Grocery Store’s - MRA Project

BY
SONA G

Contents

Executive Summary of the data
- Problem statement
- Executive Summary & Data Dictionary

Exploratory Analysis & Insights
- Univariate, Bivariate, Multivariate & other analyses
- Weekly, Monthly, Quarterly, and Yearly Weekday Trends in Sales count
- Products counts & Year Wise top products
- Summary and Recommendations

Market Basket Analysis
- Market Basket Analysis
- MRA KNIME Workflow & Output Table

Associations Identified
- MRA – values
- Association Rule Parameters
- Association Rules Table
- Table View

Inference & Recommendation
- Key insights from the analysis
- Business Recommendation

1. Executive Summary of the data

- Problem statement
- Executive Summary & Data Dictionary

Problem Statement

Business Context:

In the highly competitive grocery retail industry, understanding customer buying patterns is crucial for enhancing sales, increasing customer satisfaction, and improving profitability. By identifying frequently purchased item combinations, grocery stores can craft effective marketing strategies, optimize inventory management, and tailor promotions to meet customer needs. Leveraging Point of Sale (POS) data can unlock valuable insights that drive customer-centric offerings, such as combo packs, discounts, and targeted promotions, which can increase basket size and improve customer retention. This analysis aligns with business goals by maximizing revenue, reducing operational costs, and boosting customer loyalty.

Objective:

As a business analyst, the goal is to analyze the POS transactional data to identify frequently purchased item combinations. Using association rule mining or similar techniques, the aim is to uncover patterns that will help the store create targeted combo offers and discounts, ultimately driving revenue growth by increasing customer purchases and average basket size.

Executive Summary & Data Dictionary

Data: from 01-01-2018 to 26-02-2020

Dataset: 20641 Rows, 3 columns

Data Dictionary: Date, Order_id, Product

Missing values : 0

Duplicate values: 4730

Dropping duplicate rows is typically a good practice to avoid skewed analysis, but in this case, it may not be appropriate due to the lack of a unique identifier. Since multiple customers can purchase the same product on the same date, removing duplicates could eliminate valid data. Therefore, duplicates are retained in the dataset.

Assumptions: The data consists of a record of items bought at a grocery store on different dates.

Each entry corresponds to a single item purchased. The first column indicates the purchase date, the second column identifies the customer, and the third column lists the item purchased. Different customers may buy the same item on various dates. There is no information available regarding the quantity or price of the items. **Duplicate values have been retained** in the dataset.

2. Exploratory Data Analysis

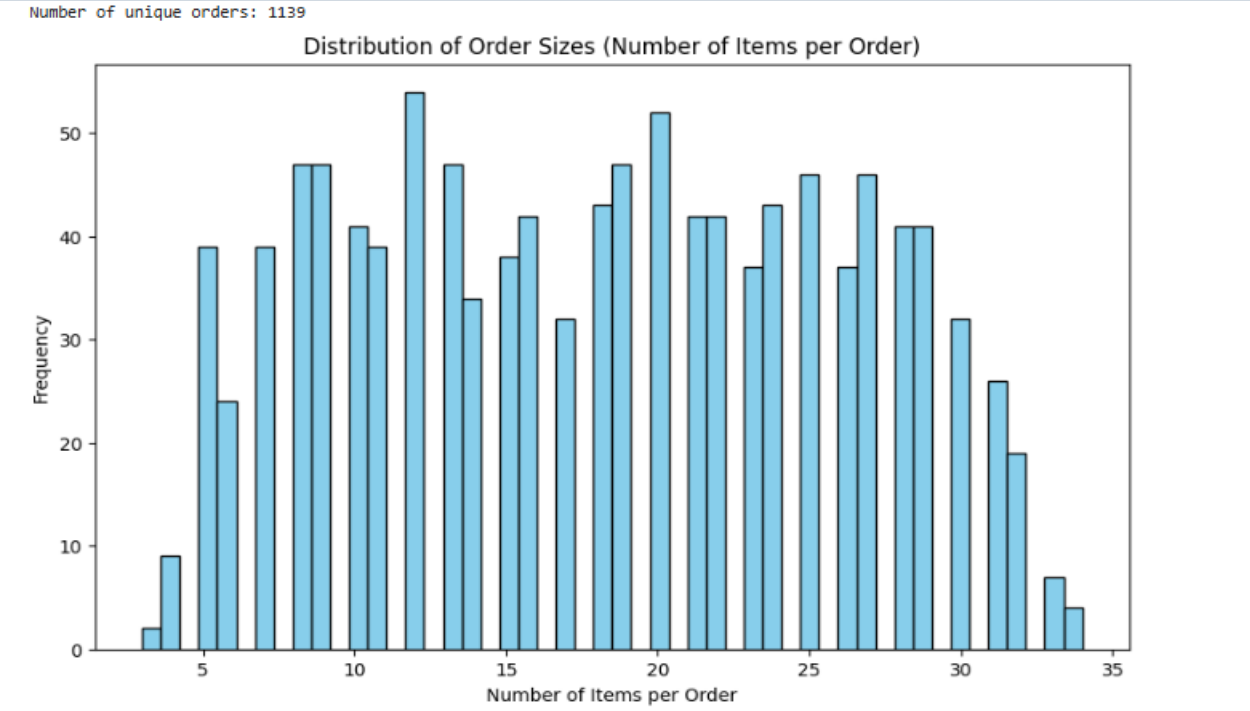
- Univariate, Bivariate, Multivariate & other analyses
- Weekly, Monthly, Quarterly, and Yearly Trends in Sales count
- Products counts & Year Wise top products
- Summary and Recommendations

