MRA PPT PROJECT 2

- Retail Market Basket & Time Series Analysis using KNIME & Tableau: Consumer Behavior, Product Affinity, and Sales Forecasting -

Name – Sibi Arvind J

Agenda

01 Introduction

02 Dataset Summary & Problem Statement

03 Univariate Analysis

04 Bivariate Analysis

Agenda

05 Multivariate Analysis

06 Time-Based Trend Analysis

07 Market Basket Analysis

08 Inferences and Recommendations

01

Introduction

Business Context :

In grocery retail, identifying frequently purchased product combinations can improve promotions, inventory, and customer satisfaction.

Objective:

Use transactional data to discover item patterns and enable data-driven combo offers.

Tools Used :

- **Tableau** (for Exploratory Data Analysis & Trend Analysis)
- **KNIME** (for Market Basket Analysis)

01

Dataset Summary & Problem Statement

Problem Statement:

- Identify high-frequency item combinations from POS data to guide promotional offers and basket growth.

Dataset Summary :

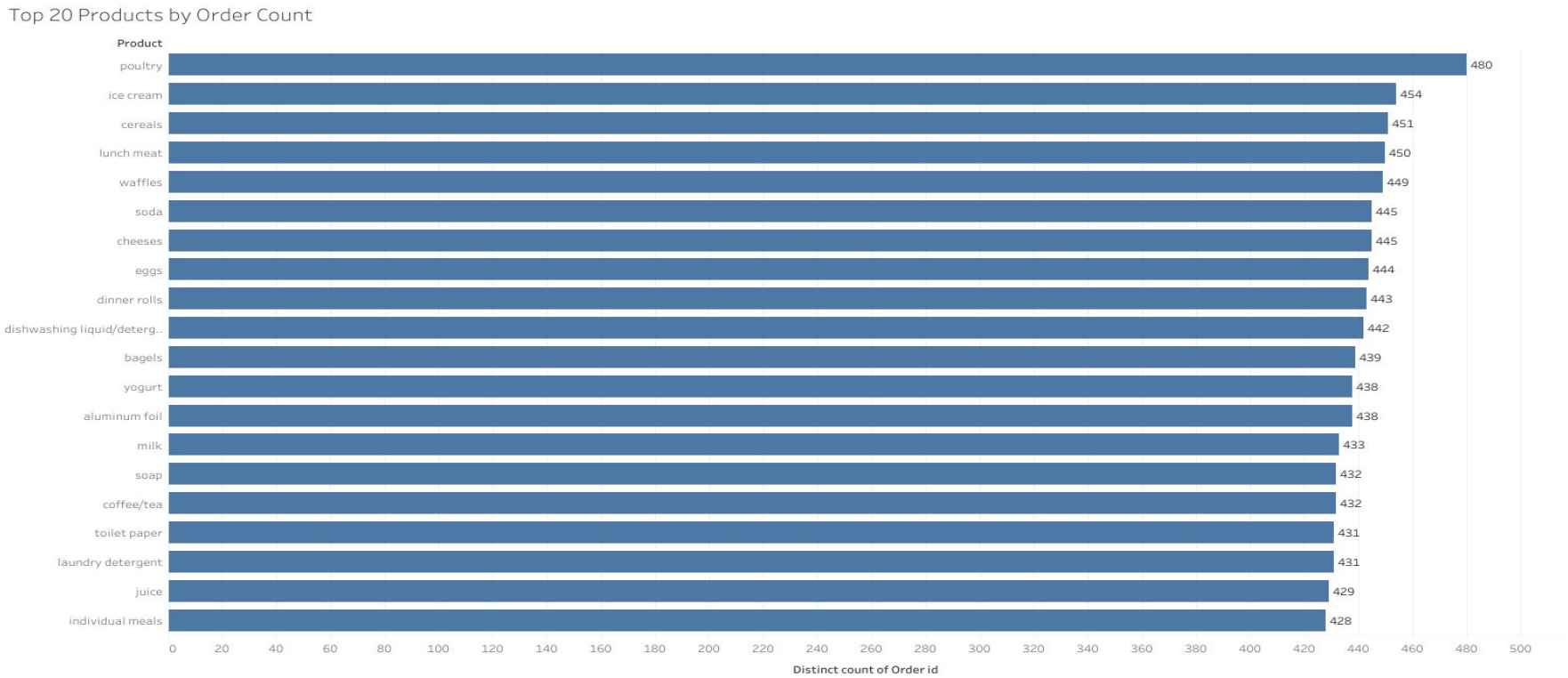
- **Total Transactions**: 20,641
- **Unique Orders**: 603
- **Unique Products**: 37

Columns:

- `Date` (dd-MM-yyyy format)
- `Order_id`
- `Product`

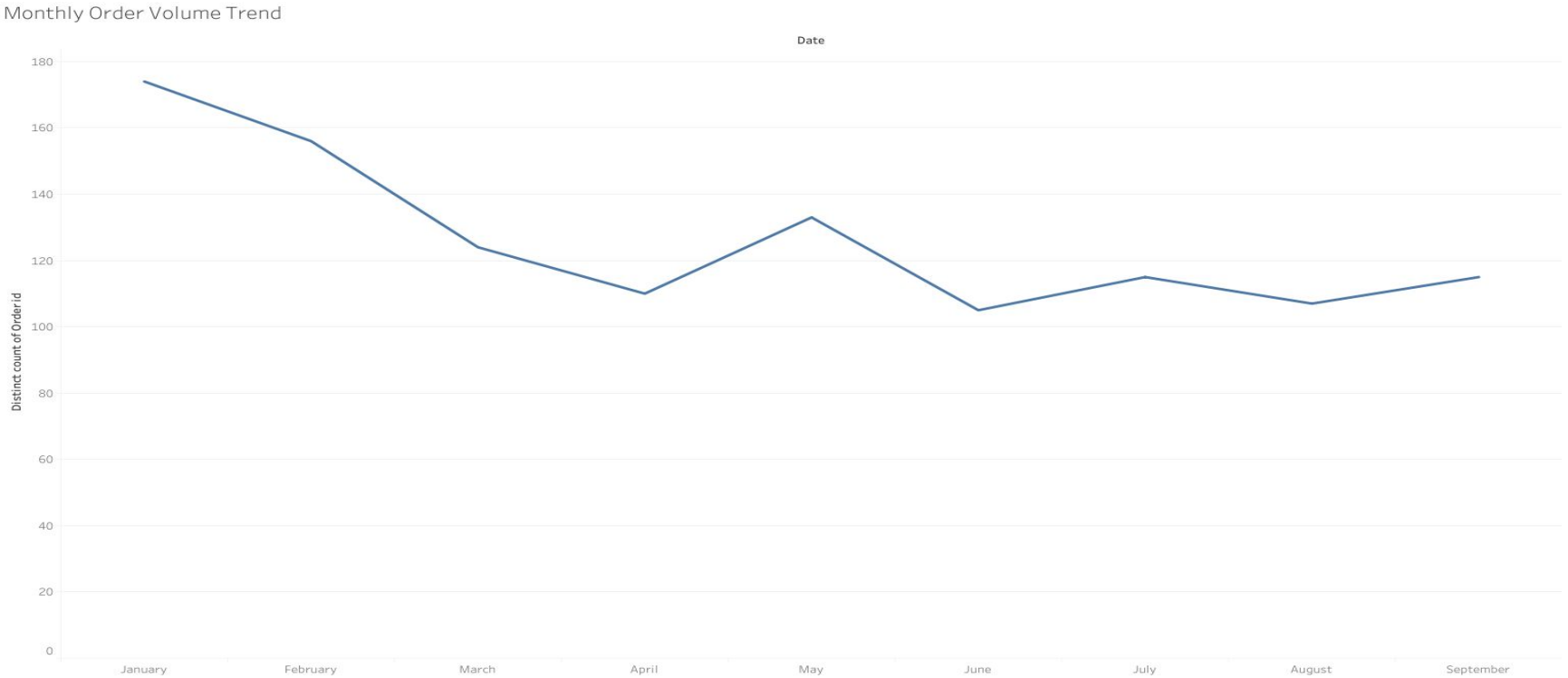
Univariate analysis

Top 20 Products by Order Count



This plot highlights the frequency of orders for the **top 20** products. "Poultry" is the most popular product with 480 orders, followed closely by "Ice Cream" at 454 orders. At the 20th position, "Individual Meal" has 428 orders.

Monthly Order Volume Trend

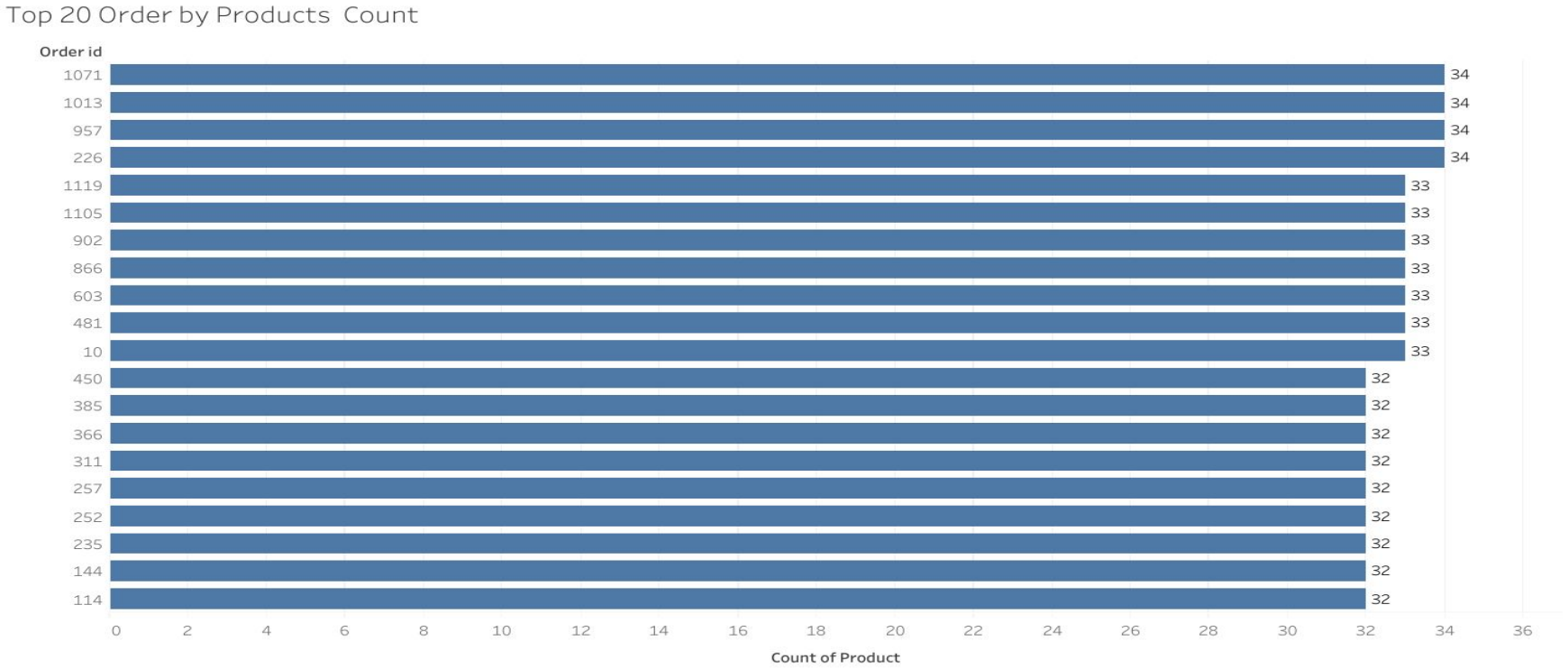


Order volume peaks in January (175) and declines through April (110). A notable increase occurs in May (135), followed by fluctuations in June, July, and August. The trend ends with a modest rise in September. This suggests a seasonal dip early in the year and recovery towards the fall.

