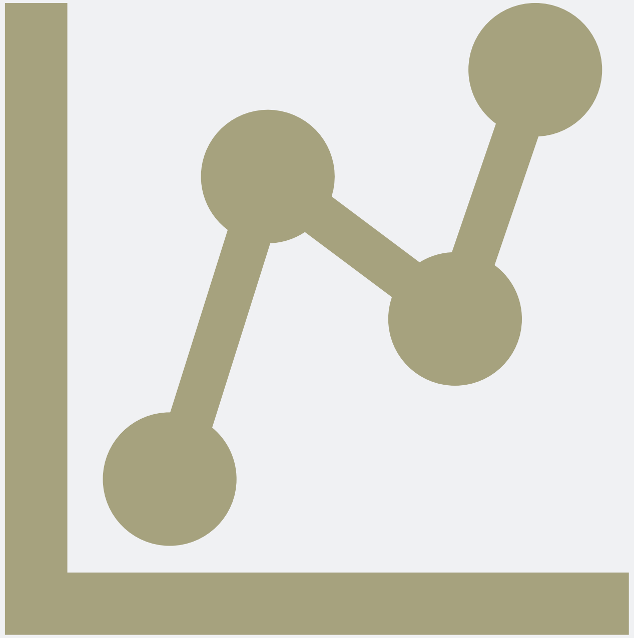


Analyzing Conagra's Sales Data: Insights for Future Business Decisions

A MACHINE LEARNING APPROACH

CONTENTS

MARKET RESEARCH AND STATS

The table spreads market in the US was valued at $1.1 billion in 2020, and it is projected to grow at a CAGR of 3.5% from 2021 to 2028. Conagra has a market share of around 22%.

Consumer preferences: According to a Mintel survey, 54% of US consumers prefer margarine and spreads that are made from natural ingredients, while 37% prefer low-fat products.

Competitive landscape: Other major players in the table spreads market include Unilever, Kerry Group, and Land O'Lakes.

Pricing strategy: Conagra has adopted a competitive pricing strategy in the table spreads market to remain competitive and gain market share.

KEY QUESTIONS

How do our loyalty and sales compare to our competitors in the table spreads category?

Who is our target audience based on region, gender, income, marital status, etc. and how can we tailor our marketing efforts to maximize sales?

What is our sales forecast for the table spreads category and how can we improve it?

What is the impact of merchandising on our table spreads sales?

What is the impact of product features on our table spread sales?

How do cooking spray, cooking oils, and table spreads interact with each other in terms of sales?

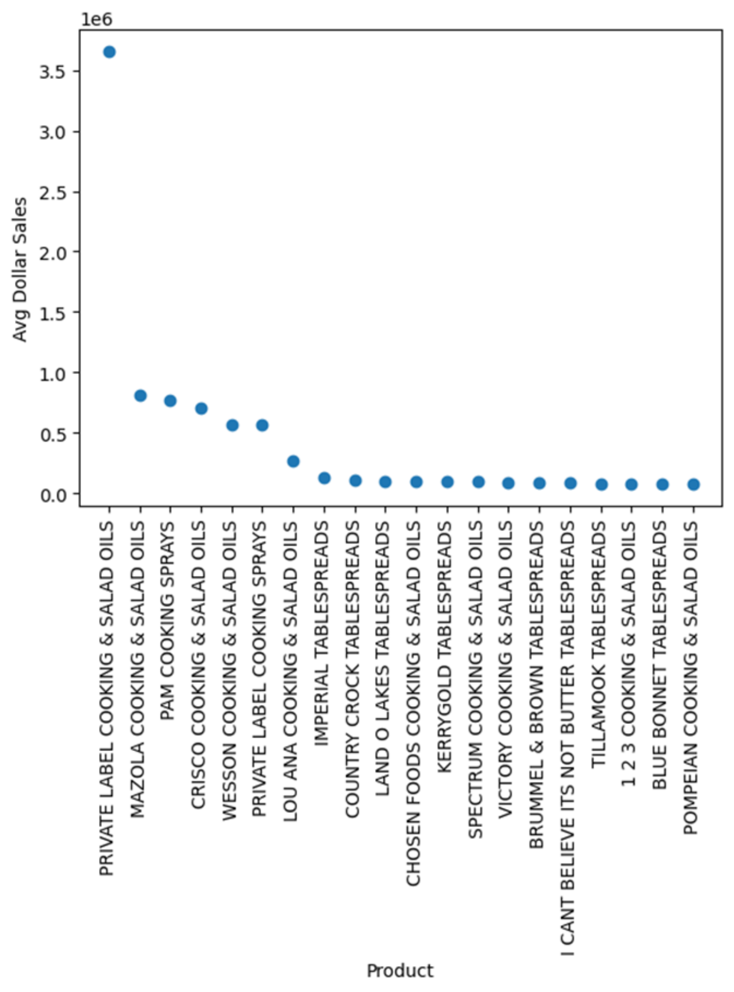
COMPETITORS

Brands **like Lodge, Primal, and La Tourangelle** have established a **loyal customer base with high conversion rates**.

