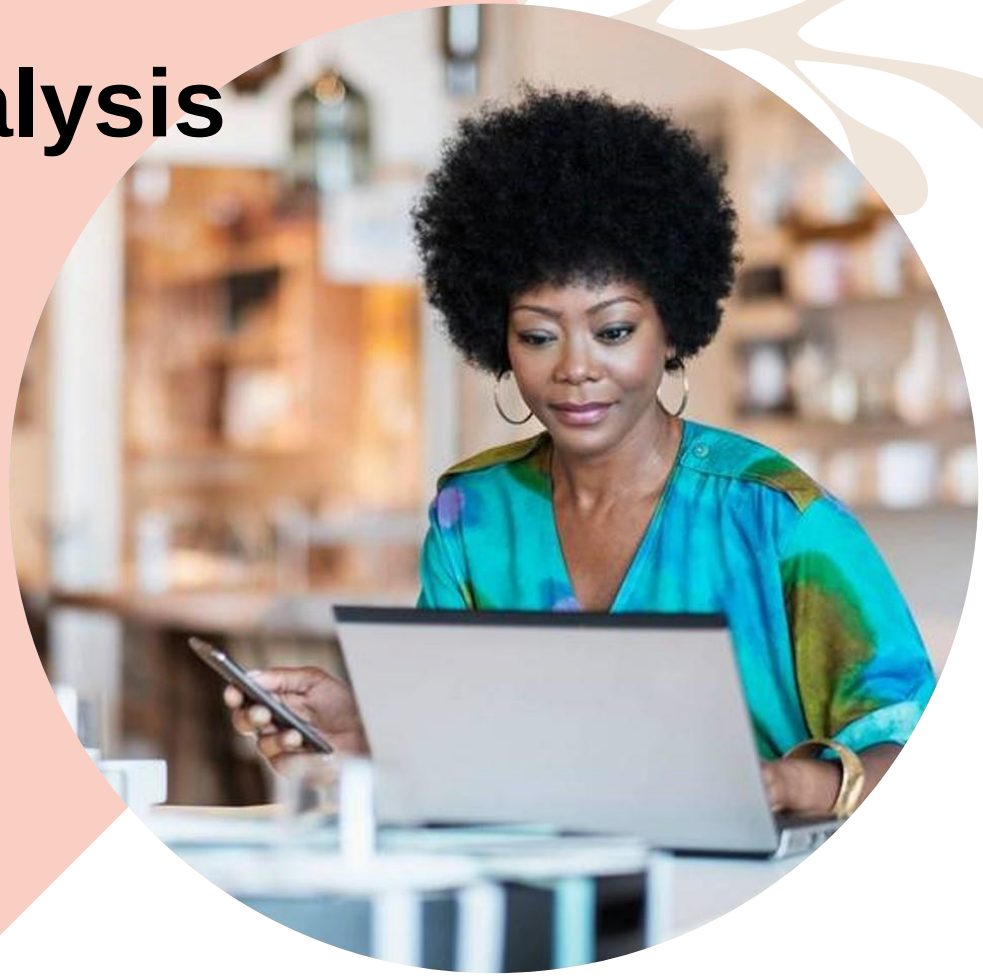
Marketing and Retail Analysis

Project 2
Market Basket Analysis



Sayyed Abdul Khaliq
Email : abdulkhaliq01112001@gmail.com

CONTENTS

- Problem Statement
- Data Overview
- Exploratory Data Analysis (EDA)
- Market Basket Analysis
- Associations Identified
- Inferences and Recommendations

Problem Statement:

A grocery store shared the transactional data with you. Your job is to conduct a thorough analysis of Point of Sale (POS) data, identify the most commonly occurring sets of items in the customer orders, and provide recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.

Overview

Objective:

- To analyze the Point of Sale (POS) data from a grocery store, we can use market basket analysis, a data mining technique that identifies the relationships between items purchased together.
- This method helps us understand the purchasing behavior of customers and identify frequent itemsets.
- Also Provide actionable recommendations for implementing popular combo offers and discounts, aiming to increase the store's revenue and enhance customer satisfaction.

Data Description

Column Name	Description
Date	This column represents the Date of product sold.
Order_id	It indicates the ID of the order
Product	This column specifies the Name of Product in the order.

Table 1 – Data Dictionary

Summary of Data

- The analysed dataset consists of 20,641 rows and 3 columns, representing the Point of Sale (POS) data from a grocery store.
- The dataset has information of all orders in a super market starting January 1st 2018 to 25th Febraury 2020.
- There are no null values in the dataset.
- There are 37 distinct products in the dataset. They are below:

['yogurt', 'pork', 'sandwich bags', 'lunch meat', 'all- purpose','flour', 'soda', 'butter', 'beef', 'aluminum foil', 'dinner rolls', 'shampoo', 'mixes', 'soap', 'laundry detergent', 'ice cream', 'toilet paper', 'hand soap', 'waffles', 'cheeses', 'milk', 'dishwashing liquid/detergent', 'individual meals', 'cereals', 'tortillas', 'spaghetti sauce', 'ketchup', 'sandwich loaves', 'poultry', 'bagels', 'eggs', 'juice', 'pasta', 'paper towels', 'coffee/tea', 'fruits', 'sugar']

Duplicates:

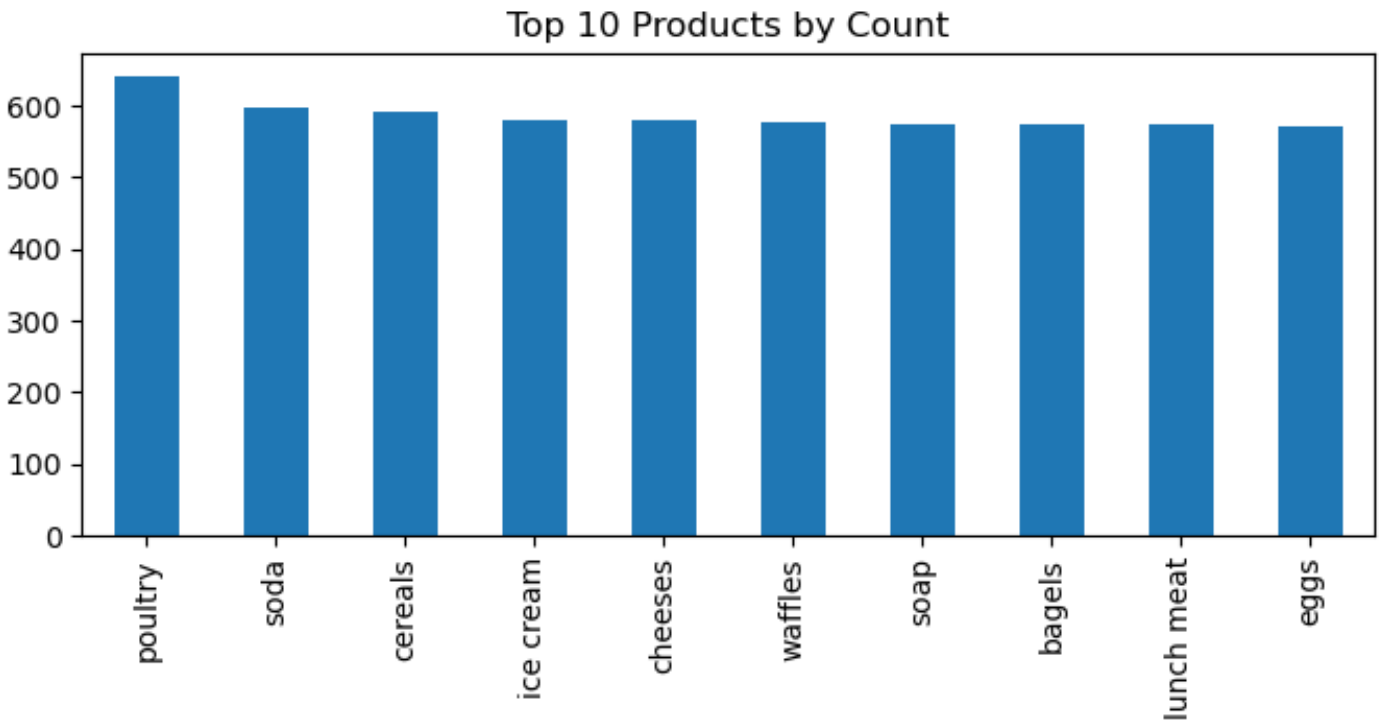
- There are 4,730 duplicate values present in the dataset.
- Although there are duplicate values in the dataset, it is not necessary to remove them.
- Each duplicate order ID may correspond to purchanse multiply quantity of the same product, indicating that these duplicates represent unique transactions.
- Therefore, the duplicate values do not need to be removed as they provide valuable insights into customer purchasing behaviour.