03

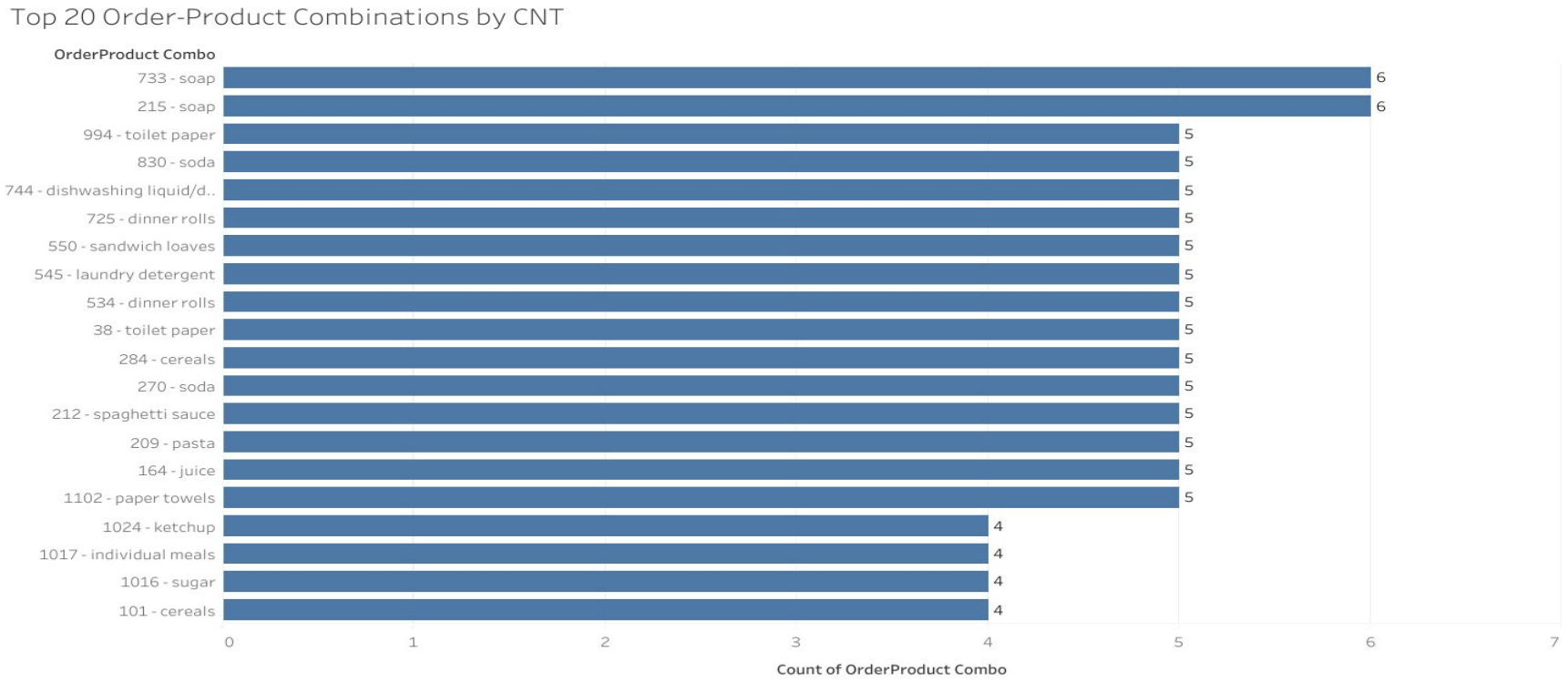
Bivariate Analysis

Top 20 Order by Products Count



Products with IDs 1072, 1013, 957, and 226 lead with 34 orders each. Several others, including 1119, 1105, and 902, follow closely with 33 orders. The chart highlights the frequency of product orders, showing that a few products are consistently ordered more than others, reflecting customer preferences.

