

Marketing & Retail Analysis Project

Grocery Store - Part 2

Vinish Vincent

Table of Content

- Executive Summary of the data
 - Problem statement & Executive Summary
 - Data Dictionary & Duplicate values
 - Data Statistics
 - Assumptions about data
- Exploratory Analysis & Insights
 - Weekly, Monthly, Quarterly, Yearly and Weekday Trends in Sales count
 - Products counts & Year Wise top products
 - Summary & Recommendations
- Market Basket Analysis
 - Market Basket Analysis – Association Rules
 - MBA KNIME WorkFlow
 - MBA KNIME WorkFlow Output
 - Association Rule Parameters
 - Market basket analysis, support, confidence, and lift values
- Recommendation
 - Recommendation

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

- The dataset contains information about product orders. It has 20,641 rows and three columns: Date, Order_id, and Product
- The dataset spans a range of dates from 01-01-2018 to 26-02-2020
- The most frequent date in the dataset is 08-02-2019, occurring 183 times
- The average order_id is approximately 576
- There are 37 unique products in the dataset
- The most frequently ordered product is "poultry," occurring 640 times
- There is no missing values
- There is 4730 duplicate values
- 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

Column Name	Description
Date	Dates of products sold
Order_id	Id of the order
Product	Name of the product

Duplicate Values

In this dataset, it's not advisable to remove duplicate rows because there is no unique identifier for each row. Each row represents a combination of a date, a customer ID, and a product purchased. Since multiple customers can purchase the same product on the same date, removing duplicates could result in the loss of valid data. Therefore, we choose not to remove duplicate values from the dataset

Data Statistics

- The data types of these columns include 1 int64, and 2 object
- There are no null values in dataset
- There are 4730 duplicate values which we choose to not to remove

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 20641 entries, 0 to 20640  
Data columns (total 3 columns):  
#   Column      Non-Null Count  Dtype  
---  ---  
0   Date        20641 non-null  object  
1   Order_id    20641 non-null  int64  
2   Product     20641 non-null  object  
dtypes: int64(1), object(2)  
memory usage: 483.9+ KB
```

	Order_id
count	20641.000000
mean	575.986289
std	328.557078
min	1.000000
25%	292.000000
50%	581.000000
75%	862.000000
max	1139.000000

