Market Basket Analysis Report

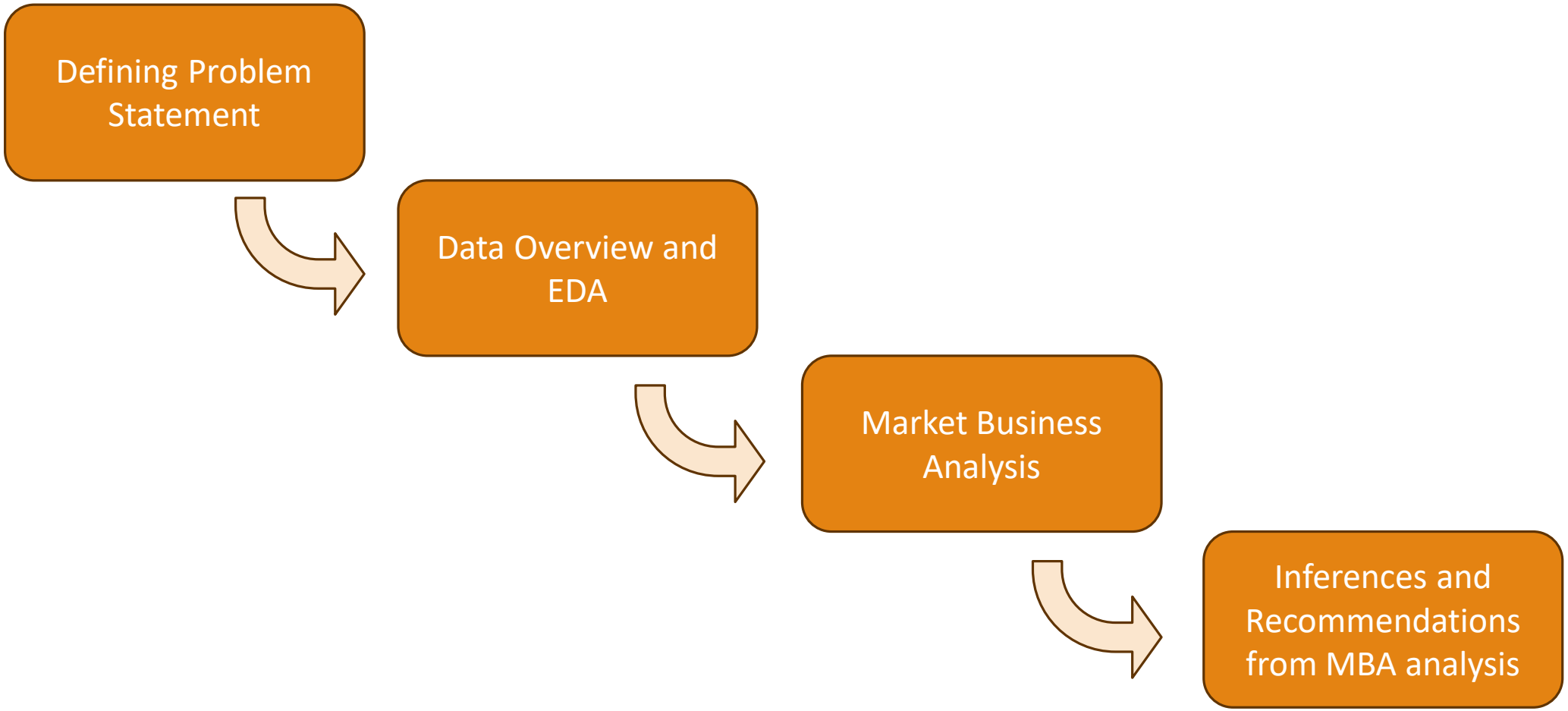
MRA Project Part-B

Sanjay Rajan J

Batch: PGP-DSBA (PGPDSBA.O.JULY24.A)

sanjaysujana@gmail.com

AGENDA


Defining Problem Statement
Data Overview and EDA
Market Business Analysis
Inferences and Recommendations from MBA analysis

PROBLEM STATEMENT :

Business Context:

In the highly competitive grocery retail industry, understanding customer buying patterns is crucial for enhancing sales, increasing customer satisfaction, and improving profitability. By identifying frequently purchased item combinations, grocery stores can craft effective marketing strategies, optimize inventory management, and tailor promotions to meet customer needs. Leveraging Point of Sale (POS) data can unlock valuable insights that drive customer-centric offerings, such as combo packs, discounts, and targeted promotions, which can increase basket size and improve customer retention. This analysis aligns with business goals by maximizing revenue, reducing operational costs, and boosting customer loyalty.

Objective:

As a business analyst, the goal is to analyze the POS transactional data to identify frequently purchased item combinations. Using association rule mining or similar techniques, the aim is to uncover patterns that will help the store create targeted combo offers and discounts, ultimately driving revenue growth by increasing customer purchases and average basket size.

