MRA PROJECT ML 2

USER NAME : DINYA ANTONY

BATCH : PGPDSBA ONLINE AUG_A 2021

EMAIL : VOLLAFREEBIRD@GMAIL.COM

TABLE OF CONTENTS

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 menu items in the customer orders. The Store doesn't have any combo meals. Can you suggest the best combo meals?

ANALYSING THE DATA

❖ *Data information:*

RangeIndex: 20641 entries, 0 to 20640

Data columns (total 3 columns):

#	Column	Non-Null Count	Dtype
---	------	----------	-----
0	Date	20641 non-null	int64
1	Order_id	20641 non-null	int64
2	Product	20641 non-null	float64

dtypes: int64(1), object(2)

memory usage: 483.9+ KB

ANALYSING THE DATA

❖ **Data head:**

Date	Order_id	Product
01-01-2018	1	yogurt
01-01-2018	1	pork
01-01-2018	1	sandwich bags
01-01-2018	1	lunch meat
01-01-2018	1	all- purpose

ANALYSING THE DATA

- *Data shape: (20641, 3)*
- Describe the data:

	count	unique	top	freq	mean	std	min	25.00%	50.00%	75.00%	max
Date	20641	603	08-02-2019	183	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Order_id	20641	NaN	NaN	NaN	575.986289	328.557078	1	292	581	862	1139
Product	20641	37	poultry	640	NaN	NaN	NaN	NaN	NaN	NaN	NaN

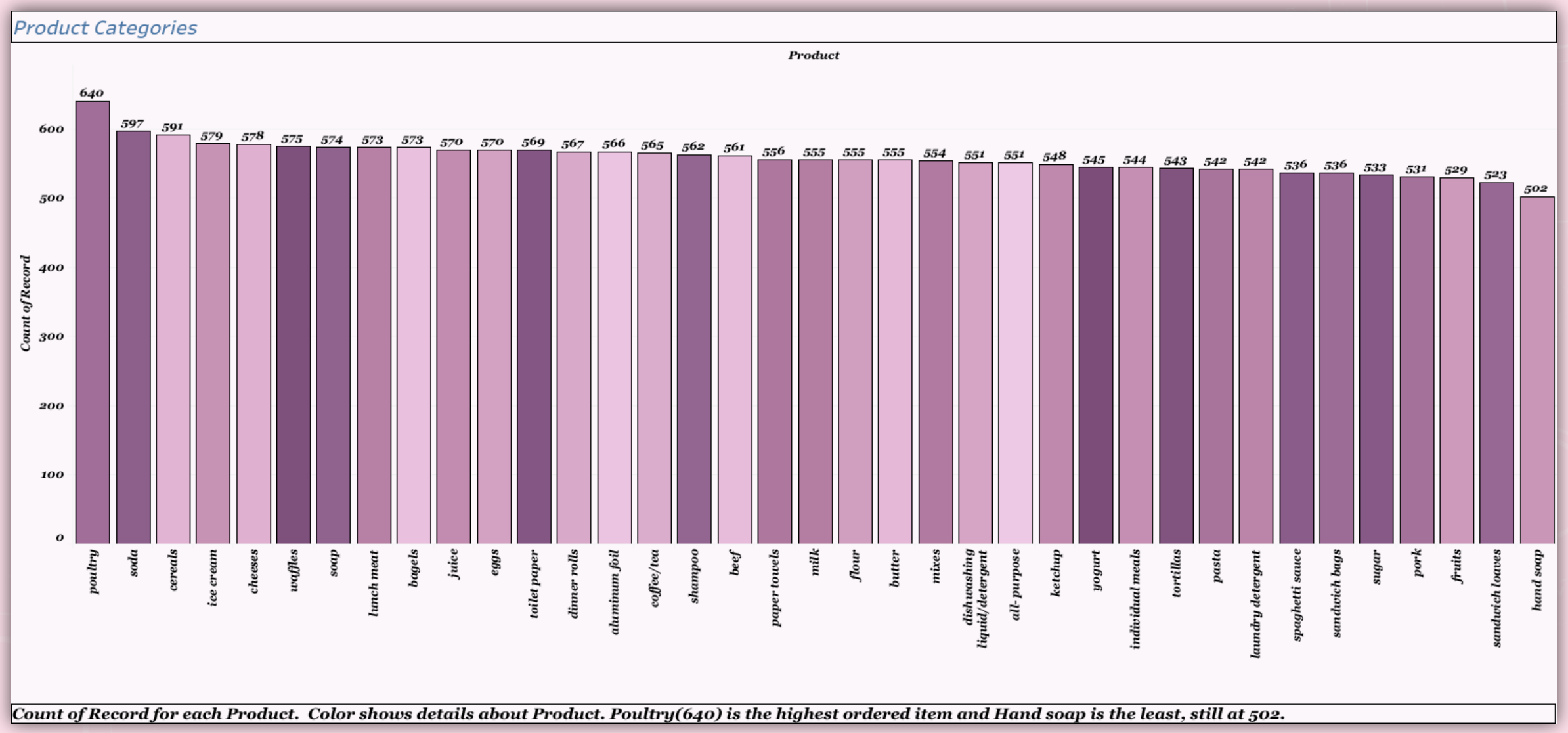
ANALYSING THE DATA

❖ **Interpretation:**

- *The data has 20641 rows and 3 columns with int and object as the data type.*
- *We have no non-null data with 3 variables. Ordinal 1 variables, 1 date-time and 1 object types.*
- *The summary stats: 37 unique items in the dataset.*
- *The orders date with most orders is 08-02-2019.*
- *The most orders item is: Poultry.*

