



MRA PROJECT

MILESTONE 2

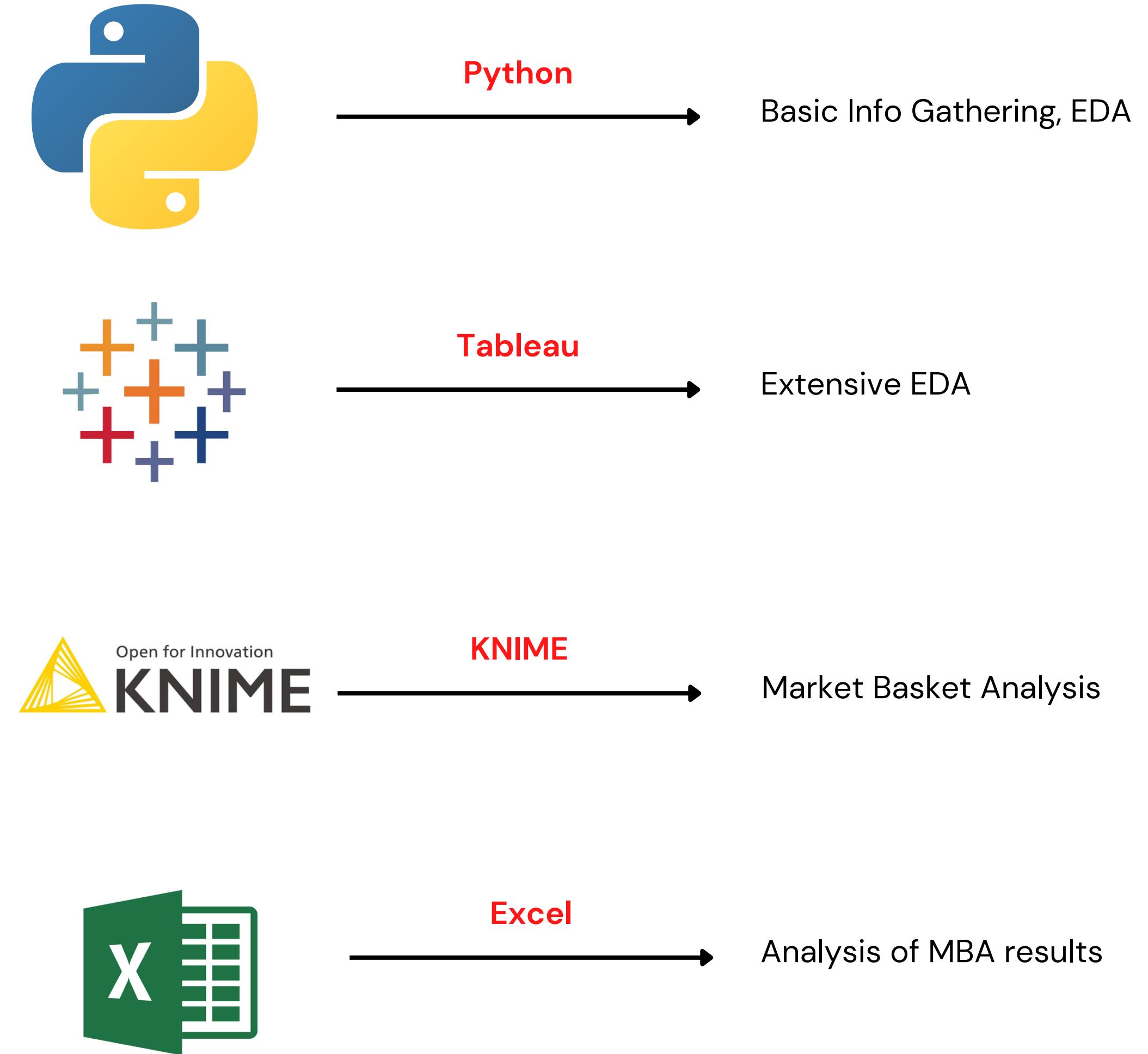


PGP DSBA Online – March' 22
Date : 08.01.23

BALAJI M P

Tools Used

The following are the tools used for this analysis



Agenda

- Problem Statement
- Getting to know the Data
- Exploratory Analysis & Inferences
- Market Basket Analysis
- Recommendations & Actionable Insights



[Back to Agenda Page](#)

Problem Statement



Given two years of transactional data from a Grocery Store that does not currently offer combo meals, the goal is to identify the most popular combinations of menu items that could be offered as combo meals. This will be done through analysis of the most commonly occurring sets of menu items in customer orders, and the resulting combo meals will be suggested to the store

[Back to Agenda Page](#)

Getting to know the Data

```
<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
```

No of missing values

0

Total no of records in dataset:

20641

Total no of features:

3

Data type of features:

Numeric

String

Date Time

1

1

1

Total Unique Orders:

1139

Total Unique Items Sold:

37

Note: Date column will be converted to DateTime for further analysis

Summary Statistics

All Features

	count	unique		top	freq	first	last	mean	std	min	25%	50%	75%	max
Date	20641	603	2019-02-08 00:00:00	183	2018-01-01	2020-02-26		NaN	NaN	NaN	NaN	NaN	NaN	NaN
Order_id	20641.0	NaN		NaN	NaN	NaT	NaT	575.986289	328.557078	1.0	292.0	581.0	862.0	1139.0
Product	20641	37		poultry	640	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN

- **Order_id** is an integer type, **Date** is a DateTime type, and **Product** is a string type
- We have access to every transactional record from **January 1, 2018, to February 26, 2020**
- The food store offers **37 distinctive items**
- The **most popular item** is poultry, with 640 units sold

