



# Marketing and Retail Analysis – Project 2

## Grocery Store Data

Soumalya Sen

BATCH: PGPDSBA.O.JULY23.A



## **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.

# 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 analyse the Point of Sale (POS) data from a grocery store to identify the frequently occurring combinations of items in customer orders. Provide actionable recommendations for implementing popular combo offers and discounts, aiming to increase the store's revenue and enhance 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:** 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 corresponds to the purchase of different products, 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.
- **Market Basket Analysis** helps the grocery store discover which items are commonly bought together by customers. This information can be used to create special deals and discounts on popular combinations of products, encouraging customers to buy related items. By using market basket analysis, the store can improve its sales strategies and better meet the preferences of its customers

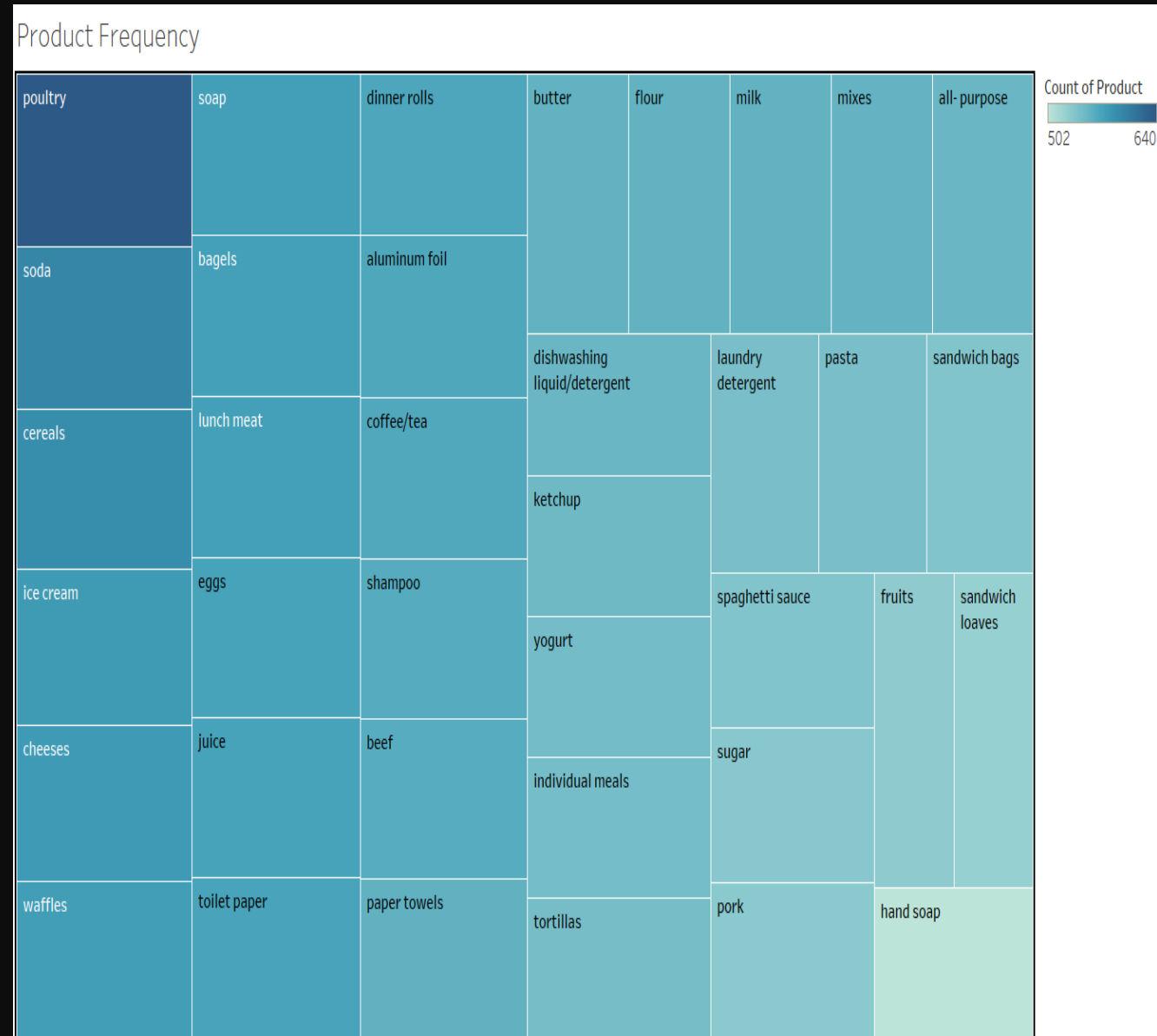


# Exploratory Analysis of data



## Product Frequency

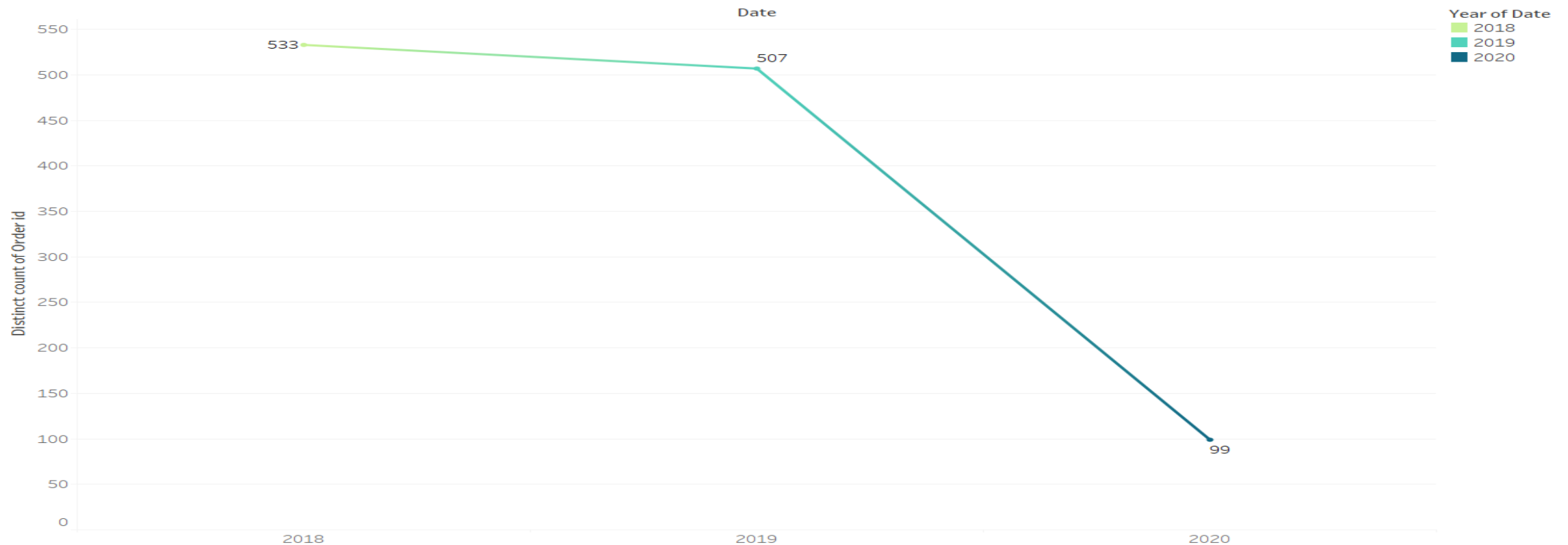
- The treemap analysis reveals the top most occurring products in the dataset, with Poultry having the highest count of 640, followed by Soda with a count of 597, Cereal with 591, Ice-cream with 579, Cheese with 578, and Waffles with 575. These findings highlight the popularity and demand for these specific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



## Orders by Year

- Based on the below data, it can be inferred that there was a relatively stable level of order activity from 2018 to 2019, with a slight decrease observed. However, further analysis is required to understand the trends and patterns in the order data for 2020, since it is important to consider that the dataset only includes data for the first two months of 2020.