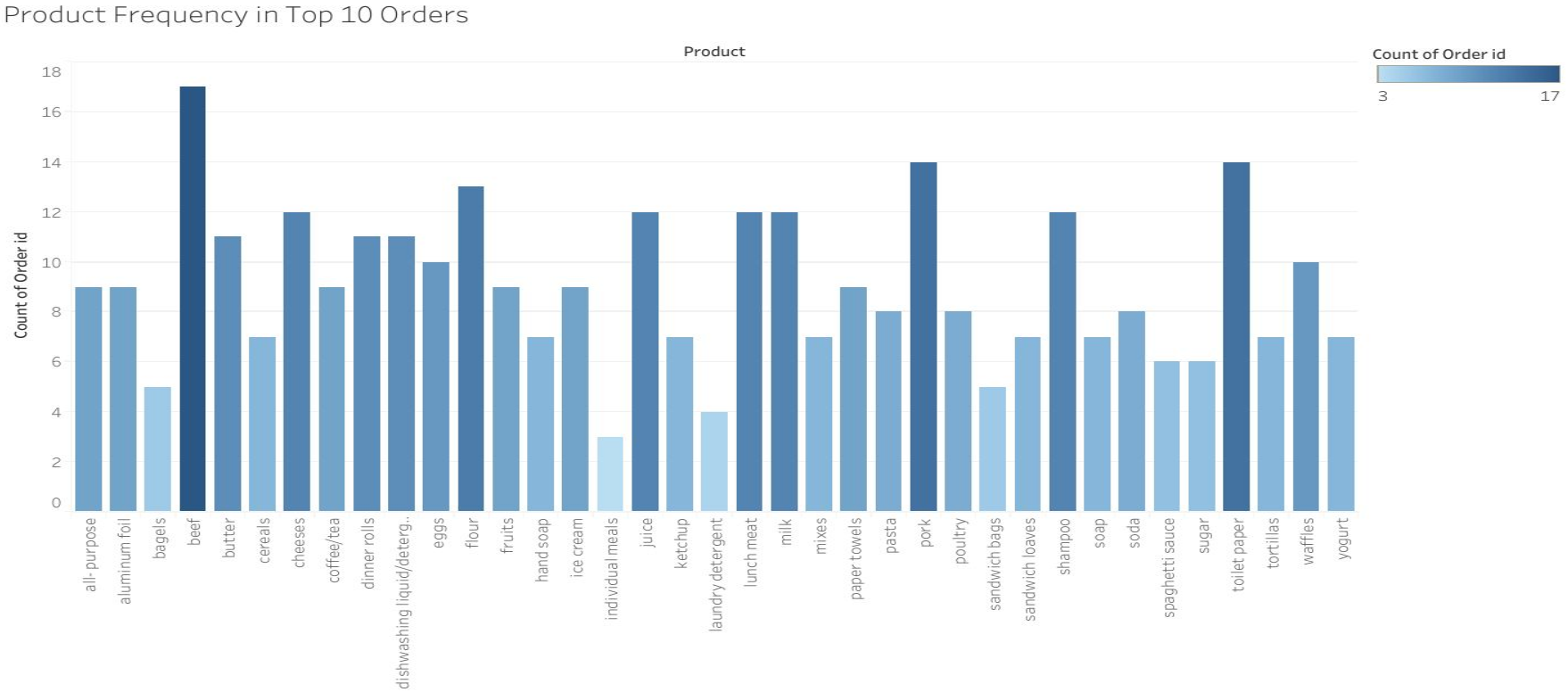
Top 20 Order-Product Combinations by CNT



This chart highlights the top 20 most frequent order-product pairs. 733-soap and **233-soap** lead with **6 orders each**, followed by **994-toilet paper** and others with **5 orders**. These repeated combinations indicate consistent customer preferences and popular purchase patterns.

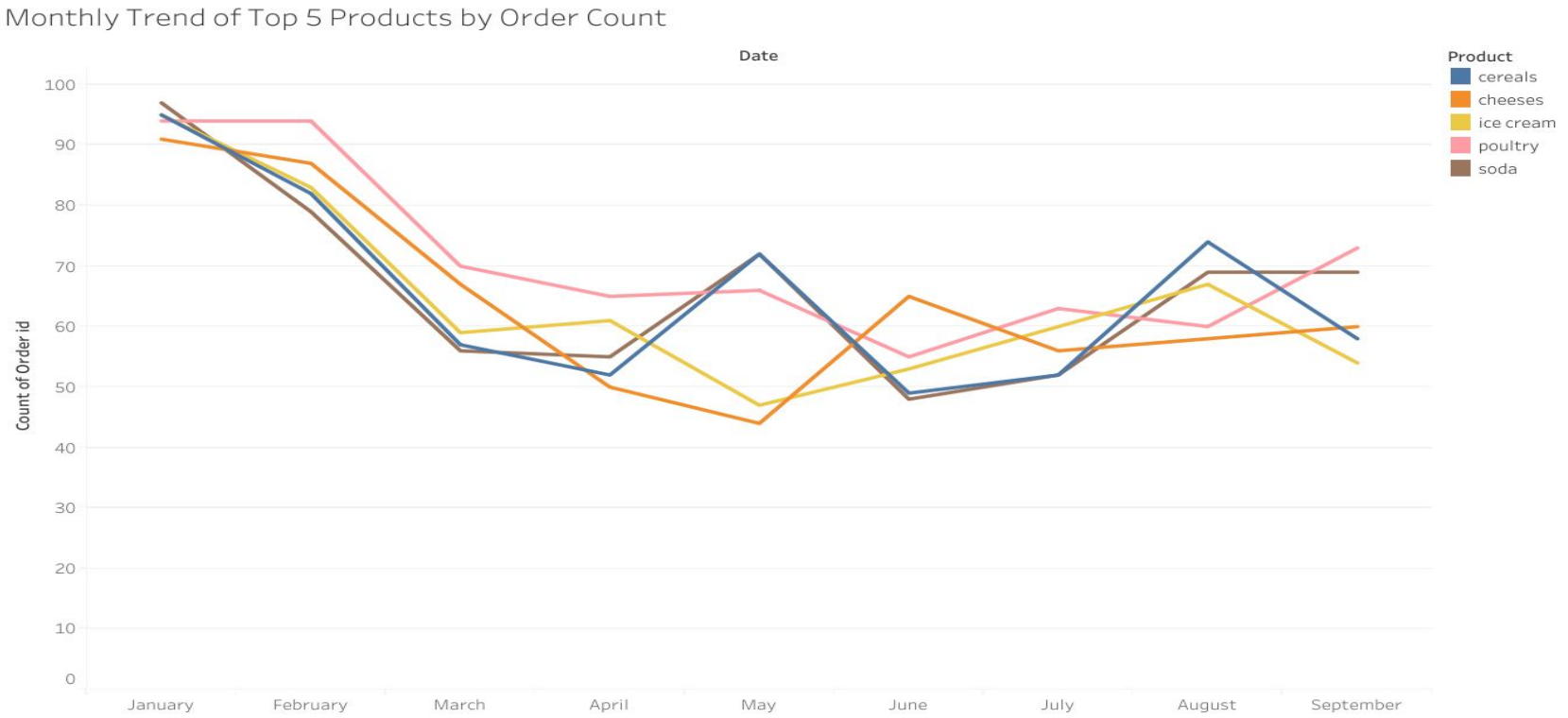
Multivariate analysis

Product Frequency in Top 10 Orders



Among the top 10 orders, **Beef** appears most frequently with **17 orders**, followed by **Pork** and **Toilet Paper** with **14 each**. **Individual Meals** show the least presence, with just **3 orders**, highlighting variation in product popularity within high-value orders.