Exploratory Analysis & Inferences

Total Orders

1,139

Single Order - Maximum Unique Items Bought

26

Single Order - Minimum Unique Items Bought

3

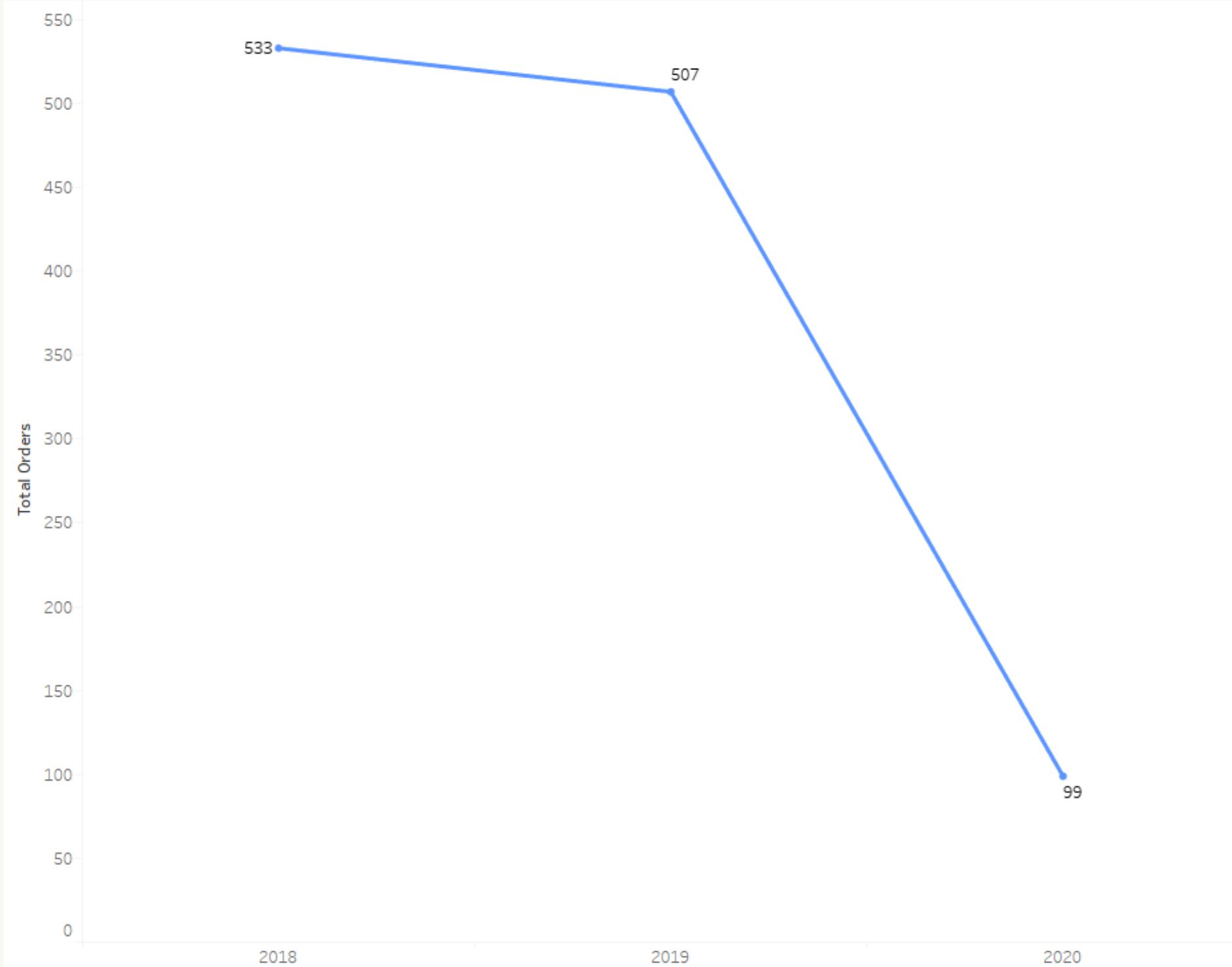
Items in Menu

37

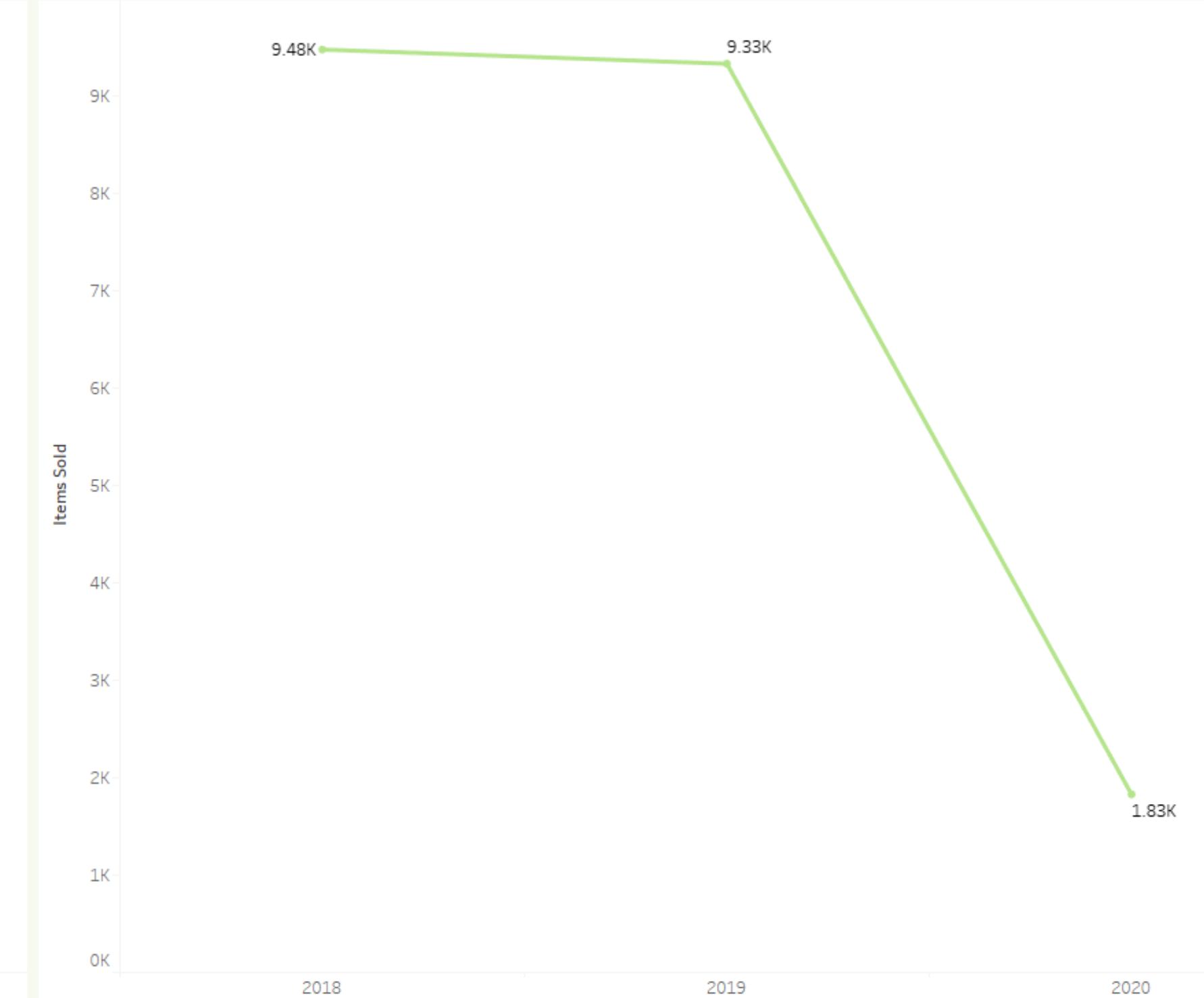
Average Unique Items Bought Together

14

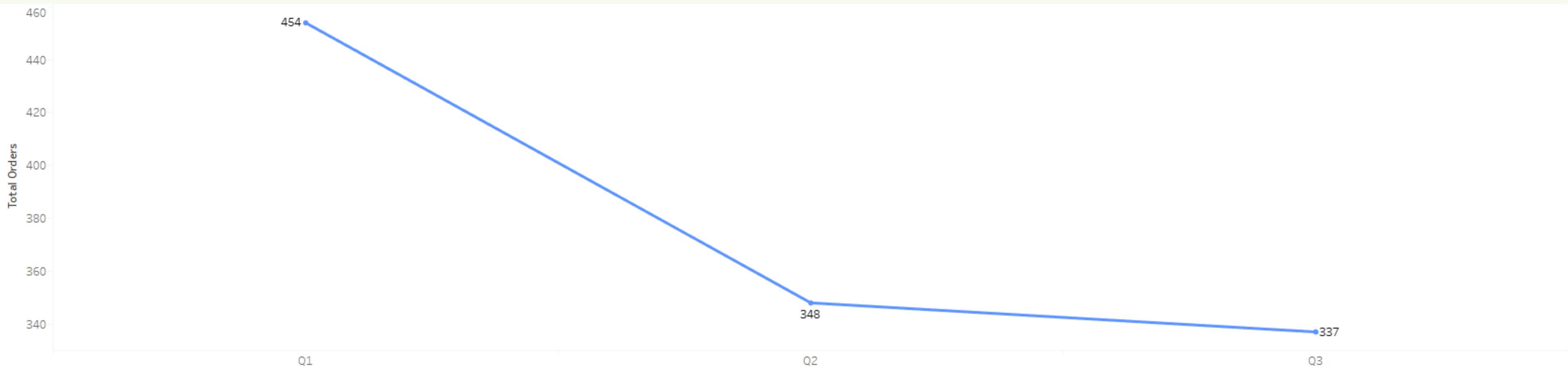
Yearly Orders



Yearly Items Sold



Quarterly Orders

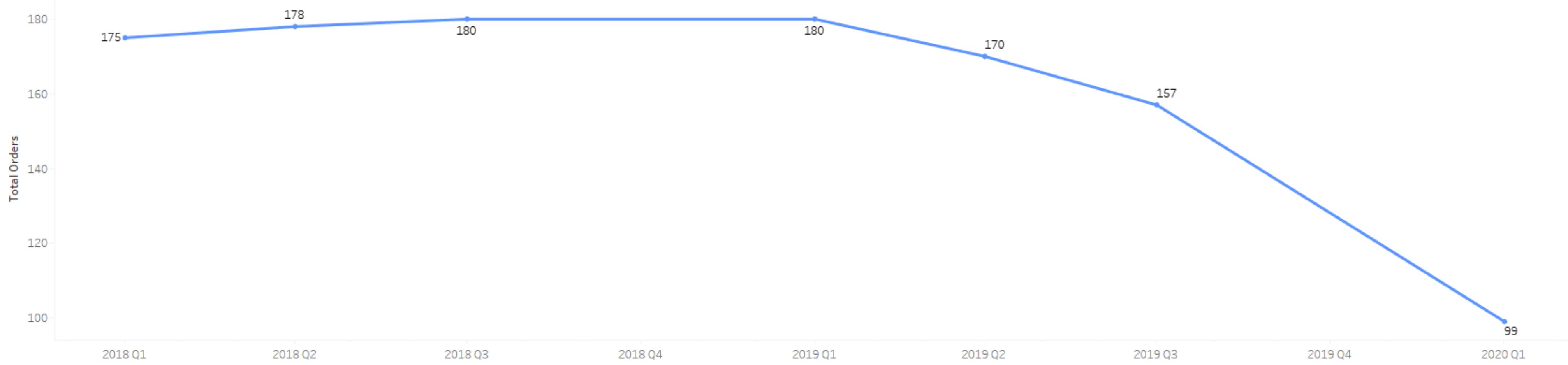


Quarterly Items Sold

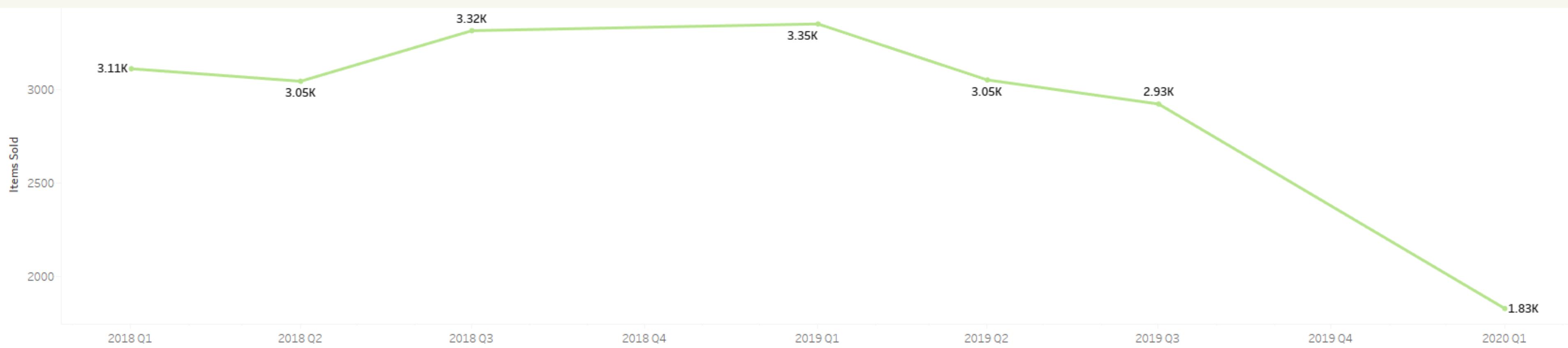


[Back to Agenda Page](#)