Data assumptions:

- The data represents a list of items purchased at a grocery store on various dates.
- Each entry in the data represents a single item purchased.
- The first column in the data represents the date the item was purchased.
- The second column represents the customer who made the purchase.
- The third column represents the item purchased.
- The same item can be purchased by multiple customers on different dates.
- There is no information provided about the quantity or price of each item.
- We have not dropped the duplicated values

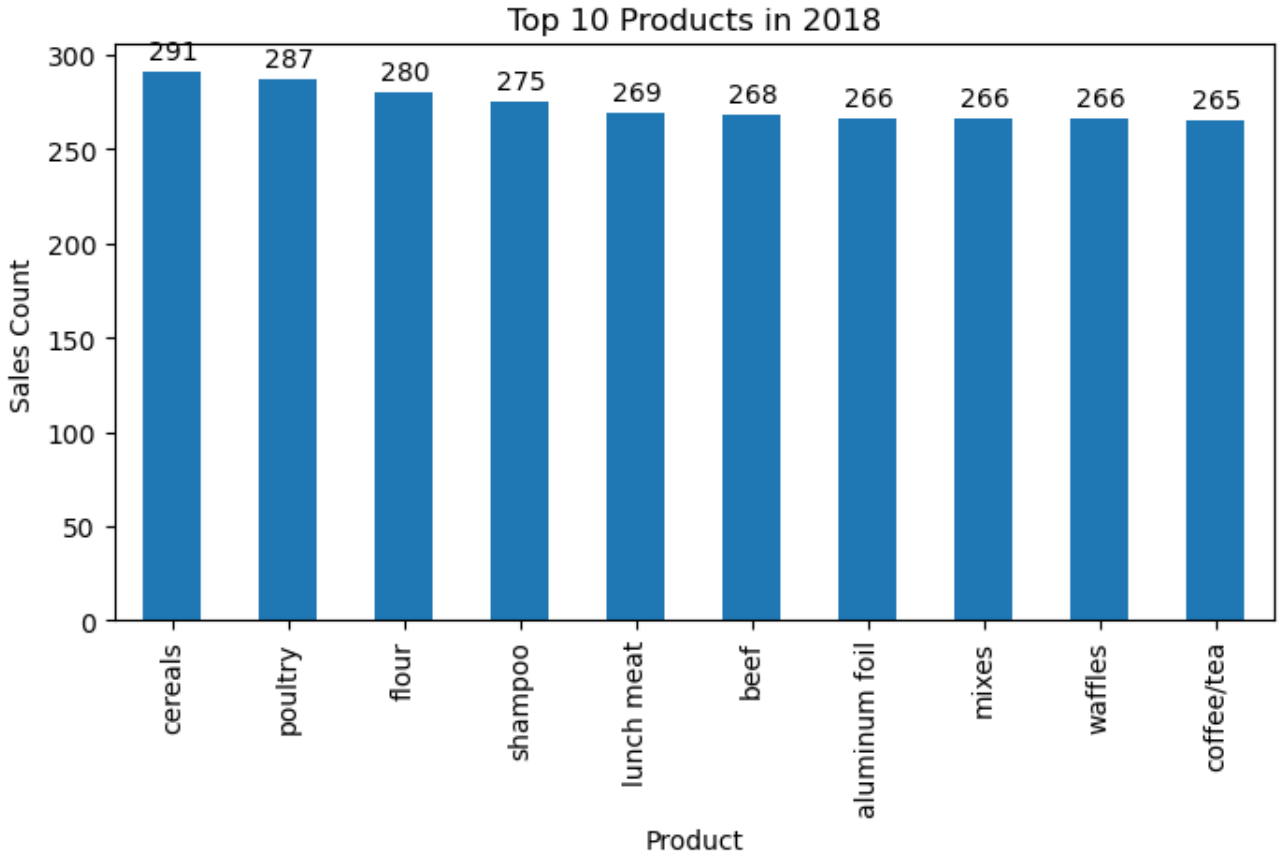
Exploratory Data Analysis

Univariate Analysis



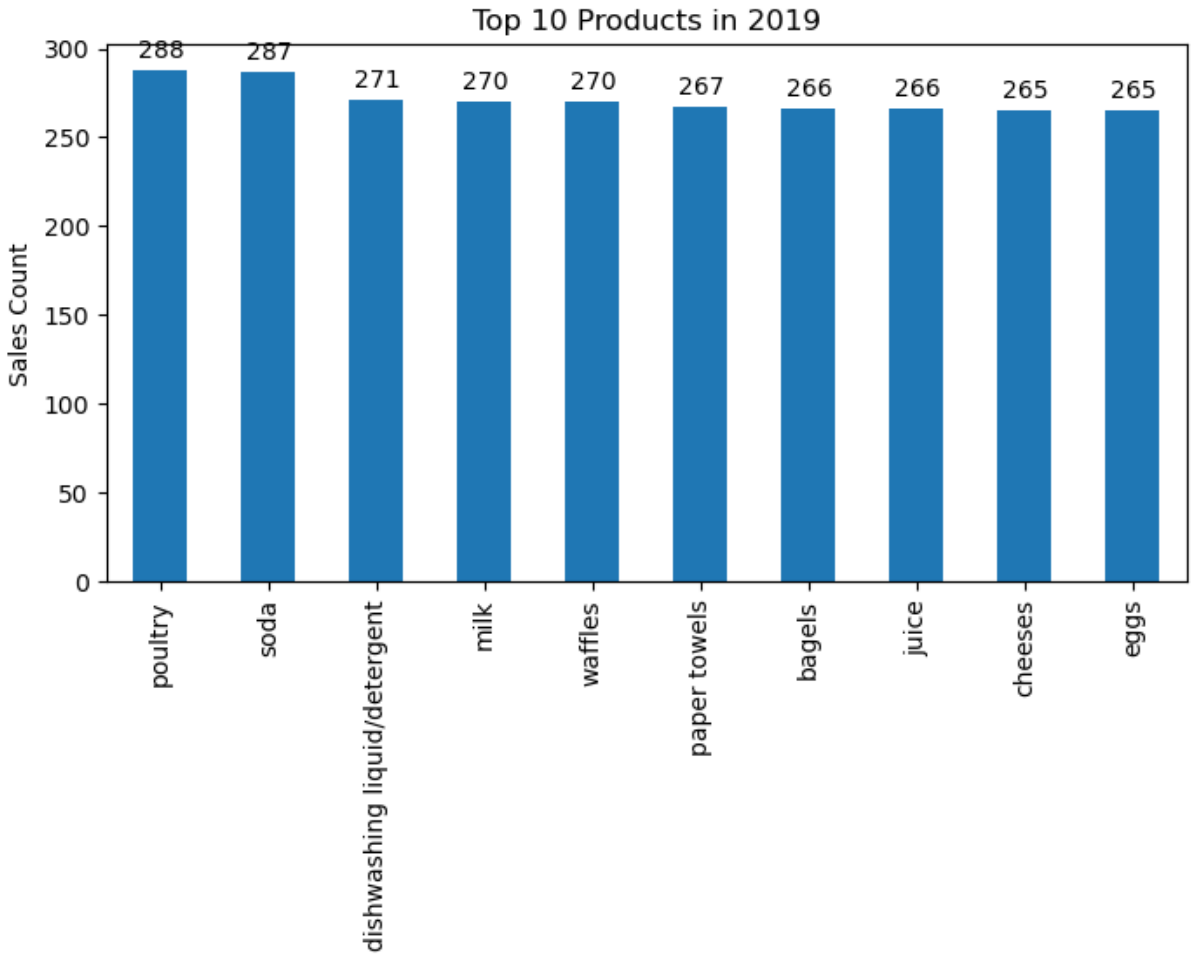
- The top 10 products purchased include poultry, soda, and cereals, indicating a diverse range of essential and indulgent items.
- Ice cream, cheeses, and waffles also feature prominently, suggesting a mix of everyday staples and treats.