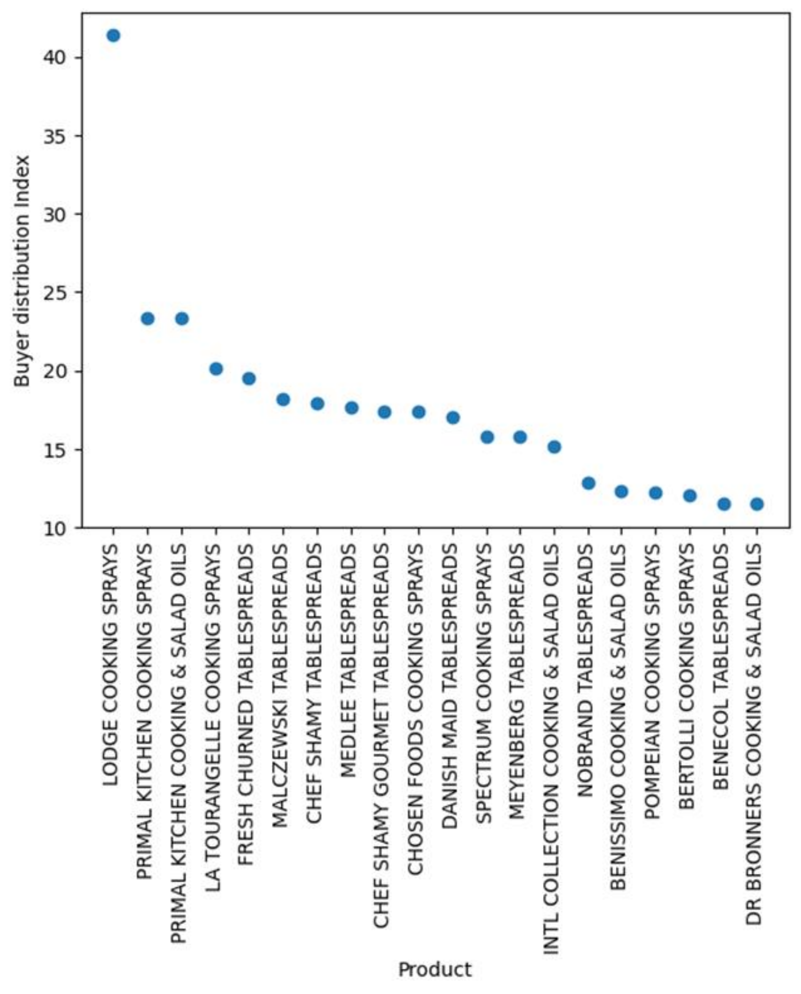
Brands like **Private Label, Mazola, PAM, and Crisco** have **high Sales**.