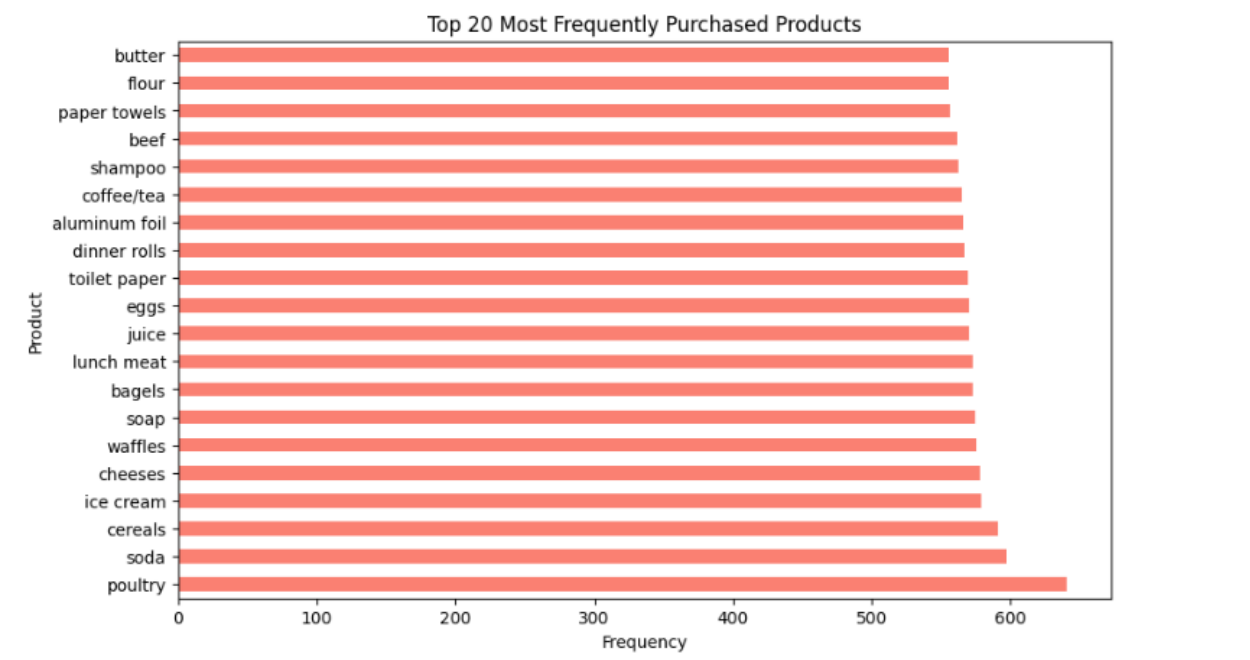
Univariate Analysis

- **Peak Sales in 2018 & 2019** – Strong transaction volumes, likely due to promotions or seasonal trends.
- **Sharp Decline After Late 2019** – Significant drop, possibly due to external factors like COVID-19.
- **Minimal Transactions in 2020** – Business impact, potential store closure or shift to online shopping.

This histogram visualizes the **distribution of order sizes** (number of items per order) for **1,139 unique orders**.

- **Most orders contain between 5 to 30 items.**
- **Peaks observed around 10, 15, 20, and 25 items per order.**
- **Few orders have very low (<5) or very high (>30) items.**

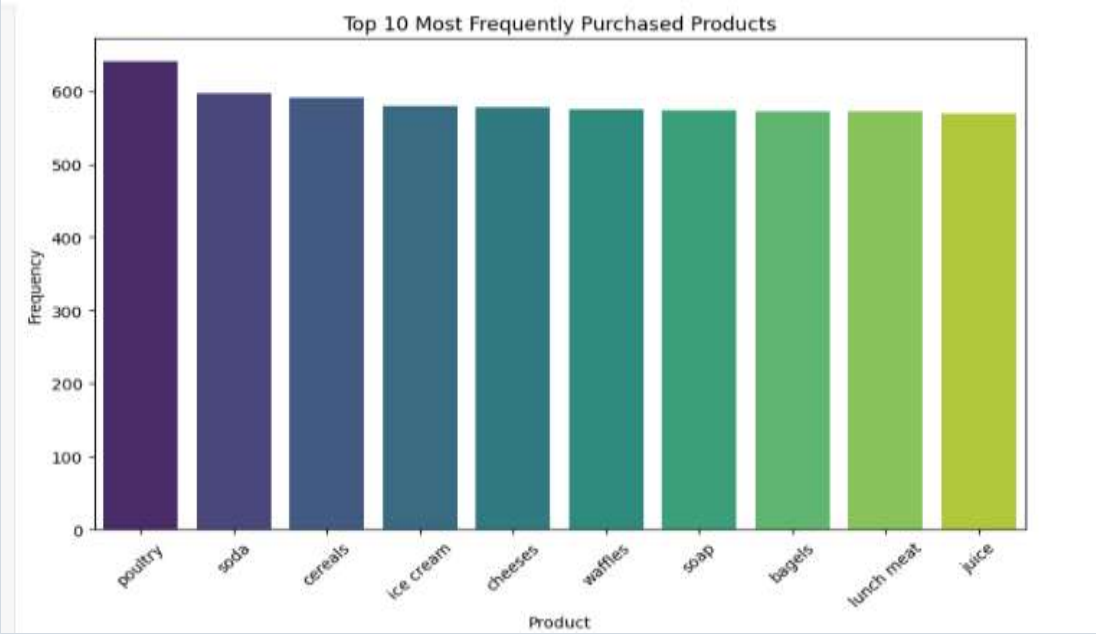




Bivariate Analysis

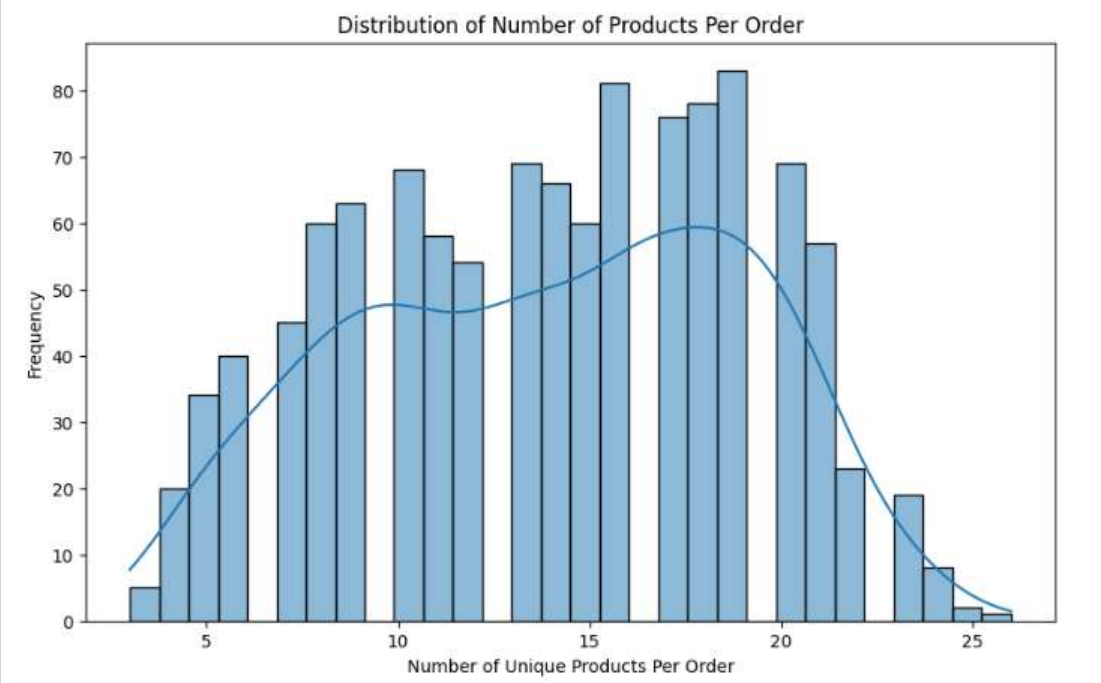
This bar chart displays the **top 20 most frequently purchased products**.

- Poultry**, soda, and cereals** are the most purchased items.
- **Essential groceries** like eggs, flour, butter, and toilet paper are among the top.
- **Non-food items** (shampoo, aluminum foil, paper towels) also have high demand.