Univariate Analysis



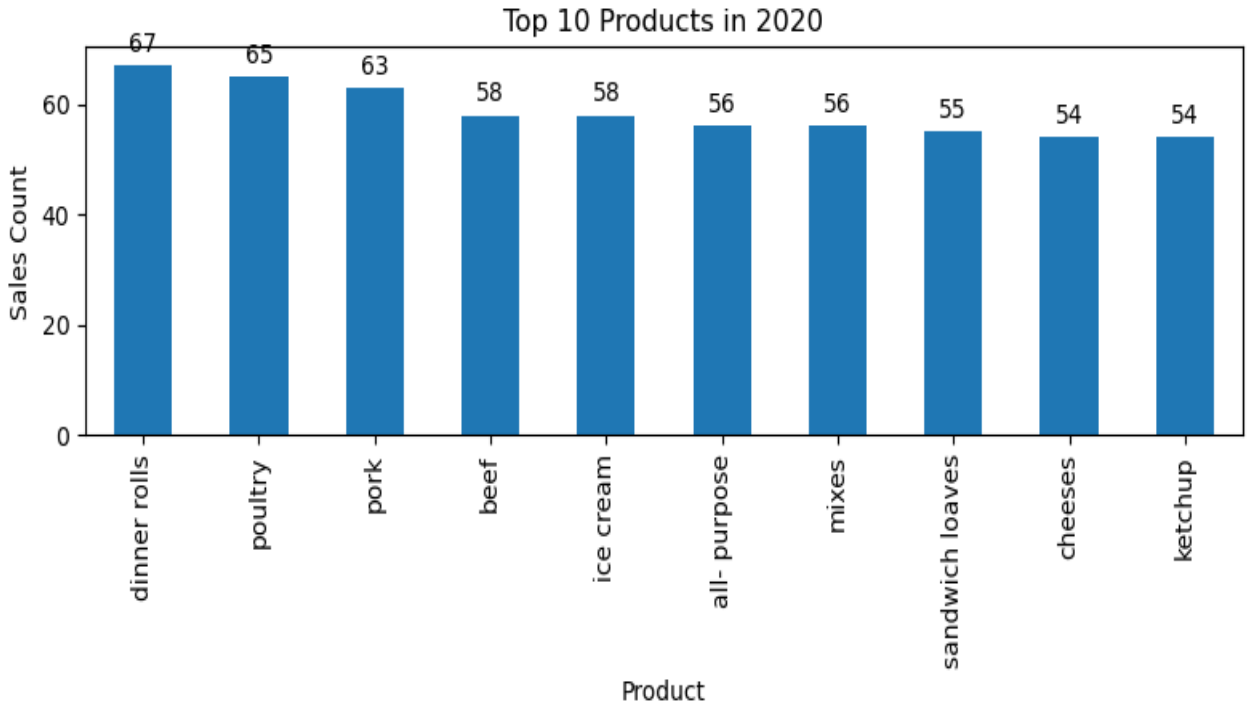
The top 10 products in 2018 are led by cereals with 291 sales, followed closely by poultry and flour.
Items like shampoo, lunch meat, and beef also feature prominently, reflecting a mix of food staples and personal care products.

Univariate Analysis



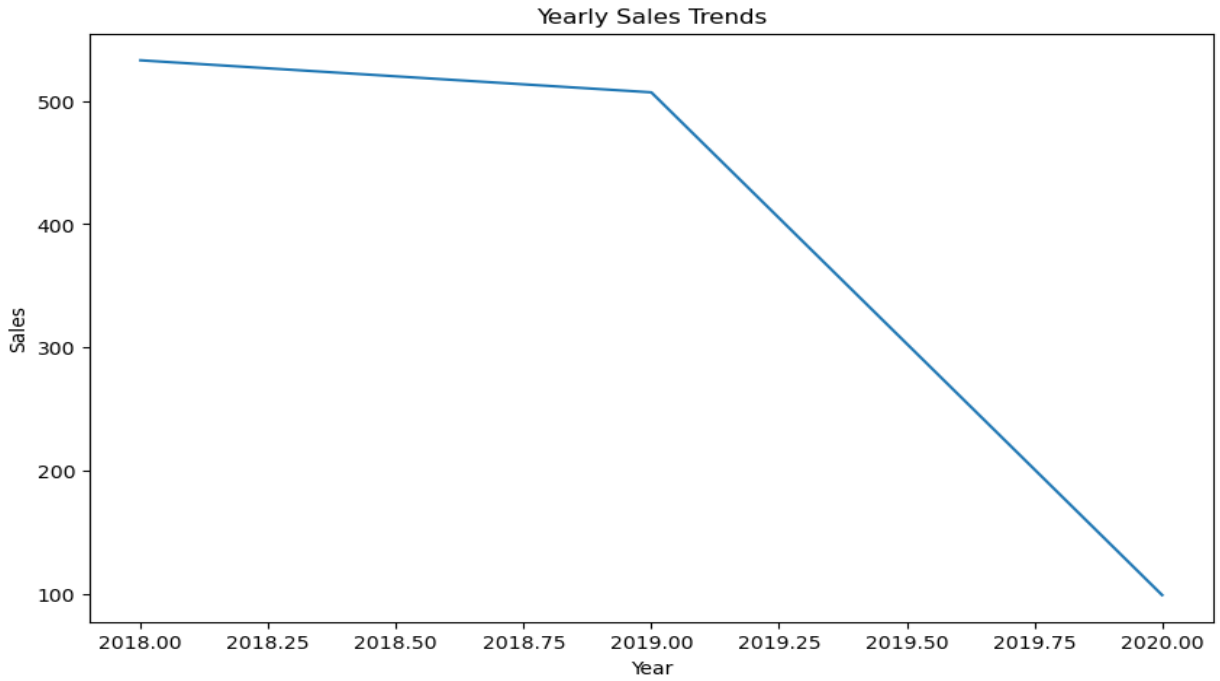
- In 2019, poultry and soda topped the sales chart, with poultry at 288 and soda at 287 sales, showing a slight shift from 2018 where cereals led.
- There was a notable shift in product preference with cereals dropping out of the top spot and poultry becoming the leading product.
- In 2019, dishwashing liquid/detergent emerged as a top-selling product with 271 sales, reflecting a significant household necessity.
- This item did not appear in the top 10 list in 2018, indicating a shift in consumer purchasing behavior or an increased focus on cleanliness.
- This change underscores the importance of offering household essentials alongside popular food items to meet diverse customer needs.

