K_picrgle _lb Pcr_gj ?l_jwqgq Npmhcar 2

By: Sangram Keshari Patro

Great Learning
POWER AHEAD

Problem Statement:

In the highly competitive grocery retail industry, understanding customer purchasing behavior is essential for driving revenue, increasing customer satisfaction, and optimizing operations. A grocery store has provided its Point of Sale (POS) transactional data with the objective of uncovering frequently purchased item combinations.

As a business analyst, your task is to perform a comprehensive analysis using association rule mining or similar techniques to identify popular item sets and generate actionable insights. The goal is to develop targeted combo offers and discount strategies that increase the average basket size, enhance customer retention, and support data-driven marketing decisions.

CONTENT

1. **Exploratory Analysis of data**
2. **Market Basket Analysis**
3. **Associations Identification**
4. **Suggestion of Possible Combos with Lucrative Offers**

01. Exploratory Analysis

- **Executive summary of the data**
- **Exploratory Analysis of data (Trends across months/years/quarters/days)**
- **Summary**

Executive Summary of the data:

Objective: To analyze Point of Sale (POS) data from a grocery store to uncover frequently purchased item combinations within customer orders. The objective is to deliver actionable recommendations for introducing effective combo offers and discounts that can boost store revenue and improve overall customer satisfaction.

- The analysed dataset consists of 20,641 rows and 3 columns, representing the Point of Sale (POS) data from a grocery store. The columns include 'Date', 'Order_id', and 'Product'. The dataset does not contain any missing values.

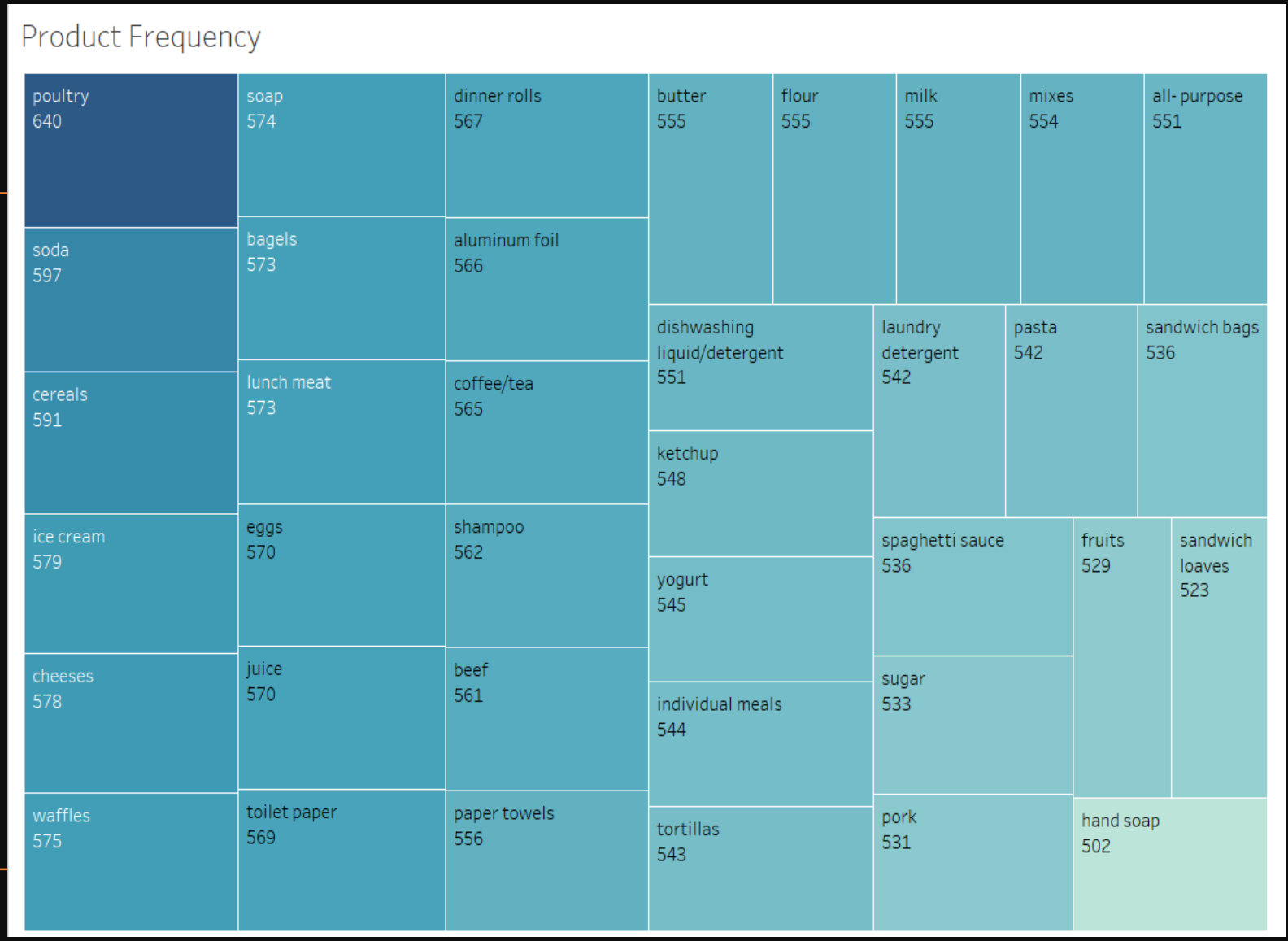
- **Duplicate Values:** The dataset contains 4,730 duplicate values; however, it is not necessary to remove them. Each duplicate Order ID reflects the purchase of a different product, meaning these entries represent individual items within the same transaction. As such, these duplicates capture essential details about customer purchasing behavior and offer valuable insights, making them important for analysis.

