

RETAIL SALES ANALYSIS

PROBLEM :

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.

DATA DESCRIPTION

- ❖ The dataset has 20641 entries and 3 variables.
- ❖ It has 1 numerical variable and 2 categorical variable.
- ❖ Date datatype has to be changed from object to datetime64(ns).
- ❖ So finally we have 1 datetime64(ns) , 1 numerical variable and 1 categorical variable.
- ❖ No null values in the given dataset
- ❖ 4730 duplicates present in the given dataset and can be removed .
Finally we got 15911 rows and 3 columns to proceed.
- ❖ No Outliers present in the dataset.
- ❖ Dataset provide orders placed from 01-01-2018 to 26-02-2020.
- ❖ **Market Basket Analysis** using association rules was performed to identify the relationships between the products purchased by the customers.It helped to identify the products that are frequently purchased together.

DATA DICTIONARY

DATE	DATE IN WHICH THE PRODUCT SOLD
ORDER_ID	ID FOR ORDERS PLACED
PRODUCT	PRODUCTS SOLD

DATA DESCRIPTION

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

Tail

	Date	Order_id	Product
20636	25-02-2020	1138	soda
20637	25-02-2020	1138	paper towels
20638	26-02-2020	1139	soda
20639	26-02-2020	1139	laundry detergent
20640	26-02-2020	1139	shampoo

shape

(20641, 3)