Summary Statistics:

Shape: 20641 rows and 3 Columns

Basic Info

```
<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
```

Numerical Statistics

	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

Null Value Summary

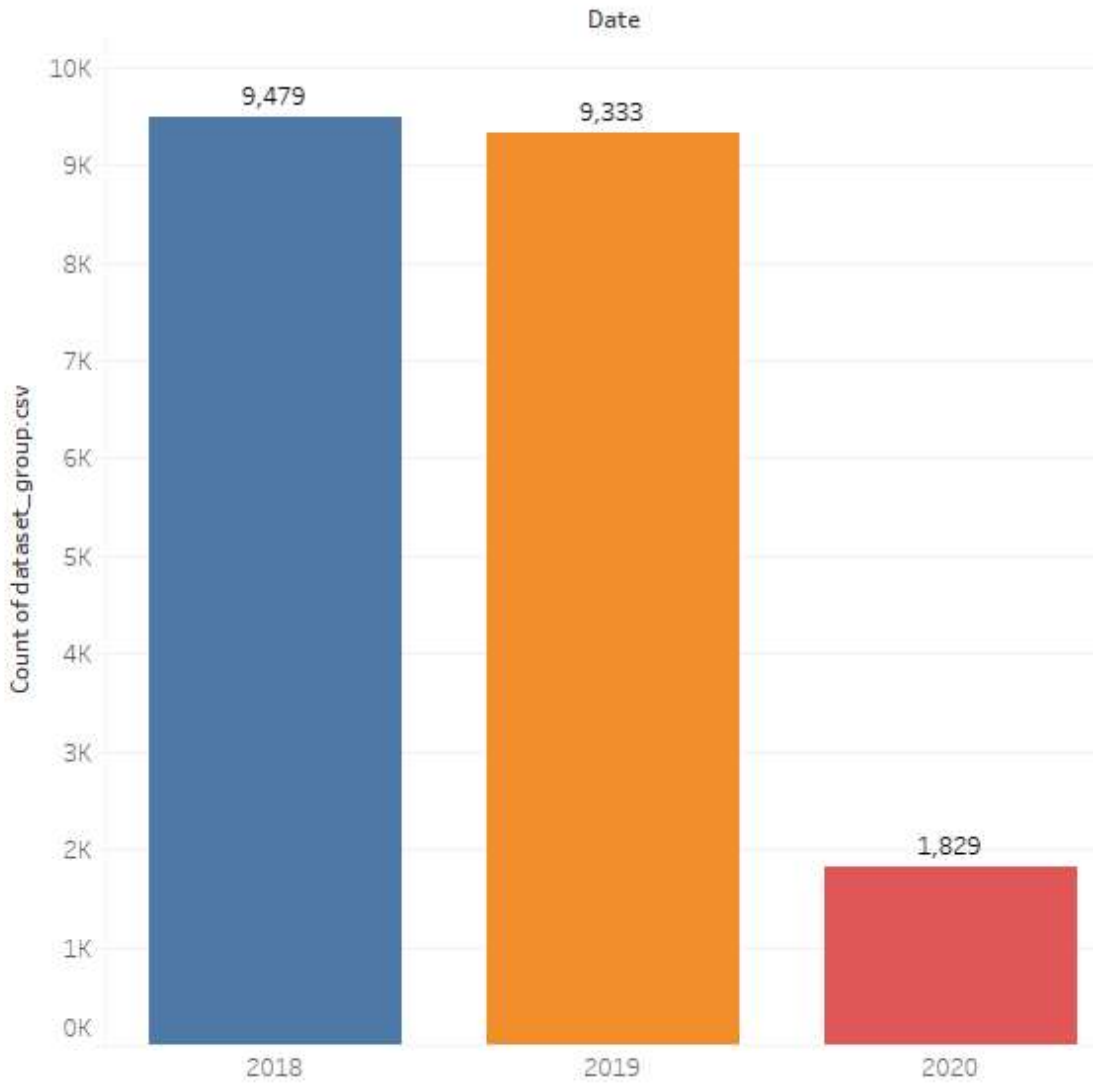
```
Date        0
Order_id    0
Product     0
dtype: int64
```