Quarterly Orders Across Years

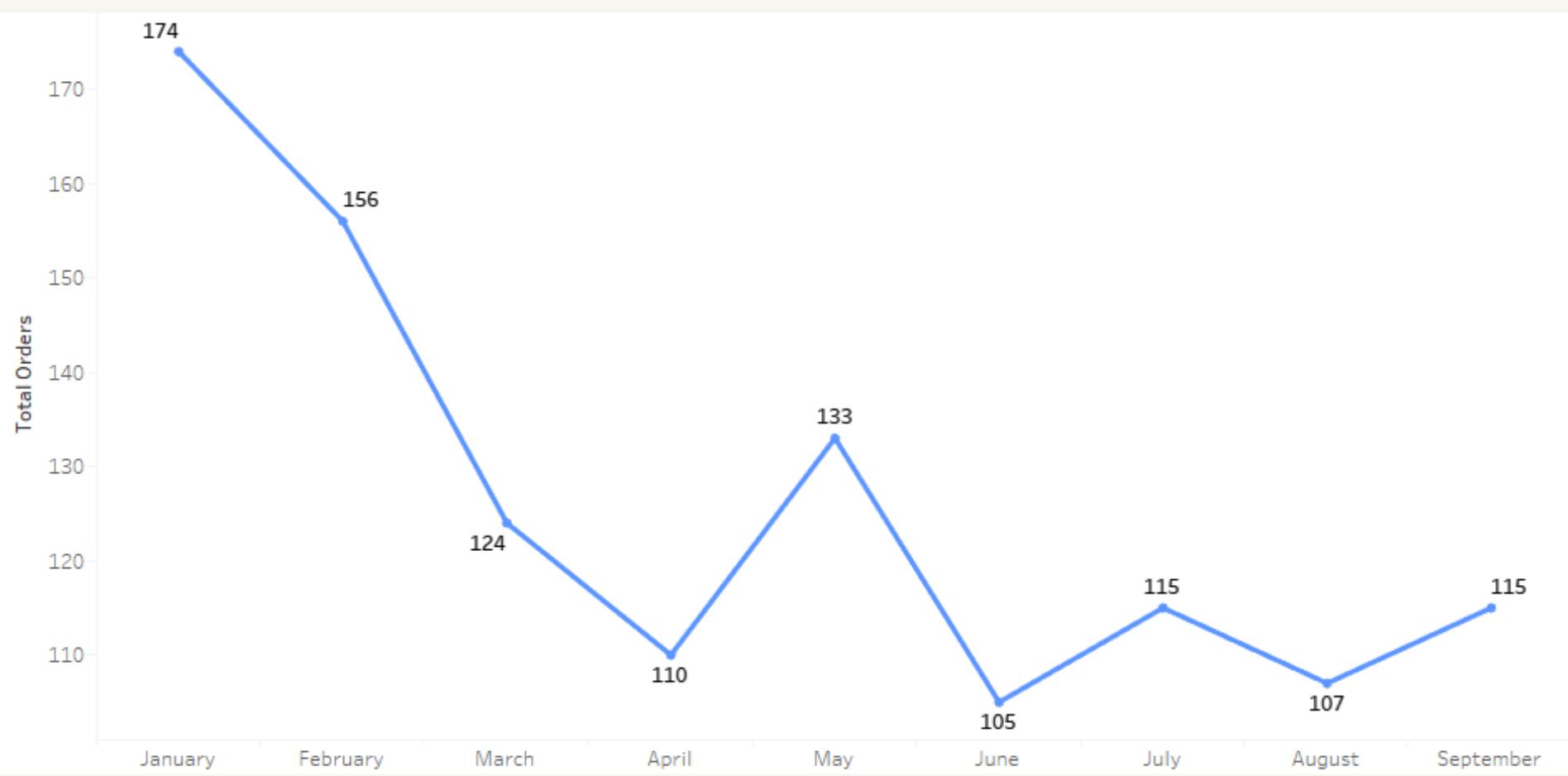


Items Sold Quarterly Across Years

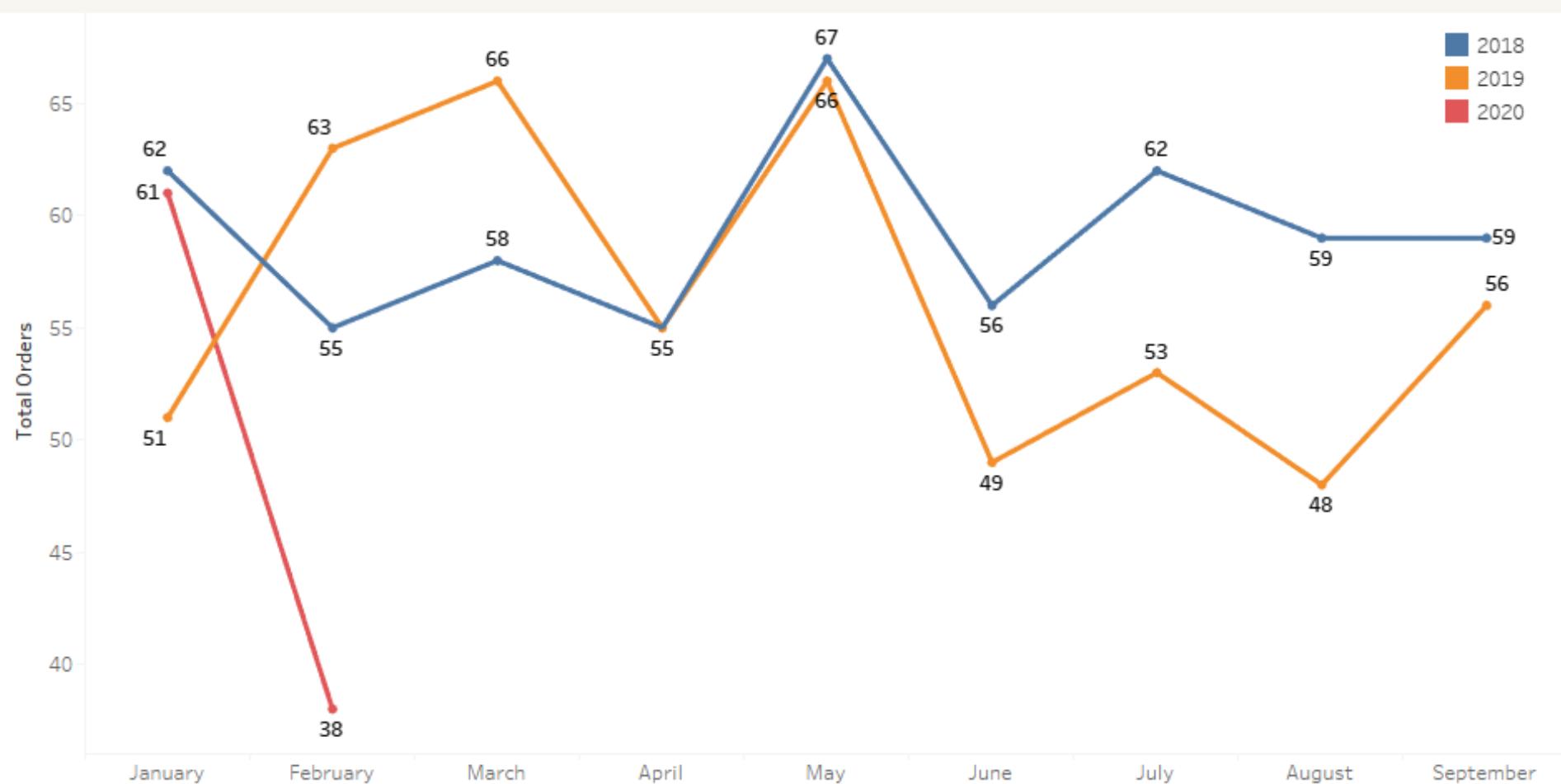


[Back to Agenda Page](#)

