



MRA PROJECT

- BARKHA AGARWAL

TABLE OF CONTENTS

- ❖ Problem statement.
- ❖ Analyzing the data.
 - Info, Shape, Describe, datatype, null values
- ❖ EDA
 - Univariate, Bivariate, and multivariate analysis.
- ❖ Sales Trends
 - Weekly, Monthly, Quarterly, Yearly
- ❖ Market Basket Analysis
- ❖ Associations Identified
 - Support, confidence, & lift values are calculated
- ❖ Inferences and recommendations
 - Possible Combos with Lucrative Offers.

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:

```
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
```

❖ Data head:

```
]:
```

	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

❖ Data shape: (20641, 3)

❖ Describe the data:

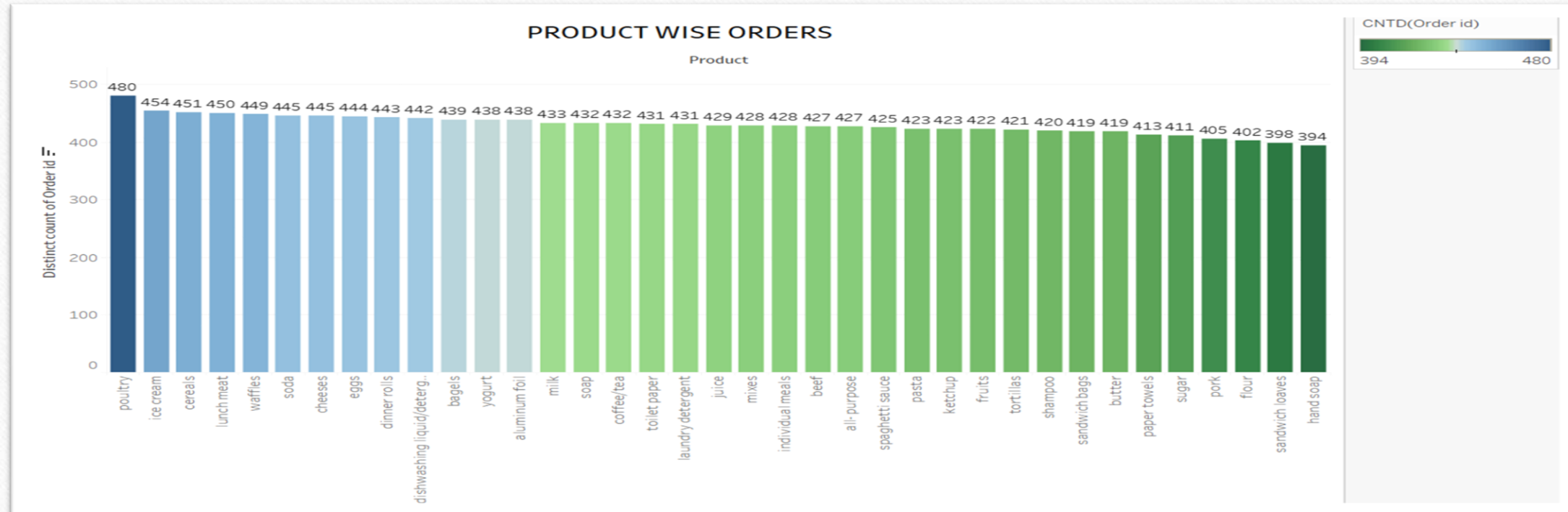
	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Date	20641	603	08-02-2019	183	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Order_id	20641.0	NaN	NaN	NaN	575.986289	328.557078	1.0	292.0	581.0	862.0	1139.0
Product	20641	37	poultry	640	NaN	NaN	NaN	NaN	NaN	NaN	NaN

❖ 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.
- There are 4730 duplicate values.
- The summary stats: 37 unique items in the dataset.
- The orders date with most orders is 08-02-2019.
- The most order item is: Poultry .