Null values

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

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  datetime64[ns]
1   Order_id    20641 non-null  int64
2   Product     20641 non-null  object
dtypes: datetime64[ns](1), int64(1), object(1)
memory usage: 483.9+ KB
```

After removing duplicates

(15911, 3)

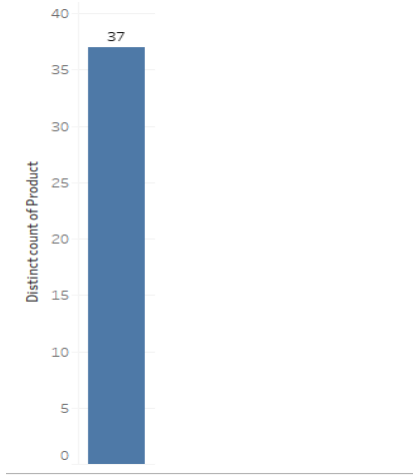
4730 Dulpicates

	Date	Order_id	Product
10	2018-01-01	1	all- purpose
13	2018-01-01	1	all- purpose
18	2018-01-01	1	dinner rolls
29	2018-01-01	2	waffles
31	2018-01-01	2	hand soap
...
20616	2020-02-24	1137	paper towels
20632	2020-02-25	1138	sandwich bags
20633	2020-02-25	1138	toilet paper
20635	2020-02-25	1138	soda
20636	2020-02-25	1138	soda

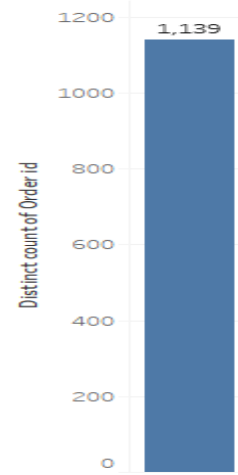
4730 rows × 3 columns

EDA

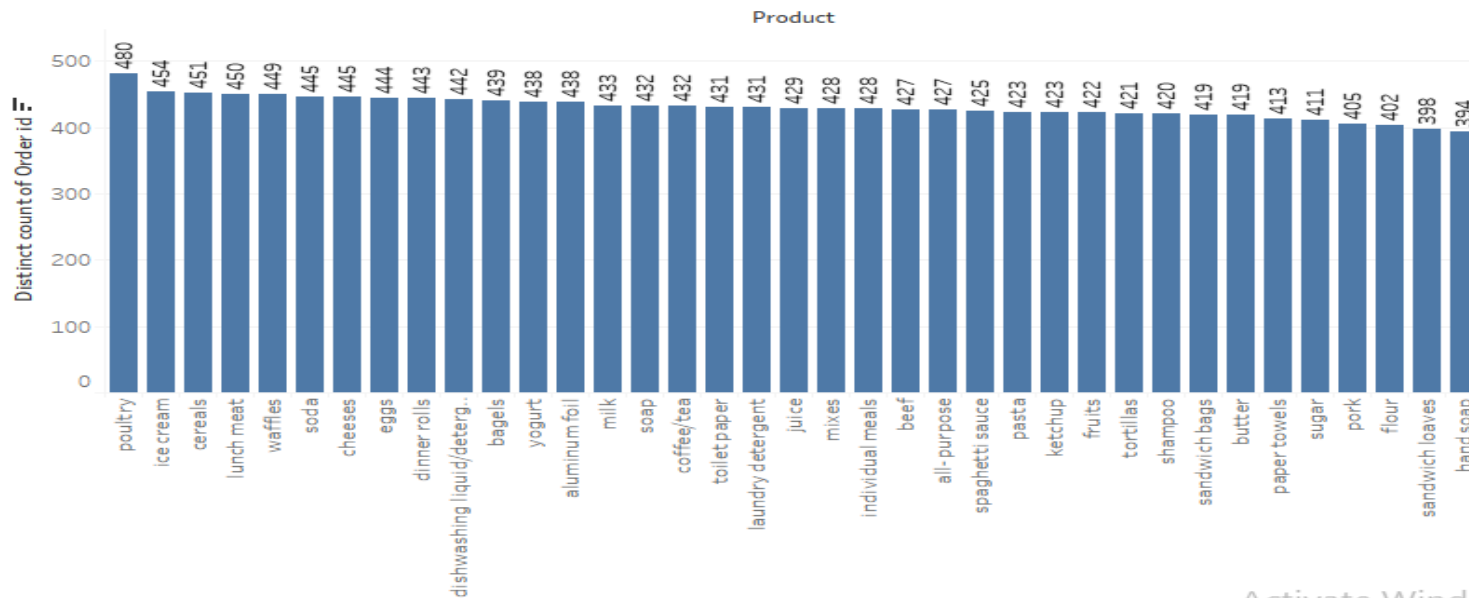
TOTAL PRODUCTS IN THE STORE



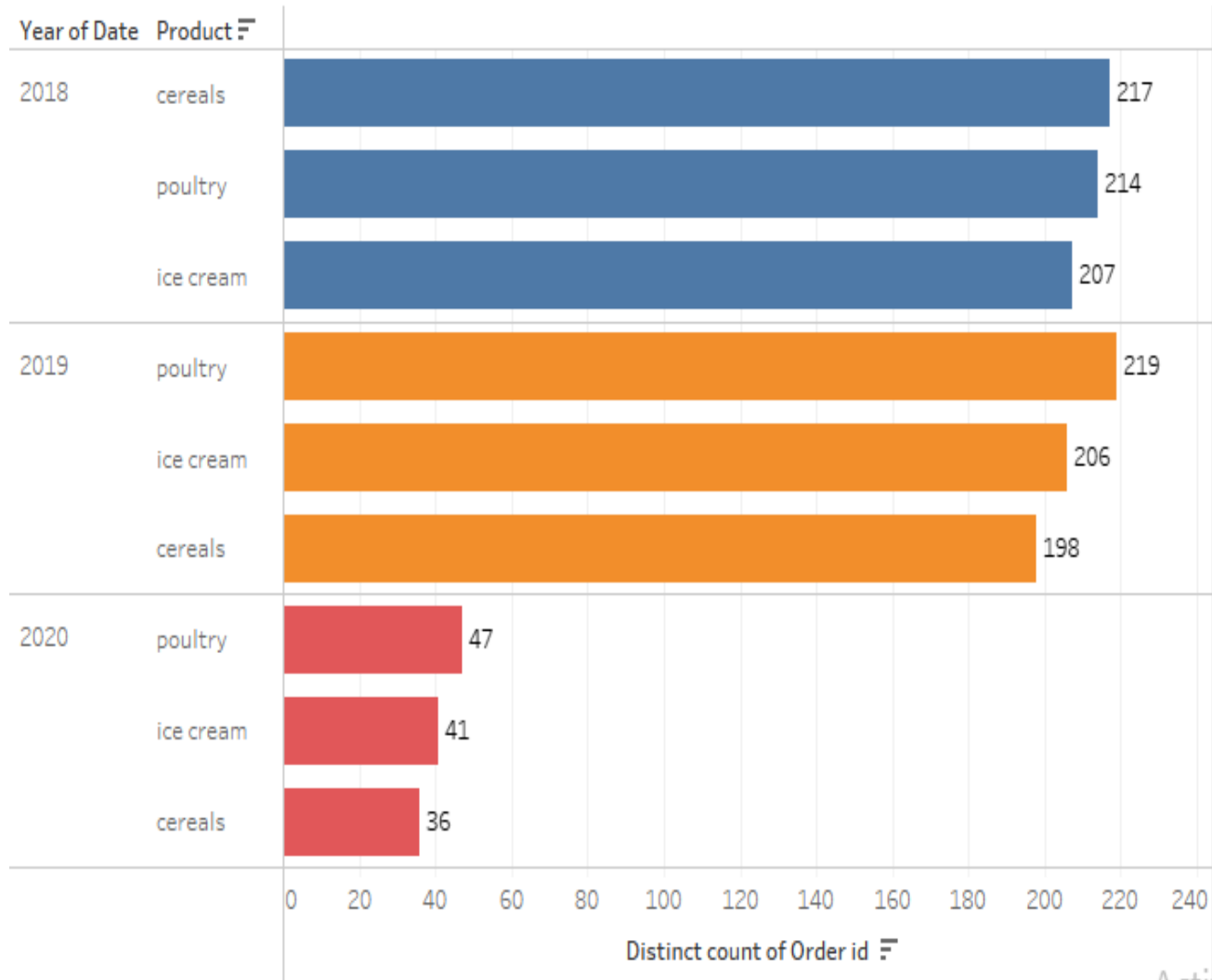
DISTINCT ORDER_ID



PRODUCTS SOLD



TOP 3 PRODUCT SALES IN 3 YEAR



YEAR(Date)

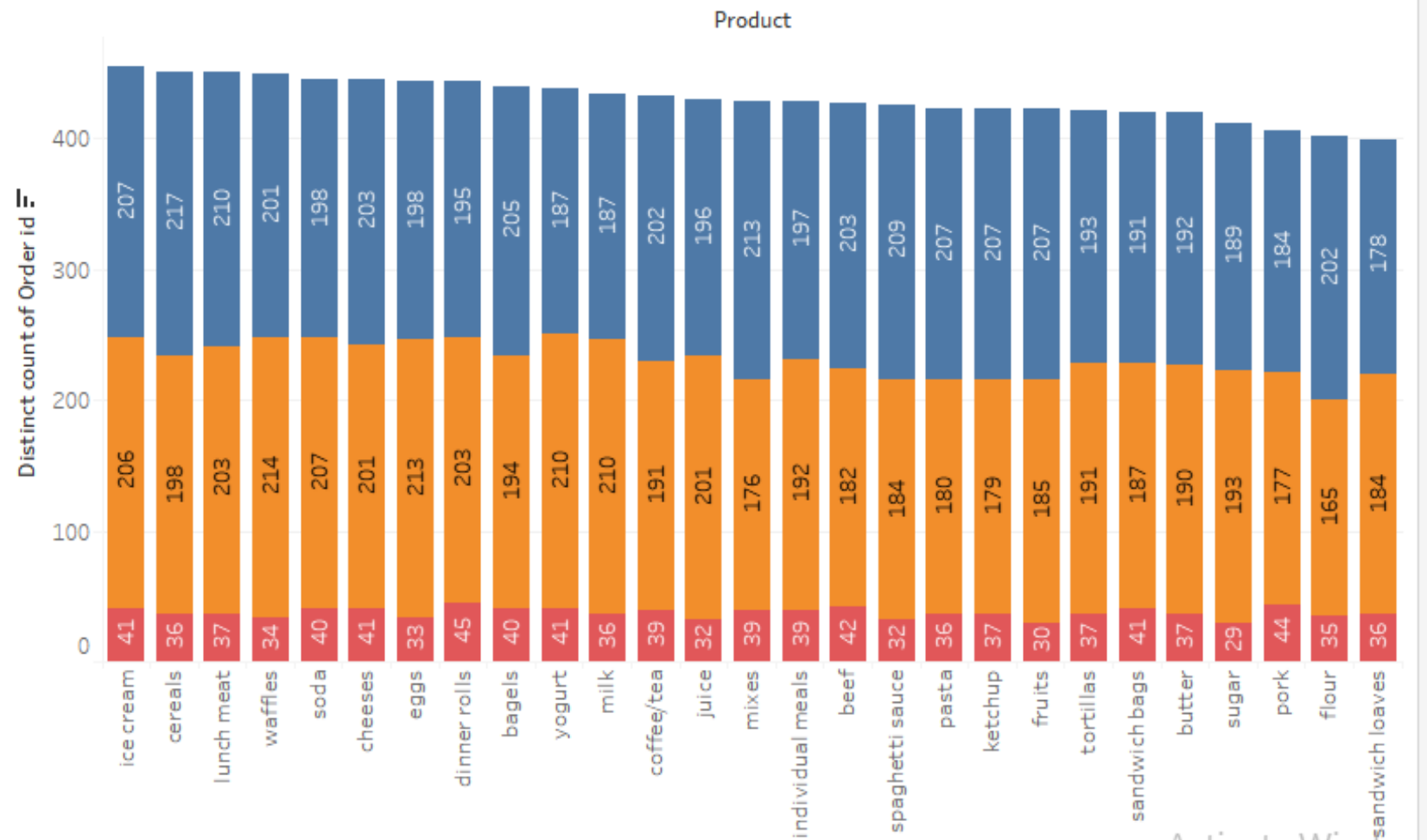
2018

2019

2020

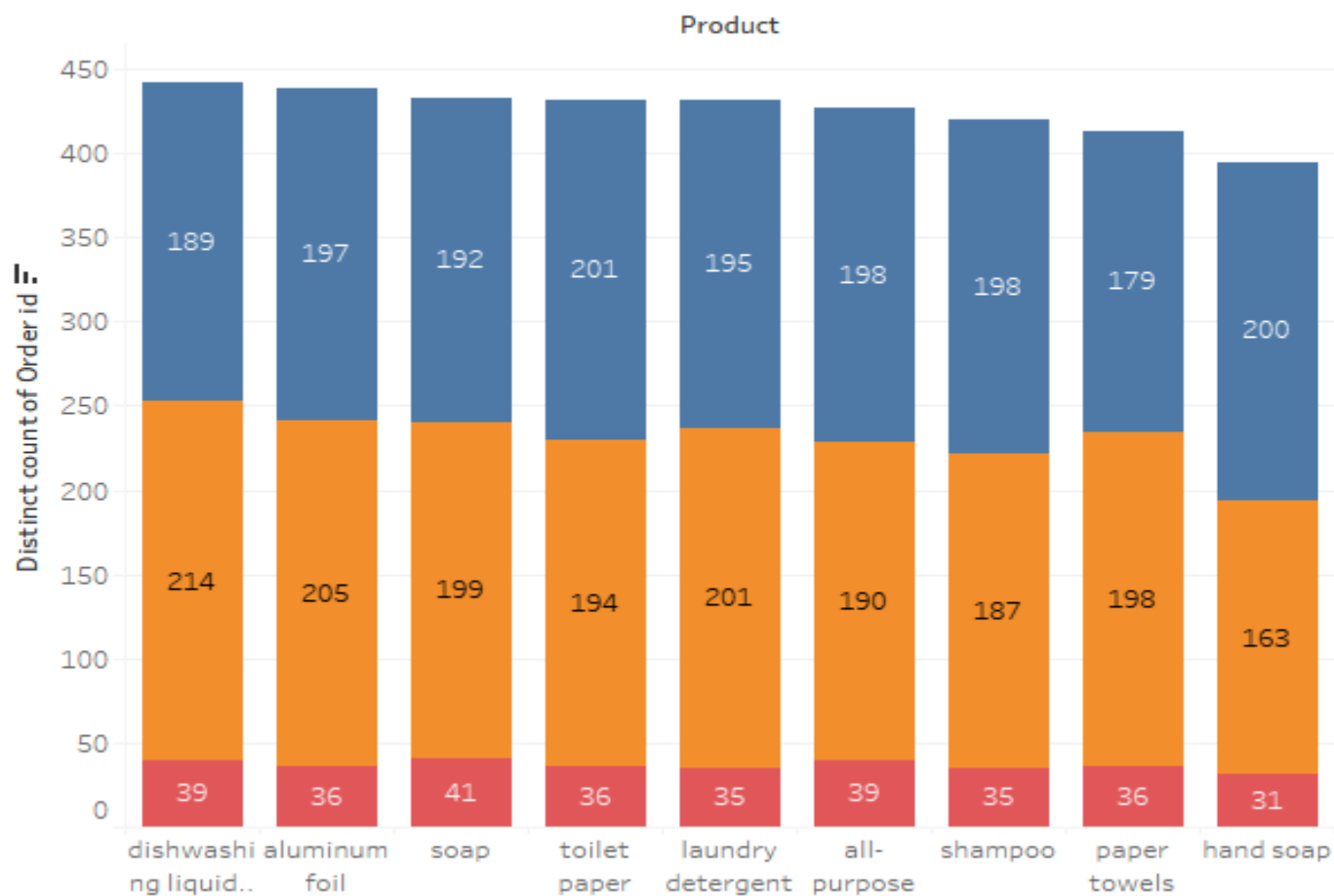
Activate Windows

EDIBLE PRODUCTS



Activate Window

NON EDIBLE PRODUCTS

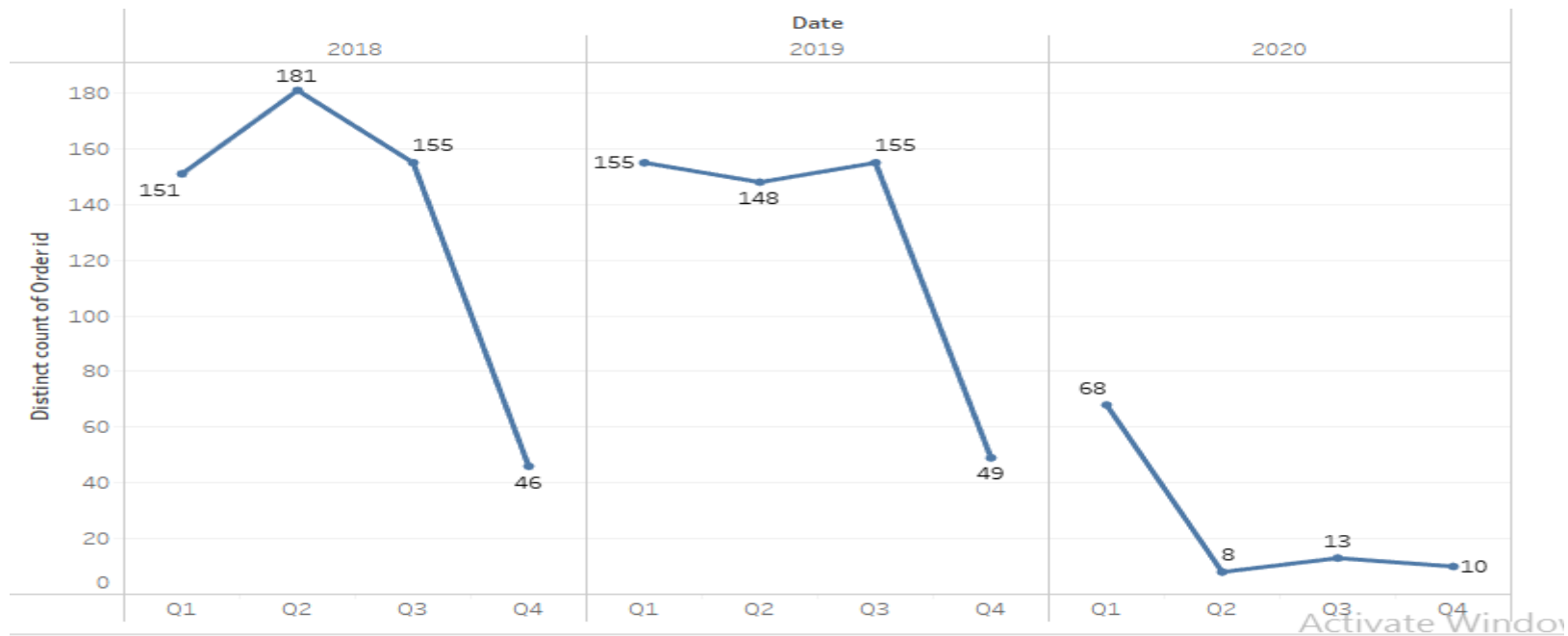
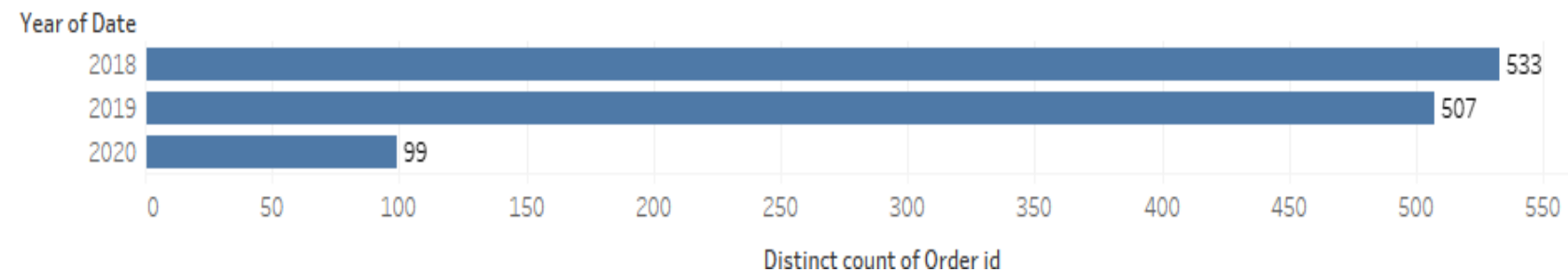


EXECUTIVE SUMMARY FROM EDA

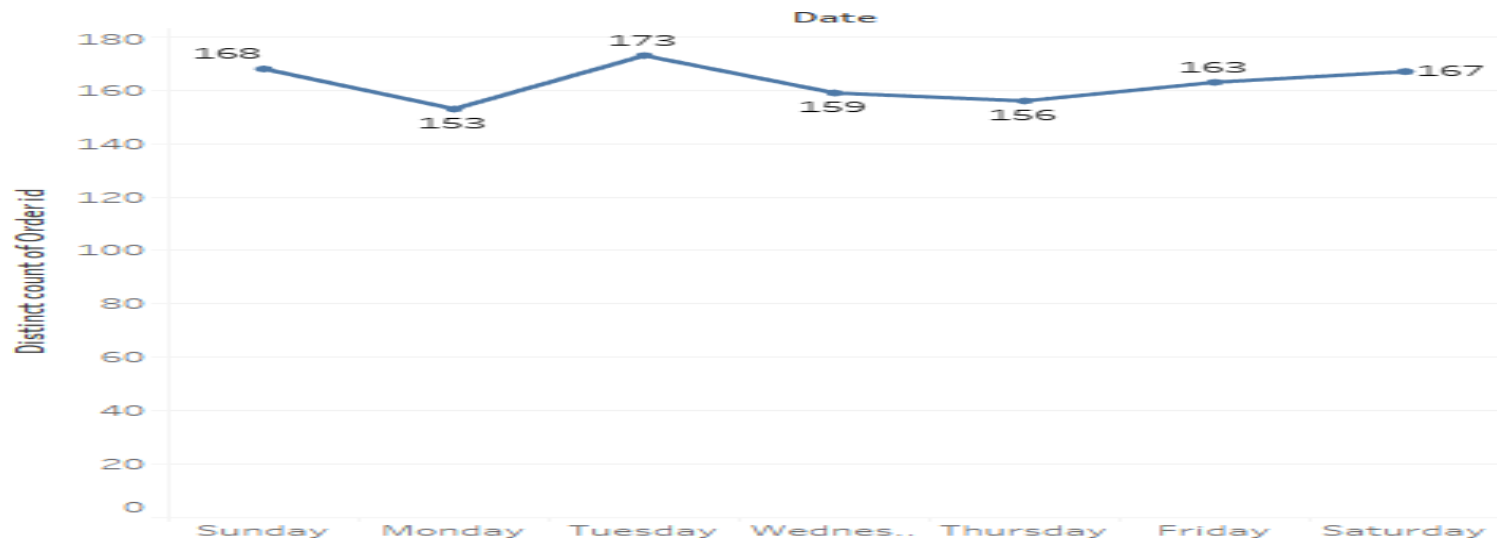
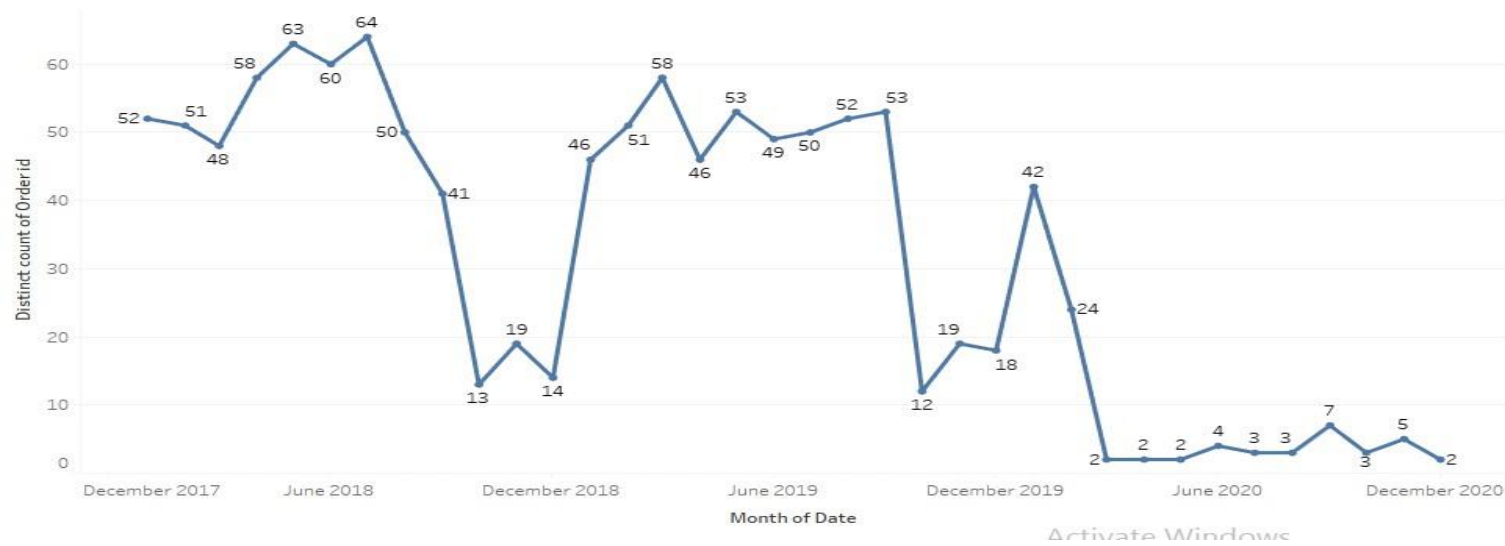
- ❖ 4730 duplicates present in the given dataset and can be removed .
Finally we got 15911 rows and 3 columns to proceed.
- ❖ Total Distinct Order id : 1139
- ❖ Number of products sold from 2018 to 2020 in the store are 37 among them , Edible products are 27 and Non-edible products are 10.
- ❖ On the whole , Poultry , Ice cream , Cereals are the Top 3 products sold whereas Handsoap, Flour are least sold.

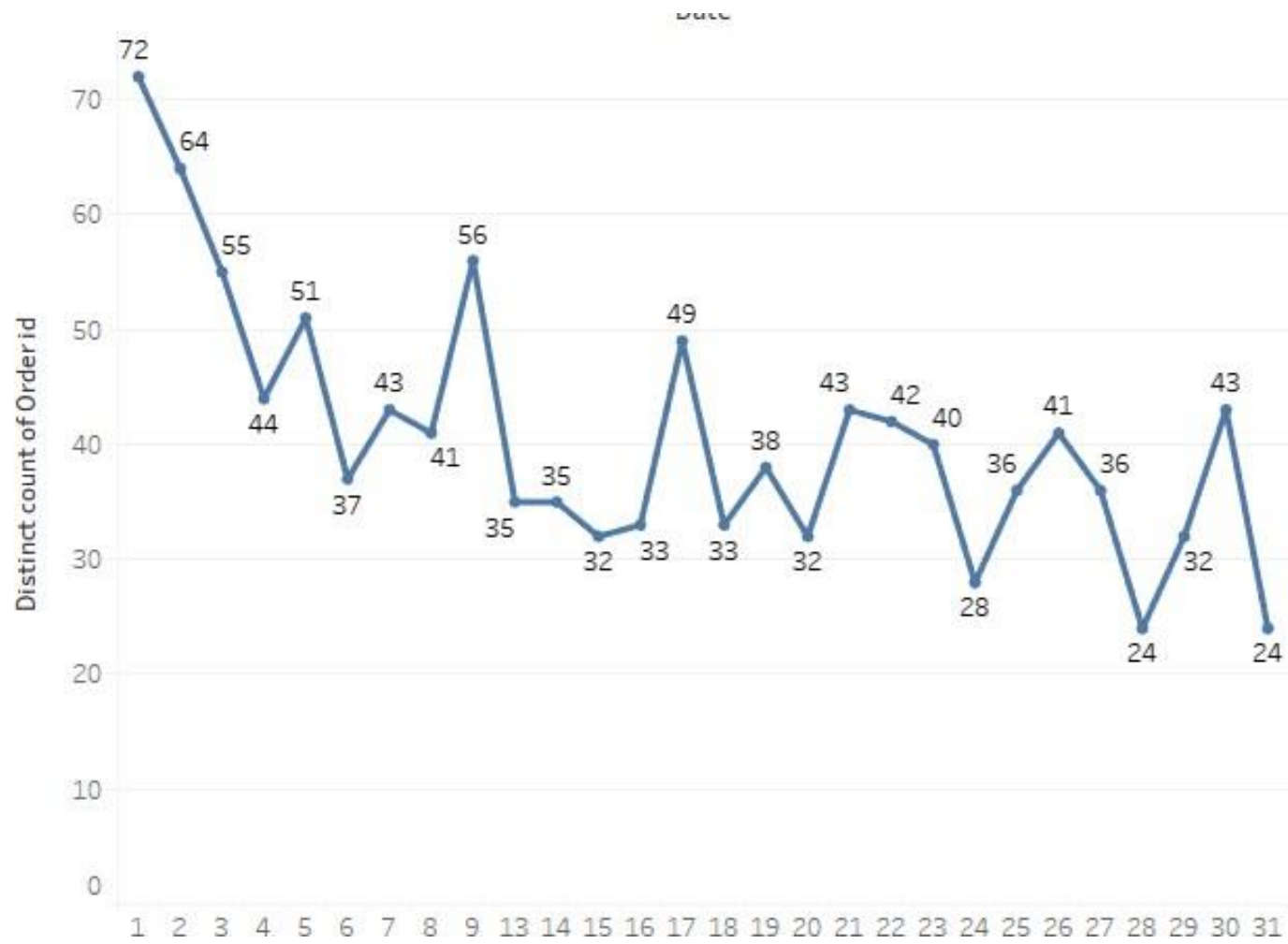
TREND ACROSS SALES