Orders by Year

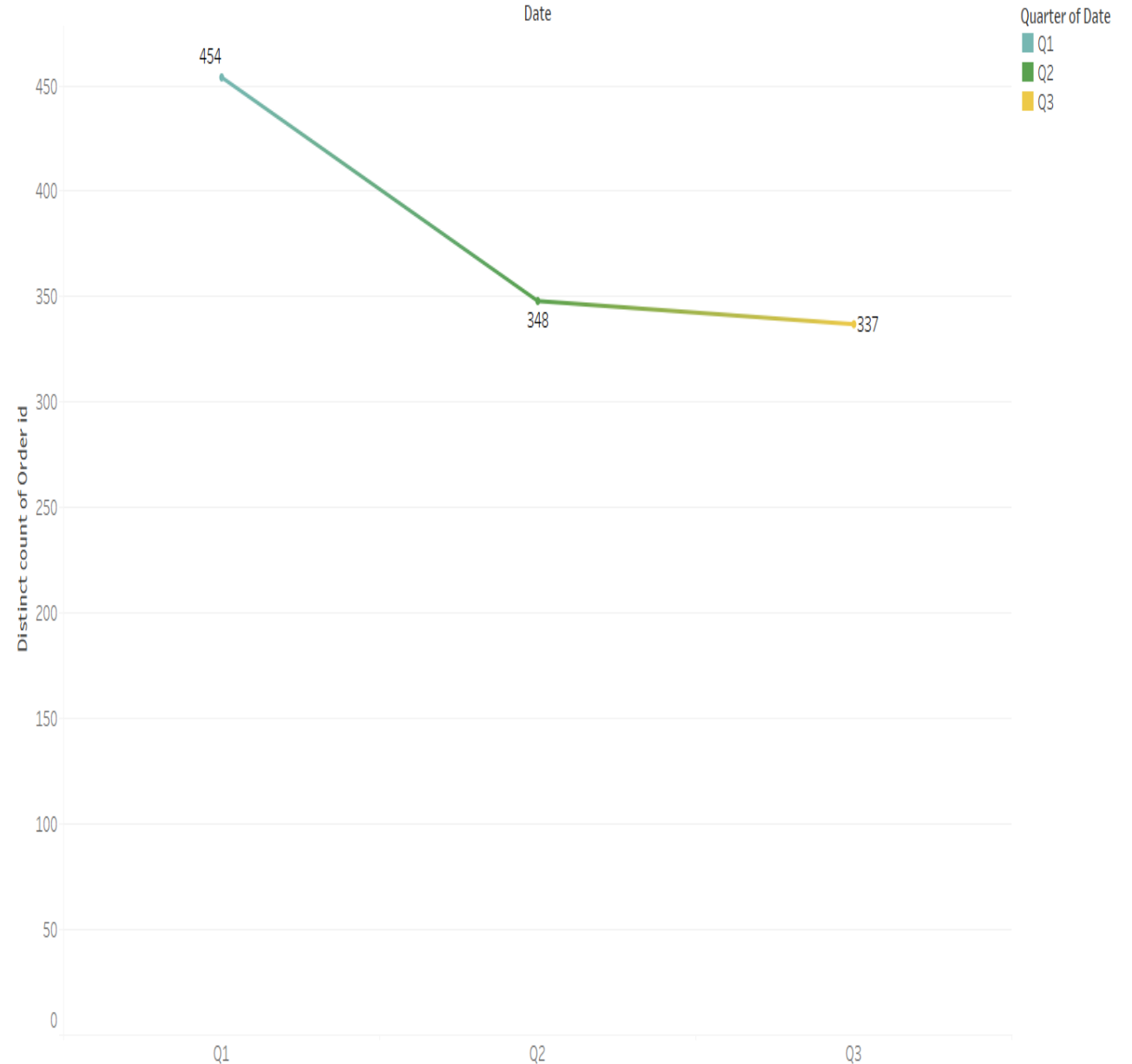




## Orders by Quarterly

- It can be inferred that there was a general downward trend in the number of unique order IDs from Q1 to Q3. This may indicate a potential decrease in overall order activity during this period. However, further analysis is required to understand the complete picture of order trends throughout the year, especially in the absence of Q4 data.