First 5 rows

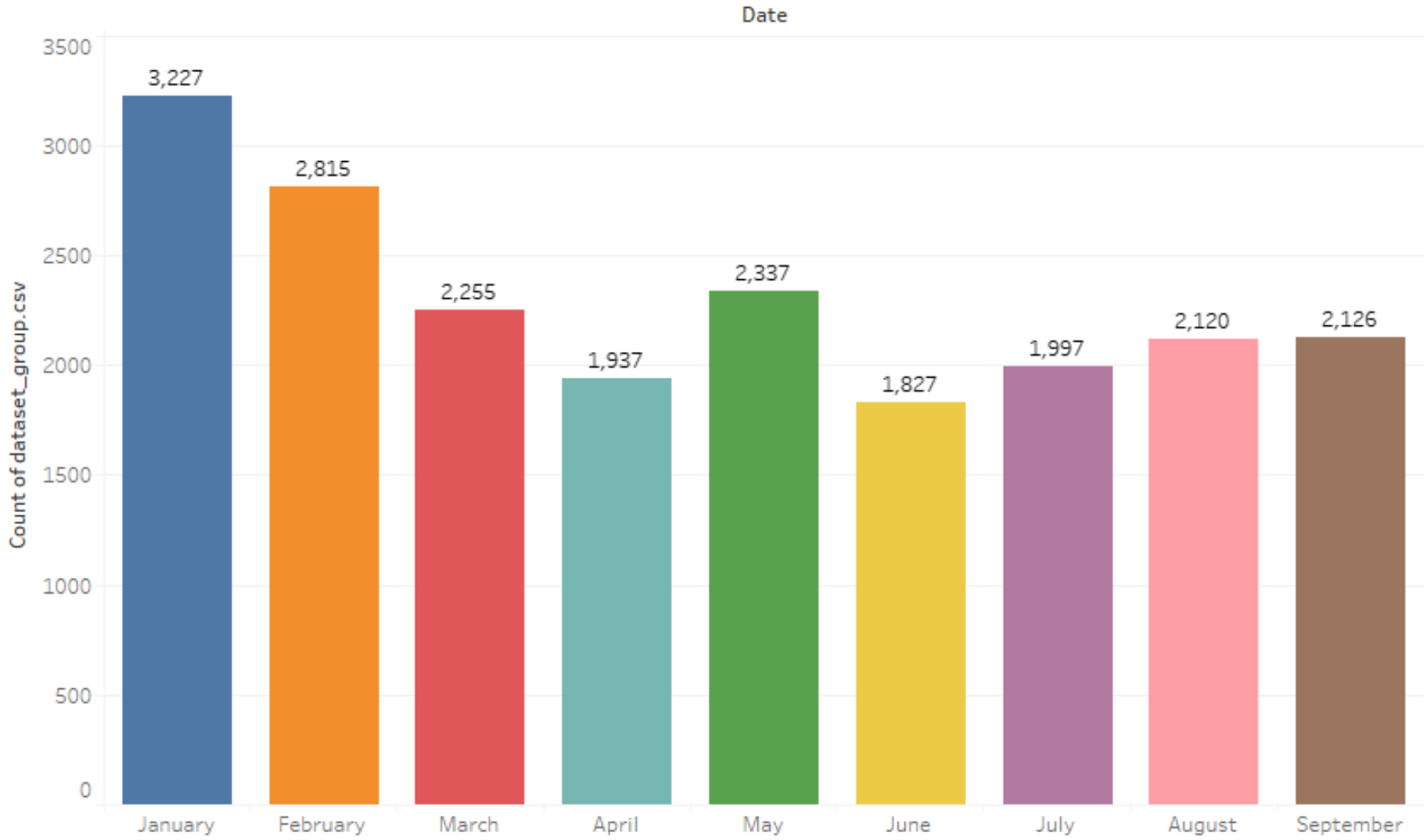
	Date	Order_id	Product
0	01-01-2018	1	yogurt
1	01-01-2018	1	pork
2	01-01-2018	1	sandwich bags
3	01-01-2018	1	lunch meat
4	01-01-2018	1	all- purpose

