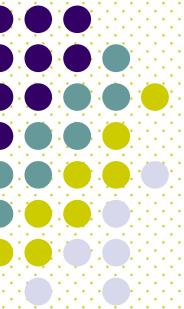


# Marketing and Retail Analysis

MBA

Vikash Kumar



## Agenda & Executive Summary of the data

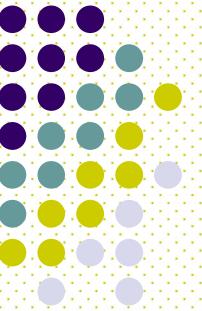
### Agenda

The project involves conducting a thorough analysis of Point of Sale (POS) Data for providing 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 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 items in the customer orders. The Store doesn't have any combo offers.

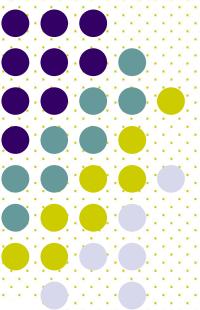
We have to suggest the best combos & offers?



# Contents of the presentation

- ❖ Exploratory Analysis
- ❖ Use of Market Basket Analysis (Association Rules)
- ❖ Associations Identified
- ❖ Possible Combos with Lucrative Offers
- ❖ Tools used

# Executive Summary of the data



## Top Findings

TOP 5	Date	Order_id	Product
0	2018-01-01	1	yogurt
1	2018-01-01	1	pork
2	2018-01-01	1	sandwich bags
3	2018-01-01	1	lunch meat
4	2018-01-01	1	all- purpose

