Grocery Store's Data
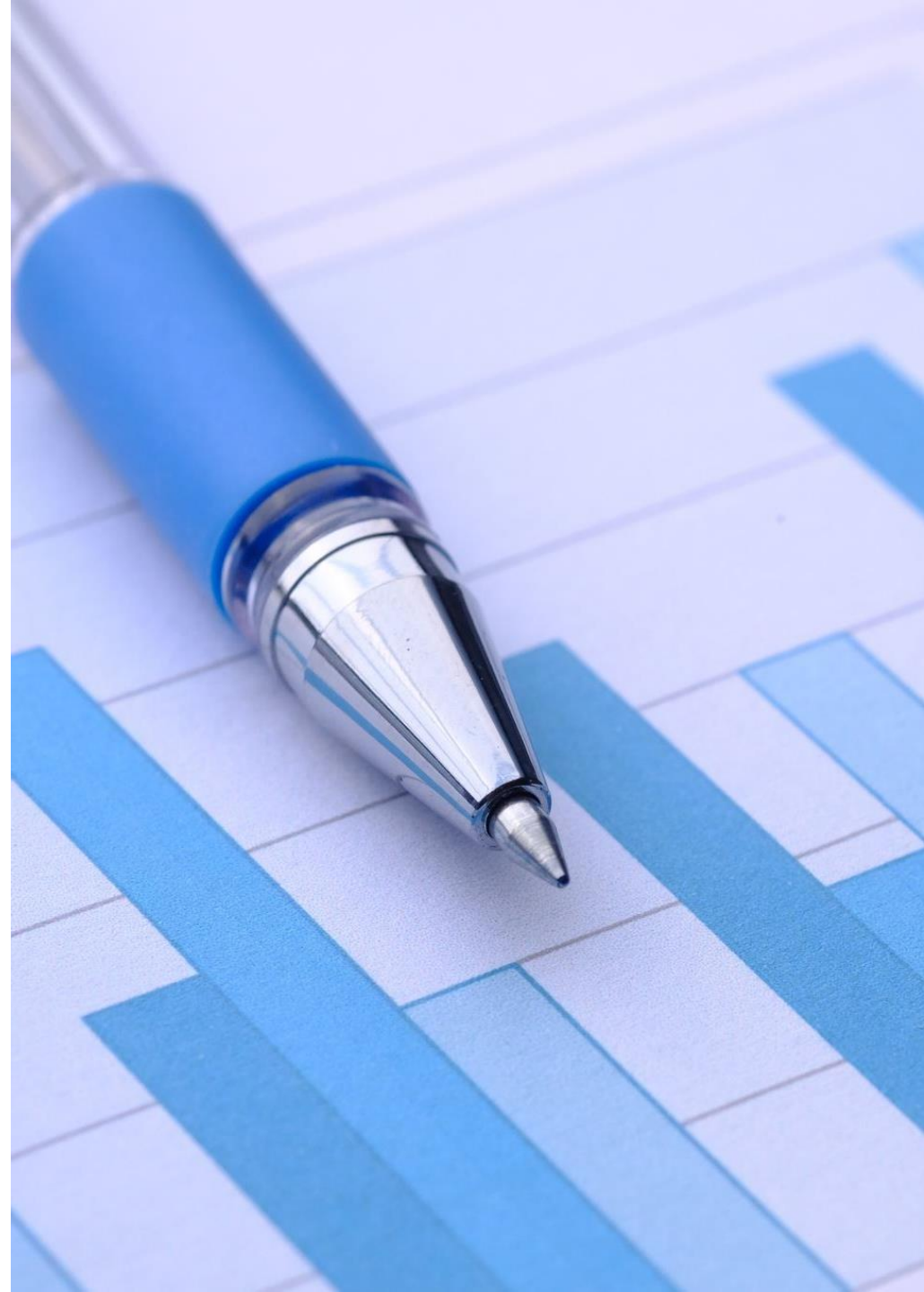MRA Project-Milestone 2

CONTENTS:

AGENDA

- Executive Summary of the data

- Exploratory Data Analysis

- Market Basket Analysis

- Associations Identified

- Recommendation

EXECUTIVE SUMMARY

- Problem statement

- Executive Summary & Data Dictionary

- Assumptions about data

PROBLEM STATEMENT

A grocery store has provided its transactional data for analysis. Your task is to uncover the most frequently purchased item combinations from customer orders. The store currently doesn't offer any combo deals. Based on your analysis, you are expected to recommend the most effective product combos and promotional offers to help boost sales and enhance customer experience.

DATA 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

Feature	Description
Date	Date of product sold
Order_id	ID of the order
Product	Name of the product sold

Duplicate Values:

While it's usually recommended to remove duplicate rows to avoid skewing analysis, in this case, it's not appropriate. The dataset lacks a unique identifier, and rows are based on combinations of date, customer ID, and product purchased. Since multiple customers can buy the same product on the same date, removing duplicates could lead to loss of valid information.
Hence, duplicate rows were retained in the dataset.

ASSUMPTIONS:
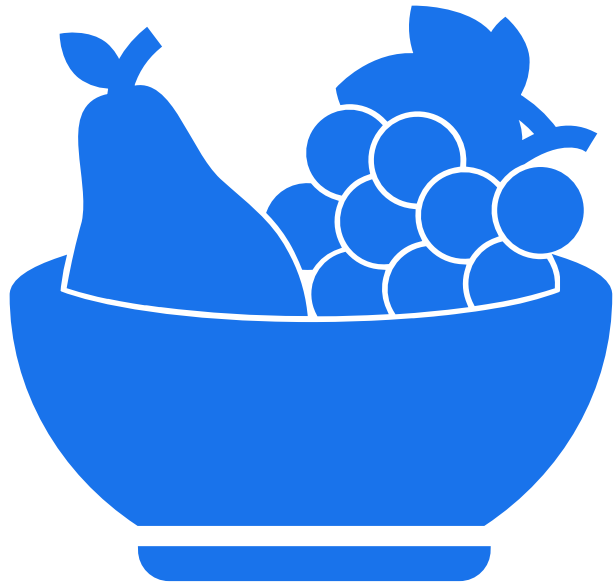
- The data represents a list of items purchased at a grocery store on various dates.

- 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.

Exploratory
Data Analysis
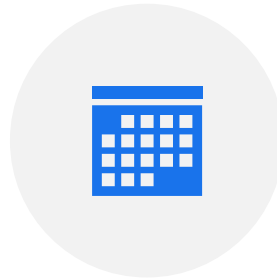
WEEKLY, MONTHLY, QUARTERLY, YEARLY WEEKDAY TRENDS IN SALES COUNT