YEAR VS DISTINCT ORDER_ID



YEAR VS DISTINCT ORDER_ID





INSIGHTS FROM TRENDS

- ❖ On the whole , Nearly 47 % of total orders placed in the year 2018 itself (533 out of 1139).
- ❖ The given dataset provide sales details from 01-01-2018 to 26-02-2020.
- ❖ No sales details available for 4th Quarter of every year.

YEAR	HIGHEST QUARTERLY	HIGHEST MONTHLY
2018	Q2 (181)	JULY
2019	Q1 (155)	MARCH
2020	Q1 (68)	JANUARY

- ❖ In General, huge orders placed during the month of January and also 1st of every month.
- ❖ On the whole , huge orders placed on Tuesday whereas least orders placed on Monday. After 2019 Q3 , the sales has been drastically reduced till Feb 2020 .

YEAR	HIGHEST SOLD PRODUCTS	LEAST SOLD PRODUCTS
2018	CEREALS, POULTRY	SANDWICH LOAVES, PAPER TOWELS
2019	POULTRY, WAFFLES	HANDSOAP , FLOUR
2020	POULTRY, DINNER ROLLS	SUGAR, FRUITS

- ❖ Tableau Link

https://public.tableau.com/app/profile/abinaya.m8348/viz/MRA_GROCERY/Sheet3

MARKET BASED ANALYSIS

- ❖ **Market Basket Analysis** is a statistical technique that analyzes customer purchase patterns to identify associations between different products. It helps businesses understand which products are frequently purchased together and how customers' buying habits affect sales.
- ❖ To conduct market basket analysis, businesses need transactional data that includes details such as customer ID, product ID, and transaction date. This data is then used to create a matrix that represents the relationships between different products.

ASSOCIATION RULES

- ❖ **Association rules** are used to identify the strength of the relationship between different products. These rules are expressed in terms of support, confidence, and lift.
- ❖ Support refers to the frequency of co-occurrence of items in a transaction
- ❖ Confidence measures the probability that if a customer buys one item, they will also buy another.