Data Description	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

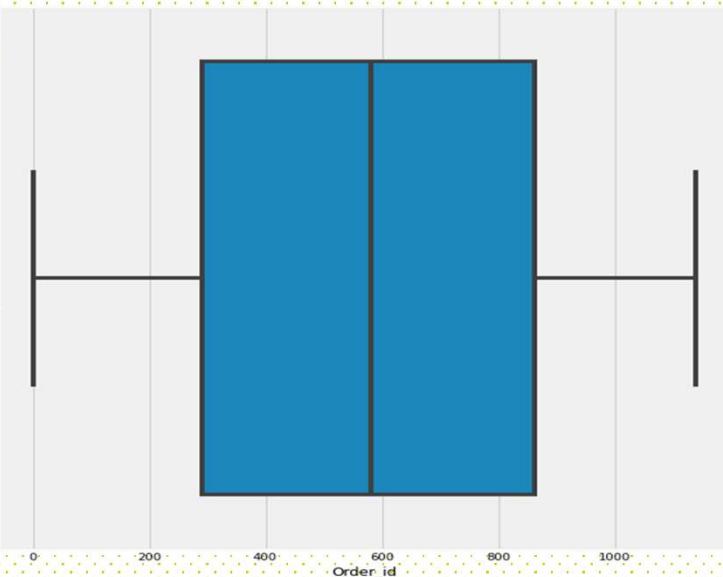
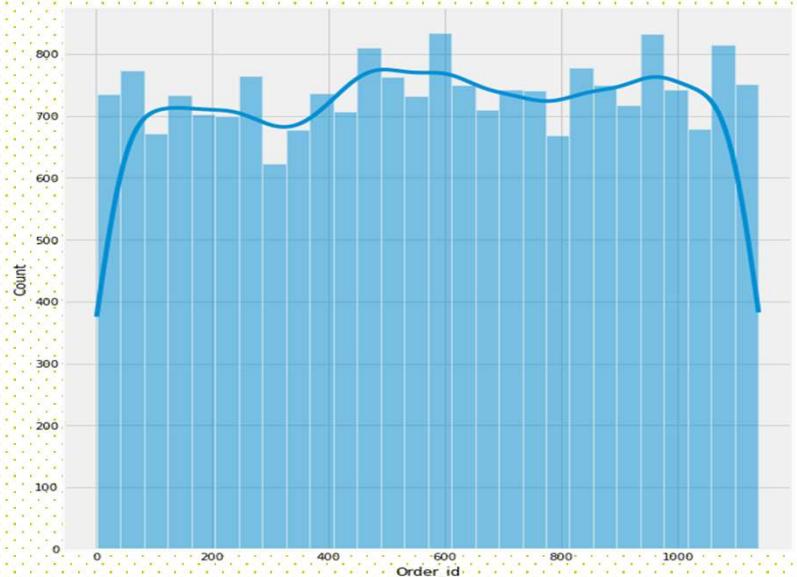
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20641 entries, 0 to 20640
Data columns (total 3 columns):
 #   Column      Non-Null Count Dtype  
 0   Date        20641 non-null  datetime64[ns]
 1   Order_id    20641 non-null  int64   
 2   Product     20641 non-null  object  
dtypes: datetime64[ns](1), int64(1), object(1)
memory usage: 483.9+ KB
```

- The shape of dataset is : (20641, 3)
- There is NO Null values in Data set
- The Duplicated in dataset is : 4730
- Unique value Count:

Order_id	1139
Product	37

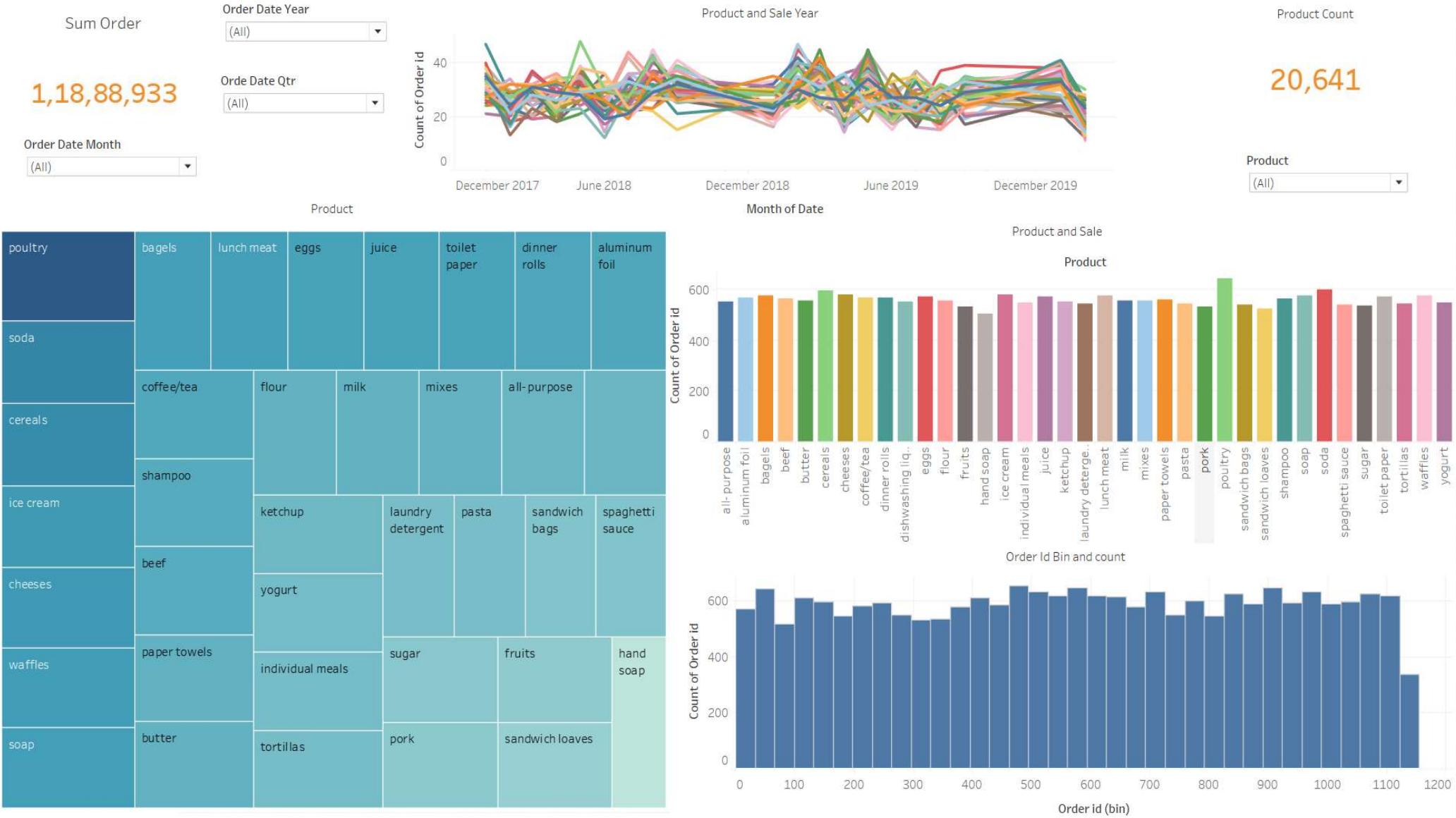
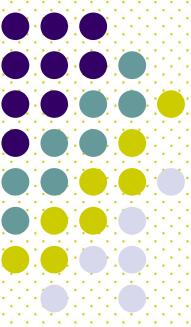
# Exploratory Analysis and Inferences

## Univariate Analysis

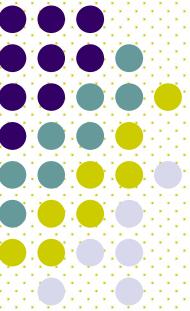


Data of Order id not having the normal distribution Count varies from minimum to more than 250 aprox to 1100 with mean value of 580 aprox.

