Grocery Store's Data MRA Project-Milestone 2

BY: Pratheek U

Contents:

Agenda:

- Executive Summary of the data
- Exploratory Data Analysis
- Market Basket Analysis
- Associations Identified
- Recommendation

01
Executive Summary

- Problem statement
- Executive Summary & Data Dictionary
- Assumptions about data

Problem Statement

A Grocery Store shared the transactional data with you. Your job is to identify the most popular combos that can be suggested to the Grocery Store chain after a thorough analysis of the most commonly occurring sets of items in the customer orders. The Store doesn’t have any combo offers. Can you suggest the best combos & offers

Executive Summary:

- Data: from 01-01-2018 to 26-02-2020
- Objective: project involves conducting a thorough analysis of Point of Sale (POS) Data for providing recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.
- Dataset: `20641 Rows, 3 columns`,
- Missing values : None
- Duplicate values: `4730`
- The exploratory analysis and insights provide a clear understanding of the data and highlight the key trends and patterns in sales.
- Market Basket Analysis using association rules was performed to identify the relationships between the products purchased by the customers.
- This analysis helped to identify the products that are frequently purchased together, which can be used to create lucrative offers for the customers.

Data Dictionary

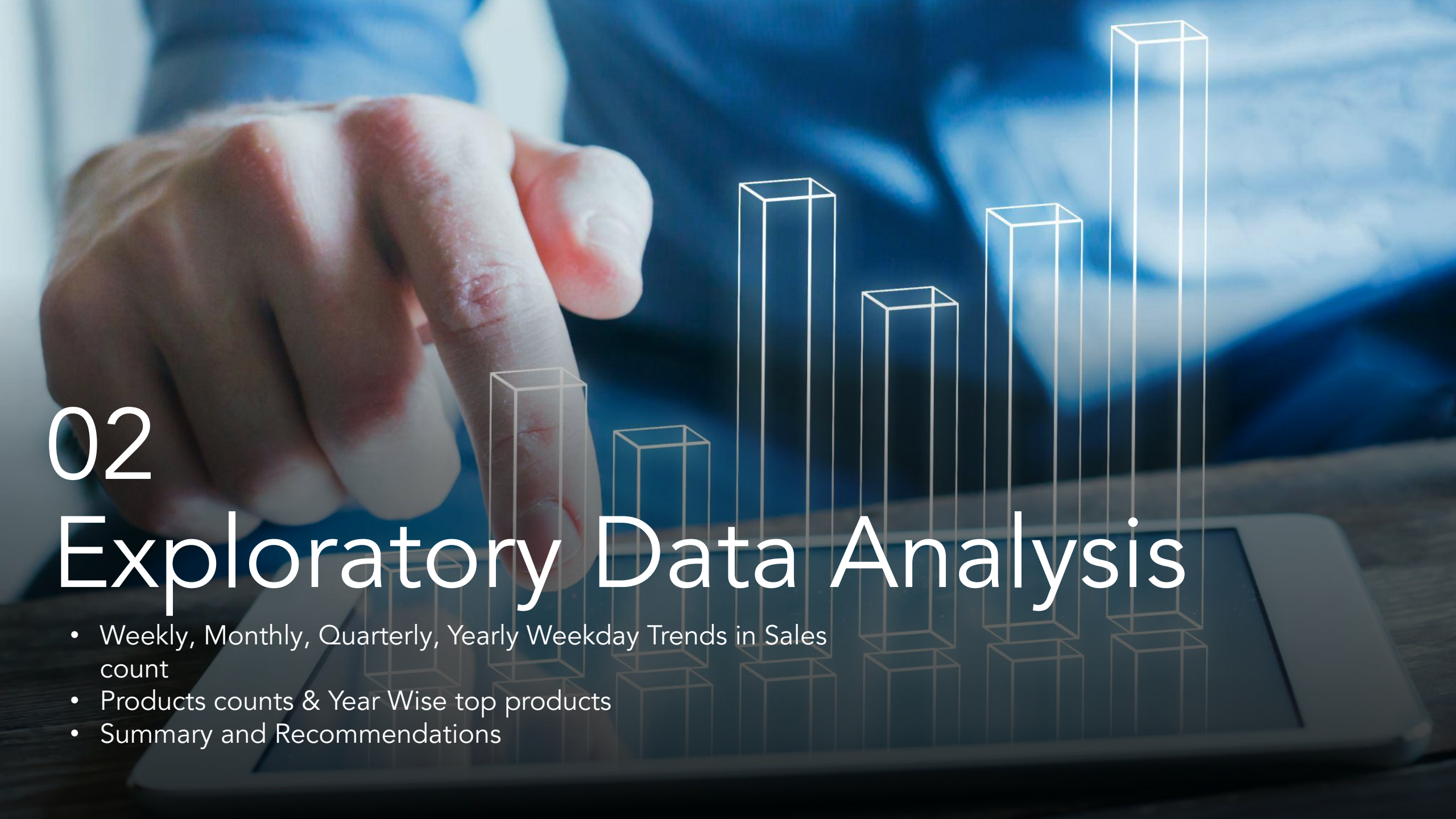
Date	date of product sold
Order_id	Id of the order
Product	Name of the Product sold

Duplicate Values:

- It is generally a good practice to drop duplicate rows in a dataset as they do not provide any additional information and can skew the results of any analysis performed on the dataset.
- However, in this particular case, dropping duplicate rows may not be appropriate as there is no unique identifier for each row.
- Each row consists of a date, a customer ID, and a product purchased, but the same product can be purchased by multiple customers on the same date.
- Therefore, we drop duplicate rows, it may inadvertently remove valid information from the dataset.
- So duplicate values are not removed from the dataset.

Assumptions:

- The data represents a list of items purchased at a grocery store on various dates.
- Each entry in the data represents a single item purchased.
- The first column in the data represents the date the item was purchased.
- The second column represents the customer who made the purchase.
- The third column represents the item purchased.
- The same item can be purchased by multiple customers on different dates.
- There is no information provided about the quantity or price of each item.
- We have not dropped the duplicated values.

