BUAN 6337.501

Group Number: 36

Predictive Analytics

Project Title: Create data-driven strategies to help Conagra unlock future growth potential in the Table Spreads category.

SUBMITTED BY:

Bala Krishna Bobbili	bxb220009
Sneha Thunga	sxt220053
Sri Dharshan Vasan Senthil Kumaran	sxv220012
Arshiya Anjum Yerur	axy220012

INTRODUCTION:

The primary goal of this project is to provide Conagra with optimal and reliable options to accelerate sales and growth in the market. To achieve this objective, we conducted a comprehensive analysis of sales across different geographical locations, forms of table spreads, brands, target audiences based on race, income, age, and time periods. This analysis helped us gain a deep understanding of the market's performance and customer demographics. We generated a complete market structure to identify opportunities for Conagra to leverage and implement strategies to upscale production, marketing, and sales.

We are providing various market strategies to aid Conagra in enhancing their performance. However, to ensure the effectiveness of these strategies, we plan to perform hypothesis testing for each of them. This involves generating regression analysis of the data and regression equations to predict future outcomes. We also use the trends to provide prescriptive actions that Conagra can take to perform better.

The ultimate objective of this project is to assist Conagra in achieving its goals and outperforming its competitors in the market. By using the insights gained from our analysis and testing, we aim to provide Conagra with actionable recommendations that are tailored to their specific needs and challenges. Our approach ensures that Conagra has access to reliable information and strategies to improve its market position, increase sales, and accelerate growth.

DATA DESCRIPTION:

The primary focus is to analyze the Table spreads category and identify sales growth opportunities across Conagra's product portfolio. Adjacent categories such as Cooking Spray and Cooking & Salad Oils are also considered as they offer substitutable goods. Conagra provided three different data sets for these categories,

- **IRi's Point of Sales:** Store and item level scan data aggregated to both IRI Standard Regions and MULO+C.
- **IRi's Panel:** Data collected by a network of panelists who are representative of the total US Population
- **NPD's National Eating Trends:** Consumption habits including the who, what, where, when, why and how for all foods and beverages

PROPOSAL FOR THE PROJECT:

By performing an in-depth analysis using common modeling algorithms, we aim to determine a set of go-to-market strategies to accelerate sales growth.

Our study will primarily focus on the category of spreads, with special attention given to the less popular product, **margarine**, which has the potential for growth and acceptance by the end customer. We plan to apply various techniques to this study, including Linear and Multiple Regressions. Our aim is to choose the best model based on the Model Selection Criteria and implement the Best Subset Selection Criteria, which involves the calculation of adjusted-R2, AIC, BIC, etc.

DATA MANIPULATION AND CLEANING:

The analysis aims to merge and join the Table spread, Cooking and Salad Oils, and Cooking Sprays data from 2018 to 2022 with their respective attribute files using Excel, Python, and Alteryx software. The goal is to test hypotheses and perform regression analysis on the data.

To ensure the reliability of the regression model, null values present in the data will be dropped. Null values in the sales data represent a lack of sales in that category. Imputing null values may increase the endogeneity of the data, leading to a less reliable regression model. New columns may be added to calculate Total Sales, %_inc_in_sales, and reduced_price. Discounts or offers during a specific period, referred to as the Merch, will also be considered.

Overall, the analysis will examine sales trends for Table spread, Cooking and Salad Oils, and Cooking Sprays over the past five years. Regression analysis will identify any factors that may have influenced these trends and help develop strategies to improve future sales performance.

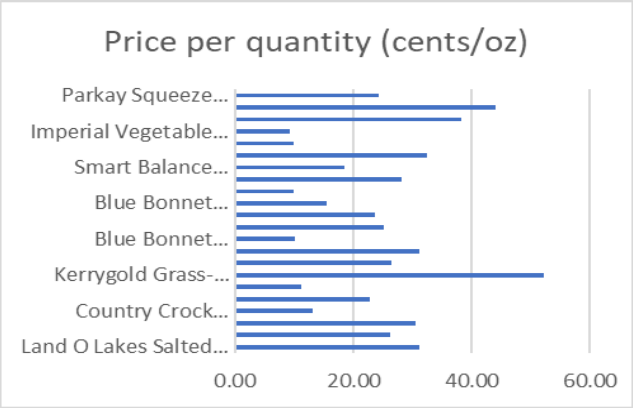
MARKET ANALYSIS:

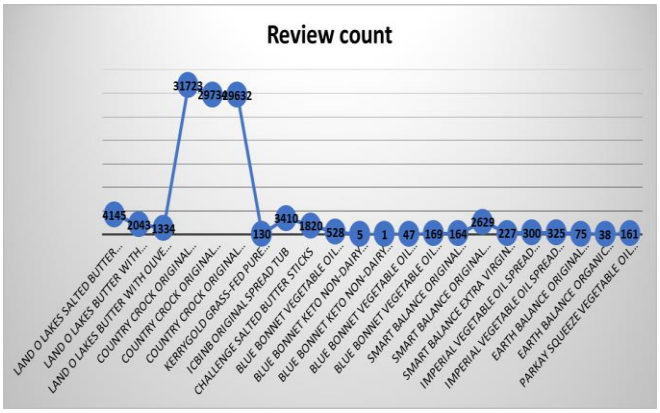
Prior to analyzing the data and performing statistical analysis, we conducted market research to gain a deeper understanding of the Table spreads market and identify key factors that influence buyers' decisions when purchasing these products. To conduct the research, we visited several Walmart stores in the Frisco/McKinney/Richardson areas and interviewed shoppers during a busy weekend period to obtain valuable insights for the project. The study was conducted at noon on a Friday, which is typically lunchtime in North Dallas, to attract a larger number of consumers for analysis. After analyzing the compiled data, we identified several noteworthy findings.

The market research revealed several insights into Table spreads products and customer preferences (market analysis-data.xlsx) *. Here are the key findings:

- The Country Crock brand was the most popular product, followed by Great Value. Half of the Country Crock products were sold out.
- White women in their late 30s and early 40s were the largest group of purchasers of table spreads, cooking oils, and cooking sprays. The second largest group was Hispanic, followed by African American.
- Customers who were concerned about health paid attention to the ingredients and percentages of fat and cholesterol on the label. Imperial and Blue Bonnet vegetable oil spread tubs were popular choices because their percentages were between 5-9.
- Important factors for customers when choosing a table spread product included brand value, reviews from a well-known source, flavor, price, discounts, marketing, and nutritional information.
- Land O Lakes products had an average price, with Country Crock coming in second. Kerry Gold and Earth Balance Organic products were more expensive.
- Customers had different preferences for the size/quantity of products depending on their needs.

- Except for some Conagra products, no company explicitly mentioned the proportion of oil mixed up in their butter products.
- Country Crock had average levels of oil concentration but was still the leading brand in terms of high average rating and review count.
- Earth Balance had lower sales compared to all other brands, including other Conagra products, due to a higher proportion of oil content.
- The Smart Balance Original Buttery Spread Tub had a higher fat and oil percentage due to an increase in tub size, but the reasons for this were unknown.
- Although Land O Lakes products clearly mentioned cholesterol percentage on the Nutrition Facts label, they were still one of the key players in sales.
- Smart Balance had the lowest average rating compared to all other brands, including other Conagra products.





