARKAY	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.3260	0.035	-9.335	0.000	-0.394	-0.258
covid_segree	-0.0644	0.024	-2.718	0.007	-0.111	-0.018
Geo_California - IRI Standard - Multi Outlet + Conv	-0.4420	0.054	-8.164	0.000	-0.548	-0.336
Geo_Great Lakes - IRI Standard - Multi Outlet + Conv	-0.0471	0.046	-1.026	0.305	-0.137	0.043
Geo_Mid-South - IRI Standard - Multi Outlet + Conv	-0.0624	0.046	-1.363	0.173	-0.152	0.027
Geo_Northeast - IRI Standard - Multi Outlet + Conv	0.1942	0.044	4.376	0.000	0.107	0.281
Geo_Plains - IRI Standard - Multi Outlet + Conv	0.0705	0.046	1.529	0.126	-0.020	0.161
Geo_Southeast - IRI Standard - Multi Outlet + Conv	0.1047	0.045	2.349	0.019	0.017	0.192
Geo_West - IRI Standard - Multi Outlet + Conv	0.0509	0.047	1.089	0.276	-0.041	0.142

Major_Brands_upd=PRIVATE LABEL	coef	std err	z	P>|z|	[0.025	0.975]
const	2.5122	0.023	107.832	0.000	2.467	2.558
covid_segree	0.0015	0.016	0.098	0.922	-0.029	0.032
Geo_California - IRI Standard - Multi Outlet + Conv	-0.0741	0.033	-2.240	0.025	-0.139	-0.009
Geo_Great Lakes - IRI Standard - Multi Outlet + Conv	0.3670	0.030	12.163	0.000	0.308	0.426
Geo_Mid-South - IRI Standard - Multi Outlet + Conv	0.7087	0.030	23.705	0.000	0.650	0.767
Geo_Northeast - IRI Standard - Multi Outlet + Conv	0.7057	0.030	23.468	0.000	0.647	0.765
Geo_Plains - IRI Standard - Multi Outlet + Conv	0.1601	0.031	5.178	0.000	0.100	0.221
Geo_Southeast - IRI Standard - Multi Outlet + Conv	0.2604	0.030	8.683	0.000	0.202	0.319
Geo_West - IRI Standard - Multi Outlet + Conv	0.1087	0.031	3.472	0.001	0.047	0.170

Major_Brands_upd=LAND O LAKES	coef	std err	z	P>|z|	[0.025	0.975]
const	1.3767	0.025	54.908	0.000	1.328	1.426
covid_segree	-0.0002	0.017	-0.013	0.990	-0.033	0.033
Geo_California - IRI Standard - Multi Outlet + Conv	0.2728	0.035	7.782	0.000	0.204	0.342
Geo_Great Lakes - IRI Standard - Multi Outlet + Conv	0.1205	0.032	3.710	0.000	0.057	0.184
Geo_Mid-South - IRI Standard - Multi Outlet + Conv	0.1076	0.032	3.330	0.001	0.044	0.171
Geo_Northeast - IRI Standard - Multi Outlet + Conv	0.1545	0.032	4.762	0.000	0.091	0.218
Geo_Plains - IRI Standard - Multi Outlet + Conv	0.0777	0.033	2.338	0.019	0.013	0.143
Geo_Southeast - IRI Standard - Multi Outlet + Conv	0.0755	0.032	2.338	0.019	0.012	0.139
Geo_West - IRI Standard - Multi Outlet + Conv	0.1283	0.034	3.824	0.000	0.063	0.194

Major_Brands_upd=OTHER	coef	std err	z	P>|z|	[0.025	0.975]
const	3.0673	0.023	133.773	0.000	3.022	3.112
covid_segree	0.0475	0.016	3.061	0.002	0.017	0.078
Geo_California - IRI Standard - Multi Outlet + Conv	0.4343	0.032	13.427	0.000	0.371	0.498
Geo_Great Lakes - IRI Standard - Multi Outlet + Conv	0.4728	0.030	15.910	0.000	0.415	0.531
Geo_Mid-South - IRI Standard - Multi Outlet + Conv	0.1601	0.030	5.401	0.000	0.102	0.218
Geo_Northeast - IRI Standard - Multi Outlet + Conv	0.6395	0.030	21.529	0.000	0.581	0.698
Geo_Plains - IRI Standard - Multi Outlet + Conv	0.2794	0.030	9.187	0.000	0.220	0.339
Geo_Southeast - IRI Standard - Multi Outlet + Conv	0.0752	0.030	2.541	0.011	0.017	0.133
Geo_West - IRI Standard - Multi Outlet + Conv	0.3220	0.031	10.472	0.000	0.262	0.382