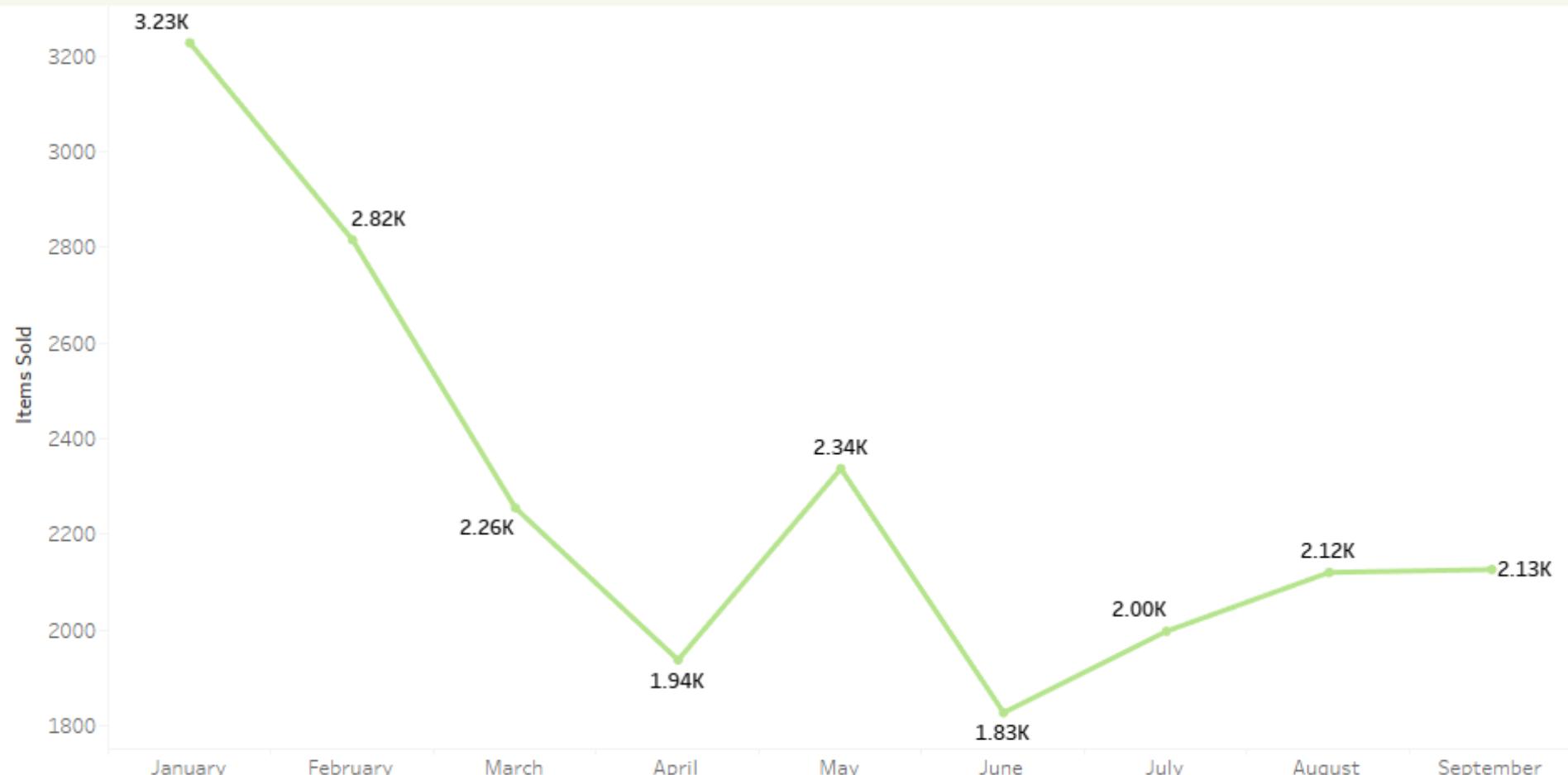
Monthly Orders



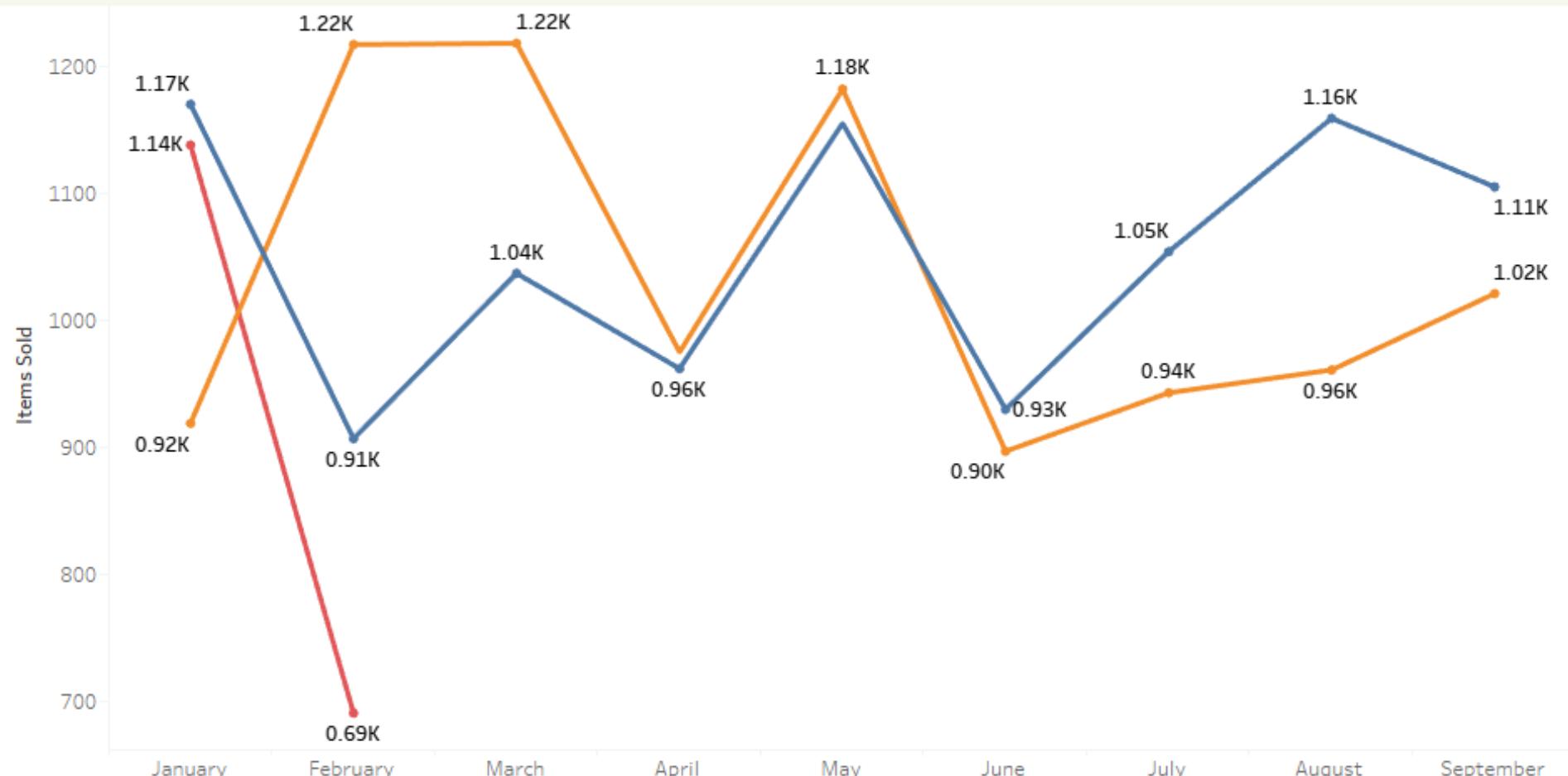
Monthly Orders - Year Wise



Items Sold Monthly



Items Sold Monthly - Year Wise



[Back to Agenda Page](#)