DATA ANALYSIS AND HYPOTHESIS TESTING:

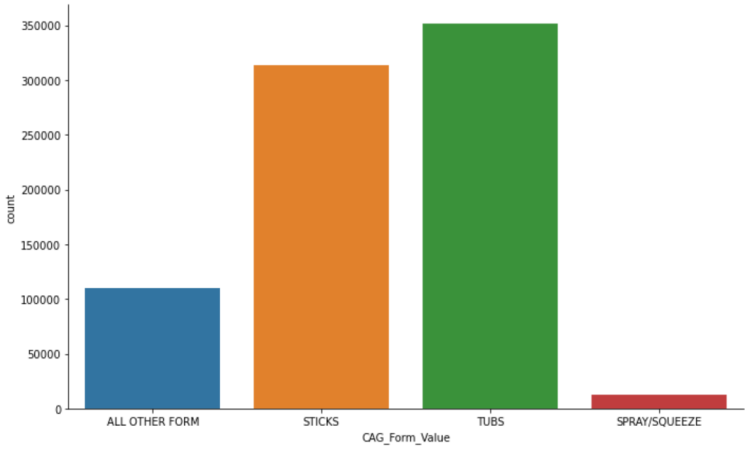
The data provided to Conagra appears to be of high quality, having been thoroughly processed and cleaned, with no further data cleansing or sampling required. However, after analyzing the different data sources, the team has identified the type of data being dealt with and summarized their findings as follows:

** For IRI Point of Sales (POS) data, the Table spreads data set contains a panel of observations of Brand, manufacturer, sub-category, category, CAG Count, CAG Ounces, CAG Form, and CAG Tier Value over a period of 5 years on a weekly basis. However, the panel is unbalanced as not all weeks have data available for each product description feature. Similarly, the Cooking & Salad Oils and Cooking Spray data set also contains a panel of observations of Brand, manufacturer, and category over 5 years on a weekly basis, with an unbalanced panel for product description features.

The IRI Panel data includes two cross-sectional datasets, namely the Buyer Distribution data set and the HH Buying data set. The Buyer Distribution data set is a summary of observations covering 52 weeks that ended on January 1st, 2023, broken down by Category, subcategory, and brand for various demographic categories.

To test the hypothesis, linear regression and OLS models have been created. Categorical variables are transformed using one hot encoding to enable regression analysis. Regression and hypothesis testing are conducted across various categories to gain insights into the market and make predictions about sales.

1. Anova test for Form and Total Sales: To understand the relationship between the Form of Table spread and the Total sales.
H0: There is no significant difference between the means of total sales of all different forms.
H1: There is a significant difference between the means of total sales of all different forms.



F-statistic: 2109.94286
P value: 0.0
R squared: 0.008
Adj R squared: 0.008.

As the P value is less than 0.05 at 95% confidence interval, we can reject the null hypothesis. We can conclude that there is a significant difference between the means of total sales of all different forms.
This means there is at least one form among four that has a different mean value when compared to the rest of the forms' means. The F statistic 2109.9 compares the variance between the Forms, and this is considered high which indicates a larger mean difference in each form. 8% of the variance in the dependent variable, Total Sales, can be explained by the independent variable, Form. From this hypothesis we can understand that there is an effect of form on the sales.
To understand which forms, have a high and low effect, we need to perform pairwise Tukey HSD test on our model. This helps in implementing new marketing strategies that would help to increase the sales of products in a particular form that hasn't been performing well.

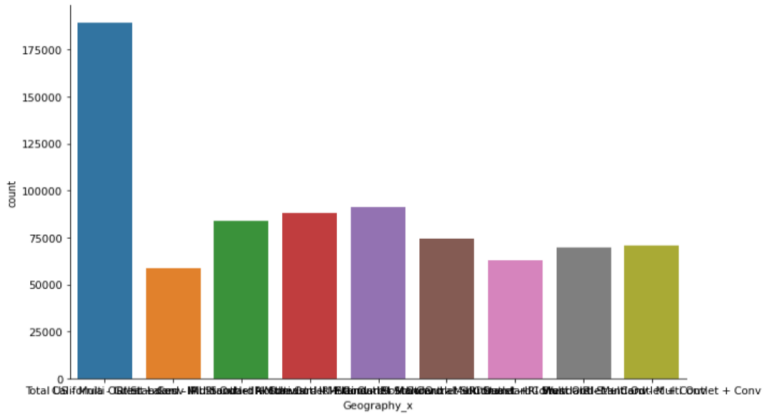
Tukey Test results specifying which form has significantly different sales:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
ALL OTHER FORM	SPRAY/SQUEEZE	33712.4161	0.0	28816.3274	38608.5048	True
ALL OTHER FORM	STICKS	48781.9245	0.0	46989.5204	50574.3286	True
ALL OTHER FORM	TUBS	18939.7899	0.0	17172.4144	20707.1653	True
SPRAY/SQUEEZE	STICKS	15069.5084	0.0	10333.9477	19805.0691	True
SPRAY/SQUEEZE	TUBS	-14772.6263	0.0	-19498.7704	-10046.4821	True
STICKS	TUBS	-29842.1346	0.0	-31098.107	-28586.1622	True

From the results we can specify that there is a statistical difference for every group of forms as the p value is less than 0.05 in every case.

2. Anova test for Geography and Total Sales: To understand the relationship between the geographical location and the Total sales.
H0: There is no significant difference between the means of total sales in 8 geographical locations.
H1: There is a significant difference between the means of total sales in 8 geographical locations.



F-statistic: 2976.438939
P value: 0.0
R squared: 0.029
Adj R squared: 0.029.

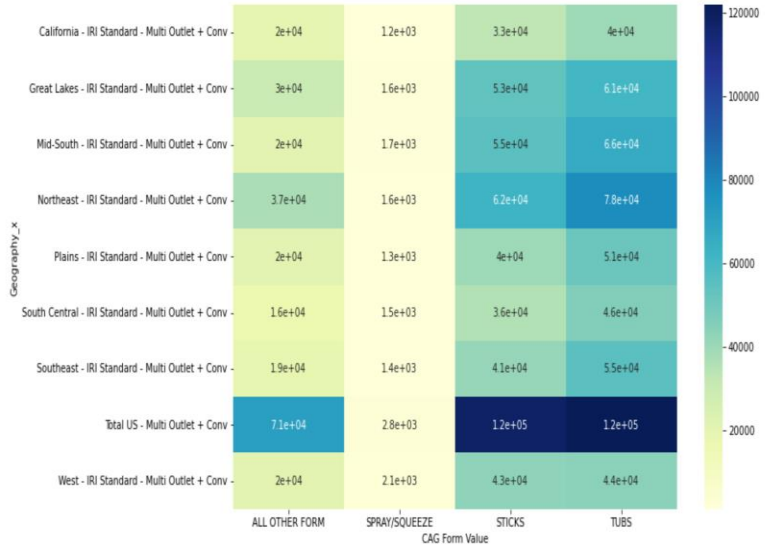
As the P value is less than 0.05 at 95% confidence interval, we can reject the null hypothesis. We can conclude that there is a significant difference between the means of total sales at all different locations. This means there is at least one geographical location among eight that has a different mean value when compared to the rest of the location's total sales means.
The F statistic 2976.438939 compares the variance between the different geographical locations, and this is considered high which indicates a larger mean difference in each location. 2.9% of the variance in the dependent variable, Total Sales, can be explained by the independent variable, Location. Performing Tukey Test to specify which geographical location has significantly different sales.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
California - IRI Standard - Multi Outlet + Conv	Great Lakes - IRI Standard - Multi Outlet + Conv	4484.8531	0.0007	1211.6083	7758.098	True
California - IRI Standard - Multi Outlet + Conv	Mid-South - IRI Standard - Multi Outlet + Conv	-3256.1046	0.0479	-6496.8934	-15.3159	True
California - IRI Standard - Multi Outlet + Conv	Northeast - IRI Standard - Multi Outlet + Conv	12888.5726	0.0	9668.0302	16109.1151	True
California - IRI Standard - Multi Outlet + Conv	Plains - IRI Standard - Multi Outlet + Conv	-13073.2045	0.0	-16430.516	-9715.893	True
California - IRI Standard - Multi Outlet + Conv	South Central - IRI Standard - Multi Outlet + Conv	185.5744	1.0	-3309.4328	3680.5816	False

From the results we can specify that the groups California - IRI Standard - Multi Outlet + Conv ~ South Central - IRI Standard - Multi Outlet + Conv has a p value of 1 which means there is no significant difference between sales among these locations. And likewise, if the p value for groups is less than 0.05, it signifies that there is significant difference between the sales among those specific groups and if the p value is greater than 0.05, it signifies that there is no significant difference among the sales of those specified groups.
From this analysis we can study the performance of table spreads in different locations and the market research can be done to understand why the sales in a particular location are less compared to other location and improve the strategies to increase sales.