PRODUCTS COUNTS & YEAR WISE TOP PRODUCTS

SUMMARY AND RECOMMENDATIONS

Yearly Count of Products Sold

Yearly Count- Sold Products

Date

Count of Order Id

10K
9K
8K
7K
6K
5K
4K
3K
2K
1K
0K

9,479
9,333
1,829
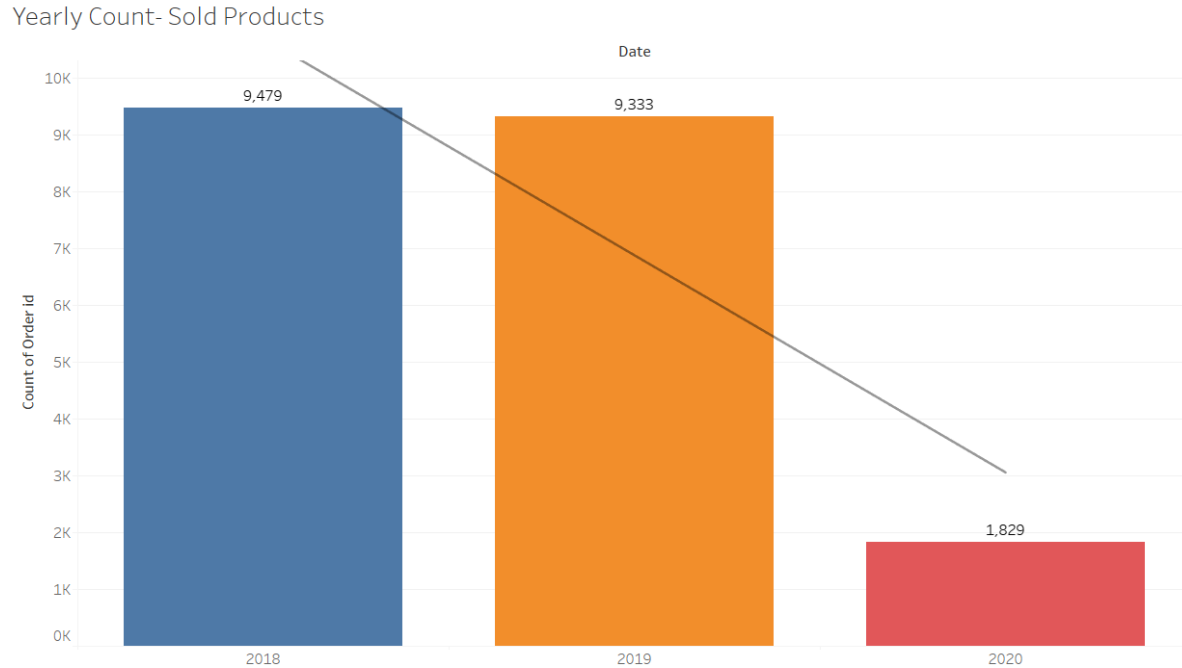
2018
2019
2020

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
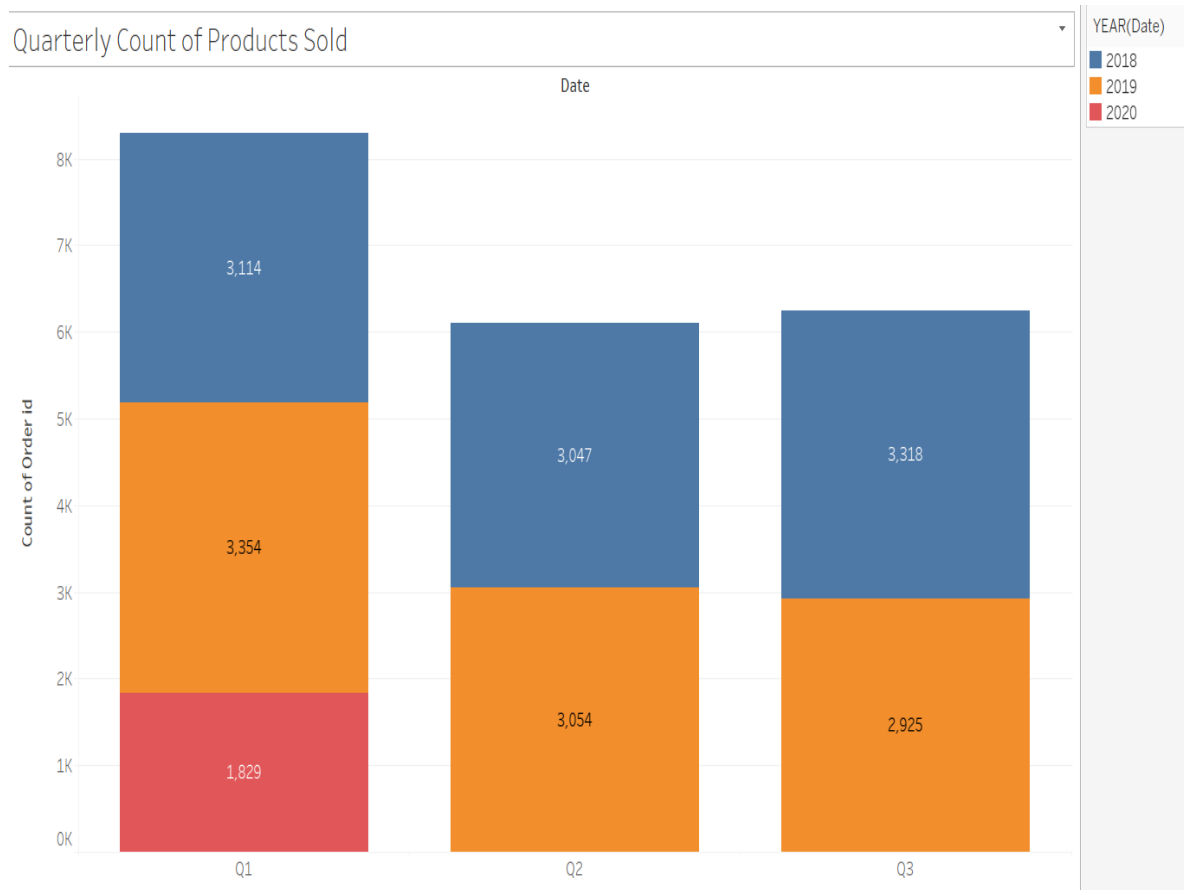
- As we have data till 26 feb 2020 that's why the count of products sold in Q1 is Hight.

- 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

Monthly Count of Products Sold
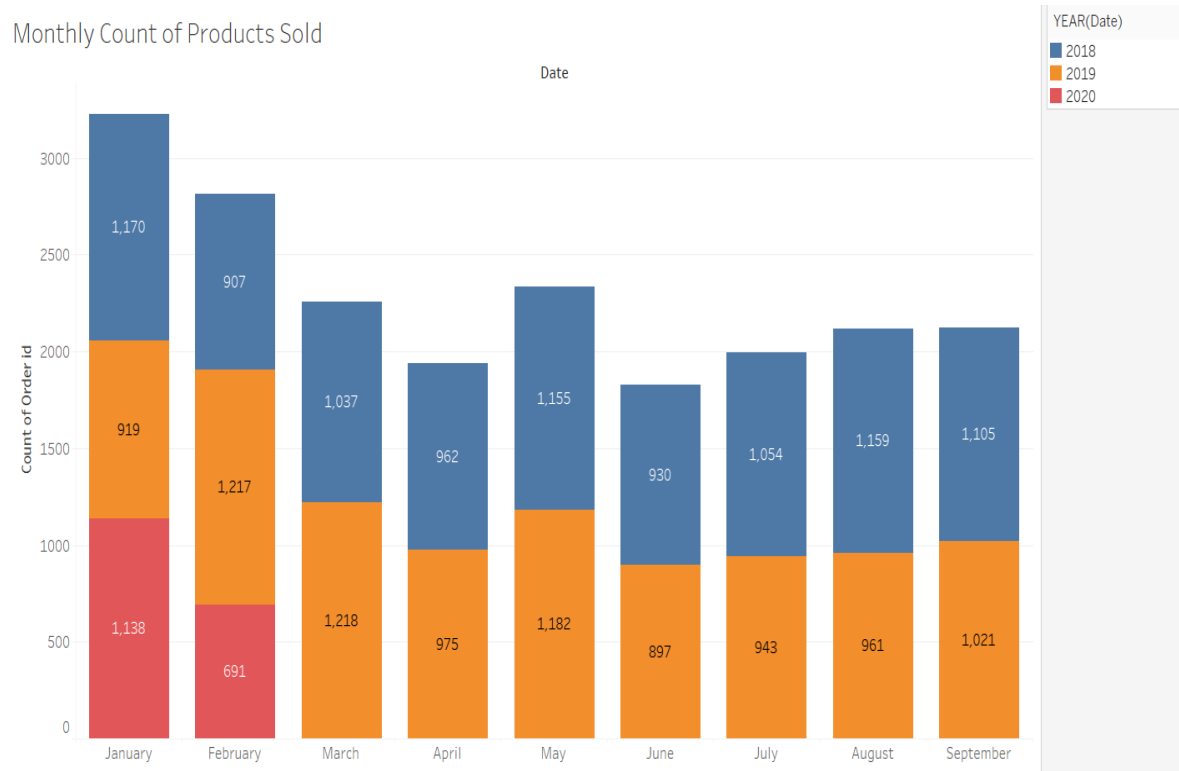
- In 2018 most of the products were sold in January and least were sold in February.