# Data Summary :



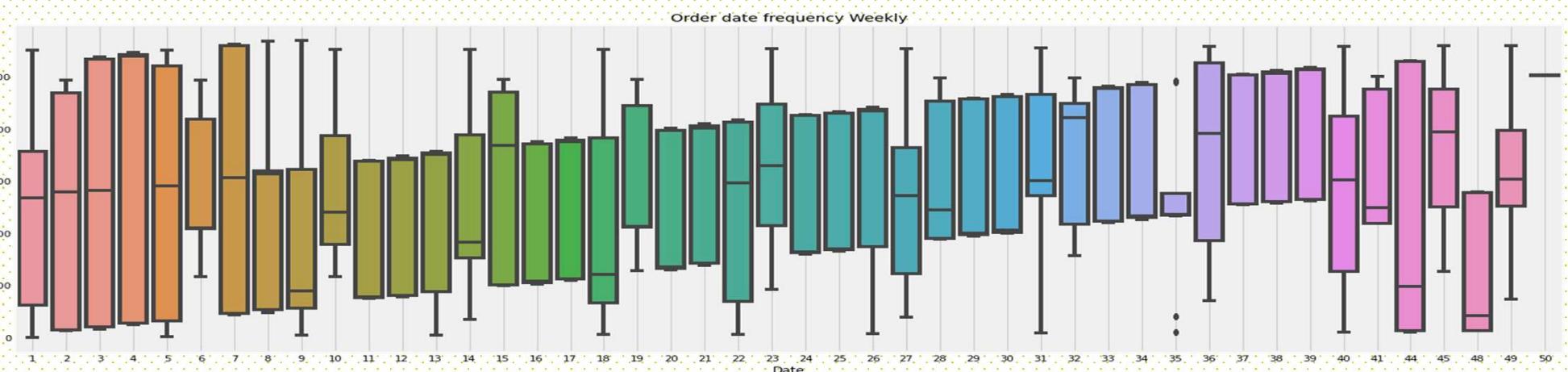
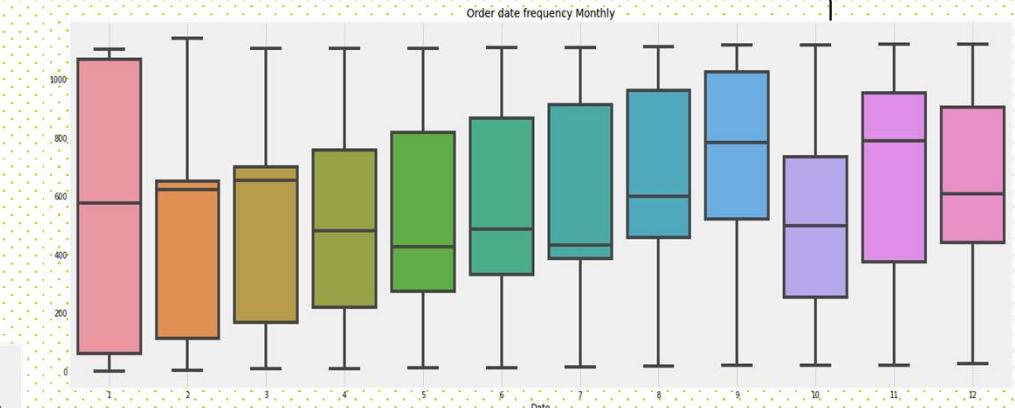
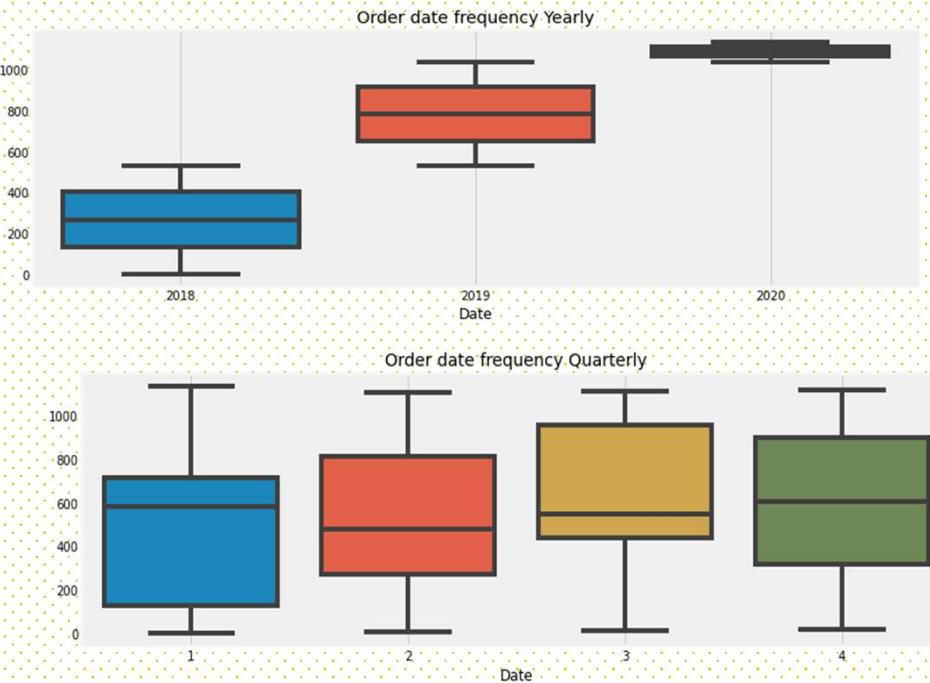
# Variables Insights



- ❖ Three types of variable in TS data set.
- ❖ Highest sale of Poultry while lowest of hand soap.
- ❖ Total products are 37
- ❖ Total order Id as 1139
- ❖ The total order generated are 1,18,88,933
- ❖ The mean (average) value of the "Order\_id" variable is approximately 575.986289.
- ❖ The standard deviation of the "Order\_id" variable is approximately 328.557078, indicating a relatively wide range of values.
- ❖ The minimum value of "Order\_id" is 1, implying that the numbering of orders starts from 1.
- ❖ The 25th percentile of "Order\_id" is 292, which means that 25% of the orders have an id lower than or equal to 292.
- ❖ The median (50th percentile) of "Order\_id" is 581, indicating that half of the orders have an id lower than or equal to 581.
- ❖ The 75th percentile of "Order\_id" is 862, implying that 75% of the orders have an id lower than or equal to 862.
- ❖ The maximum value of "Order\_id" is 1139, indicating the highest order number in the dataset.