Univariate Analysis



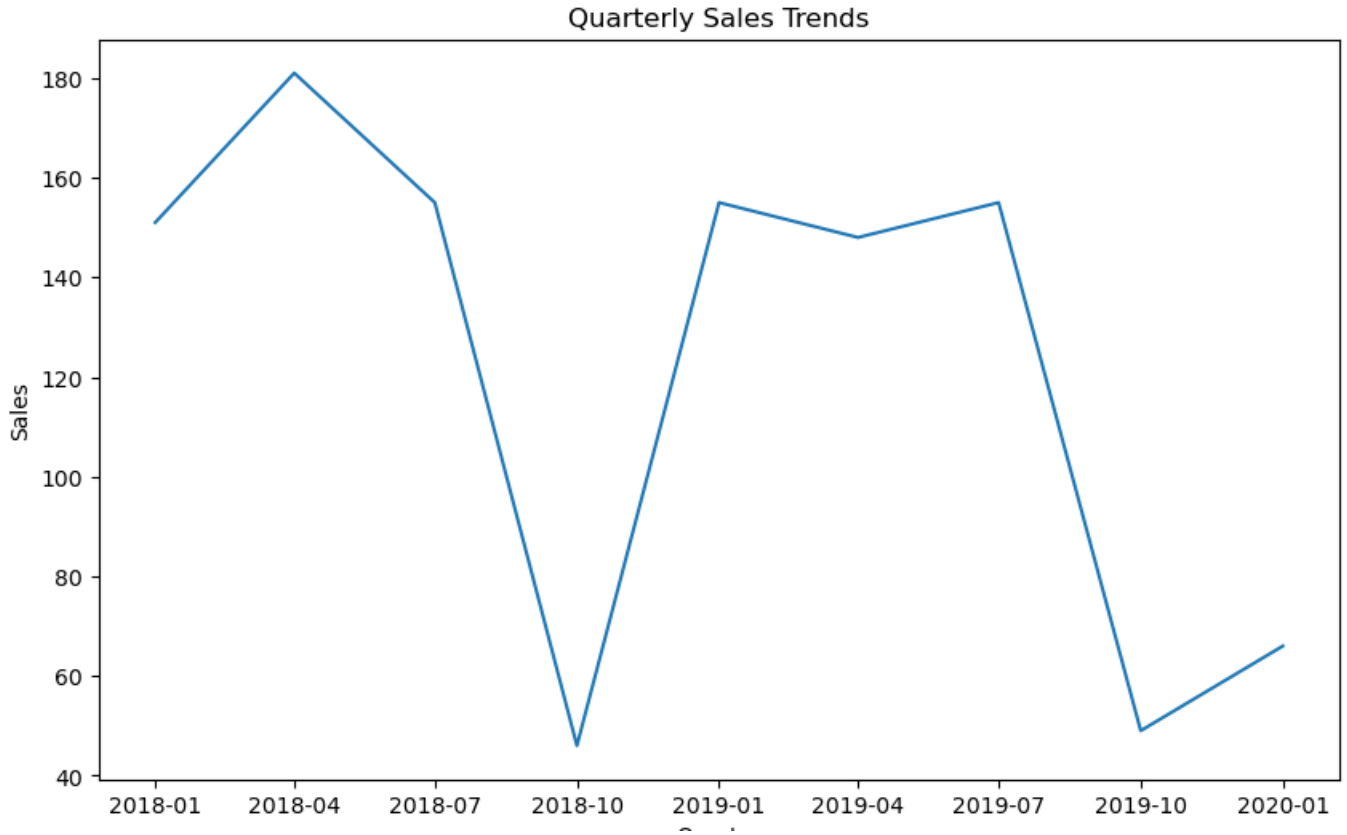
- In 2020,Dinner rolls and Pork entered into top 3, which are not in top 10 in 2018,2019. while Poultry still remains consistent.
- Also there is Significant drop in sales in 2020 because data is only available until Feb 2020.
- The sudden rise of Dinner Rolls and Pork into the top 3 in 2020 suggests a shift in consumer preferences or behavior. This could be influenced by various factors such as cultural trends, dietary habits, promotional activities, or even economic factors affecting purchasing decisions.
- Poultry's resilience in the top ranks could be attributed to its affordability, nutritional value, and versatility in cooking, making it a staple choice for many consumers across different economic conditions.

Trend Analysis - Yearly



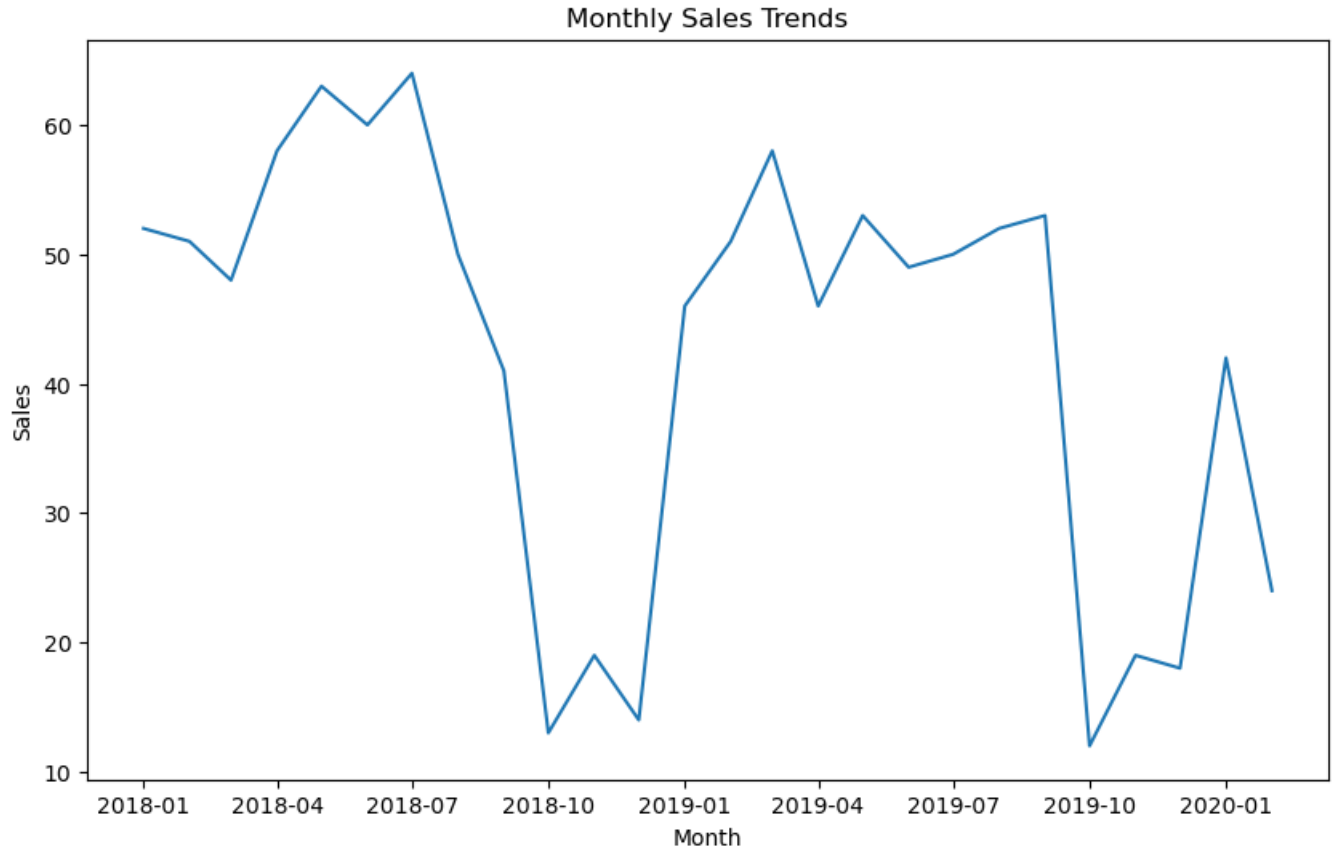
•**2018 vs. 2019:** There was a slight decrease in sales performance from 2018 to 2019. This moderate decline could be attributed to various factors such as economic fluctuations, changes in consumer preferences, or competitive pressures within the market.
•**2019 vs. 2020:** However, the trend took a drastic turn in 2020, as Data is not there after Feb 2020