EXPLORATORY DATA ANALYSIS

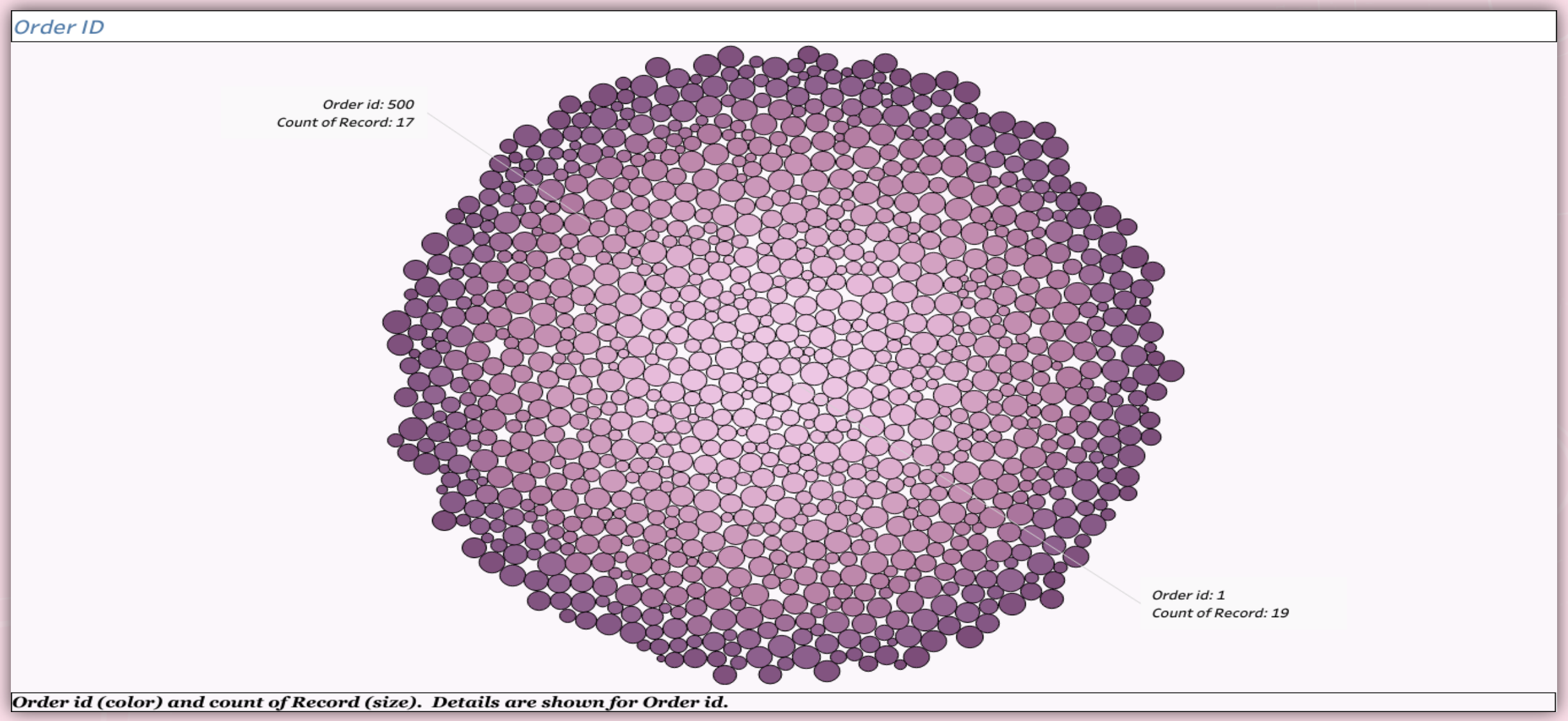
EDA - Univariate Analysis: Product Categories



Count of Record for each Product. Color shows details about Product. Poultry(640) is the highest ordered item and Hand soap is the least, still at 502.