Exploratory Data Analysis

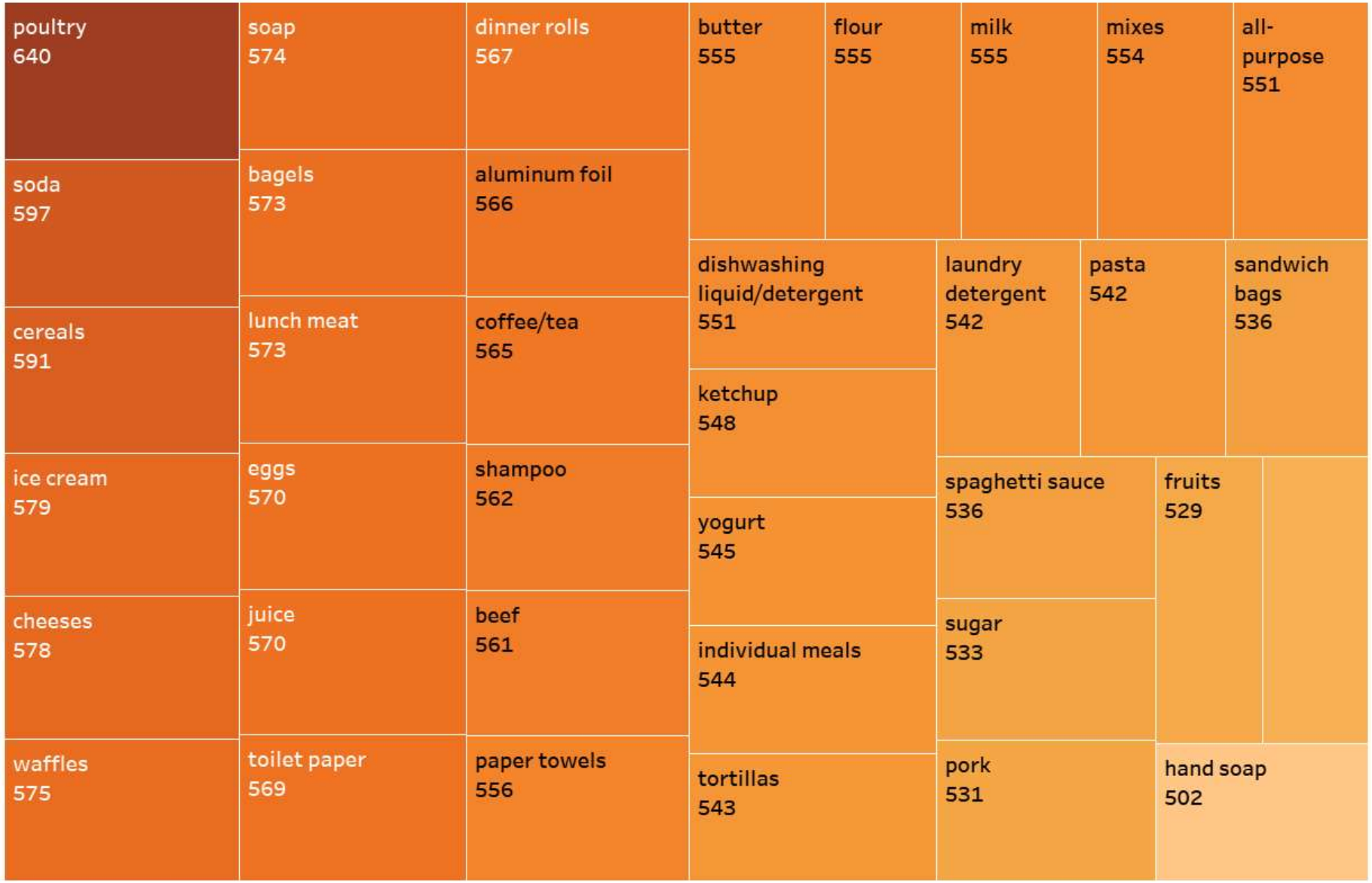
Yearly Order Counts



Monthly Order Count

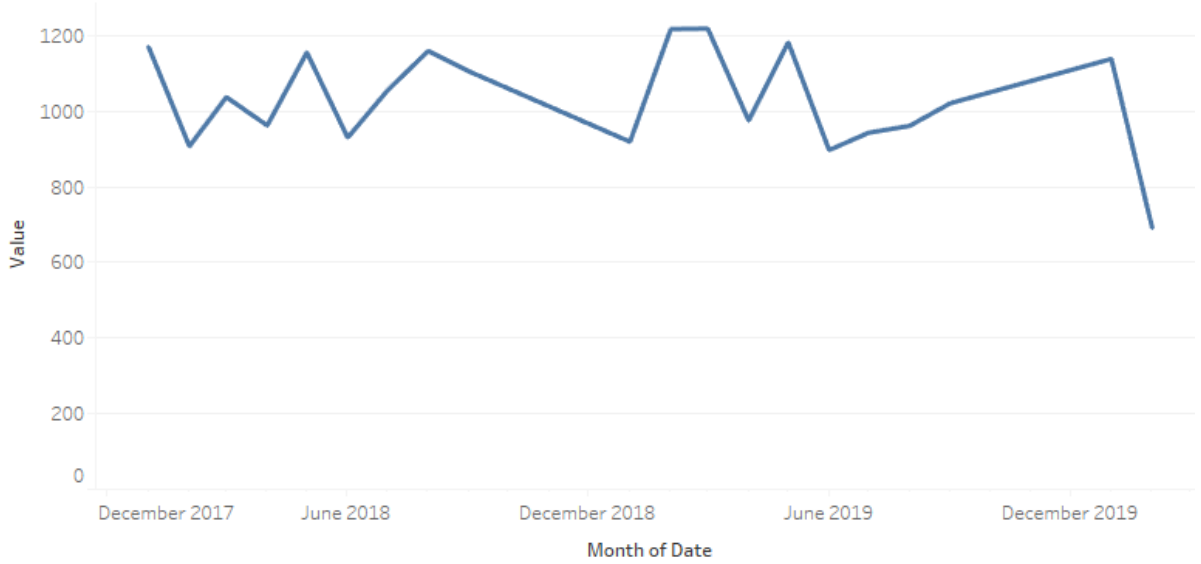