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

Day of the Month Count of Products Sold
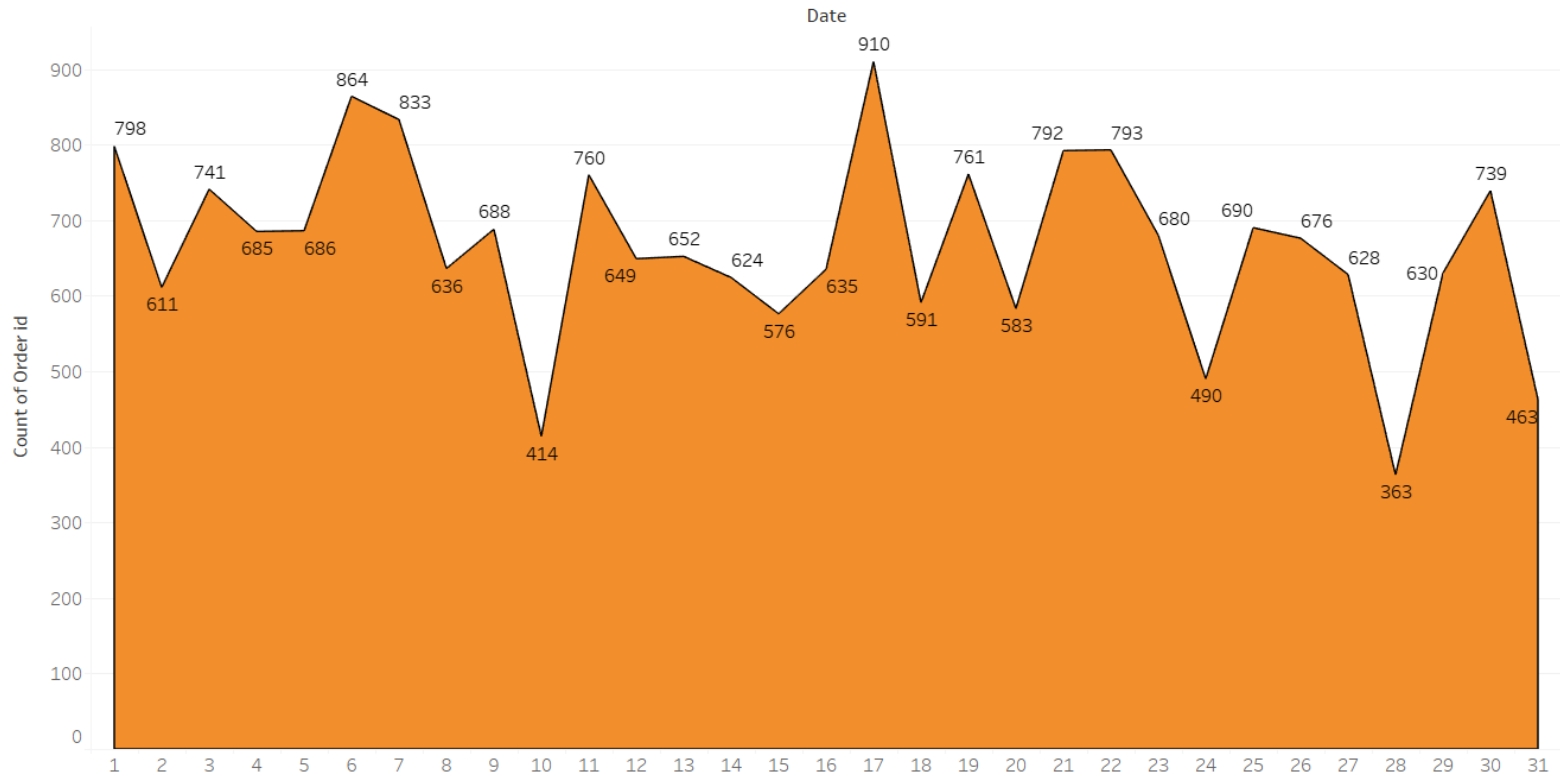
Product sales fluctuate throughout the month, with the highest volume occurring around the 17th, and significant dips around the 10th and the end of the month

Weekly Count of Products Sold

Weekly Count of Products Sold
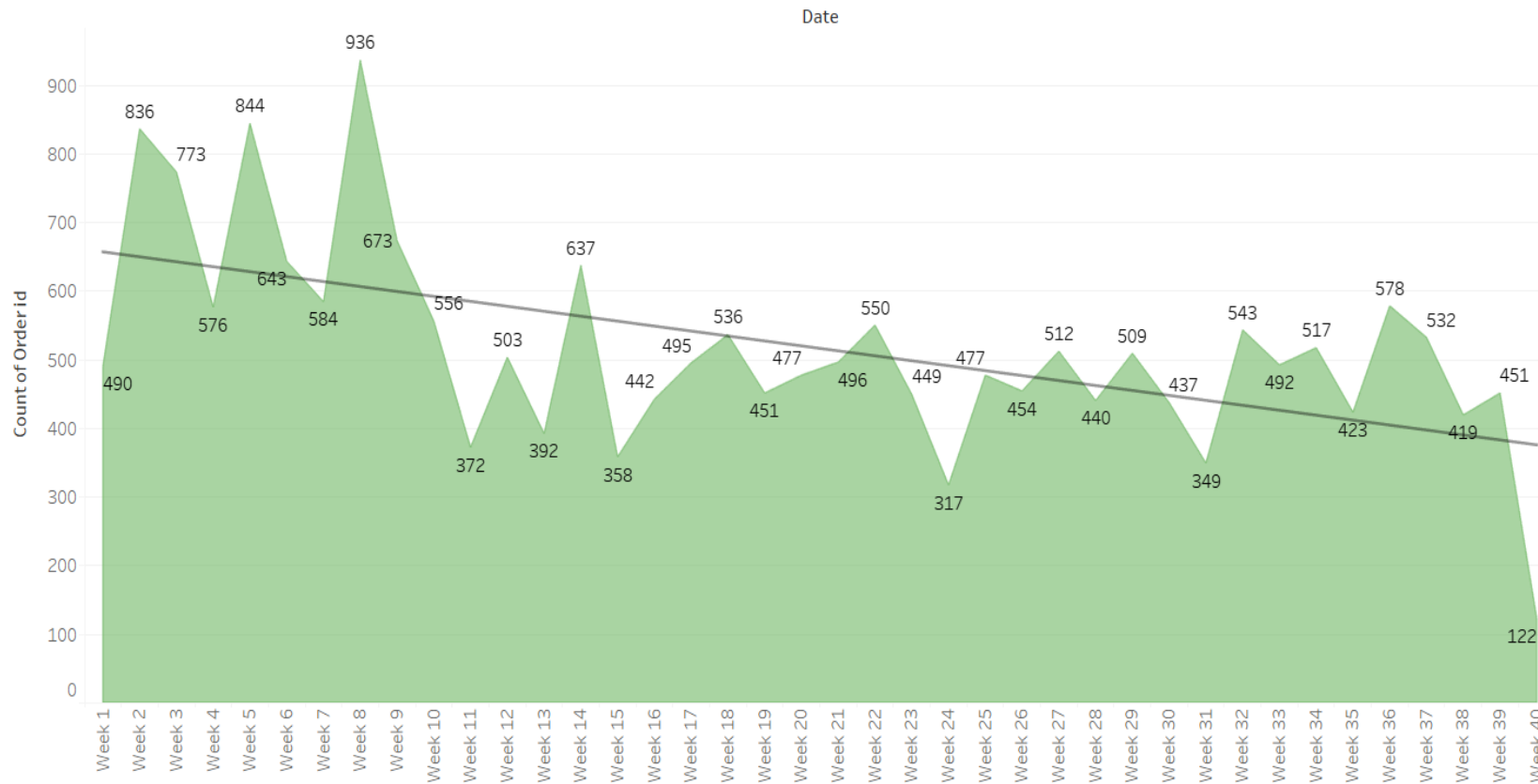
Product sales generally decline over the weeks, with a peak in the first week and a significant drop by week 40.