3. Chi Square test for Geography and Form: To understand the relationship between the geographical location and the sales of form.
H0: Geography and form are independent.
H1: Geography and form are dependent.



Chi square Statistic:
9538.940515074088

P value: 0.0
Degrees of Freedom: 24

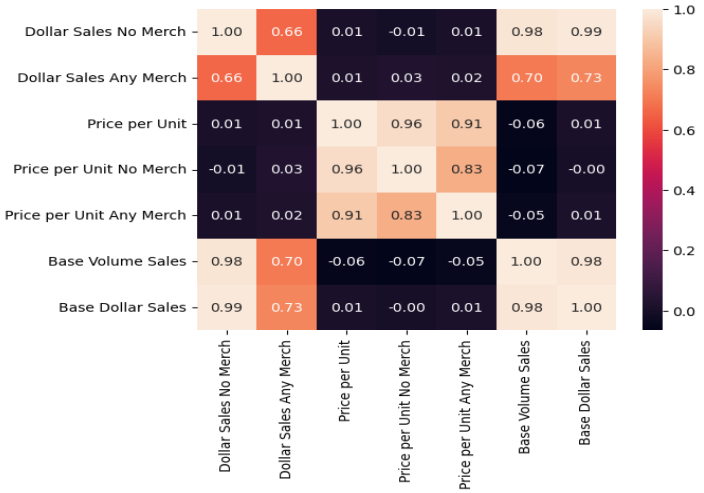
The p value 0.0 is less than 0.05 at 95% significance level. This means we can reject the null hypothesis and conclude that there is evidence to show that the geographical location and form of the table spread sales are dependent. This analysis helps in understanding which form is more popular at each location. And further this study can be used to improve the production rate of a particular popular form at a specific location and understand, improve sales of other forms.

REGRESSION ANALYSIS:

Various regression analysis has been performed to understand the relation between the variables like Total Sales, Price per volume, Price per unit (both with and without merch), Geographical Location, Form etc.,

Interaction Analysis:

The interaction map is plotted for the various product prices and dollar sales of the Margarine product across various brands from 2018 to 2022.



Hypothesis test:

H0: Optimizing the product price of margarine will not increase the dollar sales.

H1: Optimizing the product price of margarine will increase the dollar sales.

Regression Analysis-1: To understand the relationship between Total Sales and Price per Unit. Using the Margarine product data from the Spreadsheet data set 2018-2022, a regression analysis was performed by taking the 'Price per unit' as an independent variable and 'Total sales' as a dependent variable.

OLS Regression Results

Dep. Variable:	Total sales	R-squared (uncentered):	0.064
Model:	OLS	Adj. R-squared (uncentered):	0.064
Method:	Least Squares	F-statistic:	2573.
Date:	Mon, 24 Apr 2023	Prob (F-statistic):	0.00
Time:	16:11:34	Log-Likelihood:	-4.4747e+05
No. Observations:	37732	AIC:	8.949e+05
Df Residuals:	37731	BIC:	8.950e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Price per Unit	3779.9220	74.522	50.723	0.000	3633.858	3925.986

Omnibus:	71021.798	Durbin-Watson:	2.020
Prob(Omnibus):	0.000	Jarque-Bera (JB):	161360176.385
Skew:	14.389	Prob(JB):	0.00
Kurtosis:	322.073	Cond. No.	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
intercept = res.params[0]

print("Intercept:", intercept)
```

```
Intercept: 3779.922006764284
```

Y: Total sales
X: Price per unit

P value: 0.0
R squared: 0.064
Adj R squared: 0.064.

H0: Optimizing the price per unit of Margarine will not affect the Total sales.
H1: Optimizing the price per unit of Margarine will affect the Total sales.

As the P value is less than 0.05 at 95% confidence interval, we can reject the null hypothesis. We conclude that optimizing the product price of margarine will increase the dollar sales. This means that optimizing the price of at least one margarine product will increase the dollar sales. Optimizing the price may include the reduction of price or selling the product in any merchandising.
Optimization refers to reducing the price to attract the customers but in the limit that would not affect the profits of the company. The optimization ranges can be obtained by performing prescriptive analytics on the data.

Similar Regression analysis is performed for the 'No Merch' and 'Any Merch' products category. The regression results are as follows:

OLS Regression Results

Dep. Variable:	Dollar Sales No Merch	R-squared (uncentered):	0.059
Model:	OLS	Adj. R-squared (uncentered):	0.059
Method:	Least Squares	F-statistic:	2354.
Date:	Mon, 24 Apr 2023	Prob (F-statistic):	0.00
Time:	16:22:48	Log-Likelihood:	-4.4181e+05
No. Observations:	37732	AIC:	8.836e+05
Df Residuals:	37731	BIC:	8.836e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Price per Unit No Merch	2956.8965	60.946	48.517	0.000	2837.441	3076.352

Omnibus:	70541.290	Durbin-Watson:	2.025
Prob(Omnibus):	0.000	Jarque-Bera (JB):	155759999.930
Skew:	14.178	Prob(JB):	0.00
Kurtosis:	316.480	Cond. No.	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
intercept = res.params[0]

print("Intercept:", intercept)
```

```
Intercept: 2956.896459648006
```

Y: Dollar Sales No Merch
X: Price per Unit No Merch

P value: 0.0
R squared: 0.059
Adj R squared: 0.059.
H0: Optimizing the price per unit-no merch of Margarine will not affect the Dollar sales- no merch.
H1: Optimizing the price per unit-no merch of Margarine will affect the Dollar sales- no merch.

As the P value is less than 0.05 at 95% confidence interval, we can reject the null hypothesis.

OLS Regression Results

Dep. Variable:	Dollar Sales Any Merch	R-squared (uncentered):	0.040
Model:	OLS	Adj. R-squared (uncentered):	0.040
Method:	Least Squares	F-statistic:	1580.
Date:	Mon, 24 Apr 2023	Prob (F-statistic):	0.00
Time:	16:25:05	Log-Likelihood:	-3.8591e+05
No. Observations:	37732	AIC:	7.718e+05
Df Residuals:	37731	BIC:	7.718e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Price per Unit Any Merch	640.5235	16.114	39.750	0.000	608.940	672.107