- **Market Basket Analysis** helps the grocery store identify which products are frequently purchased together by customers. This insight can be leveraged to design special deals and discounts on popular item combinations, encouraging customers to buy related products. By applying market basket analysis, the store can enhance its sales strategies and more effectively align with customer preferences.

Exploratory Analysis of data

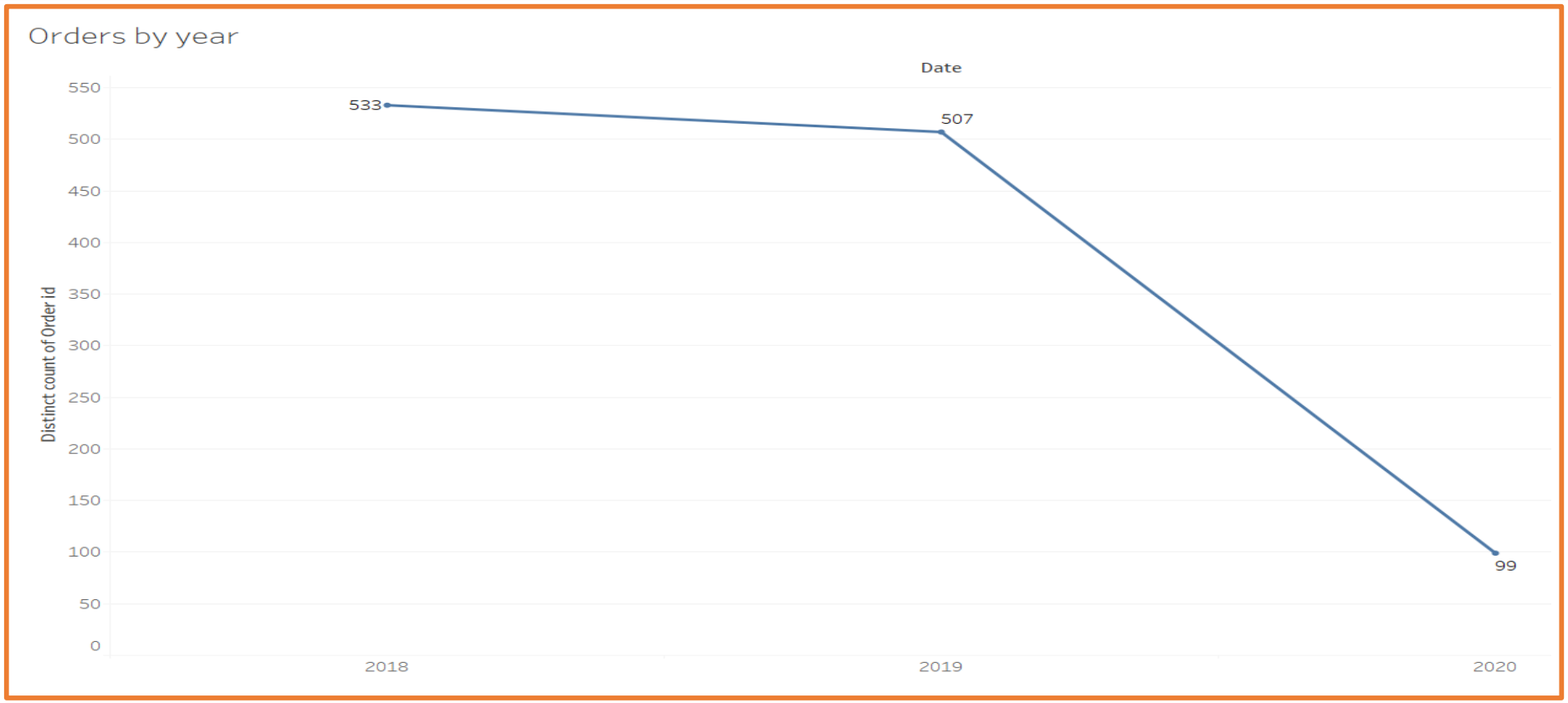
Product Frequency

- The treemap analysis reveals the top most occurring products in the dataset, with Poultry having the highest count of 640, followed by Soda with a count of 597, Cereal with 591, Ice-cream with 579, Cheese with 578, and Waffles with 575. These findings highlight the popularity and demand for these specific products.



Orders by Year

- The data indicates that order activity remained relatively stable from 2018 to 2019, with a slight decline observed. However, a deeper analysis is needed to interpret the trends for 2020, as the dataset only covers the first two months of that year and may not reflect the full annual pattern.



