PROJECT REPORT
ON
MARKETING & RETAIL ANALYTICS
(PART - B)

BY

SHAJIL FERNANDEZ

04-02-2024

Table of Contents

- Problem statement

- Information about data

- EDA and Inferences

- Market Basket analysis

- Association rules and possible combos

Agenda & Executive Summary of the data

- To conduct a thorough analysis of Point of Sale (POS) data.

- To identify the most commonly occurring sets of items in the customer orders.

- To provide recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.

Problem Statement

A grocery store shared the transactional data with you. Your job is to conduct a thorough analysis of Point of Sale (POS) data, identify the most commonly occurring sets of items in the customer orders, and provide recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.

Information about data

- There are 20641 rows and 3 columns in the given dataset.

- There are 1 int64 and 2 object datatypes.

- There are no missing values in the given data.

- There are 4730 duplicate values.

- There are 37 unique products, of which frequency of poultry is the most.
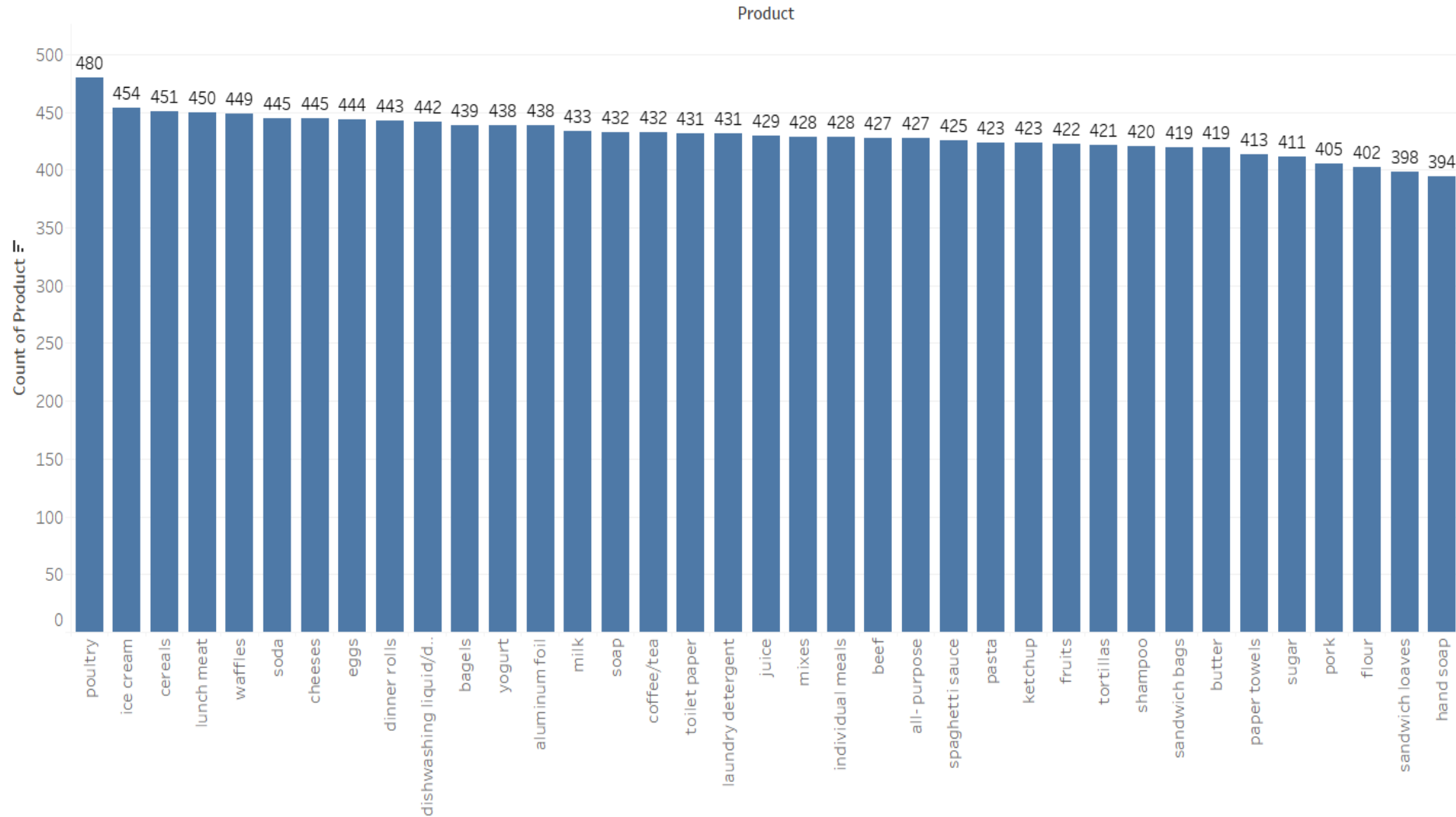
Assumptions:

- For EDA, we have removed duplicates and analysed the data.

- The data is available only till February 2020, and for 2018 and 2019 Q4 data is missing, however, incompleteness of the data have been ignored.