Omnibus:	80284.223	Durbin-Watson:	1.949
Prob(Omnibus):	0.000	Jarque-Bera (JB):	449144630.179
Skew:	18.738	Prob(JB):	0.00
Kurtosis:	536.180	Cond. No.	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
intercept = res.params[0]

print("Intercept:", intercept)
```

```
Intercept: 640.5235161062312
```

Y: Dollar Sales Any Merc
X: Price per Unit Any Merch

P value: 0.0
R squared: 0.059
Adj R squared: 0.059.

H0: Optimizing the price per unit-any merch of Margarine will not affect the Dollar sales-any merch.
H1: Optimizing the price per unit -any merch of Margarine will affect the Dollar sales-any merch.

As the P value is less than 0.05 at 95% confidence interval, we can reject the null hypothesis. Hence, we can see that there is a significant correlation between the price per unit and Dollar sales. Training and testing need to be performed on the price data by predicting the price values and observing the effect of decrease or increase in price on the Dollar sales of the Margarine products.

OLS Regression Results

Dep. Variable:	Total sales	R-squared (uncentered):	0.031
Model:	OLS	Adj. R-squared (uncentered):	0.031
Method:	Least Squares	F-statistic:	1215.
Date:	Sun, 07 May 2023	Prob (F-statistic):	4.70e-262
Time:	15:22:48	Log-Likelihood:	-4.4812e+05
No. Observations:	37732	AIC:	8.962e+05
Df Residuals:	37731	BIC:	8.962e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Price per Volume	1729.6579	49.620	34.858	0.000	1632.401	1826.914

Omnibus:	70479.064	Durbin-Watson:	1.957
Prob(Omnibus):	0.000	Jarque-Bera (JB):	154596526.568
Skew:	14.152	Prob(JB):	0.00
Kurtosis:	315.302	Cond. No.	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Y: Total Sales
X: Price per Volume

P value: 0.0
R squared: 0.031
Adj R squared: 0.031.
H0: Optimizing the price per volume of Margarine will not affect the Total sales.
H1: Optimizing the price per volume of Margarine will affect the Total sales.

As the P value is less than 0.05 at 95% confidence interval, we can reject the null hypothesis. Hence, we can see that there is a significant correlation between the price per volume and Total Dollar sales. Training and testing need to be performed on the price data by predicting the price values and observing the effect of decrease or increase in price on the Dollar sales of the Margarine products.

Regression Analysis-2: To understand the relation between the total volume sold, price of product and the geographical location of Conagra products when no merch is provided.

The data from the Table spreads - POS 2018-2022 has been merged and joined with attribute file for this analysis. The required columns: Volume Sales No Merch, Price per Volume No Merch,

Geography. One hot encoder is used to map the Categorical Variable Geography to 1 or 0 values. The train test split ratio is taken to be 80:20 with a random state of 42. Linear regression analysis was used to fit the model. The obtained incept and coefficients are:

```
# provide a regression equation for predictions
intercept = reg_model.intercept_
coefficients = reg_model.coef_
```

```
print(intercept)
print(coefficients)
```

```
31671.18885126884
[-6892.09361255  3426.22805639  7482.9318352   3151.26130755
   818.43864286  3458.19398109  7699.82402937 48338.45943016
 -1020.05242531]
```

The R2 value obtained is 0.1.

The prediction regression equation can be written as Volume Sales No Merch = 31671.18 - 6892.09(Price) + 3426.23(a) + 7482.93(b)+ 3151.26(c)+ 818 .44 (d)+ 3458.19(e) + 7699.82(f) + 48338.46(g) - 1020.05(h).
To predict the Volume sales for a given price in a particular geographical location, Input price in price and 1 for the location, 0 for rest of the locations as given below:

Great Lakes - IRI Standard - Multi Outlet + Conv ~ a=1; rest 0
Mid-South - IRI Standard - Multi Outlet + Conv ~ b=1; rest 0
Northeast - IRI Standard - Multi Outlet + Conv ~ c=1; rest 0
Plains - IRI Standard - Multi Outlet + Conv ~ d=1; rest 0
South Central - IRI Standard - Multi Outlet + Conv ~ e=1; rest 0
Southeast - IRI Standard - Multi Outlet + Conv ~ f=1; rest 0
Total US - Multi Outlet + Conv ~ g=1; rest 0
West - IRI Standard - Multi Outlet + Conv ~ h=1; rest 0
California - IRI Standard - Multi Outlet + Conv ~ all 0 This analysis helps in optimizing the price based on location to improve the sales.

Regression Analysis-3: To understand the relation between the total volume sales with merchandise and without any merchandise based on the price per volume across various forms.
The data from the Table spreads - POS 2018-2022 has been merged for this analysis. The required columns: "Volume Sales No Merch", "Volume Sales Any Merch", "Price per Volume No Merch", "Price per Volume Any Merch", "CAG Form Value".
<u>New Columns:</u>
%_inc_in_sales = Volume sales Any Merch/Volume sales No Merch * 100
Reduced_price = Price per Volume No Merch – Price per Volume Any Merch
One hot encoding is used to map the categorical variable "CAG Form Value."

The train test split ratio is taken to be 80:20 with a random state of 42. Linear regression analysis was used to fit the model. The obtained incept and coefficients are:

```
# provide a regression equation for predictions
intercept_merch = reg_model_merch.intercept_
coefficients_merch = reg_model_merch.coef_