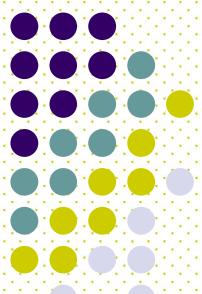
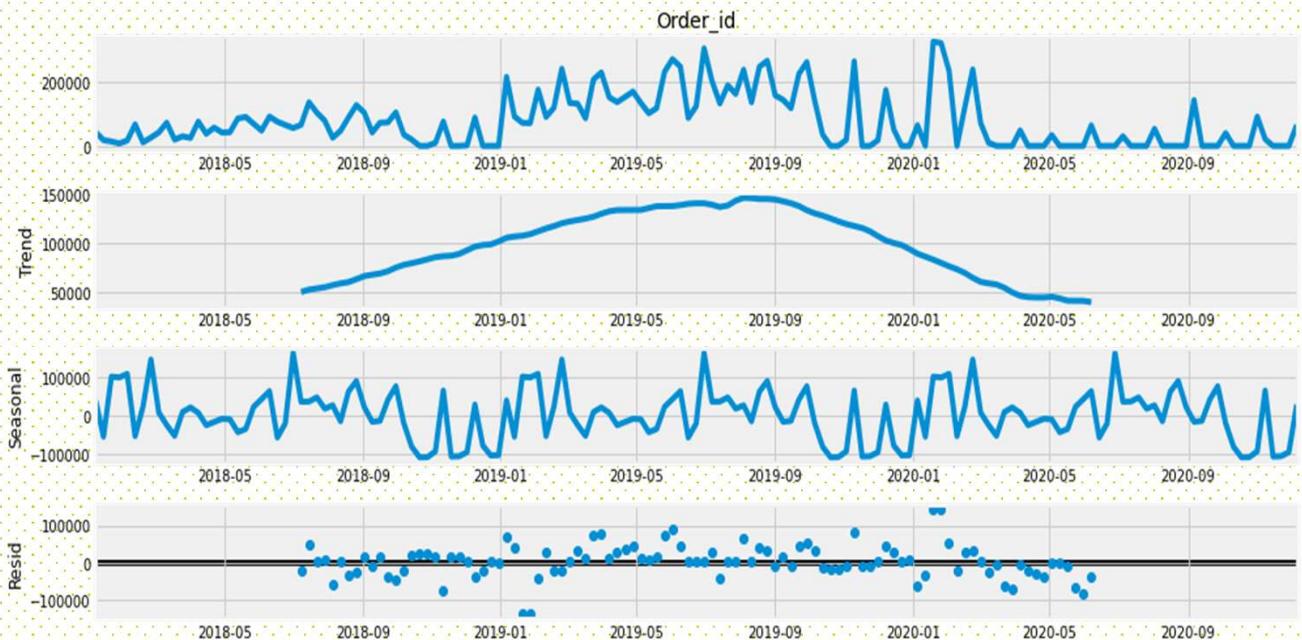
# Trend Analysis

Order\_Id

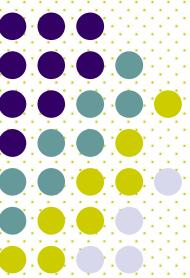


# Trend Analysis

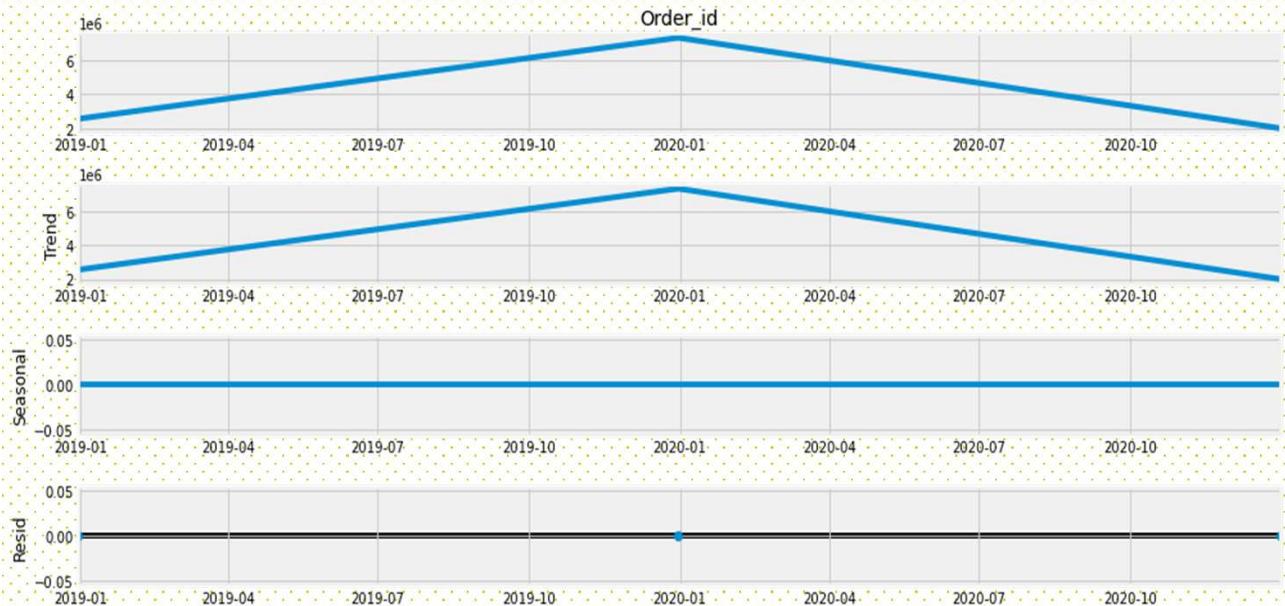
Monthly and weekly



# Trend Analysis



Quarterly and Yearly



← Decomposition of Yearly sales

Decomposition of Quarterly sales

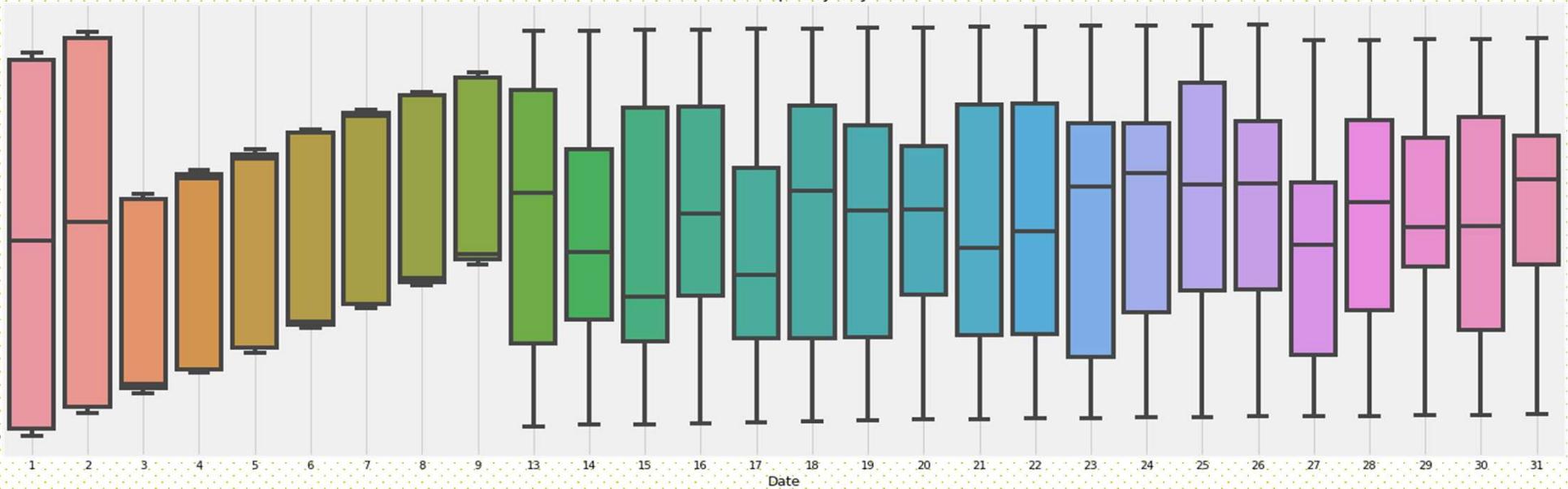


# Trend Analysis

Daily Sales



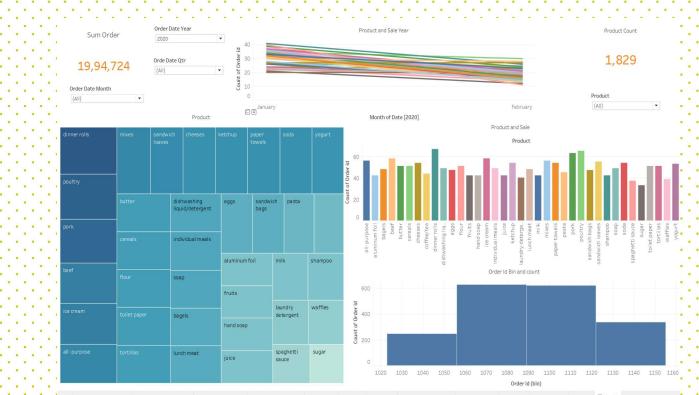
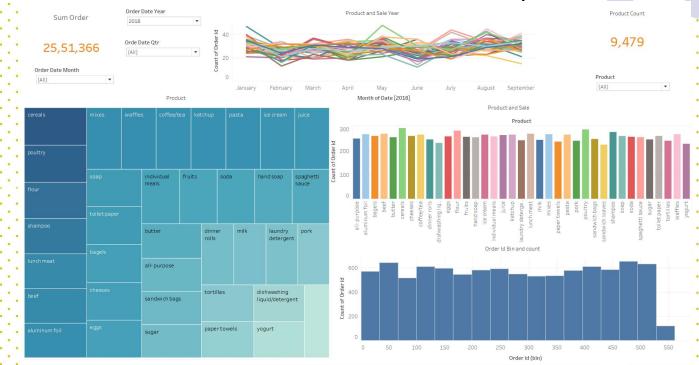
Order date frequency Day wise



# Trend Analysis

- ❖ There is trend in sales observed: From year 2019-01 on steady increasing trend till 2020-10. Decreasing trend after that in Yearly Decomposition
- ❖ Increasing Trend in sales can be observed From year 2018-10 to 2019-7, while on steady till 2019-10 .
- ❖ The decreasing trend till 2020-07 in Monthly Decomposition
- ❖ Same trend but Seasonality are now granular through out the year in Weekly Decomposition

## Plot Insights

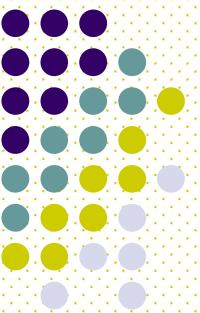


## Year wise Major and Minor sales Analysis

Year	Dominant Product	Minor Product
2018	Cereal and poultry	Sanwich Loaves
2019	Poultry and Soda	Hand Soap
2020	Dinner Role, Poultry, Pork	Sugar

# Market Basket Analysis

Association rules



**Market basket analysis** is a data mining technique used to discover relationships between products that are frequently purchased together in a transactional dataset.