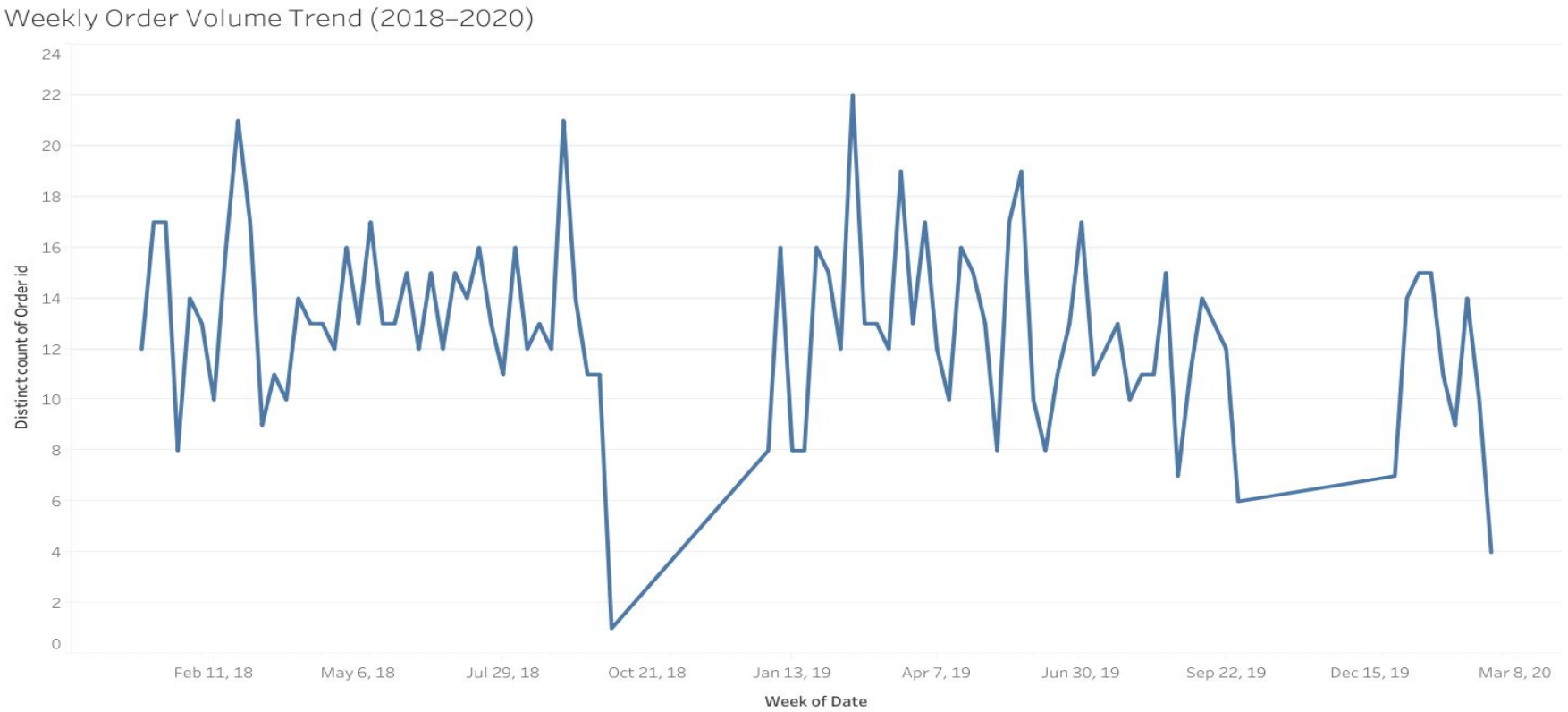
Monthly Trend of Top 5 Products by Order Count



Cereals and **Soda** show similar trends — starting high around **95 orders**, dipping to around **54** by April, rising in **May**, dropping again in **June**, and peaking in **August** (**Cereals: 74**, **Soda: 69**). In **September**, **Cereals** drop to **58**, while **Soda** remains steady at **69**. This indicates seasonal or demand-based fluctuations in order volume.