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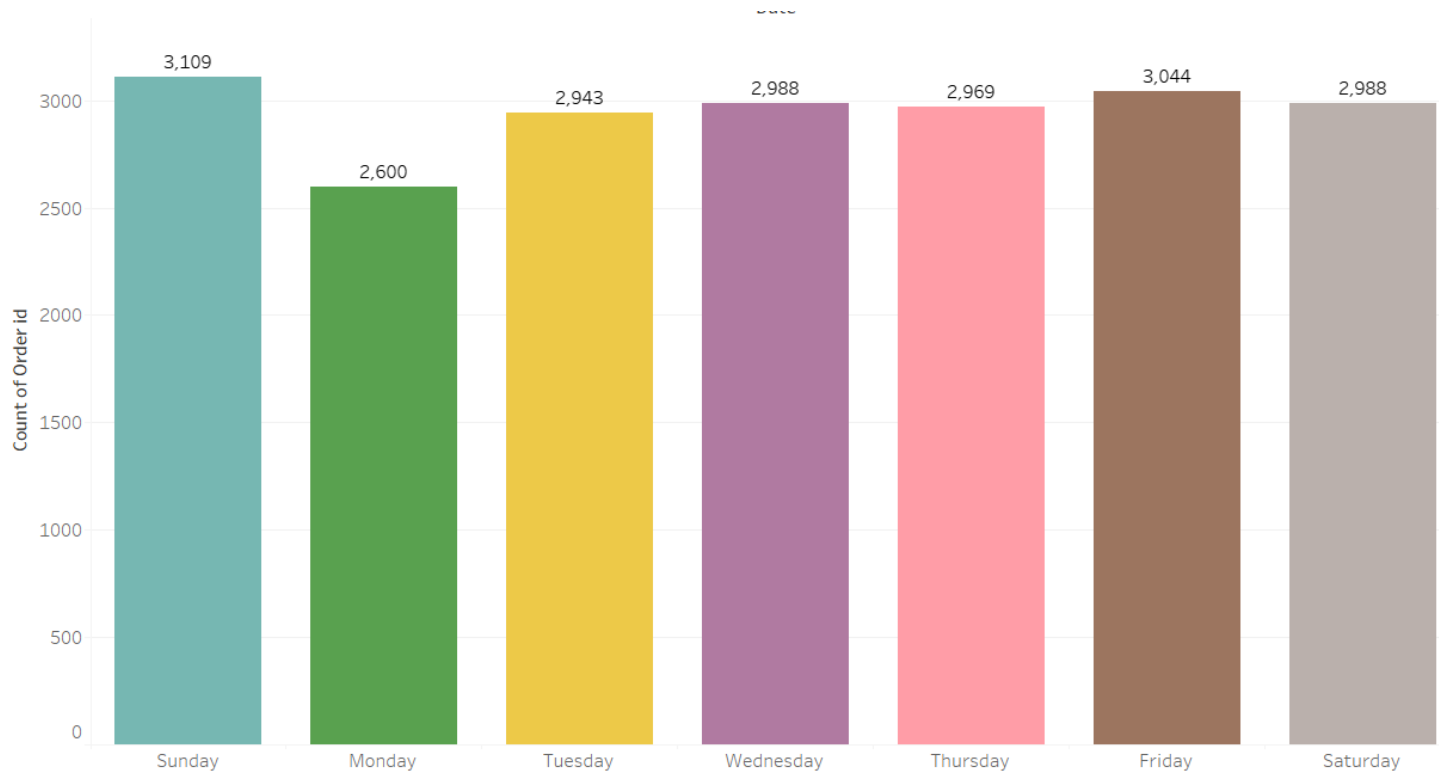
- Most of the products were sold on Sundays.
- Least products were sold on Mondays.
- On other days sales in consistent.

Daily Count of Sales
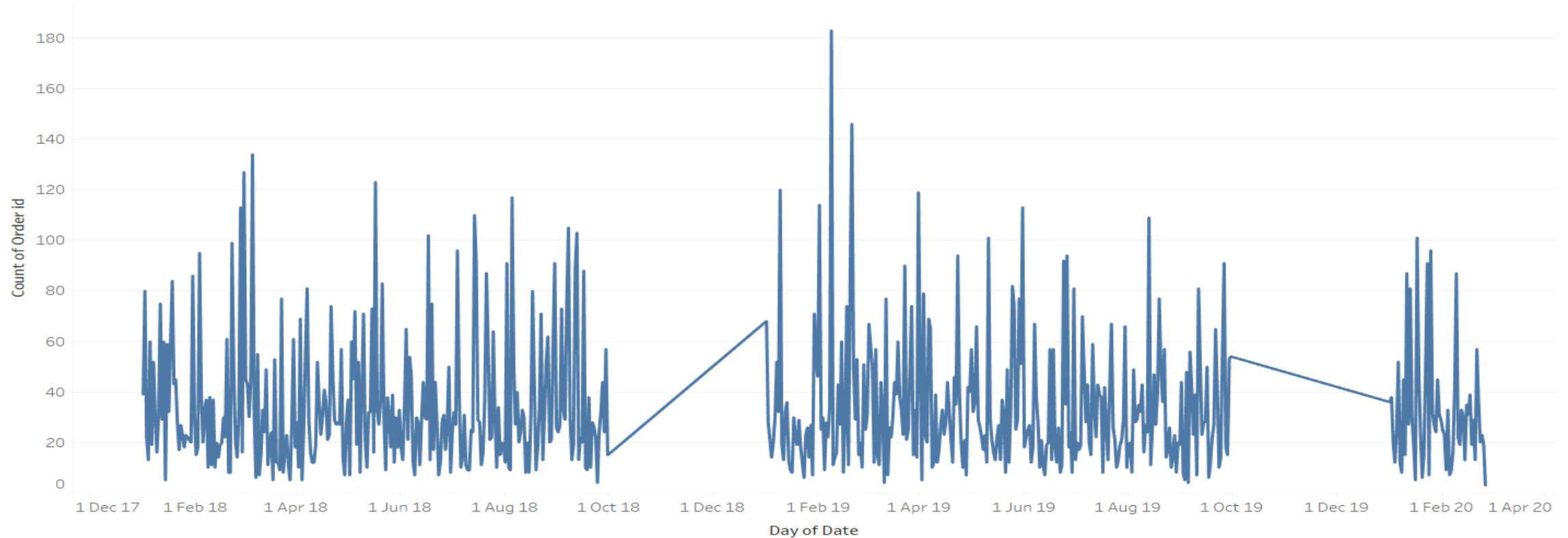
Daily sales fluctuate significantly with several peaks and troughs, showing no clear trend but indicating high variability

