Marketing and Retail Analysis – Project 2

(Grocery Store Data)

ANALYSIS DONE BY

- ANAY TIWARI

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**

1. Exploratory Analysis

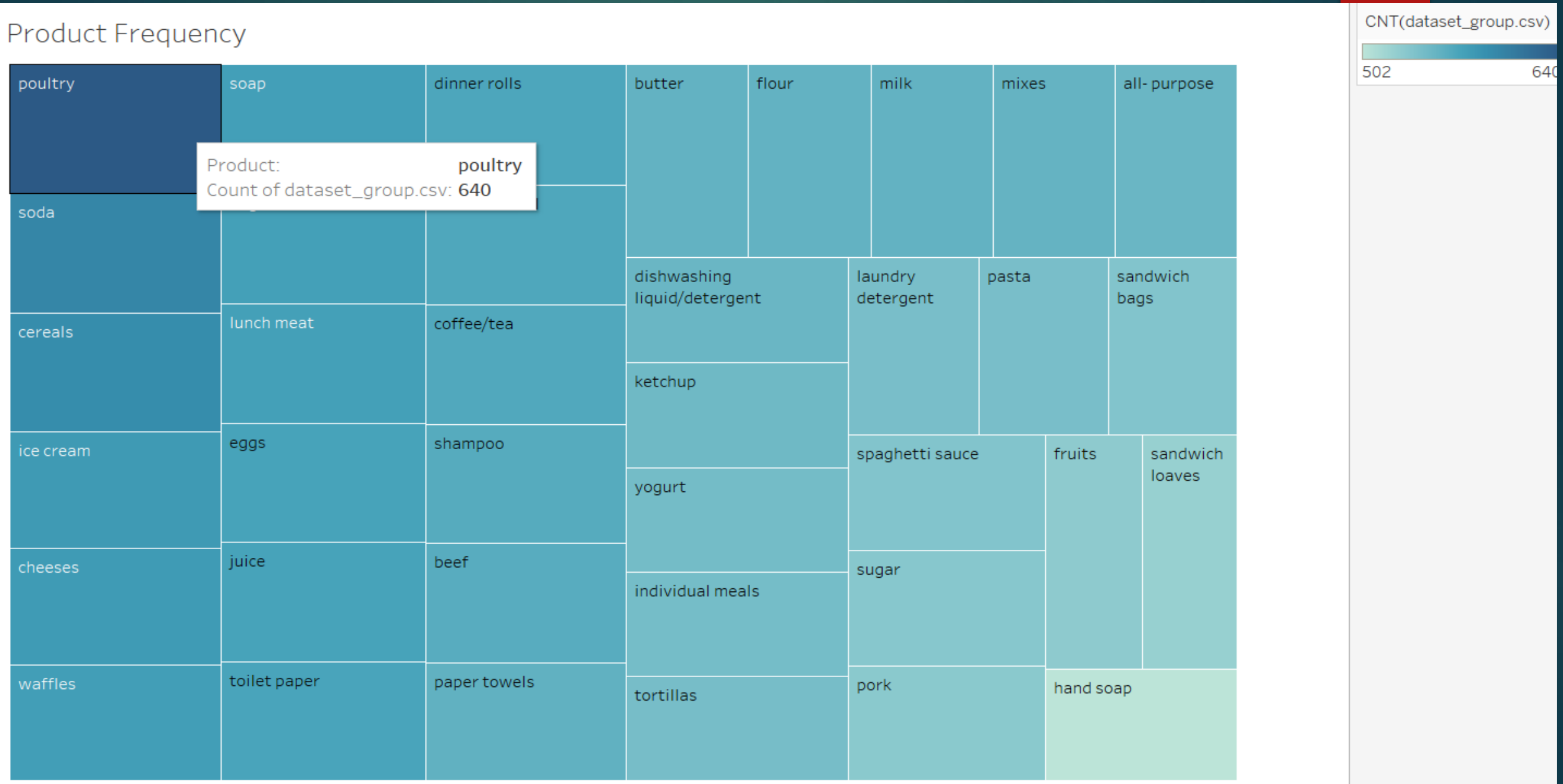
- Executive summary of the data
- Exploratory Analysis of data (Trends across months/years/quarters/days)
- Summary

Executive Summary of Data :

- **Objective:** Analyze Point of Sale (POS) data from a grocery store to identify frequently occurring item combinations in customer orders and suggest actionable recommendations for implementing popular combo offers and discounts.
- **Dataset Description:** The dataset contains 20,641 rows and 3 columns ('Date', 'Order_id', and 'Product'), representing POS data from a grocery store. There are no missing values in the dataset, and it includes 4,730 duplicate values, each corresponding to unique transactions showcasing different product purchases.
- **Utilization of Duplicate Values:** Despite the presence of duplicate values, they do not need removal as they offer insights into customer purchasing behavior. Each duplicate order ID signifies distinct product purchases, contributing valuable information for analysis.
- **Market Basket Analysis Benefits:** Market basket analysis aids in identifying commonly bought together items by customers, facilitating the creation of special deals and discounts on popular product combinations. This approach encourages customers to purchase related items, thereby enhancing sales strategies and aligning with customer preferences.

Exploratory Data Analysis :

Product Frequency



- **Treemap Analysis Results:** The analysis of the dataset using a treemap visualization shows the most frequently occurring products, with Poultry having the highest count of 640, followed by Soda (597), Cereal (591), Ice-cream (579), Cheese (578), and Waffles (575). These findings emphasize the significant popularity and demand for these specific products among customers.
- **Top Products Identified**: The treemap analysis has identified the most commonly purchased items, showcasing Poultry, Soda, Cereal, Ice-cream, Cheese, and Waffles as the top-ranking products based on their occurrence counts in the dataset.
- **High Demand Products**: The results highlight the high demand and popularity of certain products such as Poultry, indicating a strong preference among customers for these items.
- **Valuable Insights**: These findings provide valuable insights into customer preferences and can guide strategic decisions related to product promotions, inventory management, and sales strategies.