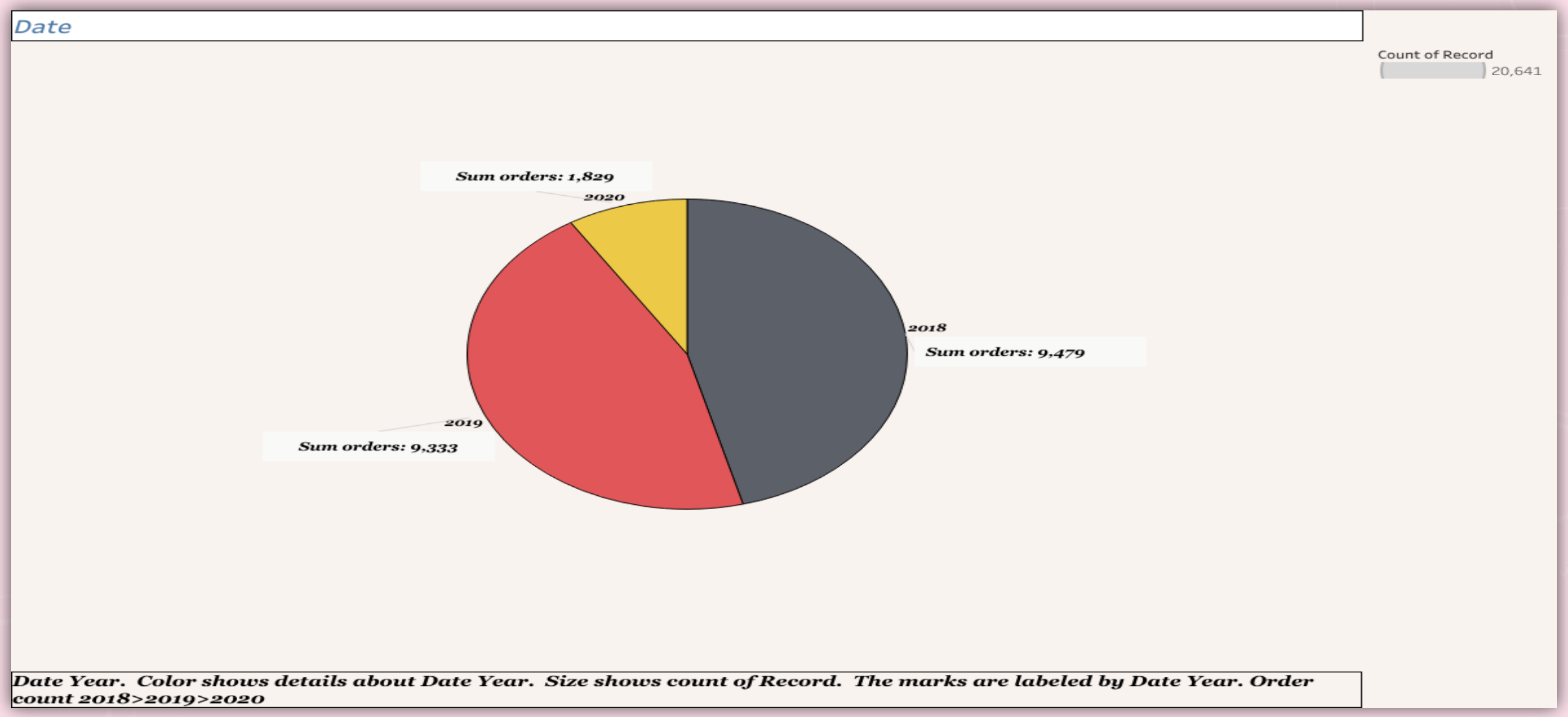
EXPLORATORY DATA ANALYSIS

Univariate Analysis: Order ID



EXPLORATORY DATA ANALYSIS

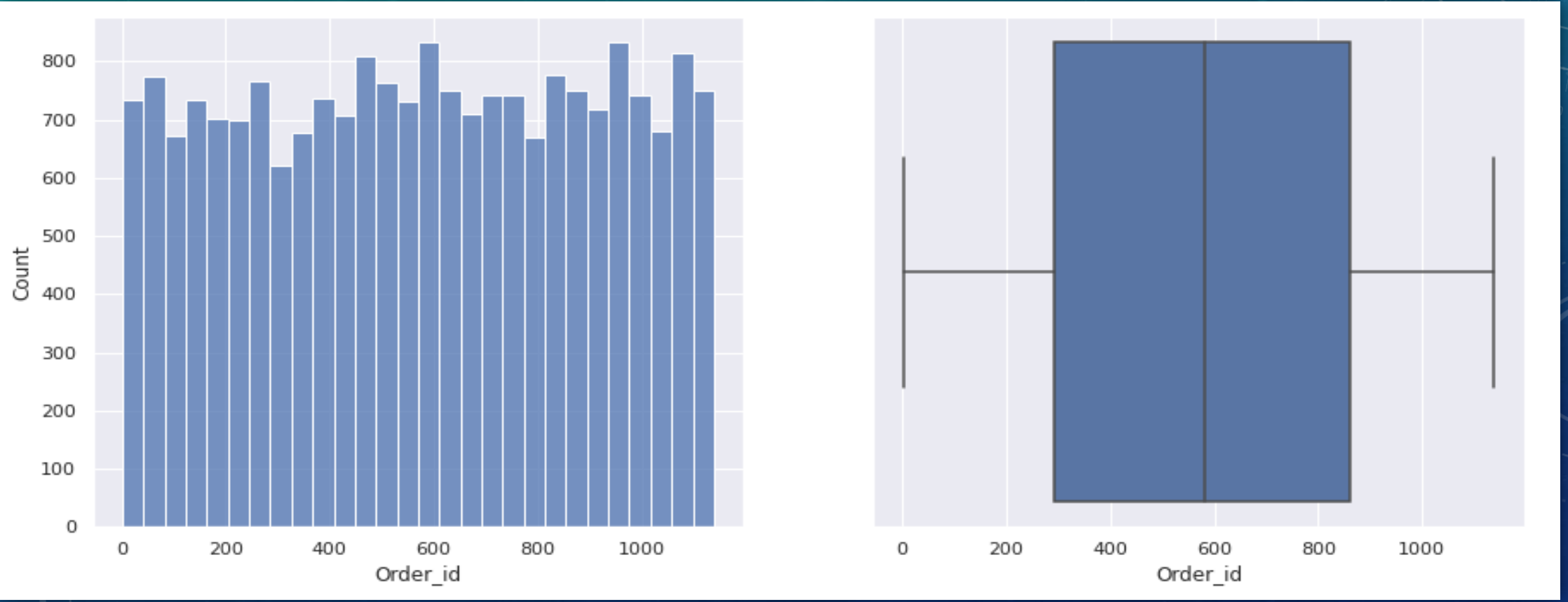
Univariate Analysis: Date



EXPLORATORY DATA ANALYSIS