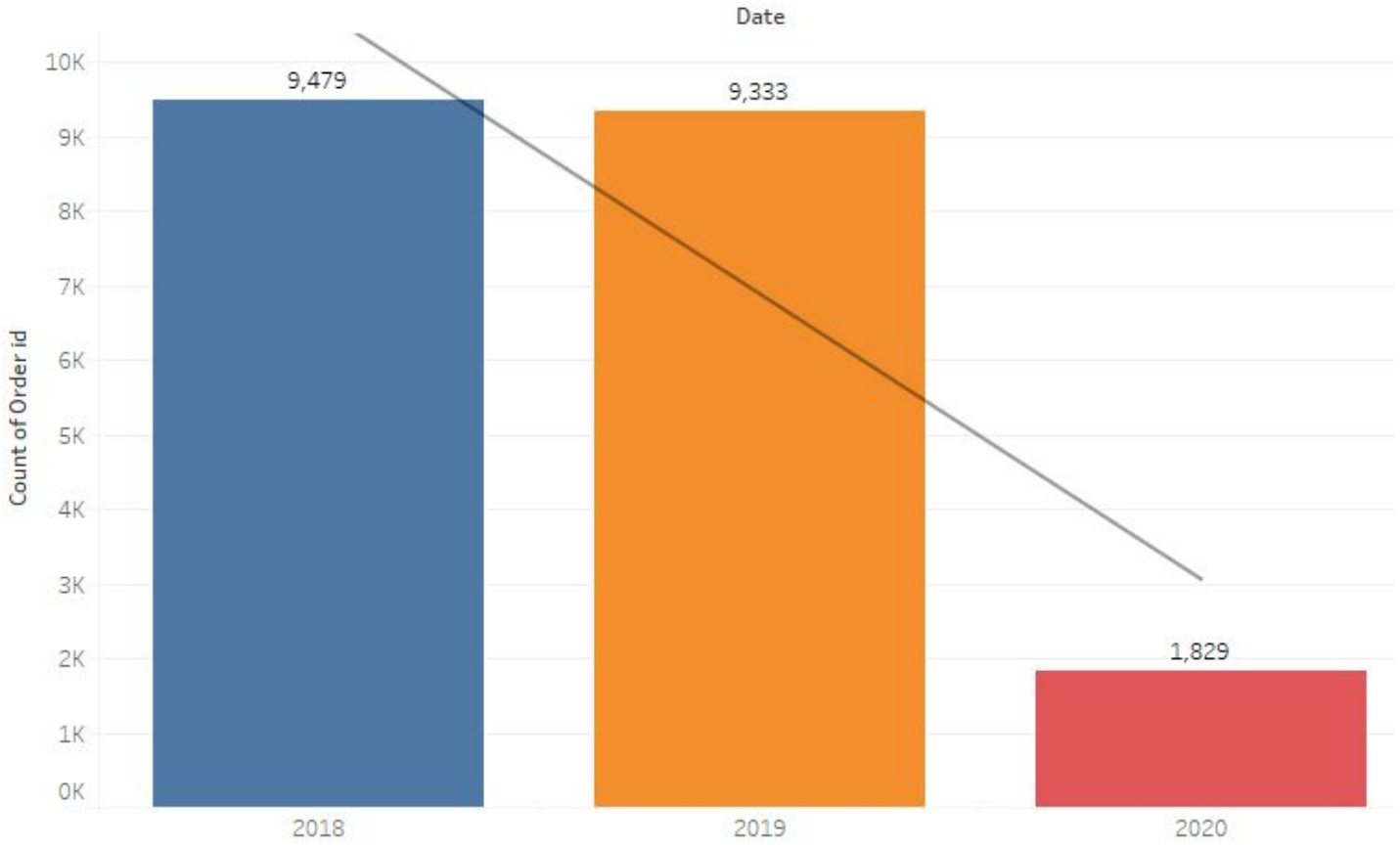


02
Exploratory Data Analysis

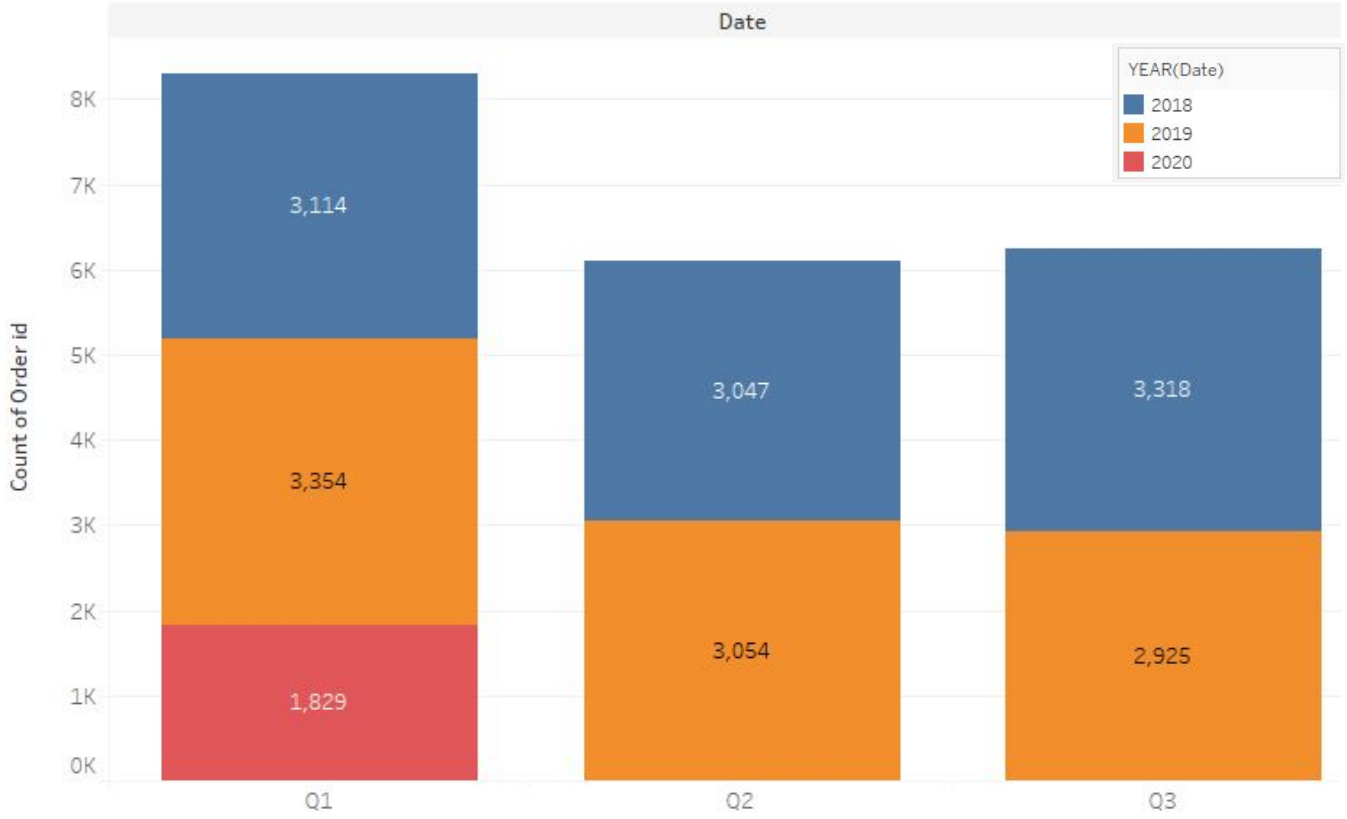
- Weekly, Monthly, Quarterly, Yearly Weekday Trends in Sales count
- Products counts & Year Wise top products
- Summary and Recommendations

Yearly Count of Products Sold

As we have data till 26 feb 2020 that's why the count of products sold in 2020 is low.



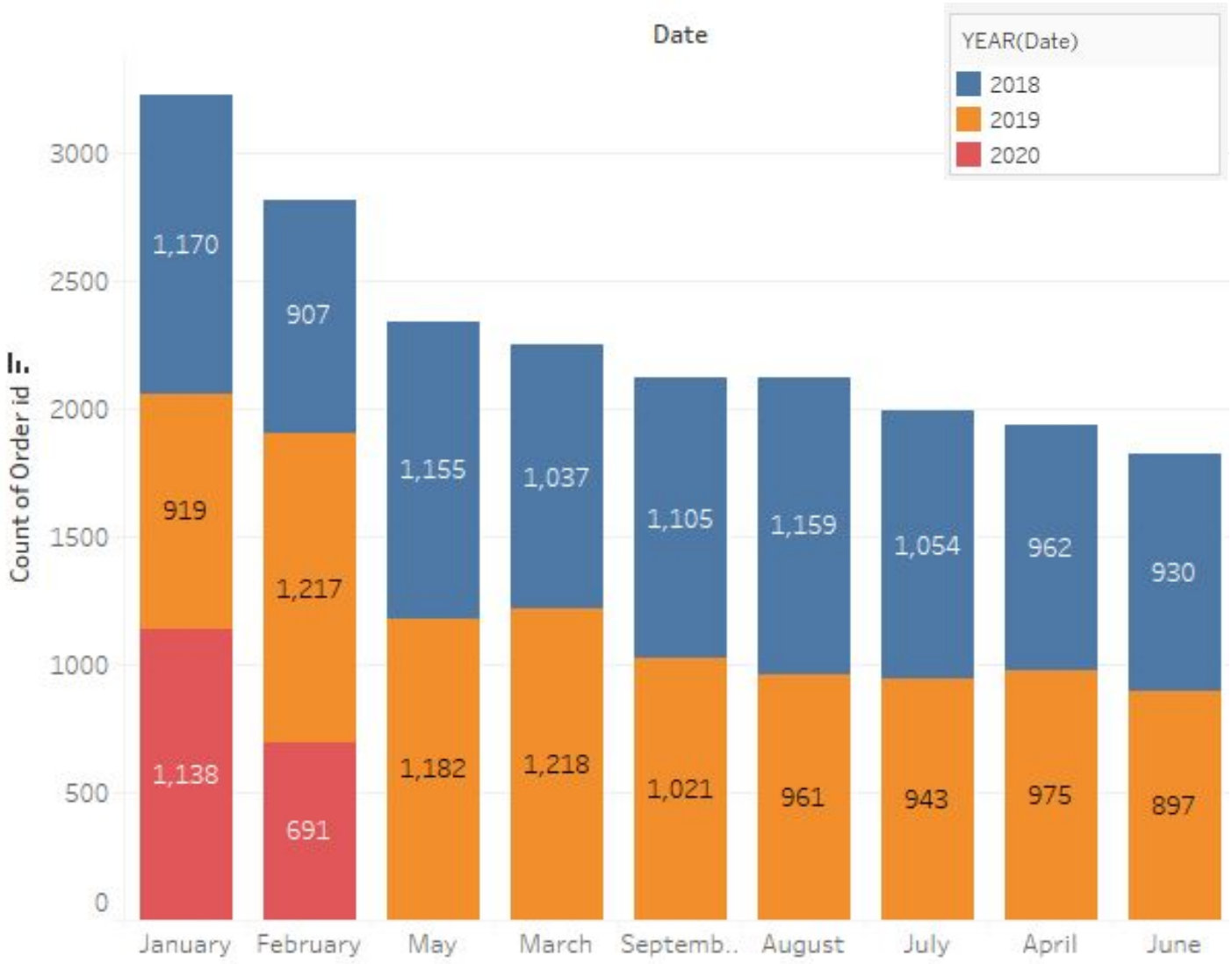
Quarterly Count of Products Sold



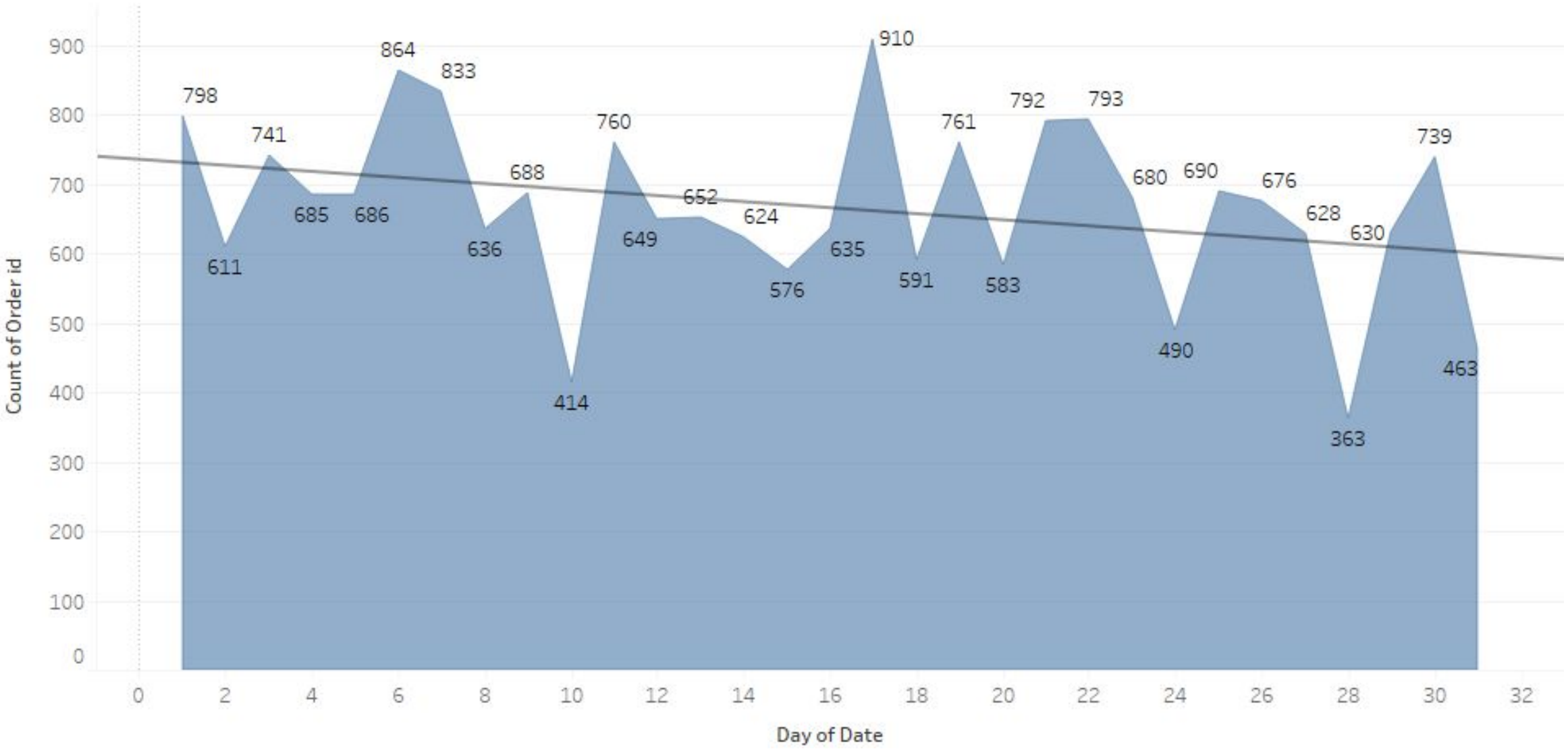
- As we have data till 26 feb 2020 that's why the count of products sold in Q1 is Hight.
- In 2019 Q1 sales was highest
- In 2018 Q3 sales was highest
- Count of product sold in Q2 is aproxx same in 2019 and 2018.