print(intercept_merch)
print(coefficients_merch)
```

```
-11.390012000302335
[380.15317761 372.84851657 162.49948683 -12.62521698]
```

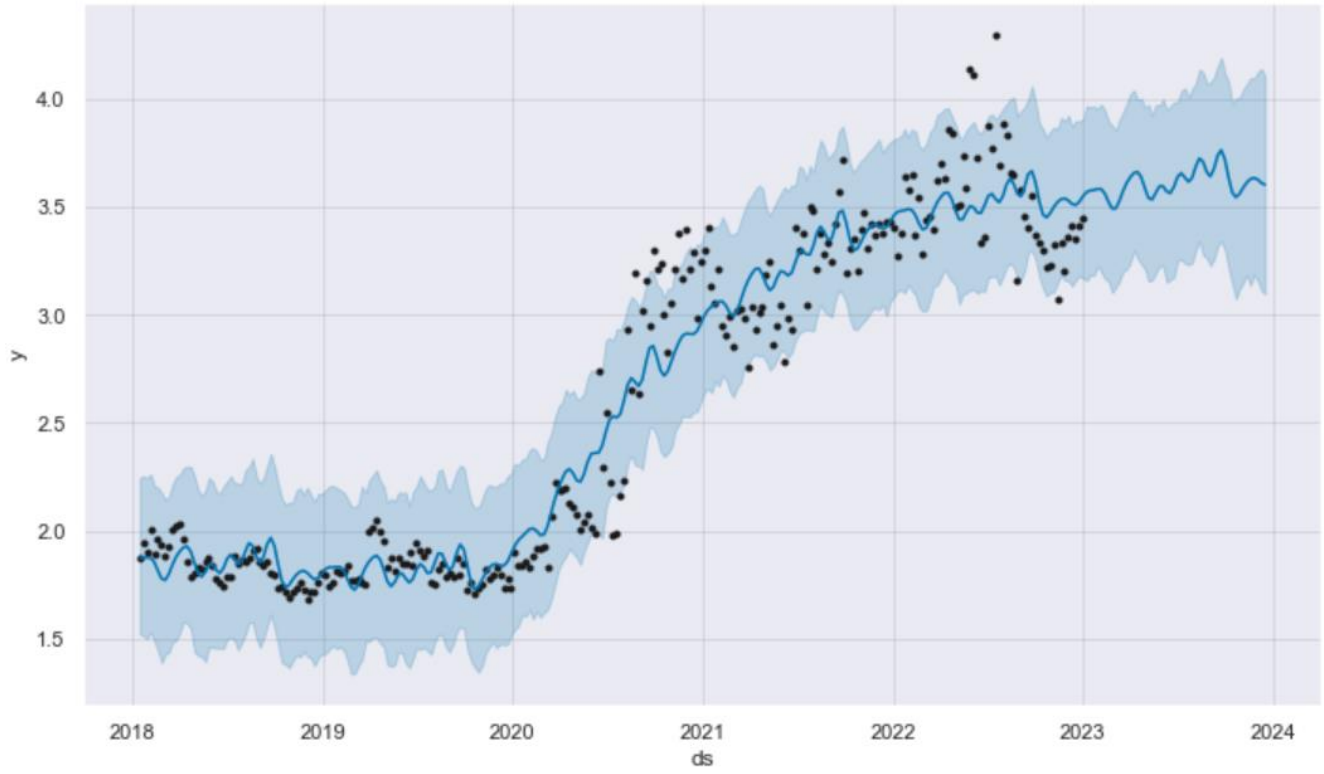
The Prediction equation can be written as: %_inc_in_sales = -11.39 + 380.15(reduced_price) + 372.84(a) + 162.5(b) -12.625(c)
To predict the % inc in sales for a given price reduction in a particular form, Input reduced price in reduced_price and 1 for the form, 0 for rest of the forms as given below.

Spray/Squeeze ~ a=1; rest 0
Sticks ~ b=1; rest 0
Tubs ~ c=1; rest 0
All Other Forms ~ all 0

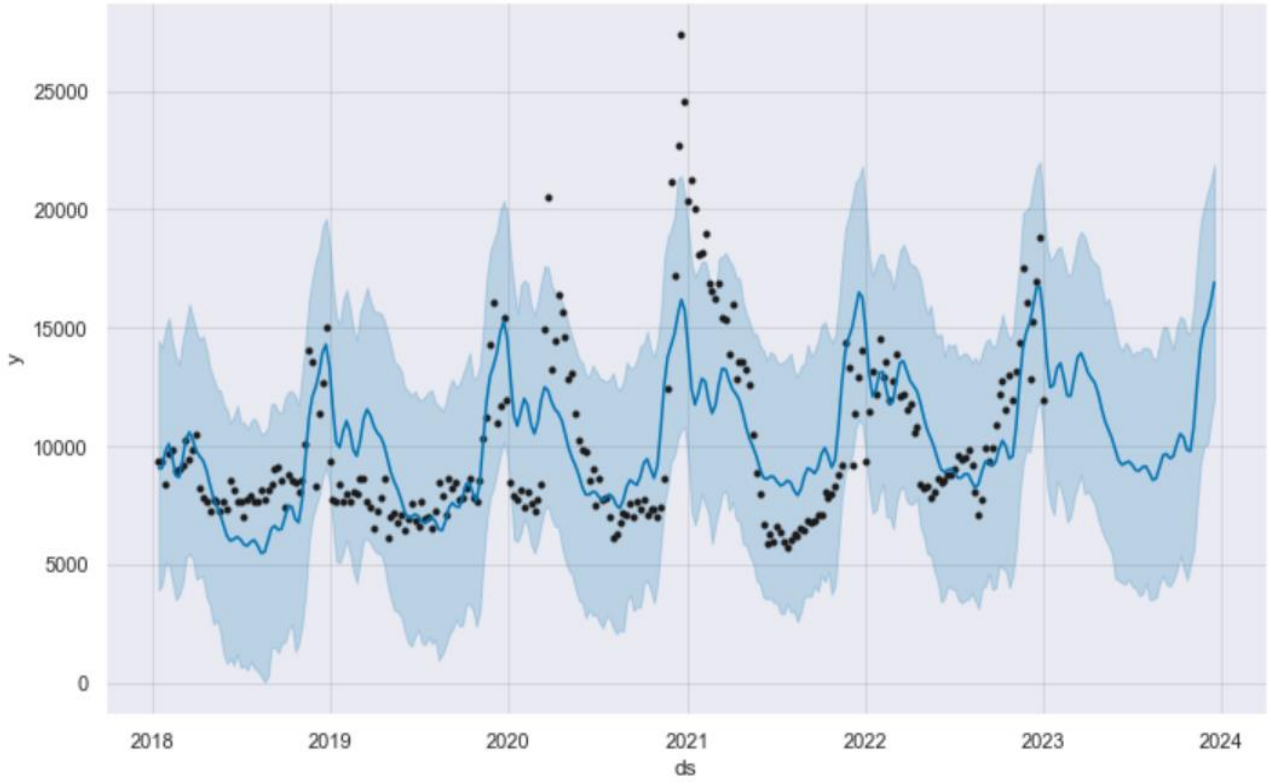
Regression Analysis-4: To understand the trend and forecast the possible near future values in the Total Sales and trend in the Price per Volume and Price per Unit for the **Margarines**.
The Prophet library is used for this analysis. Prophet is an open-source library developed by Facebook and designed for automatic forecasting of univariate time series data.

Weekly seasonality is interpreted, and the forecast is generated for the next 50 periods.
Forecast-a: Price per Unit

Forecast-b: Price per Volume



Forecast-c: Total Sales

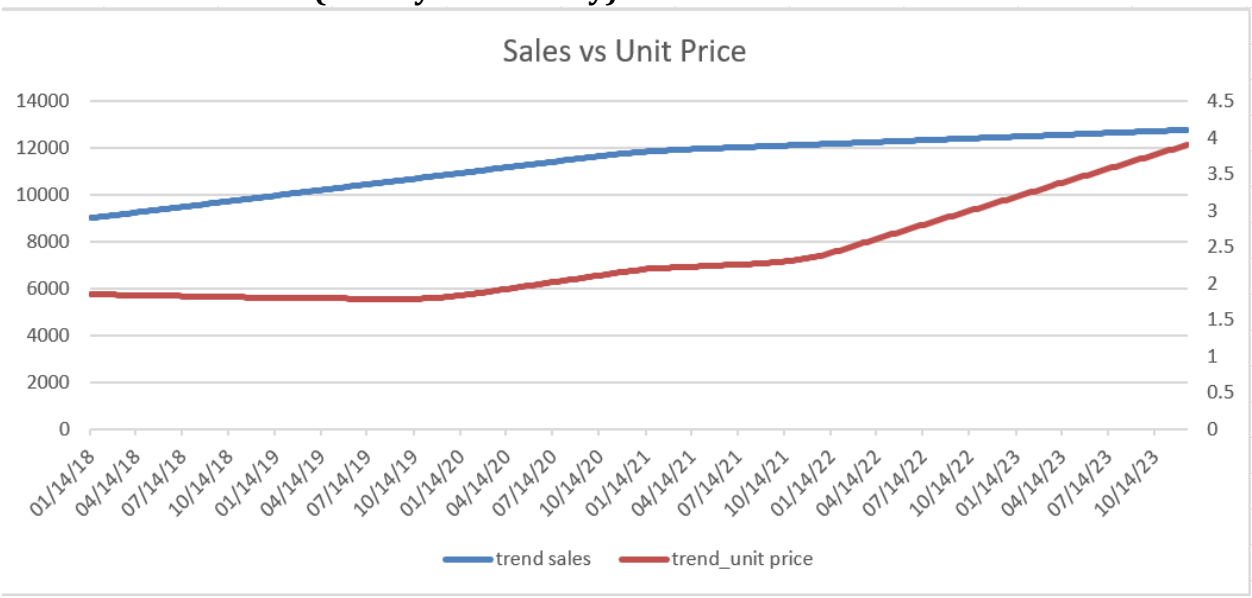


Overall average trend for each month:

From the visualization representing the trend for each month, we can interpret that the Total sales are rapidly increasing from November and are high during December of every year. And the next peak in sales is during April of every year.
The least sales can be observed during the months June, July, August, and September.

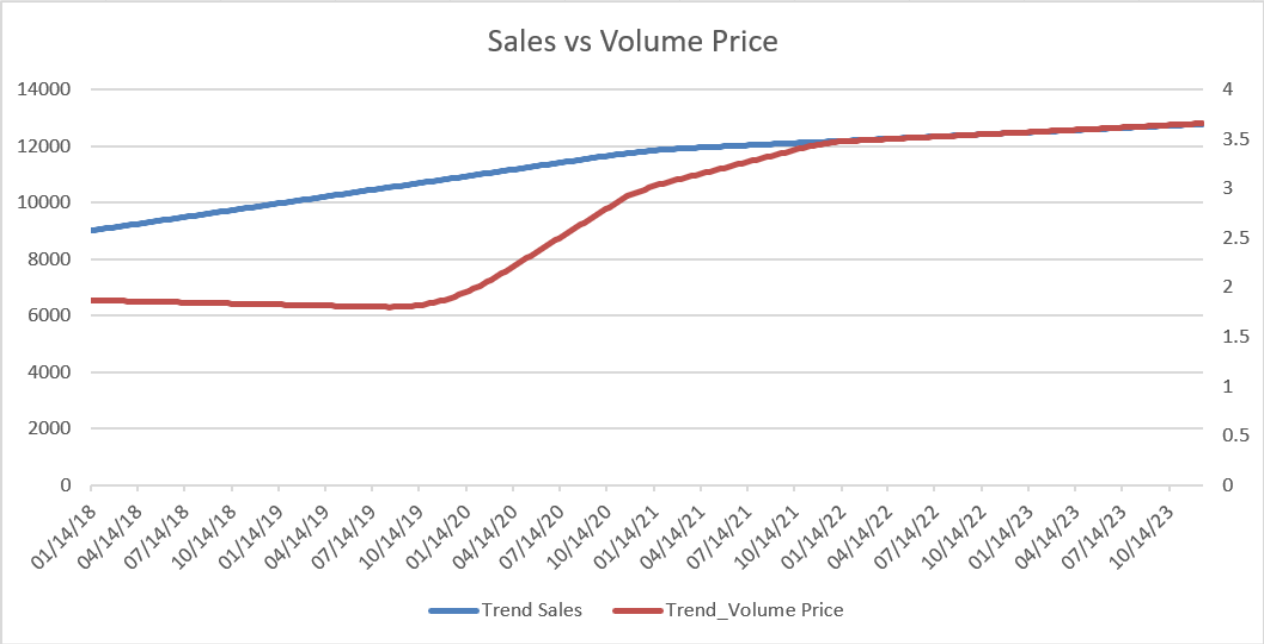
We can observe that the months with high sales are the festivity months with Thanksgiving, Christmas, and New Year. This could also be a potential reason for the sudden increase in total sales of Margarine.

Unit Price vs Total Sales (weekly seasonality):



As we can observe that the trend values generated using the prophet library for the Total sales and the unit price seem to be increasing for each period.
One can clearly observe that the sales were increasing when the prices were moderate and when there is a hike in price per unit, the total sales tend to get saturated.

Volume Price vs Total Sales (weekly seasonality):



There seems to be a very steep hike in price per volume from the beginning of 2020 to the beginning of 2022 after which the prices are running almost constant. And the period being the world pandemic, the price hike is justified. Though there was an increase in price per volume, the total sales were gradually increasing. This represents that the margarine consumption among the people was not much affected by the price hike during pandemic, but the sales are seeing a decrement in the trend forecast.