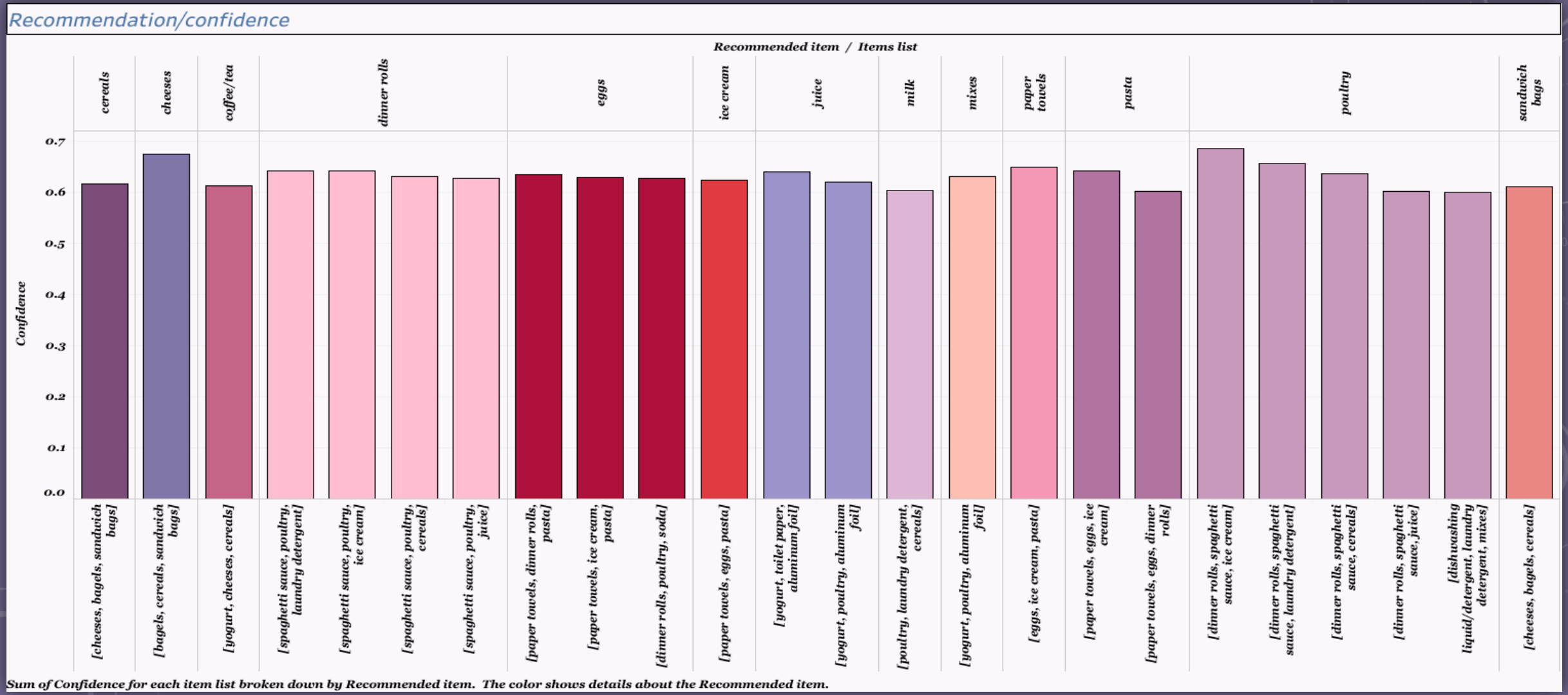
Univariate Analysis: Order ID

- *The Boxplot tells us there are no outliers.*
- *The distplot distribution can be said to be a normal distribution. The distribution ranges mostly between 1 to 1139.*