Monthly Count of Products Sold

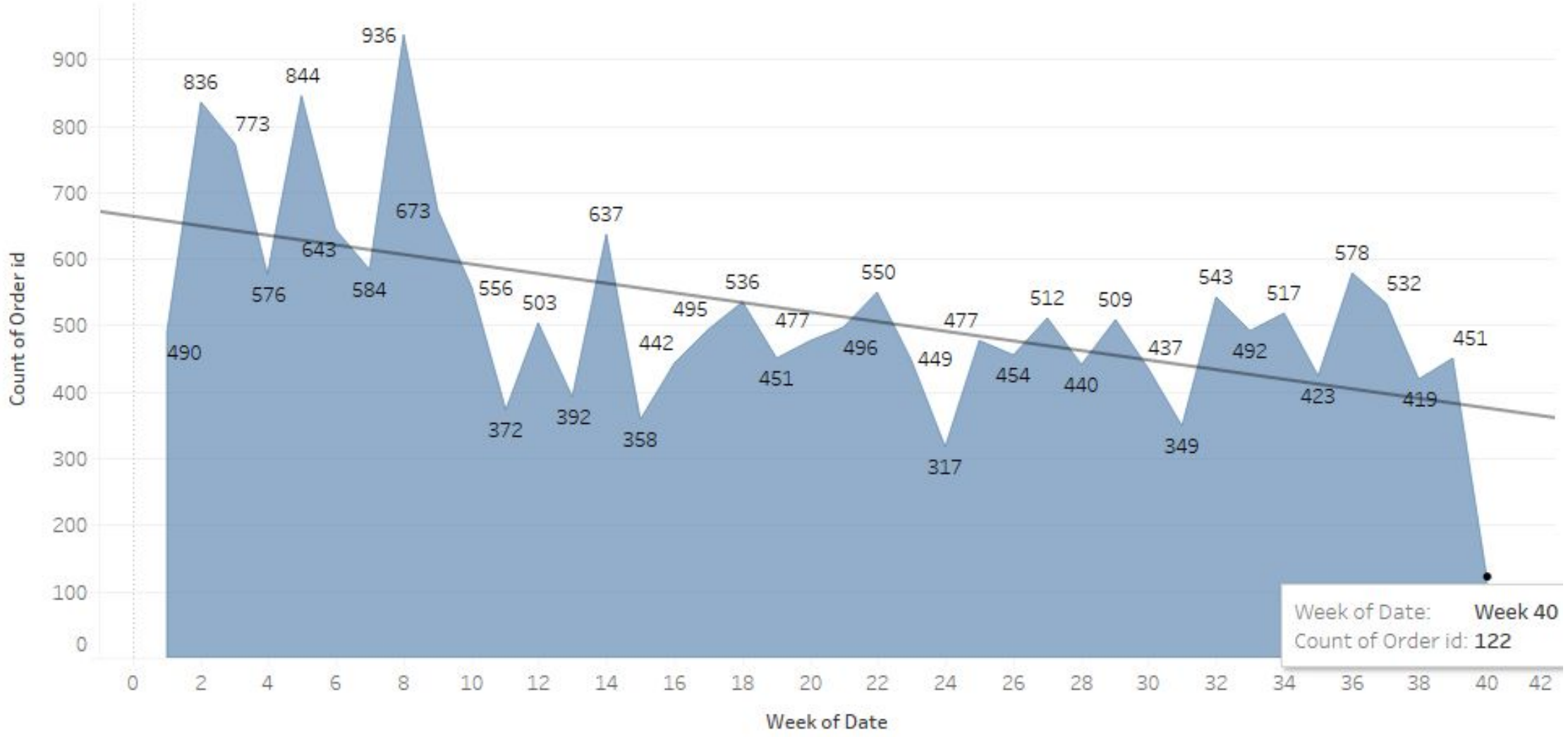
- In 2018 most of the products were sold in January and least were sold in February.
- In 2019 most of the products were sold in March and least were sold in January.