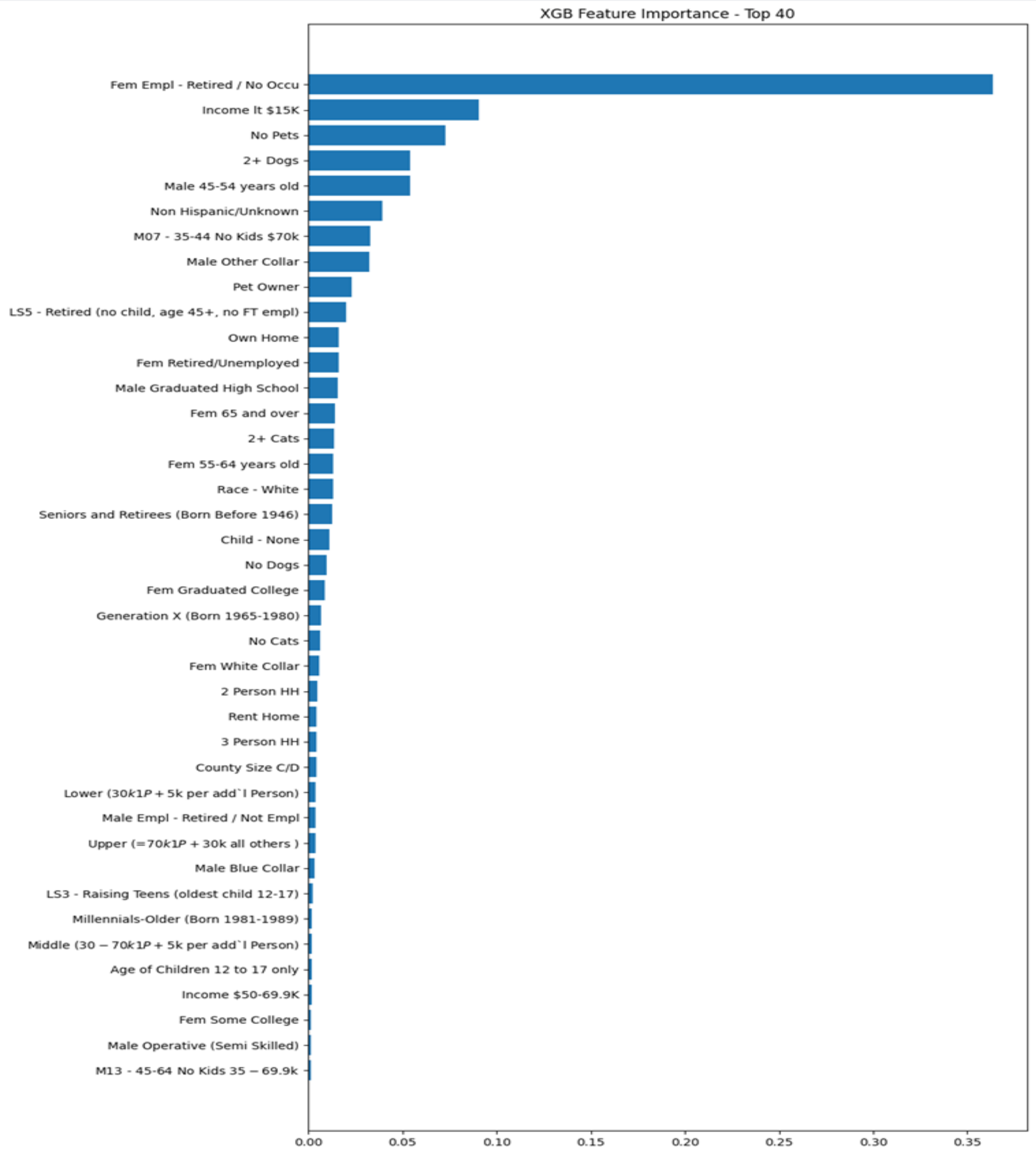
Competitor analysis can help identify gaps in the market, potential threats, and opportunities for growth, and provide insights into customer behavior, industry trends, and best practices in the market.

SALES BASED COMPETITORS



LOYALTY BASED COMPETITORS





DEMOGRAPHIC

- Employed XGBoost based feature selection method for analyzing demographic data
- The largest user group/ buyers for table ads and its substitution categories like cooking sprays and cooking oils are the **middle-class females who are either housewives or retired**.

REGRESSION

Sales forecasting can use both regression analysis and time series analysis techniques, depending on the nature of the sales data and business needs.

Combining these techniques can result in a more accurate and comprehensive sales forecast by analyzing the relationship between sales and various factors and identifying patterns and trends in historical sales data.

REGRESSION

	variables	VIF
0	const	13.119500
1	Price per Unit Any Merch(Tablespreads)	1.221101
2	Price per Volume Any Merch(Tablespreads)	1.221653
3	ACV Weighted Distribution Any Merch(Tablespreads)	1.384866
4	ACV Weighted Distribution No Merch(Tablespreads)	1.391680
5	Price per Unit(Cooking Salads)	4.093471
6	Price per Unit Any Merch(Cooking Salads)	4.089563
7	ACV Weighted Distribution No Merch(Cooking Sal...	2.973614
8	ACV Weighted Distribution Any Merch(Cooking Sa...	2.987521
9	ACV Weighted Distribution No Merch(Cooking Spr...	2.369447
10	ACV Weighted Distribution Any Merch(Cooking Sp...	2.589958
11	Incremental Units(Cooking Sprays)	1.275753
12	Incremental Dollars(Cooking Sprays)*Incrementa...	1.106077
13	Incremental Dollars(Cooking Sprays)*Incrementa...	1.025335

The given sales data has high multicollinearity

We first created a linear regression model and calculated the VIF for each variable.

We then set a threshold for high VIF and identify the variables that have a VIF above that threshold.

If the VIF value for a variable is greater than 5 or 10, it indicates that there is a high degree of multicollinearity