Weekly Orders

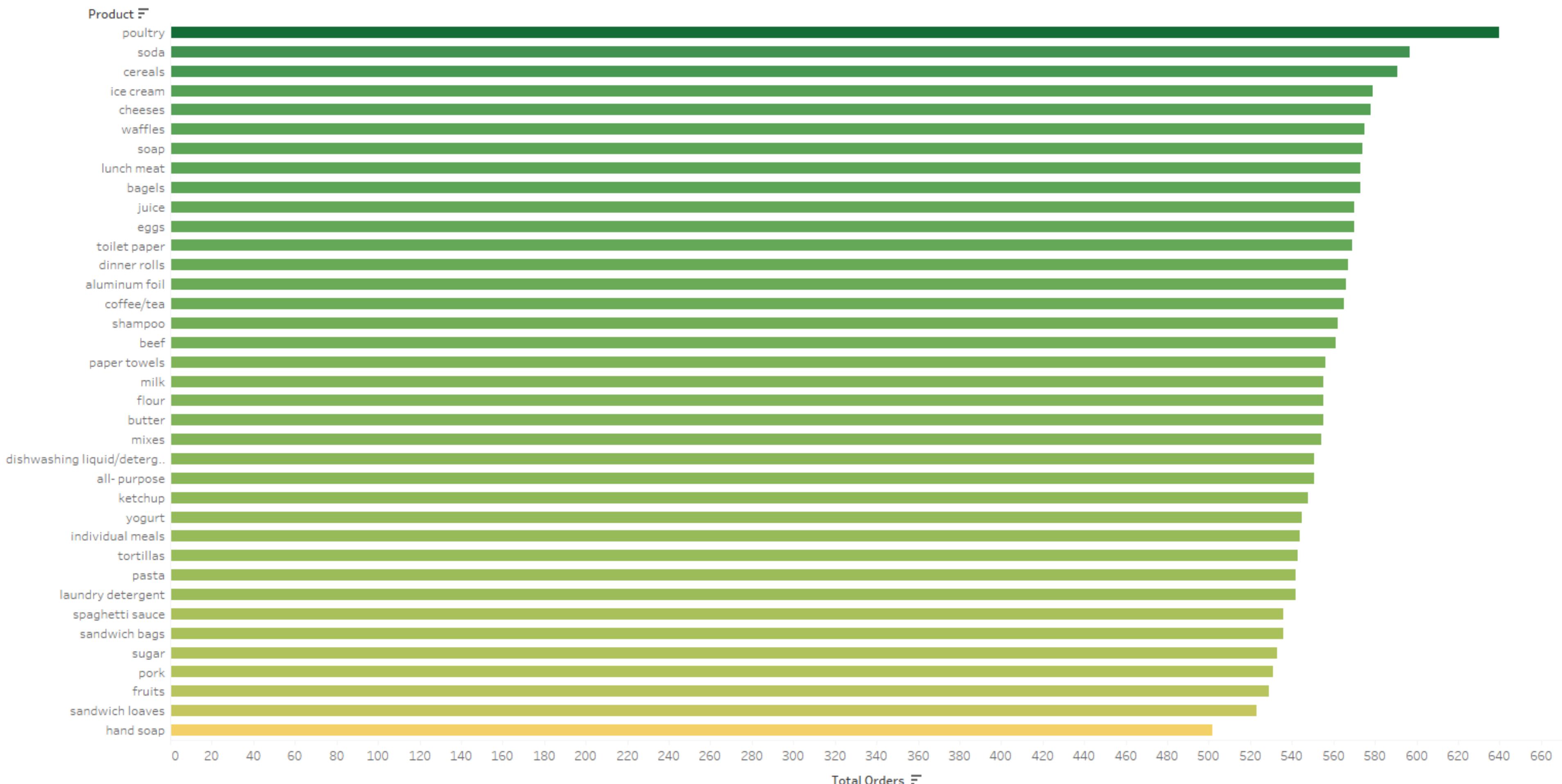


Items Sold Weekly



[Back to Agenda Page](#)

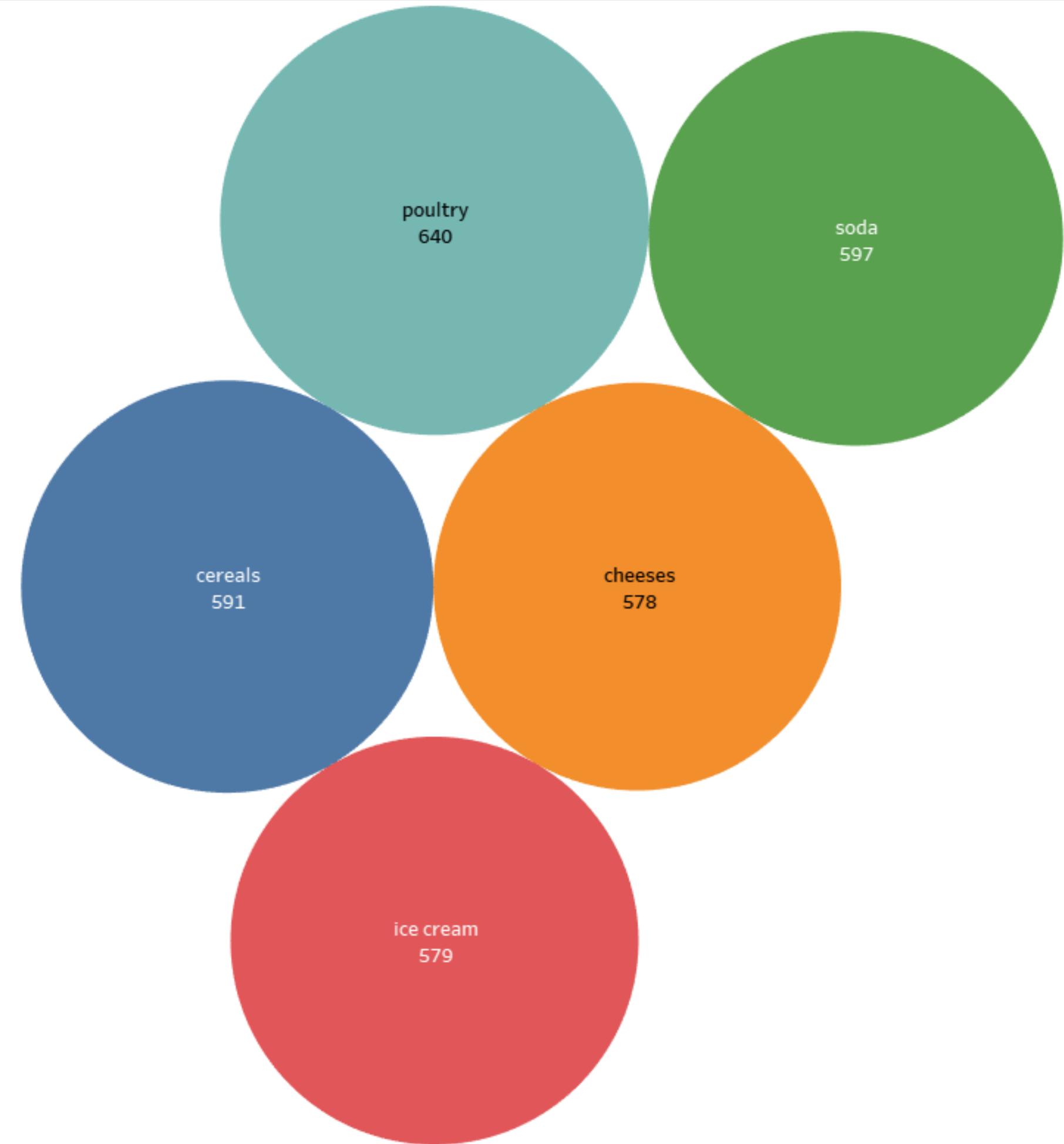
Overall Products Sold



Total Orders

[Back to Agenda Page](#)

Top Selling Items



[Back to Agenda Page](#)



EDA Inferences

- **1139 orders in total** were placed between January 1st, 2018 and February 26th, 2020.
- There are **presently 37 unique items** in the food store.
- The minimum product count is three and the maximum number of distinct products that can be purchased in a single order is 26. This implies that **anytime a customer visits the grocery shop, they purchase at least 3 different products**. An **average customer purchases 14 different products**.
- The **most orders and items were purchased in 2018**, then they somewhat decreased the following year and then **decreased by five times in 2020**. There appears to be a **worrying downward tendency in the overall trend**.
- The **quarterly orders** count and items sold follow the same pattern as the overall trend, but **data for Q4 is missing from the dataset**. It's possible that the store was closed during that quarter, but it would be **worth checking with the business to confirm the reason for the missing data**.