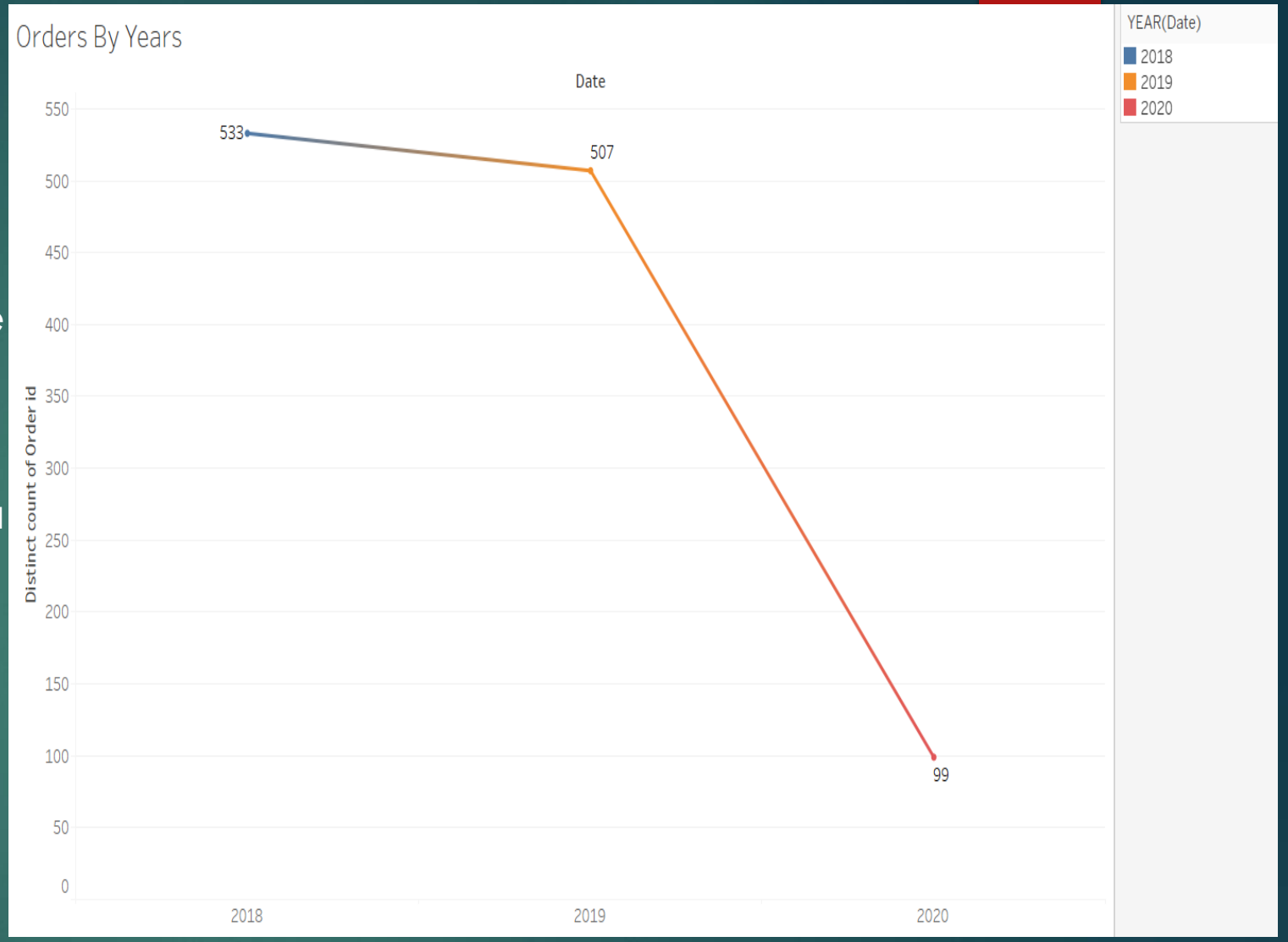
Orders by Years

•**Order Activity Analysis**: The data indicates a relatively consistent level of order activity from 2018 to 2019, showing a minor decrease over this period. However, a more in-depth analysis is necessary to fully comprehend the trends and patterns in the order data for 2020.

•**Stable Order Levels**: The dataset suggests a stable order activity trend from 2018 to 2019, with a slight downward trend observed.