- ❖ Lift measures the degree of correlation between two items.
- ❖ Applications of MBA :

Retail, e-commerce, and marketing etc. Retailers use this technique to optimize product placement and promotions. E-commerce companies use it to personalize product recommendations, and marketers use it to develop targeted advertising campaigns.

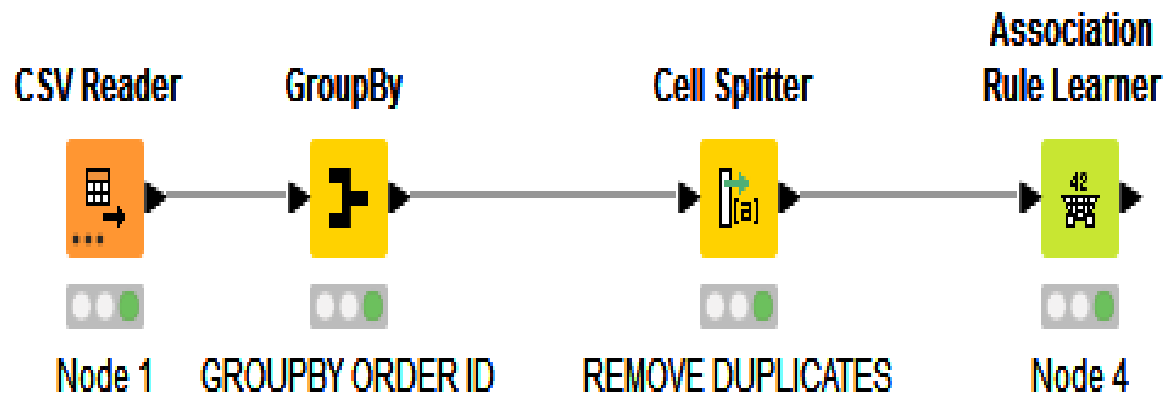
USES OF MBA

- ❖ Market Basket Analysis helps business to increase revenue by identifying cross-selling opportunities and developing targeted promotions.
- ❖ It also helps improve customer satisfaction by providing personalized recommendations.
- ❖ It also improving the overall shopping experience.

- ❖ Market basket analysis can increase sales and customer satisfaction.

Using data to determine that products are often purchased together, retailers can optimize product placement, offer special deals and create new product bundles to encourage further sales of these combinations.

KNIME WORKFLOW



OUTPUT TABLE

Table "default" - Rows: 24						
Spec - Columns: 6		Properties		Flow Variables		
Row ID	D Support	D Confide...	D Lift	S Conseq...	S implies	[...] Items
rule0	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum f...
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,l...
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,ce...
rule10	0.051	0.604	1.589	milk	<---	[poultry,laundry detergent,cer...
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 cr...
rule13	0.052	0.686	1.628	poultry	<---	[dinner rolls,spaghetti sauce,ic...
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,jui...
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,laund...
rule19	0.054	0.656	1.556	poultry	<---	[dinner rolls,spaghetti sauce,la...
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]