```
Date      0  
Order_id  0  
Product    0  
dtype: int64
```

```
df.duplicated().sum()  
4730
```

Assumptions



- The data consists of grocery store purchase records with three columns: Date, Order ID, and item purchased(Product).
- Each entry in the data represents a item or items purchased
- Multiple customers can buy the same item on different dates, and there is no information about quantity or price.
- We have retained duplicate values in the dataset

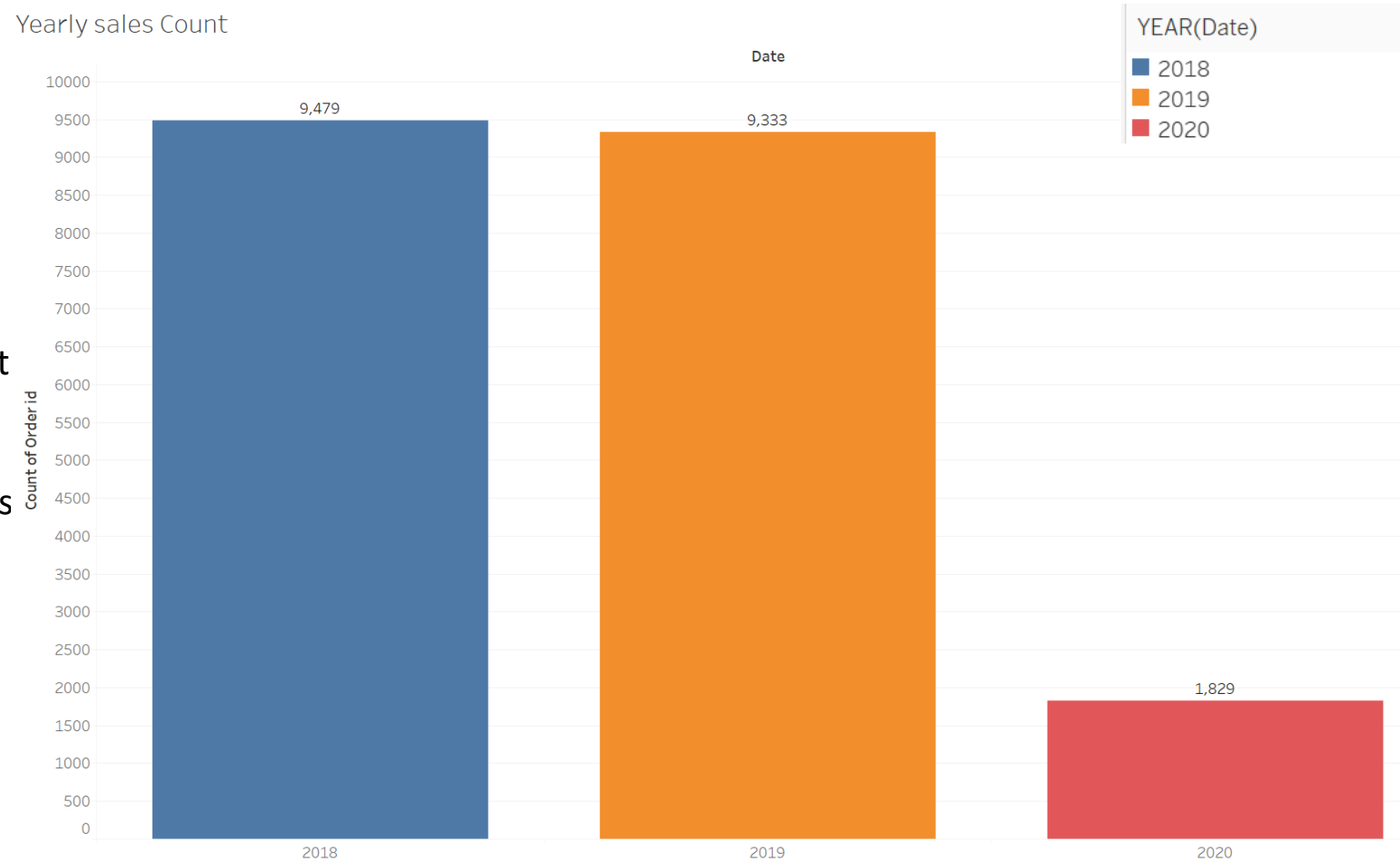
Exploratory Analysis & Insights

Weekly, Monthly, Quarterly, Yearly Weekday Trends in Sales count

Products counts & Year Wise top products

Yearly Count of Products Sold

- Year 2018 has highest product sales count, 9479 products
- There was a small dip in product sales in 2019, 9333