Product Heatmap

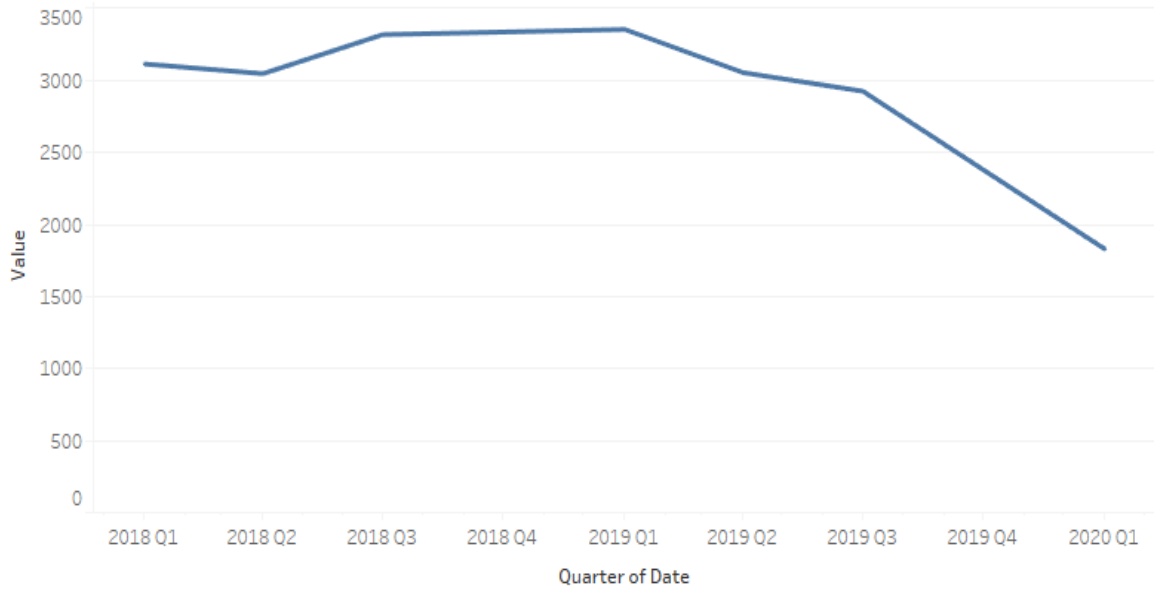


Week Trend

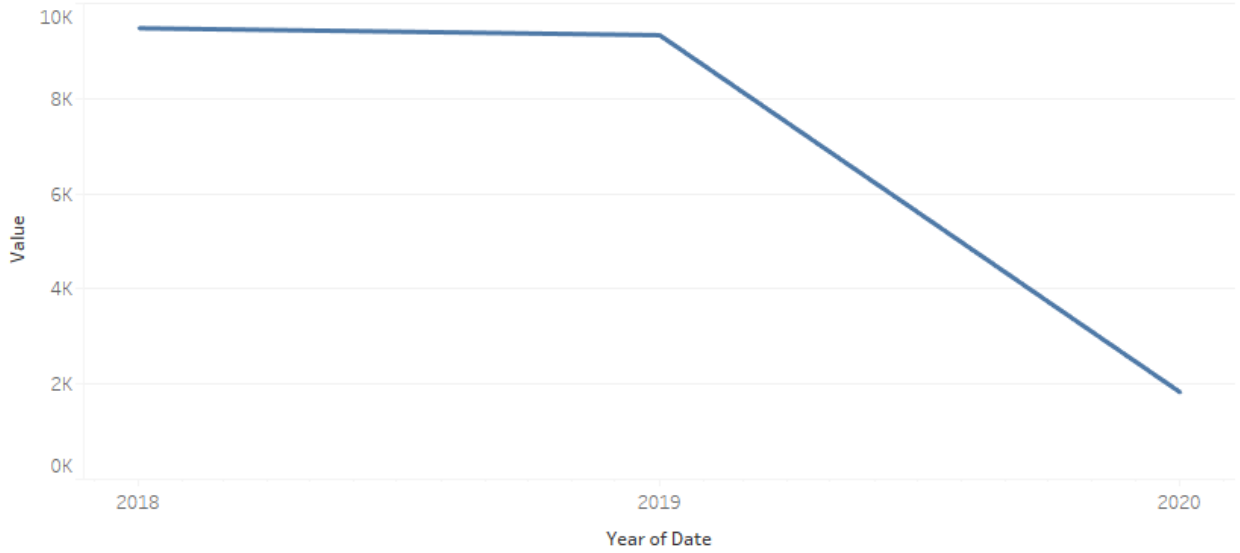
Month Trend



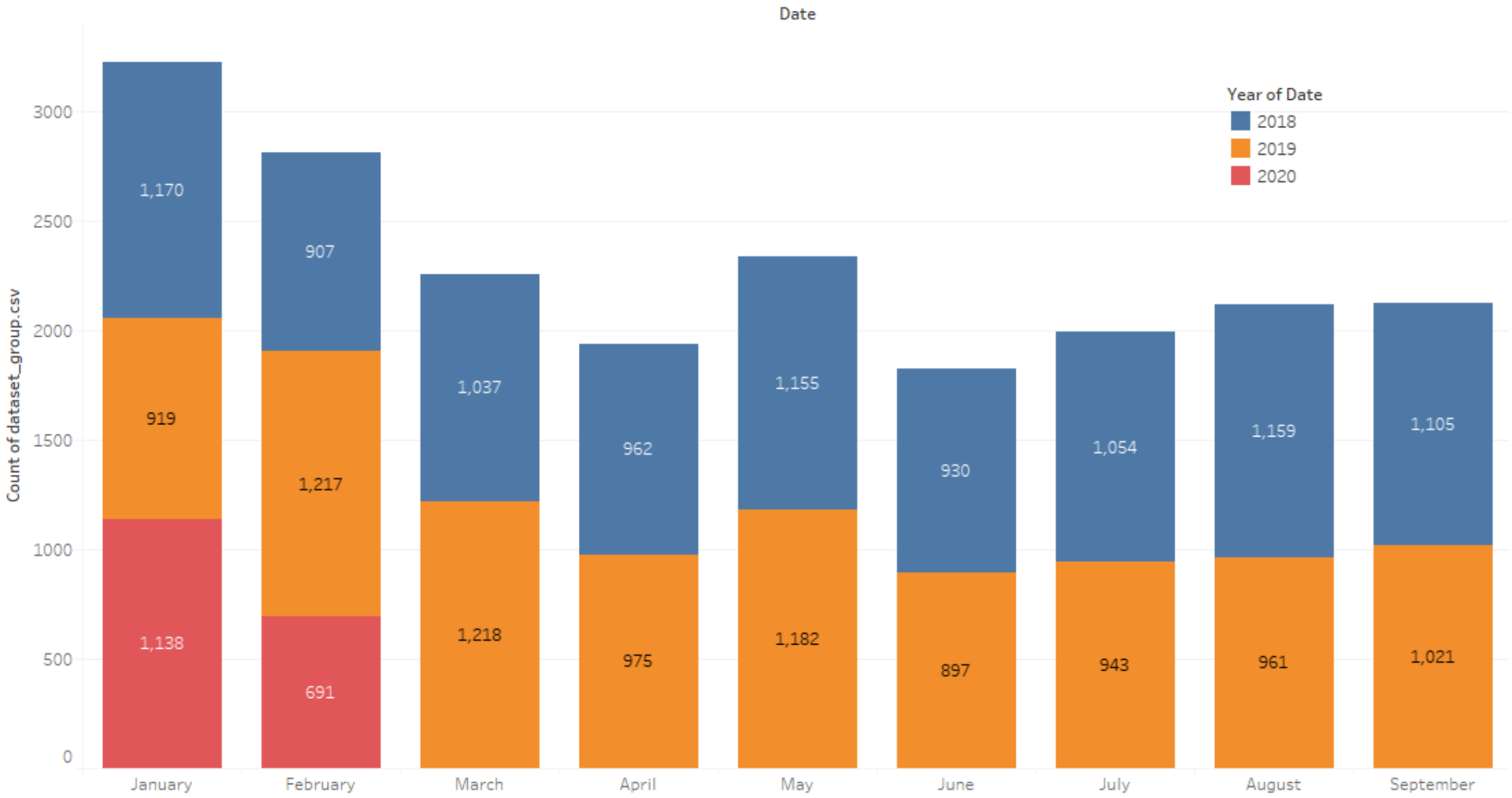