SOME PRESCRIPTIVE ACTIONS FOR CONAGRA TO INCREASE THE SALES:

- From the trend analysis of total sales, it is observed that the sales are high during the months of November and December. Conagra must ensure to have huge production budgets and material to meet the demand of supply during the final quarter of the year.
- During the final quarter, as the sales are predicted to be high, Conagra must work on improving sales by adding new customers. This can be done by offering enticing promotions and discounts within the profitable brackets.
- In the light of pandemic, people have grown more inclined towards healthy eating habits. Providing healthy options such as low-fat products can bring greater changes to the brand sales.
- Conagra may be interested in targeting the Hispanic community in some of the cities they reside the most, such as Fresno, Stockton, Salinas, Bakersfield, Lancaster, Palm Moreno, Valley Santa, and Clarita. These cities report an average household income that goes from 30- 75k (figure 5). It's also known Hispanic people include the use of table spreads in their diet (bread/toast, eggs dishes, vegetables, and potatoes etc.).
- From our analysis, there exists a relation between the form and Total sales. The form Tubs has the highest sales and spray/squeeze has lowest sales. Conagra must focus on improving the sales of spray/squeeze. One simple and cost-efficient way would be to add all the other forms available for the product on the label of the product.

- Incorporate better marketing strategies at the locations that have less sales such as ad campaigns and TV advertisements to reach people.
- Optimize the prices in a profitable bracket to attract new customers. Customers in income range greater than or equal to $100,000 account for the majority of sales. New strategies such as lower volume products or higher volume at a comparatively lower price must be used to attract the customer with lower income ranges.

Some general strategies to improve sales:
Offering discounts and reducing prices can be a useful strategy to increase sales, but it is not always the best solution.
Here are some strategies to consider:

1. Improve product quality: Customers are willing to pay more for high-quality products that meet their needs. Investing in research and development to improve product quality can lead to higher sales and profits in the long run.