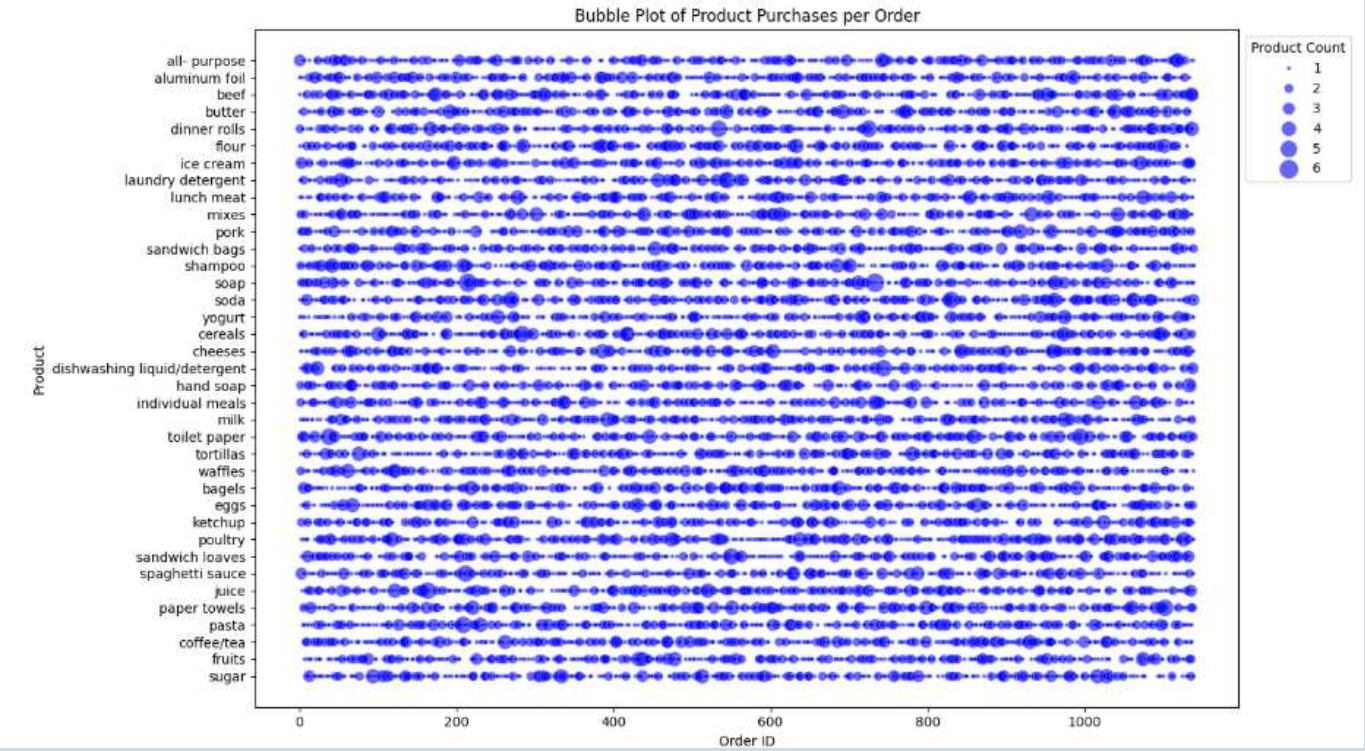
This histogram represents the **distribution of the number of unique products per order**.

- Most orders contain around **10-20 unique products**.
- The distribution is **right-skewed**, meaning a few orders contain a very high number of unique products.
- There is a **peak around 15 products per order**, indicating that many customers tend to purchase this quantity.



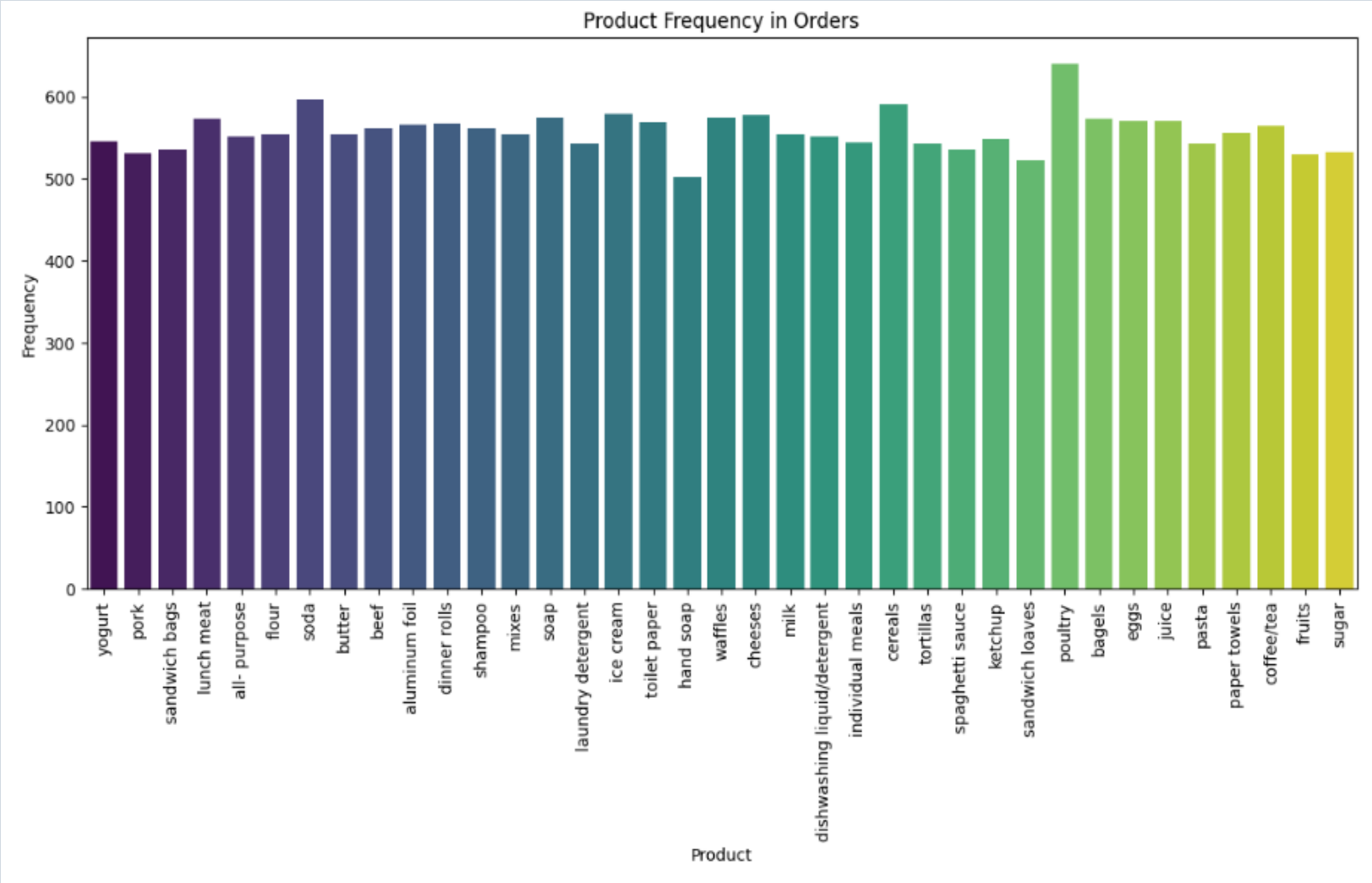
Multivariate Analysis

- The distribution is **dense and varied**, showing that customers order a diverse range of products.
- Larger bubbles indicate products that are frequently bought in **higher quantities**.
- Some products might be consistently **purchased in small quantities**, while others have a wider variation.