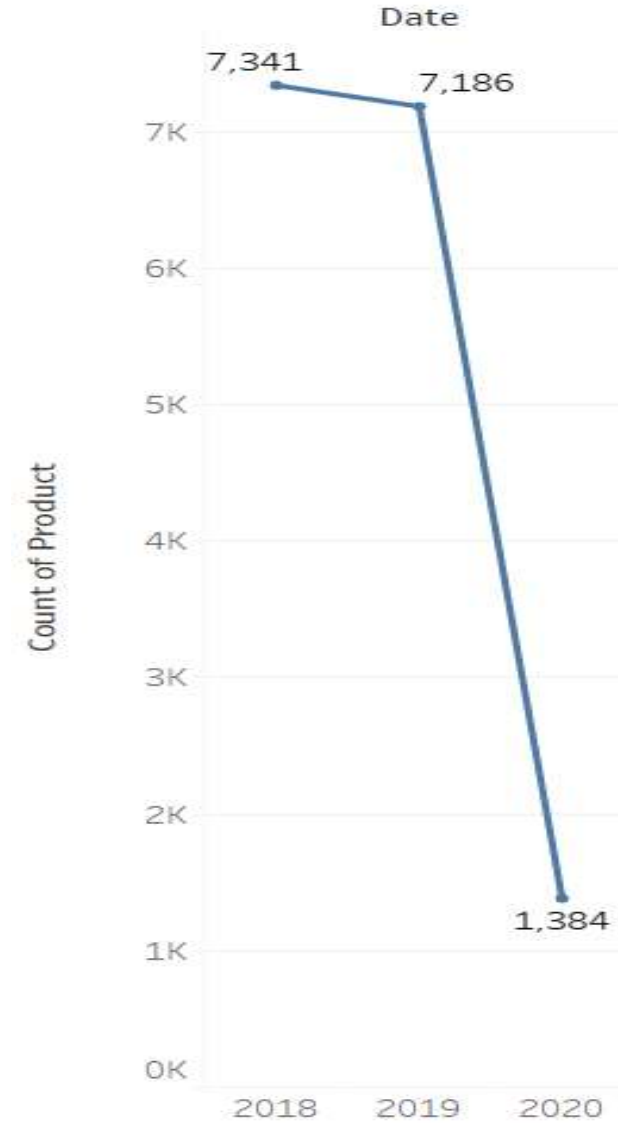# EDA and Inferences

Insights:

- Poultry is the highest with product count of 480.
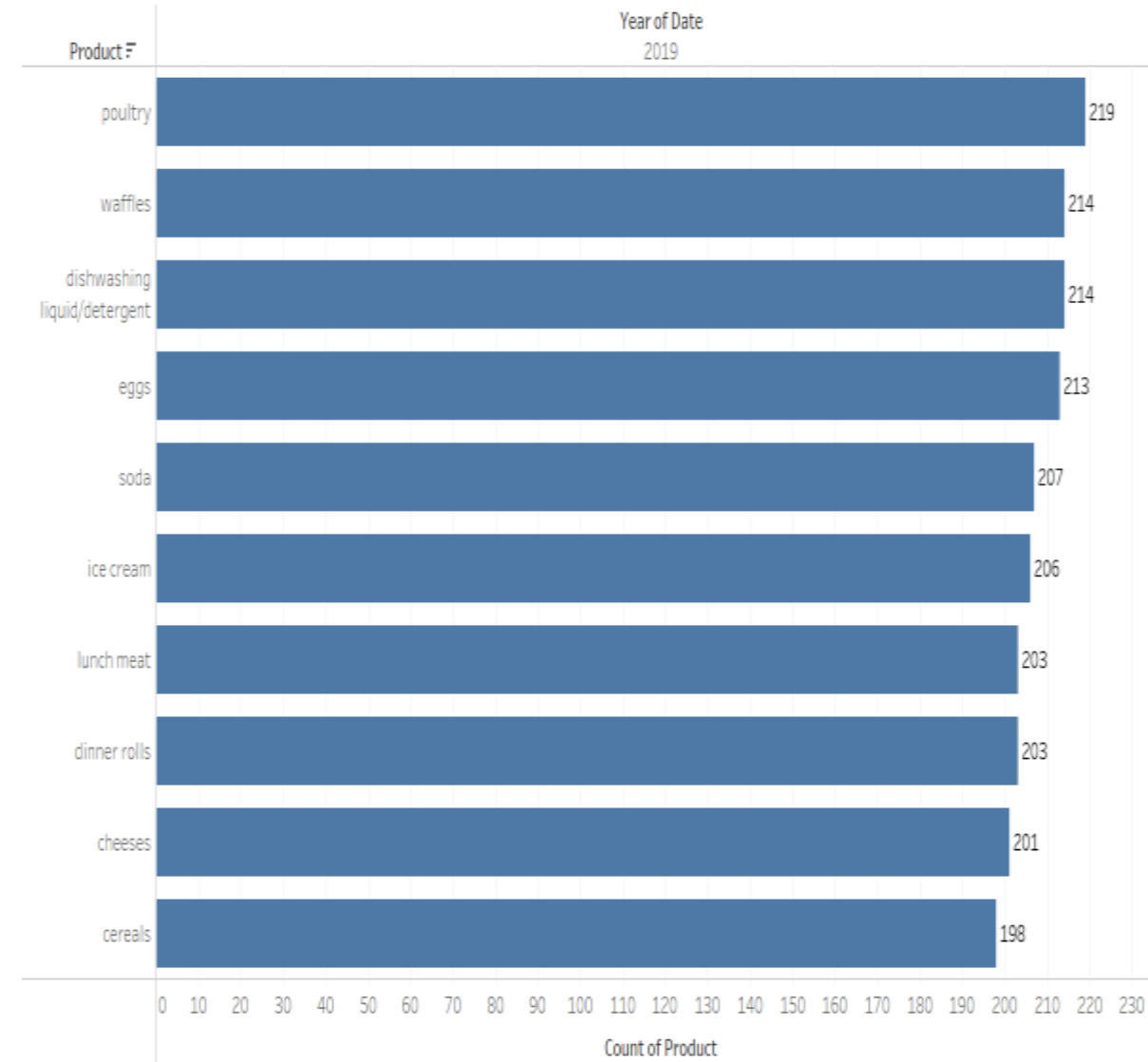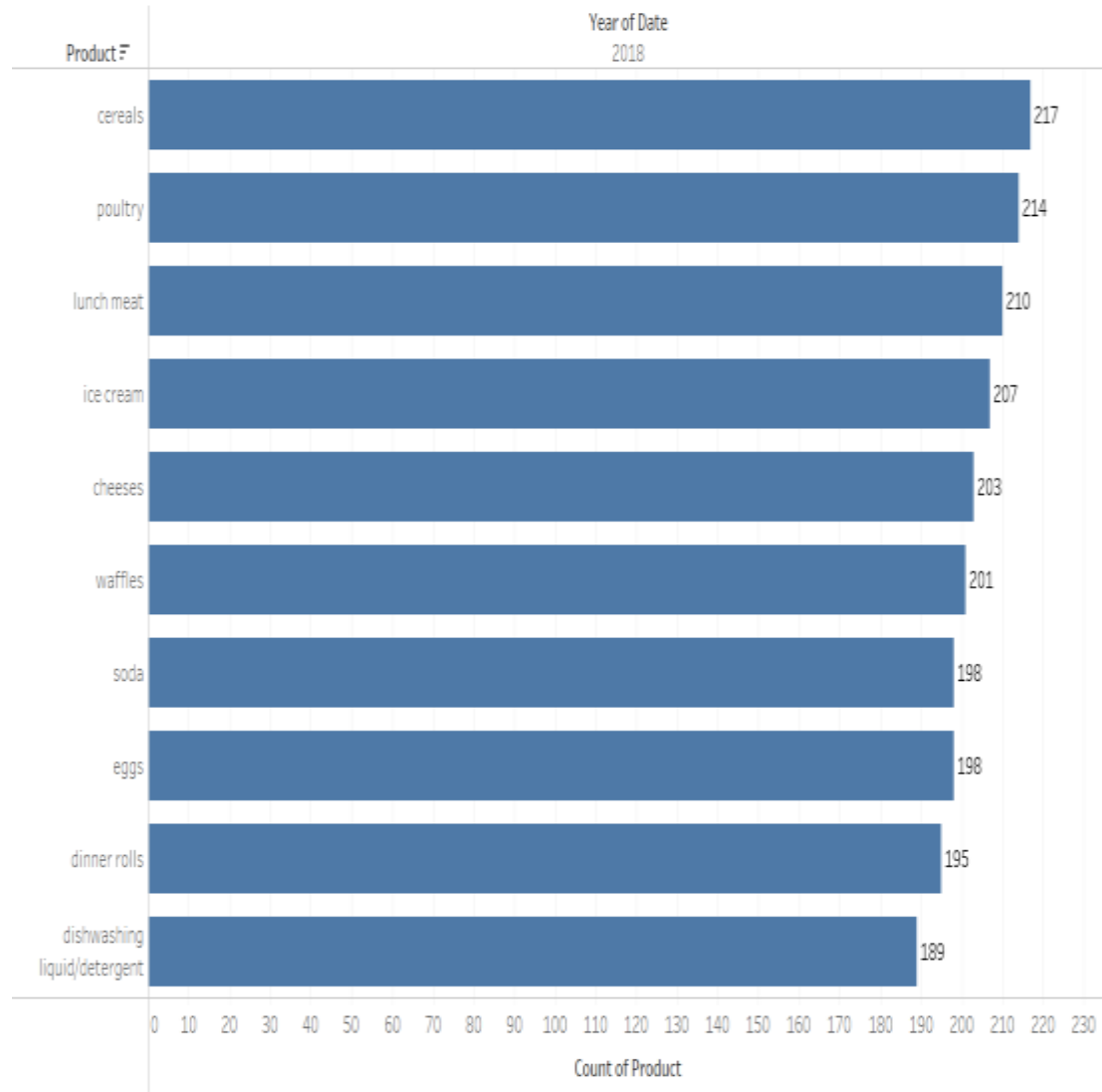- Hand