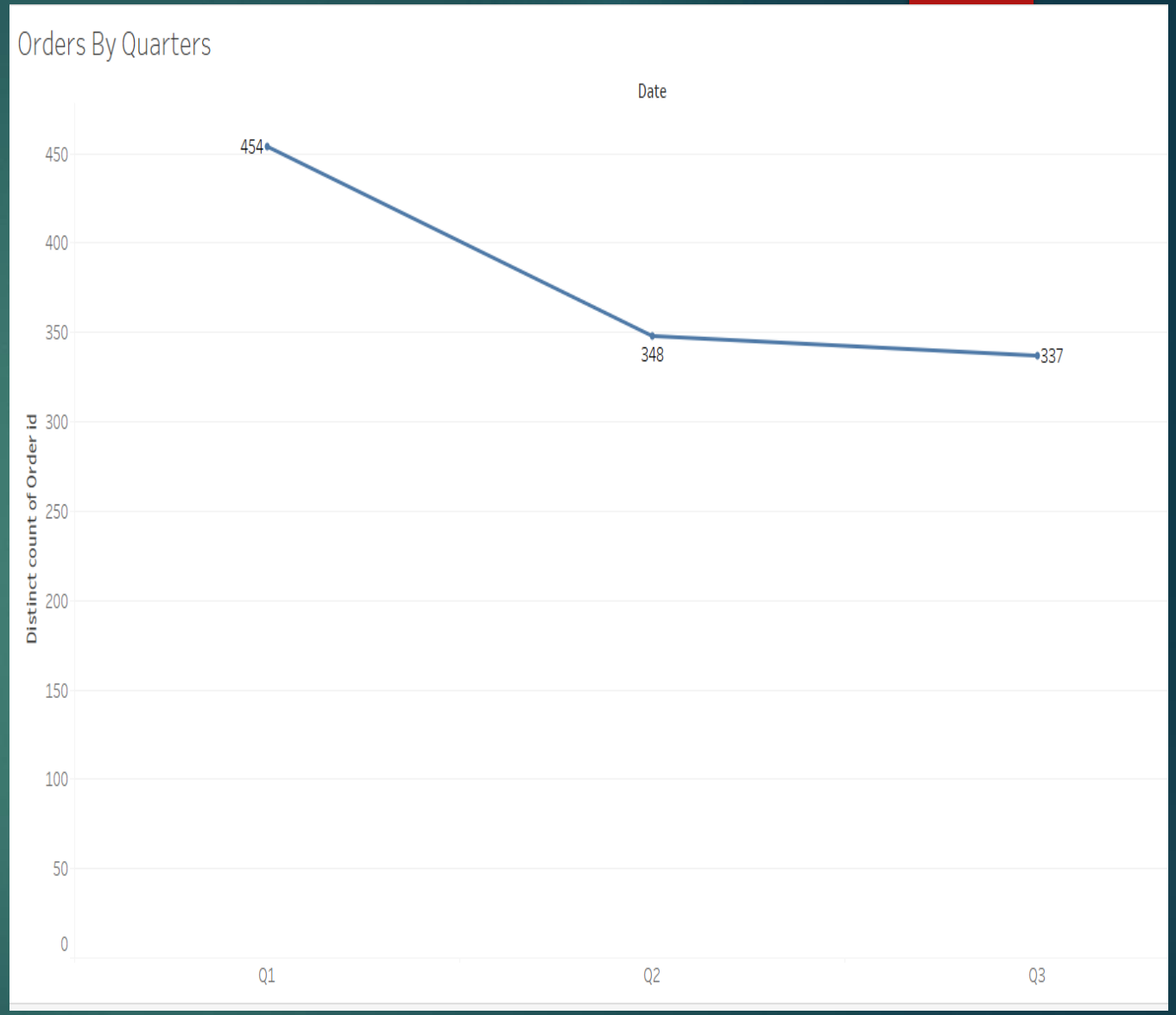
•**Limited 2020 Data**: It's crucial to note that the dataset only includes data for the first two months of 2020, which limits the understanding of the overall trends for that year.

•**Need for Further Analysis**: Due to the incomplete data for 2020, additional analysis is required to gain a comprehensive understanding of the order activity trends and patterns during that period.



Orders by Quarters

- **Downward Trend in Unique Order IDs:** An analysis of the data suggests a gradual decrease in the number of unique order IDs from Q1 to Q3, indicating a potential decline in overall order activity during this timeframe.
- **Potential Decrease in Order Activity**: The observed trend of decreasing unique order IDs may imply a reduction in overall order volume during the specified quarters.
- **Incomplete Picture Without Q4 Data**: It's important to note that the analysis lacks data for Q4, which is essential for understanding the complete order trends throughout the year.
- **Further Analysis Needed**: Given the absence of Q4 data, additional analysis is necessary to gain a comprehensive understanding of the order trends and patterns over the entire year.