- As we have data till 26 feb 2020 that's why the count of products sold in 2020 is low

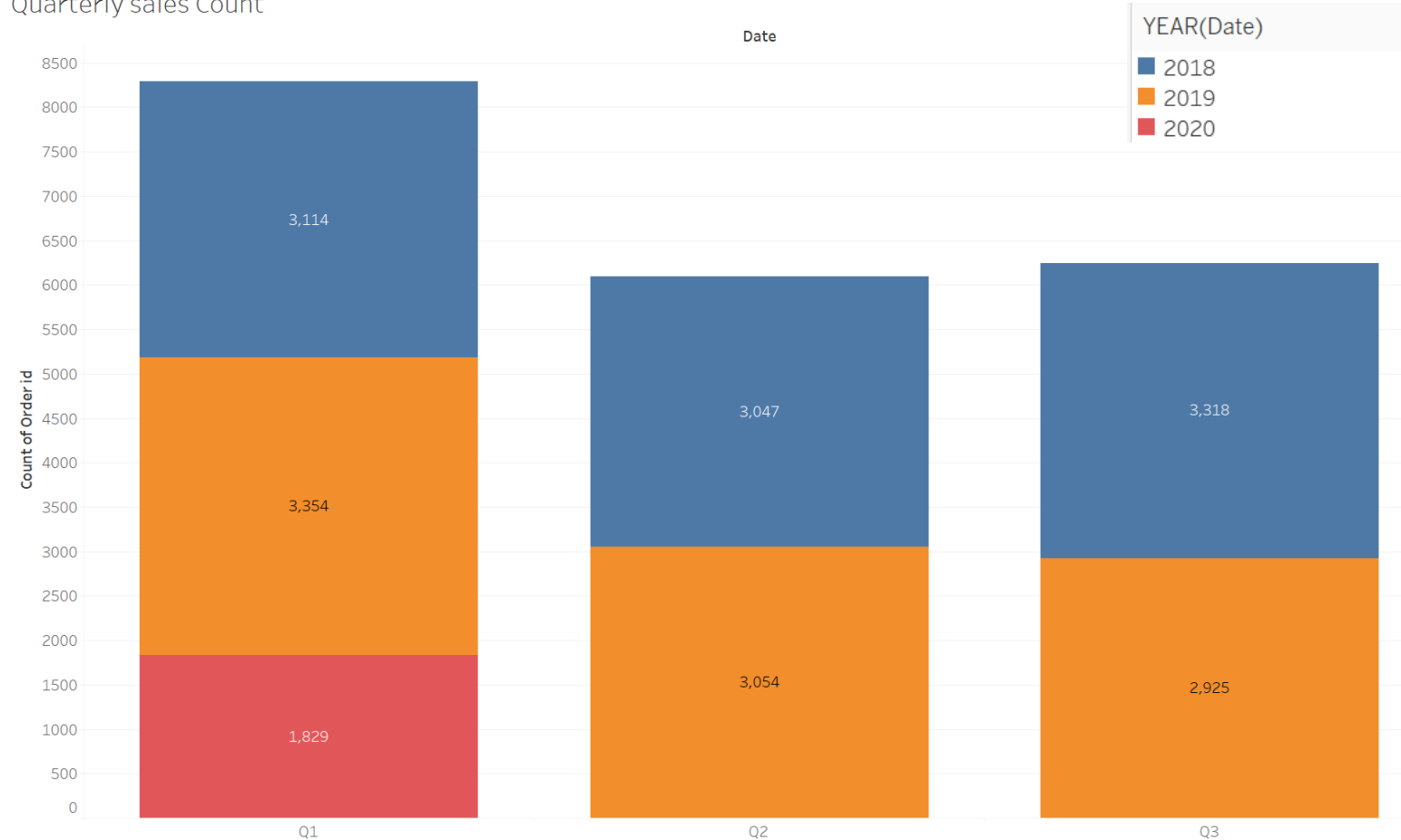


Note: Data consists of full years of 2018 and 2019. 2020 only have data till 26/02/2020

Quarterly Count of Products Sold

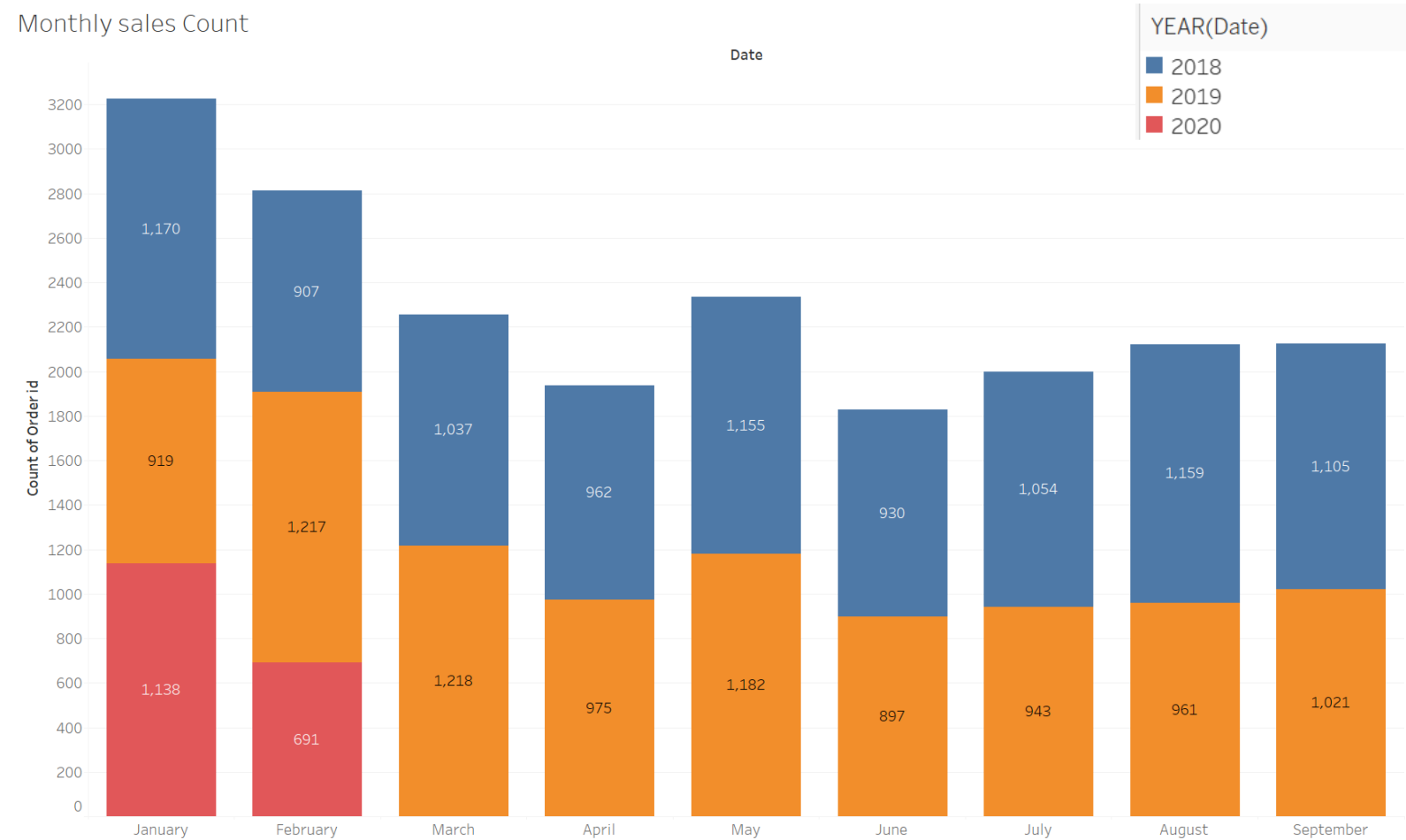
- As we have data till 26 Feb, 2020 that's why the count of products sold in Q1 is Highest
- In 2019 Q1 sales was highest
- In 2018 Q3 sales was highest
- Count of product sold in Q2 is approx. same in 2019 and 2018

Quarterly sales Count



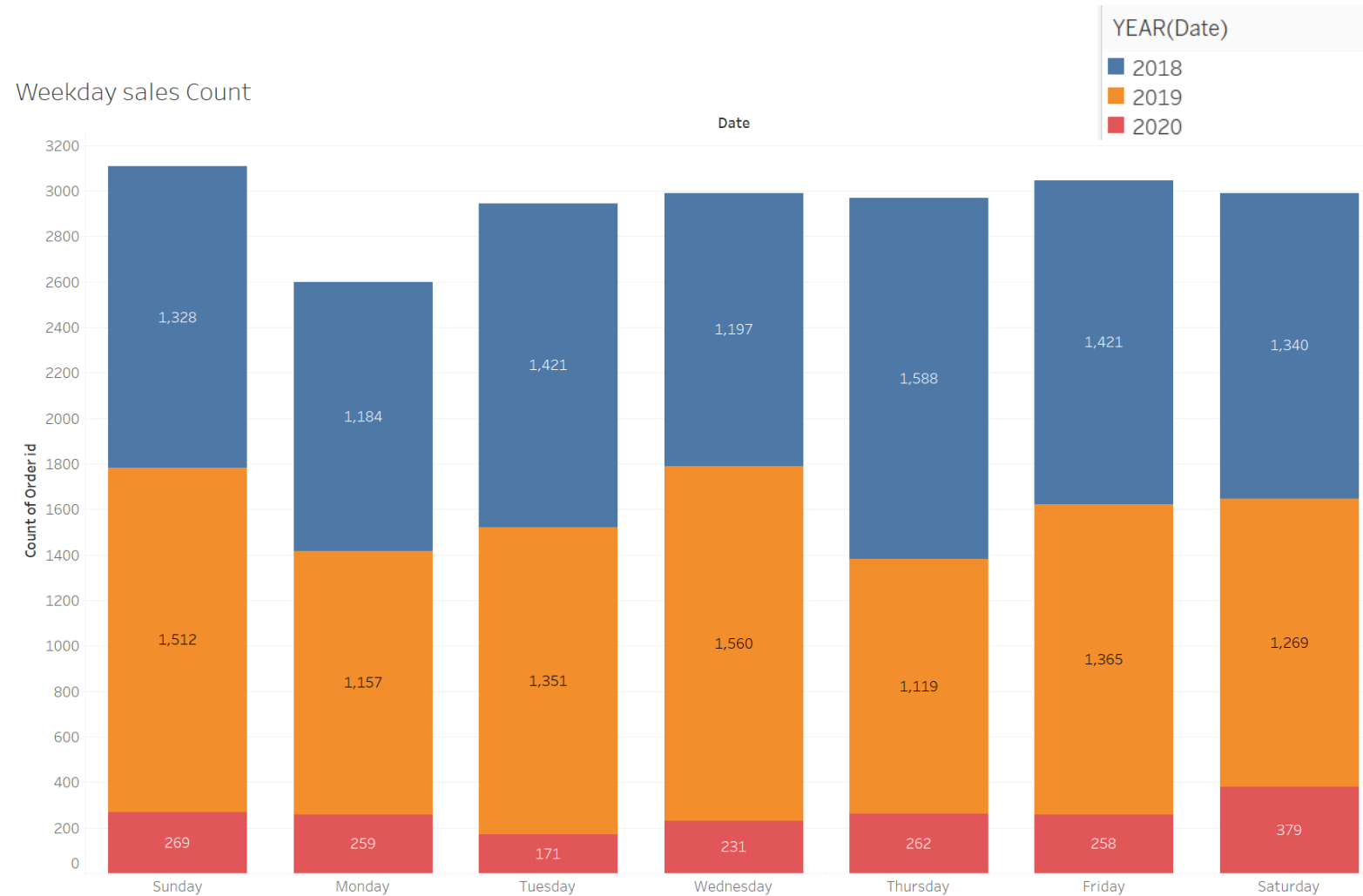
Monthly Count of Products Sold

- 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



Weekly Count of Products Sold

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



Count of products sold

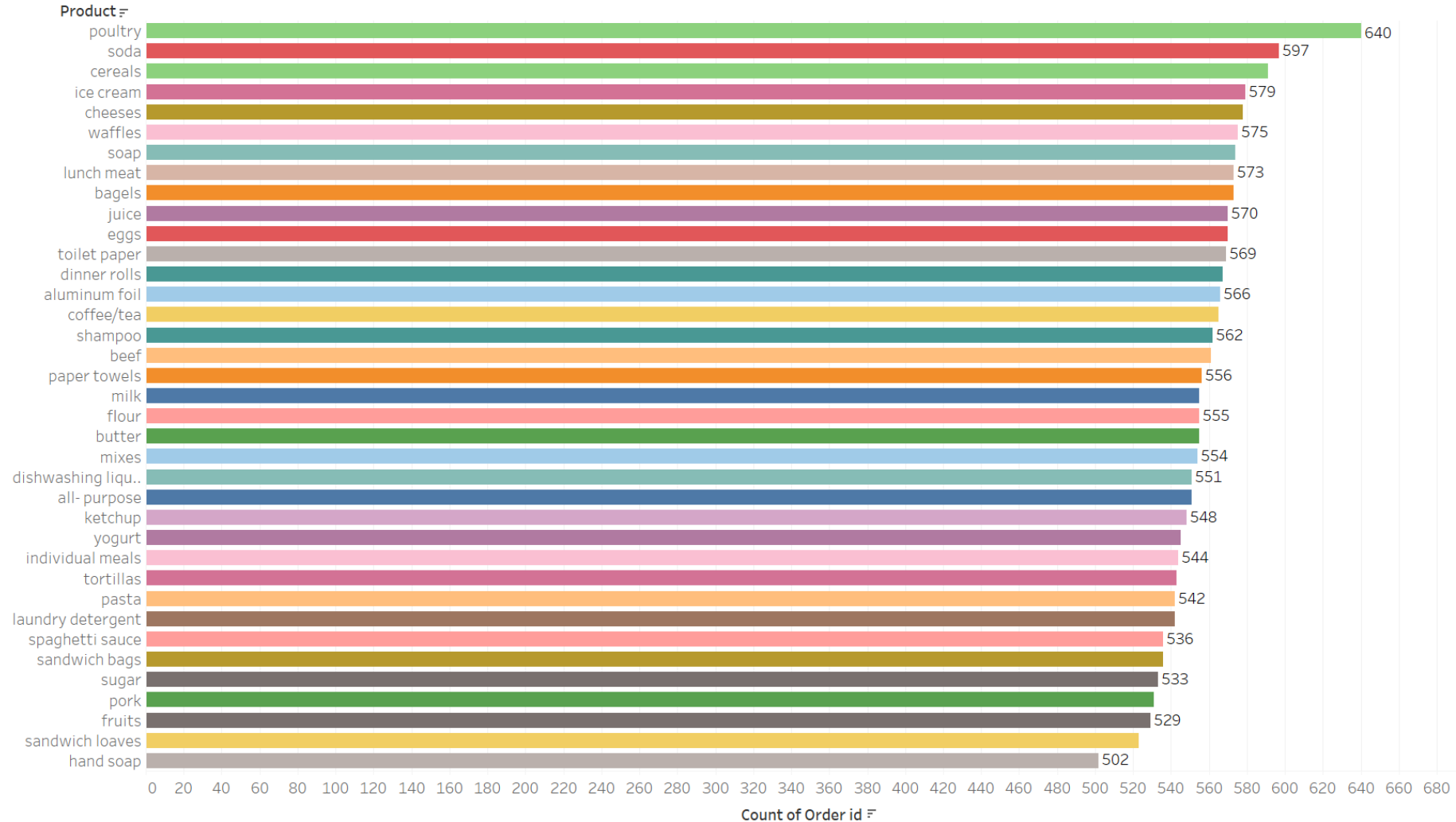
Product by Order Count(Box)