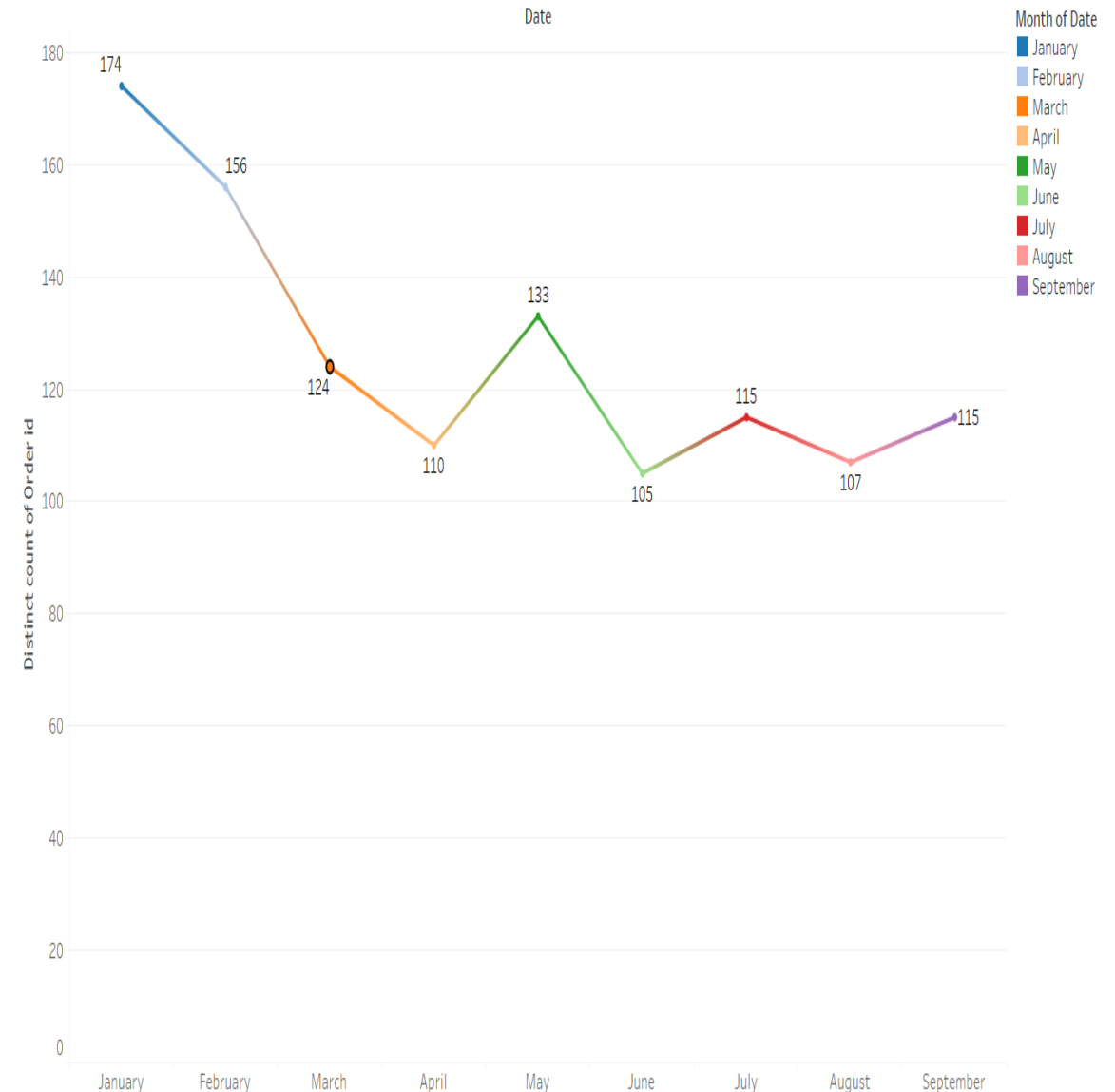
Orders by Quarterly



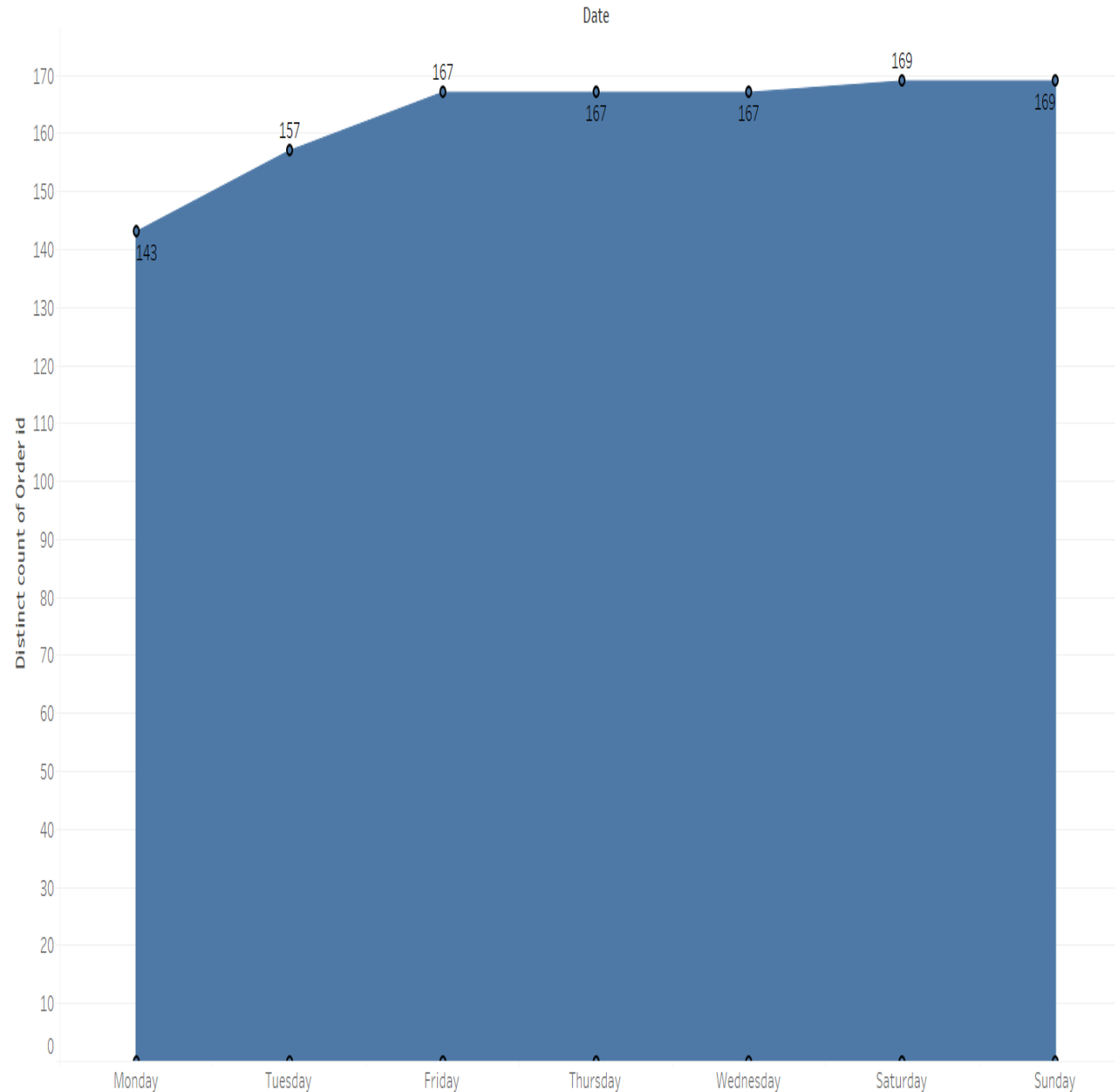
# Orders by Month

- The data indicates variations in customer ordering patterns throughout the year.
- The highest number of unique orders was observed in January, followed by February and May. June had the lowest count of unique orders. It is worth noting that there is a relatively consistent level of order activity during the summer months (June, July, and August) and a slight increase in September