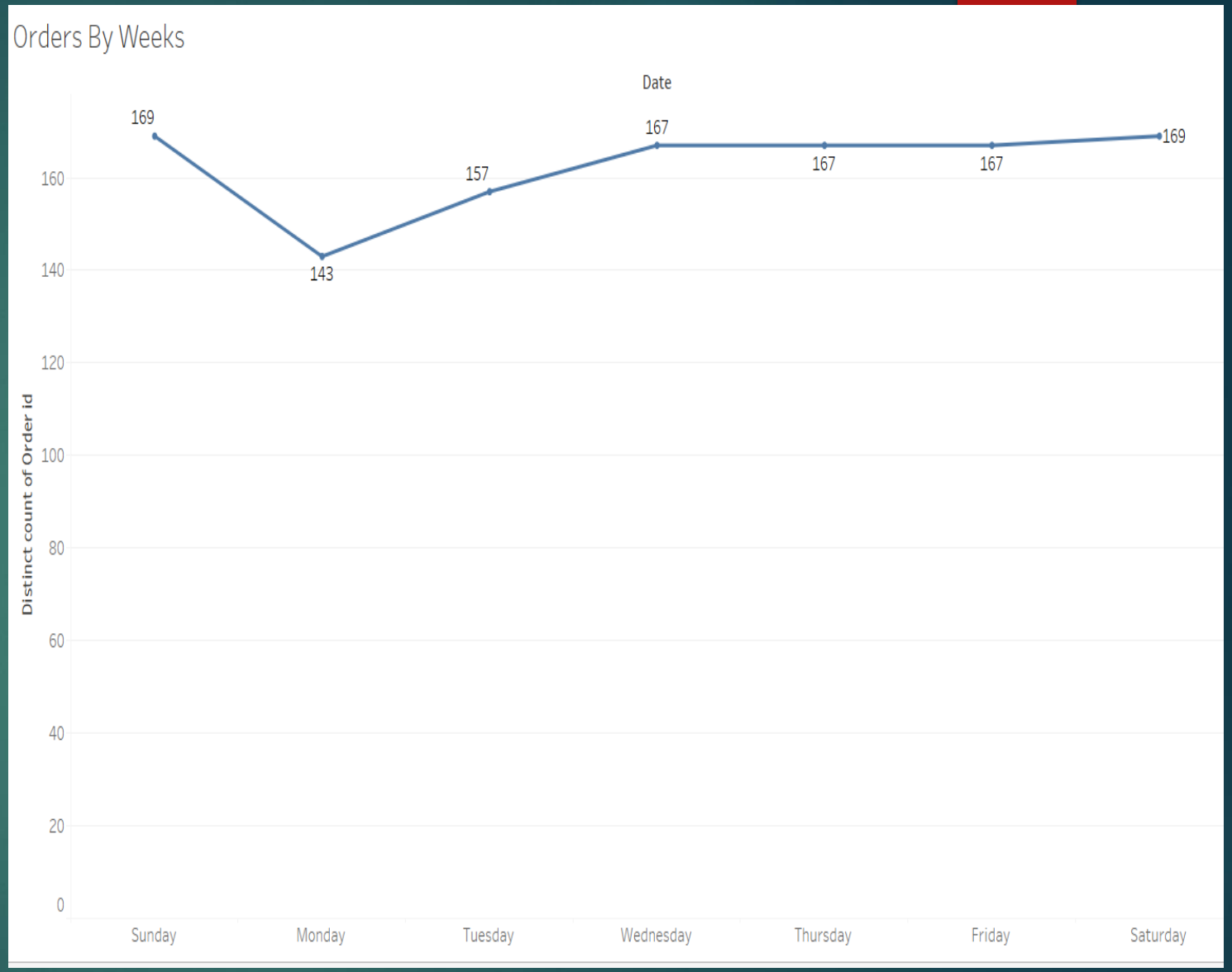


Orders by Months

- **Customer Ordering Patterns Analysis**: The data reveals fluctuations in customer ordering behaviors across different months of the year.
- **Peak Ordering Months**: January recorded the highest number of unique orders, with February and May following closely behind. In contrast, June exhibited the lowest count of unique orders.
- **Consistent Summer Activity:** During the summer months (June, July, and August), there was a relatively steady level of order activity, indicating consistent customer engagement during this period.
- **September Increase**: Notably, there was a slight uptick in unique orders observed in September, suggesting a potential shift or increase in customer ordering patterns towards the end of summer.