Day of the Month Count of Products Sold

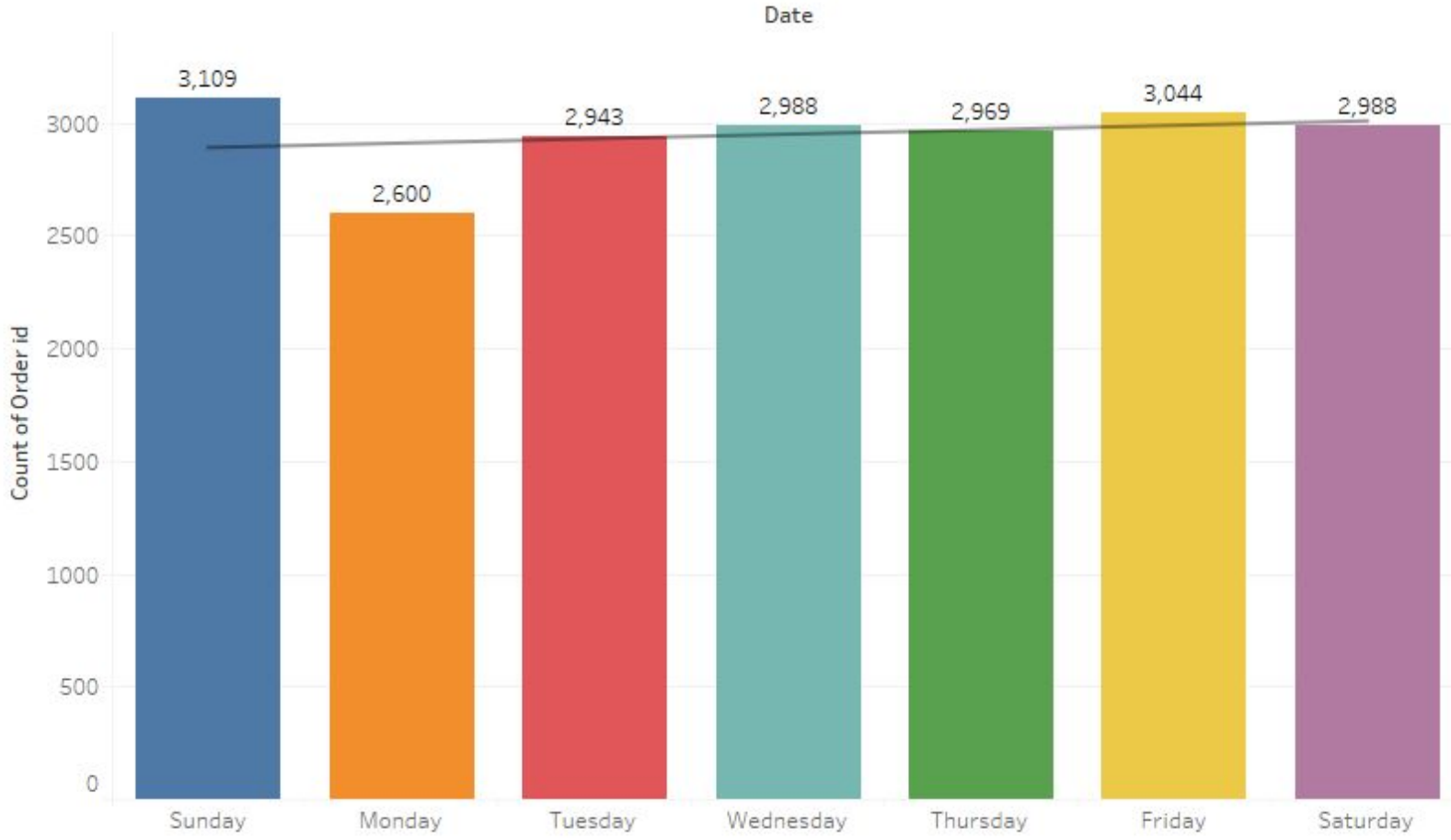


Weekly Count of Products Sold

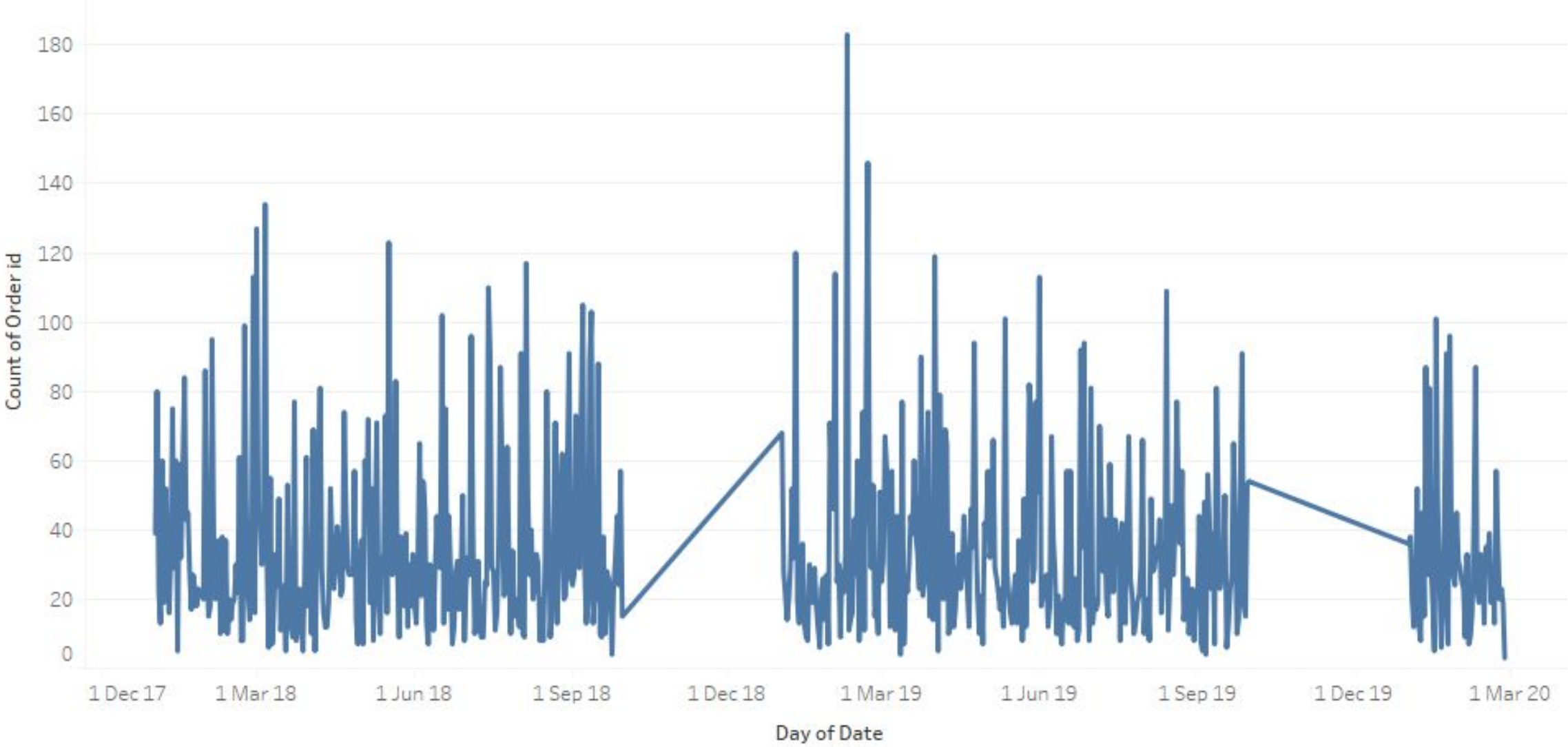


Weekday Count of Products Sold

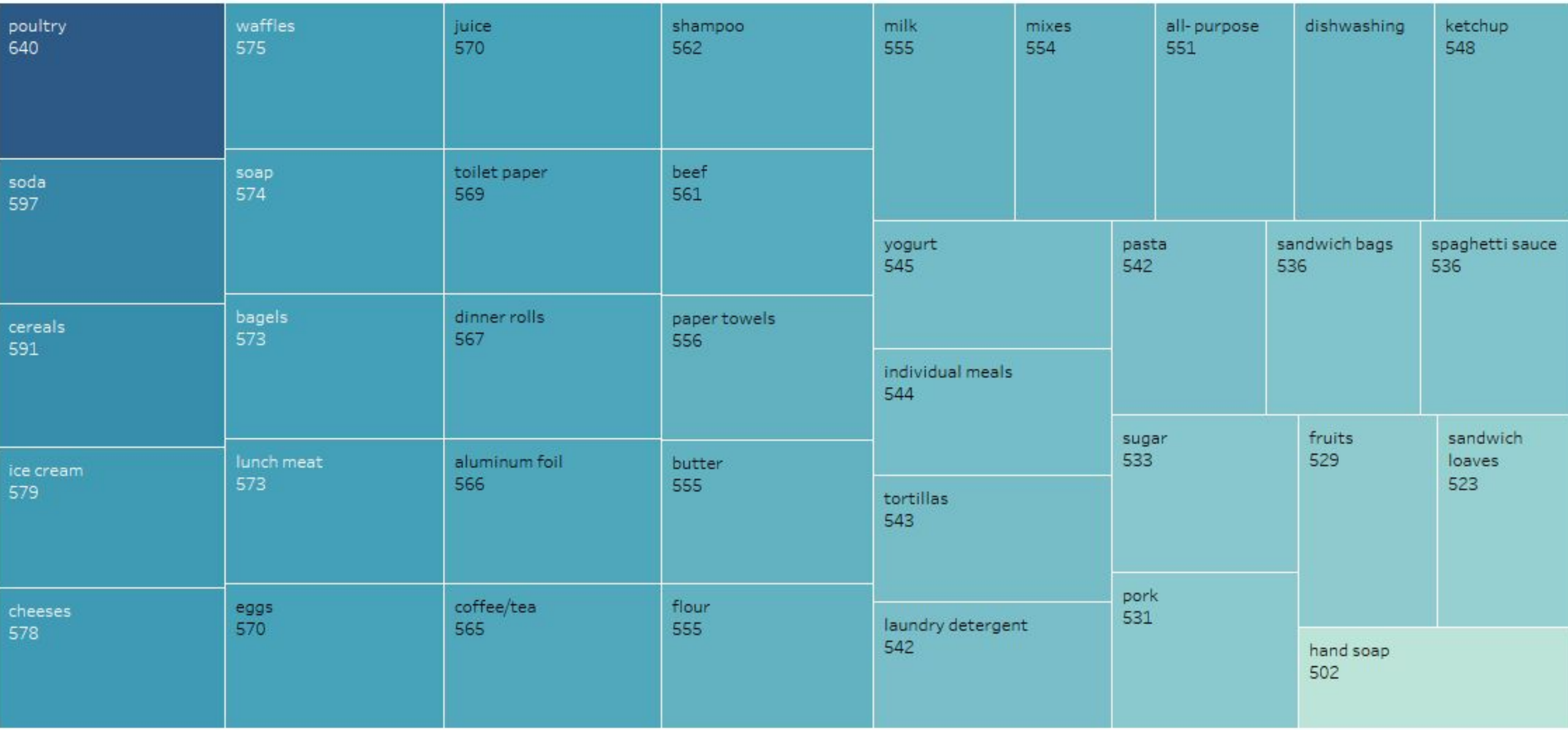
- Most of the products were sold on Sundays.
- Least products were sold on Mondays.
- On other days sales in consistent.

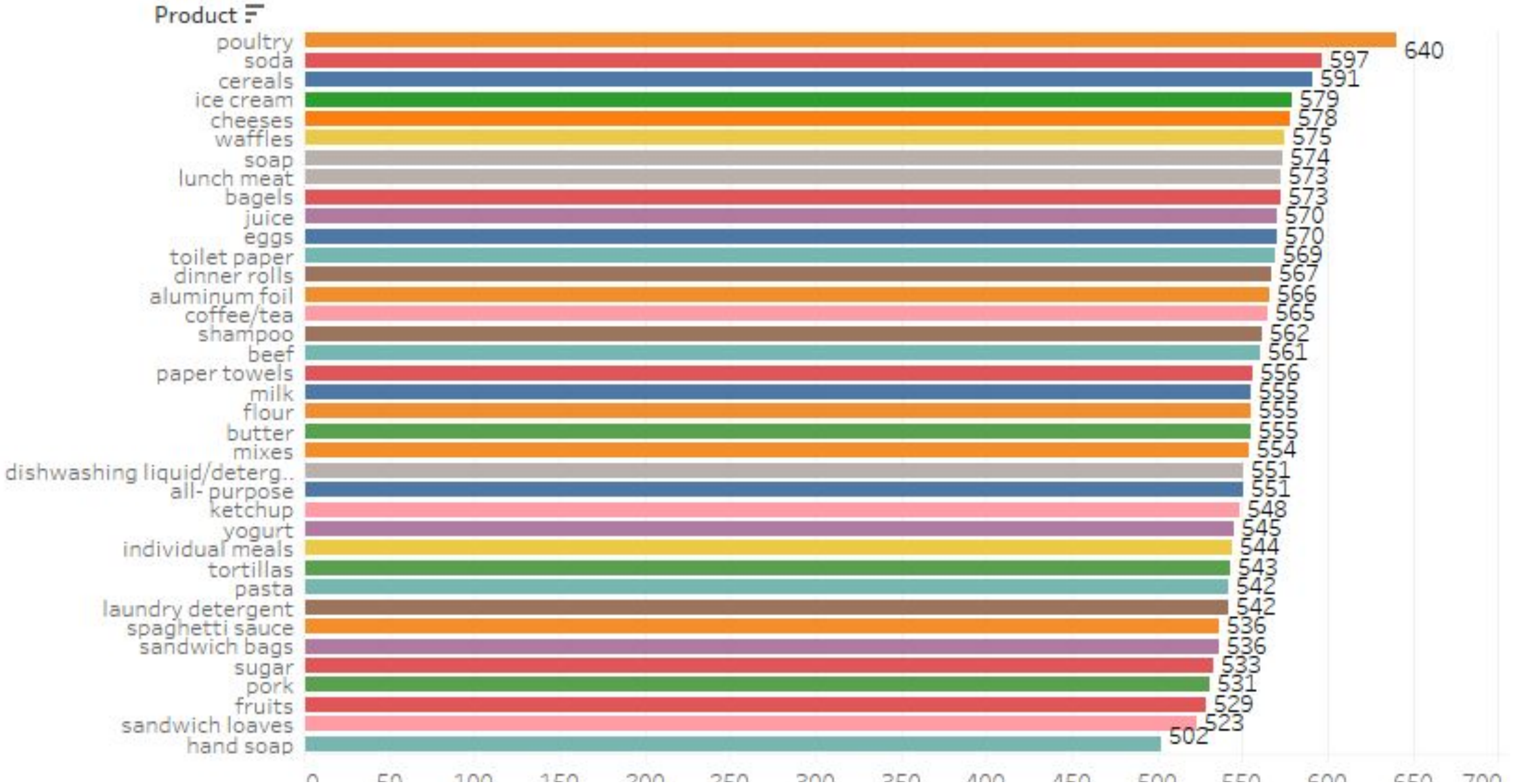


Daily Count of Sales



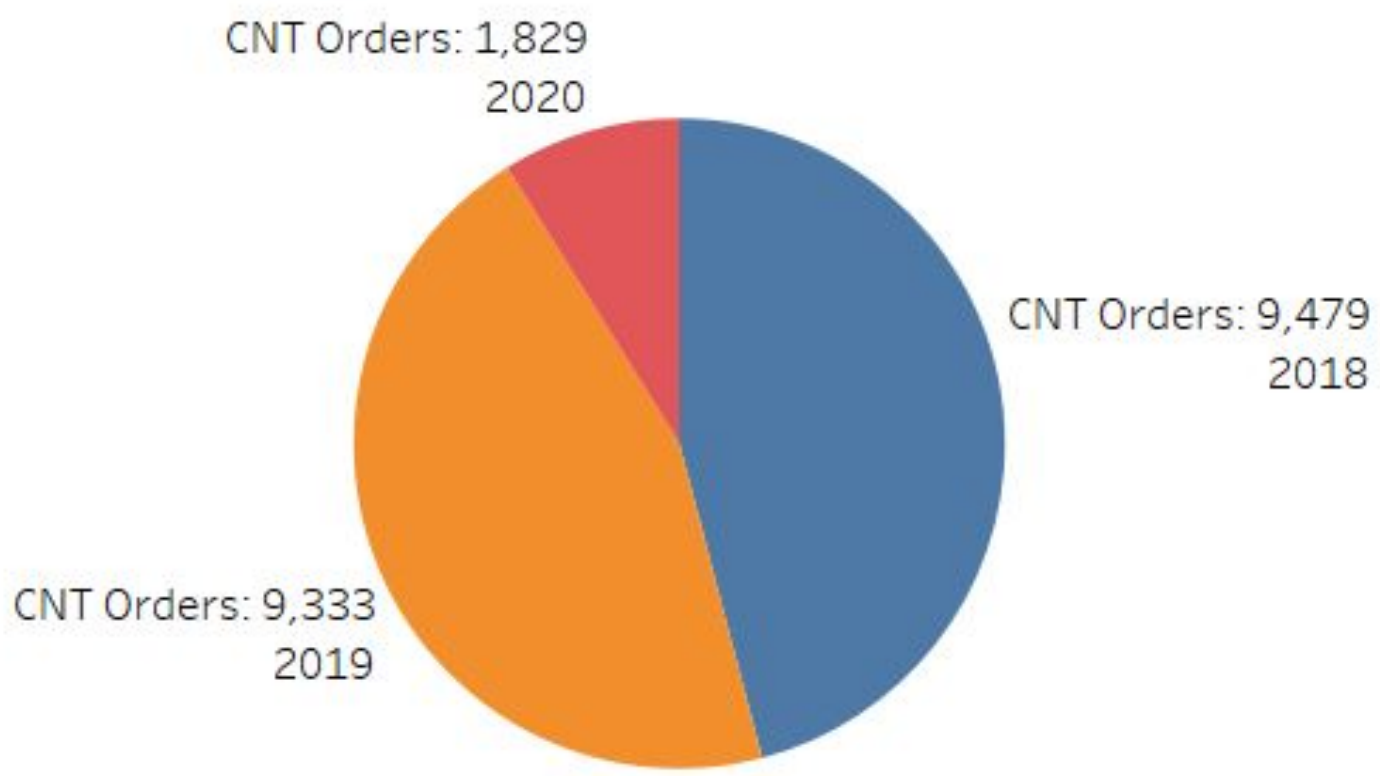
Count of Products Sold




Product
poultry 640
soda 597
cereals 591
ice cream 579
cheeses 578
waffles 575
soap 574
lunch meat 573
bagels 573
juice 570
eggs 570
toilet paper 569
dinner rolls 567
aluminum foil 566
coffee/tea 565
shampoo 562
beef 561
paper towels 556
milk 555
flour 555
butter 555
mixes 554
dishwashing liquid/deterg.. 551
all- purpose 551
ketchup 548
yogurt 545
individual meals 544
tortillas 543
pasta 542
laundry detergent 542
spaghetti sauce 536
sandwich bags 536
sugar 533
pork 531
fruits 529
sandwich loaves 523
hand soap 502
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700