Analysis

The coefficients indicate the magnitude and direction of the effect of each variable on the log-odds of belonging to the corresponding major brand. The "covid_segree" variable represents the impact of the COVID-19 severity, and its coefficient shows the effect of the severity on the log-odds of brand affiliation. Positive coefficients indicate an increase in the log-odds, while negative coefficients indicate a decrease.

Among the major brands analyzed, the best performing brand in terms of the largest coefficient is Private Label. It has the highest coefficients among all the brands, indicating a stronger association with the selected variables.

Private Label shows a positive coefficient of 0.7057 for the combination of Geo_Northeast - IRI Standard - Multi Outlet + Conv and a coefficient of 0.7087 for the combination of Geo_Mid-South - IRI Standard - Multi Outlet + Conv. This suggests that these combinations of variables have the strongest positive impact on the log-odds of Private Label affiliation.

The impact of the "covid_segree" variable on brand affiliation varies across the brands. For Earth Balance, it has a positive impact with a coefficient of 0.2573. In contrast, Smart Balance and Parkay show negative impacts of -0.2025 and -0.0644, respectively. Land O Lakes has a minimal impact with a coefficient of -0.0002. Other brands show a slightly positive impact, with coefficients ranging from 0.0015 to 0.0475.

Considering the coefficients of their corresponding variables, it appears that Private Label has the strongest association with the selected variables, making it the best performing brand among those analyzed. However, it's important to consider additional factors such as market share, sales volume, and consumer preferences to obtain a comprehensive evaluation of brand performance.

4. Potential Solutions based on analysis:

Question 1: which combination of form factor and tier value has the best unit sales per geographical location out of top 5 Manufacturers (Upfield, Land-O-Lakes, Conagra, Ornua, Challenge Dairy)

Based on the results of the Generalized Linear Regression analysis of Model 1, it can be observed that Conagra has already established a strong presence in the 'ALL OTHER FORMS' and 'SPRAY/SQUEEZE' form types. This is indicated by the large coefficients associated with these form types, as mentioned earlier. Additionally, the Histogram comparison further supports the notion that Conagra has significant market share in these form types.

Considering the above information, it can be inferred that Conagra has an opportunity to capitalize on STICK and TUB sales across all regions of the United States. The coefficients associated with the different geographies in Model 1 suggest that there is potential for stick sales in various regions, despite some regions having weaker coefficients compared to others. The positive coefficient for the Southeast region indicates a strong positive effect on stick sales, suggesting that Conagra's stick products are performing well in that region.

While some regions, such as California, the Great lakes, Plains, and West, have weaker coefficients for stick sales, it does not necessarily mean that there is no market for sticks in these areas. Rather, the coefficients indicate a relatively weaker effect compared to other regions. Therefore, Conagra could focus on targeted marketing and sales strategies to further enhance stick sales in these regions.

In summary, considering the significant market presence in 'ALL OTHER FORMS' and 'SPRAY/SQUEEZE' form types, as well as the opportunity to tap into stick sales across all regions of the United States, Conagra has the potential to leverage its strengths and expand its market share in stick products.

Question 2: What is the optimal combination of Sub-Category, Price, and Tier Value in terms of most Unit Sales?

Based on the multinomial logistic regression analysis, the most optimal product offering for Conagra to focus on the medium-priced TUB form category with a MAINSTREAM tier value positioning. This is because the coefficient for price_medium is statistically significant for all levels of CAG_Form_Value, except for SPRAY/SQUEEZE form, indicating that consumers are more likely to choose a medium-priced product over a low-priced one. Additionally,

consumers in this market segment are more likely to choose a product with a mainstream tier value. Therefore, by offering high-quality margarine and spread products in the medium-priced TUB form category with a mainstream positioning, Conagra can capture a significant market share and drive growth and profitability in the long run. It's also important to note that the coefficient for price bucket of high is statistically significant for all levels of CAG_Form_Value, so Conagra should avoid positioning its products as premium or super premium to ensure they remain affordable for the target market.