Orders by Quarterly

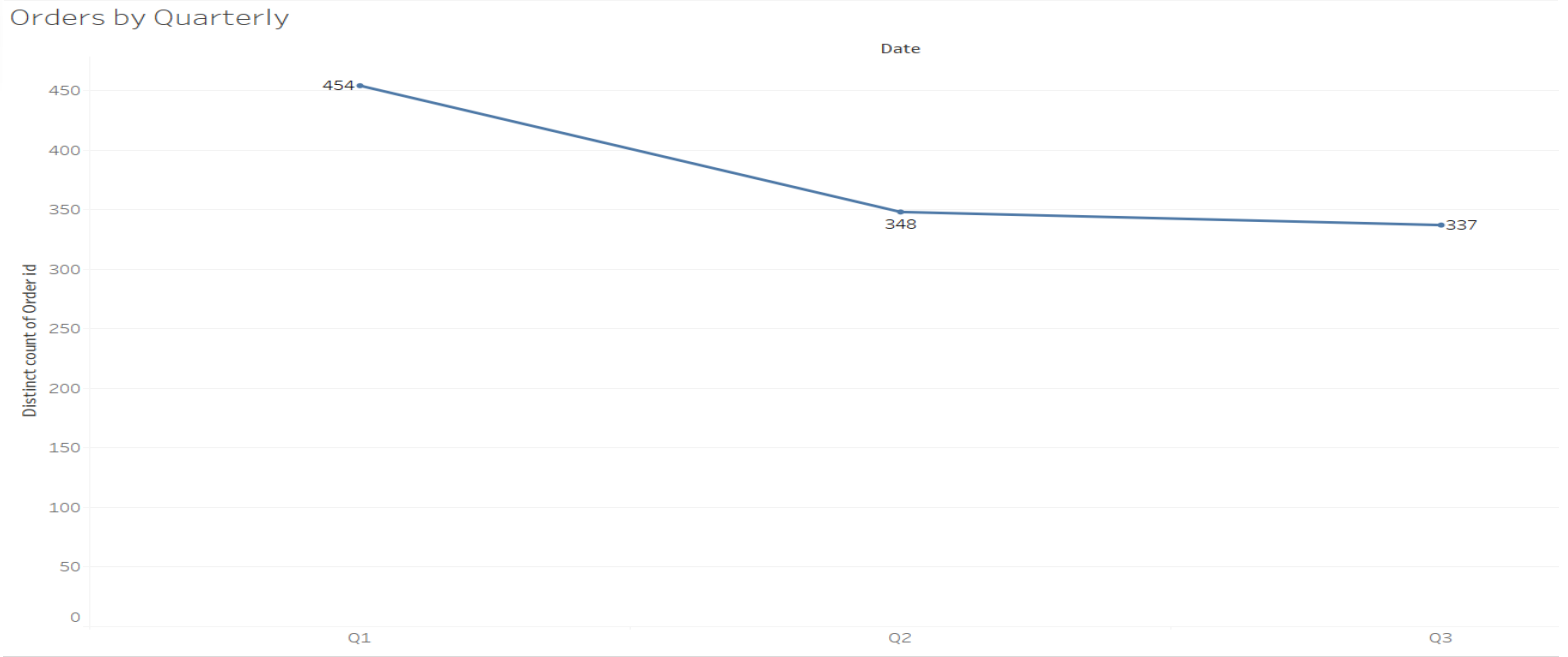
Insights Summary

Chart 1 (Year-wise Quarterly Orders):

- 2018: Gradual rise (Q1–Q3: 175→180).
- 2019: Steady decline (Q1–Q3: 180→157).
- 2020: Only Q1 (99 orders); incomplete data.
- Q1 orders dropped in 2020 vs prior years.

Chart 2 (Overall Quarterly Orders):

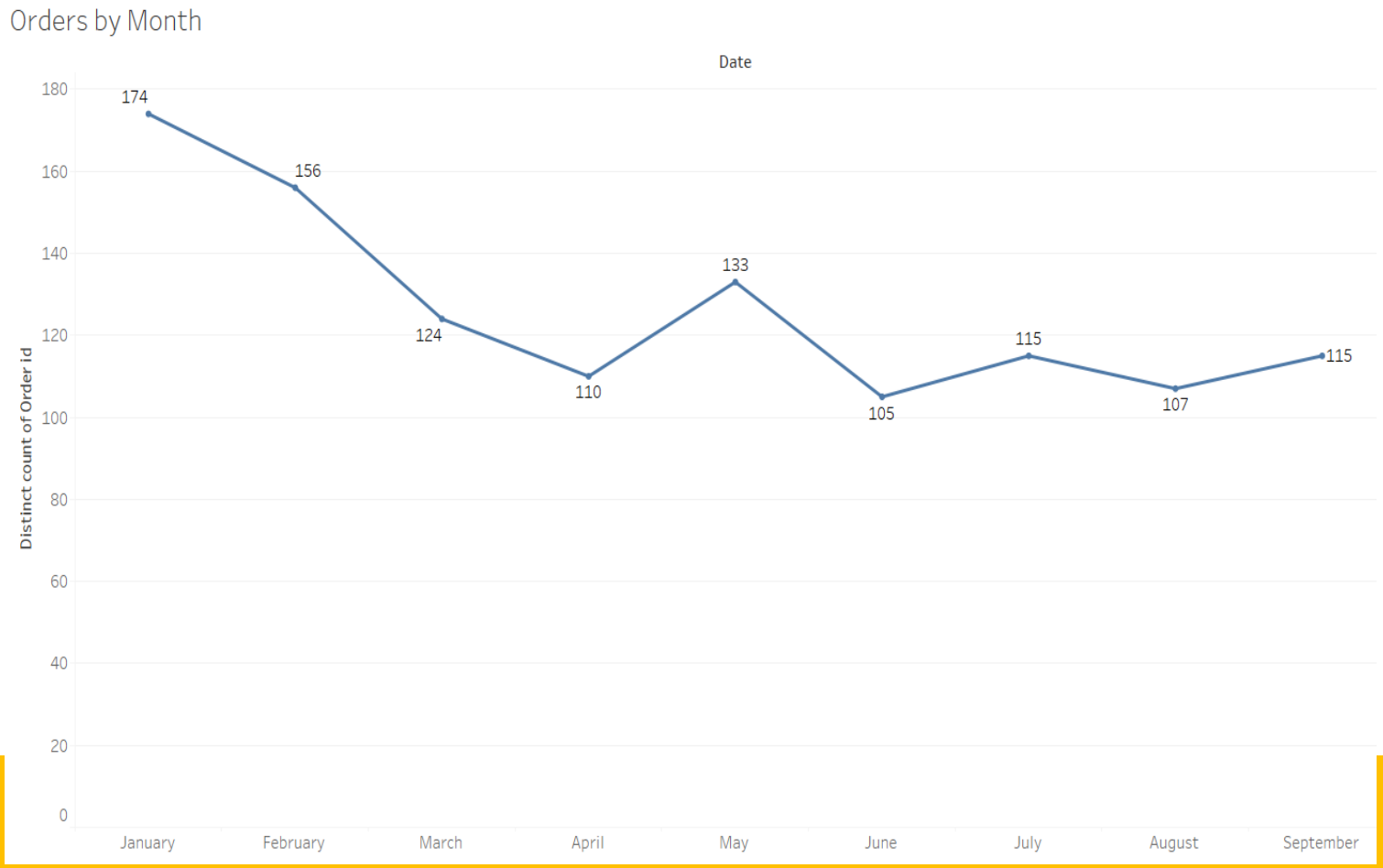
- Q1–Q2: Sharp fall (454→348).
- Q2–Q3: Minor dip (348→337).
- Q1 has highest activity;
- Q3 is lowest.