poultry 640	soap 574	dinner rolls 567	butter 555	flour 555	milk 555	mixes 554	all-purpose 551	
soda 597	bagels 573	aluminum foil 566	dishwashing liquid/detergent 551		laundry detergent 542	pasta 542	sandwich bags 536	
cereals 591	lunch meat 573	coffee/tea 565						
ice cream 579	eggs 570	shampoo 562	ketchup 548		spaghetti sauce 536	fruits 529	sandwich loaves 523	
cheeses 578	juice 570	beef 561	individual meals 544		sugar 533			
waffles 575	toilet paper 569	paper towels 556	tortillas 543		pork 531		hand soap 502	

- Top 5 product sold are Poultry (640), Soda (597), Cereals (591), Ice Cream (597), and Cheeses (578)

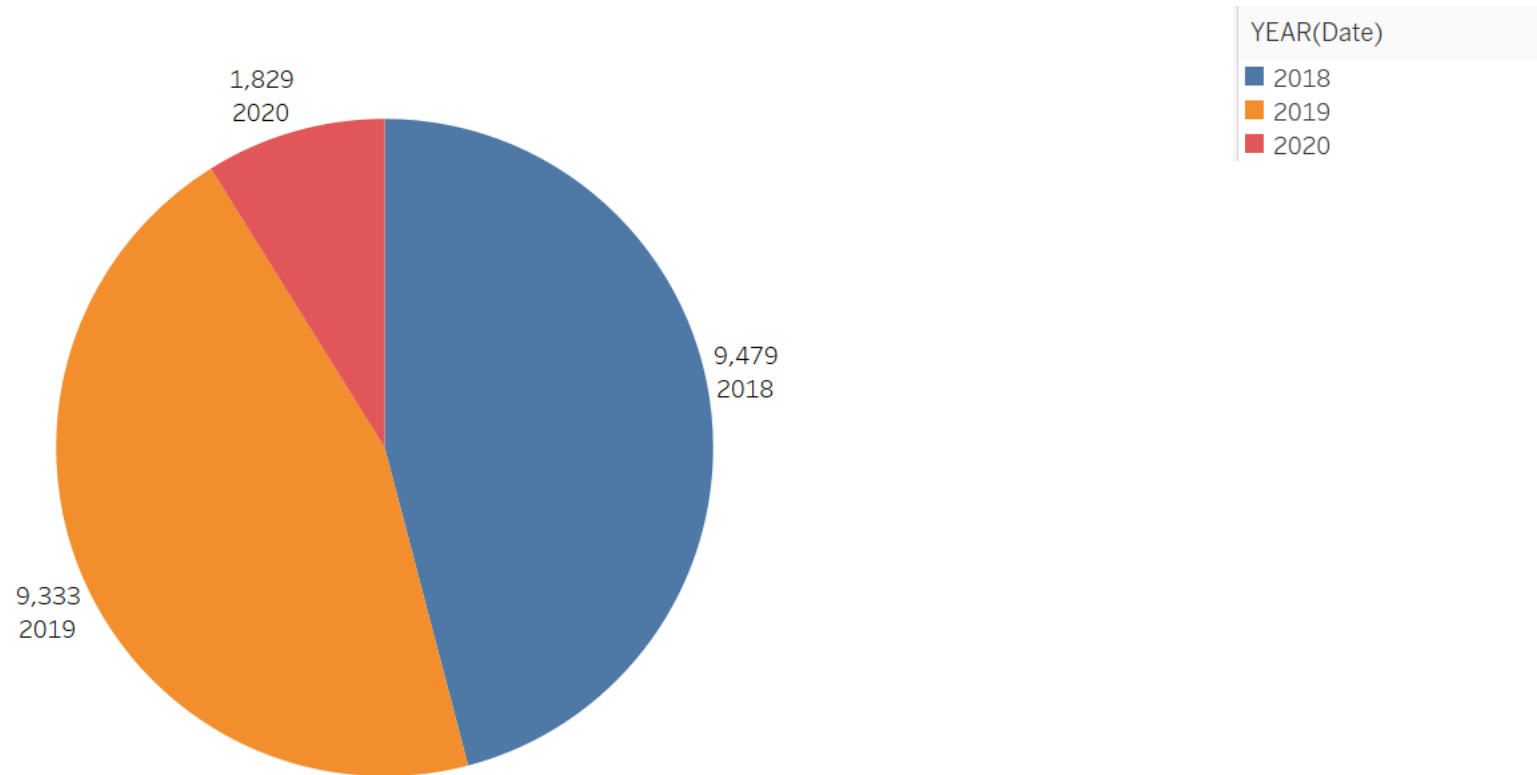
Count of products sold

Product by Order Count



- Top 5 product sold are Poultry (640), Soda (597), Cereals (591), Ice Cream (597), and Cheeses (578)

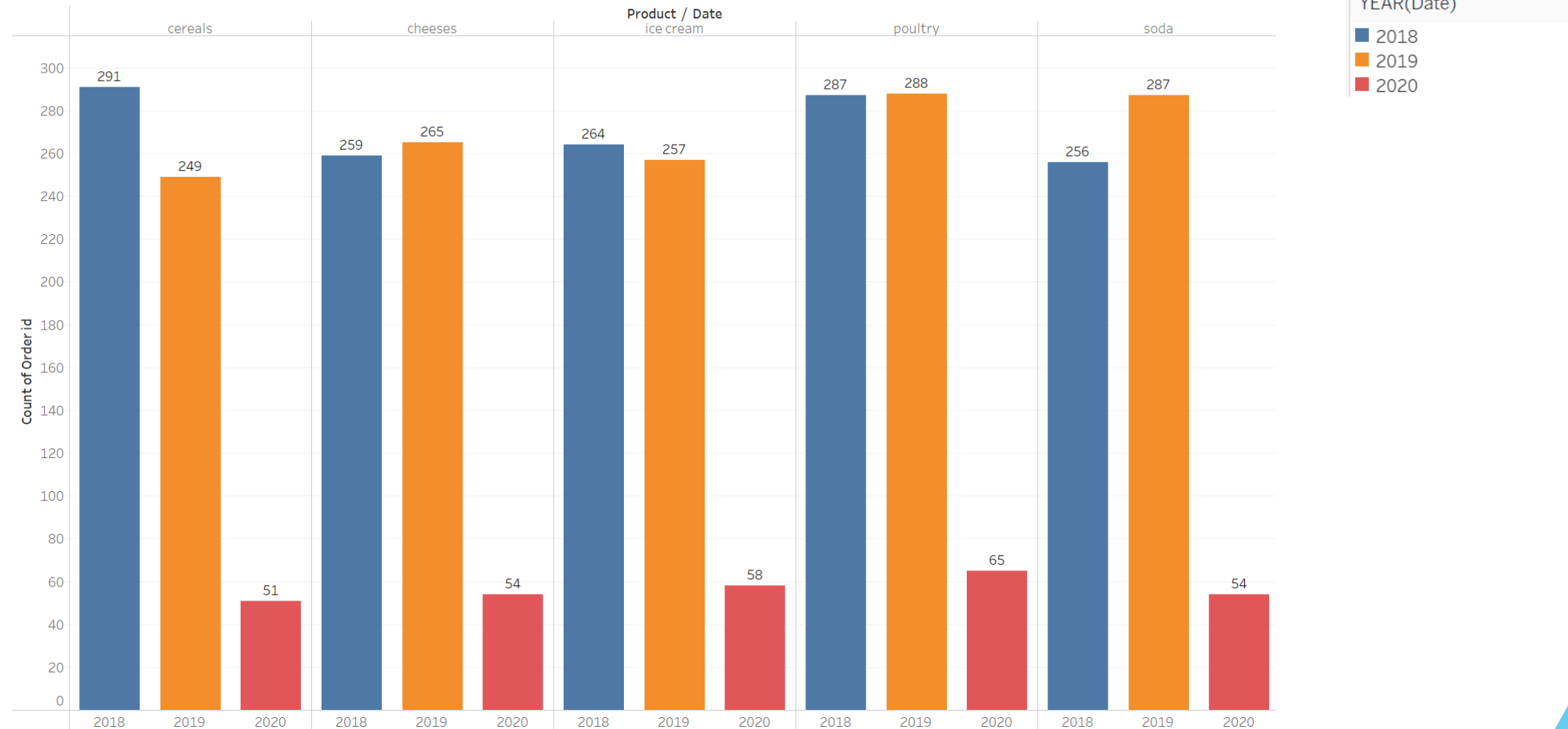