Trend Analysis - Quarterly



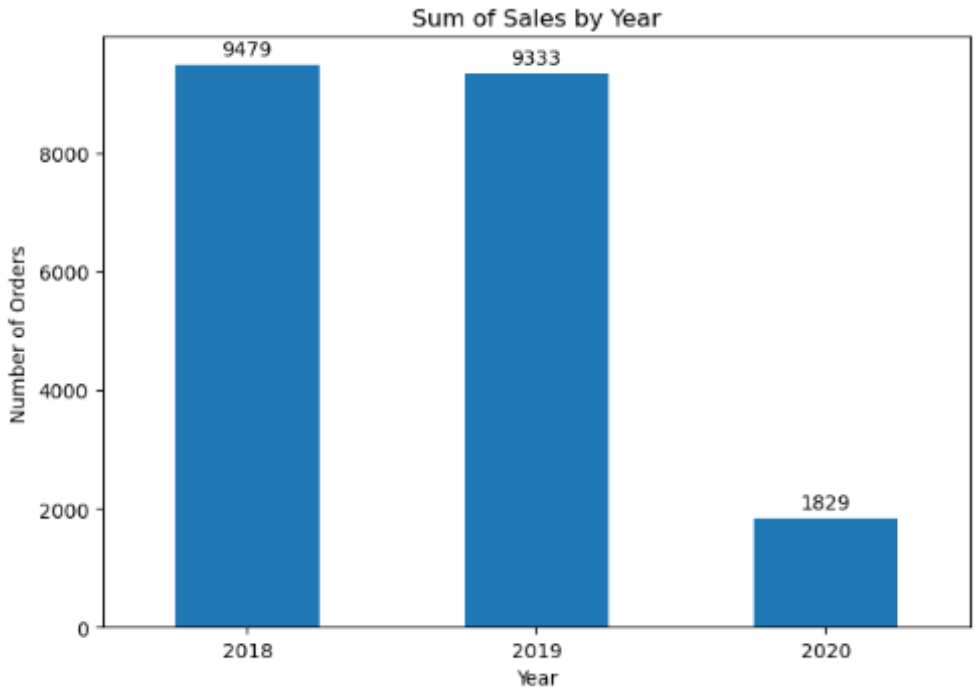
- Last quarter of the year (October to December) shows significant decline in 2018 and 2019.

Trend Analysis - Monthly



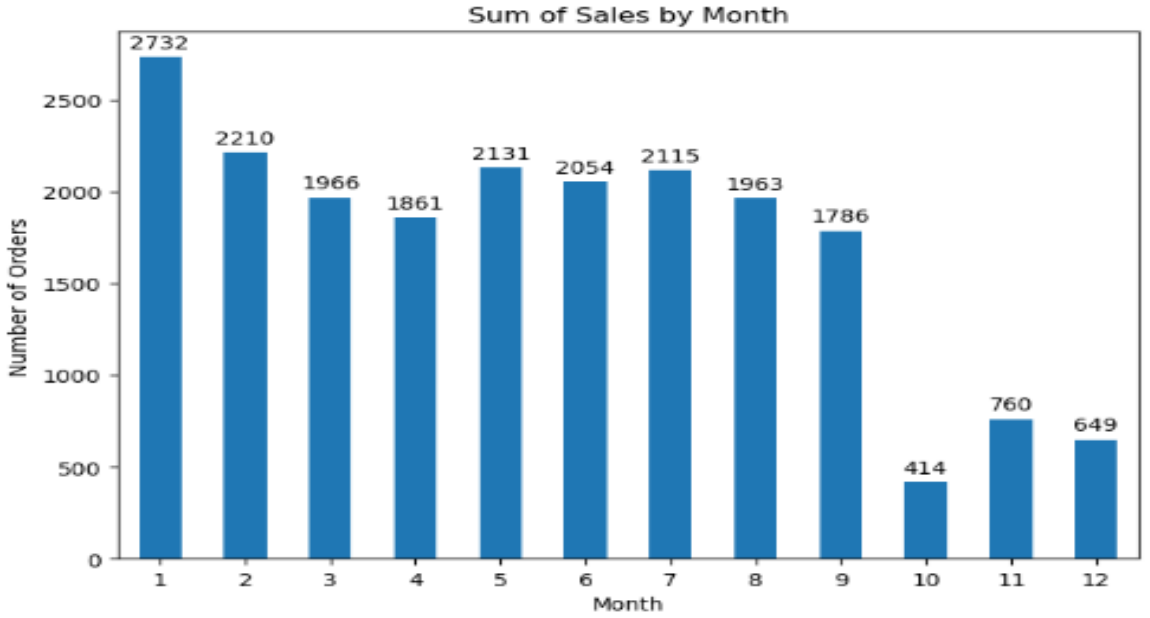
•During the months of October to December in both 2018 and 2019, there is a noticeable decline in sales. This period typically coincides with the end of the year, where consumer spending might shift towards holiday-related purchases rather than regular goods or services.
•Starting from March 2020, there is no data available

Sum of Sales by year-



•**2018 vs. 2019:** There was a slight decrease in sales performance from 2018 to 2019. This moderate decline could be attributed to various factors such as economic fluctuations, changes in consumer preferences, or competitive pressures within the market.
•**2019 vs. 2020:** 2020 saw an historic low marked by a significant decline in sales compared to 2019.