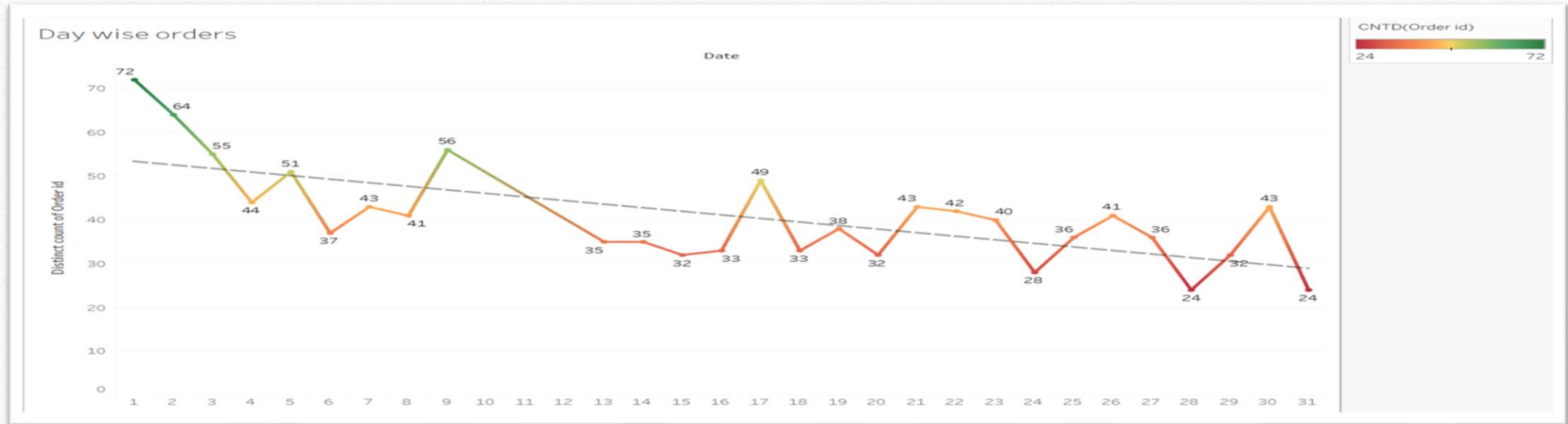
EXPLORATORY DATA ANALYSIS

- Univariate Analysis: Product Categories



- It is evident that Poultry has been ordered the highest with 480 followed by ice cream with 454 orders.
- The lowest is hand soap 394 orders and sandwich loaves with 398 orders.
- The milk, soap, coffee/tea, soda, cheese are more or less holds the same amount of orders.

Daily Wise Quantity ordered



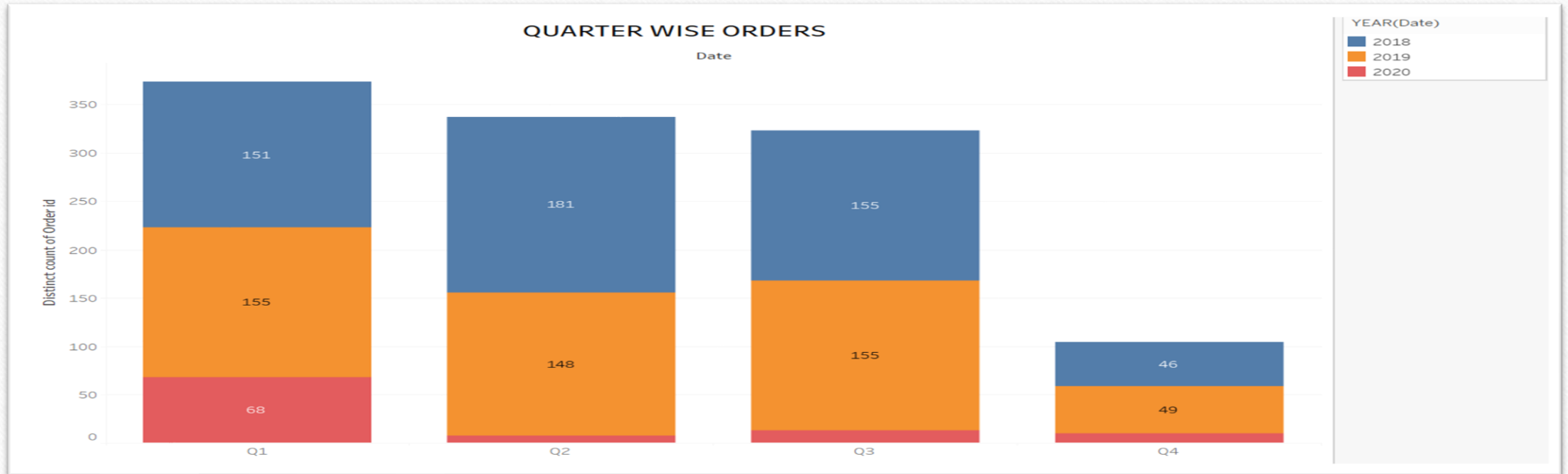
- The highest distinct count of ordered date is 1.
- No proper trend can be analyzed using daily wise orders.