Question 3: What are the ideal price brackets for Conagra to sell different product form factors based on different major US regions?

Based on the multinomial logistic regression analysis, the ideal price brackets for Conagra product to be in medium bracket if they want to target production in TUB form and Subcategory Margarine/Spread product. Per our regression model, price bucket 2 is product falling into range of 34%-66% price bucket. This means that the Conagra brand should aim to price their product within this range to optimize sales and market share. Especially for the Mid-South US regions, it will help Conagra capture a significant market share in this region by providing high-quality, affordable table spread products that are specifically suitable and fit preferences of consumers in the Mid-South base on the analysis conducted.

Additionally, Conagra should focus on highlighting the health benefits of its margarine and spread products, positioning them as a healthier alternative to traditional butter while maintaining an affordable price point. By implementing marketing solutions, Conagra can successfully capture a significant market share in the Mid-South US regions and drive long-term growth and profitability.

Question 4: Is there any variation in sales before and after covid in terms of changes in regional sales, price per unit, and Total Dollar Sales?

Based on the multinomial logistic regression analysis, Earth Balance has a positive impact from the "covid_segree" variable, suggesting that it may have benefitted from the COVID-19 situation. Conagra could leverage this positive association by further promoting and emphasizing Earth Balance highlighting its health benefits, adaptability to changing consumer preferences, and sustainable production practices.

As for both Smart Balance and Parkay since "covid_segree" variable is showing a negative impact Conagra should investigate the reasons behind this negative association and assess whether there are any specific challenges or consumer perceptions affecting these brands. They could consider conducting market research to understand consumer preferences and identify areas for improvement or repositioning.

5. Conclusion

Based on the market and regression analysis our group performed, we believe that Question 1 and Question 2 solutions will be the most valuable to Conagra. The questions compare product offerings versus Top 5 competitors as well as provide specific categorical combinations for multiple price ranges. We look to provide Conagra with the ideal combination of Product Features for Price, Form factor, Tier value, and Subcategory value. Identifying these optimal combinations will give insights into where Conagra can improve their product offerings and capture increased Market Share in the process. Combining these insights with Conagra's shortcomings in terms of current product offerings, we can be sure to provide data driven recommendations that are tailored specifically to Conagra.

Some of the most notable opportunities for Conagra to increase Unit Sales are Market Share arbitrage in the Tub and Stick Form factors, as well as offering Mainstream and Super Premium products which Conagra currently has a limited selection of. Among the 3 of our regressions that include form factor categories, TUB has been a consistent top recommendation in our outputs. Including the initial market analysis, we can see that Conagra is sitting at the 3rd largest TUB Manufacturer behind Upfield's Country Crock and I Can't Believe it's not Butter as well as Land O' Lakes.

Based on the optimal combination of Product features for TUB, our question 2 analysis recommends a Medium priced, Mainstream, Margarine/Spread offering as the most competitive option for gaining market share and increasing sales. The Medium price bracket was defined in our group analysis as falling in range of 34%-66% price bucket 2 (based on Price Per Unit variable of all Conagra brands). This ideal price indicates possible savings for the consumer while still maintaining the quality that Conagra is known for. As for The Mainstream Tier Value products, these products are priced affordably and offer good quality, making them attractive to cost-conscious consumers who do not want to compromise on quality. Moreover, Expanding the customer base is a key factor in improving profitability for any business. By introducing a Mainstream Tier Value product, Conagra can attract a new segment of customers who may not have previously considered its products due to pricing concerns. It can help have a significant impact on profitability as it can increase overall unit sales. By combining the benefits of both Tub form, Medium priced and Mainstream Tier Value, Conagra can create a unique product that offers both convenience and affordability to consumers.

Furthermore, on initial Market Share discovery, our group was able to identify Conagra's Smart Balance, Blue Bonnet, and Earth Balance have already been able to fulfill a majority of the TUB competitive combination: Medium priced, Mainstream, Margarine/Spread. However, Conagra's Tub Brands are generally labeled as Premium products. Conagra has a great opportunity to provide a Mainstream version of its already popular premium products to diversify its current product offerings. As initially stated, the Mainstream recommendation is backed by regression analysis as the ideal consumer accepted product.