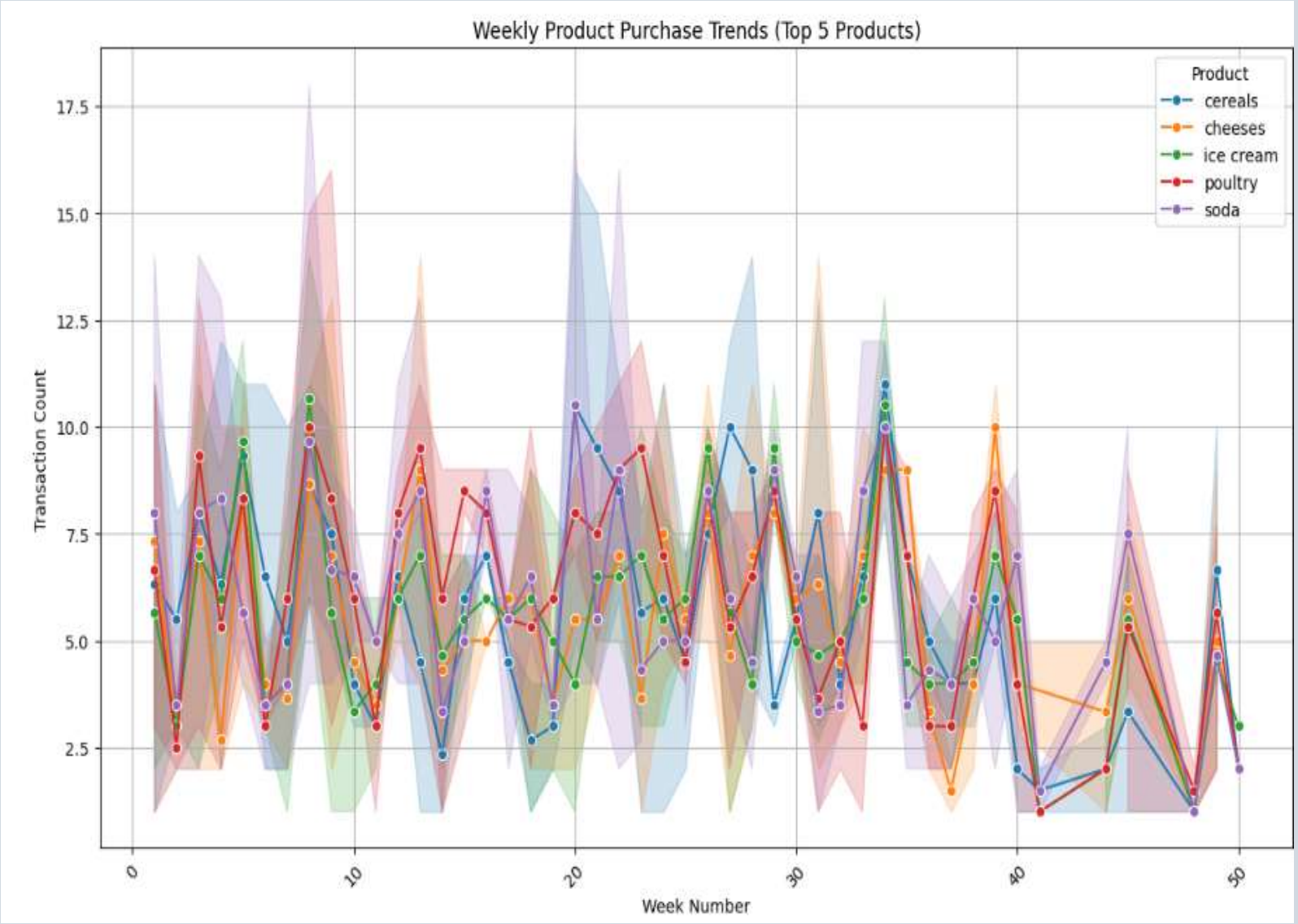
Product Frequency

- The chart shows the frequency of each product appearing in orders.
- The distribution is fairly even, but some products (e.g., cereals, poultry, and dishwashing liquid detergent) have a higher frequency.
- This suggests that these products are more commonly purchased by customers.
- Products with lower frequencies might indicate niche or less popular items.



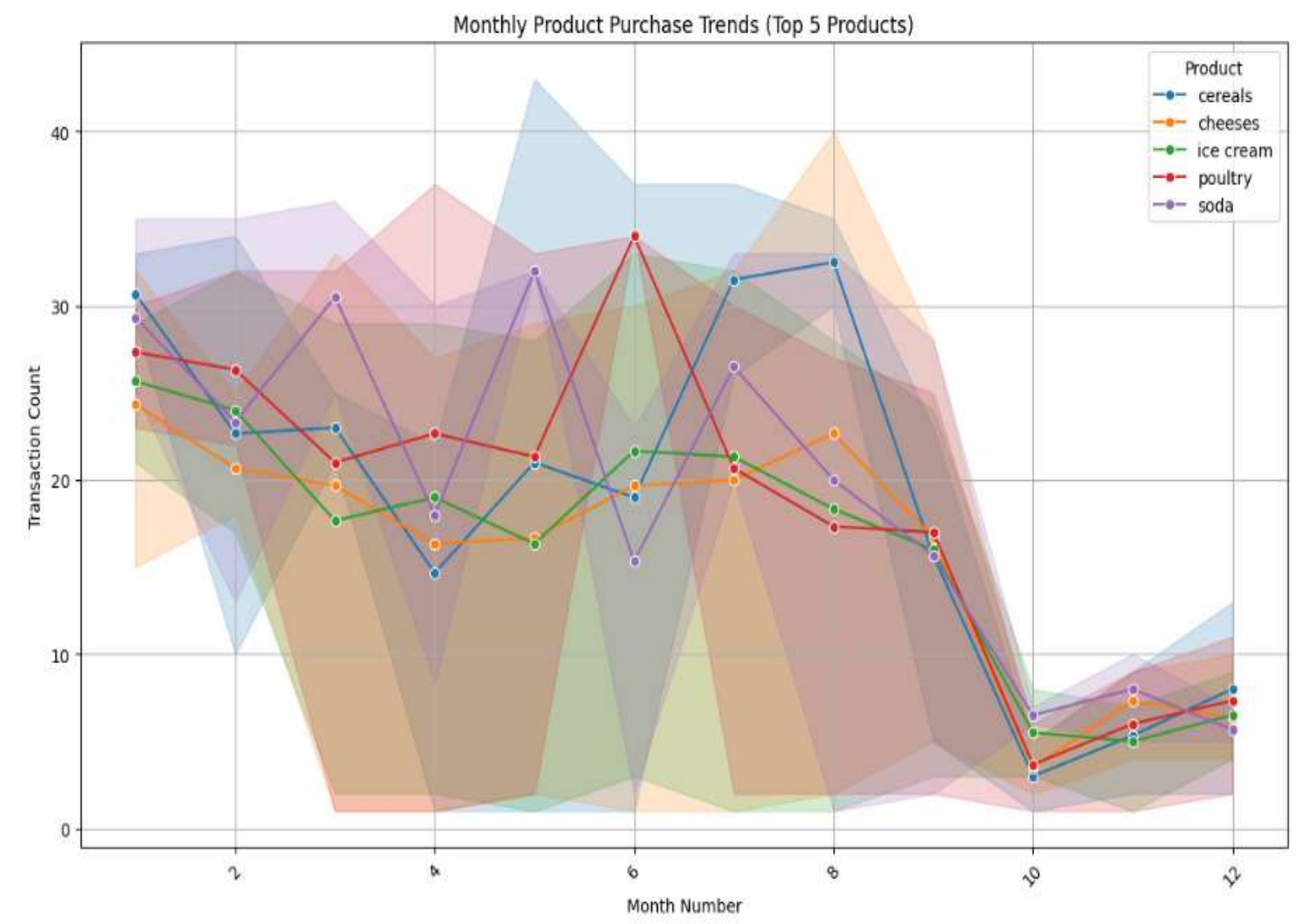
Weekly Trend

- The chart tracks weekly purchase trends for the top 5 products: cereals, cheeses, ice cream, poultry, and soda.
- The transaction counts fluctuate significantly week by week, indicating a seasonal or periodic demand for these products.
- Peaks and troughs suggest promotional periods, seasonal demand changes, or external factors affecting sales. The shaded areas indicate variance, meaning that in some weeks, there was high unpredictability in the demand.
- The trend for all products follows a similar cyclical pattern, possibly influenced by consumer behavior (e.g., increased purchases at the beginning or end of a month).