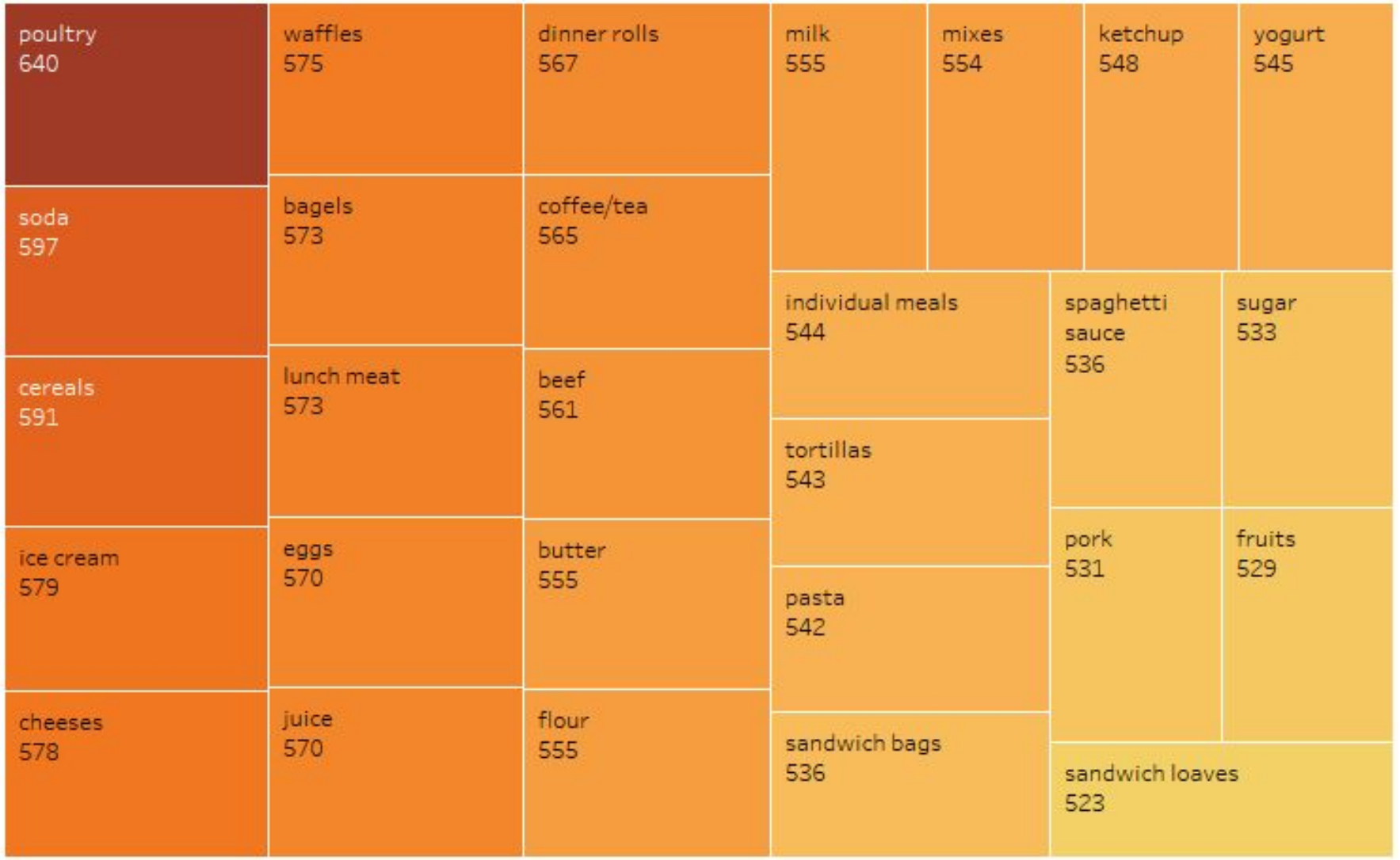
Count of Products Sold Yearly

Most of the Products were sold in 2018.



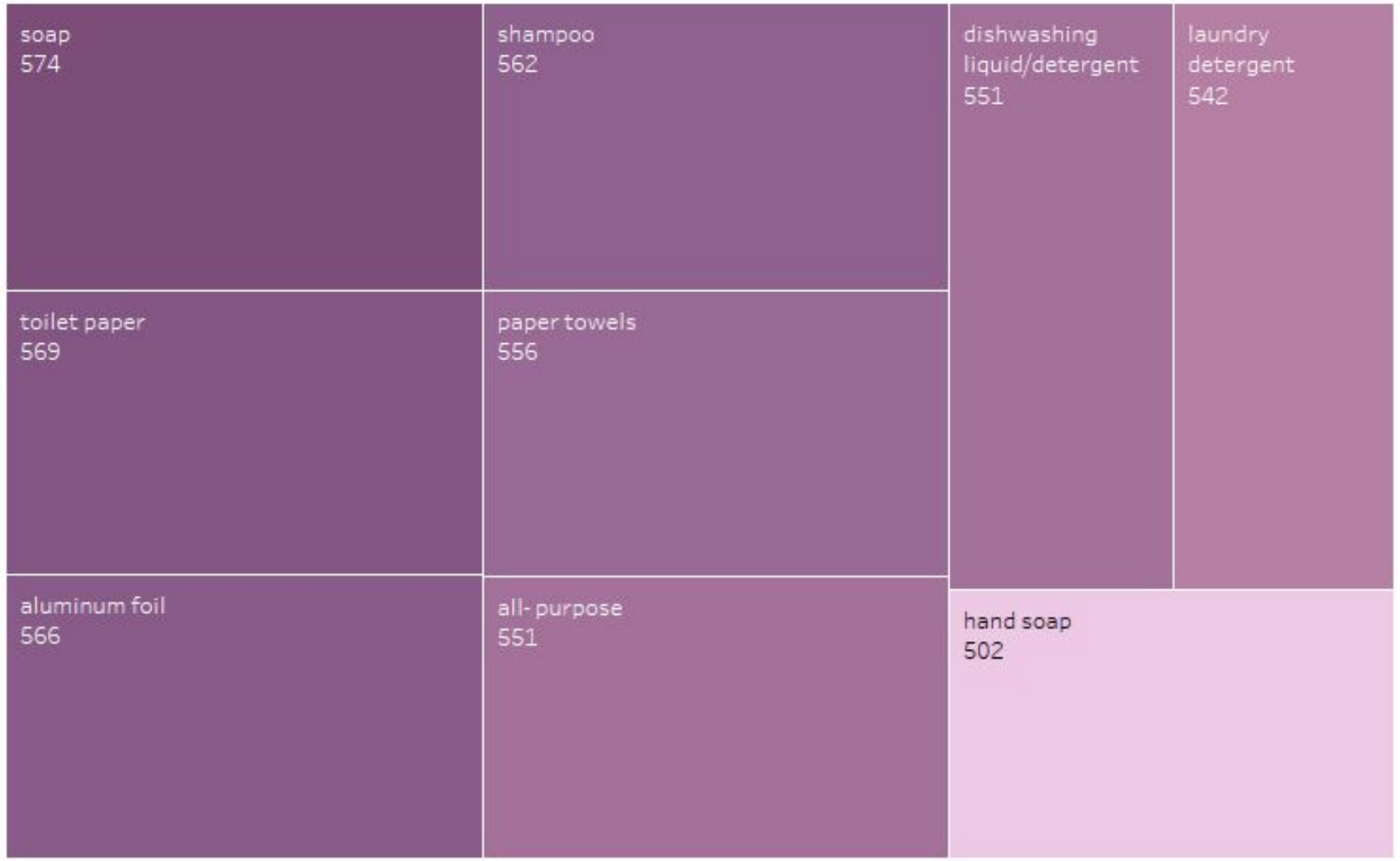
Count of Eatable Products Sold:

- There are total 28 products in this category.
- Highest sold : poultry , Soda, cereals
- Least sold : pork, fruits, sandwich loaves



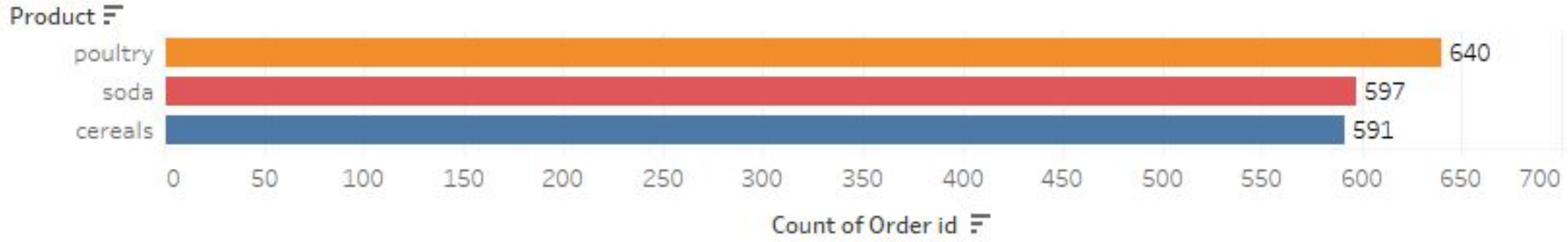
Count of Non-Eatables Products Sold:

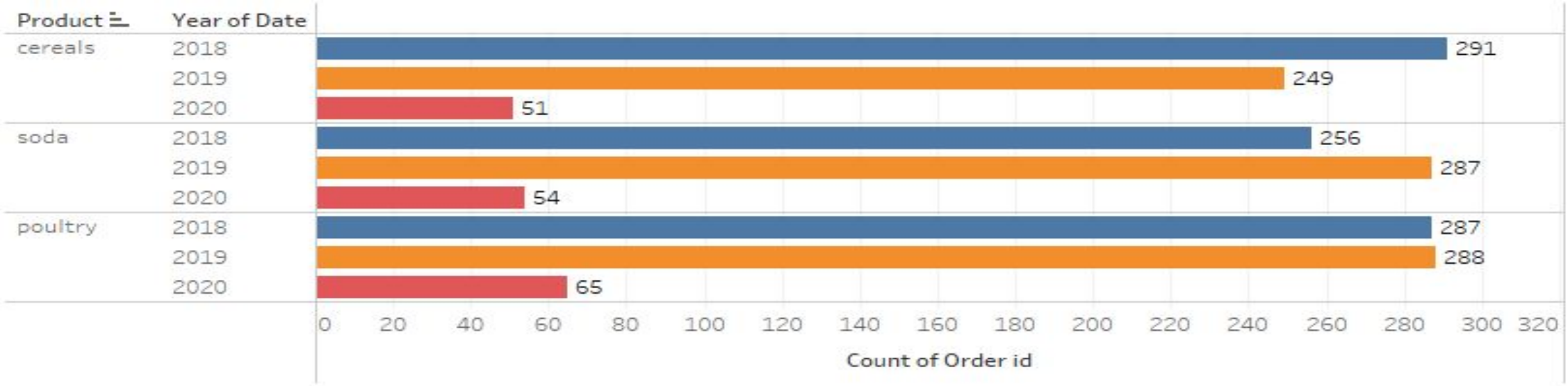
- There are 9 products in this categoary.
- Highest Sold Products : Soap, Toilet Paper.
- Lowest Sold Products: Hand soap.