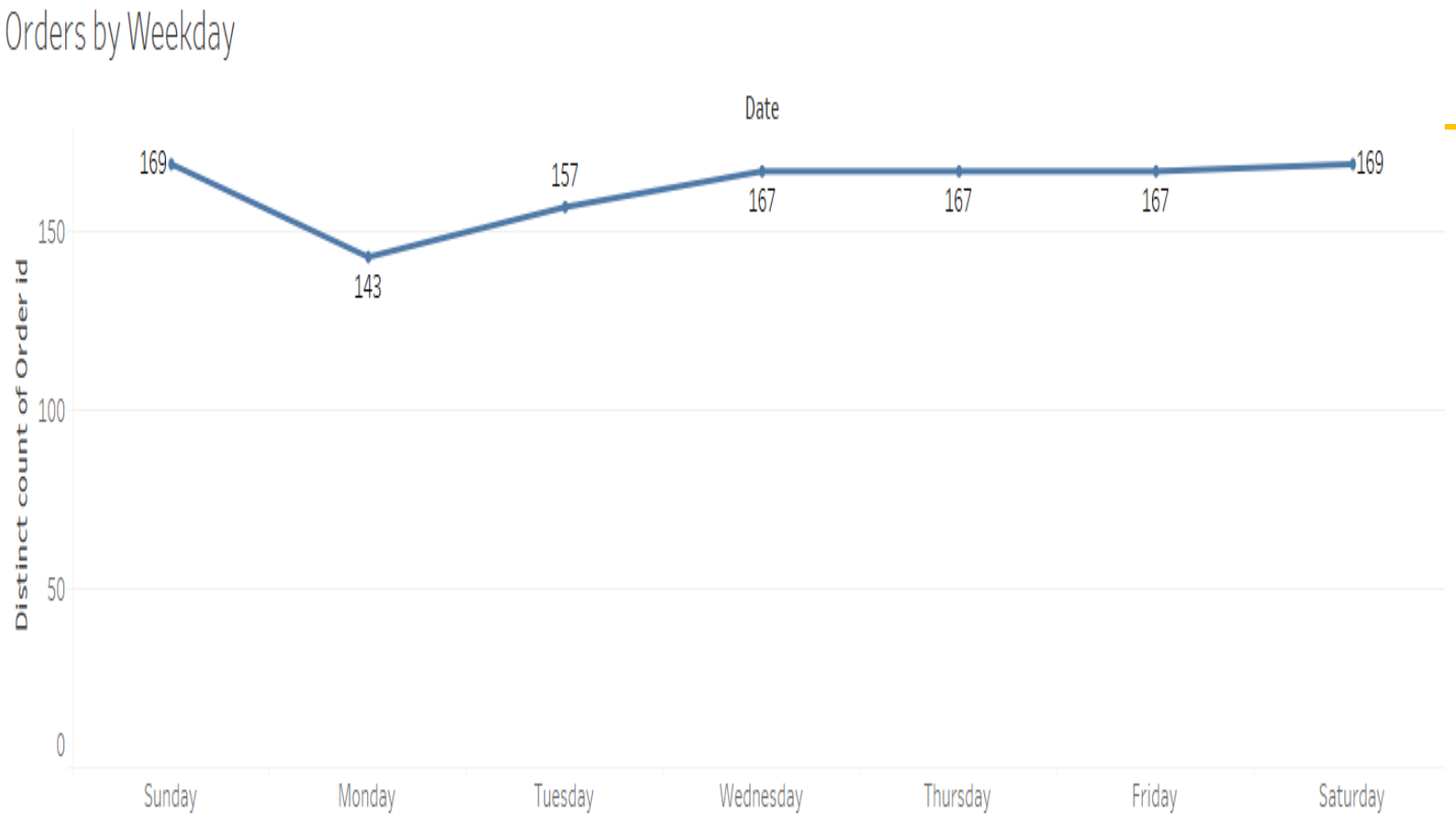


Orders by Month

Customer ordering patterns vary across the year.

- January saw the highest number of unique orders.
- Followed by February and May with strong order volumes.
- June recorded the lowest number of unique orders.
- Order activity remained relatively stable during summer months (June–August).
- A slight increase in orders was observed in September.