soap is the least with count of 394.
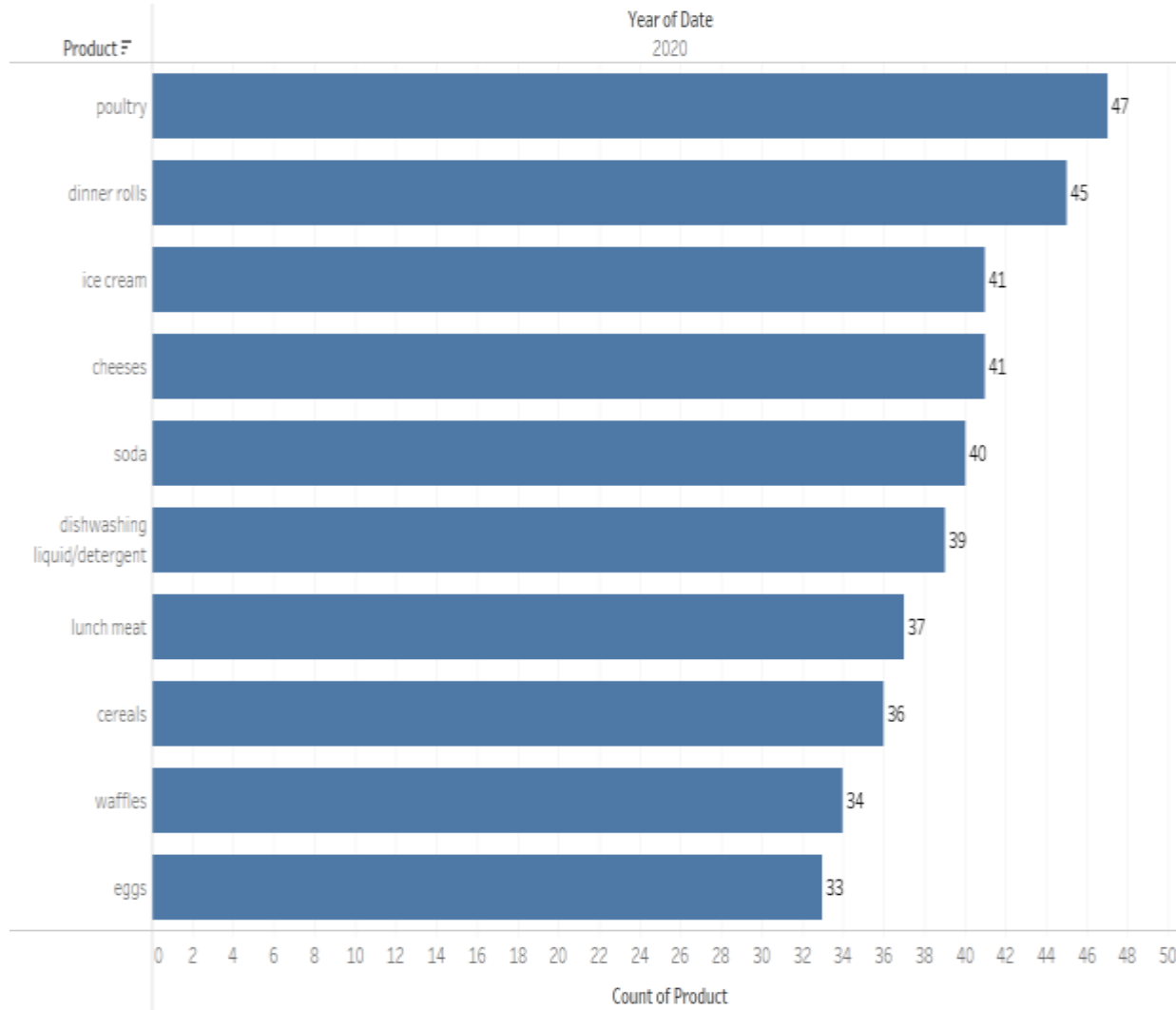
Year wise product count