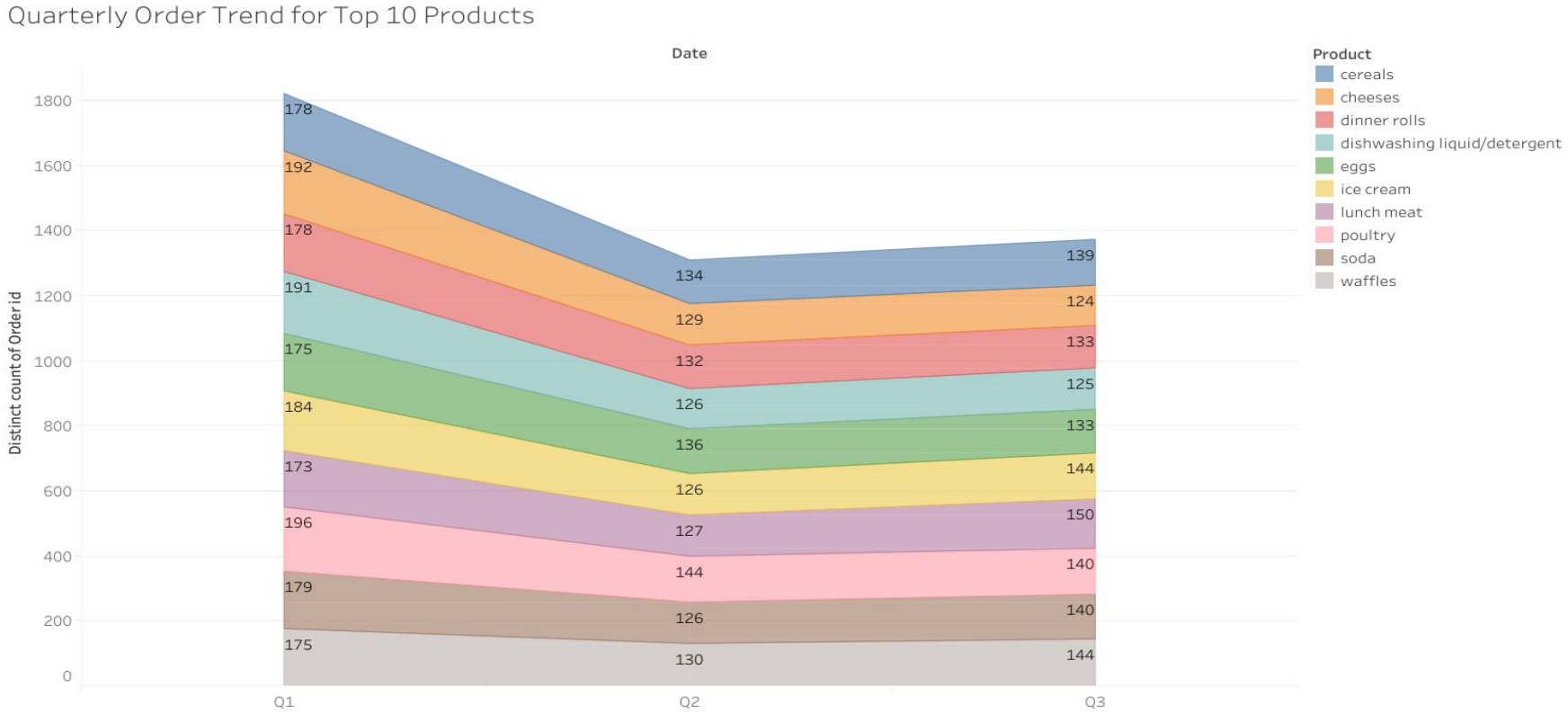
Time-Based Trends analysis

Weekly Order Volume Trend (2018–2020)



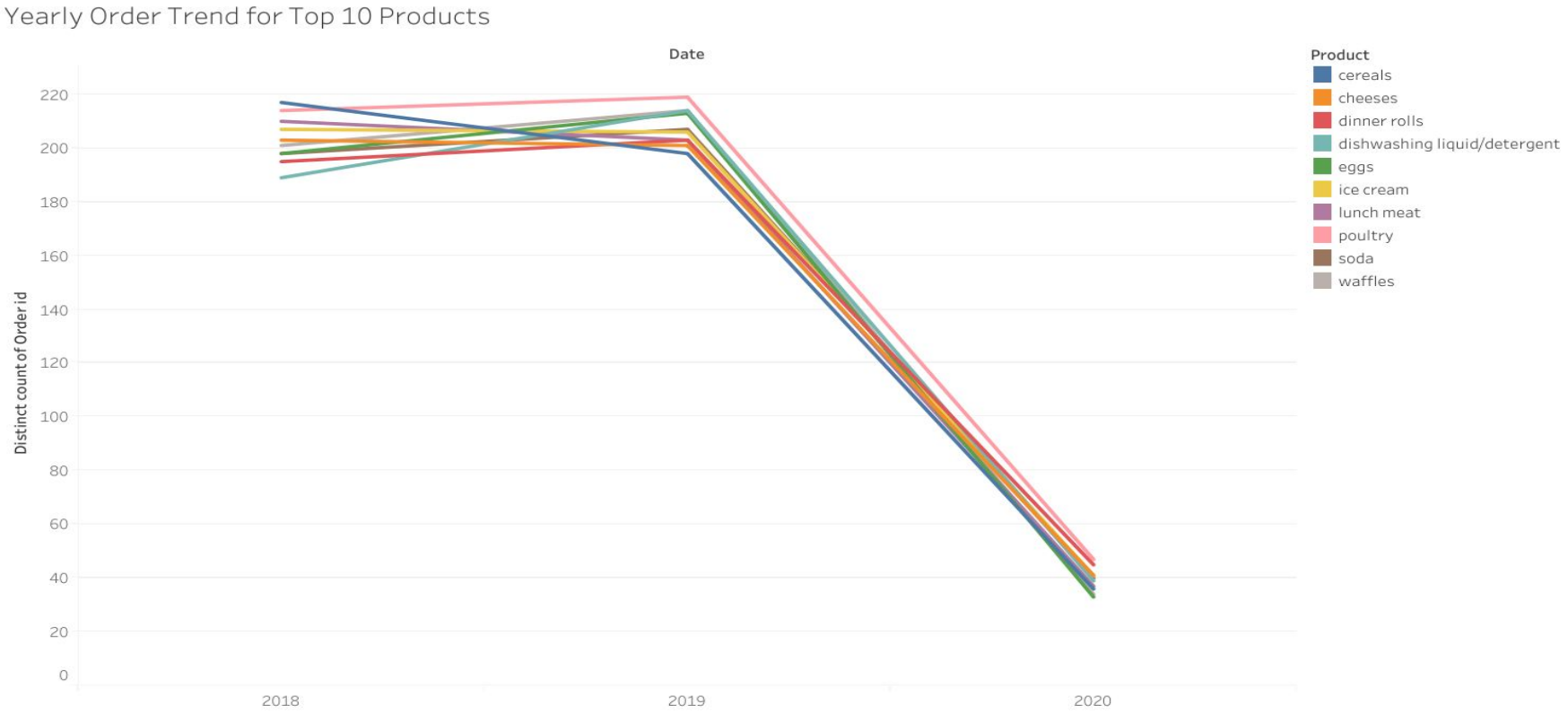
Early 2018: Variable order volume with initial peaks and dips. **Mid–Late 2018:** Stable phase, followed by a **sharp drop in late September**. **Late 2018 – Early 2019:** Recovery begins with a steady upward trend. **2019:** Increased **volatility** with sharper weekly highs and lows. **Early 2020:** Continued **fluctuations**, marked by noticeable spikes and dips in order activity.