EDA Inferences

- Upon closer examination of the quarterly metrics over the years, we can see that **transactions were stable until Q1 of 2019**, after which they began to decline. There was a particularly **steep drop after Q3 of 2019**. On the other hand, the **number of products sold saw a steady increase from Q2 of 2018 until Q2 of 2019**, at which point it started to decrease gradually
- Analysis of the monthly metrics across all years reveals that the **highest number of transactions occurred in May**, while the **most products were sold in February and March**
- There appears to be a **seasonal pattern in the transaction count and items sold, with values reaching a peak every alternative month and dropping in the corresponding month**. The business may want to consider addressing the drop in sales





EDA Inferences

- A review of the weekly orders count and items sold reveals that **Monday is the weakest day in terms of sales**, while **Sunday is the best-performing** day. Sales tend to gradually increase as the week progresses, with the **weekends being the busiest time**
- An analysis of the products sold over the years shows that the **overall sales for all products remained relatively consistent**
- The **top 5 selling product categories**, in order of highest sales, were **Poultry, Soda, Cereals, Ice Cream, and Cheeses**
- In general, the **sales of the grocery store have been gradually declining**, and there is a **seasonal pattern in terms of sales**



MARKET BASKET ANALYSIS





What is it?

- Market basket analysis is a **technique used to identify which items are frequently purchased together by customers**. It can help retailers understand the types of products that tend to be bought together, so they can optimize their store layouts, product placements, and promotional offers

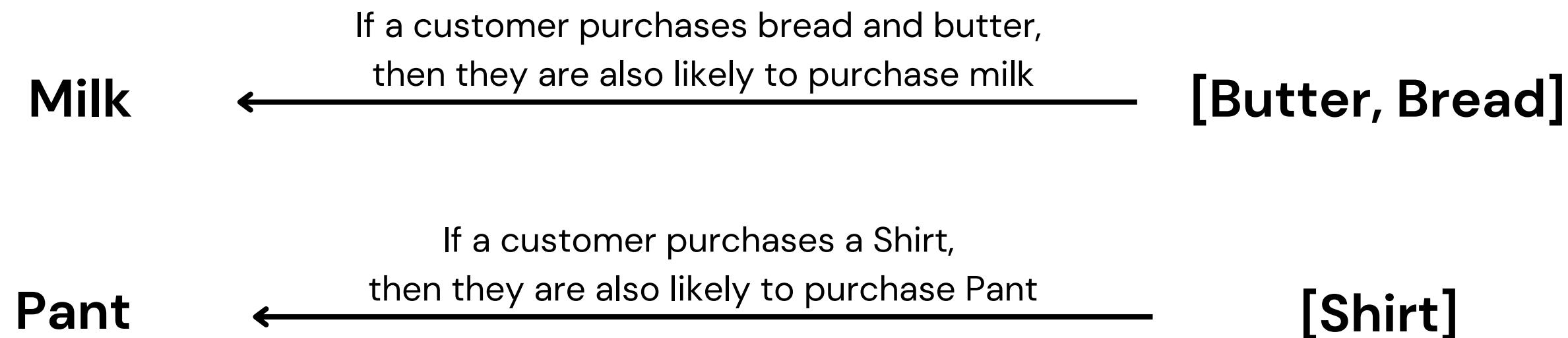
Association Rules

- Association rules in market basket analysis **identify relationships between items frequently purchased together.**
- These rules, written in "if-then" statements, help retailers optimize store layouts, product placements, and promotions to increase sales and customer satisfaction.
- **For example**, if an association rule reveals that customers who purchase apples also tend to purchase peanut butter, the retailer might place these items near each other and offer a discount on peanut butter to customers who purchase apples.
- Association rules are important because they allow retailers to better **understand their customers' purchasing habits** and **make informed decisions about how to effectively market and sell their products.**

Association Rules Example

- Association rules are **typically written in the form of "if-then" statements**, and they specify that if a certain item is purchased, then another item is also likely to be purchased

Example:

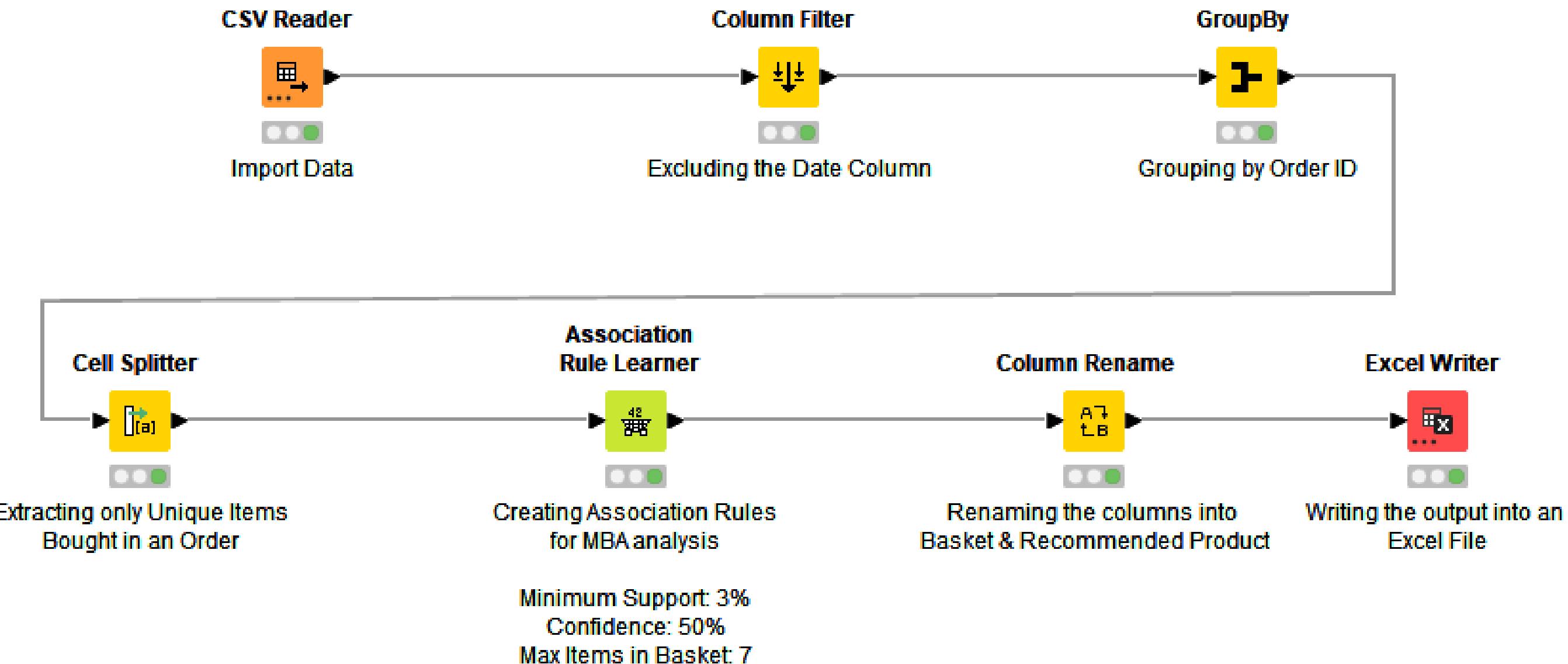


MBA - Key Metrics



- **Support** refers to the percentage of transactions in which a particular item appears. It measures how frequently an item is purchased.
- **Confidence:** This is the percentage of transactions in which a particular item appears, given that another item is also present in the transaction. It measures the strength of the relationship between two items.
- **Lift:** This is the ratio of the observed support for an item pair to the expected support, if the items were independent. It measures the strength of the relationship between two items, taking into account the overall popularity of the items.