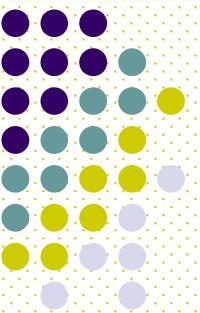
**Association rules** are derived from MBA, which provide insights into the co-occurrence and dependencies among items. Association rules are derived from this analysis, which provide insights into the co-occurrence and dependencies among items.

It is used to identify patterns or relationships between items that occur frequently together in a transactional database. The association rule is an implication expression of the form  $X \rightarrow Y$ , where  $X$  and  $Y$  are any two itemset. The rule evaluation metrics include support(s), confidence and lift<sup>12</sup>.

It consists of two parts: **antecedent** (the item(s) that appear in the left-hand side of the rule) and **consequent** (the item(s) that appear in the right-hand side of the rule). The rules are represented as "If antecedent, then consequent" and are typically measured using support, confidence, and lift metrics:



# Market Basket Analysis



Association rules and their relevance in this case

- **Support:** Measures the frequency or proportion of transactions that contain both the antecedent and consequent items. It indicates the popularity of the rule. A higher support indicates a stronger association between the items.
- **Confidence:** Indicates the likelihood that the consequent item(s) will be purchased when the antecedent item(s) are already in the basket. It measures the conditional probability of the rule. Higher confidence values indicate stronger relationships between the items.
- **Lift:** Compares the observed support of a rule to the expected support if the items were independent of each other. Lift greater than 1 indicates a positive association between the items.



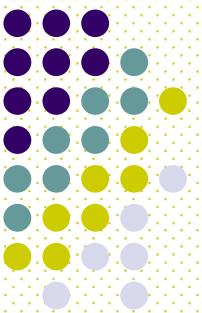
$$\text{Support} = \frac{\text{frq}(X, Y)}{N}$$

Rule:  $X \Rightarrow Y$

$$\text{Confidence} = \frac{\text{frq}(X, Y)}{\text{frq}(X)}$$
$$\text{Lift} = \frac{\text{Support}}{\text{Supp}(X) \times \text{Supp}(Y)}$$

# Market Basket Analysis

Association rules and their relevance in this case



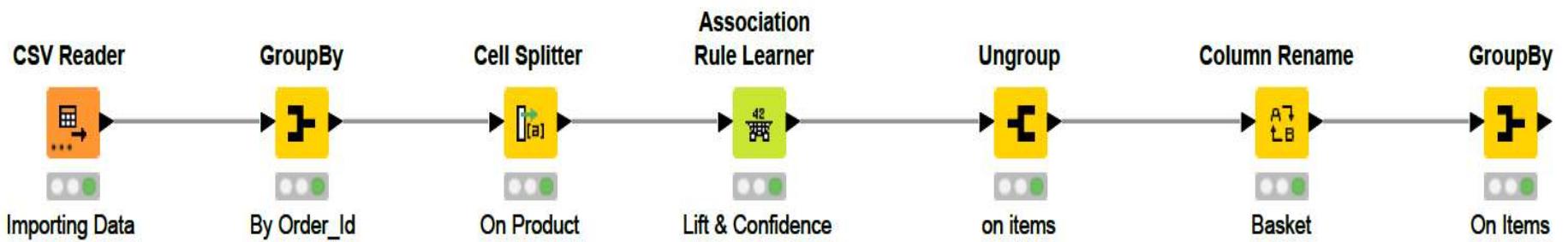
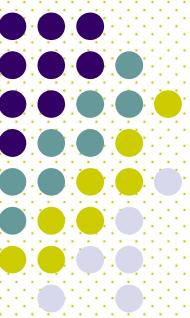
Association rules are considered relevant if they meet certain predefined thresholds for support, confidence, and lift. The choice of these thresholds depends on the specific application and the desired level of significance. High support indicates that the rule is based on a substantial number of transactions, high confidence indicates a strong relationship between the items, and lift indicates the strength of the association beyond what would be expected by chance.

Relevance in association rule mining refers to the usefulness or significance of the discovered rules in the context of the problem at hand. Relevant rules can provide valuable insights for various applications such as market basket analysis, cross-selling, recommendation systems, and targeted marketing campaigns. By analyzing the association rules, businesses can make informed decisions on product placement, promotions, and customer segmentation to optimize their strategies and improve overall business performance.