Quarterly Order Trend for Top 10 Products



Cereals, **Cheese**, and **Dinner Rolls** show a **decline in order volume from Q1 to Q2**.
While **Cereals** and **Dinner Rolls** see a slight **recovery**, **Cheese** continues to **decline into Q3**, indicating reduced demand over time.

Yearly Order Trend for Top 10 Products



All top 10 products show **high order counts in 2018**, followed by a **slight increase or stability in 2019**. In **2020**, there's a **sharp decline** across all products, indicating a significant drop in overall demand or activity..

Summary of EDA & Time-Based Trends :

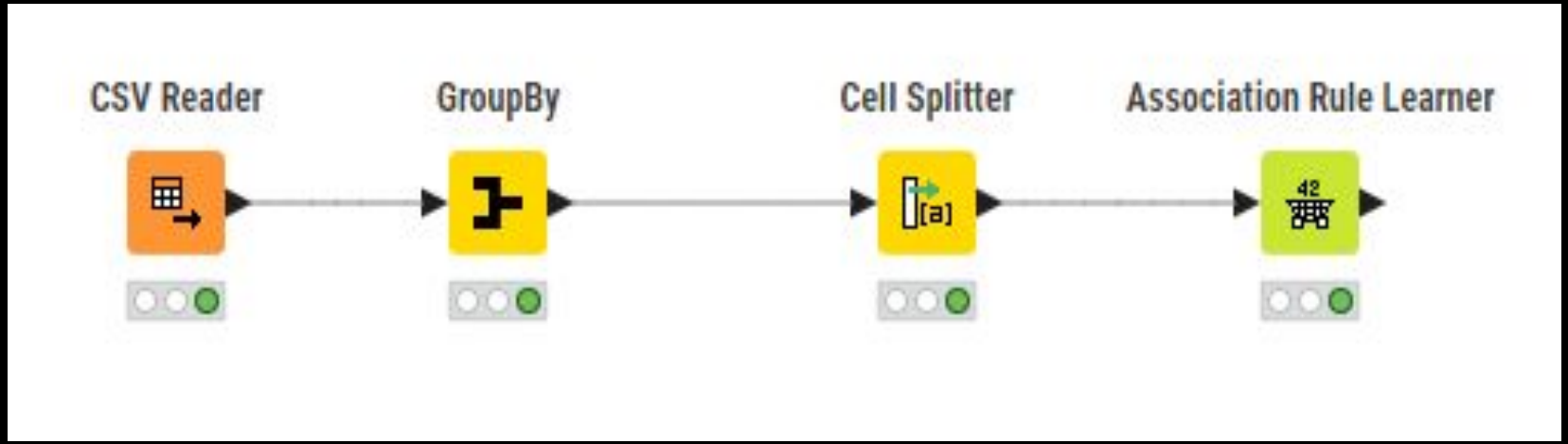
- **Univariate Analysis:**
  - **Top Products**: Poultry (480), Ice Cream (454), Individual Meal (428).
  - **Monthly Volume**: Peaks Jan→dips Apr, rebounds May, fluctuates, small rise by Sept.
- **Bivariate Analysis:**
  - **Product Orders**: IDs 1072/1013/957/226 at 34 orders; next tier at 33.
  - **Order-Product Pairs**: 733-soap & 233-soap at 6; 994-toilet paper at 5.
- **Multivariate Analysis:**
  - **Monthly Trends**: Cereals & Soda follow similar dips Apr, peaks Aug, diverge in Sept.
  - **High-Volume Orders**: Beef (17), Pork & Toilet Paper (14), Individual Meals (3).
- **Time-Based Trends:**
  - **Weekly (’18–’20)**: Early volatility → Sept ’18 drop → recovery into ’19 → high volatility in ’19 → fluctuations into ’20.
  - **Quarterly**: Cereals & Rolls dip Q1→Q2, rebound Q3; Cheese keeps falling.
  - **Yearly**: High ’18 → slight rise/stable ’19 → sharp drop in ’20.

Key Takeaways: Clear seasonality (early-year dip, late-summer rebound), a handful of top performers, bulk-order favorites, and a notable 2020 demand collapse.

Market Basket Analysis

- **Objective:**
  - **Discover frequently purchased product combinations to support combo offers, targeted promotions, and inventory decisions.**
- **Approach:**
  - Used **Association Rule Mining** on transactional data.
  - Applied KNIME workflow to identify strong product associations.
- **Technique Used:**
  - **Apriori algorithm** via KNIME's Association Rule Learner node.
  - Transactions were prepared by **grouping Product values by Order ID**.

KNIME Workflow for Market Basket Analysis:



Explanation:

- CSV Reader: Imported transactional data.
- GroupBy: Aggregated products under each `Order_id`.
- Cell Splitter: Split concatenated items into individual elements.
- Association Rule Learner: Extracted association rules using set thresholds.