REGRESSION

```
print(model.summary())

Output exceeds the size limit. Open the full output data in a text editor
                            OLS Regression Results
==============================================================================
Dep. Variable:    Dollar Sales Any Merch(Tablespreads)   R-squared:                 0.114
Model:                                             OLS   Adj. R-squared:            0.113
Method:                                  Least Squares   F-statistic:               226.6
Date:                                 Mon, 08 May 2023   Prob (F-statistic):         0.00
Time:                                         16:15:43   Log-Likelihood:      -2.9127e+05
No. Observations:                                22923   AIC:                  5.826e+05
Df Residuals:                                    22909   BIC:                  5.827e+05
Df Model:                                           13
Covariance Type:                             nonrobust
==================================================================================================================
                                                    coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------
const                                          2.689e+04   1910.166     14.076      0.000    2.31e+04    3.06e+04
Price per Unit Any Merch(Tablespreads)         -314.7410    425.546     -0.740      0.460   -1148.840     519.358
Price per Volume Any Merch(Tablespreads)      -1936.6483    199.052     -9.729      0.000   -2326.804   -1546.493
ACV Weighted Distribution Any Merch(Tablespreads)  9661.4220  220.879  43.741      0.000    9228.485    1.01e+04
ACV Weighted Distribution No Merch(Tablespreads)  -400.5059   73.050     -5.483      0.000    -543.689    -257.323
Price per Unit(Cooking Salads)                  -76.4750    263.445     -0.290      0.772    -592.845     439.895
Price per Unit Any Merch(Cooking Salads)       -352.5952    264.016     -1.336      0.182    -870.085     164.895
ACV Weighted Distribution No Merch(Cooking Salads)   31.0526   61.722    0.503      0.615     -89.926     152.031
ACV Weighted Distribution Any Merch(Cooking Salads)  69.9232  166.209    0.421      0.674    -255.858     395.704
ACV Weighted Distribution No Merch(Cooking Sprays)   25.4878   33.914    0.752      0.452     -40.985      91.961
ACV Weighted Distribution Any Merch(Cooking Sprays) -133.6703  95.939   -1.393      0.164    -321.718      54.377
...
```

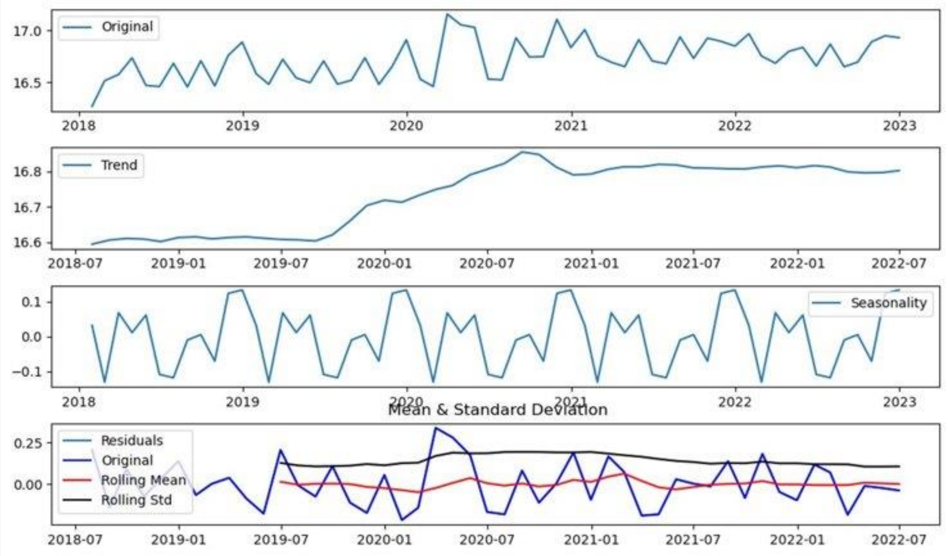
- The coefficients for Price per Unit Any Merch (Table spreads) and Price per Unit (Cooking Salads) are negative, indicating that an increase in these variables leads to a decrease in the dollar sales of table spreads.

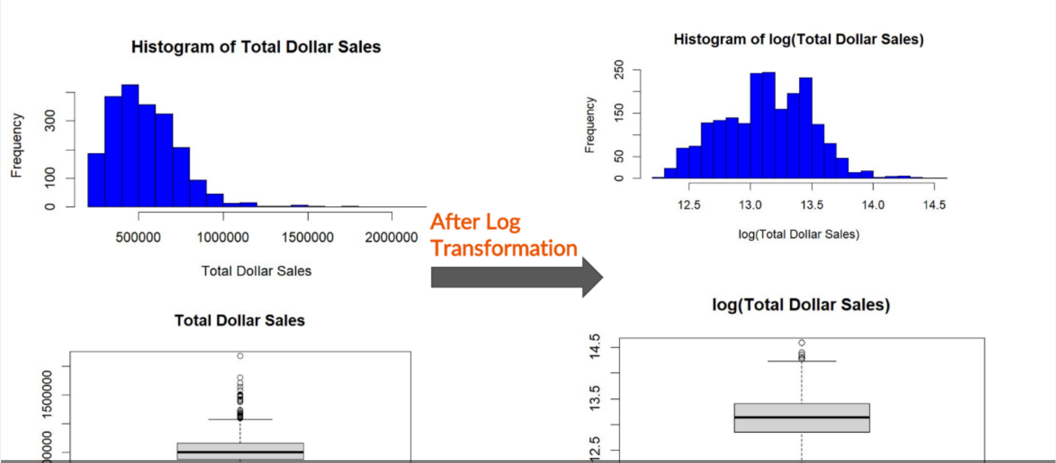
- the ACV Weighted Distribution Any Merch (Table spreads) and ACV Weighted Distribution No Merch (Table spreads) variables have a positive and negative coefficient, respectively, indicating that increasing the distribution of table spreads in stores with merchandising leads to higher dollar sales.