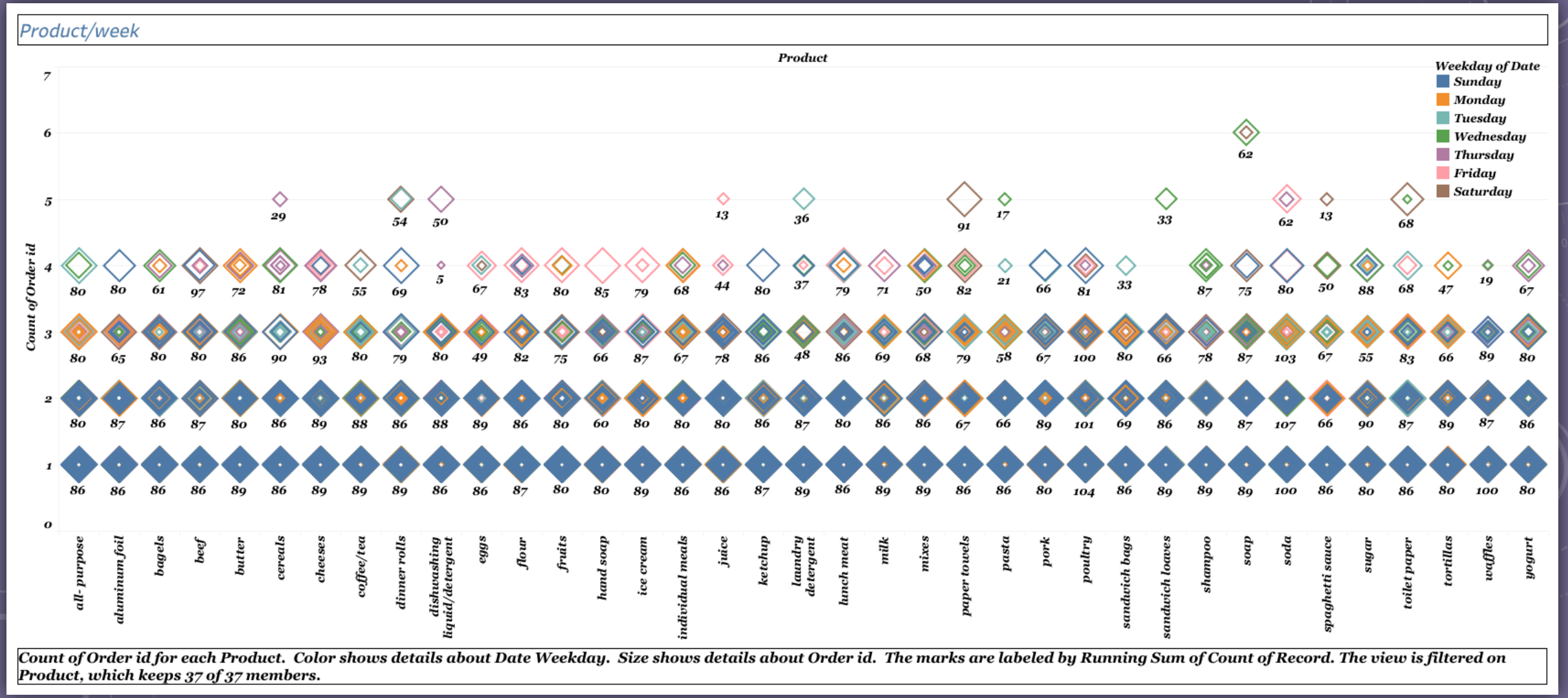
EXPLORATORY DATA ANALYSIS

Bivariate Analysis: Product recommendation



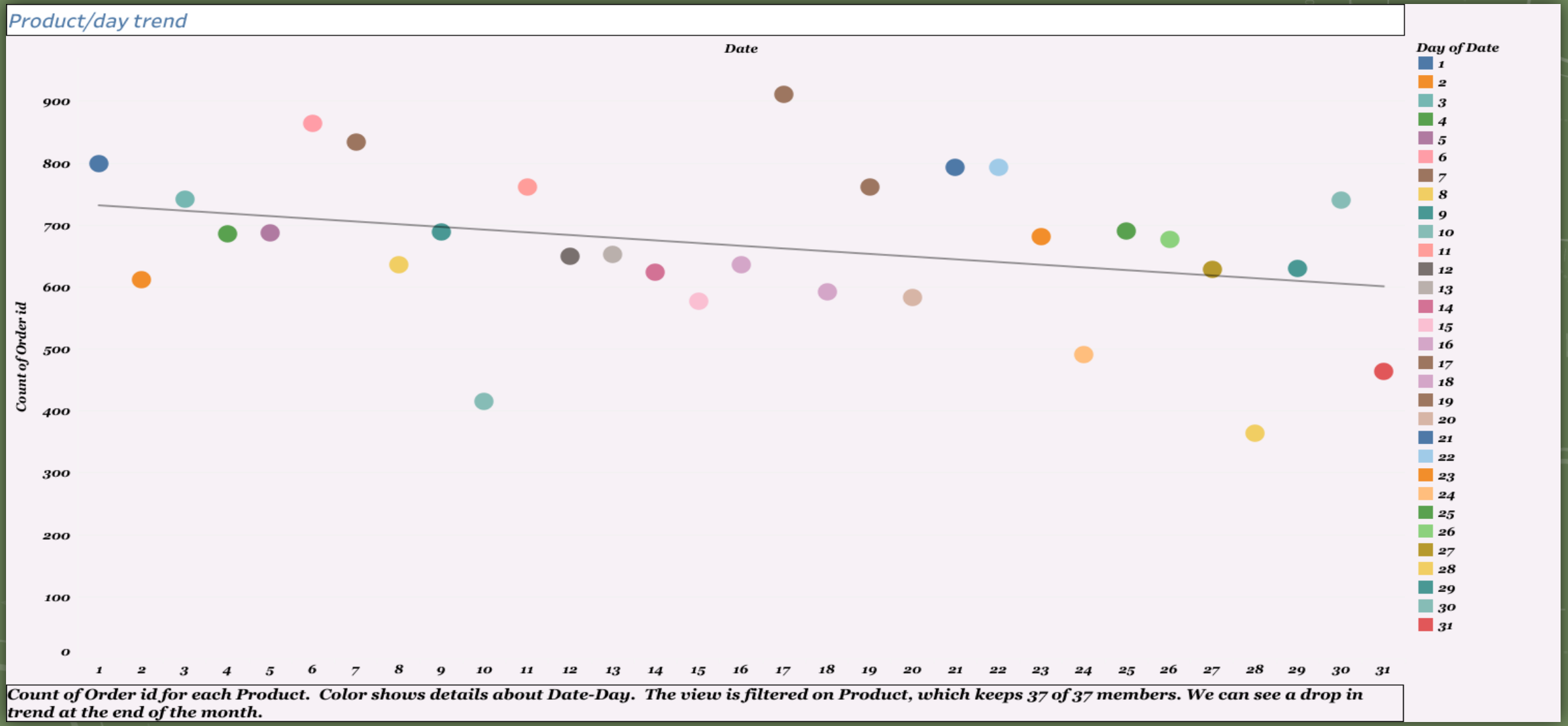
EXPLORATORY DATA ANALYSIS

Multivariate Analysis: Product/orderID running total



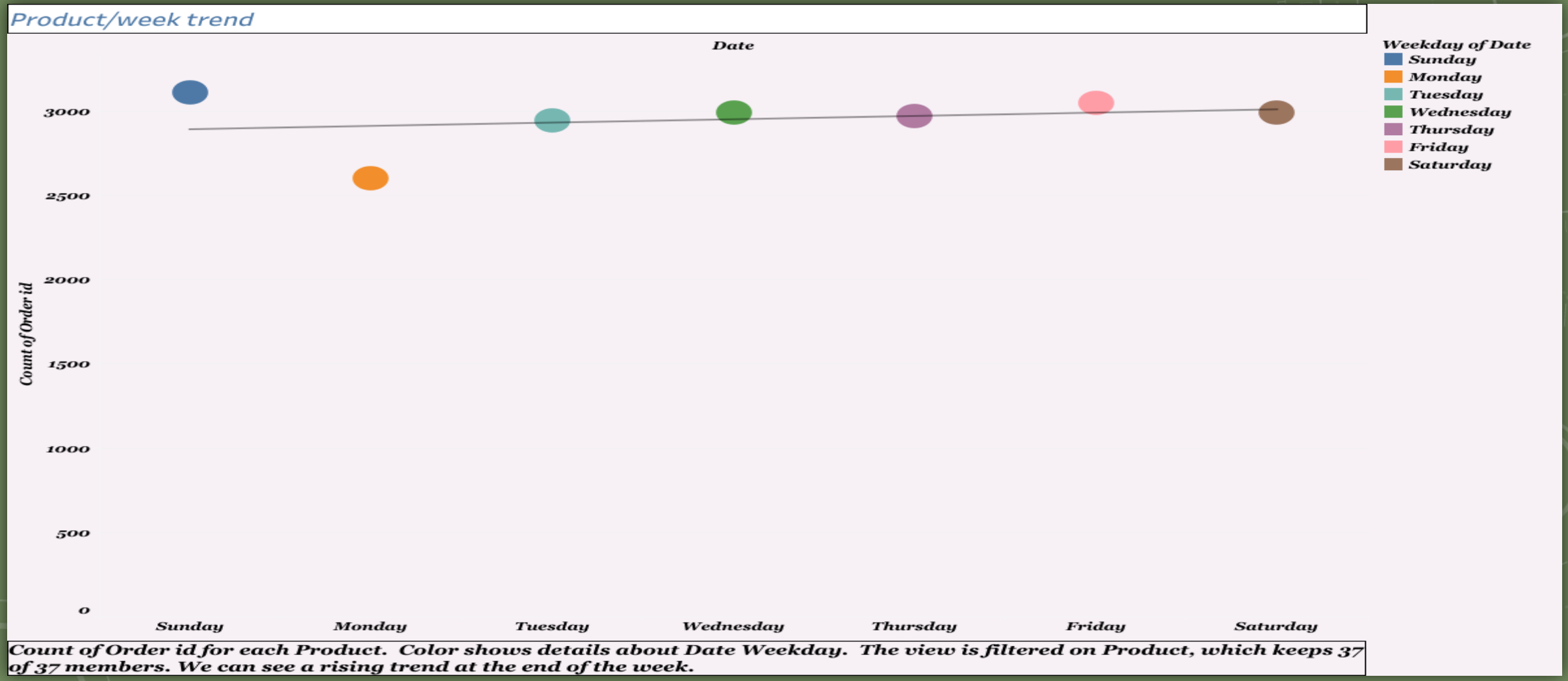
SALES TRENDS

❖ **Days:**



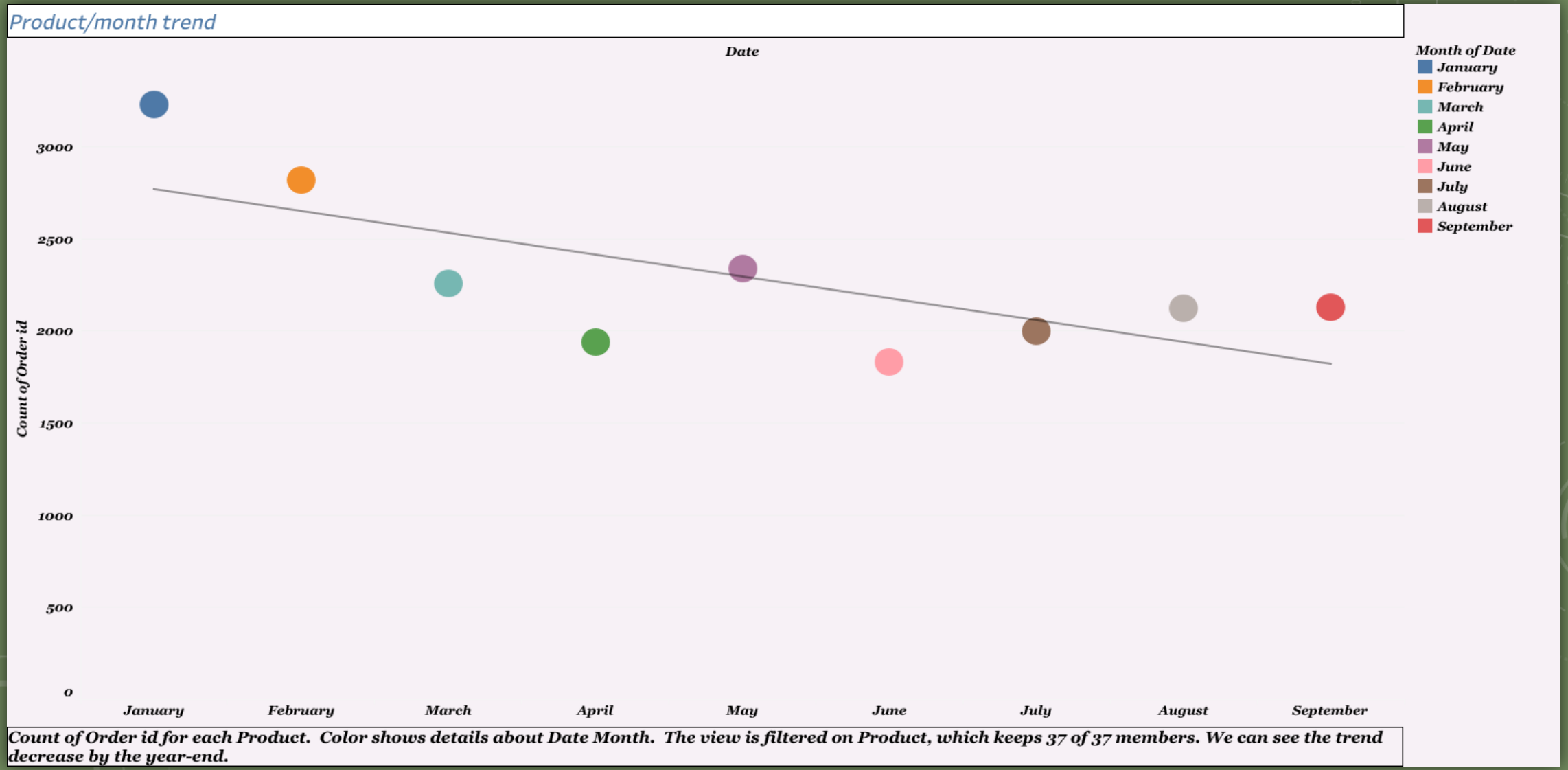
SALES TRENDS

- **Weekly:**