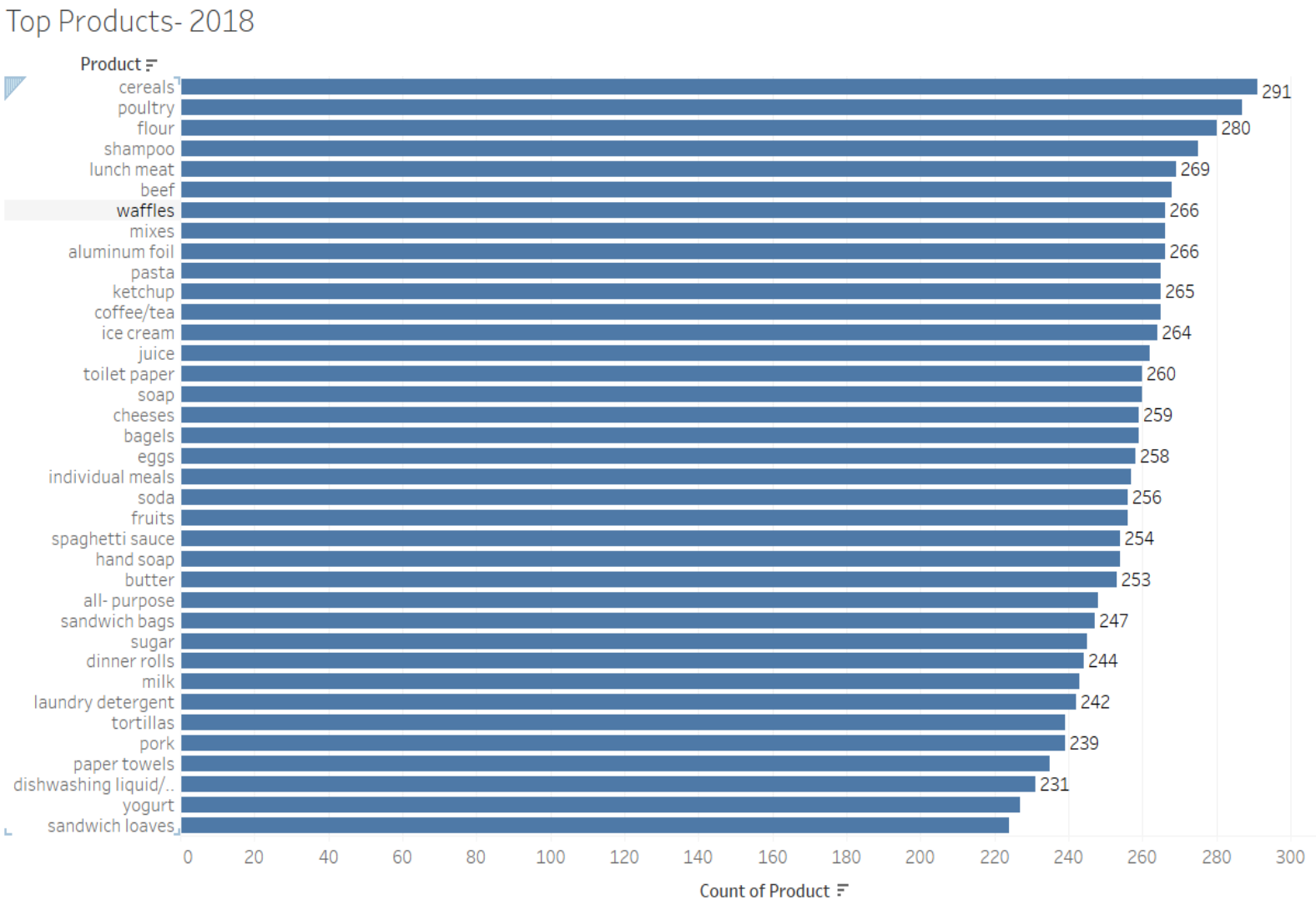


Orders by Weekday

- **Order activity is relatively consistent across weekdays.**
- **Wednesday, Thursday, and Friday show slightly higher unique orders compared to other weekdays.**
- **Saturday and Sunday have the highest number of unique orders.**
- **This indicates greater customer engagement and purchasing behavior on weekends.**