Count of products sold



- Most of the Products were sold in 2018 (9,479), followed by 2019 (9,333)

Top 5 products over the years

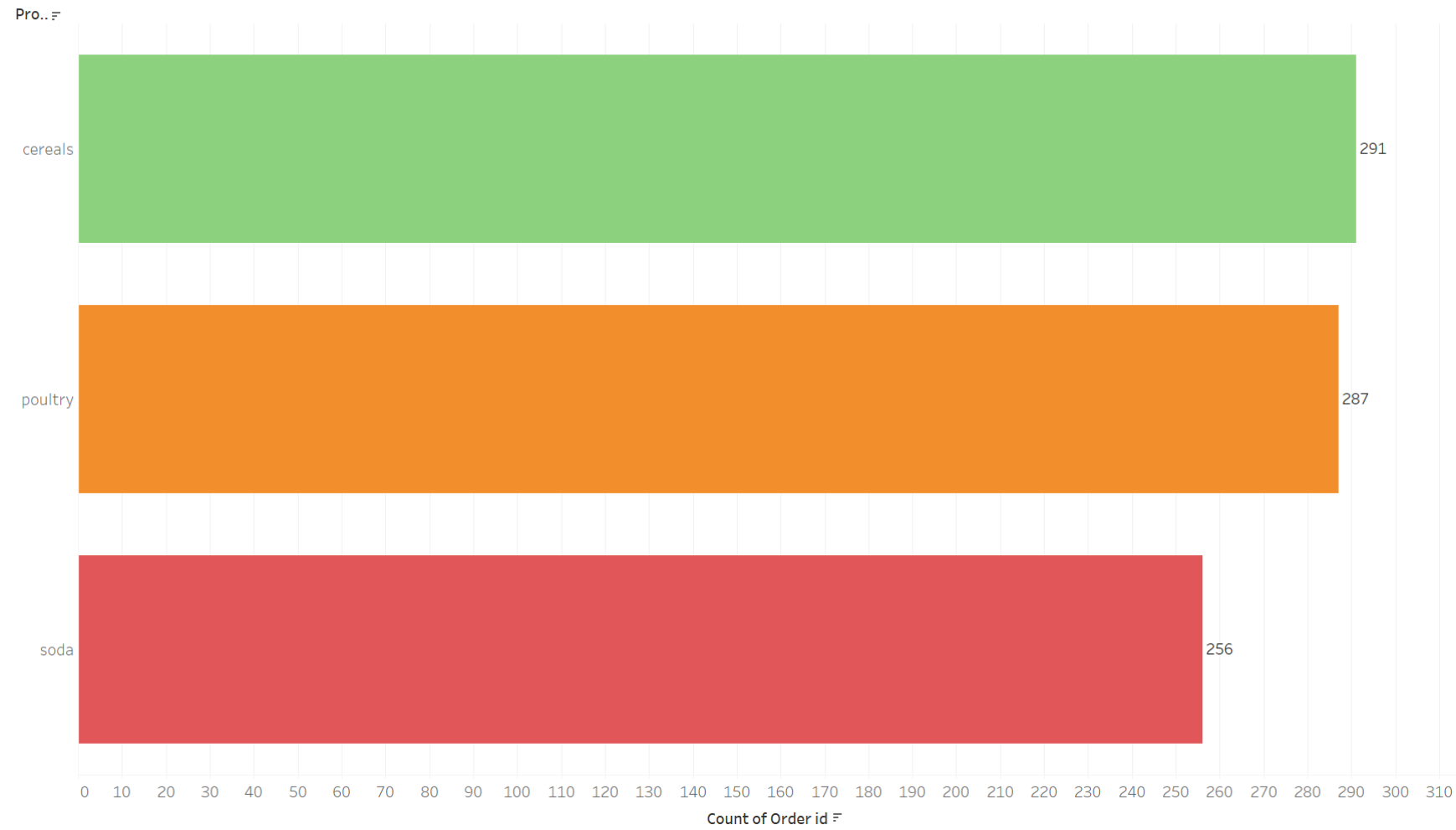
Top 5 Products of the year



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

Top 3 products for year 2018

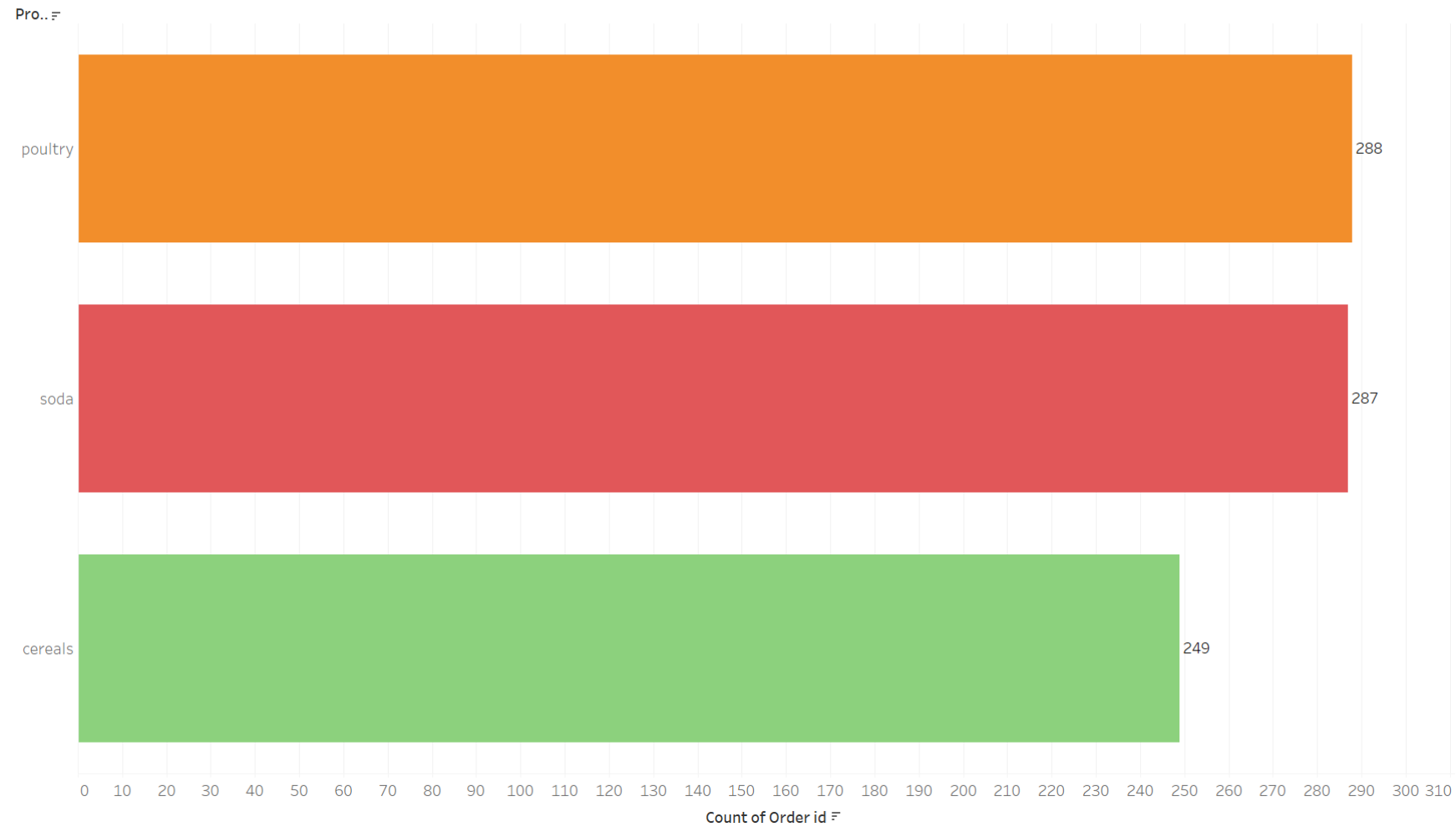
Top 3 products - 2018



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

Top 3 products for year 2019

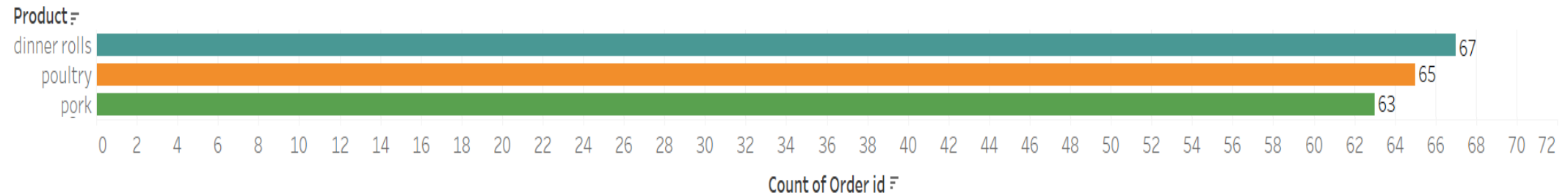
Top 3 products - 2019



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

Top 3 products for year 2020

Top 3 products - 2020



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

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
- 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

Recommendations