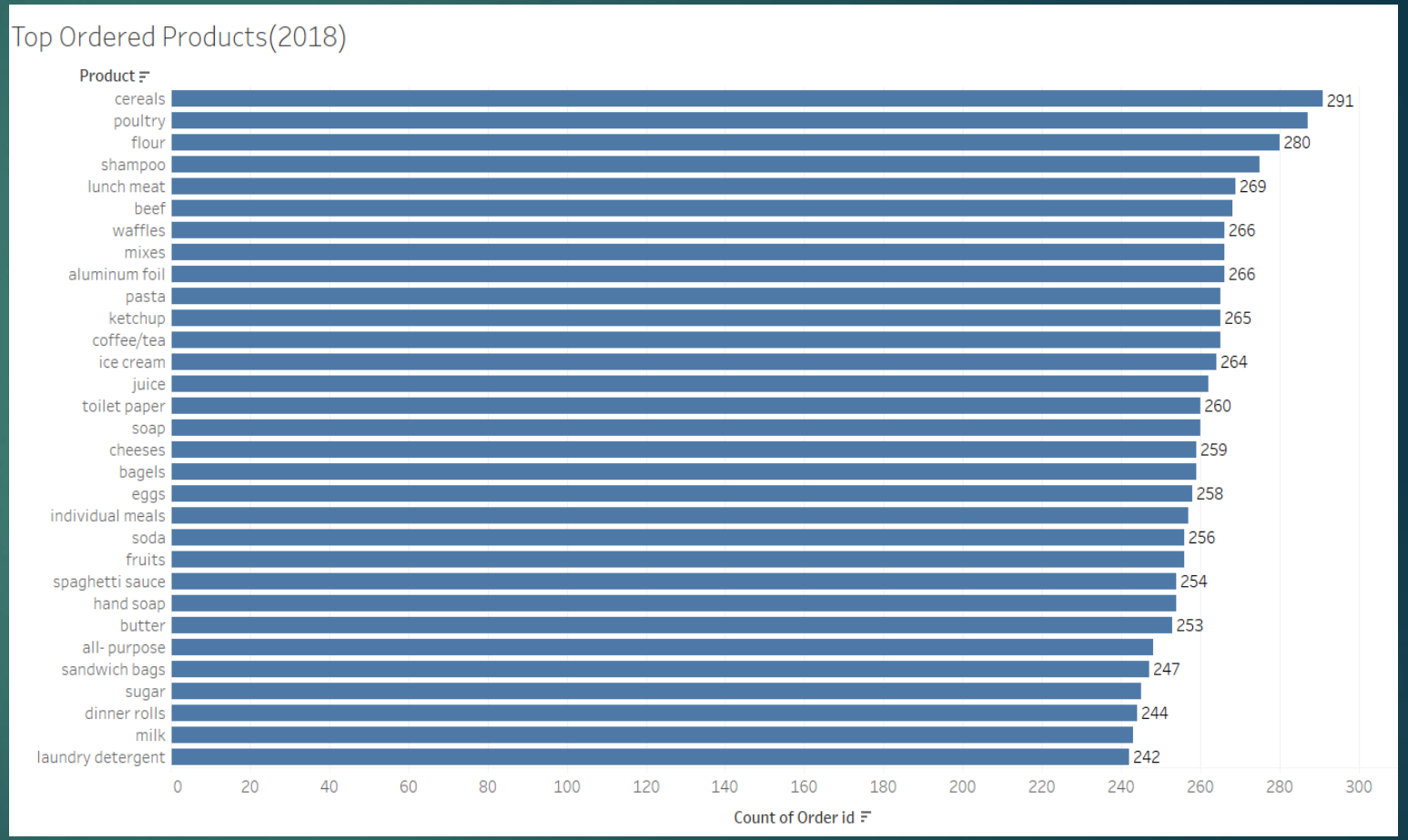
Orders by Weeks

- **Order Activity Analysis by Weekday**: The data indicates a consistent level of order activity throughout the weekdays, with Wednesday, Thursday, and Friday exhibiting slightly higher numbers of unique orders.
- **Weekend Engagement**: Saturdays and Sundays stand out with the highest count of unique orders, highlighting increased customer engagement and purchasing behavior over the weekends.
- **Weekday Stability**: Monday to Friday shows stable order activity levels, suggesting regular customer interaction and consistent ordering patterns during the workweek.
- **Weekend Peaks**: The weekends, particularly Saturdays and Sundays, experience heightened customer activity, indicating a significant surge in purchasing behavior and overall engagement during these days.



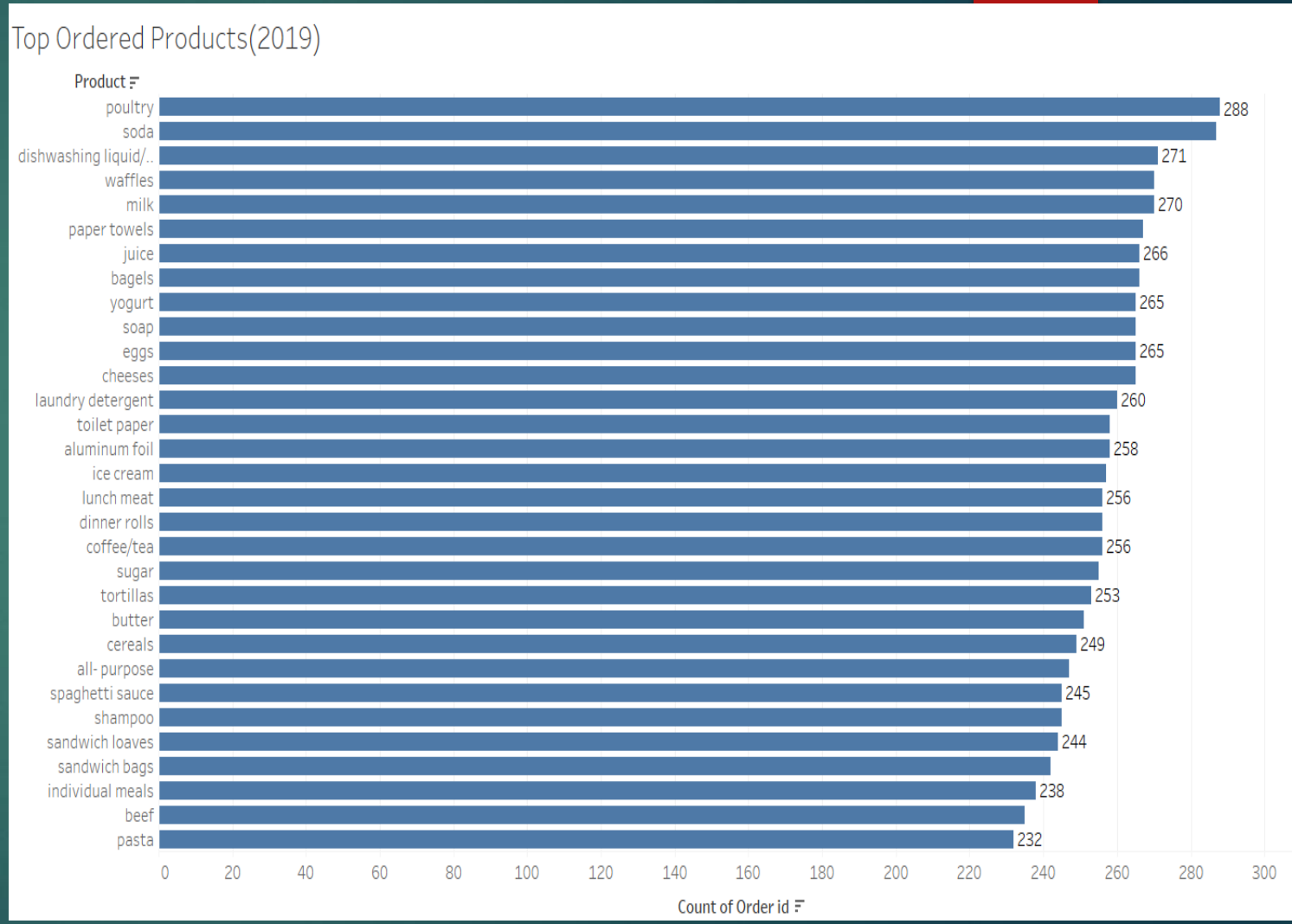
Top Ordered Products (2018)

- **Popular Products in 2018**: The top five products frequently ordered by customers in 2018 were cereals, poultry, flour, shampoo, and lunch meat.
- **Leading Product**: Cereals topped the list with 291 orders, closely followed by poultry with 287 orders, showcasing their high demand among customers.
- **Significant Demand**: Flour, shampoo, and lunch meat also demonstrated substantial popularity, with 280, 275, and 269 orders respectively, indicating strong market appeal and customer preference for these items during the year.
- **Popularity and Market Demand**: These products' consistent order counts highlight their widespread popularity and the strong market demand they enjoyed throughout 2018.