# Market Basket analysis

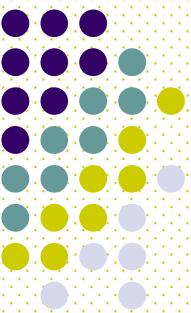
KNIME Work Flow



Node Name	Description
Read CSV	Read the CSV file
Group by	Grouping on the basis of Order_Id
Cell Splitter	Convert the Product data to Set
Association Rule:	Market basket Analysis to generate the frequent item list
Ungroup	On the basis of item
Column Rename	Change the name of Item and remane as Basket
Groupby	Grouping on the basis of items for granularity in output

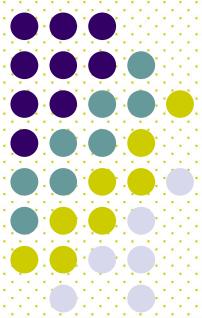
# Threshold values of Support and Confidence

## Outline

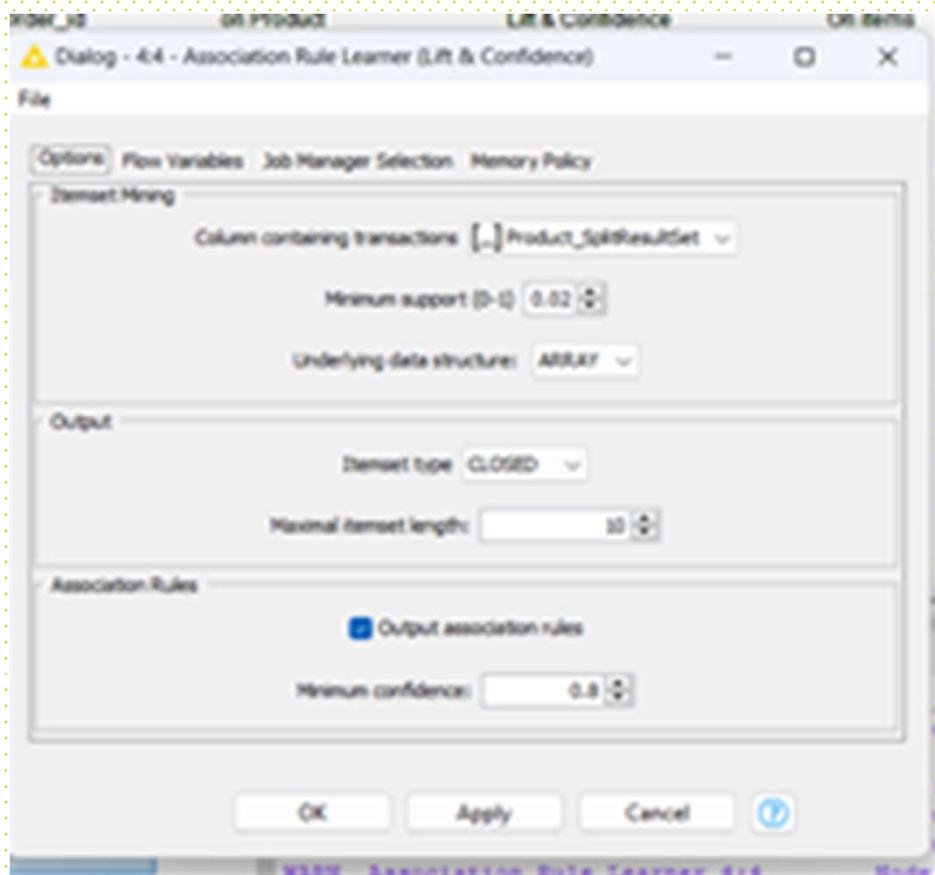


- ❖ Threshold values for support and confidence play a crucial role in determining the significance and usefulness of discovered association rules. Support and confidence are two common measures used to quantify the strength and reliability of associations between items in a dataset.
- ❖ Determining suitable threshold values for support and confidence involves a trade-off between rule quality and quantity. Higher thresholds result in fewer but more reliable rules, whereas lower thresholds generate a larger number of rules, which may include both significant and less significant associations. The choice of threshold values is influenced by the specific requirements of the analysis, domain knowledge, and the desired balance between precision and recall in the rule discovery process.
- ❖ It's worth mentioning that the selection of appropriate threshold values is often an iterative process, where we adjust the thresholds, analyze the generated rules, and refine the values based on their relevance and usefulness in the given context.

# Associations Identified

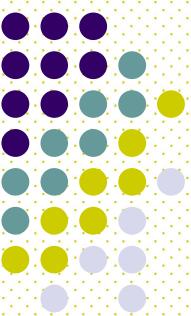


Support, confidence, & lift values that are calculated



- ❖ The Threshold values are found out by various calculation run in background of tool KNIME
- ❖ We fine Tune the Confidence and Support to get the best combination out of it.