Orders by Month



Orders by weekday



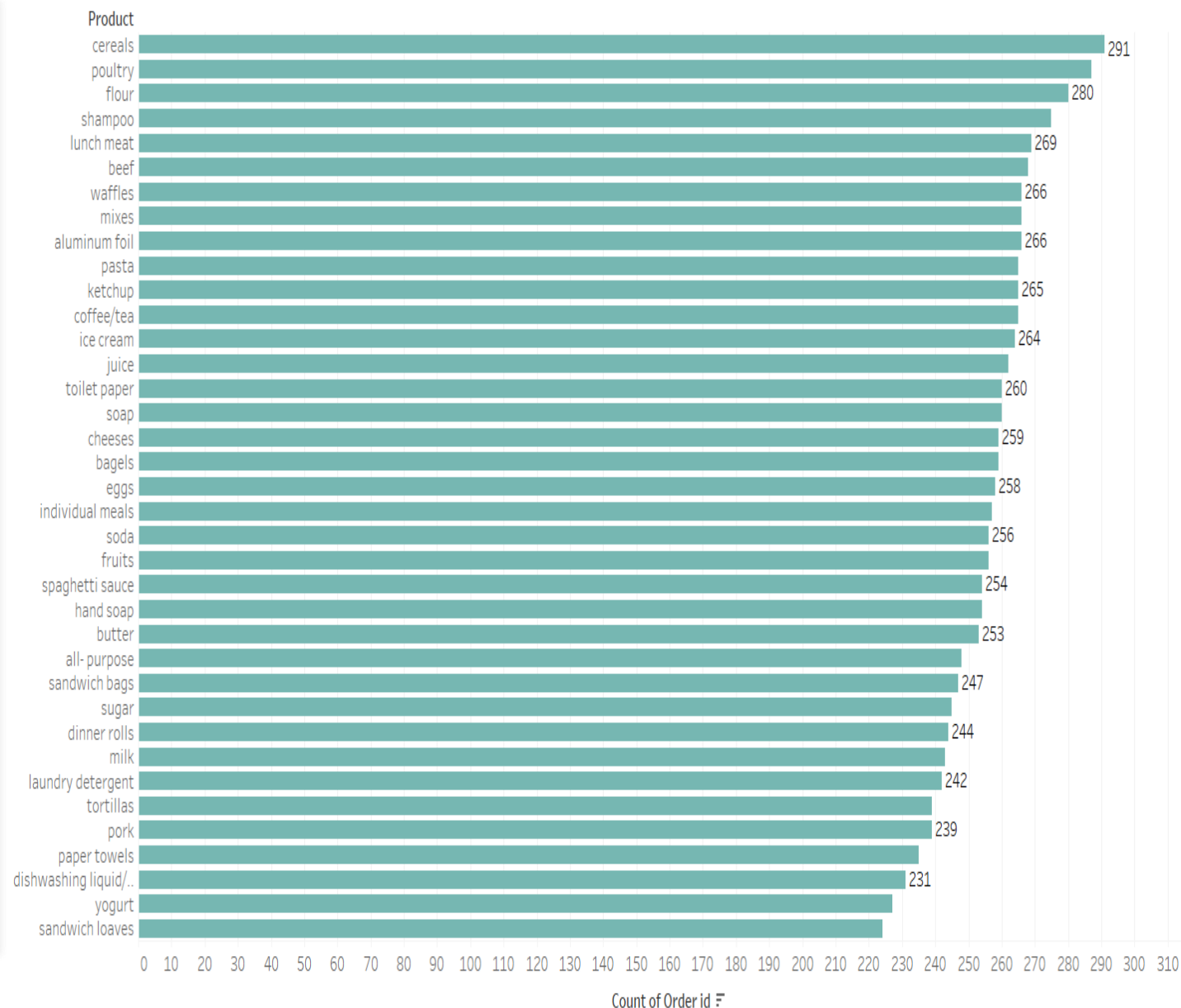
## Orders by Weekday

- This data suggests that there is relatively consistent order activity throughout the weekdays, with Wednesday, Thursday, and Friday showing a slightly higher number of unique orders. Saturdays and Sundays have the highest count of unique orders, indicating increased customer engagement and purchasing behaviour over the weekends.