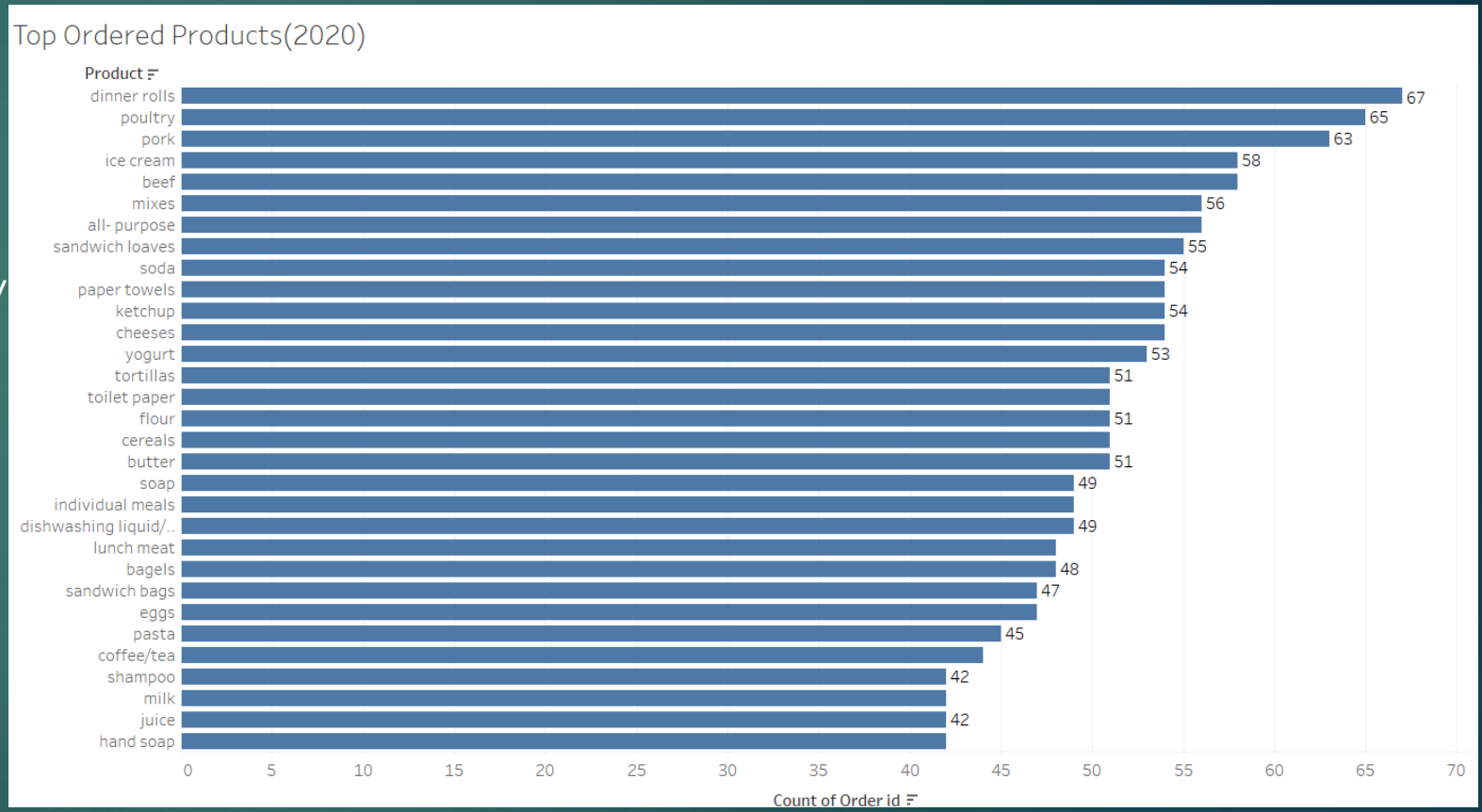
Top Ordered Products (2019)

- **Popular Products in 2019**: The top five highly ordered products by customers in 2019 included poultry, soda, dishwashing liquid, waffles, and milk.
- **Leading Product**: Poultry led the list with 288 orders, closely followed by soda with 287 orders, highlighting their significant demand among customers.
- **Substantial Orders**: Dishwashing liquid, waffles, and milk also received a substantial number of orders, with 271, 270, and 270 respectively, showcasing their popularity and customer preference throughout the year.
- **Market Demand**: These products' consistent order counts in 2019 indicate their strong appeal and the robust market demand they experienced, emphasizing their popularity among customers.



Top Ordered Products (2020)

- **Popular Products in 2020**: The top five highly ordered products by customers in 2020 were dinner rolls, poultry, pork, ice cream, and beef.
- **Leading Product**: Dinner rolls led the list with 67 orders, closely followed by poultry with 65 orders, indicating their strong appeal and customer preference.
- **Notable Choices**: Pork and ice cream were also popular among customers, with 63 and 58 orders respectively, showcasing their appeal and demand throughout the year.
- **Significant Orders**: Beef also had a notable number of orders, with 58, highlighting its popularity and consistent demand among customers in 2020.