Focusing on Conagra's opportunity to increase TUB Sales, Question 1 regression displays that Conagra has the greatest opportunity to increase Tub Sales in the West and California Regions. With this understanding, Conagra would benefit from increased

merchandising of its Premium and Mainstream products in California and West locations. Conagra already has relatively great Tub sales in the Midsouth, Southeast, and Northeast.

When innovating the existing Tub form products of other competitors, Conagra should introduce new flavors and ingredients to cater to the evolving preferences of consumers. Furthermore, the brand-new packaging should also be convenient and stored. Conagra will benefit from Tub form butter products due to its form being easy to use, also as they are pre-portioned and can be easily spread on bread or other foods. Conagra could introduce a tub form butter product with a garlic or herb flavor, or a product with added nutritional benefits such as omega-3 fatty acids or vitamins. They could also offer a larger tub form butter product for families or individuals who use a lot of butter or offer individual portion-sized tubs for convenience.

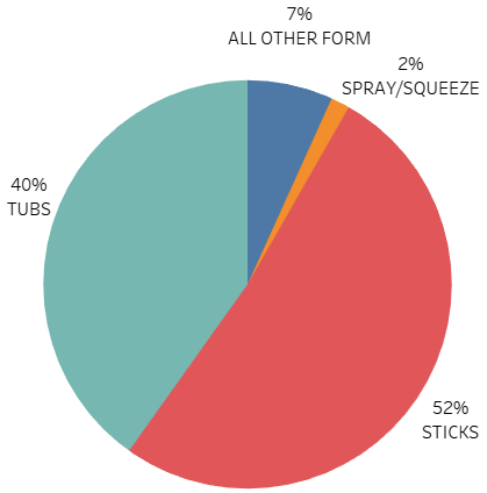
Lastly, Conagra can optimize these features into a brand-new product, as in freshly packaging Tub form along with Mainstream Tier value for quality for a price range in medium bracket.

In conclusion, optimizing the features of Tub form and Mainstream Tier Value into a new product can help Conagra capitalize on market trends and increase its sales and profitability. By offering a unique product that combines convenience, affordability, and health benefits, Conagra can attract a new segment of customers and improve its overall brand perception in the market.

6. Appendix

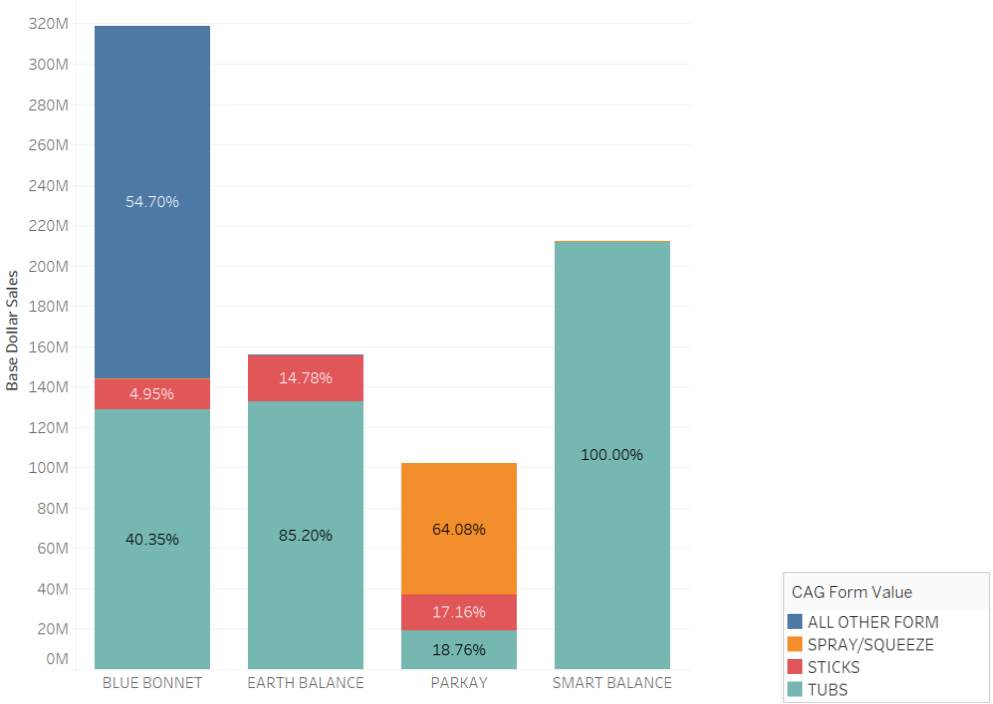
Data Visualization:

- Tablespread share by form.