RELEVANCE OF ASSOCIATION

Association Rules is the set of rules where the likelihood of buying the product is greater with a set of products.

Eg) In our case from the output table , the paper towel can be bought along with the set of [eggs, ice cream , pasta] is 1.791 times higher than being bought individually which is just 0.055.

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]

ASSOCIATION RULE PARAMETERS

- ❖ Minimum Support value used : 0.05
- ❖ Maximum itemset length : 10
- ❖ Minimum Confidence used : 0.6

The screenshot shows a software dialog box titled "Dialog - 5:4 - Association Rule Learner". It has a pink title bar with standard window controls. The dialog is divided into several sections:

- File**: A menu bar at the top.
- Options**: A tabbed interface with four tabs: "Options" (selected), "Flow Variables", "Job Manager Selection", and "Memory Policy".
- Itemset Mining**: A section containing:
 - "Column containing transactions": A dropdown menu showing "[...] Product_SplitResultSet".
 - "Minimum support (0-1)": A numeric input field set to "0.05".
 - "Underlying data structure": A dropdown menu showing "ARRAY".
- Output**: A section containing:
 - "Itemset type": A dropdown menu showing "CLOSED".
 - "Maximal itemset length": A numeric input field set to "10".
- Association Rules**: A section containing:
 - A checked checkbox labeled "Output association rules".
 - "Minimum confidence": A numeric input field set to "0.6".

At the bottom of the dialog are four buttons: "OK", "Apply", "Cancel", and a help button (a circle with a question mark).

ASSOCIATIONS IDENTIFIED

We observed that in the given dataset, we found out 24 associations .

Table "default" - Rows: 24 Spec - Columns: 6 Properties Flow Variables						
Row ID	D Support	D Confide...	D Lift	S Conseq...	S implies	[...] Items