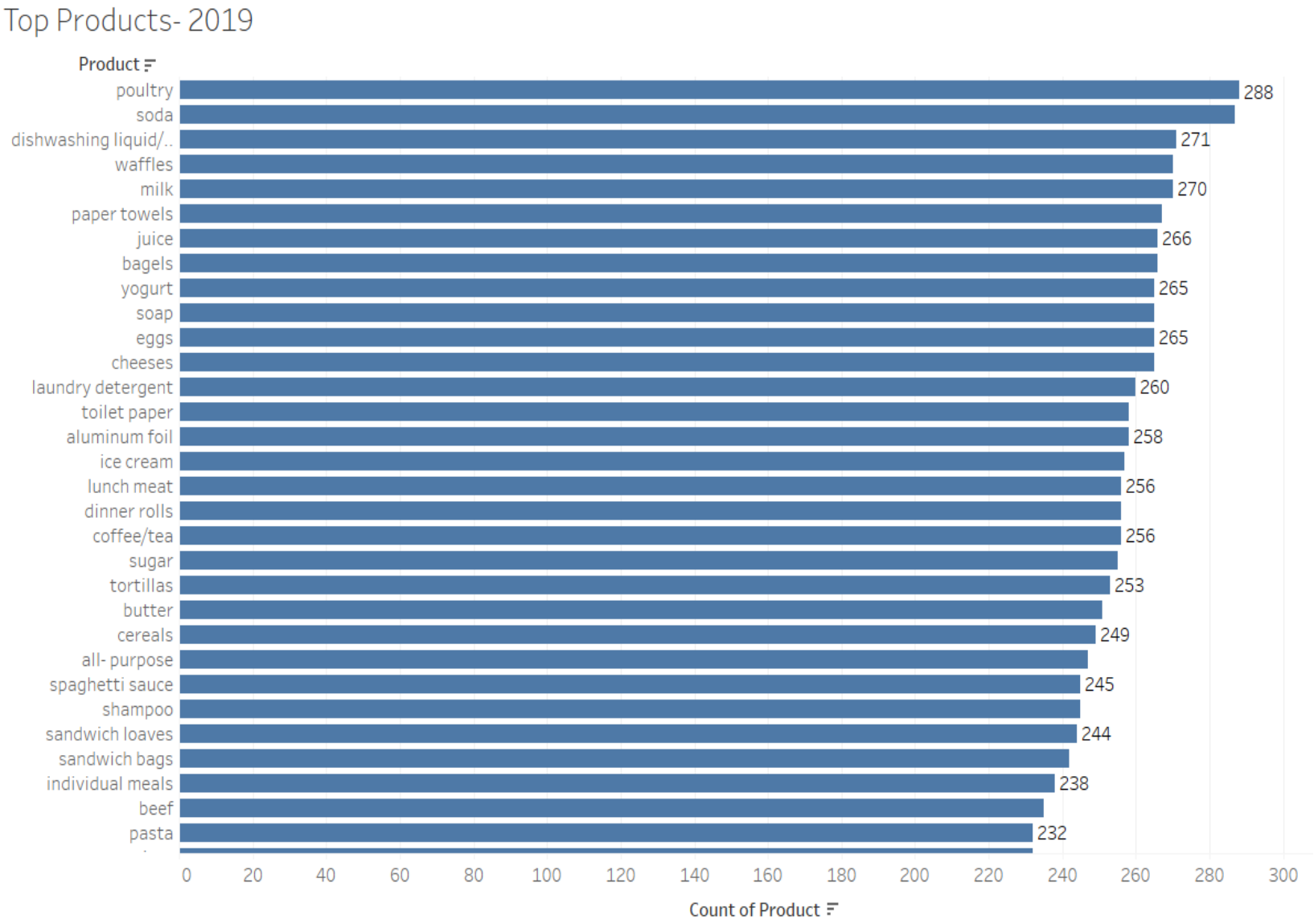
Top ordered products- 2018

- In 2018, cereals were the most frequently ordered product with 291 orders.
- Poultry followed closely with 287 orders.
- Flour ranked third with 280 orders.
- Shampoo had 275 orders, showing high demand among non-food items.
- Lunch meat completed the top five with 269 orders.
- These products reflect strong customer preference and consistent market demand throughout the year.



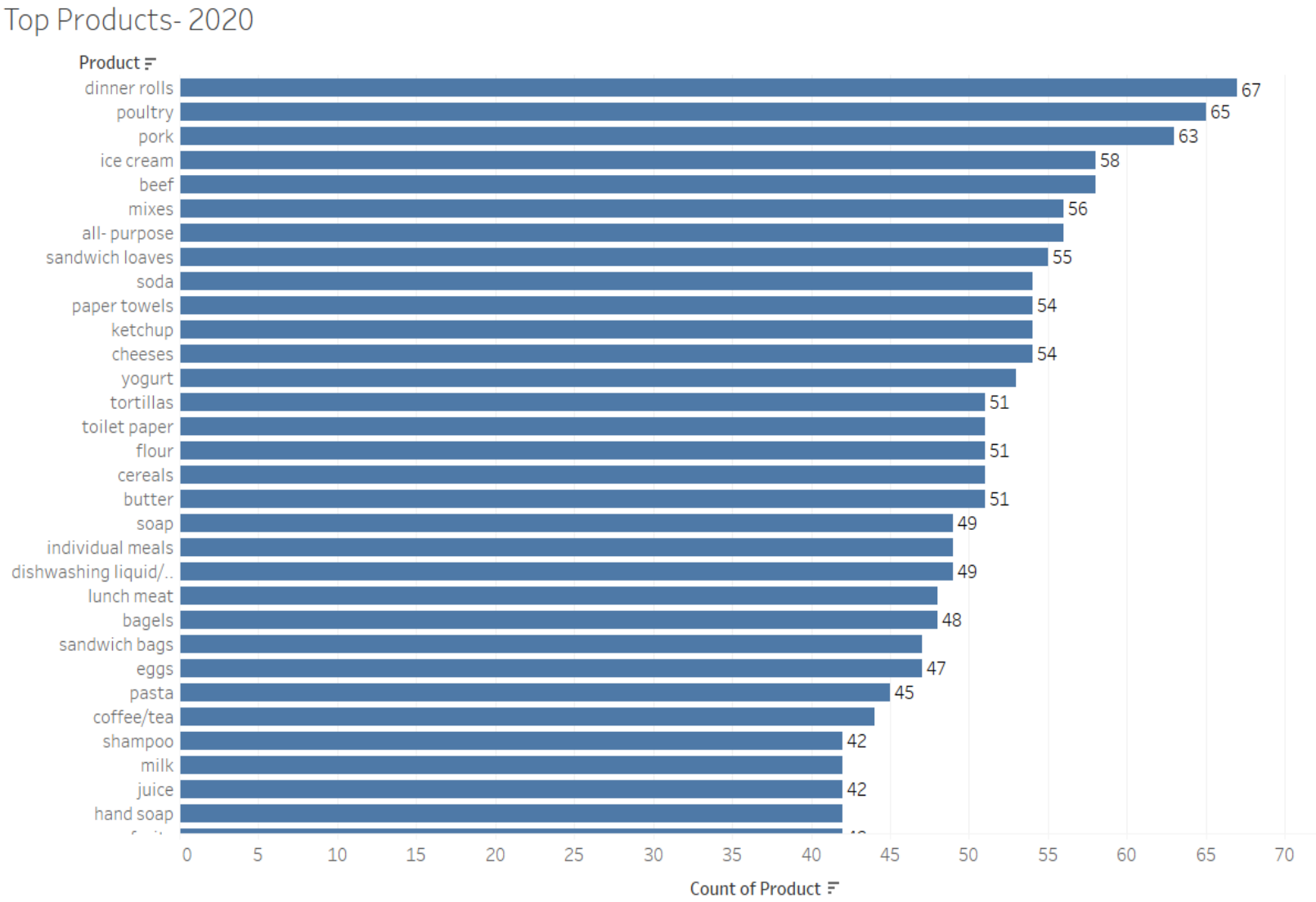
Top ordered products- 2019

- In 2019, poultry was the most ordered product with 288 orders.
- Soda was a close second with 287 orders.
- Dishwashing liquid received 271orders, showing strong demand for household items.
- Waffles and milk both recorded 270 orders.
- These top products highlight strong customer preference and steady demand throughout the year.