From the 2022 Tablespreads dataset, Sticks are found to be the largest form in the category valued at 52%.

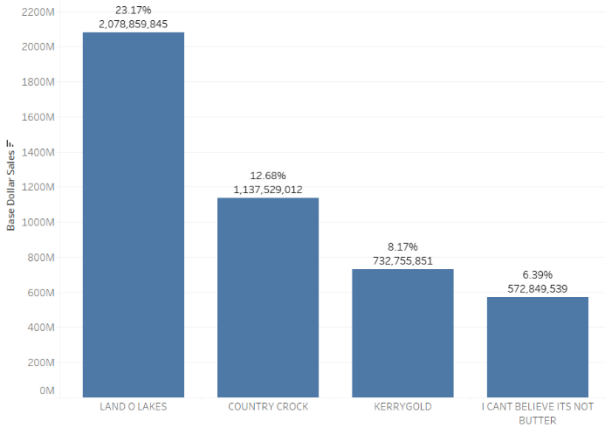
- Tablespread share by form in Major brands.



- In the brand Blue Bonnet, Tubs have major share (40%)
- In the brand Earth Balance, there are only tubs and sticks and tubs have largest share (85%)
- In Parky brand, Sticks and Tubs have almost same share whereas Spray/Squeeze have largest share (64%)

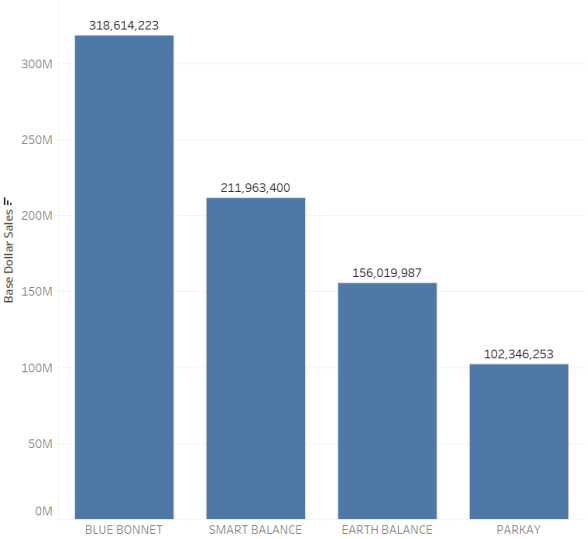
- Smart Balance brand has only Tubs
- Sales split by brand.

Showing on top 4 brands:

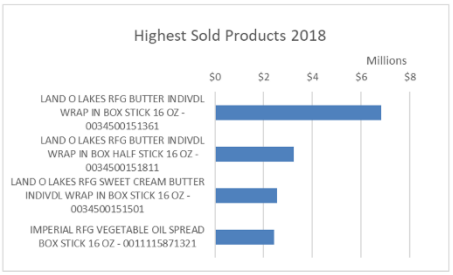


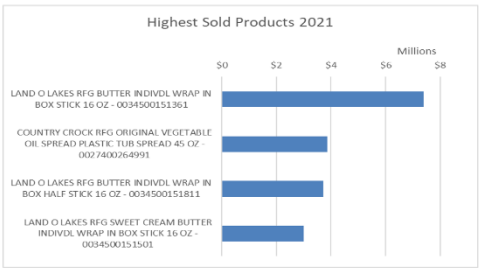
The Base Dollar Sales for Land O Lakes are the highest valued at $2.07B followed by Country Crock valued at $1.13B.