Count of Products Sold

poultry Product Sold : 640	soap Product Sold : 574	dinner rolls Product Sold : 567	butter Product Sold : 555	flour Product Sold : 555	milk Product Sold : 555	mixes Product Sold : 554	all-purpose Product Sold : 551

soda
Product Sold : 597

bagels
Product Sold : 573

aluminum foil
Product Sold : 566

cereals
Product Sold : 591

lunch meat
Product Sold : 573

coffee/tea
Product Sold : 565

dishwashing
liquid/detergent
Product Sold : 551

laundry
detergent
Product Sold
: 542

pasta
Product Sold
: 542

sandwich
bags
Product Sold
: 536

ketchup
Product Sold : 548

ice cream
Product Sold : 579

eggs
Product Sold : 570

shampoo
Product Sold : 562

yogurt
Product Sold : 545

spaghetti sauce
Product Sold : 536

fruits
Product
Sold :
529

cheeses
Product Sold : 578

juice
Product Sold : 570

beef
Product Sold : 561

individual meals
Product Sold : 544

sugar
Product Sold : 533

waffles
Product Sold : 575

toilet paper
Product Sold : 569

paper towels
Product Sold : 556

tortillas
Product Sold : 543

pork
Product Sold : 531

hand soap
Product Sold : 502

Poultry has the highest sales volume, while "Hand Soap" has the lowest, with most products selling in a relatively narrow range between 500 and 600 units.

Count of Products Sold Yearly
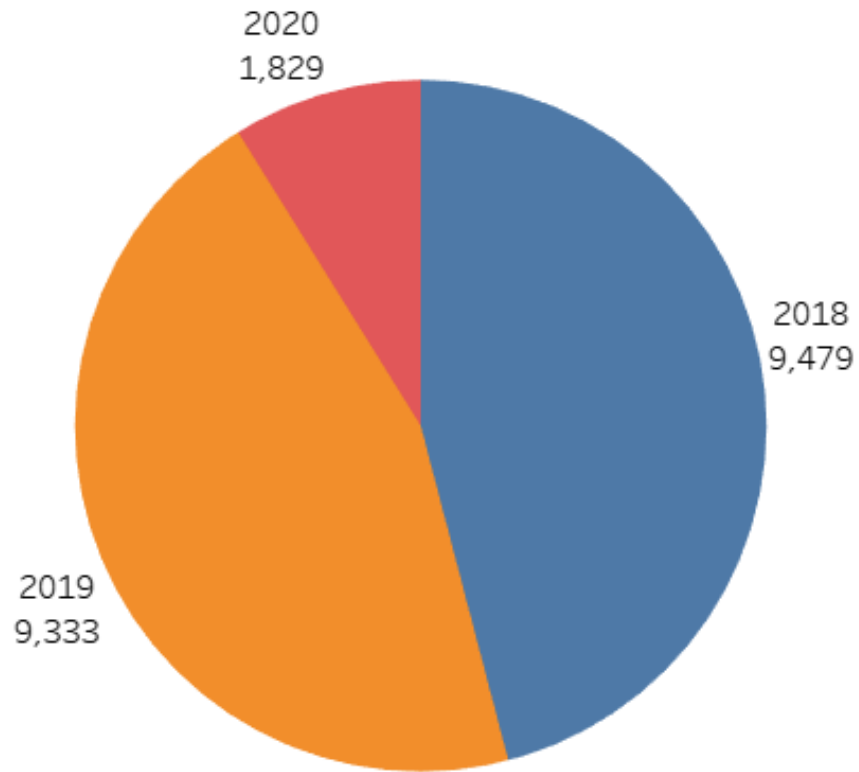
In 2018, Most products were sold

Count of Eatable Products Sold

| poultry 640 | waffles 575 | dinner rolls 567 | coffee/tea 565 | beef 561 | butter 555 | flour 555 |

| soda 597 | bagels 573 | milk 555 | individual meals 544 | tortillas 543 | pasta 542 | spaghetti sauce 536 |

| cereals 591 | lunch meat 573 | mixes 554 | | | | |

| ice cream 579 | eggs 570 | ketchup 548 | sugar 533 | | fruits 529 | sandwich loaves 523 |

| cheeses 578 | juice 570 | yogurt 545 | pork 531 | | | |

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

Count of Non-Eatable Products Sold

Product	Count
soap	574
shampoo	562
dishwashing liquid/detergent	551
laundry detergent	542
toilet paper	569
paper towels	556
sandwich bags	536
aluminum foil	566
all-purpose	551
hand soap	502
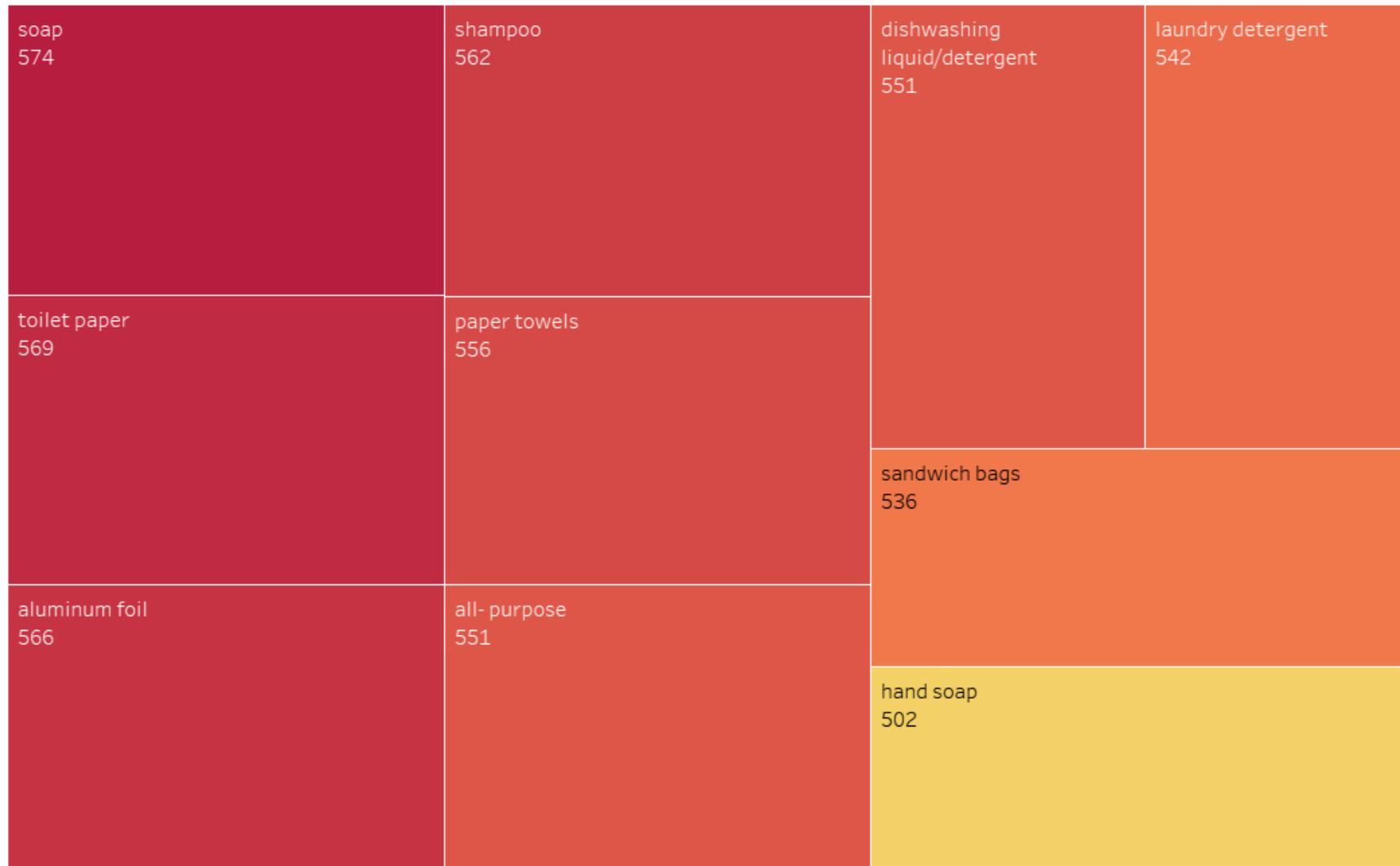
- There are 9 products in this category.
- Highest Sold Products : Soap, Toilet Paper.
- Lowest Sold Products: Hand soap.