Top three products over the years

We can see poultry, cereals and soda are highly sold products over the years

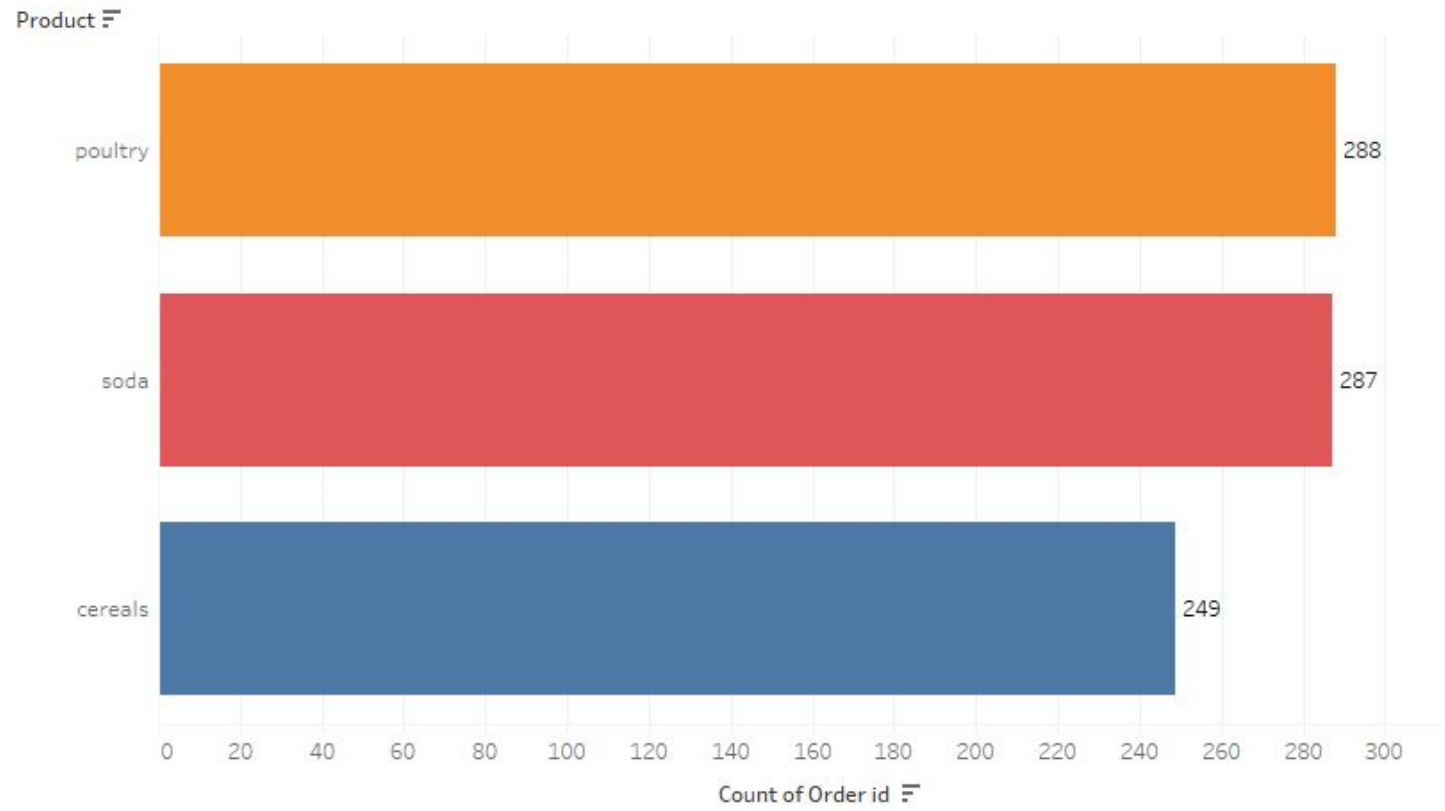




Count of Products sold in 2020 January and February

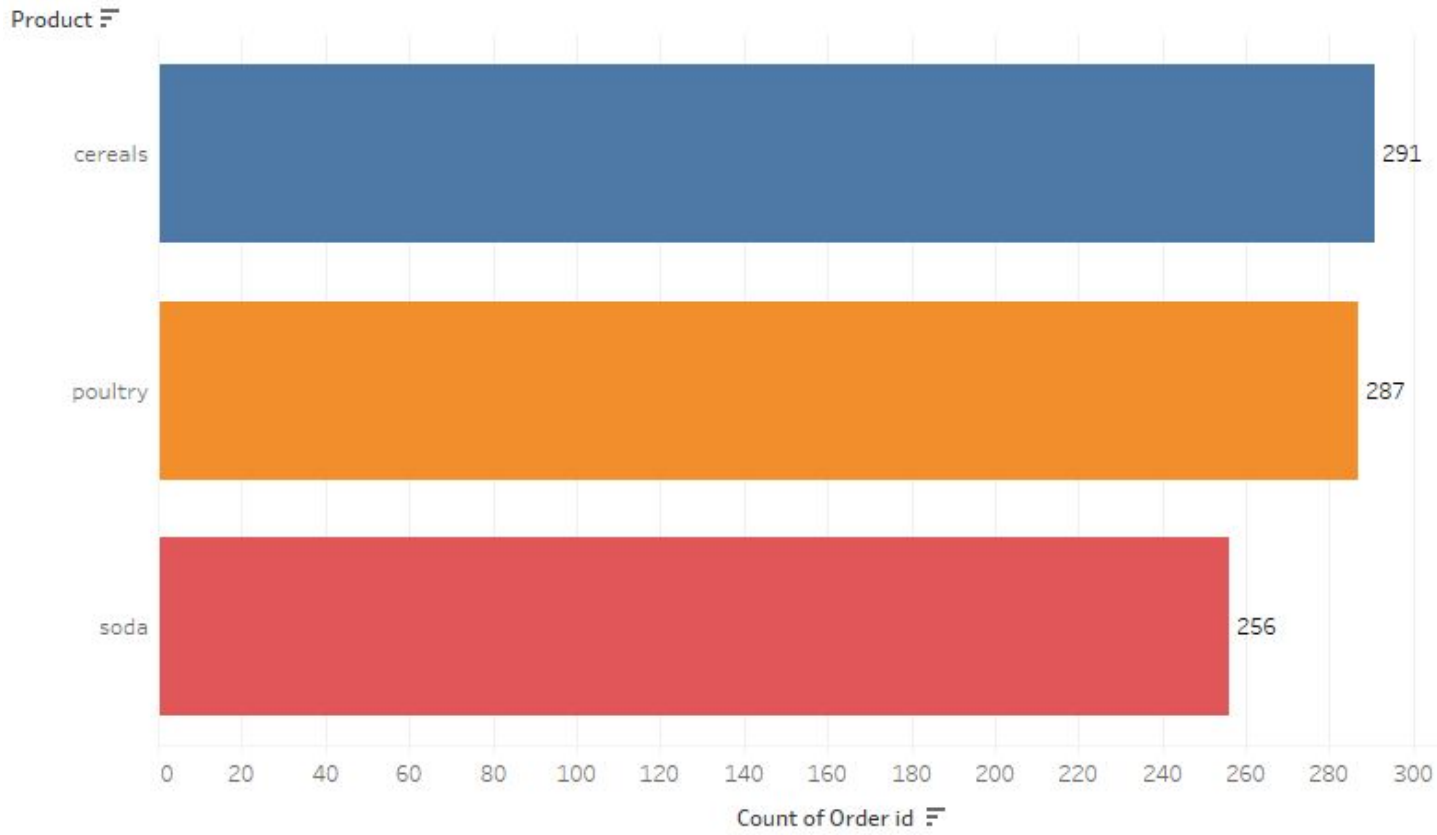
Top 3 products sold in 2020 January and February are dinner role, poultry and pork

Count of Products sold in 2019



Poultry, Soda and Cereal are the top 3 Products sold in 2019

Count of Products sold in 2018



Cereals, Poultry and Soda are the top 3 Products sold in 2018

Summary

- Cereals, Poultry, and Soda are the top 3 products sold in 2018 and 2019.
- Dinner rolls, Poultry, and Pork are the top 3 products sold in January and February 2020.
- Poultry, Cereals, and Soda are the highly sold products over the years.
- The highest sold products are Soap and Toilet Paper, and the lowest sold product is Hand Soap in non eatables products.
- The highest sold products in the eatables category are Poultry, Soda, and Cereals, and the least sold products are Pork, Fruits, and Sandwich Loaves.
- Most of the products were sold on Sundays, and the least were sold on Mondays.
- In 2018, the most products were sold in January, and the least were sold in February. In 2019, the most products were sold in March, and the least were sold in January.
- The sales were the highest in Q1 2019 and Q3 2018.
- The count of products sold in Q2 is approximately the same in 2019 and 2018.
- The count of products sold in 2020 is low, possibly due to the data being only until 26th February.

Recommendation:

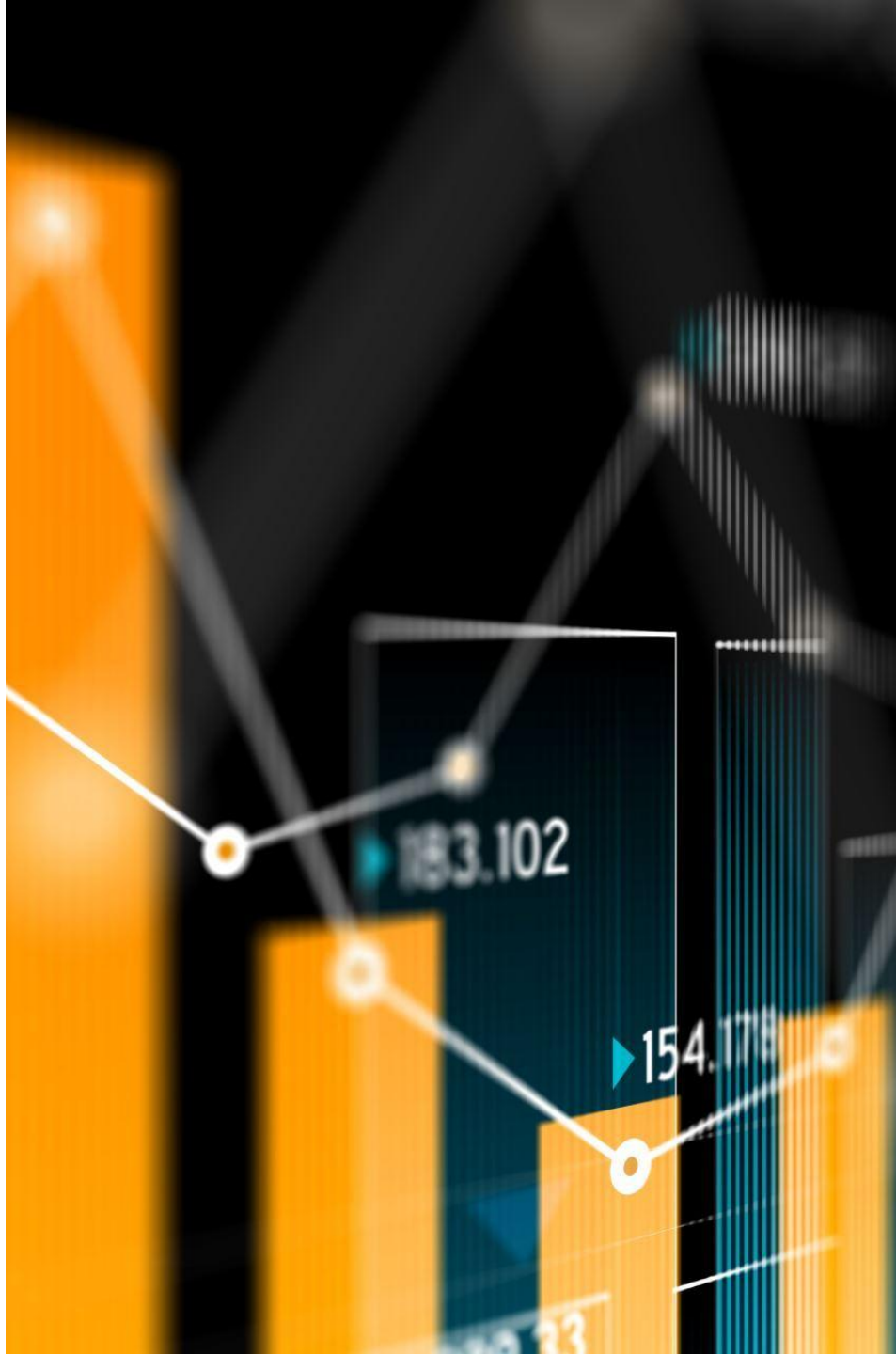
- Focus on promoting and stocking up on poultry, soda, and cereals as they are consistently top-selling products.
- Consider increasing the stock of soap and toilet paper as they are the highest sold non-eatable products.
- Evaluate the reasons behind the low sales of hand soap and take measures to increase its sales.
- Schedule promotions and offers on Sundays to maximize sales on the day with the highest sales.
- Plan marketing campaigns and discounts during February to increase sales during the historically low-sales month.
- Plan marketing campaigns and discounts during January and March to increase sales during the historically high-sales months.
- Aim to replicate the sales patterns of Q1 2019 and Q3 2018.
- Keep the stock of products sold in Q2 consistent with the previous years to maintain sales levels.
- Keep in mind the limited data for 2020 while making sales and marketing decisions.

03
Market Basket Analysis

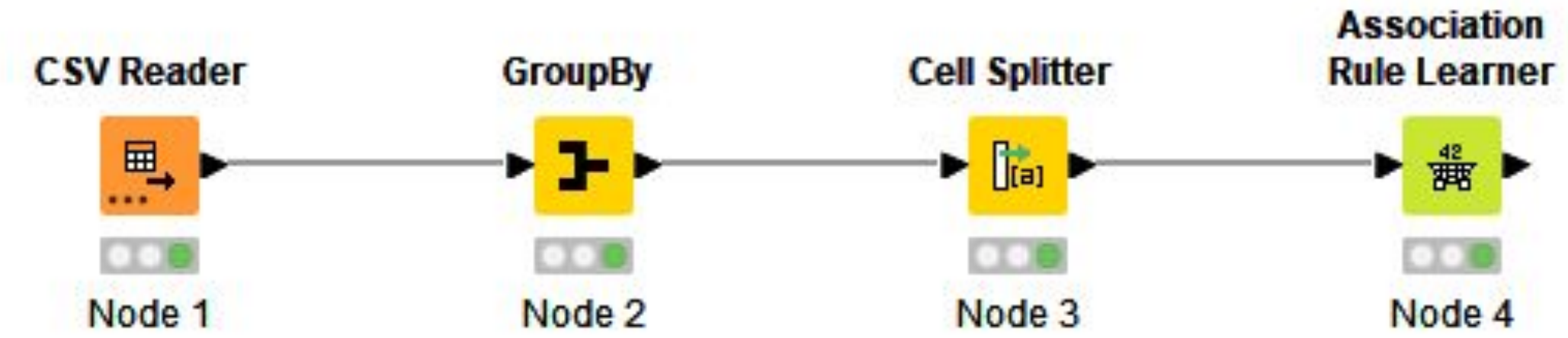
- Market Basket Analysis Meaning
- MRA KNIME WorkFlow & Qutput Table

Market Basket Analysis

- Definition: Market Basket Analysis is a statistical technique that analyzes customer purchase patterns to identify associations between different products. It helps businesses understand which products are frequently purchased together and how customers' buying habits affect sales.
- Data: To conduct market basket analysis, businesses need transactional data that includes details such as customer ID, product ID, and transaction date. This data is then used to create a matrix that represents the relationships between different products.
- Association Rules: Association rules are used to identify the strength of the relationship between different products. These rules are expressed in terms of support, confidence, and lift. Support refers to the frequency of co-occurrence of items in a transaction, while confidence measures the probability that if a customer buys one item, they will also buy another. Lift measures the degree of correlation between two items.
- Applications: Market Basket Analysis is used in a variety of industries, including retail, e-commerce, and marketing. Retailers use this technique to optimize product placement and promotions. E-commerce companies use it to personalize product recommendations, and marketers use it to develop targeted advertising campaigns.
- Benefits: Market Basket Analysis helps businesses increase revenue by identifying cross-selling opportunities and developing targeted promotions. It also helps improve customer satisfaction by providing personalized recommendations and improving the overall shopping experience