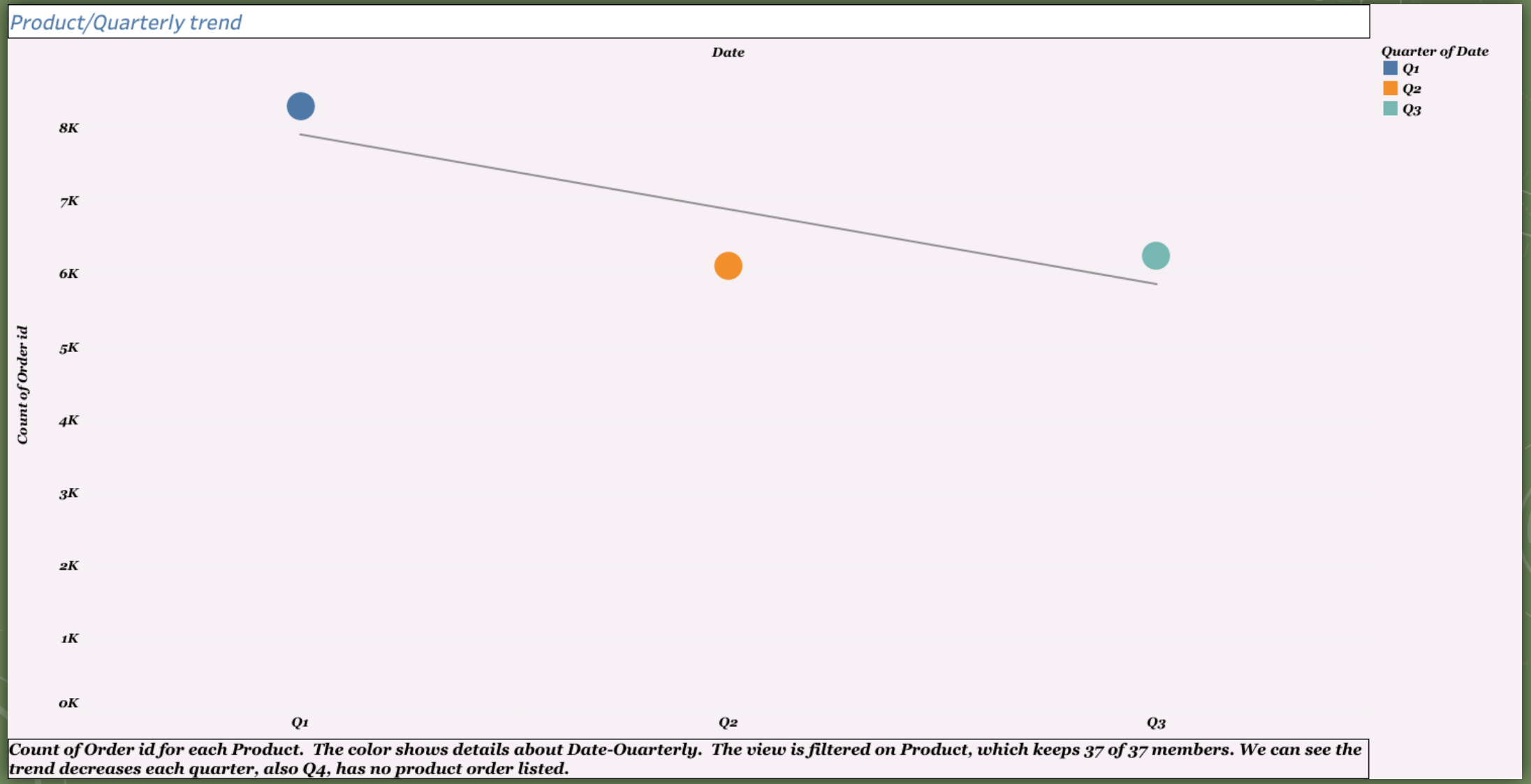
SALES TRENDS

❖ **Monthly:**



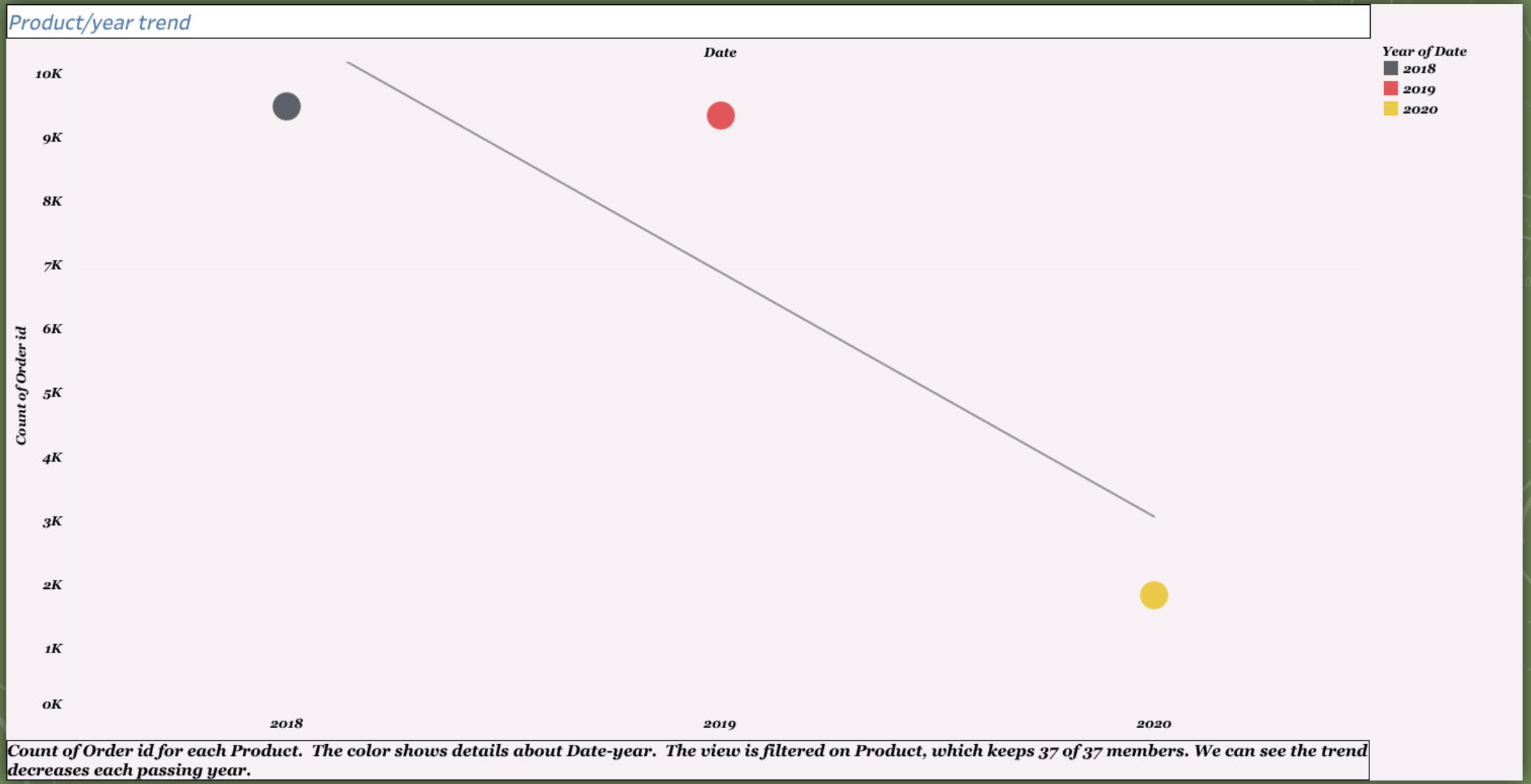
SALES TRENDS

- ***Quarterly***:



SALES TRENDS

- ***Yearly***:



EDA AND TRENDS

❖ **Interpretation:**

- ***The orders most received are for Poultry and the least is for hand soap. Though there isn't much difference in the total. The orders of medium size are received mostly.***
- ***The store receives order Id varies from 1-500. 34 is the highest item list on orders and 3 is the least.***
- ***The store is facing a Deline in order on 2020. Compared to the last 2 years.***
- ***The end of the week we see a raise in trend.***
- ***Same cannot be said for the end on the month, there is a decrease.***
- ***There no record are there for Q4 and Q1> Q3> Q2 we see the decrease.***