MARKET BASKET ANALYSIS USING KNIME



Setting the Thresholds

Minimum Support: 3%

Confidence: 50%

Max Items in Basket: 7

- In order to identify the most strongly associated item pairs, we set the threshold for minimum support to 3% and the point for confidence to 50%
- This means we only considered item pairs that appeared in at least 3% of transactions (based on the support threshold) and had a minimum confidence value of at least 50% (based on the confidence threshold)
- This allowed us to focus on item pairs that were relatively frequent and had a strong relationship, and to make informed decisions

Support, Confidence & Lift

Table "default" - Rows: 50743		Spec - Columns: 6		Properties	Flow Variables			
Columns: 6	Column Type	Column Index	Color Handler	Size Handler	Shape Han...	Filter Handler	Lower Bound	Upper Bound
Support	Number (double)	0					0.031	0.195
Confidence	Number (double)	1					0.5	0.795
Lift	Number (double)	2					1.186	2.194

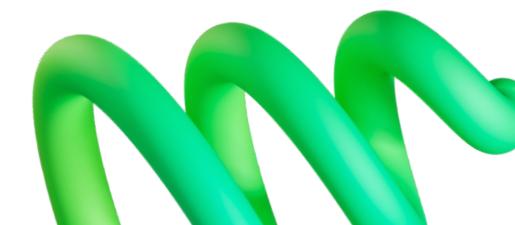
- We were able to obtain **50743 associations** using the specified thresholds.
- **Support** has a bottom and upper bounds of 0.031 and 0.195, respectively.
- The minimum and upper bounds for **Confidence** are 0.5 and 0.795, respectively.
- The **Lift** has a bottom and upper bounds of 1.186 and 2.194, respectively.



OUTPUT HEAD OF KNIME WORKFLOW

Row ID	D Support	D Confidence	D Lift	S Recommended Product	S implies	S Basket
rule2790	0.031	0.795	2.194	paper towels	<---	[eggs, ice cream, pasta, lunch meat]
rule6038	0.032	0.783	2.158	paper towels	<---	[eggs, ice cream, pasta, cereals]
rule2791	0.031	0.729	2.066	flour	<---	[dishwashing liquid/detergent, cheeses, waffles, soda]
rule9828	0.032	0.74	2.041	paper towels	<---	[eggs, dinner rolls, ice cream, pasta]
rule6043	0.032	0.72	1.986	paper towels	<---	[eggs, poultry, ice cream, pasta]
rule2787	0.031	0.778	1.951	ice cream	<---	[paper towels, eggs, pasta, lunch meat]
rule2792	0.031	0.761	1.947	soda	<---	[dishwashing liquid/detergent, cheeses, flour, waffles]
rule14095	0.033	0.717	1.931	pasta	<---	[paper towels, dishwashing liquid/detergent, eggs, ice cream]
rule42417	0.04	0.697	1.922	paper towels	<---	[all-purpose, individual meals, toilet paper]
rule2803	0.031	0.714	1.914	spaghetti sauce	<---	[dinner rolls, poultry, laundry detergent, juice]
rule14094	0.033	0.745	1.911	eggs	<---	[paper towels, dishwashing liquid/detergent, ice cream, pasta]
rule14096	0.033	0.691	1.905	paper towels	<---	[dishwashing liquid/detergent, eggs, ice cream, pasta]
rule6042	0.032	0.706	1.901	pasta	<---	[paper towels, eggs, poultry, ice cream]
rule6041	0.032	0.72	1.847	eggs	<---	[paper towels, poultry, ice cream, pasta]
rule9827	0.032	0.685	1.845	pasta	<---	[paper towels, eggs, dinner rolls, ice cream]
rule24706	0.036	0.641	1.833	sandwich loaves	<---	[all-purpose, flour, individual meals]
rule9826	0.032	0.712	1.825	eggs	<---	[paper towels, dinner rolls, ice cream, pasta]
rule42521	0.04	0.676	1.822	pasta	<---	[hand soap, soda, aluminum foil]
rule42709	0.04	0.676	1.822	ketchup	<---	[butter, aluminum foil, soap]
rule33318	0.038	0.632	1.81	sandwich loaves	<---	[paper towels, flour, individual meals]
rule43941	0.041	0.671	1.808	ketchup	<---	[pork, sandwich bags, soap]
rule14492	0.034	0.629	1.8	sandwich loaves	<---	[yogurt, hand soap, soap]
rule2799	0.031	0.7	1.796	eggs	<---	[dishwashing liquid/detergent, ice cream, pasta, soda]
rule48820	0.046	0.65	1.793	paper towels	<---	[ice cream, pasta, lunch meat]
rule49555	0.055	0.649	1.791	paper towels	<---	[eggs, ice cream, pasta]
rule32563	0.038	0.662	1.786	fruits	<---	[all-purpose, beef, lunch meat]
rule38886	0.039	0.688	1.784	bagels	<---	[sandwich loaves, fruits, juice]

[Back to Agenda Page](#)





RESULTING ASSOCIATIONS (TOP 20)

SORTED BY DESCENDING ORDER OF LIFT & CONFIDENCE

RowID	Support	Confidence	Lift	Recommended Product	implies	Basket
rule2790	0.030728709	0.795454545	2.19375963	paper towels	<---	[eggs, ice cream, pasta, lunch meat]
rule6038	0.031606673	0.782608696	2.158332456	paper towels	<---	[eggs, ice cream, pasta, cereals]
rule2791	0.030728709	0.729166667	2.065972222	flour	<---	[dishwashing liquid/detergent, cheeses, waffles, soda]
rule9828	0.032484636	0.74	2.040823245	paper towels	<---	[eggs, dinner rolls, ice cream, pasta]
rule6043	0.031606673	0.72	1.98566586	paper towels	<---	[eggs, poultry, ice cream, pasta]
rule2787	0.030728709	0.777777778	1.951297112	ice cream	<---	[paper towels, eggs, pasta, lunch meat]
rule2792	0.030728709	0.760869565	1.947484123	soda	<---	[dishwashing liquid/detergent, cheeses, flour, waffles]
rule14095	0.033362599	0.716981132	1.930594585	pasta	<---	[paper towels, dishwashing liquid/detergent, eggs, ice cream]
rule42417	0.040386304	0.696969697	1.922151295	paper towels	<---	[all- purpose, individual meals, toilet paper]
rule2803	0.030728709	0.714285714	1.914285714	spaghetti sauce	<---	[dinner rolls, poultry, laundry detergent, juice]
rule14094	0.033362599	0.745098039	1.911411411	eggs	<---	[paper towels, dishwashing liquid/detergent, ice cream, pasta]
rule14096	0.033362599	0.690909091	1.905436936	paper towels	<---	[dishwashing liquid/detergent, eggs, ice cream, pasta]
rule6042	0.031606673	0.705882353	1.90070922	pasta	<---	[paper towels, eggs, poultry, ice cream]
rule6041	0.031606673	0.72	1.847027027	eggs	<---	[paper towels, poultry, ice cream, pasta]
rule9827	0.032484636	0.685185185	1.844978548	pasta	<---	[paper towels, eggs, dinner rolls, ice cream]
rule24706	0.035996488	0.640625	1.83334642	sandwich loaves	<---	[all- purpose, flour, individual meals]
rule9826	0.032484636	0.711538462	1.825320513	eggs	<---	[paper towels, dinner rolls, ice cream, pasta]
rule42521	0.040386304	0.676470588	1.821513002	pasta	<---	[hand soap, soda, aluminum foil]
rule42709	0.040386304	0.676470588	1.821513002	ketchup	<---	[butter, aluminum foil, soap]
rule33318	0.037752414	0.632352941	1.809673367	sandwich loaves	<---	[paper towels, flour, individual meals]

[Back to Agenda Page](#)



Inferences

The **RowID column** in the output of the Association Rule Learner step in KNIME will **contain a list of association rules** that have been identified based on the input data and the specified parameters. These rules can be used to understand the relationships between different items and make informed decisions

Let us look at some of the rules obtained

Rule 2790

RowID	Support	Confidence	Lift	Recommended Product	implies	Basket
rule2790	0.030728709	0.795454545	2.19375963	paper towels	<-->	[eggs, ice cream, pasta, lunch meat]

Rule 2790

paper towels

[eggs, ice cream, pasta, lunch meat]

If a customer purchases these items [eggs, ice cream, pasta, lunch meat], then they are also likely to purchase paper towels if recommended to them

- The **support value of 0.031** indicates that the item set "eggs, ice cream, pasta, and lunch meat" appears in 3.1% of transactions.
- The **confidence value of 0.795** indicates that the item "paper towels" chance of buying is 79.5% if the item set containing "eggs, ice cream, pasta, and lunch meat" is already bought.
- The **lift value of 2.194** indicates that the purchase of the item set "eggs, ice cream, pasta, and lunch meat" increases the chances of the customer buying paper towels by 2.194 times.
- In other words, If "eggs, ice cream, pasta, and lunch meat" are purchased, paper towels are 2.19 times more likely to be purchased with an 80% confidence.