2. Expand product line: Offering a wider range of products can attract new customers and increase sales to existing customers. Introducing complementary products or expanding into new markets can also increase profits.
3. Enhance customer experience: Providing excellent customer service can build customer loyalty and increase sales. Offering a convenient and enjoyable shopping experience, providing fast and reliable shipping, and having a user-friendly website can also improve customer satisfaction and retention.
4. Target new customers: Identifying new customer segments and targeting them with personalized marketing campaigns can lead to increased sales. Conducting market research to understand customer needs and preferences can help businesses develop effective marketing strategies.
5. Reduce costs: Reducing production costs can increase profit margins and allow businesses to lower prices while maintaining profitability. Strategies to reduce costs may include optimizing the supply chain, streamlining operations, and reducing waste.
6. Invest in marketing: Increasing brand awareness through marketing and advertising can help businesses attract new customers and increase sales. Developing a strong brand identity and leveraging social media can also help businesses connect with customers and build loyalty.
7. Adopting technology: Investing in technology such as e-commerce platforms, customer relationship management systems, and artificial intelligence can streamline operations, improve customer experience, and increase sales and profitability.

A company can consider implementing strategies that can improve profits from sources other than just Sales.

1. Upselling and cross-selling: Encouraging customers to purchase additional items or to upgrade to a more expensive product can increase sales and profitability. Offering discounts or incentives for purchasing bundles or complementary products can also be effective.
2. Implementing a loyalty program: Rewarding customers for repeat purchases and referrals can increase customer loyalty and encourage repeat business.

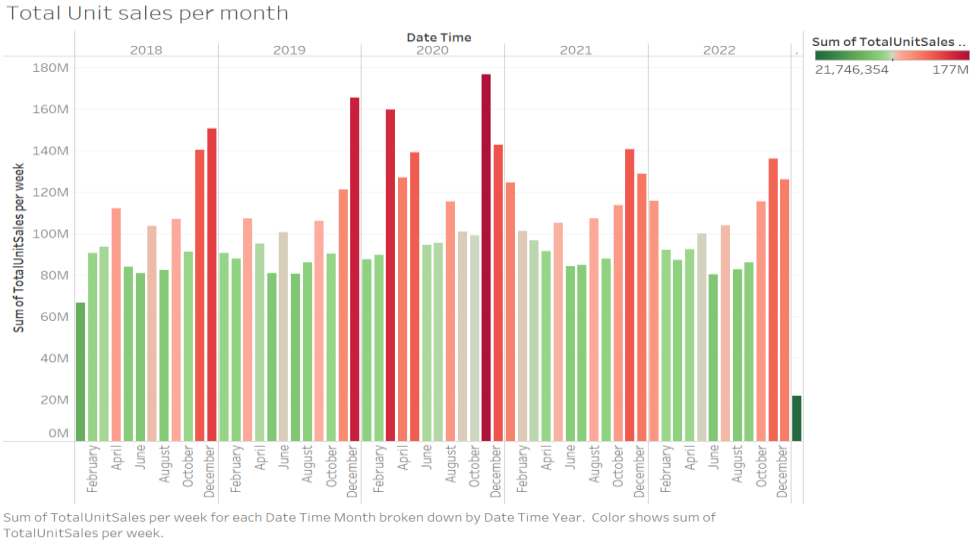
3. Partnering with other businesses: Collaborating with other businesses in the same or related industries can expand the customer base and increase sales. For example, a fitness apparel company could partner with a gym to offer a discount on membership to customers who purchase their products.
4. Offer financing options: Providing financing options can make expensive products more accessible to customers who may not have the funds to purchase upfront. Offering flexible payment plans can increase sales and customer loyalty.

APPENDIX