Monthly orders



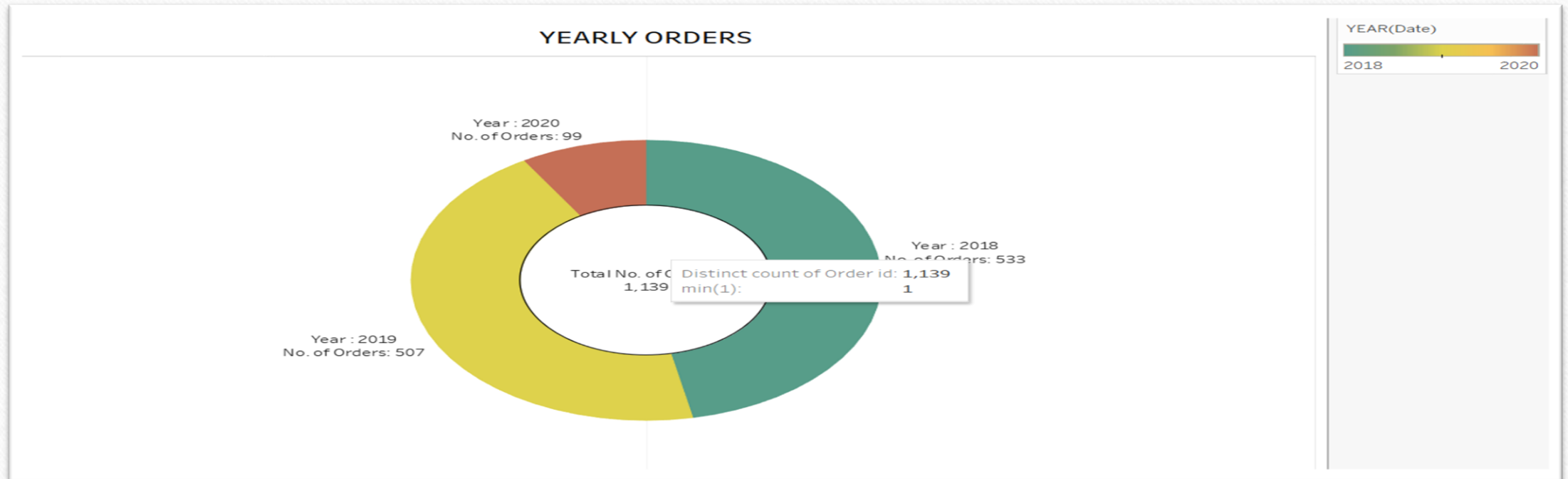
- January, February, May, June and July month has the highest numbers of sales.

Quarterly orders



- In 2018, Q2 had the highest number of orders with 181.
- In 2019, Q1 and Q3 had the highest with 155.
- This started decreasing as quarters passed by. The lowest was recorded in Q2 of 2020.
- Also, it is evident that Q4 of every year doesn't have good sales.

Yearly Orders



- The total number of orders are highest in the year 2018.
- Followed by 507 orders in 2019.
- 2020 has registered only 99 orders. As only two months data is given in the data set , this might be a reason for low order count.
- Trends shows that the number of orders placed has been decreasing yearly.
- Also it shows a downward trend and no forecast has been generated with this data.

• The

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 Decline 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.