rule0	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum f...
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,l...
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,ce...
rule10	0.051	0.604	1.589	milk	<---	[poultry,laundry detergent,cer...
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 cr...
rule13	0.052	0.686	1.628	poultry	<---	[dinner rolls,spaghetti sauce,ic...
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,jui...
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,laund...
rule19	0.054	0.656	1.556	poultry	<---	[dinner rolls,spaghetti sauce,la...
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]

ASSOCIATIONS IN TABULAR MANNER

product 1	product 2	product 3	combinations
yogurt	Toilet paper	aluminum foil	juice
yogurt	poultry	aluminum foil	juice
yogurt	cheeses	cereals	coffee/tea
dishwashing liquid/detergent	laundry detergent	mixes	poultry
yogurt	poultry	aluminum foil	mixes
cheeses	bagels	cereals	sandwich bags
bagels	cereals	sandwich bags	cheeses
cheeses	bagels	sandwich bags	cereals
spaghetti sauce	poultry	cereals	dinner rolls
dinner rolls	spaghetti sauce	cereals	poultry
poultry	laundry detergent	cereals	milk
dinner rolls	poultry	soda	eggs
spaghetti sauce	poultry	ice cream	dinner rolls
dinner rolls	spaghetti sauce	ice cream	poultry
spaghetti sauce	poultry	juice	dinner rolls
dinner rolls	spaghetti sauce	juice	poultry
paper towels	dinner rolls	pasta	eggs
paper towels	eggs	dinner rolls	pasta
spaghetti sauce	poultry	laundry detergent	dinner rolls
dinner rolls	spaghetti sauce	laundry detergent	poultry
paper towels	eggs	pasta	ice cream
paper towels	ice cream	pasta	eggs
paper towels	eggs	ice cream	pasta
eggs	ice cream	pasta	paper towels

CALCULATIONS USED

- ❖ Support of minimum used = 0.05