Summary Statistics for each geographical location: (summary_statistics_for_region.xlsx)

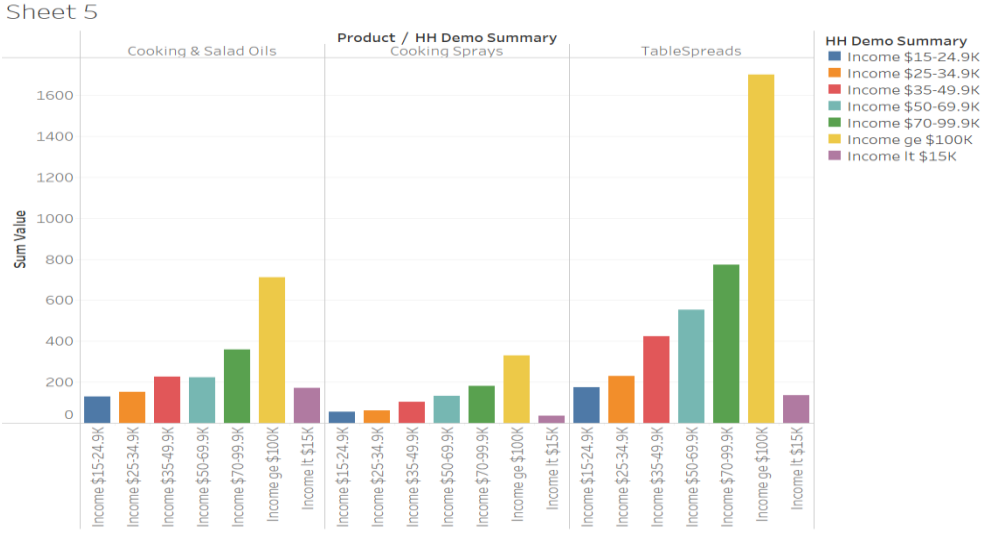
	Total Sales							
Geography	**Count**	**Mean**	**Std. Dev.**	**Min**	**25%**	**50%**	**75%**	**Max**
California - IRI Standard - Multi Outlet + Conv	56,801	33,386	74,300.84	1.88	2,509	8,691	33,565	2,696,891.28
Great Lakes - IRI Standard - Multi Outlet + Conv	78,452	38,235	107,574.01	1.51	1,404	7,008	31,012	3,511,234.61
Mid-South - IRI Standard - Multi Outlet + Conv	82,852	30,622	83,468.52	2.06	1,430	6,206	23,921	2,717,802.66
Northeast - IRI Standard - Multi Outlet + Conv	85,984	46,743	122,324.94	1.82	2,471	10,818	39,613	4,951,354.24
Plains - IRI Standard - Multi Outlet + Conv	70,679	20,402	53,321.51	2.84	996	4,201	16,192	1,258,688.87
South Central - IRI Standard - Multi Outlet + Conv	59,621	33,894	76,237.56	1.29	1,774	7,470	28,982	1,924,306.38
Southeast - IRI Standard - Multi Outlet + Conv	66,295	44,294	109,593.71	2.12	2,322	11,248	41,596	4,219,204.05
Total US - Multi Outlet + Conv	181,791	115,066	373,285.61	1.08	3,094	13,862	74,522	17,570,600.31
West - IRI Standard - Multi Outlet + Conv	67,761	33,939	74,417.17	2.50	1,544	6,921	29,639	1,561,605.08

Total Unit Sales per Month:



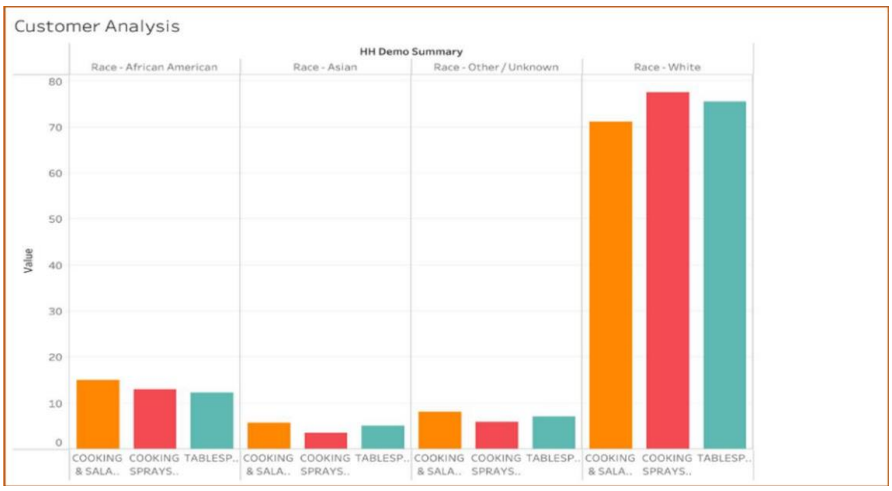
Sum of TotalUnitSales per week for each Date Time Month broken down by Date Time Year. Color shows sum of TotalUnitSales per week.

Income vs Product Buyer Distribution:

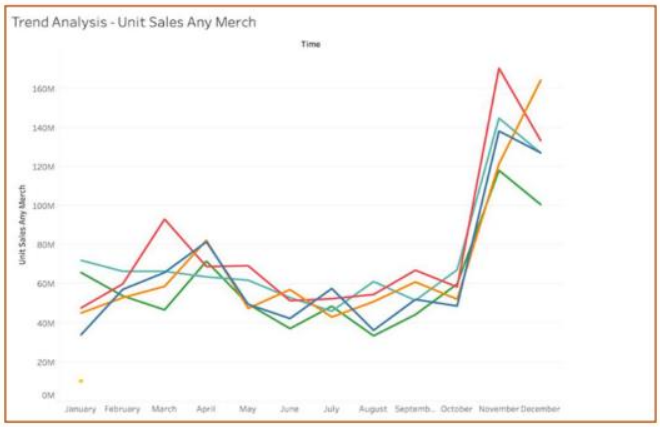


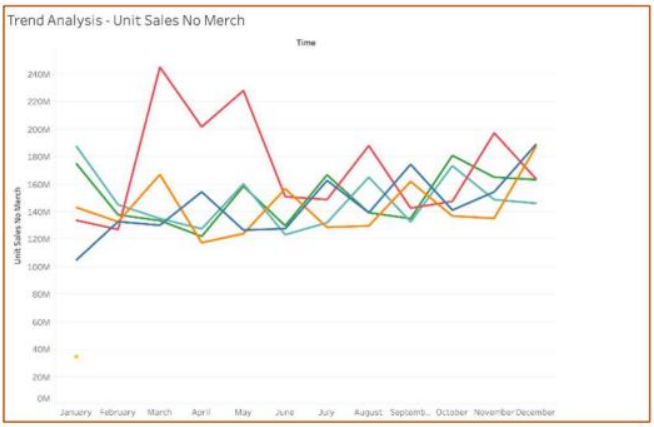
Sum of Sum Value for each HH Demo Summary broken down by Product. Color shows details about HH Demo Summary.

Race vs Form



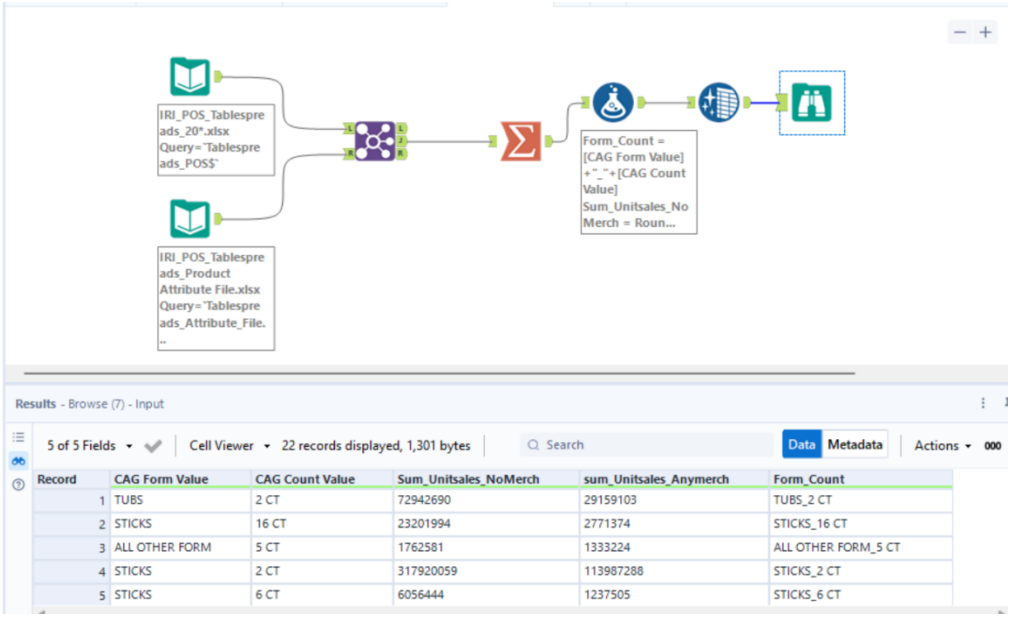
Trend Analysis- Unit sales- Any Merch vs No Merch:





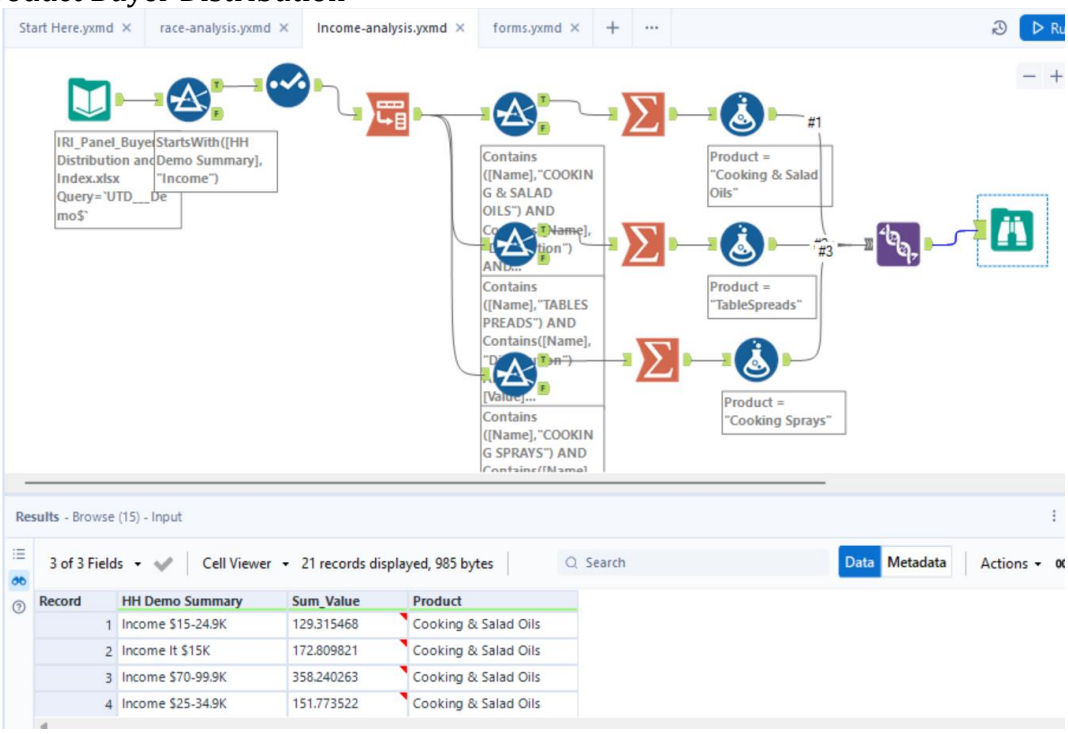
Alteryx workflows:

Forms-Units sales:



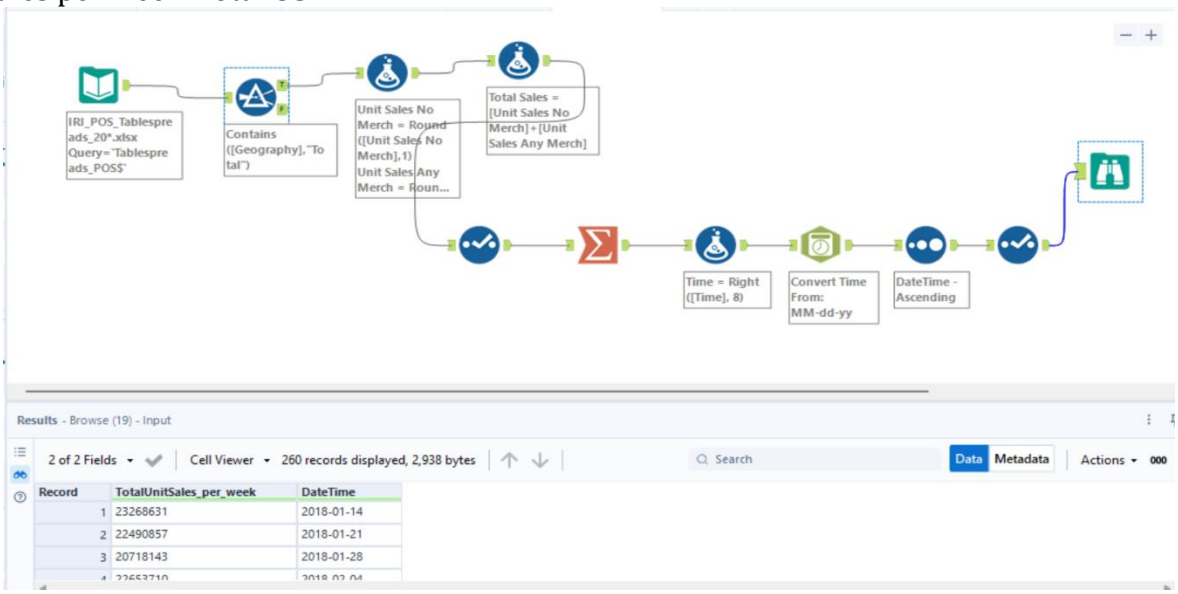
Record	CAG Form Value	CAG Count Value	Sum_Unitsales_NoMerch	sum_Unitsales_Anymerch	Form_Count
1	TUBS	2 CT	72942690	29159103	TUBS_2 CT
2	STICKS	16 CT	23201994	2771374	STICKS_16 CT
3	ALL OTHER FORM	5 CT	1762581	1333224	ALL OTHER FORM_5 CT
4	STICKS	2 CT	317920059	113987288	STICKS_2 CT
5	STICKS	6 CT	6056444	1237505	STICKS_6 CT

Income-Product Buyer Distribution



Record	HH Demo Summary	Sum_Value	Product
1	Income $15-24.9K	129.315468	Cooking & Salad Oils
2	Income lt $15K	172.809821	Cooking & Salad Oils
3	Income $70-99.9K	358.240263	Cooking & Salad Oils
4	Income $25-34.9K	151.773522	Cooking & Salad Oils

Unit sales per week-Total US



Record	TotalUnitSales_per_week	DateTime
1	23268631	2018-01-14
2	22490857	2018-01-21
3	20718143	2018-01-28