Insights:
- In 2018, there were 7341 product count, which decreased to 7186 in 2019.
- In 2020, product count for two months sums up to 1384.

# Top 10 product count of individual years

Top 10 product count of individual years (continue)

Year of Date
2020

Product	Count of Product
poultry	47
dinner rolls	45
ice cream	41
cheeses	41
soda	40
dishwashing liquid/detergent	39
lunch meat	37
cereals	36
waffles	34
eggs	33

Insights:
- In 2018, cereals were the most preferred product, followed by poultry.
- In 2019, poultry was the highest and cereals were 10th among top 10 products based on the count.
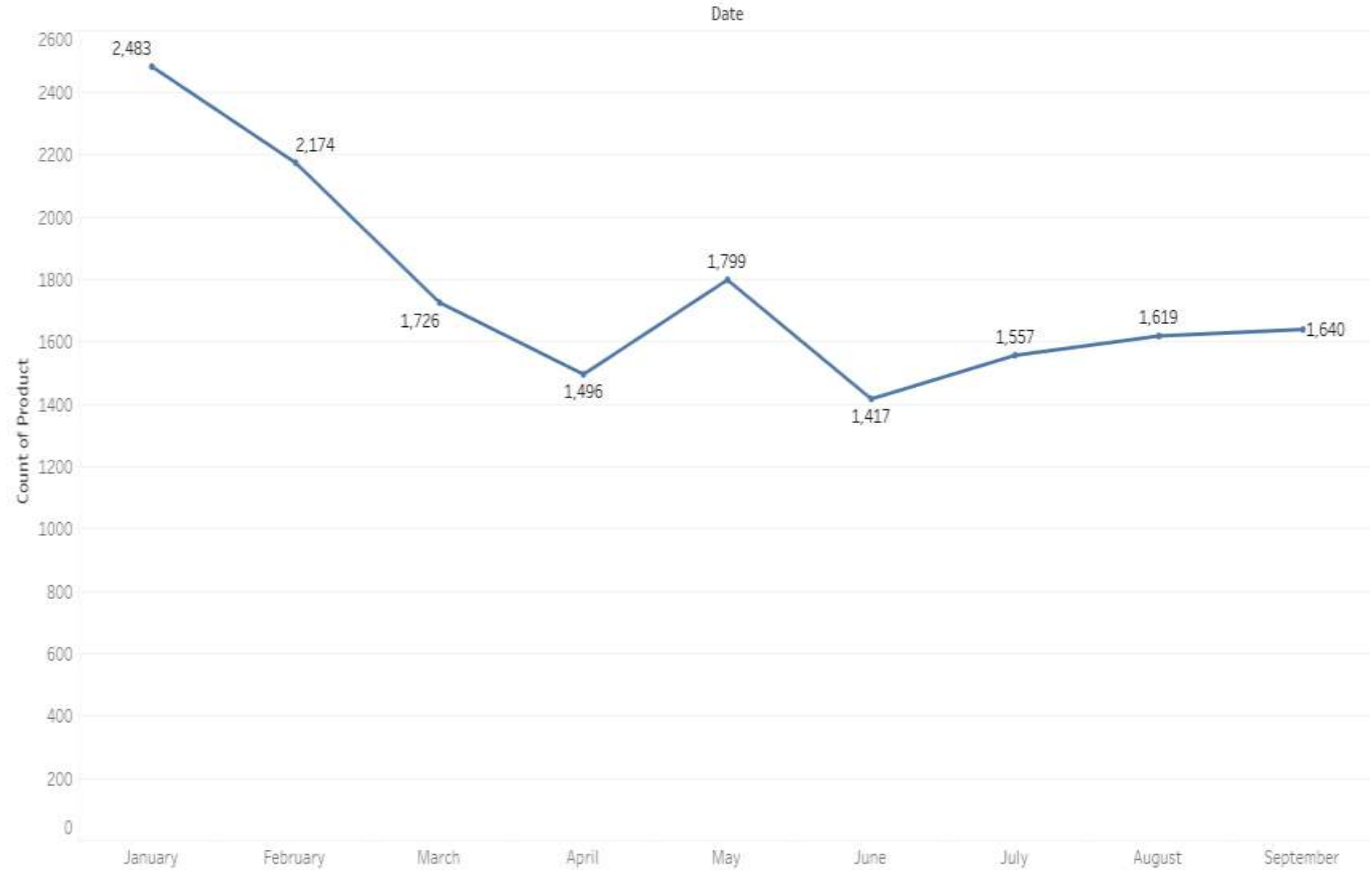
Quarter wise product count
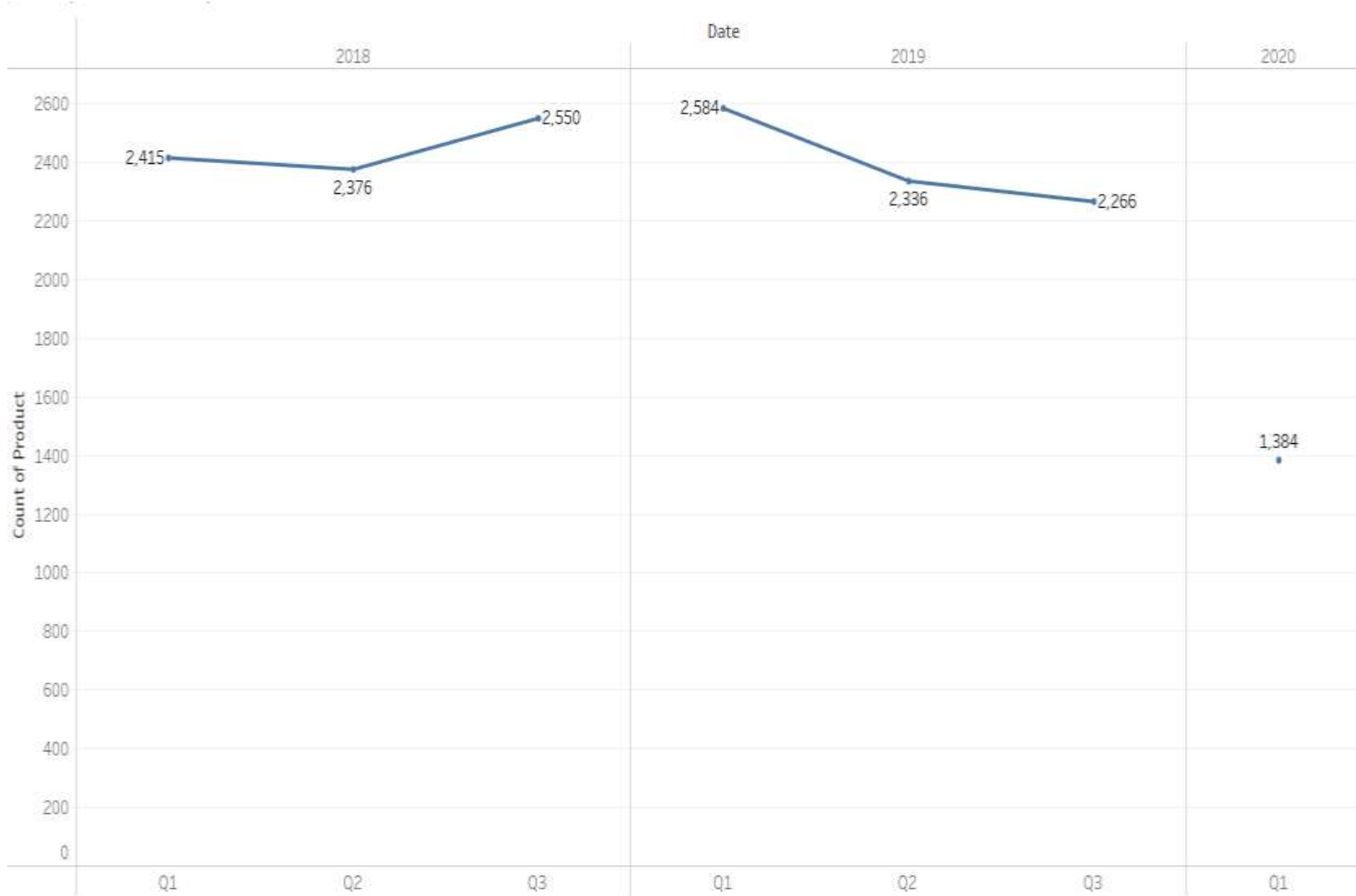
Insights:
- In 2018, Q2 has the least product count and Q3 has the highest.
- In 2019, Q1 has the highest product count and Q3 has the least.