- However, low R-square and presence of multicollinearity makes Regression unsuitable for Sales data forecasting but useful for understanding the relationship between the variables.
- This is where time series comes into place.

TIME-SERIES FORECASTING

As we see the Table Spread Data on the Dollar Sales is Right Skewed, Having More outliers as shown in the box plot above. To make the data normal, we applied a natural log to the data to make the data normal. With this data we can eliminate the bias and omit necessary outliers which pull the regression line.



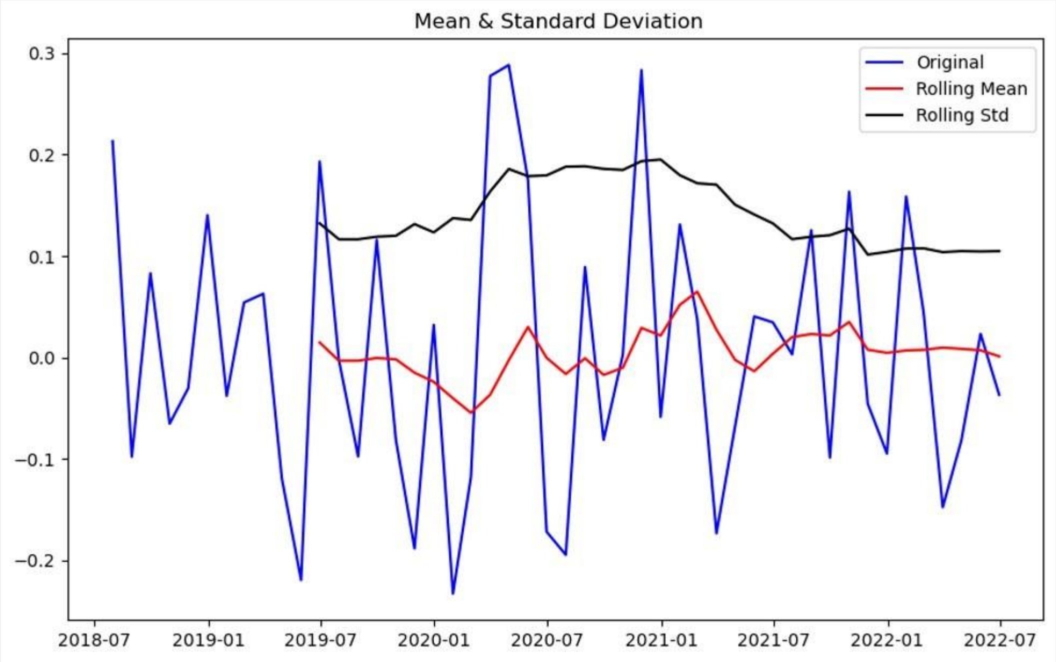


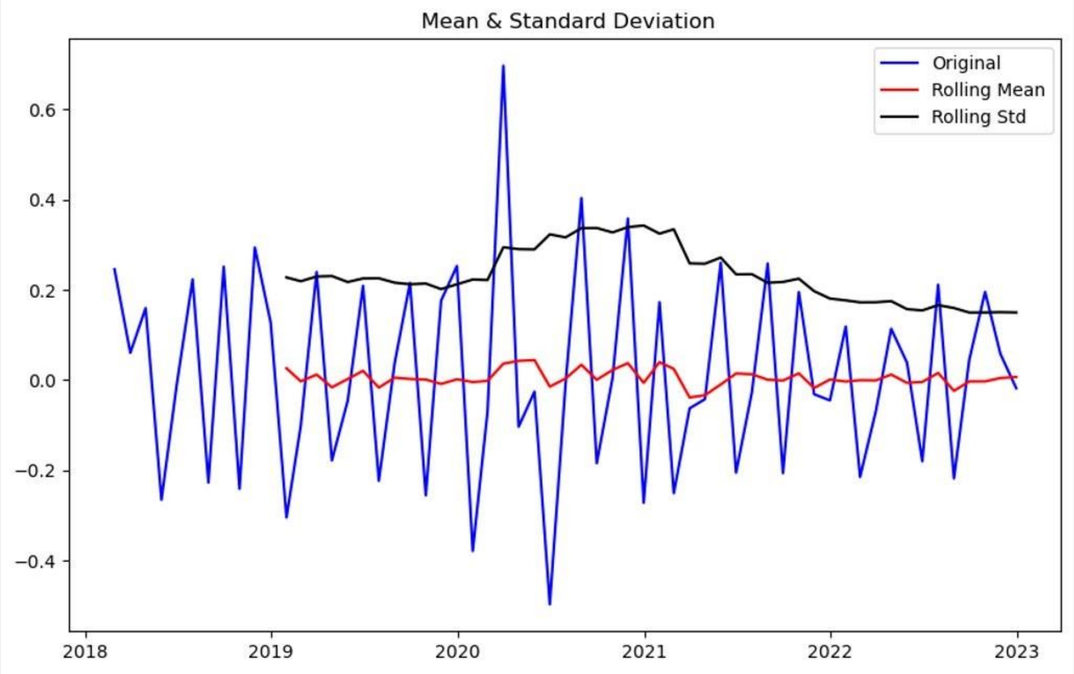
From the above graphs we can see the original data, Seasonality, Trend along with Rolling Mean and Std wr.t data.