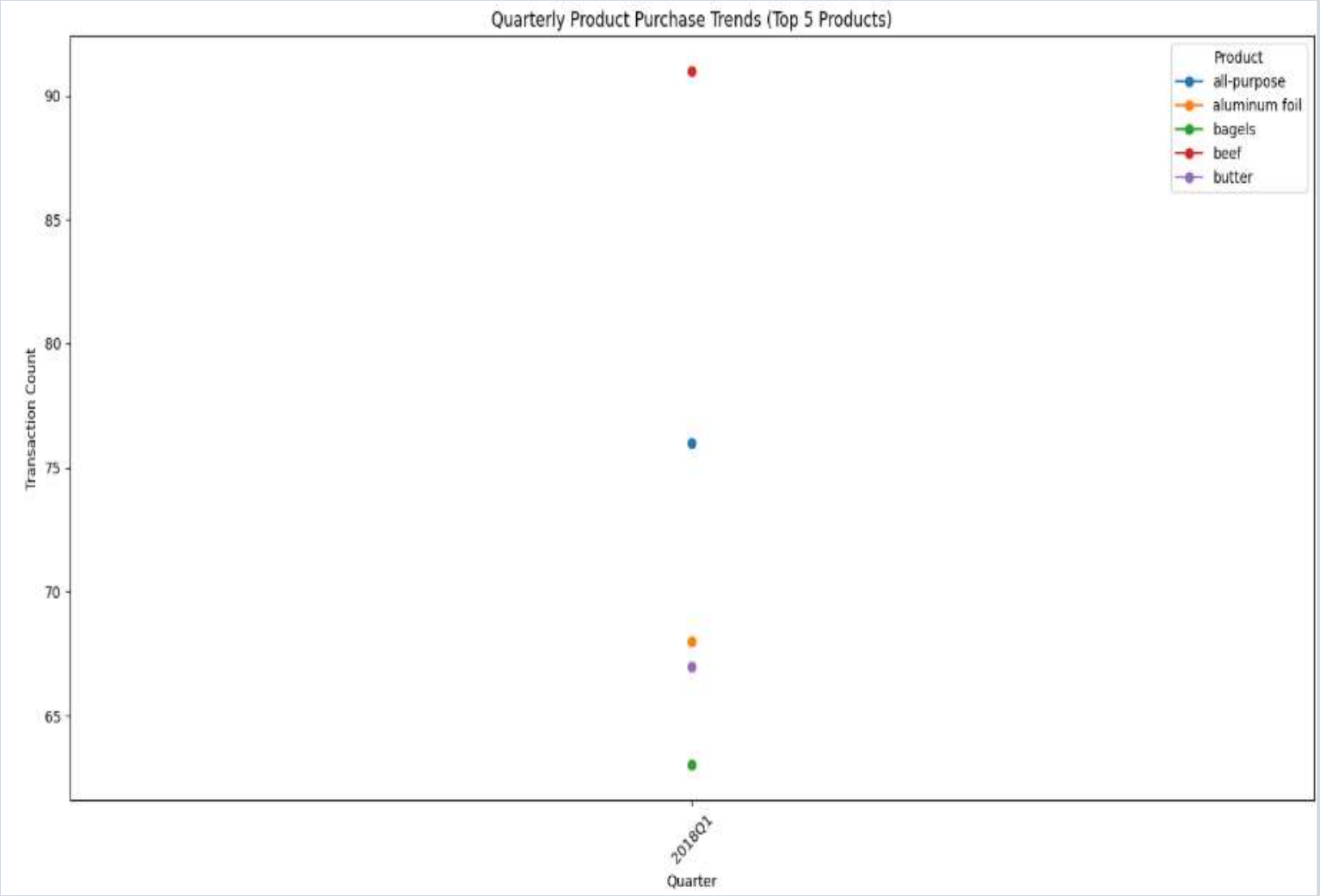
Monthly Trend

- **Fluctuating Sales** – Transaction counts vary across months, with notable peaks and dips.
- **Year-End Decline** – Sharp drop in sales from months 10-12, possibly due to seasonality.
- **Top Performers** – Poultry and soda peak around months 3 & 6, while cereals remain stable.
- **Business Impact** – Plan inventory, adjust marketing, and analyze external factors for better sales strategies.