Top ordered products- 2020

- In 2020, dinner rolls were the most ordered product with 67 orders.
- Poultry followed closely with 65 orders.
- Pork received 63 orders, indicating strong customer interest.
- Ice cream and beef were also popular, each with 58 orders.
- These products reflected high customer demand and remained popular choices throughout the year.



Summary

- Poultry, soda, cereal, ice cream, cheese, and waffles were the most popular products.
- Order activity was stable from 2018 to 2019, with a slight dip.
- Incomplete 2020 data limits full trend analysis.
- Unique orders declined from Q1 to Q3, hinting at reduced activity.
- January, February, and May had the most unique orders; June had the fewest.
- Weekdays showed steady orders, peaking midweek; weekends had the highest.
- Product trends shifted over the years:
- 2018: High demand for cereals, poultry, flour, shampoo, lunch meat.
- 2019: Poultry, soda, dishwashing liquid, waffles, milk gained popularity.
- 2020: Focus moved to dinner rolls, poultry, pork, ice cream, beef.
- Overall, customer preferences evolved from basics to indulgent items.

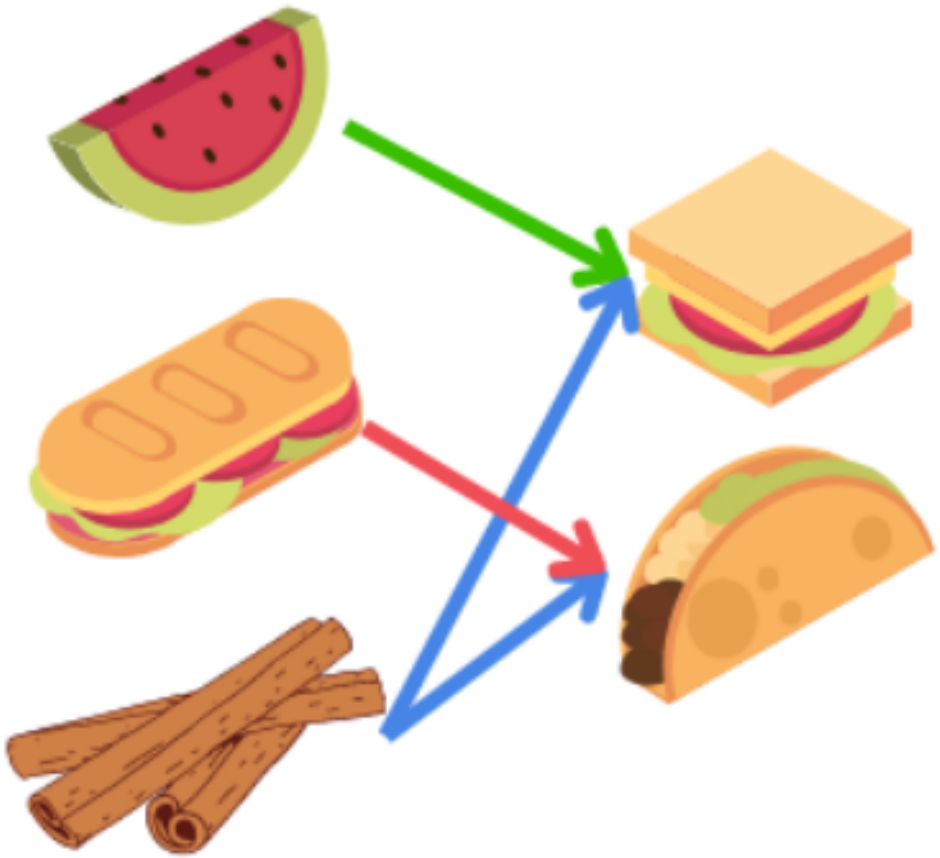
2. Market Basket Analysis

- **Write Something about the association rules and its relevance in this case**
- **Add KNIME workflow image**
- **Write about threshold values of Support and Confidence**

Market Basket Analysis

- Market Basket Analysis identifies patterns and associations between products often bought together.
- It helps businesses understand customer buying behavior and product relationships.
- The technique uses transactional data to uncover item co-occurrence and generate association rules.
- Insights from the analysis support strategies like product bundling, cross-selling, and targeted marketing.
- It enables data-driven decisions to boost sales and revenue.
- Understanding product combinations improves customer satisfaction.
- Overall, it supports business growth by aligning offerings with customer preferences.