Quarterly Trend



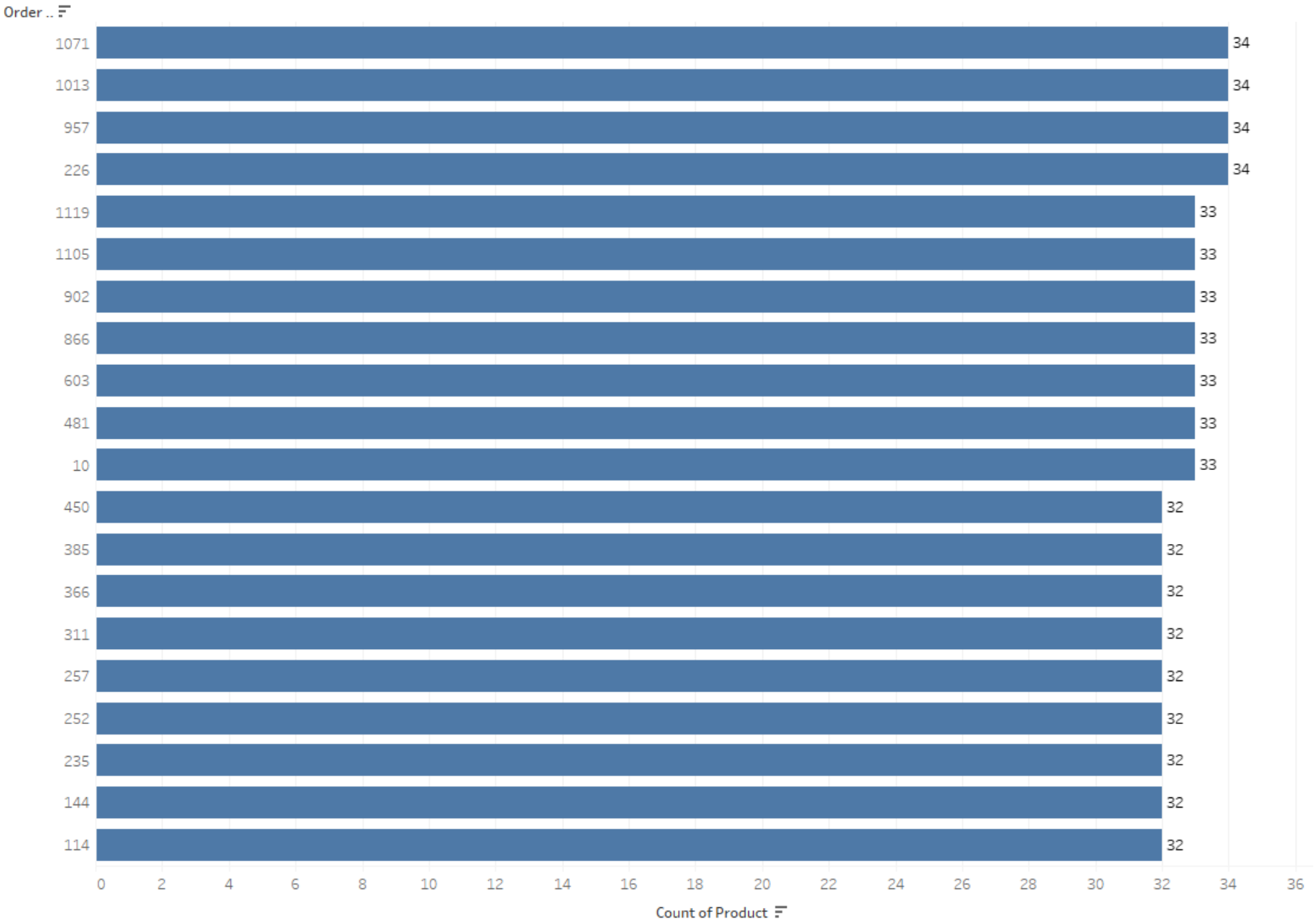
Yearly trend



Monthly Distribution of Orders by years



Top 20 Orders by Product Count



Inferences:

- The year **2018** has the highest order counts and the year **2020** has the lowest order counts.
- **Poultry** is the most purchased product in this dataset (640).
- **Soda, lunch meat, cereals, ice cream, cheeses and waffles** also perform well.
- The years 2018 and 2019 were consistent in orders.
- **January** month has the highest order counts compared to other months (3227).
- Weekly trend suggests high volatility with constant ups and downs.
- Even though sales was recovered after dips, stability and consistency of sales are absent.
- Order count was increased from **Q1 to Q3 in 2018**, but it was drastically reduced in **2019** and **2020**.
- Sales was gradually increased from **February 2019** and attains its peak.

Tools used for Market Basket Analysis:

- Tableau (For EDA)

- KNIME

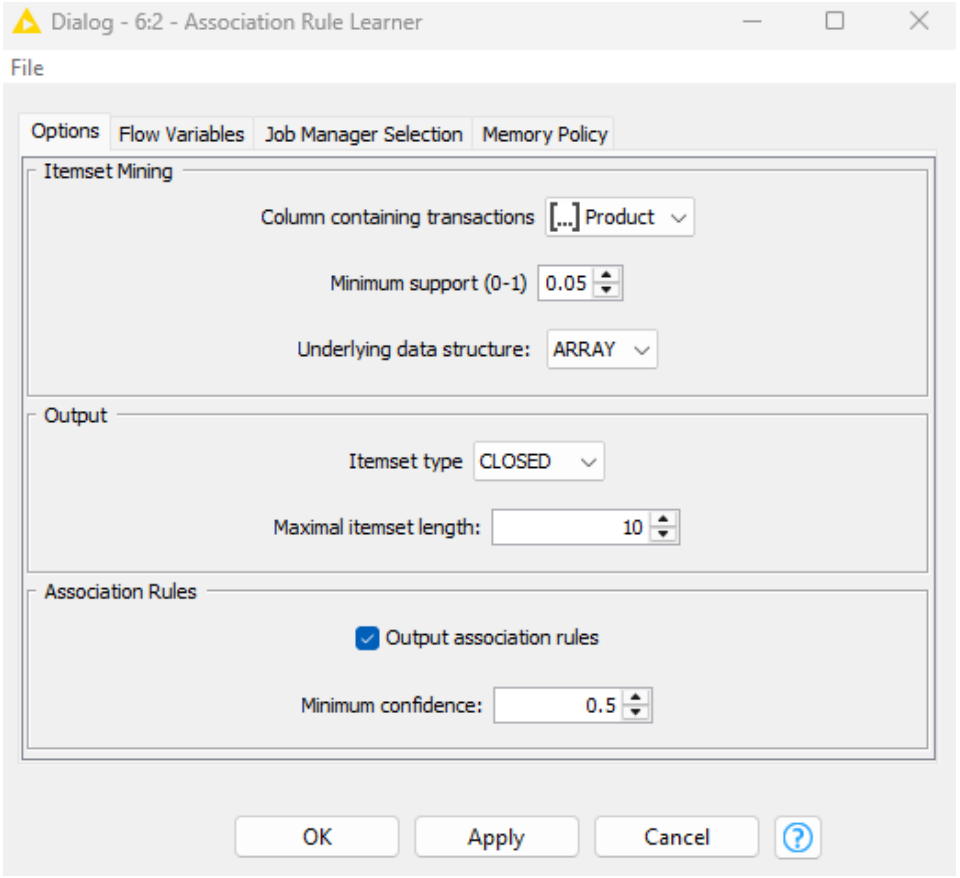
Market Basket Analysis:

- **Market basket analysis (MBA)** is a data mining technique that is used to uncover purchase patterns in any retail setting.
- MBA is a set of statistical affinity calculations that help business leaders better understand – and ultimately serve – their customers by highlighting purchasing patterns.
- In simplest terms, MBA looks for what combinations of products most frequently occur together in transactions. These relationships can be used to increase profitability through cross-selling, recommendations, promotions, or even the placement of items on a menu or in a store.

Parameters :

- **Support**: The proportion of transactions that contain a particular product or product combination.
- **Confidence**: The likelihood that a customer who buys item A will also buy item B.
- **Lift**: How much more likely item B is purchased when item A is purchased, compared to random chance.