# Top ordered products- 2018

- In 2018, the top five products that were most frequently ordered by customers were cereals, poultry, flour, shampoo, and lunch meat. Cereals ranked first with 291 orders, followed closely by poultry with 287 orders. Flour, shampoo, and lunch meat also had significant demand, with 280, 275, and 269 orders respectively. These products demonstrated their popularity and appeal to customers, indicating a strong market demand for these items during that year.

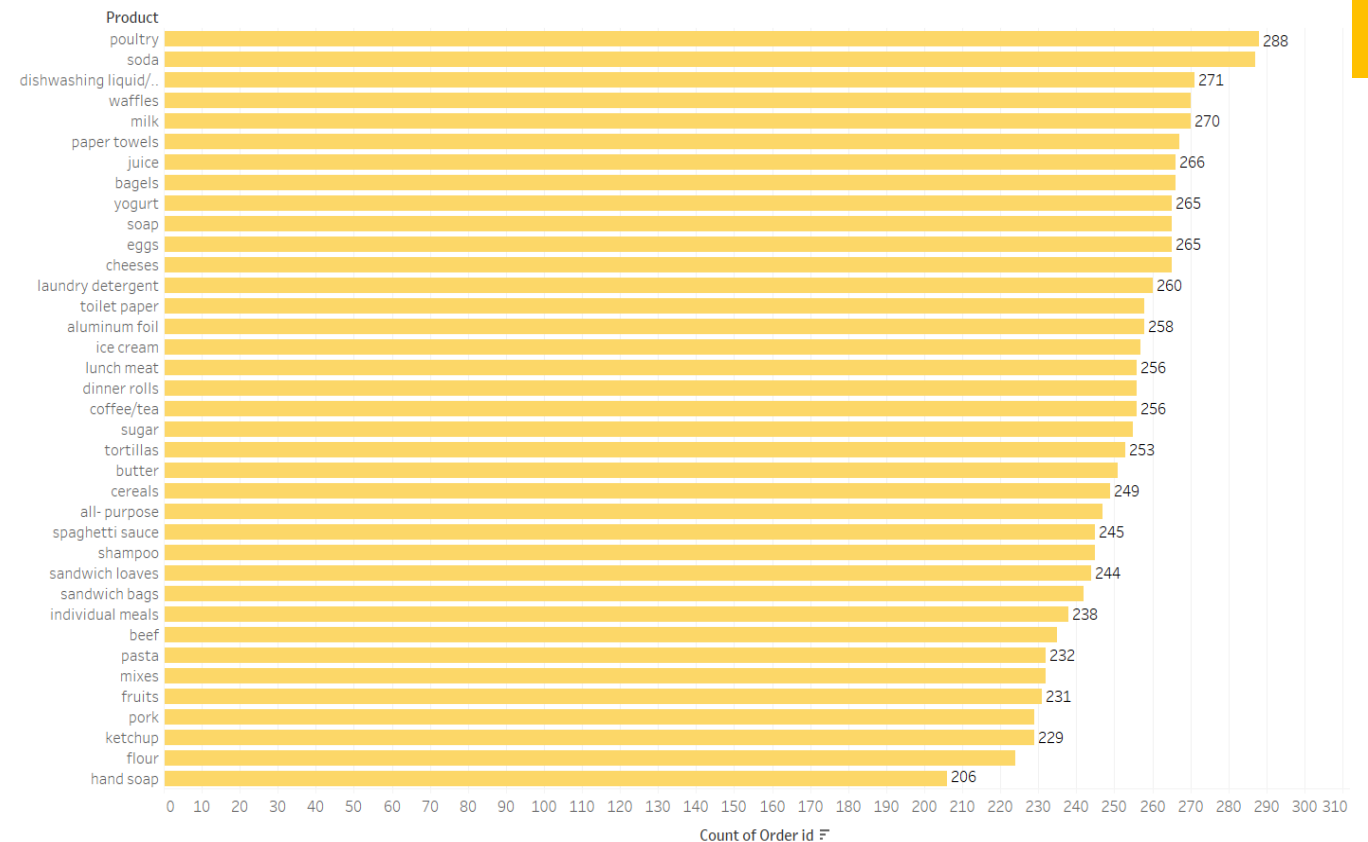
2018-Top Products



# Top ordered products- 2019

- In 2019, the top five products that were highly ordered by customers were poultry, soda, dishwashing liquid, waffles, and milk. Poultry took the lead with 288 orders, closely followed by soda with 287 orders. Dishwashing liquid, waffles, and milk also had a substantial number of orders, with 271, 270, and 270 respectively. These products showcased their popularity and customer preference, indicating a strong demand for them throughout the year.