KNIME workflow image

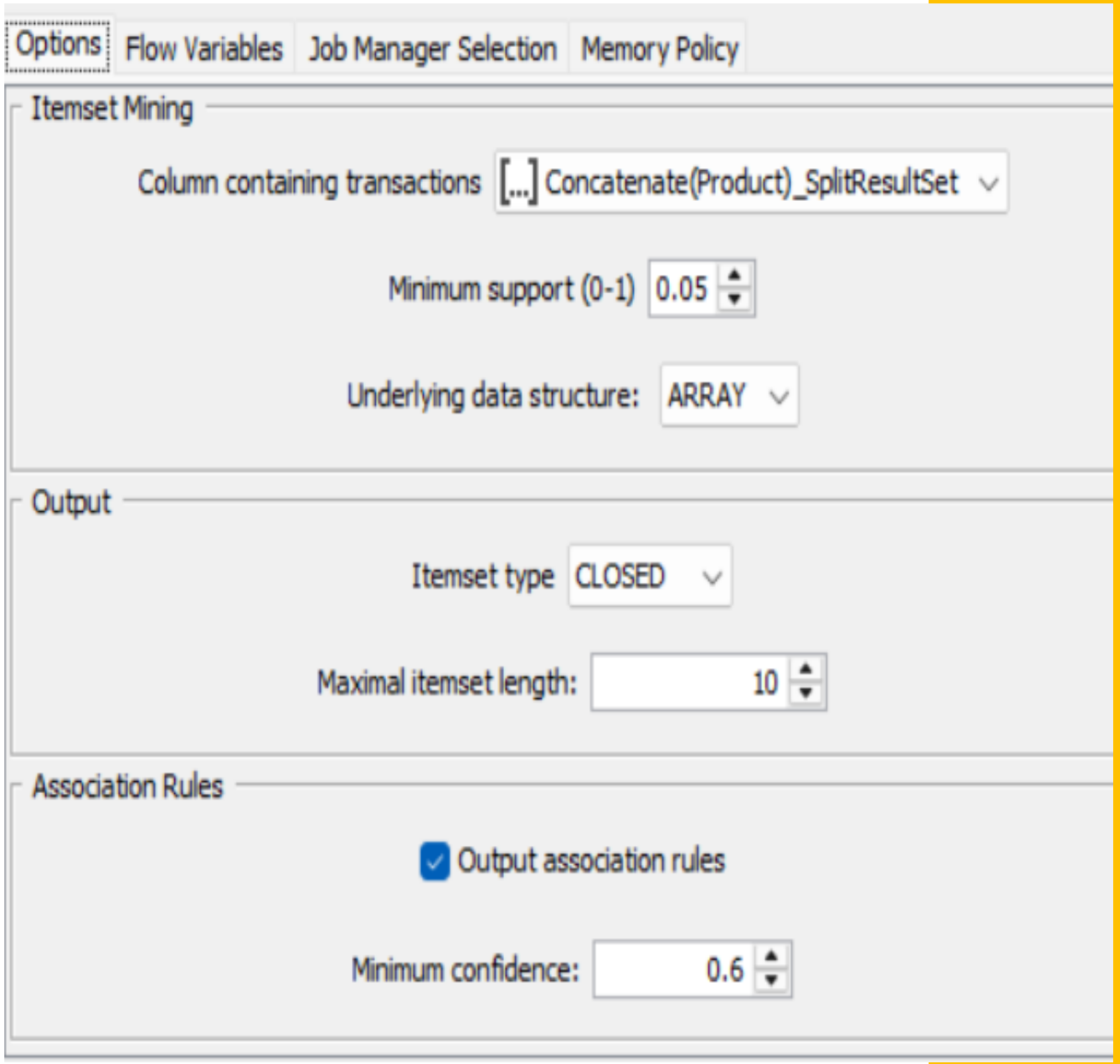


Association Rules & its relevance.

- Association rules in Market Basket Analysis show how items are frequently bought together, offering insights into customer preferences and behavior.
- These rules help businesses improve product placement, design targeted promotions, and apply cross-selling and upselling strategies.
- Using these insights can boost customer satisfaction and drive higher revenue.

Threshold values of Support and Confidence

- Threshold value for Minimum Support is 0.05
- Threshold value for Minimum Confidence is 0.6
- 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.



3. Associations Identification

- **Put the associations in a tabular manner**
- **Explain about support, confidence, & lift values that are calculated**

Association represented in tabular form

#	RowID	Support *Number (double)*	Confidence *Number (double)*	Lift *Number (double)*	Recommended_item *String*	Recommended_with *String*	Items_list *String*
1	rule0	0.05	0.64	1.7	juice	<---	[yogurt, toilet paper, aluminum fo
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, la
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, cer
11	rule10	0.051	0.604	1.589	milk	<---	[poultry, laundry detergent, cerea
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 crea
14	rule13	0.052	0.686	1.628	poultry	<---	[dinner rolls, spaghetti sauce, ice
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, juic
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
20	rule19	0.054	0.656	1.556	poultry	<---	[dinner rolls, spaghetti sauce, lau
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]

Support, Confidence, and Lift Values: Metrics for Association Analysis

- **Support** indicates how frequently an itemset appears in the dataset, showing how common a combination of items is in customer transactions.
- **Confidence** reflects the probability that a customer who buys one item will also buy another, calculated by dividing transactions with both items by transactions with the first item.
- **Lift** shows the strength of the relationship between two items, comparing their joint purchase probability to their individual probabilities. A lift above 1 indicates a strong, positive association.

4. Suggestion of Possible Combos with Lucrative Offers

- Write recommendations

Make discount offers or combos (or buy two get one free) based on the associations and your experience

Recommendations

- Combo Deal: Offer a discounted combo that includes yogurt, poultry, aluminum foil, and juice to encourage multi-item purchases.
- Buy Two Get One Free: Apply a “Buy 2 Get 1 Free” deal on dinner rolls, spaghetti sauce, and ice cream to boost bundled sales.
- Bundle Promotion: Create a discounted bundle featuring paper towels, eggs, and pasta to attract value-seeking customers.
- Cross-Selling Offer: Offer discounts on cheese, bagels, and sandwich bags when customers buy cereals to promote related item purchases.
- Limited-Time Promotion: Provide percentage savings for customers who buy poultry, laundry detergent, and mixes during a limited-time offer.
- Loyalty Program: Launch a rewards program for frequent buyers of recommended combos, encouraging repeat purchases and long-term loyalty.
- These strategies are based on identified item associations and aim to boost revenue, customer satisfaction, and product exploration.