Association Rule Learner parameters:

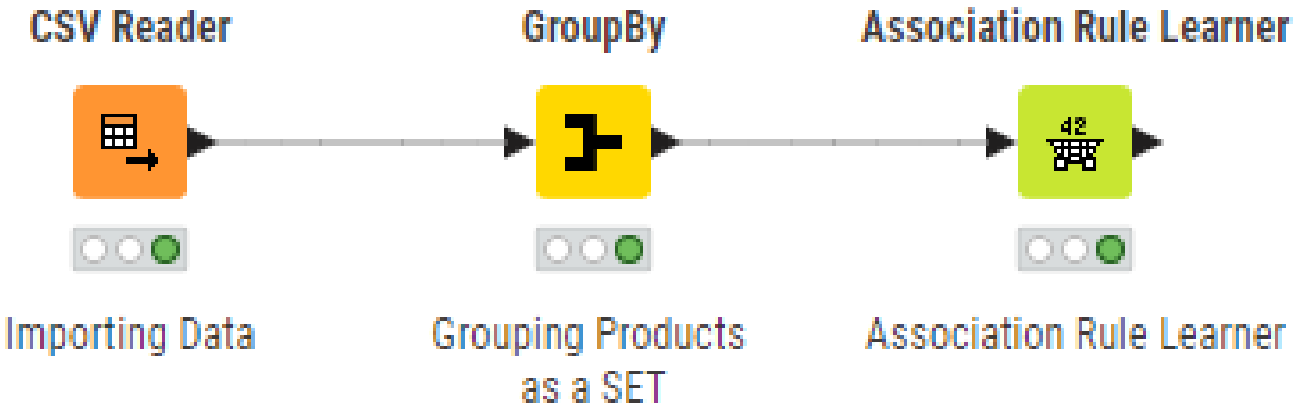


Minimum Support = 0.05

Minimum Confidence = 0.5

Maximum Itemset length = 10

KNIME Workflow:



Association Rule Output Table :

Rows: 1247 | Columns: 6

#	RowID	Support *Number (double)*	Confidence *Number (double)*	Lift ↓ *Number (double)*	Consequent *String*	implies *String*	Items *Set*
60	rule59	0.055	0.649	1.791	paper towels	<---	[eggs,ice cream,pasta]
59	rule58	0.055	0.643	1.731	pasta	<---	[paper towels,eggs,ice cream]
22	rule21	0.051	0.674	1.726	cheeses	<---	[bagels,cereals,sandwich bags]
4	rule3	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum foil]
19	rule18	0.051	0.63	1.678	mixes	<---	[yogurt,poultry,aluminum foil]
21	rule20	0.051	0.611	1.66	sandwich bags	<---	[cheeses,bagels,cereals]
53	rule52	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce,poultry,laundry detergent]
41	rule40	0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce,poultry,ice cream]
8	rule7	0.05	0.62	1.645	juice	<---	[yogurt,poultry,aluminum foil]
44	rule43	0.052	0.686	1.628	poultry	<---	[dinner rolls,spaghetti sauce,ice cream]
50	rule49	0.052	0.634	1.627	eggs	<---	[paper towels,dinner rolls,pasta]
51	rule50	0.052	0.602	1.621	pasta	<---	[paper towels,eggs,dinner rolls]
25	rule24	0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce,poultry,cereals]
58	rule57	0.055	0.63	1.616	eggs	<---	[paper towels,ice cream,pasta]
12	rule11	0.05	0.613	1.616	coffee/tea	<---	[yogurt,cheeses,cereals]
45	rule44	0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce,poultry,juice]
36	rule35	0.052	0.628	1.61	eggs	<---	[dinner rolls,poultry,soda]
55	rule54	0.054	0.598	1.603	spaghetti sauce	<---	[dinner rolls,poultry,laundry detergent]

Inferences:

Strongest Rule (Sorted by Highest Lift)

- Rule: {eggs, ice cream, pasta} ---> {paper towels}
- Lift: 1.791
- Confidence: 64.9%
- Support: 5.5%
- When customers buy **eggs**, **ice-cream**, and **pasta** together, they are **1.79** times more likely to also buy **paper towels** than random chance.

Top 5 Rules (Sorted by Highest Lift) :

1. {eggs, ice cream, pasta} ---> {paper towels}
2. {paper towels, eggs, ice cream} ---> {pasta}
3. {bagels, cereals, sandwich bags} ---> {cheeses}
4. {yogurt, toilet paper, aluminium foil} ---> {juice}
5. {yogurt, poultry, aluminium foil} ---> {mixes}

Bottom 5 Rules (Sorted by Lowest Lift) :

1. {pasta} ---> {poultry}
2. {hand soap, toilet paper} ---> {poultry}
3. {fruits, toilet paper} ---> {poultry}
4. {cheeses, hand soap} ---> {poultry}
5. {hand soap, ketchup} ---> {poultry}

Actionable Insights and Business Recommendations

- Bundle offers can be created for the customer segment buying **eggs, ice-cream and pasta**. For example: **“Buy 3, get 5% off for paper towels”.** This will boost the sales of all 4 products.
- Cross-Promotion can be implemented in the grocery stores to boost the frequency.
- For high lift combinations, the consumers can be targeted with personalized offers and discount.
- Stock Availability should be regularly checked and maintained to prevent cross-sell opportunities.
- Investing in Marketing Campaigns and offers for the products are good rather than low-lift combinations.
- As we clearly see that a numerous low lift rules has **“poultry”** as consequent. Customers are weakly associating poultry with other combinations.
- Local grocery stores can create a custom designed bundle involving products specifically for the locals. For example: “**Family Essentials bundle”** involving {paper towels, eggs, ice-cream, pasta} or **“Sanitation Bundles”** involving {Dishwashing liquid, detergent, soap, paper towels}