2019-Top Products

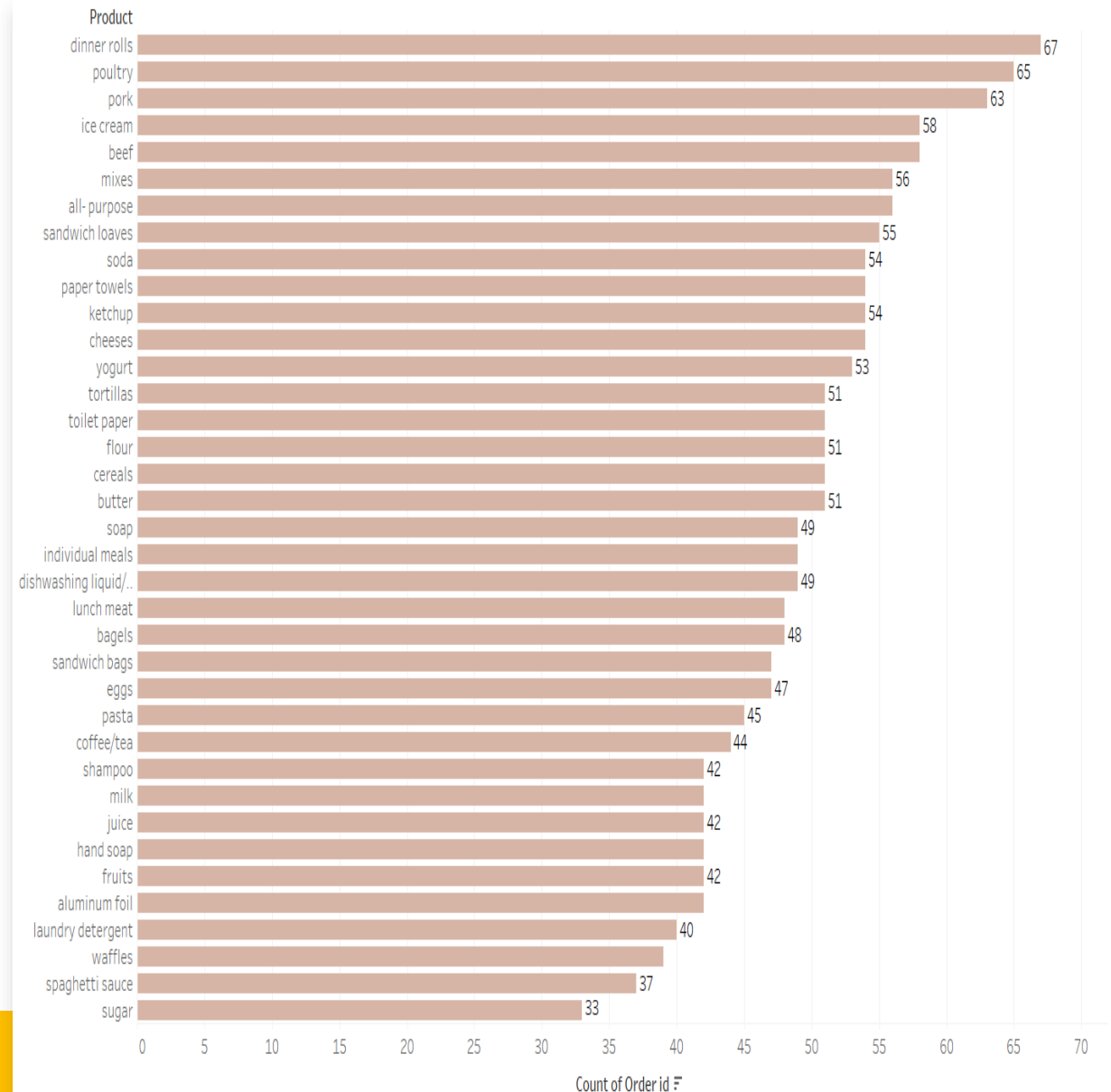




# Top ordered products- 2020

- In 2020, the top five products that were highly ordered by customers were dinner rolls, poultry, pork, ice cream, and beef. Dinner rolls took the lead with 67 orders, closely followed by poultry with 65 orders. Both pork and ice cream were popular choices, each receiving 63 and 58 orders respectively. Beef also had a notable number of orders with 58. These products demonstrated their appeal and customer demand, showcasing their popularity of 2020.

2020-Top Products





# Summary

- Poultry, Soda, Cereal, Ice-cream, Cheese, and Waffles were the top products, indicating high popularity and demand.
- Order activity remained relatively stable from 2018 to 2019, with a slight decrease observed.
- Limited data for 2020 prevents a comprehensive understanding of order trends for that year.
- There was a general downward trend in unique order IDs from Q1 to Q3, suggesting a potential decrease in overall order activity.
- January, February, and May had the highest number of unique orders, while June had the lowest.
- Weekdays showed consistent order activity, with higher counts on Wednesday, Thursday, and Friday, and weekends had the highest overall counts, indicating increased customer engagement
- Over the course of three years, the analysis of customer orders reveals changing preferences and trends. In 2018, cereals, poultry, flour, shampoo, and lunch meat were popular choices, indicating a demand for everyday household items. In 2019, poultry, soda, dishwashing liquid, waffles, and milk took the lead, suggesting a mix of food and cleaning products. However, in 2020, there was a shift towards dinner rolls, poultry, pork, ice cream, and beef, showcasing a desire for more indulgent food options. This suggests a shift in customer preferences between the two years, indicating changing market trends.