TIME-SERIES FORECASTING

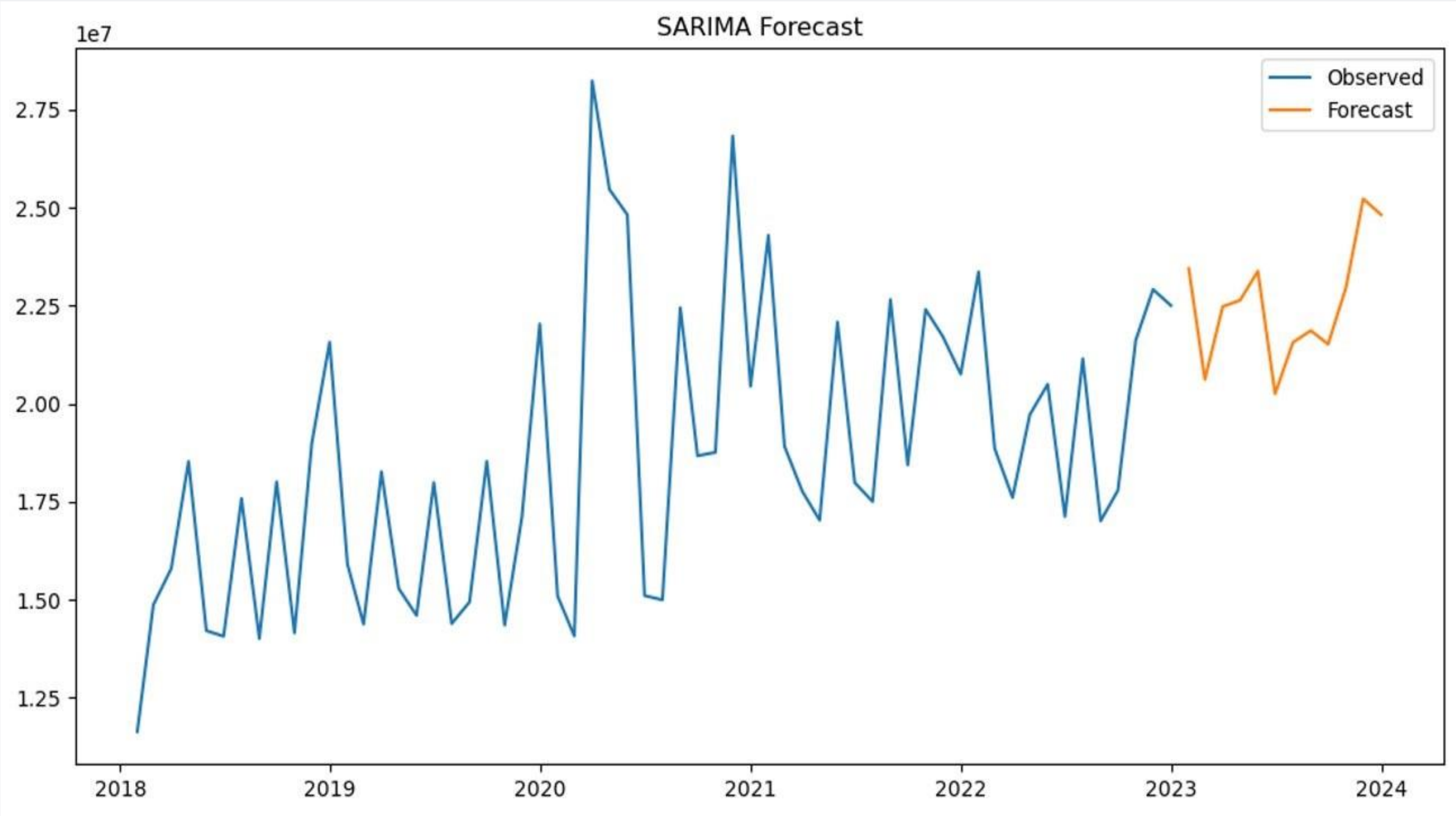
As per the moving average concept, We take the average of the first four months and then so on till the end, then take the average of that differentiated mean provides us with following figure

Here, we can see that the mean and standard deviation of the data is almost in a straight line (conditions for Data to be stationary). This is how you make the data Stationary.