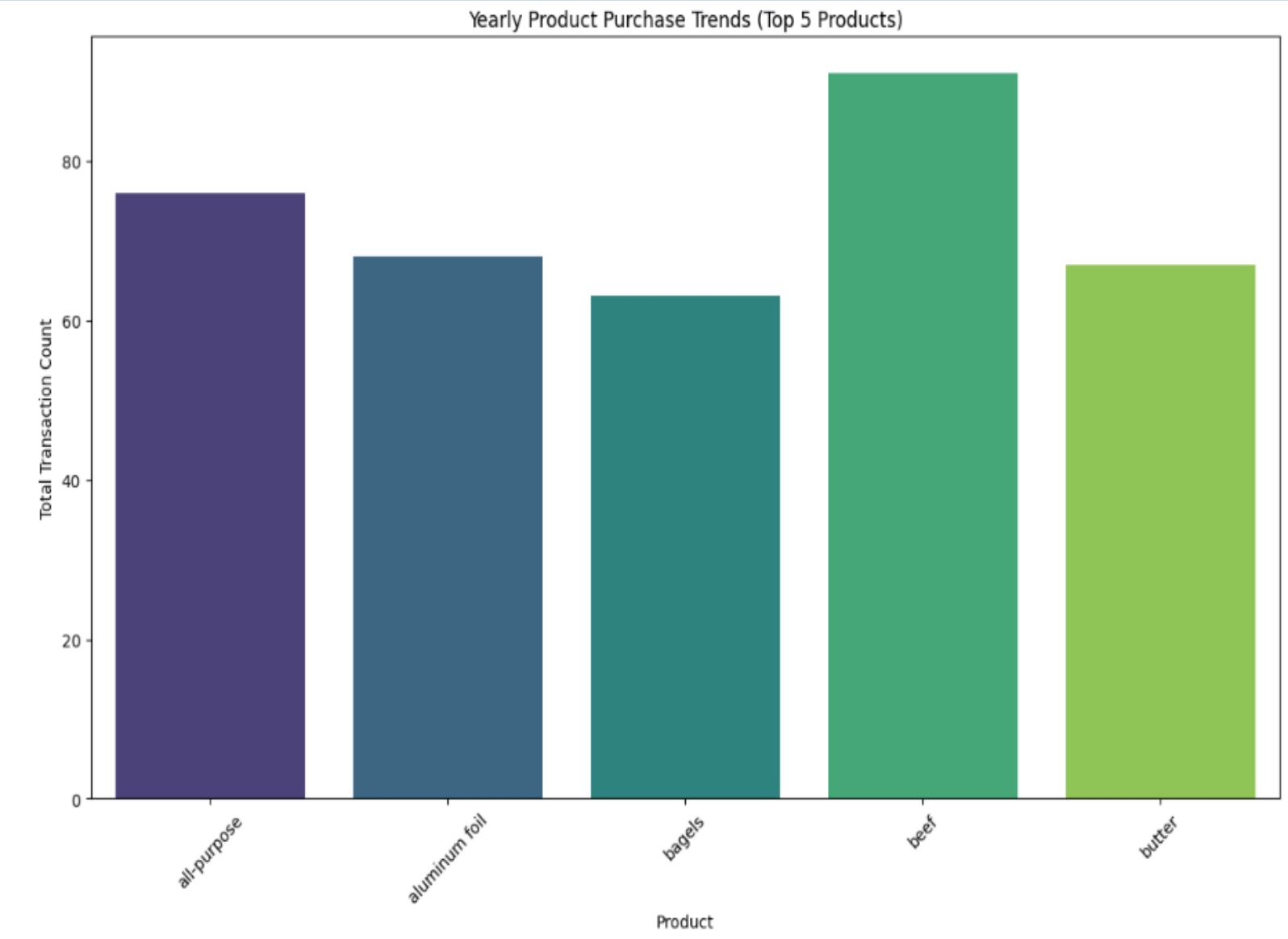
Quarterly Trend

- The data includes only **one quarter (2018Q1)**, making it impossible to observe trends over time.
- **Beef** had the highest transaction count (**91**), while **bagels** had the lowest (**63**).
- The scatter plot lacks time variation, resulting in overlapping data points that reduce clarity.
- A **line chart with multiple quarters** would better visualize product purchase trends.



Yearly Trend

- **Beef has the highest transactions**, indicating it is the most frequently purchased product among the top 5.
- **All-purpose and aluminum foil follow closely**, suggesting consistent demand.
- **Bagels and butter have relatively lower transactions**, but still rank among the top 5.
- **The variation in purchase frequency** highlights product popularity differences, which could inform inventory or marketing strategies.



Summary

- In 2018 and 2019, the top three products sold were Cereals, Poultry, and Soda. In January and February of 2020, the leading products were Dinner Rolls, Poultry, and Pork.
- Over the years, Poultry, Cereals, and Soda consistently ranked as the highest-selling items.
- Among non-edible products, Soap and Toilet Paper were the best-sellers, while Hand Soap had the lowest sales.
- For edible products, Poultry, Soda, and Cereals were the top sellers, while Pork, Fruits, and Sandwich Loaves saw the least sales.
- Sunday was the day with the highest sales, while Monday had the lowest.
- In 2018, January saw the most products sold, while February had the least. In 2019, March had the highest sales, and January had the lowest.
- Sales peaked in Q1 of 2019 and Q3 of 2018. The number of products sold in Q2 was similar in both 2019 and 2018. Sales in 2020 were low, likely due to data being available only up to February 26th.

Recommendation

- Prioritize promoting and stocking poultry, soda, and cereals, as these consistently top the sales charts.
- Consider increasing inventory of soap and toilet paper, as these are the highest-selling non-edible products.
- Analyze the reasons behind the low sales of hand soap and implement strategies to boost its performance.
- Schedule promotions and discounts on Sundays to take advantage of the highest sales day.
- Plan marketing campaigns and special offers for February to drive sales during the typically slow month.
- Focus marketing efforts and discounts on January and March to capitalize on the historically high-sales months. Aim to replicate the sales success seen in Q1 2019 and Q3 2018.
- Maintain consistent stock levels for Q2 products to ensure steady sales. Take into account the limited data available for 2020 when making sales and marketing decisions.

3. Market Basket Analysis

- Market Basket Analysis
- MRA KNIME Workflow & Output Table

Market Basket Analysis

Definition: Market Basket Analysis is a statistical method that examines customer purchasing behaviors to uncover patterns and relationships between products. It helps businesses understand which items are commonly bought together and how customers' buying habits influence sales.

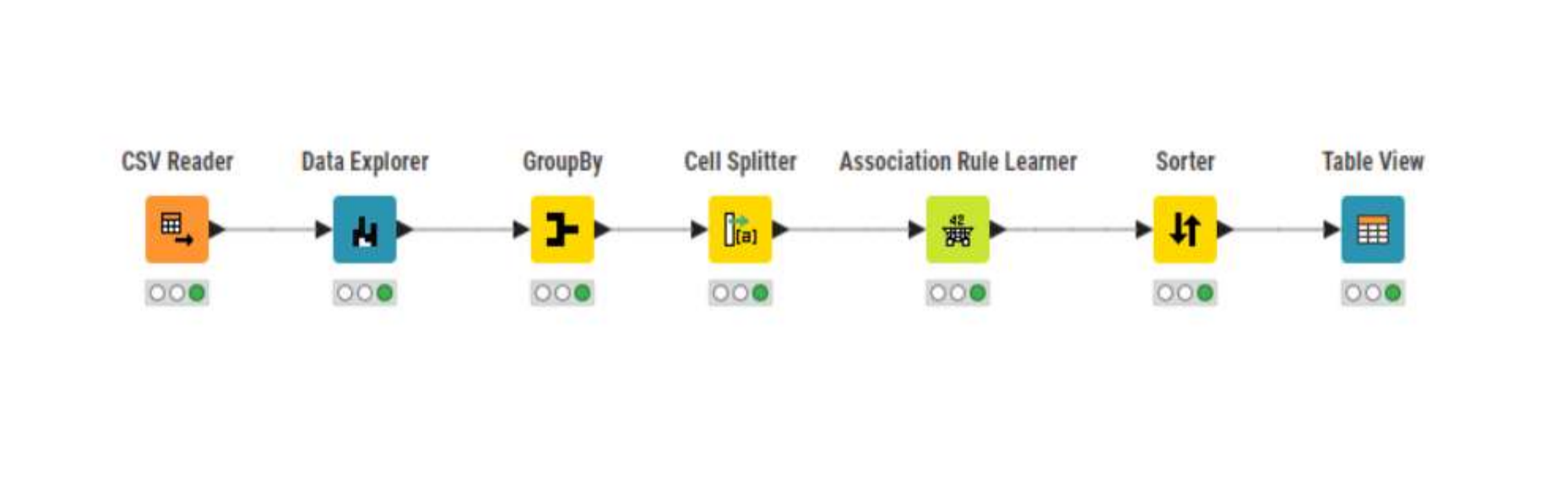
Data: To perform market basket analysis, businesses need transactional data that includes customer ID, product ID, and transaction date. This data is used to create a matrix that highlights the connections between various products.

Association Rules: Association rules are used to quantify the strength of relationships between different products. These rules are expressed through support, confidence, and lift. Support indicates how often items are bought together in a transaction, confidence measures the likelihood that a customer who buys one item will buy another, and lift evaluates the strength of the correlation between two products.