Rule 33318

RowID	Support	Confidence	Lift	Recommended Product	impli	Basket
rule33318	0.037752414	0.632352941	1.809673367	sandwich loaves	<-->	[paper towels, flour, individual meals]

sandwich loaves ←———— [paper towels, flour, individual meals]

If a customer purchases these items in set
then they are also likely to purchase sandwich loaves
if recommended to them

- The **support value of 0.037** indicates that the item set "paper towels, flour, individual meals" appears in 3.7% of transactions.
- The **confidence value of 0.632** indicates that the item "sandwich loaves" chance of buying is 63.2% if the item set containing "paper towels, flour, individual meals" is already bought.
- The **lift value of 1.809** indicates that the purchase of the item set "paper towels, flour, individual meals" increases the chances of the customer buying sandwich loaves by 1.809 times.
- In other words, If "paper towels, flour, individual meals" are purchased, sandwich loaves are 1.8 times more likely to be purchased with a 63.2% confidence.

Rule 2787

RowID	Support	Confidence	Lift	Recommended Product	implies	Basket
rule2787	0.030728709	0.77777778	1.951297112	ice cream	<-->	[paper towels, eggs, pasta, lunch meat]

ice cream

[paper towels, eggs, pasta, lunch meat]

If a customer purchases these items in set
then they are also likely to purchase ice cream if
recommended to them

- The **support value of 0.03** indicates that the item set "paper towels, eggs, pasta, lunch meat" appears in 3.7% of transactions.
- The **confidence value of 0.777** indicates that the item "ice cream" chance of buying is 77.7% if the item set containing "paper towels, eggs, pasta, lunch meat" is already bought.
- The **lift value of 1.95** indicates that the purchase of the item set "paper towels, eggs, pasta, lunch meat" increases the chances of the customer buying ice cream by 1.95 times.
- In other words, If "paper towels, eggs, pasta, lunch meat" are purchased, ice cream is 1.95 times more likely to be purchased with a 77.7% confidence.

RECOMMENDATIONS & ACTIONABLE STRATEGIES

TOP 10 RECOMMENDED PRODUCTS

BASED ON ASSOCIATION RULES

Recommended Product	F
poultry	3,891
cheeses	2,332
soda	2,231
lunch meat	2,219
yogurt	2,134
dinner rolls	2,006
eggs	1,986
cereals	1,966
waffles	1,943
ice cream	1,851

These products can divided into the standard supermarket categories:

Meat and Poultry: Poultry, lunch meat

Dairy: Cheeses, yogurt

Bakery: Dinner rolls, waffles

Frozen foods: Ice cream

Beverages: Soda

Grocery: Cereals, eggs

To help the customers easily find these products, they can be placed in the spotlight of their respective sections in the store.

TOP 20 RECOMMENDATIONS

RowID	Support	Confidence	Lift	Recommended Product	impli	Basket
rule2790	0.030728709	0.795454545	2.19375963	paper towels	<---	[eggs, ice cream, pasta, lunch meat]
rule6038	0.031606673	0.782608696	2.158332456	paper towels	<---	[eggs, ice cream, pasta, cereals]
rule2791	0.030728709	0.729166667	2.065972222	flour	<---	[dishwashing liquid/detergent, cheeses, waffles, soda]
rule9828	0.032484636	0.74	2.040823245	paper towels	<---	[eggs, dinner rolls, ice cream, pasta]
rule6043	0.031606673	0.72	1.98566586	paper towels	<---	[eggs, poultry, ice cream, pasta]
rule2787	0.030728709	0.777777778	1.951297112	ice cream	<---	[paper towels, eggs, pasta, lunch meat]
rule2792	0.030728709	0.760869565	1.947484123	soda	<---	[dishwashing liquid/detergent, cheeses, flour, waffles]
rule14095	0.033362599	0.716981132	1.930594585	pasta	<---	[paper towels, dishwashing liquid/detergent, eggs, ice cream]
rule42417	0.040386304	0.696969697	1.922151295	paper towels	<---	[all- purpose, individual meals, toilet paper]
rule2803	0.030728709	0.714285714	1.914285714	spaghetti sauce	<---	[dinner rolls, poultry, laundry detergent, juice]
rule14094	0.033362599	0.745098039	1.911411411	eggs	<---	[paper towels, dishwashing liquid/detergent, ice cream, pasta]
rule14096	0.033362599	0.690909091	1.905436936	paper towels	<---	[dishwashing liquid/detergent, eggs, ice cream, pasta]
rule6042	0.031606673	0.705882353	1.90070922	pasta	<---	[paper towels, eggs, poultry, ice cream]
rule6041	0.031606673	0.72	1.847027027	eggs	<---	[paper towels, poultry, ice cream, pasta]
rule9827	0.032484636	0.685185185	1.844978548	pasta	<---	[paper towels, eggs, dinner rolls, ice cream]
rule24706	0.035996488	0.640625	1.83334642	sandwich loaves	<---	[all- purpose, flour, individual meals]
rule9826	0.032484636	0.711538462	1.825320513	eggs	<---	[paper towels, dinner rolls, ice cream, pasta]
rule42521	0.040386304	0.676470588	1.821513002	pasta	<---	[hand soap, soda, aluminum foil]
rule42709	0.040386304	0.676470588	1.821513002	ketchup	<---	[butter, aluminum foil, soap]
rule33318	0.037752414	0.632352941	1.809673367	sandwich loaves	<---	[paper towels, flour, individual meals]

RECOMMENDATIONS

- The **top recommended products** include **paper towels, flour, ice cream, soda, pasta, spaghetti sauce, eggs, sandwich loaves, and ketchup**
- We can see that **customers who purchase eggs, ice cream, and pasta are more likely to also purchase paper towels**. The store could **place a display of paper towels near the eggs, ice cream, and pasta** in the store and use signage or other promotional materials to **suggest that customers purchase all four items together**
- **Poultry is the most popular product among customers.** The store can take advantage of this demand by **promoting cross-selling offers. Offering discounted prices on poultry products when bought separately could attract more customers**, both existing and new. The store could use this strategy to also promote cross-selling of other products

LUCRATIVE COMBOS

As the grocery store currently does not offer any combo meals, it may be worth considering implementing additional promotion and marketing strategies using the following potential product combinations

COMBO 1

- Eggs, Pasta & Ice Cream

COMBO 2

- Waffles & Cheeses

COMBO 3

- Poultry, Ice Cream & Eggs

COMBO 4

- Individual Meals & Sandwich Loaves

Note: The above combos should be regularly updated with current market trends

BUCKET STRATEGY

Categorize the items "eggs, ice cream, pasta, spaghetti, lunch meat, poultry, dinner rolls" into 2 buckets:

Bucket 1: Eggs, ice cream, pasta, spaghetti

Bucket 2: Lunch meat, poultry, dinner rolls

BUY 1 GET 1

- Customers can select any one product from one group and choose a free product from the same group

BUY 2 GET 1

- Customers can select any two products from bucket 2 and choose a free product from the bucket 1

BUY 2 & GET OTHER @25% DISCOUNT

- Customers can choose two products from any group and receive a 25% discount on the third product from the same group. The discount will be applied to the item with the lowest price

Note: The above strategies can be implemented based on customer behavior and should be regularly updated with current market trends

DISCOUNT STRATEGY

- Since **poultry is the most frequently purchased item** in the store, the store could consider **offering regular discounts on it on a weekly or monthly basis**. Since **customers primarily come to the store to buy poultry items**, it could be beneficial to **place these items at the farthest reach from the entrance**. This would allow customers to see and **potentially purchase other products before reaching the poultry section**. This could lead to increased sales of other products
- **Placing products like flour, eggs, butter, pasta, milk, and sandwich loaves in close proximity to one another** in the store may encourage customers to purchase them in combination. Offering discounts on select products based on recent orders may also help increase sales
- Customers **who purchase dishwashing liquid/detergent, cheeses, waffles, and soda are twice as likely to buy flour** if it is recommended to them, The store can offer a "**buy three, get one free**" promotion on dishwashing liquid/detergent, cheeses, waffles, and soda, and include a discount on the purchase of flour when all four items are purchased together or **could offer a 10% discount on the purchase of flour** when customers buy the other items.

THANK YOU