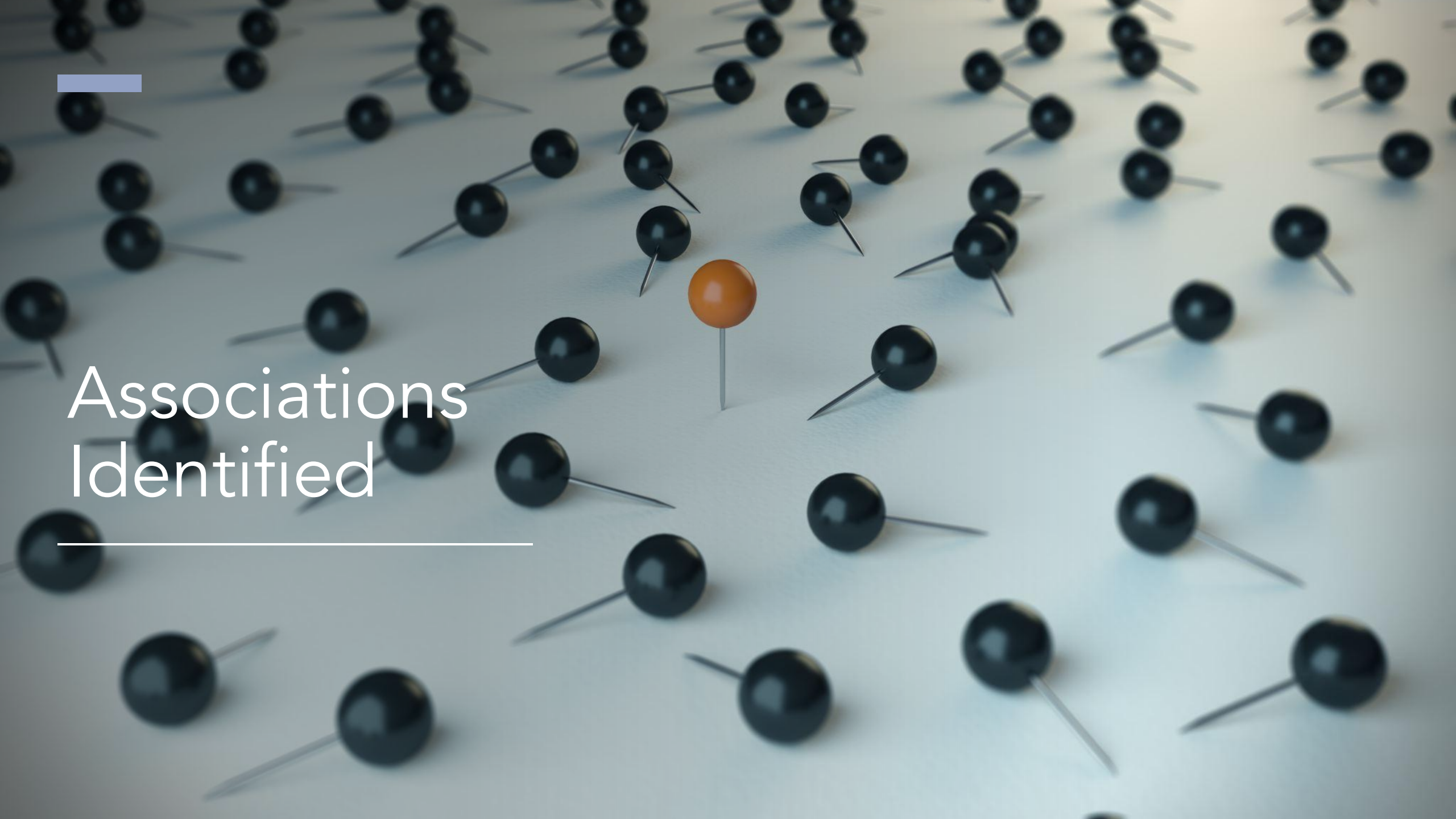


MRA KNIME Workflow



Output Table

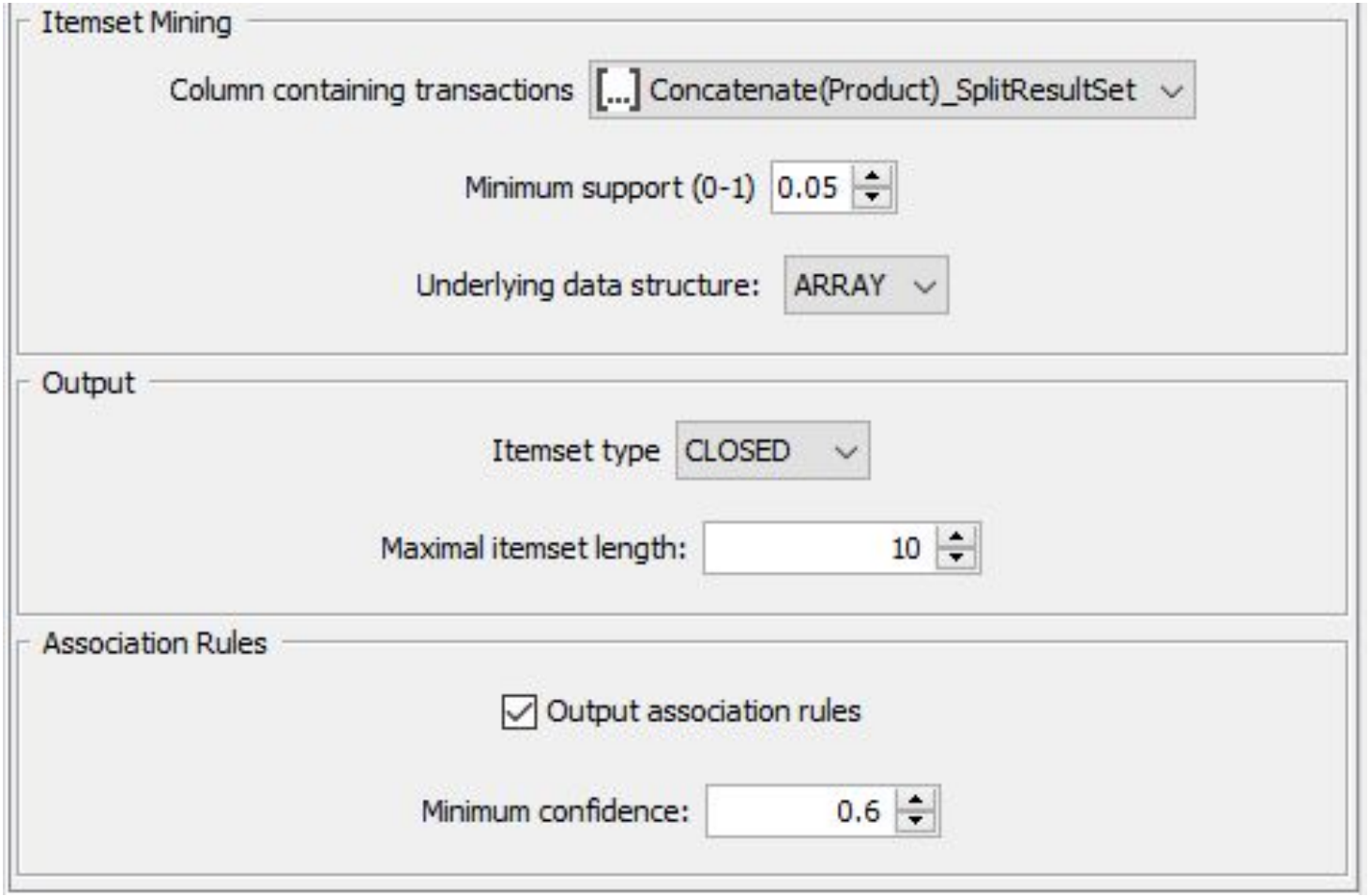
Row ID	I Order_id	S Concatenate(Product)	[...] Concatenate(Product)_SplitResultSet
Row0	1	yogurt, pork, sandwich bags, lunch meat, all- purpose, flour, soda, butter, beef, alumi...	[yogurt,pork,sandwich bags,...]
Row1	2	toilet paper, shampoo, hand soap, waffles, cheeses, mixes, milk, sandwich bags, laund...	[toilet paper,shampoo,hand soap,...]
Row2	3	soda, pork, soap, ice cream, toilet paper, dinner rolls, hand soap, spaghetti sauce, milk...	[soda,pork,soap,...]
Row3	4	cereals, juice, lunch meat, soda, toilet paper, all- purpose	[cereals,juice,lunch meat,...]
Row4	5	sandwich loaves, pasta, tortillas, mixes, hand soap, toilet paper, paper towels, flour, p...	[sandwich loaves,pasta,tortillas,...]
Row5	6	laundry detergent, toilet paper, eggs, toilet paper, bagels, dishwashing liquid/detergen...	[laundry detergent,toilet paper,eggs,...]
Row6	7	individual meals, paper towels, tortillas, milk, ice cream, juice, dishwashing liquid/deterg...	[individual meals,paper towels,tortillas,...]
Row7	8	ice cream, juice, paper towels, waffles, soda, cheeses, poultry, toilet paper	[ice cream,juice,paper towels,...]
Row8	9	juice, poultry, coffee/tea, coffee/tea, dishwashing liquid/detergent	[juice,poultry,coffee/tea,...]
Row9	10	ketchup, coffee/tea, toilet paper, pork, flour, milk, soda, dishwashing liquid/detergent, ...	[ketchup,coffee/tea,toilet paper,...]
Row10	11	sandwich loaves, ice cream, soda, bagels, dishwashing liquid/detergent, eggs, sugar, ...	[sandwich loaves,ice cream,soda,...]
Row11	12	pork, tortillas, pork, shampoo, lunch meat, pasta, juice, bagels, bagels, laundry deterg...	[pork,tortillas,shampoo,...]
Row12	13	sugar, fruits, all- purpose, aluminum foil, laundry detergent, individual meals, flour, por...	[sugar,fruits,all- purpose,...]
Row13	14	fruits, dinner rolls, individual meals, shampoo, ketchup, cereals, sandwich bags, laundr...	[fruits,dinner rolls,individual meals,...]
Row14	15	individual meals, ice cream, cereals, paper towels, bagels, mixes, lunch meat, juice, toil...	[individual meals,ice cream,cereals,...]
Row15	16	sugar, sandwich bags, flour, juice, milk, paper towels, cereals, sandwich bags, pasta, s...	[sugar,sandwich bags,flour,...]
Row16	17	milk, hand soap, pasta, individual meals, spaghetti sauce, cereals, sandwich loaves, ha...	[milk,hand soap,pasta,...]
Row17	18	sandwich bags, toilet paper, bagels, shampoo, coffee/tea	[sandwich bags,toilet paper,bagels,...]
Row18	19	individual meals, laundry detergent, coffee/tea, eggs, aluminum foil, beef, juice, flour, ...	[individual meals,laundry detergent,coffee/tea,...]
Row19	20	shampoo, dishwashing liquid/detergent, yogurt, juice, sugar, soap, sandwich loaves, b...	[shampoo,dishwashing liquid/detergent,yogurt,...]
Row20	21	waffles, fruits, all- purpose, pork, juice, bagels, mixes	[waffles,fruits,all- purpose,...]



Associations Identified

Association Rule Parameters

- Support of Minimum: 0.05
- Maximum Item Set Length : 10
- Minimum Confidence Level:0.6



Market basket analysis, support, confidence, and lift values

- In market basket analysis, support, confidence, and lift values are used to measure the strength of association between items in a transaction dataset.
- Support: It is the probability of observing the items together in a transaction. It is calculated as the number of transactions that contain both items divided by the total number of transactions. It measures how frequent the itemset occurs in the dataset. High support indicates that the itemset is popular and should be considered for promotion or placement together.
- Confidence: It is the conditional probability that a transaction containing one item also contains another item. It is calculated as the number of transactions containing both items divided by the number of transactions containing the first item. It measures the strength of the association between two items. High confidence indicates that the items are likely to be bought together, and can be used to recommend or suggest items to customers.
- Lift: It is the measure of how much more often two items occur together than expected if they were independent of each other. It is calculated as the support of the itemset divided by the product of the individual supports of the items. A lift value of 1 indicates that the items are independent, while a value greater than 1 indicates a positive association between the items. A lift value less than 1 indicates a negative association between the items. High lift indicates that the items have a strong association and can be used for cross-selling or bundling.

Association Rules

- Association rules are a technique used to find relationships or associations between items in a large dataset. These rules are based on the concept of frequent itemsets, which are sets of items that appear together frequently in a transactional dataset.
- 24 rules have been found with the dataset and set parameters.

Row ID	Support	Confide...	Lift	Conseq...	implies	Items
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 detergent,mixes]
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 detergent]
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]

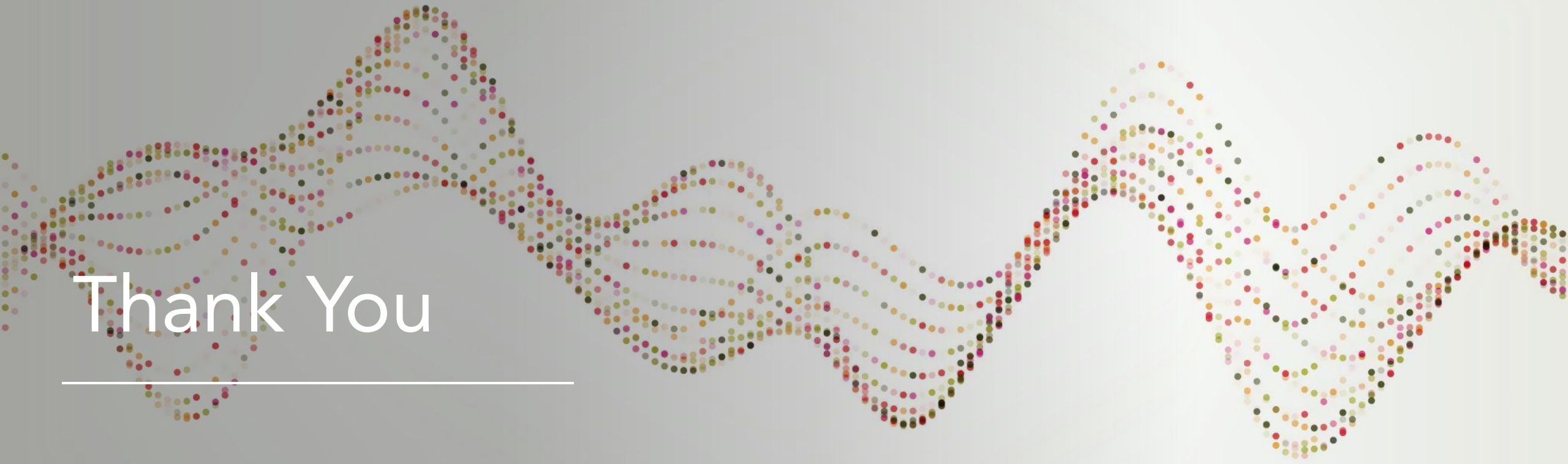
Recommendation

Recommendations:

- Offer a "Buy Two Get One Free" promotion on yogurt, poultry, and aluminum foil to encourage customers to purchase more items at once.
- Create a combo deal where customers can purchase cereals, bagels, and sandwich bags together at a discounted price.
- 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.
- Create a "Paper Products Bundle" offer that includes paper towels, toilet paper, and/or tissues at a discounted price.
- These discount offers and combos can help increase sales by providing customers with more value for their money and encouraging them to purchase more items. It is important to promote these offers through in-store signage, advertisements, and social media to ensure customers are aware of the deals available.

Summary

- The analysis identified the products that are often purchased together by customers, which can help the store to optimize its product placement and promotions.
- Yogurt, poultry, aluminum foil, cheeses, cereals, and dinner rolls are some of the most frequently purchased products.
- Some of the product associations are unexpected, such as poultry with dishwashing liquid/detergent, laundry detergent, and mixes.
- The analysis suggests that offering discounts or combos, such as "buy two get one free," on certain products can encourage customers to purchase more.
- The store can also consider placing complementary products in close proximity to each other to increase the likelihood of customers making additional purchases.
- Overall, the market basket analysis can help the store to better understand customer behavior and preferences, and to make informed decisions about product placement and promotions.



Thank You