It is the probability of observing the items together in a transaction. It is calculated as the number of transactions that contain both items divided by the total number of transactions. It measures how frequent the itemset occurs in the dataset. High support indicates that the itemset is popular and should be considered for promotion or placement together.

- ❖ Minimum confidence used = 0.6

It is the conditional probability that a transaction containing one item also contains another item. It is calculated as the number of transactions containing both items divided by the number of transactions containing the first item. It measures the strength of the association between two items. High confidence indicates that the items are likely to be bought together, and can be used to recommend or suggest items to customers.

❖ Lift :

It is the measure of how much more often two items occur together than expected if they were independent of each other. It is calculated as the support of the item set divided by the product of the individual supports of the items.

A lift value of 1 indicates that the items are independent
value greater than 1 indicates a positive association between the items.

A lift value less than 1 indicates a negative association between the items.

High lift indicates that the items have a strong association and can be used for cross-selling or bundling.

SUGGESTIONS

- ❖ Poultry could be provided as a combo offer for most of the foods such as dinner rolls, spaghetti sauce and ice cream.
- ❖ Ice cream sold more so it can be provided as combo offer.
- ❖ Top 5 combo offers suggested are :

poultry	<---	[dinner rolls,spaghetti sauce,ice cream]
cheeses	<---	[bagels,cereals,sandwich bags]
poultry	<---	[dinner rolls,spaghetti sauce,laundry detergent]
paper towels	<---	[eggs,ice cream,pasta]
pasta	<---	[paper towels,eggs,ice cream]