- Focus on promoting and stocking up on poultry, soda, and cereals as they are consistently top-selling products
- 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
- 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



Market Basket Analysis

Market Basket Analysis Meaning
MRA KNIME WorkFlow & Qutput Table

MBA – Association Rules

MBA – Association Rules

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
- 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

Like this with a minimum below threshold values rules for the given data are calculated:

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

KNIME Workflow



KNIME Workflow output



Output of KNIME workflow

- 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 product 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.

Table "default" - Rows: 24 Spec - Columns: 6 Properties Flow Variables						
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 f...
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,l...
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,ce...
rule10	0.051	0.604	1.589	milk	<---	[poultry,laundry detergent,cer...
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 cr...
rule13	0.052	0.686	1.628	poultry	<---	[dinner rolls,spaghetti sauce,ic...
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,jui...
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,laund...
rule19	0.054	0.656	1.556	poultry	<---	[dinner rolls,spaghetti sauce,la...
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 Rule Parameters

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

The screenshot shows a software dialog box titled "Dialog - 3:4 - Association Rule Learner". It has a "File" menu and four tabs: "Options", "Flow Variables", "Job Manager Selection", and "Memory Policy". The "Options" tab is active and contains three sections:

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

At the bottom of the dialog are four buttons: "OK", "Apply", "Cancel", and a help icon (a question mark inside a circle).

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 item set occurs in the dataset. High support indicates that the item set 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 item set 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.

Recommendations

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

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the right side of the frame, creating a modern, layered effect. The rest of the background is a solid, very light blue-grey color.

Thank you