Summary

- **Top Products and Market Demand:**
  - Poultry, Soda, Cereal, Ice-cream, Cheese, and Waffles emerged as the top products, indicating significant popularity and demand among customers.
- **Order Activity Trends**:
  - Order activity remained relatively stable from 2018 to 2019, with a slight decrease noted over this period.
  - Limited data for 2020 hinders a thorough understanding of order trends for that year.
- **Unique Order IDs and Overall Activity:**
  - A general downward trend in unique order IDs from Q1 to Q3 implied a potential decrease in overall order activity during that period.
- **Monthly Ordering Patterns:**
  - January, February, and May saw the highest number of unique orders, while June recorded the lowest count.
- **Weekly Order Activity**:
  - Weekdays displayed consistent order activity, with higher counts observed on Wednesday, Thursday, and Friday. Weekends had the highest overall counts, indicating heightened customer engagement.
- **Changing Preferences Over Time**:
  - Over three years, customer order analysis revealed shifting preferences and trends.
  - In 2018, popular items included cereals, poultry, flour, shampoo, and lunch meat, suggesting a demand for household essentials.
  - 2019 showed a mix of food and cleaning products with poultry, soda, dishwashing liquid, waffles, and milk leading the list.
  - However, 2020 saw a shift towards indulgent food options like dinner rolls, poultry, pork, ice cream, and beef, reflecting changing customer preferences and market trends.

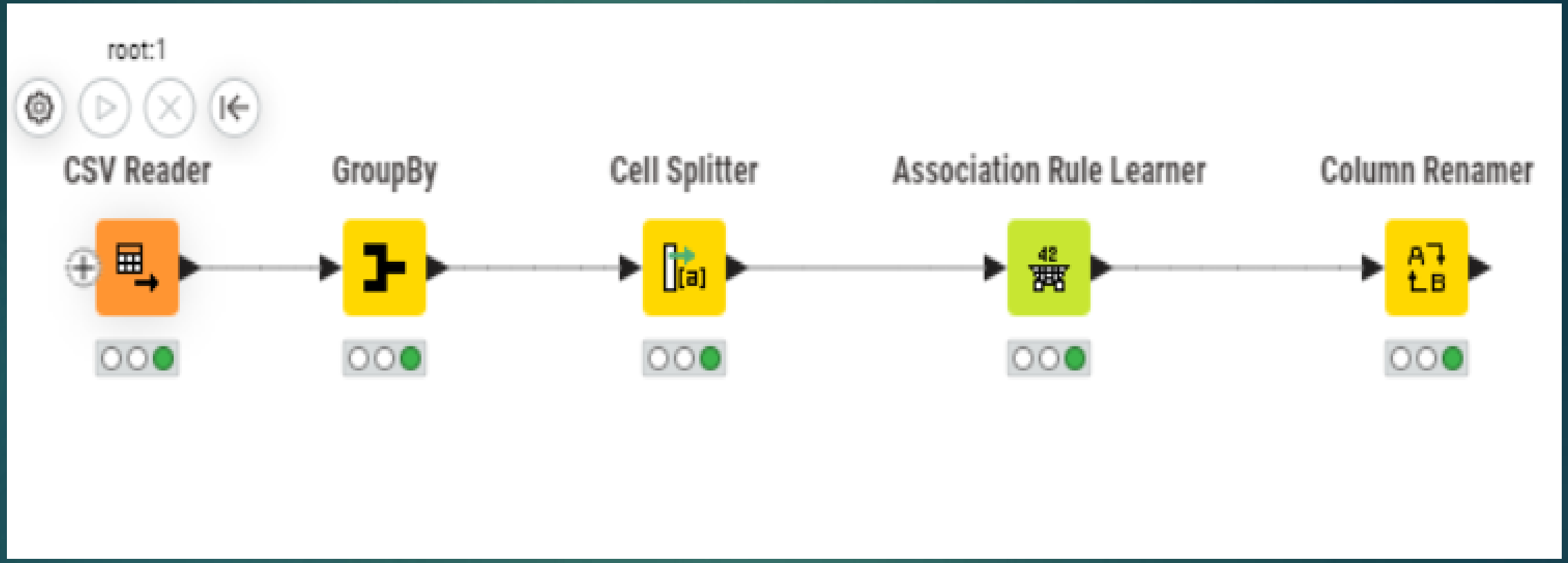
2. Market Basket Analysis

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

Association Rules and Its Relevance

- **Understanding Association Rules**:
  - Association rules in Market Basket Analysis unveil relationships and co-occurrence patterns among items, offering crucial insights into customer buying behavior and preferences.
- **Importance of Association Rules**:
  - The significance of association rules lies in their capacity to assist businesses in optimizing product placement strategies.
  - They enable the creation of targeted marketing campaigns based on customer preferences and behavior.
  - Association rules also aid in implementing effective cross-selling and upselling strategies to boost customer satisfaction and revenue generation.
- **Leveraging Insights:**
  - Businesses can leverage insights from association rules to tailor their offerings, improve customer experience, and drive sales by recommending complementary or related products based on customer purchasing patterns.
- **Strategic Decision-Making:**
  - The utilization of association rules facilitates informed decision-making in areas such as inventory management, promotional offers, and overall marketing strategies aimed at enhancing customer engagement and loyalty.