🔷 Support & Confidence – Threshold Selection :

- **Minimum Support:** `0.05`

  - *Means item combinations must appear in at least 5% of all orders to be considered.*

- **Minimum Confidence:** `0.5`

  - *Rules are only accepted if they hold true in at least 50% of the relevant cases.*

- **Why These Thresholds?**
  - Balanced enough to avoid rare or weak associations.
  - Provides meaningful rules that can drive marketing actions.

Association Rules – Top Insights

Table: Top 8 Association Rules (Sorted by Lift)

#	RowID	Support Number (double)	Confidence Number (double)	Lift Number (double)	Consequent String	implies String	Items set
60	rule59	0.05531167691	0.6494845361	1.79119343	paper towels	<---	[eggs, ice cream, pasta]
59	rule58	0.05531167691	0.6428571429	1.73100304	pasta	<---	[paper towels, eggs, ice cream]
22	rule21	0.05092186128	0.6744186047	1.726208518	cheeses	<---	[bagels, cereals, sandwich bags]
4	rule3	0.05004389816	0.6404494382	1.700400723	juice	<---	[yogurt, toilet paper, aluminum foil]
19	rule18	0.05092186128	0.6304347826	1.677722471	mixes	<---	[yogurt, poultry, aluminum foil]
21	rule20	0.05092186128	0.6105263158	1.659640749	sandwich bags	<---	[cheeses, bagels, cereals]
53	rule52	0.05355575066	0.6421052632	1.650920756	dinner rolls	<---	[spaghetti sauce, poultry, laundry detergent]
41	rule40	0.05179982441	0.6413043478	1.648861517	dinner rolls	<---	[spaghetti sauce, poultry, ice cream]

Insight: High lift values suggest strong associations between these product combos. These can inform combo packs, cross-promotions, and store layout decisions.

🔷 KNIME Workflow: Association Rules Output View: *Top 8 (Sorted by Lift)*

#	RowID	Support *Number (double)*	Confidence *Number (double)*	Lift ↓ *Number (double)*	Consequent *String*	implies *String*	Items *Set*
60	rule59	0.055	0.649	1.791	paper towels	<---	[eggs,ice cream,pasta]
59	rule58	0.055	0.643	1.731	pasta	<---	[paper towels,eggs,ice cream]
22	rule21	0.051	0.674	1.726	cheeses	<---	[bagels,cereals,sandwich bags]
4	rule3	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum foil]
19	rule18	0.051	0.63	1.678	mixes	<---	[yogurt,poultry,aluminum foil]
21	rule20	0.051	0.611	1.66	sandwich bags	<---	[cheeses,bagels,cereals]
53	rule52	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce,poultry,laundry detergent]
41	rule40	0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce,poultry,ice cream]

The above table represents the output of the **Association Rule Learner** node in KNIME, sorted by **Lift** in descending order. It highlights the top associations between products based on **Support**, **Confidence**, and **Lift** metrics. Each rule shows a combination of items that imply the likelihood of another item being purchased in the same transaction. These insights help identify potential product bundles and promotional opportunities.

Interpreting Association Metrics: Support, Confidence & Lift

Metric	Definition	What it Tells Us
Support	Proportion of total transactions that contain the itemset.	Measures how frequently the itemset appears in the dataset.
Confidence	Probability that a transaction with the antecedent also has the consequent.	Indicates how often the rule has been found to be true.
Lift	Ratio of observed support to expected support if the items were independent.	Values >1 suggest a strong association between the items.

Referring to first row of Association Rules output :

#	RowID	Support Number (double)	Confidence Number (double)	Lift Number (double)	Consequent String	implies String	Items set
60	rule59	0.05531167691	0.6494845361	1.79119343	paper towels	<---	[eggs, ice cream, pasta]

Item : `[eggs, ice cream, pasta] → paper towels`
Support: 5.5% | **Confidence:** 64.9% | **Lift:** 1.79
Interpretation: When **eggs, ice cream, and pasta** are purchased together, there's a **65% chance** that **paper towels** are also bought, and this is **1.79 times more likely** than by random chance.

Inferences and Recommendations

Key Insights from Market Basket Analysis :

- **Frequent Itemsets & Associations:**
  - The combination `[eggs, ice cream, pasta] → paper towels` had the highest lift (1.79), indicating a strong association.
  - `[bagels, cereals, sandwich bags] → cheeses` and `[yogurt, toilet paper, aluminum foil] → juice` also showed high-confidence patterns.
- **Consistent High-Confidence Rules:**
  - Most of the top rules had confidence levels above 0.60, meaning the associations are reliably observed in transactions.
- **Cross-Category Pairings:**
  - Items like **paper towels** (household) frequently co-occurred with **ice cream**, **eggs**, and **pasta** (groceries), hinting at cross-category shopping behavior.
- **Strong Potential for Combo Offers:**
  - Several triplet combinations imply the feasibility of creating meaningful product bundles or deals across different aisles.

Business Recommendations :

✅ Introduce Combo Offers:

- Create bundles such as **“Eggs + Ice Cream + Pasta → Free Paper Towels”** or **“Bagels + Cereals + Sandwich Bags → Discount on Cheese.”**

✅ Promote Cross-Selling:

- Place related household and food items together (e.g., paper towels near frozen foods) to boost basket size.

✅ Personalized Promotions:

- Use customer purchase history to offer tailored combo discounts based on common associations (e.g., “Based on your last purchase, you might like...”)

✅ Loyalty & Bulk Deals:

- For frequently co-purchased items like **poultry, toilet paper, and aluminum foil**, offer “Buy 2 Get 1 Free” schemes to drive volume sales.

✅ Optimize Store Layout:

- Arrange top associated products near each other to naturally encourage linked purchases and reduce decision fatigue.