Sum of Sales by Month-

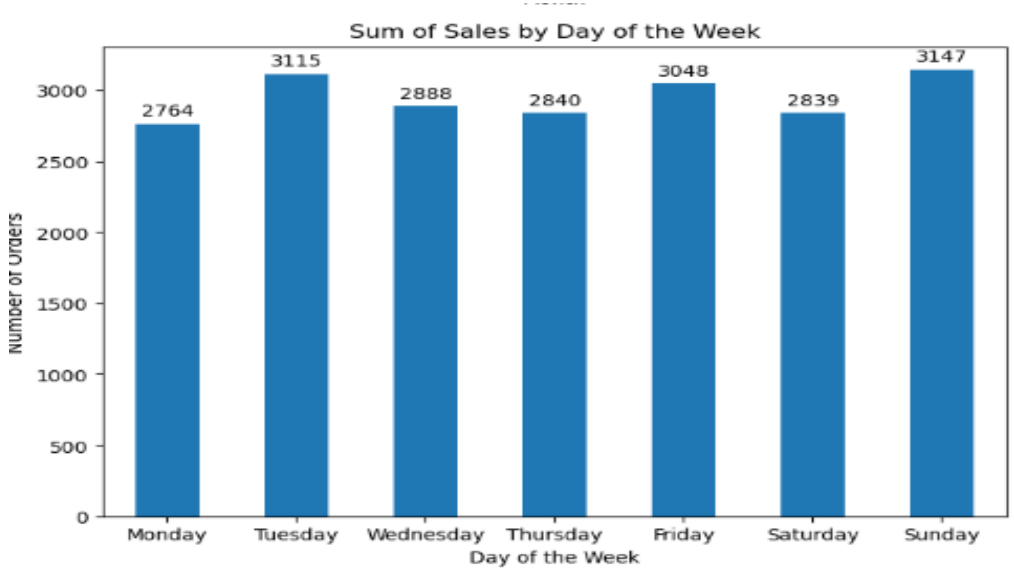


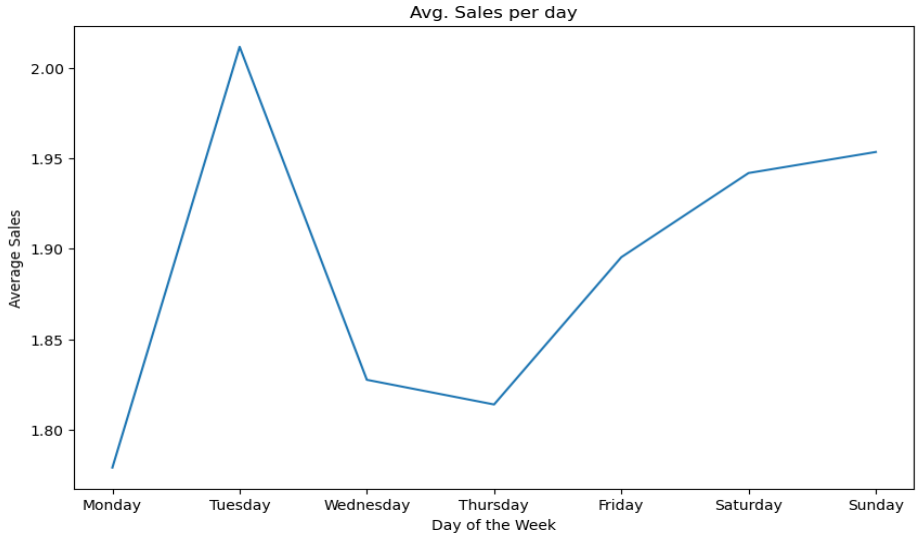
•During the months of October to December, there is a noticeable decline in sales. This period typically coincides with the end of the year, where consumer spending might shift towards holiday-related purchases rather than regular goods or services.

•Possible factors contributing to this seasonal pattern include:

- •Holiday Spending: Consumers may prioritize spending on gifts, travel, and holiday-related expenses, reducing expenditures on other products.
- •Weather Conditions: Depending on the region, colder weather or early snowfall can impact consumer behavior and reduce foot traffic or outdoor activities, affecting sales.
- •Inventory Management: Businesses may adjust their inventory levels or focus on year-end financial planning, influencing sales volumes during these months.

Sum of Sales by Day of the week-





- **Low Sales Days:** Mondays, Wednesdays, and Thursdays consistently show lower sales volumes compared to other days of the week. This pattern suggests that consumer spending or transactional activity tends to be subdued during these weekdays.
- **High Sales Days:** In contrast, Tuesdays, Fridays, and Sundays stand out with higher sales volumes. These days typically exhibit increased consumer engagement and purchasing behavior, possibly influenced by factors such as weekly routines, promotional activities, or cultural norms.

Summary of Insights:

- **Weekly Sales Patterns:** Mondays, Wednesdays, and Thursdays consistently show lower sales volumes, while Tuesdays, Fridays, and Sundays see higher sales, suggesting distinct consumer shopping behaviors across weekdays.
- **Seasonal Sales Decline:** Sales typically decline during October to December, coinciding with increased consumer spending on holiday-related purchases rather than regular goods.
- **Sales Comparison (2018 vs. 2019):** There was a slight decrease in sales from 2018 to 2019, influenced by economic fluctuations, changing consumer preferences, and competitive pressures.
- **Sales in 2020:** 2020 experienced a low sales compared to 2019, Due to unavailability of data after Feb 2020
- **Product Trends (2018 vs. 2019):** In 2019, Dinner Rolls and Pork entered the top 3 products, which were not in the top 10 in 2018, signaling shifts in consumer preferences. Poultry remained consistently popular.
- Focus on promoting and stocking up on poultry, soda, and cereals as they are consistently top-selling products.

- **Consumer Behavior Insights:** Overall, the data highlights significant changes in consumer behavior, product preferences, and sales patterns across the years, influenced by economic conditions, seasonal factors, and shifting consumer priorities.

Market Basket Analysis

What is Market Basket Analysis?

- Market Basket Analysis is a statistical technique that analyzes customer purchase patterns to identify associations between different products.
- By analyzing transactional data, Market Basket Analysis generates insights on item co-occurrence and association rules, enabling businesses to make informed decisions on product bundling, cross-selling, and targeted marketing campaigns.
- This analysis provides valuable insights into customer preferences, allowing businesses to improve customer satisfaction and drive business growth.

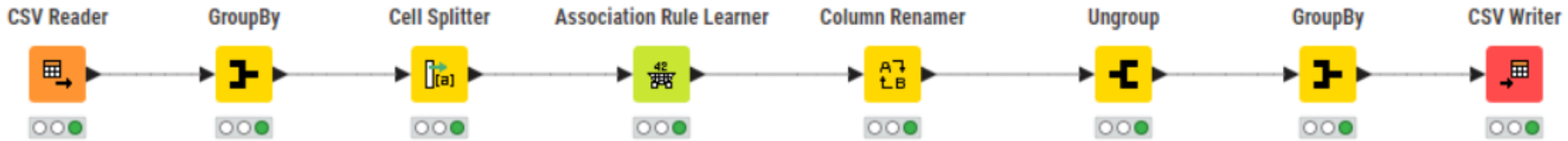
Association Rules :

- Association rules in Market Basket Analysis reveal the relationships and co-occurrence patterns between items, providing valuable insights into customer purchasing behaviour and preferences. These rules are expressed in terms of support, confidence, and lift.
    - **Support:** The proportion of transactions that contain a particular itemset.
    - **Confidence**: The likelihood that item B is purchased when item A is purchased.
    - **Lift:** The ratio of the observed support to that expected if the items were independent.

Relevance:

- The relevance of association rules lies in their ability to guide businesses in optimizing product placement, creating targeted marketing campaigns, and implementing effective cross-selling and upselling strategies to enhance customer satisfaction and increase revenue.
- Market Basket Analysis is used in a variety of industries, including retail, e-commerce, and marketing. Retailers use this technique to optimize product placement and promotions. E-commerce companies use it to personalize product recommendations, and marketers use it to develop targeted advertising campaigns.