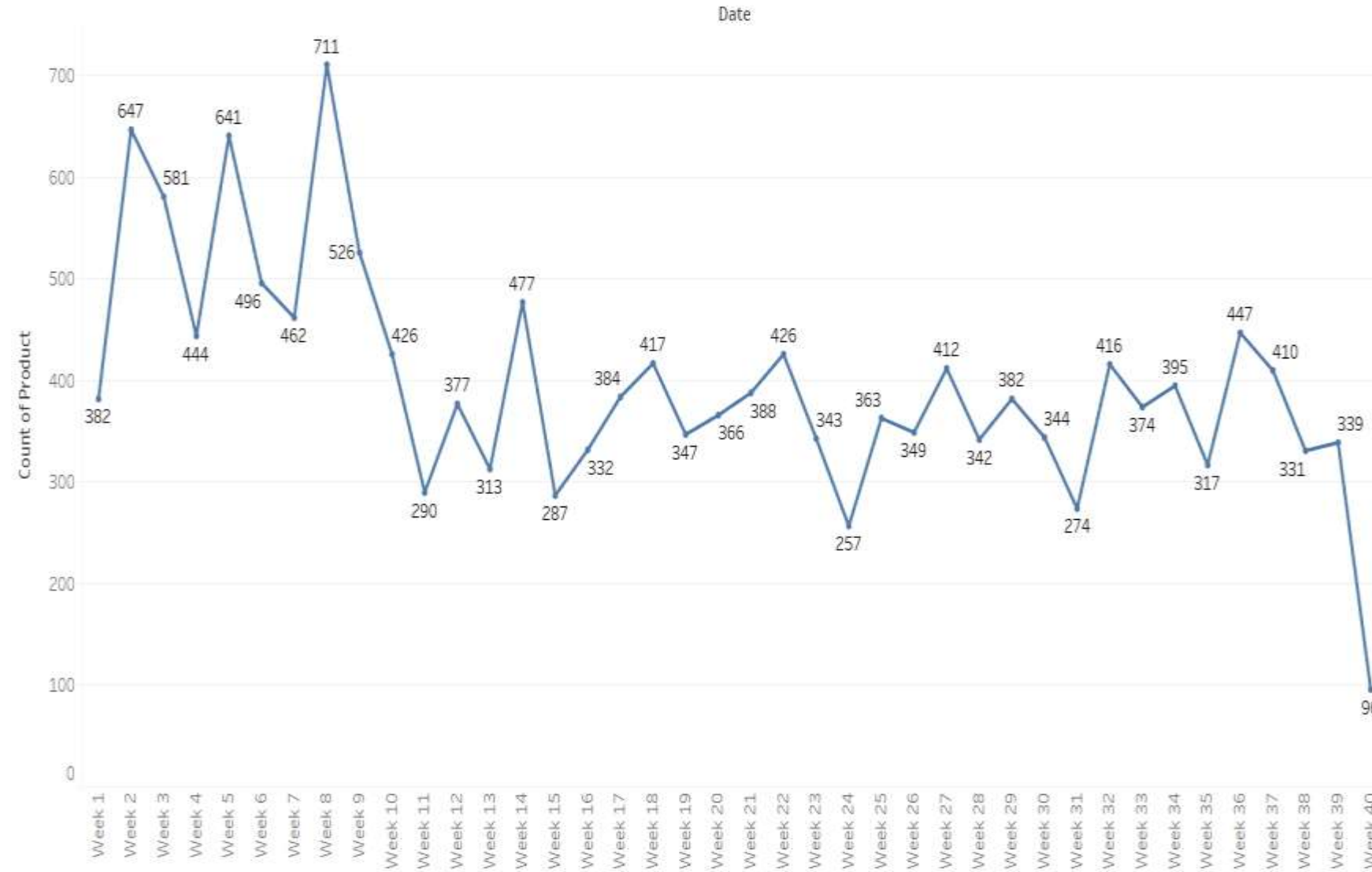# Month wise product count

Insights:

- January is the highest with 2483 product counts.
- June is the least with 1496 counts.

Week wise product count

Insights:

- Week 8 is the highest (711), which is followed by Week 2 (647).
- Week 40 is the least with 96 counts.
- We can also observe that initial weeks has the highest product counts.

Week days product count

Product		Sunday		Monday		Tuesday		Date Wednesday		Thursday		Friday		Saturday
all- purpose		72		59		60		56		58		57		65
aluminum foil		70		51		68		52		72		69		56
bagels		70		52		72		54		70		67		54
beef		65		53		56		59		55		66		73
butter		58		55		51		61		63		58		73
cereals		71		58		65		55		67		72		63
cheeses		70		53		65		62		68		67		60
coffee/tea		69		48		70		64		69		61		51
dinner rolls		56		54		69		67		61		70		66
dishwashing liquid/..		64		54		63		73		57		65		66
eggs		67		52		64		66		67		66		62
flour		59		56		60		58		53		55		61
fruits		64		49		56		66		67		58		62
hand soap		51		49		51		65		54		65		59
ice cream		70		63		63		62		72		68		56
individual meals		70		54		61		65		58		65		55
juice		69		59		64		62		55		62		58
ketchup		65		50		61		68		55		61		63
laundry detergent		72		57		53		66		59		60		64
lunch meat		74		52		65		52		69		70		68
milk		66		57		69		56		62		65		58
mixes		60		52		56		63		72		67		58
paper towels		62		50		63		52		63		58		65
pasta		56		52		56		71		56		61		71
pork		62		58		55		52		56		67		55
poultry		60		68		62		68		66		78		78
sandwich bags		62		55		55		58		59		65		65
sandwich loaves		61		49		65		58		55		62		48
shampoo		69		55		66		61		48		64		57
soap		67		59		67		63		58		53		65
soda		57		60		63		80		57		58		70
spaghetti sauce		74		51		60		62		64		56		58
sugar		53		51		56		73		55		63		60
toilet paper		62		56		64		58		64		65		62
tortillas		61		57		44		64		58		63		74
waffles		75		54		64		63		71		65		57
yogurt		64		52		71		66		57		67		61