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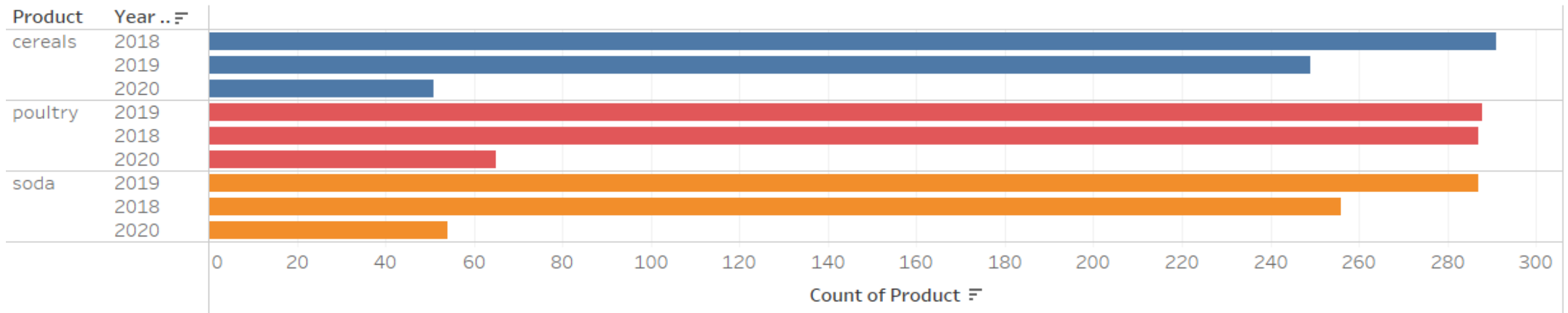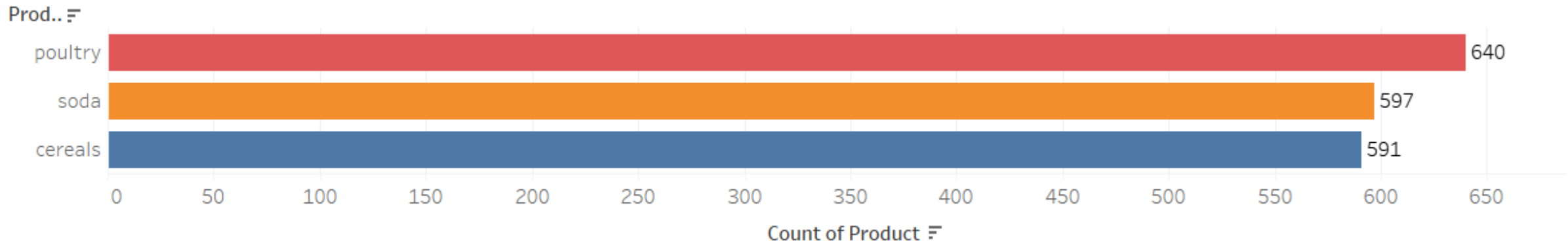
We can see poultry, cereals and soda are highly sold products over the years