dinner rolls	<---	[spaghetti sauce,poultry,laundry detergent]
.	.	.

RECOMMENDATIONS

- ❖ Provide a discount on dinner rolls when purchased with spaghetti sauce or poultry.
- ❖ Create a "Paper Products Bundle" offer that includes paper towels, toilet paper, or tissues at a discounted price.
- ❖ Create a combo deal where customers can purchase cereals, bagels, and sandwich bags together at a discounted price.
- ❖ Provide mega sale offers on least sold products to increase its sales.
- ❖ Since Poultry and Ice cream are the most sold items and Hand soap and Flour are the least combo offer of these would eventually increase a sale of Hand soap and Flour as well.
- ❖ Offer a "Buy Two Get One Free" promotion on yogurt, poultry, and aluminum foil to encourage customers to purchase more items at once.

- ❖ Yogurt, poultry, aluminum foil, cheeses, cereals, ice cream and dinner rolls are some of the most frequently purchased products so we can give them as combo offers.
- ❖ It is important to promote these offers to increase sales through in-store signage, advertisements, and social media to ensure customers are aware of the deals available because the sales of the store are in decreasing trend since end of 2019.
- ❖ Provide offers during festival season to attract customers
- ❖ Provide credit points for the customers to encourage sales.
- ❖ Offers provided to the customers who bought more. This makes the customers to retain.
- ❖ If possible provide timely delivery of those ordered products to increase sales.
- ❖ Promote online platform to attract customers in this digital world.

TOOLS USED

- ❖ For Data summary , Python is used
- ❖ For EDA , Tableau Tool is used

https://public.tableau.com/app/profile/abina_ya.m8348/viz/MRA_GROCERY/Sheet3

- ❖ For RFM Analysis , KNIME Tool used.

THANK YOU