Week days product count (continue)
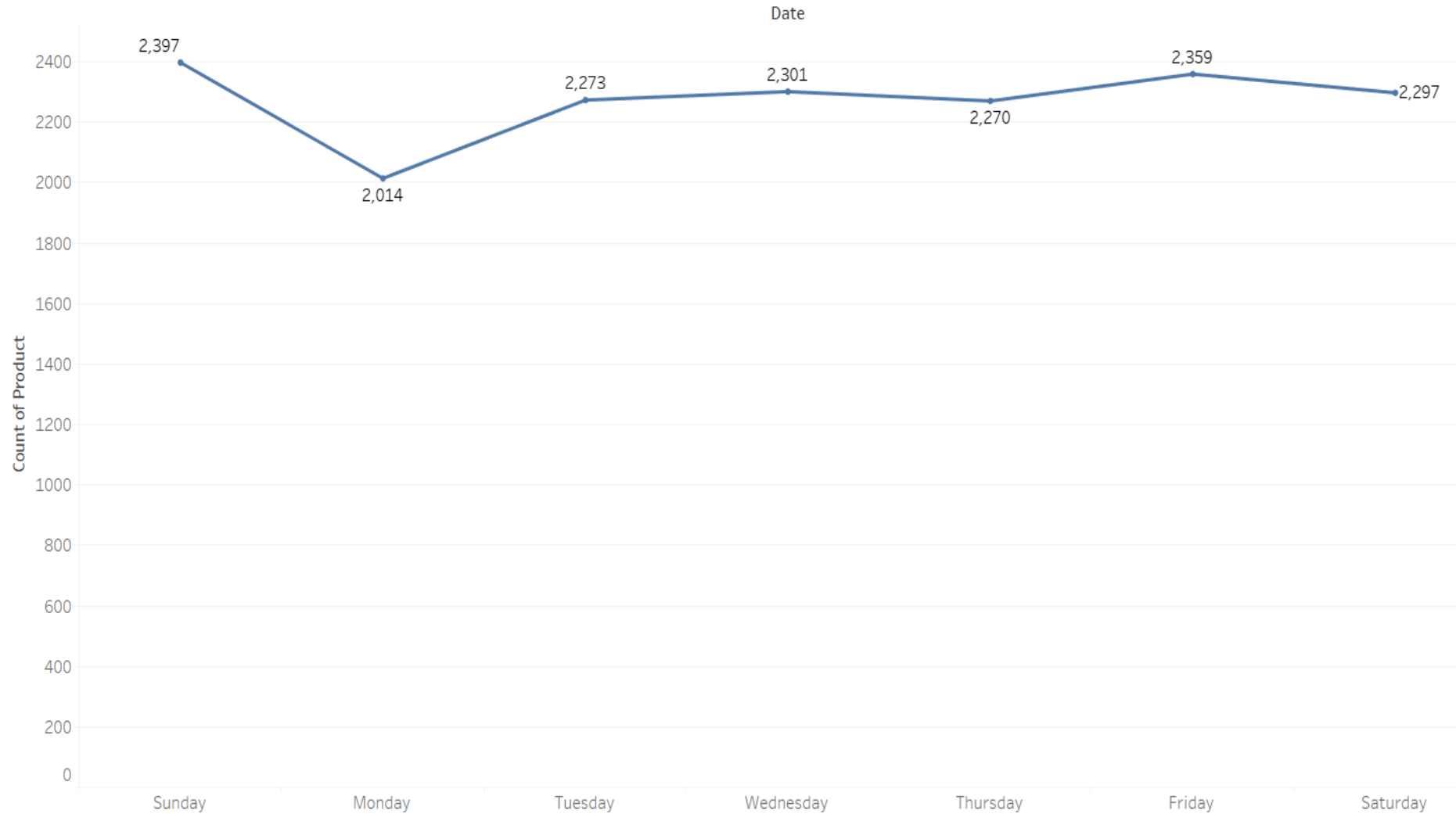
Insights:

- Highest product counts are on Sundays, i.e., 2397.
- Mondays are the least with product count of 2014.

Market Basket Analysis

- Market Basket Analysis is a technique which identifies and study the buying patterns of the customer from the historical data and groups the same, which then used to suggest the future customers based on the association between the products.

Uses of MBA:

- To study the buying behaviour of the customers.
- To boost sales and thereby increase in revenue.
- To strategize the placement of products in the retail stores.
- To make customers buy more products based on the suggestions in e-commerce.
- To plan the stock accordingly based on the chances of purchase.

Association rules:

- Association rules are rules that helps in identifying the relationship between the products, i.e., How good is the relationship between the products? What would be the confidence level of the relationship? What are the chances? etc.

- It helps in knowing the probability, that the customer would end up buying a particular product.

- In our case, Association rules helps in identifying connections between the products based on the hyper-parameter set and thereby suggesting various combos, groups or rules that helps in further analysis or actions.

KNIME workflow
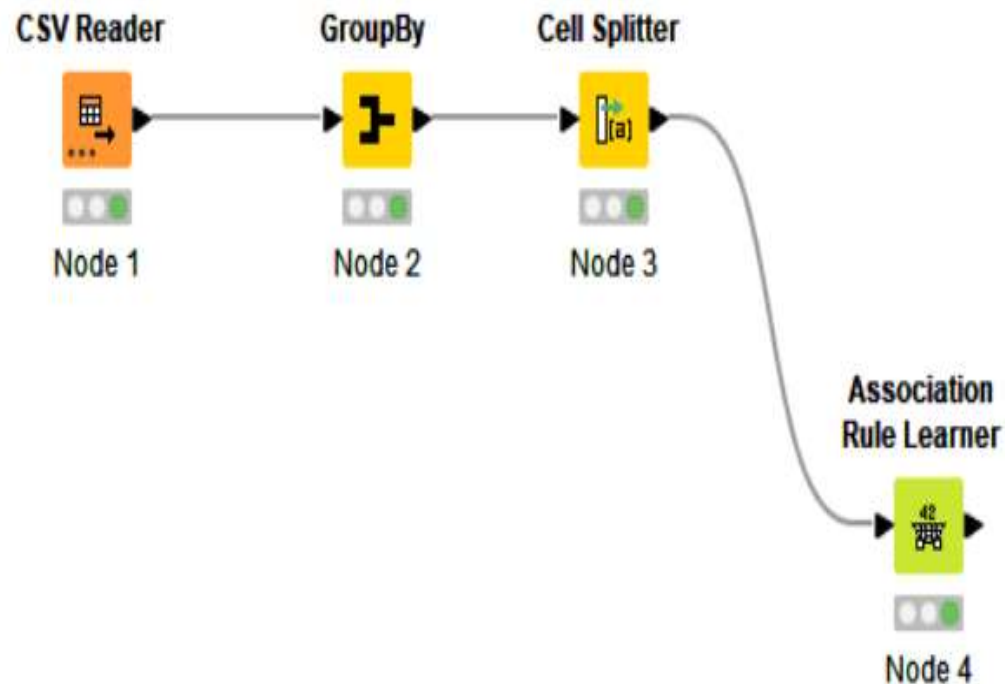
- **CSV reader**: To import and read the dataset which is in csv format.
- **GroupBy:** We are grouping by "orderid" and concatenating the products based on "orderid".
- **Cell Splitter:** Helps in removing duplicates and setting a delimiter "," for the products concatenated.
- **Association Rule Learner:** Setting up rules with support 0.07 and confidence 0.55 to get set of association rules.