Count of Products sold in 2018

Product

cereals — 291

poultry — 287

soda — 256
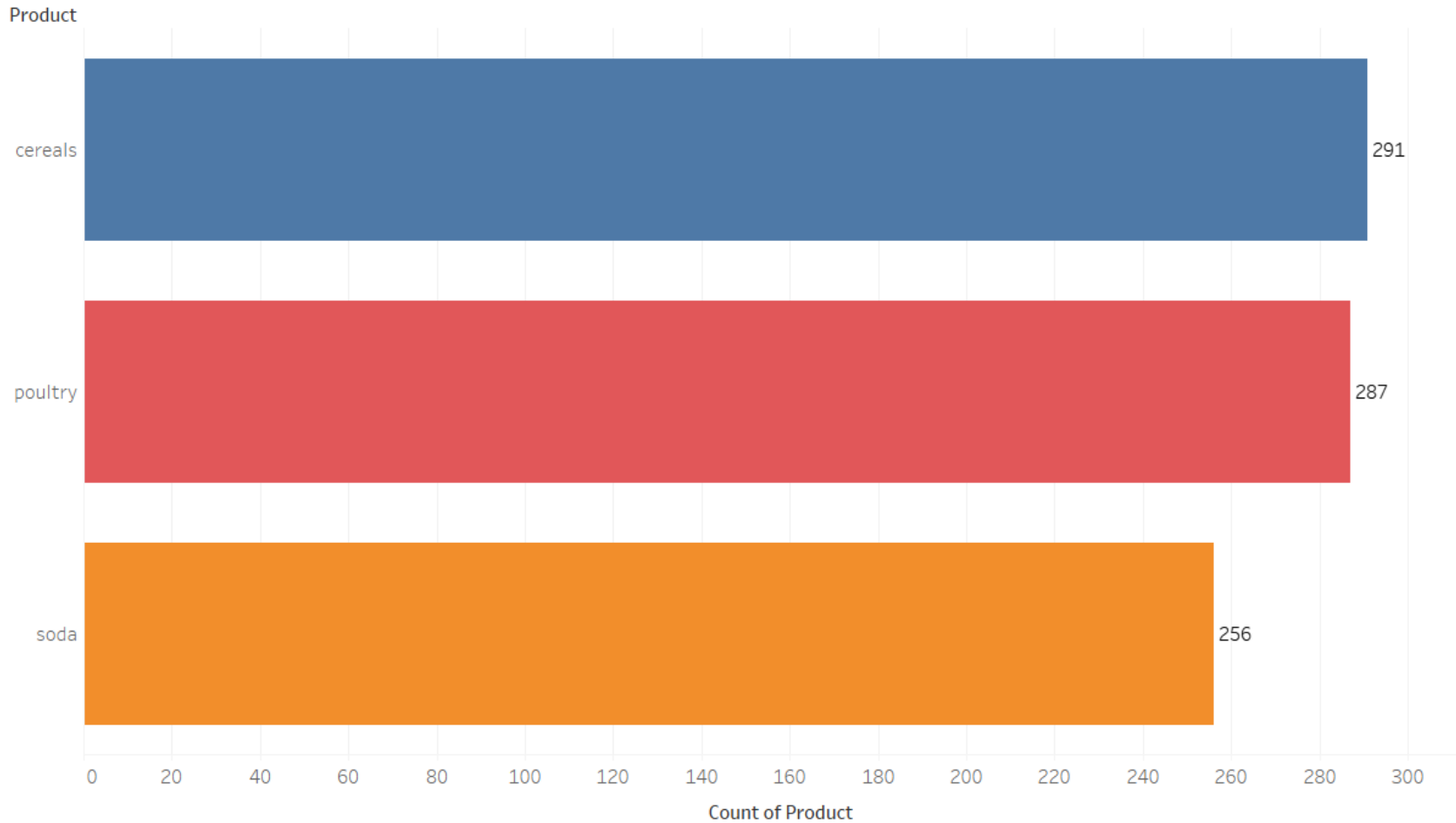
Count of Product

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

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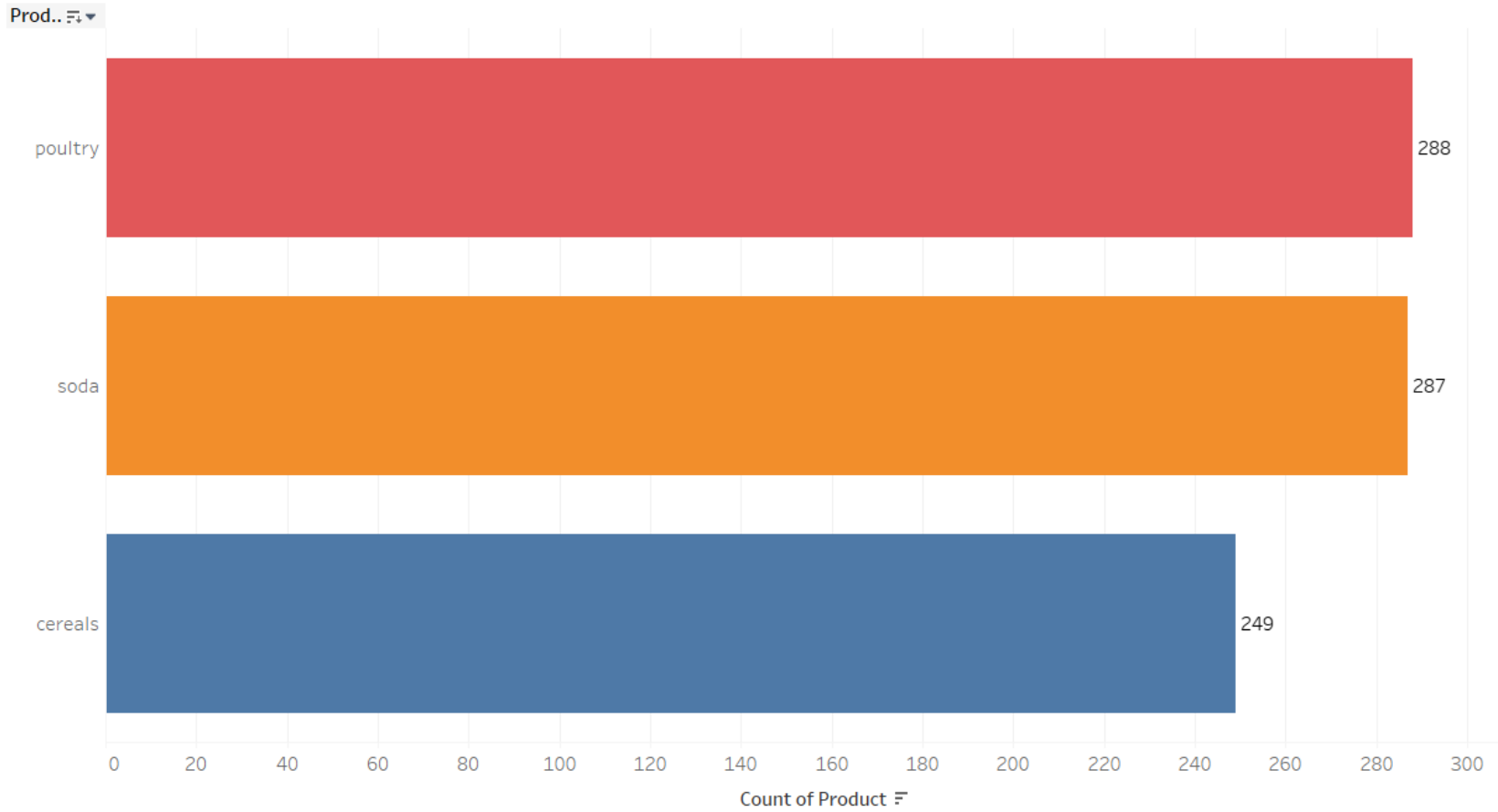
Poultry, Soda and Cereal are the top 3 Products sold in 2019

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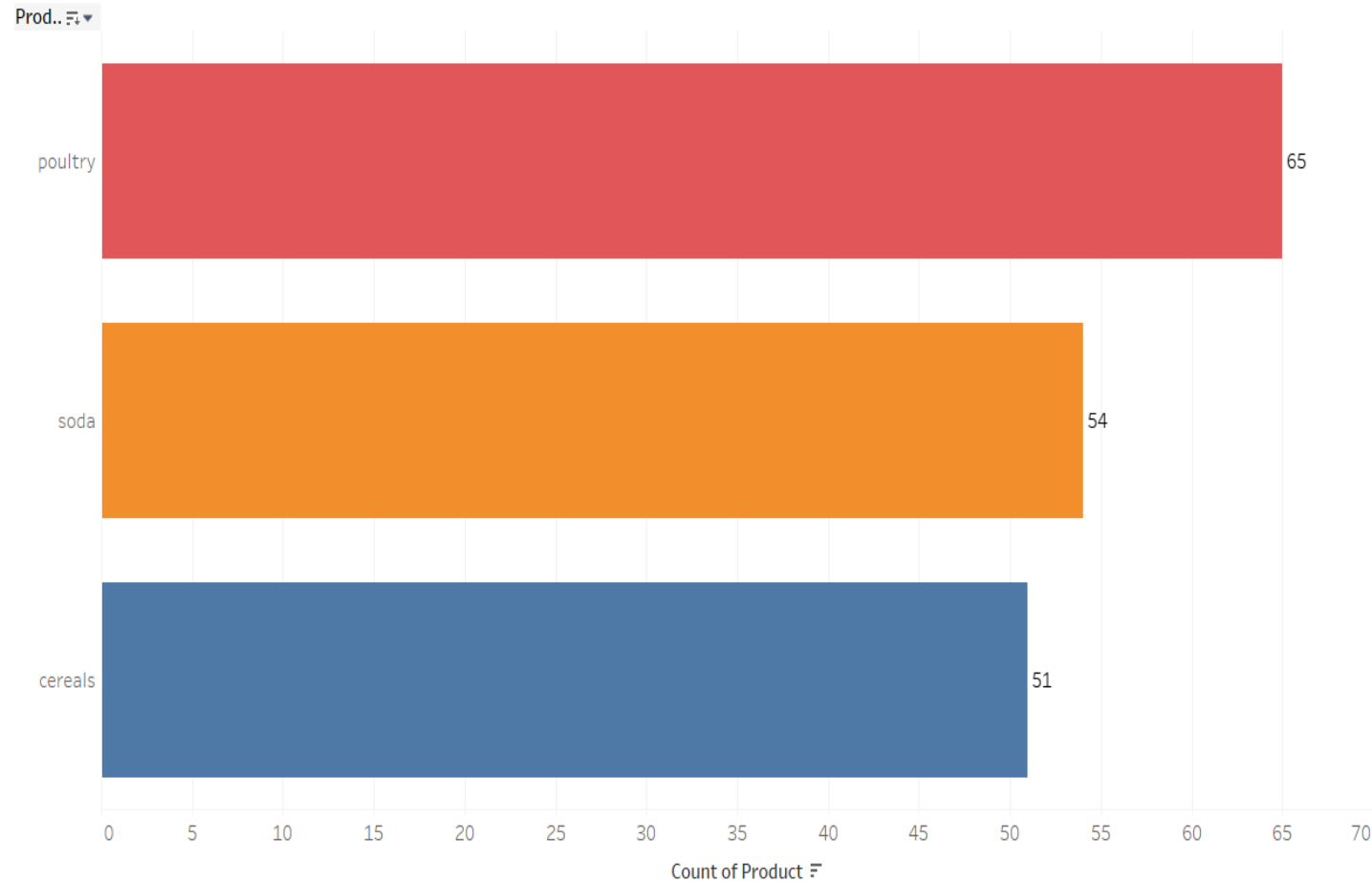
Top 3 products sold in 2020 January and February are dinner role, poultry and pork

SUMMARY

- **Top-selling products** in **2018 and 2019** were **Cereals, Poultry, and Soda**.
- For **Jan-Feb 2020**, the leading items were **Dinner Rolls, Poultry, and Pork**.
- **Overall**, the consistently high-selling products across the years were **Poultry, Cereals, and Soda**.
- In the **non-edible** category, **Soap and Toilet Paper** had the highest sales, while **Hand Soap** had the lowest.
- Among **edible items, Poultry, Soda, and Cereals** topped the list, whereas **Pork, Fruits, and Sandwich Loaves** were least sold.
- **Sunday** saw the highest product sales, while **Monday** had the lowest.
- In **2018, January** had the highest sales; **February** the lowest.
- In **2019**, the peak was in **March**, and the dip was in **January**.
- The **highest quarterly sales** were in **Q1 2019** and **Q3 2018**.
- Product sales in **Q2** were nearly the same for both **2018 and 2019**.
- **2020** shows fewer sales, likely because the data only extends up to **February 26th**.