Market Basket Analysis

❖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 analyzing 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 analyze 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

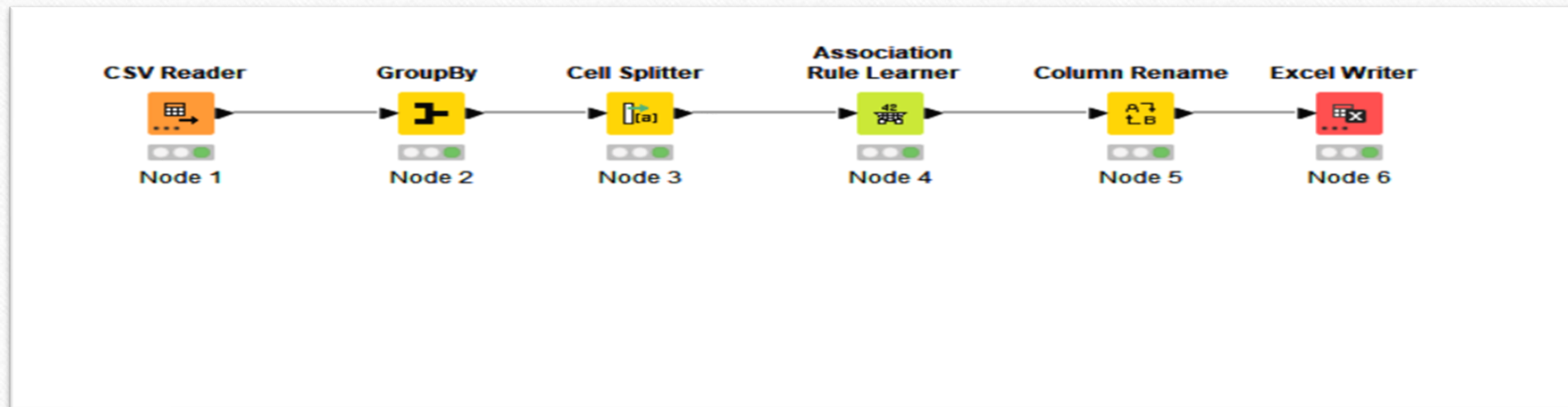
❖ I have used Python and Tableau for data read and EDA

Tableau public link-

[BarkhaAgarwal_MRA2 | Tableau Public](#)

❖ 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

Table "default" - Rows: 24 Spec - Columns: 6 Properties Flow Variables

Row ID	D Support	D Confide...	D Lift	S Recommended_item	S Recommended_with	S Items_list
rule0	0.05	0.64	1.7	juice	<---	[yogurt, toilet paper, aluminum foil]
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, laundry detergent, mixes]
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, cereals]
rule10	0.051	0.604	1.589	milk	<---	[poultry, laundry detergent, cereals]
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 cream]
rule13	0.052	0.686	1.628	poultry	<---	[dinner rolls, spaghetti sauce, ice cream]
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, juice]
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, laundry detergent]
rule19	0.054	0.656	1.556	poultry	<---	[dinner rolls, spaghetti sauce, laundry detergent]
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]

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.
- $\text{Confidence}(A \Rightarrow B) = \text{Support}(A, B) / \text{Support}(A)$
- The higher the probability more likely the combo will work.
- Lift : One product effect the sale of other in a positive direction. Increase in the sale of A when you sell B.
- $\text{Lift}(A \Rightarrow B) = \text{Confidence}(A, B) / \text{Support}(B)$
- $\text{Lift}(A \Rightarrow B) = 1$ means that there is no correlation within the itemset.
- $\text{Lift}(A \Rightarrow B) > 1$ 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.
- $\text{Lift}(A \Rightarrow B) < 1$ 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 = 0.05 at the highest possible confidence = 0.60 for this dataset.

INTERPRETATION

Recommendations:

- At support: 0.5 and confidence: 0.60, our 23 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 in 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.
- Most of the poultry purchased is with dinner rolls.
- Loyalty points to redeem at the end of the month/year regular customer would improve retention rate.
- In order to increase the sales, the grocery store can provide combo offers to its customers.



THANK YOU