Support and Confidence threshold

Threshold set:

- In this case, we have selected 0.07 as the threshold for Support i.e., 7%, as the occurrence of the particular product on other groups will be lower.

- Confidence threshold is set to 0.55, in order to have a good number of association rules.

Associations identified

Row ID	D Support	D Confide...	D Lift	S Conseq...	S implies	[...] Items
rule0	0.073	0.553	1.416	soda	<---	[sandwich bags,sugar]
rule1	0.076	0.585	1.388	poultry	<---	[sandwich loaves,laundry detergent]
rule2	0.078	0.56	1.486	juice	<---	[shampoo,spaghetti sauce]
rule3	0.079	0.552	1.31	poultry	<---	[dinner rolls,pork]
rule4	0.079	0.573	1.36	poultry	<---	[mixes,sugar]
rule5	0.081	0.551	1.41	cheeses	<---	[shampoo,sandwich bags]
rule6	0.081	0.564	1.339	poultry	<---	[juice,sugar]
rule7	0.082	0.554	1.314	poultry	<---	[tortillas,sugar]
rule8	0.083	0.563	1.498	individual m...	<---	[sandwich loaves,lunch meat]
rule9	0.083	0.553	1.312	poultry	<---	[dinner rolls,sandwich loaves]
rule10	0.083	0.556	1.432	dishwashing...	<---	[mixes,soda]
rule11	0.085	0.557	1.323	poultry	<---	[yogurt,sandwich loaves]
rule12	0.085	0.561	1.33	poultry	<---	[toilet paper,sugar]
rule13	0.085	0.554	1.315	poultry	<---	[dishwashing liquid/detergent,coffee/tea]
rule14	0.086	0.551	1.462	juice	<---	[spaghetti sauce,aluminum foil]
rule15	0.086	0.554	1.42	eggs	<---	[paper towels,dinner rolls]
rule16	0.087	0.566	1.342	poultry	<---	[lunch meat,mixes]
rule17	0.087	0.559	1.435	eggs	<---	[beef,soda]
rule18	0.087	0.562	1.446	dinner rolls	<---	[poultry,hand soap]
rule19	0.087	0.566	1.342	poultry	<---	[dinner rolls,hand soap]
rule20	0.088	0.556	1.422	cheeses	<---	[cereals,sandwich bags]
rule21	0.088	0.565	1.341	poultry	<---	[dinner rolls,all- purpose]
rule22	0.088	0.565	1.341	poultry	<---	[beef,sugar]
rule23	0.088	0.562	1.333	poultry	<---	[dinner rolls,milk]
rule24	0.089	0.564	1.339	poultry	<---	[dinner rolls,juice]
rule25	0.089	0.558	1.324	poultry	<---	[dinner rolls,coffee/tea]
rule26	0.09	0.554	1.315	poultry	<---	[lunch meat,sugar]
rule27	0.09	0.557	1.321	poultry	<---	[dinner rolls,mixes]
rule28	0.091	0.562	1.334	poultry	<---	[dinner rolls,lunch meat]
rule29	0.092	0.553	1.414	soda	<---	[eggs,soap]
rule30	0.095	0.554	1.421	eggs	<---	[dinner rolls,soda]
rule31	0.096	0.556	1.32	poultry	<---	[juice,aluminum foil]
rule32	0.099	0.579	1.49	dinner rolls	<---	[spaghetti sauce,poultry]
rule33	0.099	0.577	1.368	poultry	<---	[dinner rolls,spaghetti sauce]

Associations identified (continue)

Insights:

- Sandwich bags, sugar recommends soda with lift of 1.416, confidence of 0.553 and support of 0.073.
- Sandwich loaves, laundry detergent recommends poultry with lift of 1.388, confidence of 0.585 and support of 0.076.
- Mixes, sugar recommends poultry with lift of 1.36, confidence of 0.573 and support of 0.079.

Support: Support of one set mean, its probability of its occurrence in the given set of transactions.

Confidence: Confidence is nothing but the conditional probability that the customer would end up buying a product given that they already bought a set of product or products..

Lift: Lift means confidence divided by support of product that the customer ended up buying.

Suggestions and Recommendations

- Company can provide combo offers on dinner rolls, spaghetti sauce and poultry.
- Company can also provide combo offers on eggs, soap and soda.
- Buy two get one free offers can be introduced on products hand soap, flour and sandwich loaves, as product counts for these products are less when compared with others.
- Customers who buy dinner rolls, pork can be given an additional discount of 10% on poultry.
- Dinner rolls, milk, juice, poultry can be placed strategically in order to attract customers.
- Company can provide discount on cheeses when customer buys cereals and sandwich bags.
- Can provide combo offers on Juice, aluminium foil and poultry.

Recommendations:

- Company can increase prices of waffles, lunch meat, poultry by 2-3% and can introduce additional discount offers of same percentage on any two days in weekdays.
- Prices of poultry can be increased slightly i.e., by 1-2%, as more number of customers end up buying the product.
- Company can introduce Monday discount offers up to 10% on products like coffee/tea, sandwich loaves, as there is a dip in purchase during Mondays.
- Avoid keeping pork and beef together because of religious sentiments.
- Company can strategize and place poultry in a different place, as the product have good count across years and most of the customers come in search of the product.
- Company should place products such as dishwashing liquid/detergent, laundry detergent, hand soaps together.

THANK YOU