Applications: Market Basket Analysis is widely applied across industries such as retail, e-commerce, and marketing. Retailers use it to enhance product placement and promotions, e-commerce companies leverage it for personalized product recommendations, and marketers apply it to design targeted advertising campaigns.

Benefits: Market Basket Analysis boosts revenue by identifying cross-selling opportunities and enabling targeted promotions. It also enhances customer satisfaction by offering personalized recommendations and improving the overall shopping experience.

MRA KNIME Workflow



KNIME Output

Output Table (Table)

Rows: 1139 | Columns: 3

#	RowID	Order_id (Number (integer))	Concatenate(Product) (String)	Concatenate(Product)_SplitResultSet (Set)
1	Row0	1	yogurt, pork, sandwich bags, lunch meat, all- purpose, flour, soda, butter, be	[yogurt,pork,sandwich bags,...]
2	Row1	2	toilet paper, shampoo, hand soap, waffles, cheeses, mixes, milk, sandwich b	[toilet paper,shampoo,hand soap,...]
3	Row2	3	soda, pork, soap, ice cream, toilet paper, dinner rolls, hand soap, spaghetti s	[soda,pork,soap,...]
4	Row3	4	cereals, juice, lunch meat, soda, toilet paper, all- purpose	[cereals,juice,lunch meat,...]
5	Row4	5	sandwich loaves, pasta, tortillas, mixes, hand soap, toilet paper, paper towe	[sandwich loaves,pasta,tortillas,...]
6	Row5	6	laundry detergent, toilet paper, eggs, toilet paper, bagels, dishwashing liquic	[laundry detergent,toilet paper,eggs,...]
7	Row6	7	individual meals, paper towels, tortillas, milk, ice cream, juice, dishwashing	[individual meals,paper towels,tortillas,...]
8	Row7	8	ice cream, juice, paper towels, waffles, soda, cheeses, poultry, toilet paper	[ice cream,juice,paper towels,...]
9	Row8	9	juice, poultry, coffee/tea, coffee/tea, dishwashing liquid/detergent	[juice,poultry,coffee/tea,...]
10	Row9	10	ketchup, coffee/tea, toilet paper, pork, flour, milk, soda, dishwashing liquid/	[ketchup,coffee/tea,toilet paper,...]
11	Row10	11	sandwich loaves, ice cream, soda, bagels, dishwashing liquid/detergent, eg	[sandwich loaves,ice cream,soda,...]
12	Row11	12	pork, tortillas, pork, shampoo, lunch meat, pasta, juice, bagels, bagels, launc	[pork,tortillas,shampoo,...]
13	Row12	13	sugar, fruits, all- purpose, aluminum foil, laundry detergent, individual meals	[sugar,fruits,all- purpose,...]
14	Row13	14	fruits, dinner rolls, individual meals, shampoo, ketchup, cereals, sandwich b	[fruits,dinner rolls,individual meals,...]
15	Row14	15	individual meals, ice cream, cereals, paper towels, bagels, mixes, lunch mea	[individual meals,ice cream,cereals,...]
16	Row15	16	sugar, sandwich bags, flour, juice, milk, paper towels, cereals, sandwich bag	[sugar,sandwich bags,flour,...]
17	Row16	17	milk, hand soap, pasta, individual meals, spaghetti sauce, cereals, sandwich	[milk,hand soap,pasta,...]
18	Row17	18	sandwich bags, toilet paper, bagels, shampoo, coffee/tea	[sandwich bags,toilet paper,bagels,...]
19	Row18	19	individual meals, laundry detergent, coffee/tea, eggs, aluminum foil, beef, ju	[individual meals,laundry detergent,coffee/tea,...]
20	Row19	20	shampoo, dishwashing liquid/detergent, yogurt, juice, sugar, soap, sandwich	[shampoo,dishwashing liquid/detergent,yogurt,...]
21	Row20	21	waffles, fruits, all- purpose, pork, juice, bagels, mixes	[waffles,fruits,all- purpose,...]
22	Row21	22	cheeses, cereals, sugar, bagels, soda	[cheeses,cereals,sugar,...]
23	Row22	23	aluminum foil, bagels, shampoo, shampoo, dishwashing liquid/detergent, c	[aluminum foil,bagels,shampoo,...]
24	Row23	24	fruits, all- purpose, pasta, cheeses, juice, sandwich bags, sandwich loaves,	[fruits,all- purpose,pasta,...]
25	Row24	25	bagels, sugar, pork, sandwich loaves, tortillas, ice cream, all- purpose, yogu	[bagels,sugar,pork,...]

4. Associations Identified

- MRA – values
- Association Rule Parameters
- Association Rules Table
- Table View

MRA – values
Support, Confidence and Lift

In **Market Basket Analysis**, support, confidence, and lift are key metrics for evaluating the relationship between items in a transaction dataset.

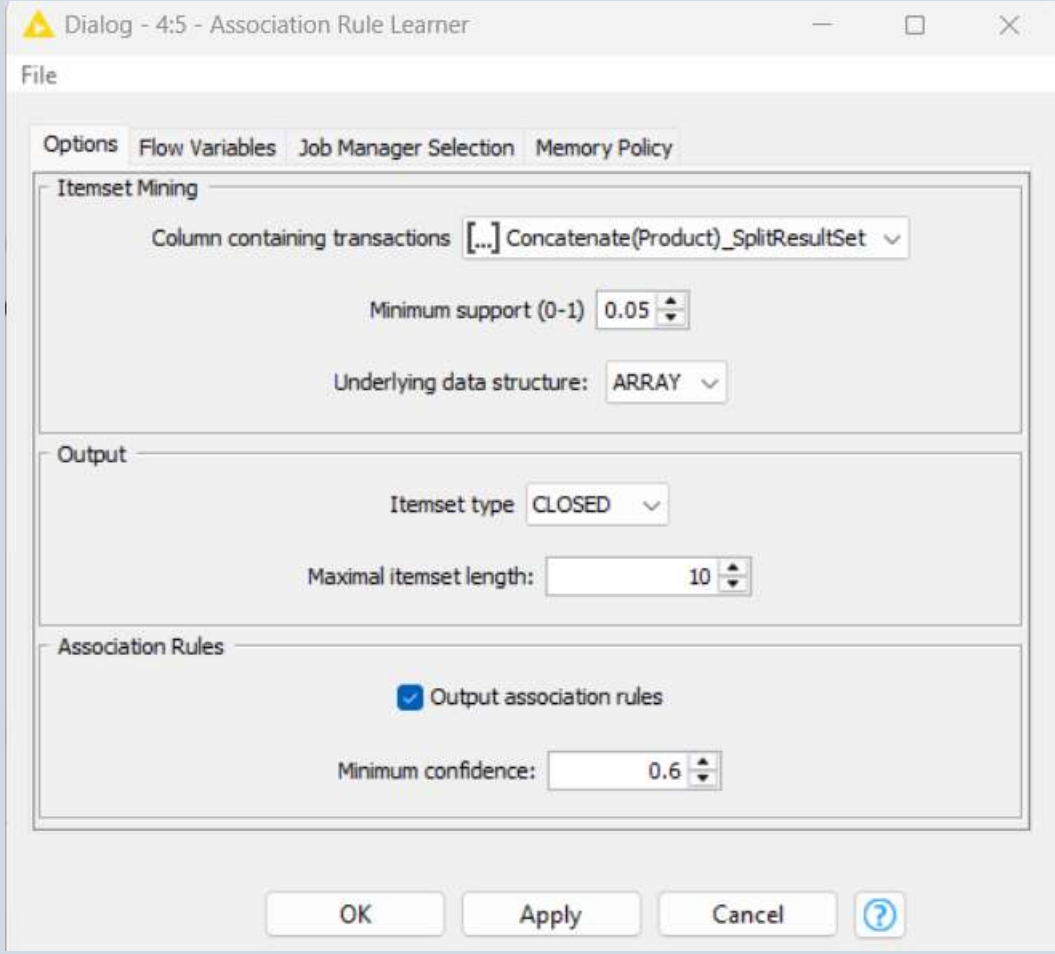
- **Support:** The probability of items appearing together in a transaction. It is calculated as the number of transactions containing both items divided by the total transactions. High support suggests that an itemset is frequent and suitable for promotions or grouping.
- **Confidence:** The likelihood that a transaction containing one item also includes another. It is computed as the number of transactions with both items divided by those containing the first item. A high confidence score indicates a strong association, making it useful for recommendations.
- **Lift:** Measures how much more frequently two items appear together compared to if they were independent. It is determined by dividing the support of the itemset by the product of individual item supports.
    - **Lift = 1:** No association.
    - **Lift > 1:** Positive association (items frequently bought together).
    - **Lift < 1:** Negative association.
      A high lift value signifies strong interdependence, making it useful for cross-selling and bundling strategies.