Knime Workflow



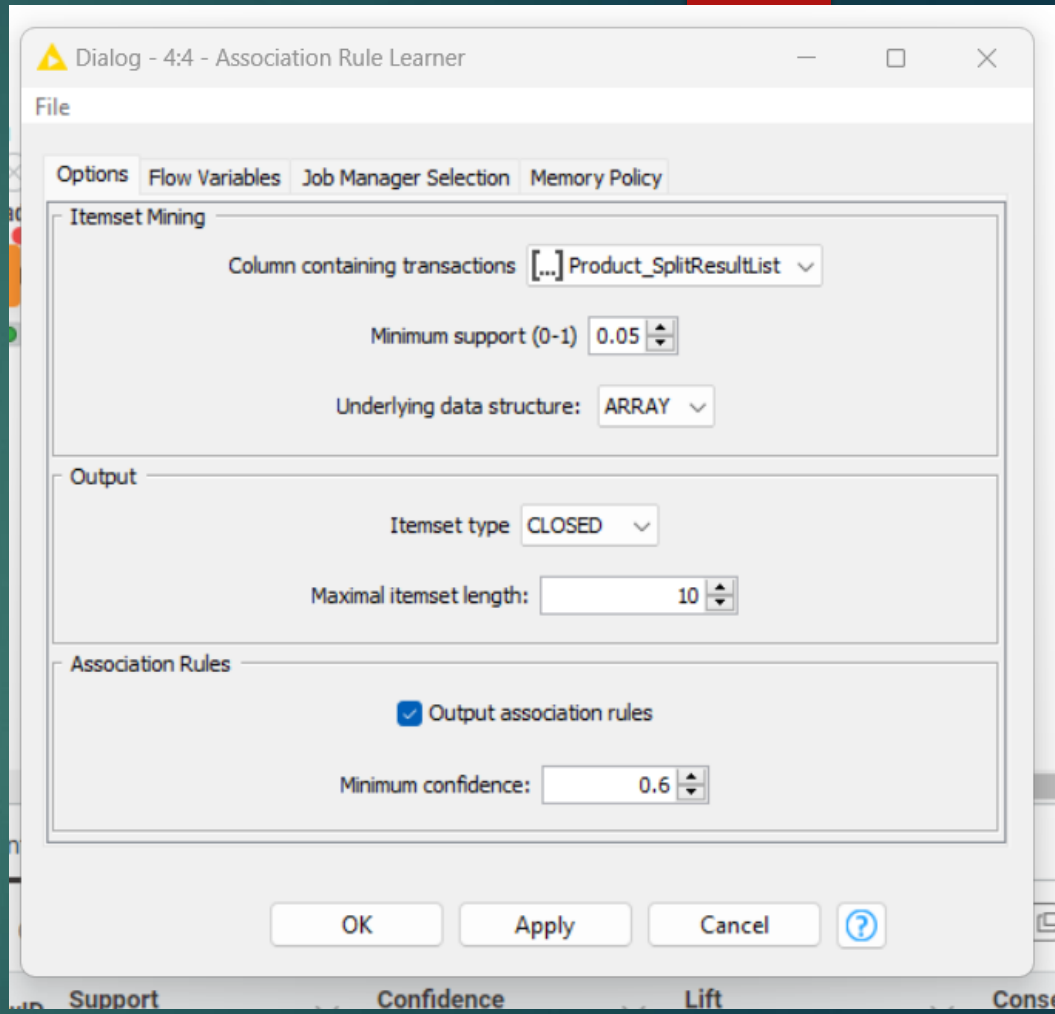
Threshold Values of Support and Confidence

- **Setting Threshold Values**:
  - For this analysis, we have established threshold values to determine the significance of association rules.
  - The minimum support threshold is set at 0.05, while the minimum confidence threshold is set at 0.6.

- **Significance Criteria:**
  - These threshold values serve as criteria for evaluating the importance and trustworthiness of association rules.
  - Rules that meet or exceed the specified support and confidence thresholds are considered significant and reliable.

- **Filtering Less Important Rules:**
  - By applying these thresholds, we can filter out association rules with lower support and confidence levels.
  - This filtering process helps us focus on the rules that demonstrate strong support and confidence, ensuring that our analysis yields meaningful and actionable insights.

- **Enhancing Decision-Making**:
  - Utilizing these threshold values enhances the quality of our analysis by emphasizing rules that are statistically significant and likely to have a meaningful impact on decision-making processes.
  - It enables us to prioritize and concentrate on the most relevant and reliable association rules for deriving actionable recommendations.



3. Association Identification

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

Association Representation in Tabular

RowID	Support *Number (double)*	Confidence *Number (double)*	Lift *Number (double)*	Recommended Products *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,laundr...
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 crea...
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 dete...
rule19	0.054	0.656	1.556	poultry	<---	[dinner rolls,spaghetti sauce,laundry ...
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 Explanation**:
  - Support value reflects how frequently a specific combination of items appears together in the dataset.
  - It measures the popularity or frequency of an itemset among customer transactions, indicating the level of association between items.
- **Confidence Explanation**:
  - Confidence quantifies the probability that if a customer buys one item, they will also purchase another item.
  - It is calculated by comparing the number of transactions where both items are bought together to the total number of transactions involving the first item.
- **Lift Interpretation**:
  - Lift evaluates the strength of association between two items in an association rule.
  - It compares the probability of the two items being bought together to their individual probabilities, indicating the extent of dependency or relationship between the items.
- **Positive Association:**
  - A lift value exceeding 1 suggests a positive association between the items, indicating that they are more likely to be purchased together.
  - This indicates a stronger dependency or correlation between the items, implying a potential opportunity for targeted marketing or bundling strategies.

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)

- **Special Combo Deal**: Offer a discounted price or additional item when customers purchase yogurt, poultry, aluminium foil, and juice together in a special combo deal.
- **"Buy Two Get One Free" Offer**: Incentivize customers to buy dinner rolls, spaghetti sauce, and ice cream together by offering a "buy two get one free" promotion.
- **Bundle Promotion**: Introduce a bundle promotion where customers can purchase paper towels, eggs, and pasta together at a discounted price, encouraging them to explore related products.
- **Cross-Selling Offer**: Provide a discount on cheese, bagels, and sandwich bags for customers purchasing cereals, encouraging cross-selling and increasing customer value.
- **Limited-Time Promotion**: Launch a time-limited promotion offering a percentage savings to customers buying poultry, laundry detergent, and mixes together, creating urgency and driving sales.
- **Loyalty Program**: Implement a loyalty program rewarding customers who frequently purchase recommended items or participate in suggested combos, fostering long-term relationships and repeat purchases while increasing customer satisfaction.


THANK
YOU