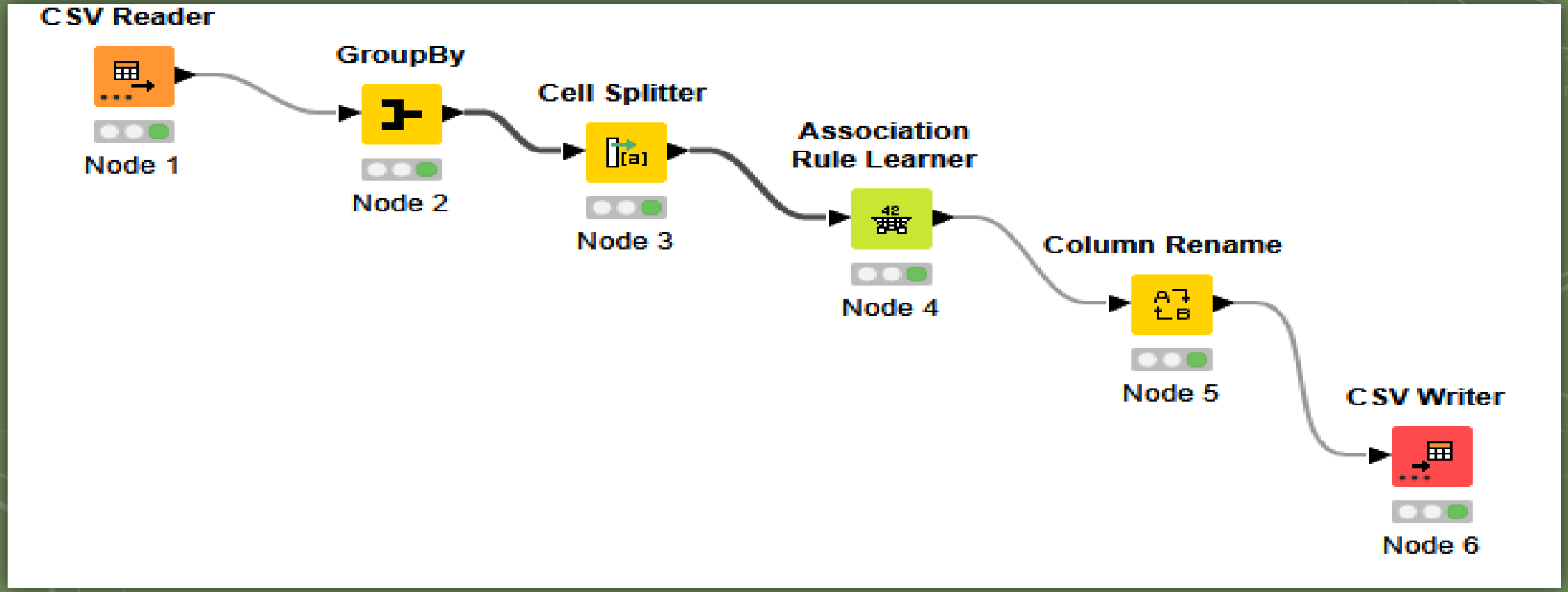
MBA

❖ What is MBA?

- Market Basket Analysis is *a data mining technique used by retailers to increase sales by better understanding customer purchasing patterns.*
- *It involves analysing large data sets, such as purchase history, to reveal product groupings, as well as products that are likely to be purchased together.*
- Association Rules*: are widely used to analyse retail basket, association between different objects in a set, find frequent patterns in a transaction database.*
- *For our Grocery Store problem, this will help us identify combo and best suggestion to improve sales.*
- *The Apriori algorithm is commonly used in research articles about market basket analysis. There are three components in Apriori algorithm:*
- *Support: It is the ratio of transactions involving number of transaction of the item by the total number of transactions made.*
- *Confidence: It is whether the product sales are popular on individual sales or through combined sales has been calculated.*
- *Lift: It calculated for knowing the ratio for the sales. (confidence percent/ support percent).*

MBA

- *I have used Python and Tableau for data read and EDA*
- *In this project I have used, KNIME is used to perform the MBA and I have used support = 0.05 and confidence =0.60 for this dataset.*
- *Here is the workflow diagram:*



MBA OUTPUT

❖ *Association Rule output: 15*

Rules	Support	Confidence	Lift	Recommended_item	Recommended_with	Items_list
Rule 1	0.050043898156278	0.640449438202247	1.70040072287263	juice	<---	[yogurt, toilet paper, aluminum foil]
Rule 2	0.050043898156278	0.619565217391304	1.64495287321374	juice	<---	[yogurt, poultry, aluminum foil]
Rule 3	0.050043898156278	0.612903225806452	1.61596475507766	coffee/tea	<---	[yogurt, cheeses, cereals]
Rule 4	0.050043898156278	0.6	1.42375	poultry	<---	[dishwashing liquid/detergent, laundry detergent, mixes]
Rule 5	0.050921861281826	0.630434782608696	1.67772247054043	mixes	<---	[yogurt, poultry, aluminum foil]
Rule 6	0.050921861281826	0.610526315789474	1.65964074864967	sandwich bags	<---	[cheeses, bagels, cereals]
Rule 7	0.050921861281826	0.674418604651163	1.72620851842174	cheeses	<---	[bagels, cereals, sandwich bags]
Rule 8	0.050921861281826	0.617021276595745	1.55828654998349	cereals	<---	[cheeses, bagels, sandwich bags]
Rule 9	0.050921861281826	0.630434782608696	1.62091471194425	dinner rolls	<---	[spaghetti sauce, poultry, cereals]
rule 10	0.050921861281826	0.637362637362637	1.51240842490842	poultry	<---	[dinner rolls, spaghetti sauce, cereals]
Rule 11	0.050921861281826	0.604166666666667	1.58925134719015	milk	<---	[poultry, laundry detergent, cereals]
Rule 12	0.051799824407375	0.627659574468085	1.61014471918727	eggs	<---	[dinner rolls, poultry, soda]
Rule 13	0.051799824407375	0.641304347826087	1.64886615173226	dinner rolls	<---	[spaghetti sauce, poultry, ice cream]
Rule 14	0.051799824407375	0.686046511627907	1.62793120155039	poultry	<---	[dinner rolls, spaghetti sauce, ice cream]
Rule 15	0.051799824407375	0.627659574468085	1.61377935737957	dinner rolls	<---	[spaghetti sauce, poultry, juice]

MBA- ASSOCIATIONS IDENTIFIED

- *Support : It's popularity of an item. In mathematical terms, the support of item A is the ratio of transactions of A to the total number of transactions.*
- *The higher support the item is more likely to be ordered.*
- *Confidence : Probability that customer who bought both A and B. It is the ratio of the number of transactions of A and B by the number of transactions involving B.*
- *Confidence(A => B) = Support(A, B)/Support(A)*
- *The higher the probability more likely the combo will works.*
- *Lift : One product effect the sale of other in a positive direction. Increase in the sale of A when you sell B.*
- *Lift(A => B) = Confidence(A, B)/Support(B)*
- *Lift (A => B) = 1 means that there is no correlation within the itemset.*
- *Lift (A => B) > 1 means that there is a positive correlation within the itemset, i.e., products in the itemset, A, and B, are more likely to be bought together.*
- *Lift (A => B) < 1 means that there is a negative correlation within the itemset, i.e., products in itemset, A, and B, are unlikely to be bought together.*
- *Mostly like we would want Lift to be higher than 1.*
- *Hence I have chosen the minimum support = 0.055 at the highest possible confidence = 0.60 for this dataset.*

INTERPRETATION

Recommendations:

- At support: 0.5 and confidence: 0.60, our 15 rules lift values are above 1. i.e. there is a positive correlation within the itemset.
- The rules have most set to have poultry as recommend item. It also being the most sold item, creating combo with poultry will be highly beneficial.
- The introductory offer for new customer with point rewarding scheme for next purchase.
- The order drop during Q4 and year end, end of season sales will be profitable during this time.
- Rise is sales are seen at the year start, discounts or limited time period event price that would be helpful during end of the year or weekdays to boost sales.
- Sandwich bag placed between Cheeses and Bagels. Would increase Sandwich bag sale.

Possible Combos with Lucrative Offers:

- Giving buy one get one free for paper towel with eggs, ice cream.
- 15% off on poultry purchased with dinner rolls.
- 10% off on hand soap with poultry purchase.
- Loyalty points to redeem at the end of the month/year regular customer. Would import retention rate.
- Free spaghetti Sause with Dinner rolls+poultry+eggs combo.