Knime Workflow :



Tool Used

Open for Innovation
KNIME

Output Table:

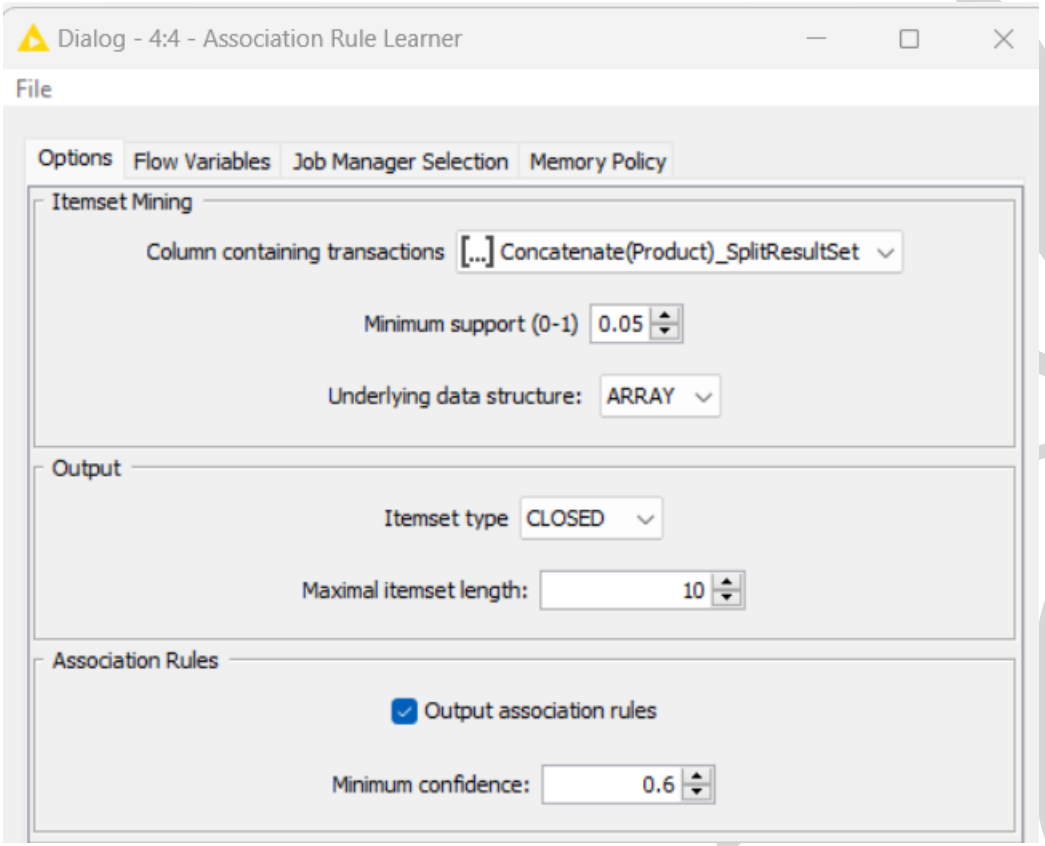
Rows: 24 | Columns: 6

#	RowID	Support *Number (double)*	Confidence *Number (double)*	Lift *Number (double)*	Recommended_Item *String*	Recommended_with *String*	Items_list` (#1) *String*
1	Row0	0.05	0.6	1.424	poultry	<---	dishwashing liquid/detergent, laundry detergent, mixes
2	Row1	0.05	0.613	1.616	coffee/tea	<---	yogurt, cheeses, cereals
3	Row2	0.05	0.62	1.645	juice	<---	yogurt, poultry, aluminum foil
4	Row3	0.05	0.64	1.7	juice	<---	yogurt, toilet paper, aluminum foil
5	Row4	0.051	0.604	1.589	milk	<---	poultry, laundry detergent, cereals
6	Row5	0.051	0.611	1.66	sandwich bags	<---	cheeses, bagels, cereals
7	Row6	0.051	0.617	1.558	cereals	<---	cheeses, bagels, sandwich bags
8	Row7	0.051	0.63	1.621	dinner rolls	<---	spaghetti sauce, poultry, cereals
9	Row8	0.051	0.63	1.678	mixes	<---	yogurt, poultry, aluminum foil
10	Row9	0.051	0.637	1.512	poultry	<---	dinner rolls, spaghetti sauce, cereals
11	Row...	0.051	0.674	1.726	cheeses	<---	bagels, cereals, sandwich bags
12	Row...	0.052	0.602	1.429	poultry	<---	dinner rolls, spaghetti sauce, juice
13	Row...	0.052	0.602	1.621	pasta	<---	paper towels, eggs, dinner rolls
14	Row...	0.052	0.628	1.61	eggs	<---	dinner rolls, poultry, soda
15	Row...	0.052	0.628	1.614	dinner rolls	<---	spaghetti sauce, poultry, juice
16	Row...	0.052	0.634	1.627	eggs	<---	paper towels, dinner rolls, pasta
17	Row...	0.052	0.641	1.649	dinner rolls	<---	spaghetti sauce, poultry, ice cream
18	Row...	0.052	0.686	1.628	poultry	<---	dinner rolls, spaghetti sauce, ice cream
19	Row...	0.054	0.642	1.651	dinner rolls	<---	spaghetti sauce, poultry, laundry detergent
20	Row...	0.054	0.656	1.556	poultry	<---	dinner rolls, spaghetti sauce, laundry detergent
21	Row...	0.055	0.624	1.565	ice cream	<---	paper towels, eggs, pasta
22	Row...	0.055	0.63	1.616	eggs	<---	paper towels, ice cream, pasta
23	Row...	0.055	0.643	1.731	pasta	<---	paper towels, eggs, ice cream
24	Row...	0.055	0.649	1.791	paper towels	<---	eggs, ice cream, pasta

Table 3– Output table

Threshold values of Support and Confidence

- Threshold value for Minimum Support is 0.05
- Threshold value for Minimum Confidence is 0.6
- Maximum Item Set Length : 10
- In this analysis, we have defined the threshold values for Support and Confidence as 0.05 and 0.6, respectively.
- These values help us determine which association rules are significant and trustworthy.
- By using these thresholds, we can filter out less important rules and focus on the ones that have strong support and confidence levels, ensuring that our analysis provides meaningful insights for decision-making.



Association Identified

Rows: 24 | Columns: 6

#	RowID	Support *Number (doub...*	Confidence *Number (double)*	Lift *Number (double)*	Consequent *String*	implies *String*	Items *Set*
1	rule0	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum foil]
2	rule1	0.05	0.62	1.645	juice	<---	[yogurt,poultry,aluminum foil]
3	rule2	0.05	0.613	1.616	coffee/tea	<---	[yogurt,cheeses,cereals]
4	rule3	0.05	0.6	1.424	poultry	<---	[dishwashing liquid/detergent,laundry deter...
5	rule4	0.051	0.63	1.678	mixes	<---	[yogurt,poultry,aluminum foil]
6	rule5	0.051	0.611	1.66	sandwich bags	<---	[cheeses,bagels,cereals]
7	rule6	0.051	0.674	1.726	cheeses	<---	[bagels,cereals,sandwich bags]
8	rule7	0.051	0.617	1.558	cereals	<---	[cheeses,bagels,sandwich bags]
9	rule8	0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce,poultry,cereals]
10	rule9	0.051	0.637	1.512	poultry	<---	[dinner rolls,spaghetti sauce,cereals]
11	rule10	0.051	0.604	1.589	milk	<---	[poultry,laundry detergent,cereals]
12	rule11	0.052	0.628	1.61	eggs	<---	[dinner rolls,poultry,soda]
13	rule12	0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce,poultry,ice cream]
14	rule13	0.052	0.686	1.628	poultry	<---	[dinner rolls,spaghetti sauce,ice cream]
15	rule14	0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce,poultry,juice]
16	rule15	0.052	0.602	1.429	poultry	<---	[dinner rolls,spaghetti sauce,juice]
17	rule16	0.052	0.634	1.627	eggs	<---	[paper towels,dinner rolls,pasta]
18	rule17	0.052	0.602	1.621	pasta	<---	[paper towels,eggs,dinner rolls]
19	rule18	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce,poultry,laundry detergent]
20	rule19	0.054	0.656	1.556	poultry	<---	[dinner rolls,spaghetti sauce,laundry deterge...
21	rule20	0.055	0.624	1.565	ice cream	<---	[paper towels,eggs,pasta]
22	rule21	0.055	0.63	1.616	eggs	<---	[paper towels,ice cream,pasta]
23	rule22	0.055	0.643	1.731	pasta	<---	[paper towels,eggs,ice cream]
24	rule23	0.055	0.649	1.791	paper towels	<---	[eggs,ice cream,pasta]

- Association rules are based on the concept of frequent itemsets, which are sets of items that appear together frequently in a transactional dataset.
- 24 rules have been found with the dataset and set parameters.