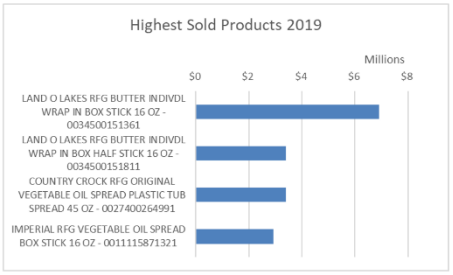
b. Showing major 4 brands

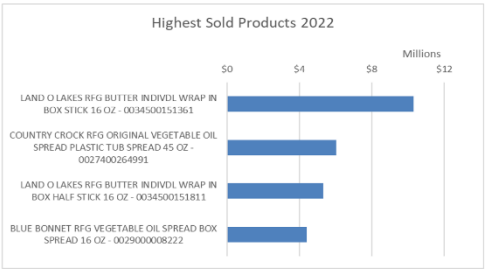


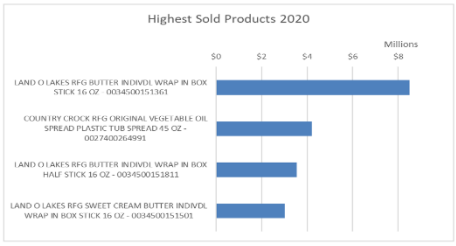
- Sales split by Form Value and categorized by CAG Manufacturer displaying Top 5 Manufacturers in 2018-2022 (Excluding Private Label)


Highest Sold Products 2018
Millions
$0
$2
$4
$6
$8
LAND O LAKES RFG BUTTER INDIVDL WRAP IN BOX STICK 16 OZ - 0034500151361
LAND O LAKES RFG BUTTER INDIVDL WRAP IN BOX HALF STICK 16 OZ - 0034500151811
LAND O LAKES RFG SWEET CREAM BUTTER INDIVDL WRAP IN BOX STICK 16 OZ - 0034500151501
IMPERIAL RFG VEGETABLE OIL SPREAD BOX STICK 16 OZ - 0011115871321


Highest Sold Products 2021
Millions
$0
$2
$4
$6
$8
LAND O LAKES RFG BUTTER INDIVDL WRAP IN BOX STICK 16 OZ - 0034500151361
COUNTRY CROCK RFG ORIGINAL VEGETABLE OIL SPREAD PLASTIC TUB SPREAD 45 OZ - 0027400264991
LAND O LAKES RFG BUTTER INDIVDL WRAP IN BOX HALF STICK 16 OZ - 0034500151811
LAND O LAKES RFG SWEET CREAM BUTTER INDIVDL WRAP IN BOX STICK 16 OZ - 0034500151501


Highest Sold Products 2019
Millions
$0
$2
$4
$6
$8
LAND O LAKES RFG BUTTER INDIVDL WRAP IN BOX STICK 16 OZ - 0034500151361
LAND O LAKES RFG BUTTER INDIVDL WRAP IN BOX HALF STICK 16 OZ - 0034500151811
COUNTRY CROCK RFG ORIGINAL VEGETABLE OIL SPREAD PLASTIC TUB SPREAD 45 OZ - 0027400264991
IMPERIAL RFG VEGETABLE OIL SPREAD BOX STICK 16 OZ - 0011115871321


Highest Sold Products 2022
Millions
$0
$4
$8
$12
LAND O LAKES RFG BUTTER INDIVDL WRAP IN BOX STICK 16 OZ - 0034500151361
COUNTRY CROCK RFG ORIGINAL VEGETABLE OIL SPREAD PLASTIC TUB SPREAD 45 OZ - 0027400264991
LAND O LAKES RFG BUTTER INDIVDL WRAP IN BOX HALF STICK 16 OZ - 0034500151811
BLUE BONNET RFG VEGETABLE OIL SPREAD BOX SPREAD 16 OZ - 0029000008222


Highest Sold Products 2020
Millions
$0
$2
$4
$6
$8
LAND O LAKES RFG BUTTER INDIVDL WRAP IN BOX STICK 16 OZ - 0034500151361
COUNTRY CROCK RFG ORIGINAL VEGETABLE OIL SPREAD PLASTIC TUB SPREAD 45 OZ - 0027400264991
LAND O LAKES RFG BUTTER INDIVDL WRAP IN BOX HALF STICK 16 OZ - 0034500151811
LAND O LAKES RFG SWEET CREAM BUTTER INDIVDL WRAP IN BOX STICK 16 OZ - 0034500151501

7. References

https://www.grandviewresearch.com/press-release/global-margarine-market

https://www.statista.com/outlook/cmo/food/oils-fats/butter/united-states#revenue

https://www.grandviewresearch.com/industry-analysis/butter-market-report