TIME-SERIES FORECASTING



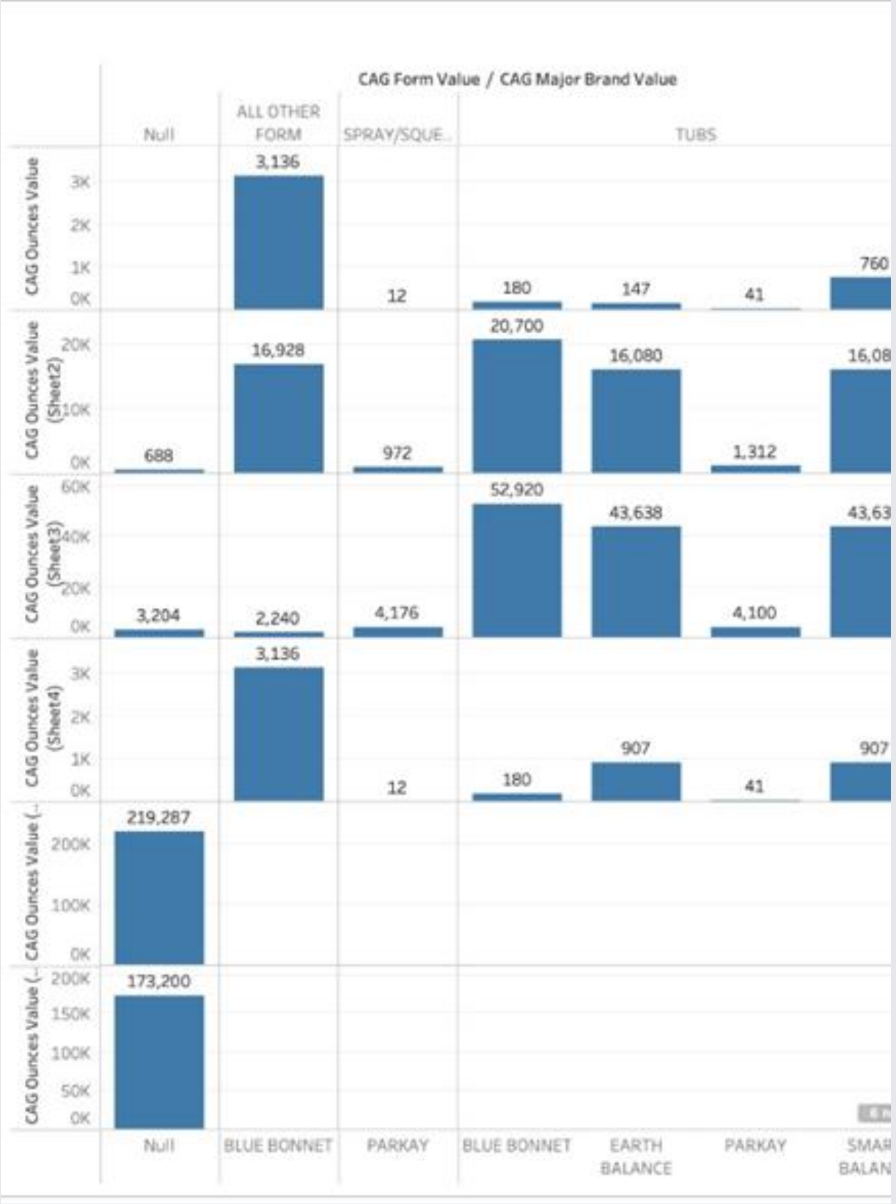
By using the SARIMAX code, we predicted the Values of sales for 2023 and 2024.

In This SARIMA process, We Find the Moving Average of the Data, then find the Seasonality component, Find the Trend component, Then Average Both Seasonality and Trend components for respective Month Variable, with this ST, we remove seasonality of the data which then use to predict the future values