## 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 is a technique used to identify patterns and associations among products that are frequently purchased together in customer transactions. It helps businesses understand customer behaviour and uncover product relationships, which can be used to optimize sales strategies and increase revenue. By analysing 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.





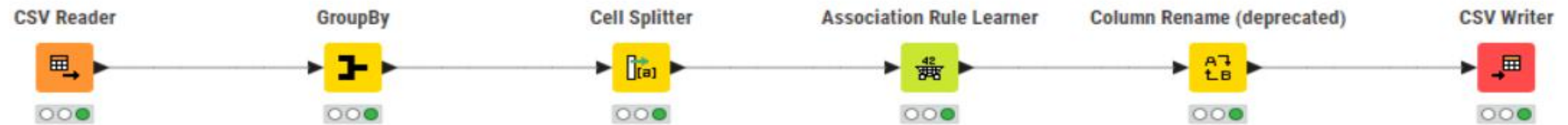
# Association Rules & its relevance.

---

- Association rules in Market Basket Analysis reveal the relationships and co-occurrence patterns between items, providing valuable insights into customer purchasing behaviour and preferences.
- 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.



# KNIME workflow image



# 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.

The screenshot shows a software window titled "Dialog - 7:4 - Association Rule Learner". It has four tabs: "Options", "Flow Variables", "Job Manager Selection", and "Memo". The "Options" tab is active and contains three sections:

- Itemset Mining**:
  - Column containing transactions: `[...] Concatenate(Product)_SplitR`
  - Minimum support (0-1):
  - Underlying data structure:
- Output**:
  - Itemset type:
  - Maximal itemset length:
- Association Rules**:
  - ☒ Output association rules
  - Minimum confidence:



# 3. Associations Identification

---

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

# Association represented in tabular form

Row ID	<b>D</b> Support	<b>D</b> Confidence	<b>D</b> Lift	<b>S</b> Recommended_item	<b>S</b> Recommended_w...	<b>S</b> Items_list
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 deter...
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]
rule11	0.052	0.628	1.61	eggs	<---	[dinner rolls, poultry, soda]
rule12	0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce, poultry, ice cream]
rule13	0.052	0.686	1.628	poultry	<---	[dinner rolls, spaghetti sauce, ice cream]
rule14	0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce, poultry, juice]
rule15	0.052	0.602	1.429	poultry	<---	[dinner rolls, spaghetti sauce, juice]
rule16	0.052	0.634	1.627	eggs	<---	[paper towels, dinner rolls, pasta]
rule17	0.052	0.602	1.621	pasta	<---	[paper towels, eggs, dinner rolls]
rule18	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce, poultry, laundry detergent]
rule19	0.054	0.656	1.556	poultry	<---	[dinner rolls, spaghetti sauce, laundry deterg...
rule20	0.055	0.624	1.565	ice cream	<---	[paper towels, eggs, pasta]
rule21	0.055	0.63	1.616	eggs	<---	[paper towels, ice cream, pasta]
rule22	0.055	0.643	1.731	pasta	<---	[paper towels, eggs, ice cream]
rule23	0.055	0.649	1.791	paper towels	<---	[eggs, ice cream, pasta]



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

- **Support:** The support value represents the frequency or popularity of an itemset in the dataset. It indicates how often a specific combination of items appears together in customer transactions.
- **Confidence:** Confidence measures the likelihood that a customer who buys one item will also purchase another item. It is calculated as the ratio of the number of transactions where both items are purchased together to the number of transactions where the first item is purchased.
- **Lift:** Lift measures the strength of association between two items in an association rule. It compares the probability of the two items being purchased together to the probability of them being purchased independently. A lift value greater than 1 suggests a positive association, indicating that the items are more likely to be purchased together.





## 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 (discount offers/ combos)

- **Combo Deal:** Offer a special combo deal where customers can buy yogurt, poultry, and aluminium foil along with juice to avail a discounted price or additional item.
- **Buy Two Get One Free:** Introduce a "buy two get one free" offer on dinner rolls, spaghetti sauce, and ice cream to incentivize customers to purchase these items together.
- **Bundle Promotion:** Create a bundle promotion where customers can buy paper towels, eggs, and pasta together at a discounted.
- **Cross-Selling Offer:** Provide a cross-selling offer where customers purchasing cereals can get a discount on cheese, bagels, and sandwich bags.
- **Limited-Time Promotion:** Launch a limited-time promotion where customers buying poultry, laundry detergent, and mixes can receive a percentage savings.
- **Loyalty Program:** Implement a loyalty program where customers who frequently purchase recommended items or participate in the suggested combos can earn rewards or exclusive discounts. This will incentivize customer loyalty and encourage them to continue shopping with the store, fostering long-term relationships and repeat purchases.

These recommendations are based on the association rules and the occurrence of certain items together, aiming to increase customer satisfaction and encourage them to explore additional products.







Thank You!