MARKET BASKET ANALYSIS

- Market Basket Analysis Meaning

- MRA KNIME Workflow & Output Table

What is 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
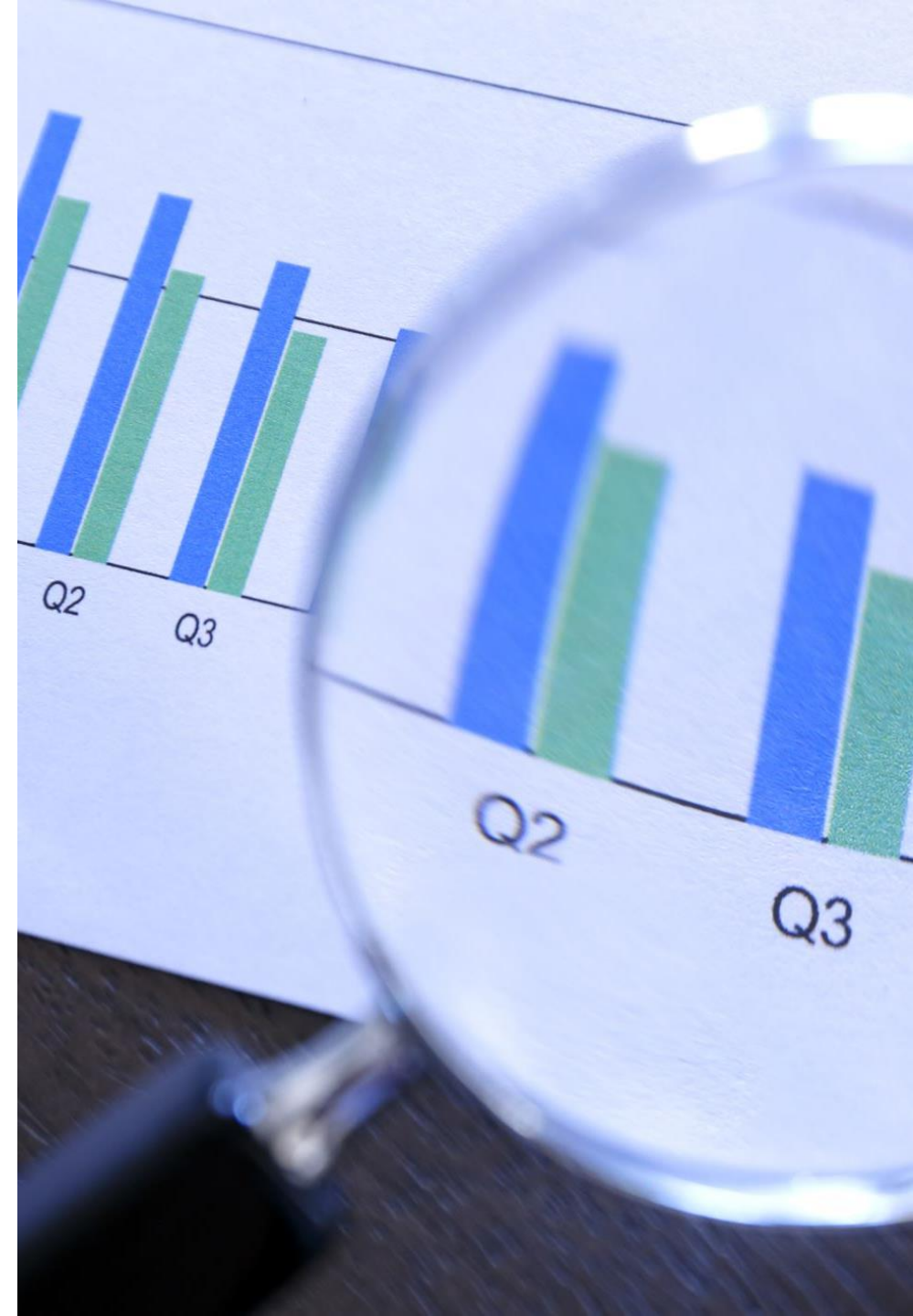
KNIME WORKFLOW

Output Table

Row ID	Order_id	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,...]

ASSOCIATION IDENTIED

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, 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

Row ID	D Support	D Confide...	D Lift	S Conseq...	S 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]

- 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 item sets, which are sets of items that appear together frequently in a transactional dataset.

- 24 rules have been found with the dataset and set parameters.
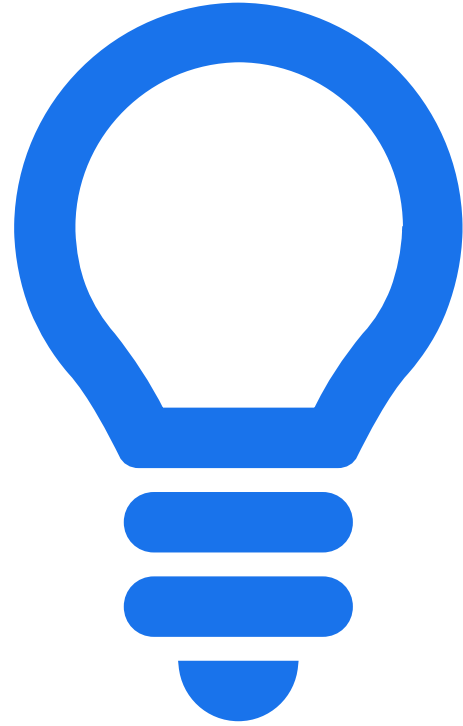
RECOMMENDATIONS

RECOMMENDATIONS

- Introduce a **"Buy 2, Get 1 Free"** offer on **yogurt, poultry, and aluminum foil** to boost bulk purchases.

- Launch a **combo deal** with **cereals, bagels, and sandwich bags** at a special price to promote bundled buying.

- Offer a **discount on mixes** when bought alongside **yogurt, poultry, or aluminum foil**.

- Provide **discounts on dinner rolls** when purchased with **spaghetti sauce or poultry**.

- Roll out a **"Paper Products Bundle"** including **paper towels, toilet paper, and tissues** at a reduced price.

•These promotions aim to **increase customer spending** and offer better value. For maximum impact, they should be **advertised through in-store displays, flyers, and social media channels** to ensure customer awareness and engagement.

SUMMARY

The analysis revealed key product combinations frequently bought together by customers, offering valuable insights for **optimizing product placement and promotional strategies**.

Popular items include **yogurt, poultry, aluminum foil, cheeses, cereals, and dinner rolls**.
Interestingly, some **unexpected product pairings** emerged—such as **poultry with dishwashing liquid, laundry detergent, and mixes**.

Introducing **promotions like "Buy Two Get One Free"** on these frequently paired products could **encourage higher purchase volumes**.

Additionally, placing **complementary items near each other** in-store can boost cross-selling opportunities.

In summary, this **market basket analysis** helps the store better understand customer preferences and make more **data-driven decisions** to enhance sales and customer satisfaction.