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:            spray_total dollar sales   No. Observations:                   60
Model:             SARIMAX(0, 1, 1)x(0, 1, 1, 12)   Log Likelihood                -786.225
Date:                            Wed, 03 May 2023   AIC                           1578.450
Time:                                    12:59:03   BIC                           1584.000
Sample:                                01-31-2018   HQIC                          1580.539
                                     - 12-31-2022
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.5374      0.154     -3.486      0.000      -0.840      -0.235
ma.S.L12      -0.6680      0.247     -2.704      0.007      -1.152      -0.184
sigma2      2.638e+13   4.51e-15   5.85e+27      0.000    2.64e+13    2.64e+13
===================================================================================
Ljung-Box (L1) (Q):                   0.37   Jarque-Bera (JB):                 0.21
Prob(Q):                              0.54   Prob(JB):                         0.90
Heteroskedasticity (H):               0.45   Skew:                            -0.11
Prob(H) (two-sided):                  0.12   Kurtosis:                         2.75
===================================================================================
```

PRODUCT ATTRIBUTES

We performed k-means clustering on the table spreads data. From the 6 clusters we chose the cluster that had the highest average sales



KEY FINDINGS

Competitor analysis shows that certain brands have established a loyal customer base with high conversion rates and others have high sales.

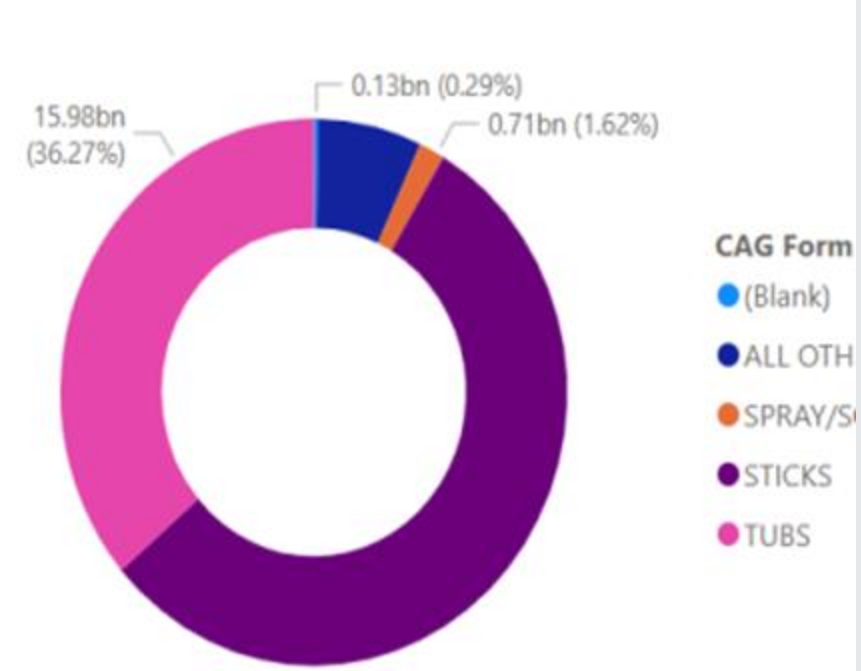
Middle-class females who are housewives or retired make up the largest buyer group for table spreads and its substitution categories like cooking sprays and cooking oils.

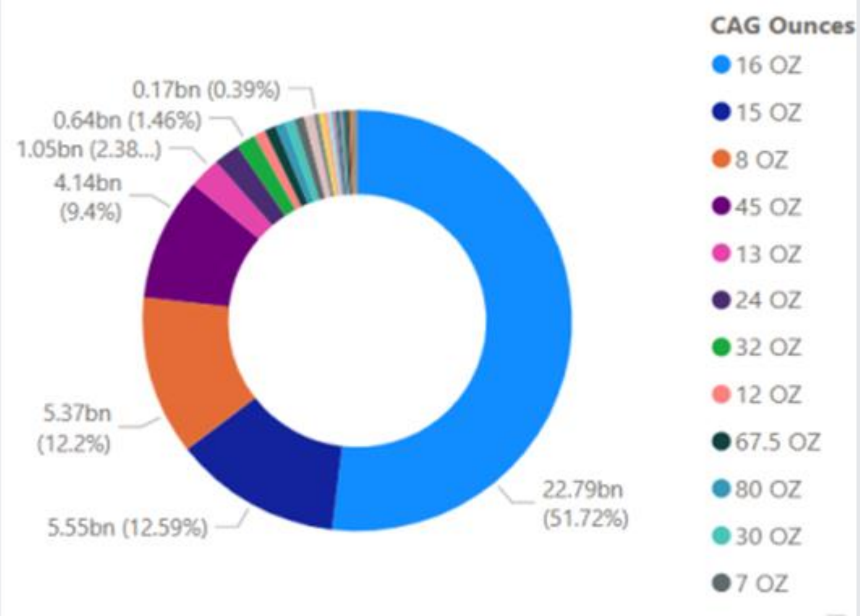
The regression analysis shows that distribution in stores with merchandising leads to higher dollar sales, while price per unit has a negative relationship with sales.

Products in Stick form, 16 ounces have the highest sales in the cluster and contribute positively to sales.

PRODUCT ATTRIBUTES

Products in Stick form, 16 ounces have the highest sales in the cluster and contribute positively to sales.





RECOMMENDATIONS

Conagra should analyze the marketing strategies, product features, and pricing strategies of successful competitor brands like Private Label, Mazola, PAM, and Crisco to understand what they are doing better to maintain brand loyalty.

Conagra should tailor their marketing and advertising strategies to target the largest buyer group for table spreads effectively i.e., Retired Middle Class Married Females

Conagra should focus on improving the distribution of table spreads in stores with merchandising to drive higher sales and increase promotional activity.

Conagra should produce and promote products in Stick form, 16 ounces, and explore the factors that make these products successful.