Association Rule Parameters

Support of Minimum: 0.05

Maximum Item Set Length: 10

Minimum Confidence Level:0.6



Association Rules Tables

- Association rules identify relationships between items in large datasets by analyzing frequent item sets—groups of items that commonly appear together in transactions.
- Based on the dataset and set parameters, 24 rules have been generated.

Frequent Items/Association rules [Table]

Rows: 24 | Columns: 6

#	RowID	Support Number (double)	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
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 crea
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 deter
20	rule19	0.054	0.656	1.556	poultry	<—	[dinner rolls,spaghetti sauce,laundry r
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]

Table View

Sorter Node:

The **Sorter** node is used to arrange the association rules in a meaningful order based on a selected metric (e.g., confidence, lift, or support). This helps in identifying the most relevant rules for business insights.

Table View Node:

The **Table View** node displays the sorted rules in a structured format, making it easier to analyze and interpret the results. This allows users to examine key metrics like support, confidence, and lift for each rule.

Interactive View: Table View

Association Rules Table

Rows: 24 | Columns: 6

R...	Support (Number (double))	Confidence (Number (double))	Lift (Number (double))	Consequent (String)	Implies (String)	Items (Set)
rule0	0.05	0.64	1.7	juice	<–	[yogurt,toilet paper,aluminum foil]
rule1	0.05	0.62	1.645	juice	<–	[yogurt,poultry,aluminum foil]
rule2	0.05	0.613	1.616	coffee/tea	<–	[yogurt,cheeses,cereals]
rule3	0.05	0.6	1.424	poultry	<–	[dishwashing liquid/detergent,laundry
rule4	0.051	0.63	1.678	mixes	<–	[yogurt,poultry,aluminum foil]
rule5	0.051	0.611	1.66	sandwich bags	<–	[cheeses,bagels,cereals]
rule6	0.051	0.674	1.726	cheeses	<–	[bagels,cereals,sandwich bags]
rule7	0.051	0.617	1.558	cereals	<–	[cheeses,bagels,sandwich bags]
rule8	0.051	0.63	1.621	dinner rolls	<–	[spaghetti sauce,poultry,cereals]
rule9	0.051	0.637	1.512	poultry	<–	[dinner rolls,spaghetti sauce,cereals]
rule10	0.051	0.604	1.589	milk	<–	[poultry,laundry detergent,cereals]

5. Inference & Recommendation

- Key insights from the analysis
- Business Recommendation

Key insights from the analysis

- The analysis has identified **frequently purchased product combinations**, providing valuable insights for optimizing **store layout** and **promotional strategies**.
- Items such as **yogurt, poultry, aluminum foil, cheeses, cereals,** and **dinner rolls** rank among the most **commonly bought products**.
- Some **product associations** are **unexpected**, such as **poultry** being purchased alongside **dishwashing liquid, laundry detergent,** and **mixes**. These insights suggest opportunities for **strategic promotions**, such as **"buy two, get one free"** offers, to encourage **higher sales**.
- Additionally, placing **complementary products** closer together can enhance **cross-selling opportunities**, increasing the likelihood of **additional purchases**.
- Overall, **market basket analysis** enables the store to gain a **deeper understanding of customer behavior and preferences**, leading to more **informed decisions** on **product placement** and **marketing efforts**.

Business Recommendation

1. Lucrative Combo Deals to Boost Sales

"Family Meal Deal": Bundle dinner rolls, poultry, spaghetti sauce, and ice cream at a discounted price.

"Breakfast Saver Pack": Offer bagels, cereals, sandwich bags, and cheese as a breakfast combo at a lower price.

2. Discount Offers Based on Strong Product Associations

"Buy Two, Get One Free" promotions:

- Buy eggs & ice cream, get pasta free.
- Buy dinner rolls & poultry, get juice at a discount.
- Buy yogurt, poultry, and aluminum foil, get an additional discount.

Offer a discount on mixes when purchased with yogurt, poultry, or aluminum foil.

Provide a discount on dinner rolls when purchased with spaghetti sauce or poultry.

3. Cross-Selling & Strategic Product Placement

- Promote **paper towels** near checkout areas when customers buy **eggs, pasta, and ice cream**, as they are frequently purchased together.
- Highlight **spaghetti sauce, poultry, and cereals** together as a **convenient meal-prep kit**.
- Create a **"Paper Products Bundle"** that includes **paper towels, toilet paper, and tissues** at a discounted price.

4. Seasonal & Holiday Promotions

- During **festive seasons**, offer deals on **poultry, pasta, and dinner rolls** as part of a **holiday meal bundle**.
- Promote **household essentials** like **laundry detergent, poultry, and spaghetti sauce** with special discounts to **encourage bulk buying**.

5. Promotional Strategies to Increase Awareness

- Advertise these **offers through in-store signage, advertisements, and social media** to ensure customers are aware of the deals available.
- Place **complementary products in close proximity** to encourage **additional purchases**.

6. Personalized & Loyalty-Based Offers

- **Targeted Discounts**: Offer **personalized discounts** to frequent buyers of **poultry, cereals, or dairy** based on their purchase history.
- **Loyalty Rewards**: Provide **bonus points** for customers purchasing **combo deals** or specific high-margin products like **spaghetti sauce and ice cream**.

7. Limited-Time Flash Sales & Bundles

- **Weekend-Only Offers**: Offer special discounts on **dinner rolls, poultry, and juices** every **Saturday & Sunday** to increase foot traffic.
- **BOGO on Essentials**: Buy **toilet paper, paper towels, or cleaning products**, get **a small grocery item free**.

8. Upselling & Digital Promotions

- **Premium Upgrades**: Suggest **organic or gourmet versions** of commonly purchased items like **bagels, cheese, and pasta** at checkout.
- **Online-Exclusive Deals**: Provide **discounts on online purchases** for **bulk orders of poultry, cereals, and household essentials**.

THANK YOU