Metrics Used:

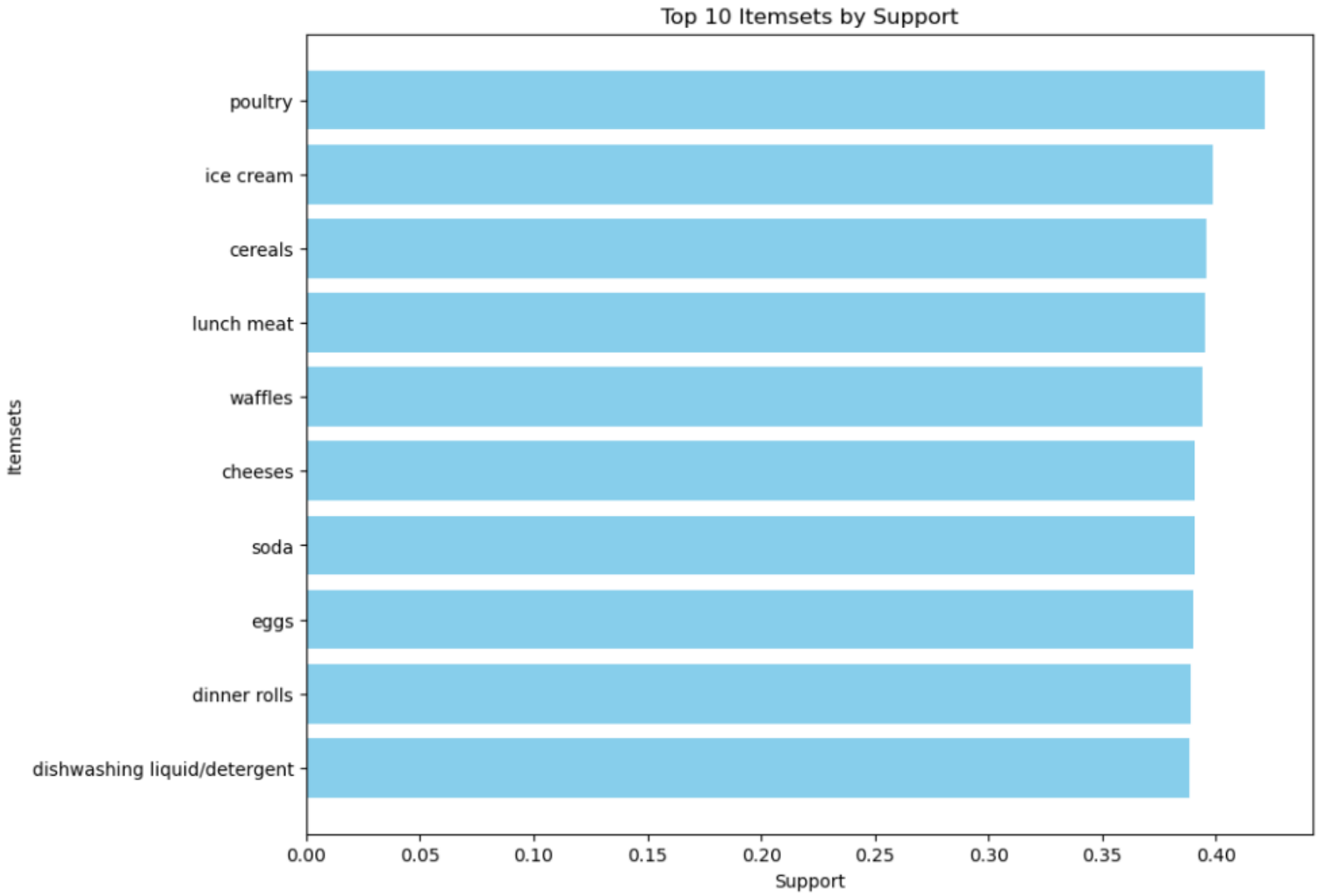
- **Support:** It is the probability of observing the items together in a transaction. It is calculated as the number of transactions that contain both items divided by the total number of transactions. It measures how frequent the itemset occurs in the dataset. High support indicates that the itemset is popular and should be considered for promotion or placement together.

- **Confidence:** It is the conditional probability that a transaction containing one item also contains another item. It is calculated as the number of transactions containing both items divided by the number of transactions containing the first item. It measures the strength of the association between two items. High confidence indicates that the items are likely to be bought together, and can be used to recommend or suggest items to customers.

- **Lift:** It is the measure of how much more often two items occur together than expected if they were independent of each other. It is calculated as the support of the itemset divided by the product of the individual supports of the items. A lift value of 1 indicates that the items are independent, while a value greater than 1 indicates a positive association between the items. A lift value less than 1 indicates a negative association between the items. High lift indicates that the items have a strong association and can be used for cross-selling or bundling.

INFERENCES AND RECOMMENDATION

Top 10 Items by support:



•Poultry, ice cream, and cereal stand out as the top-ranked items based on their high levels of support, indicating their popularity and frequent occurrence together in customer transactions.

Recommendations:

- **Bundle Promotion**: Create a special offer on juice, yogurt, poultry, and aluminum foil, encouraging customers to buy these items together at a reduced price. This strategy aligns with their common occurrence in shopping baskets, promoting a complete meal solution or household essentials bundle

- **Combo Deal:** Offer a bundled discount on poultry, including dishwashing liquid/detergent, laundry detergent, and mixes. This promotion capitalizes on the strong association between poultry and these household essentials, encouraging customers to stock up on related items in one purchase.

- **Buy Two Get One Free:** Introduce a promotion where customers purchasing coffee or tea can get yogurt, cheeses, or cereals at a discounted rate or receive one of these items for free. This leverages the frequent co-purchase patterns observed between coffee/tea and these breakfast essentials.

- **Cross-Promotion Strategy**: Implement a cross-promotion where customers buying sandwich bags can receive a discount on cheeses, bagels, and cereals. This encourages customers to explore related items and increases basket size based on their common purchase patterns.

- **Limited Time Offer:** Roll out a time-limited offer for dinner rolls, Spaghetti sauce offering them together at a discounted rate. This targets high-lift item combinations, encouraging immediate purchases while boosting overall basket value.

- These discount offers and combos can help increase sales by providing customers with more value for their money and encouraging them to purchase more items. It is important to promote these offers through in-store signage, advertisements, and social media to ensure customers are aware of the deals available.

THANK YOU

Sayyed Abdul Khaliq

abdulkhaliq01112001@gmail.com