  - ❑ The Support of Minimum =0.02
  - ❑ Maximum Item set length is =10
  - ❑ The Confidence level =0.8

# Associations Identified



Associations in a tabular manner

Group table - 46 - Groupby (On items)

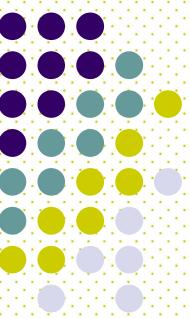
File Edit Hilitc Navigation View

Table "default" - Rows: 36 Spec - Columns: 7 Properties Flow Variables

Row ID	Items	Support	Confidence	Lift	Consequent	Implies	Basket
Row8	dishwashing liqu...	0.021	0.837	2.2	mixes	<=	[yogurt, dishwashing liquid/detergent, all-purpose, hand soap]
Row22	pasta	0.021	0.841	2.196	soda	<=	[bacon, pasta, individual meals, pork]
Row9	eos	0.021	0.838	2.194	milk	<=	[eos, poultry, beef, sandwich bao]
Row24	poultry	0.022	0.825	2.191	milk	<=	[eos, poultry, beef, sandwich bao]
Row11	fruits	0.021	0.826	2.182	beef	<=	[shampoo, fruits, lunch meat, pork]
Row1	bacon	0.02	0.852	2.18	soda	<=	[bacon, pasta, individual meals, pork]
Row18	lunch meat	0.022	0.835	2.172	beef	<=	[shampoo, fruits, lunch meat, pork]
Row23	pork	0.021	0.836	2.168	beef	<=	[shampoo, fruits, lunch meat, pork]
Row4	cereals	0.025	0.835	2.168	coffee/tea	<=	[yogurt, ice cream, tortillas, cereals]
Row32	toilet paper	0.022	0.82	2.167	bacon	<=	[sandwich loaves, fruits, toilet paper, juice]
Row25	sandwich bao	0.023	0.831	2.166	milk	<=	[eos, poultry, beef, sandwich bao]
Row13	ice cream	0.022	0.825	2.158	paper towels	<=	[eos, dinner rolls, ice cream, pasta, lunch meat]
Row0	all-purpose	0.023	0.818	2.143	mixes	<=	[yogurt, dishwashing liquid/detergent, all-purpose, hand soap]
Row35	yogurt	0.023	0.823	2.143	mixes	<=	[yogurt, dishwashing liquid/detergent, all-purpose, hand soap]
Row16	ketchup	0.024	0.809	2.133	soap	<=	[spaghetti sauce, all-purpose, sandwich bao, ketchup]
Row34	waffles	0.023	0.822	2.132	spaghetti sauce	<=	[waffles, laundry detergent, mixes, soap]
Row14	individual meals	0.02	0.837	2.13	soda	<=	[bacon, pasta, individual meals, pork]
Row21	paper towels	0.023	0.832	2.129	lunch meat	<=	[paper towels, milk, individual meals, coffee/tea]
Row33	tortillas	0.022	0.825	2.125	yogurt	<=	[cheeses, all-purpose, tortillas, coffee/tea]
Row28	soap	0.024	0.828	2.12	spaghetti sauce	<=	[waffles, laundry detergent, mixes, soap]
Row15	juice	0.023	0.824	2.114	yogurt	<=	[dishwashing liquid/detergent, eos, juice, sandwich bao]
Row20	mixes	0.022	0.818	2.111	spaghetti sauce	<=	[waffles, laundry detergent, mixes, soap]
Row10	flour	0.025	0.8	2.109	soap	<=	[all-purpose, flour, soda, ketchup]
Row29	soda	0.024	0.809	2.092	waffles	<=	[paper towels, laundry detergent, soda, sugar]
Row19	milk	0.022	0.818	2.091	lunch meat	<=	[paper towels, milk, individual meals, coffee/tea]
Row6	coffee/tea	0.022	0.82	2.087	yogurt	<=	[cheeses, all-purpose, tortillas, coffee/tea]
Row27	shampoo	0.023	0.825	2.079	beef	<=	[shampoo, fruits, lunch meat, pork]
Row7	dinner rolls	0.024	0.831	2.072	ice cream	<=	[paper towels, eos, dinner rolls, pasta, lunch meat]
Row17	laundry detergent	0.023	0.813	2.054	spaghetti sauce	<=	[waffles, laundry detergent, mixes, soap]
Row31	sugar	0.025	0.832	2.043	beef	<=	[poultry, fruits, hand soap, sugar]
Row12	hand soap	0.024	0.828	2.034	mixes	<=	[yogurt, dishwashing liquid/detergent, all-purpose, hand soap]
Row5	cheeses	0.021	0.814	2.025	yogurt	<=	[cheeses, all-purpose, tortillas, coffee/tea]
Row2	beef	0.025	0.823	2.024	milk	<=	[eos, poultry, beef, sandwich bao]
Row30	spaghetti sauce	0.025	0.819	2.001	cheeses	<=	[butter, spaghetti sauce, ice cream, lunch meat]
Row3	butter	0.022	0.809	1.997	poultry	<=	[butter, cheeses, sandwich loaves, laundry detergent]
Row26	sandwich loaves	0.024	0.814	1.977	bacon	<=	[sandwich loaves, fruits, toilet paper, juice]

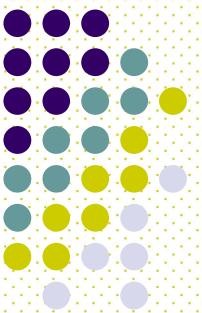
# Associations Identified

Top 5 Association by Lift Score



Row	Items	Support	Confidence	lift	Consequent	implied	Basket
Row33	tortillas	0.021729587	0.825056114	2.12511712	yogurt	<---	[cheeses, all-purpose, tortillas, coffee/tea]
Row29	soda	0.024143986	0.809090909	2.09239031	waffles	<---	[paper towels, laundry detergent, soda, sugar]
Row7	dinner rolls	0.023924495	0.831466636	2.07226287	ice cream	<---	[paper towels, eggs, dinner rolls, pasta, lunch meat]
Row30	spaghetti sauce	0.025168276	0.818732534	2.00068232	cheeses	<---	[butter, spaghetti sauce, ice cream, lunch meat]
Row3	butter	0.02238806	0.809475806	1.99663865	poultry	<---	[butter, cheeses, sandwich loaves, laundry detergent]

# Recommendations

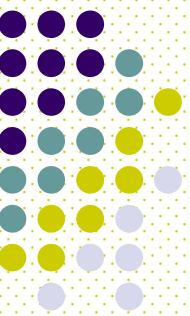


## Offer no 1 : Discount Price

Discount as Combo Offer: Buy cheeses, all-purpose, tortillas, and coffee/tea together and get a discounted price on yogurt.

Explanation:

- Based on the association rule "yogurt <-- [cheeses, all-purpose, tortillas, coffee/tea],"
- customers who purchase cheeses, all-purpose, tortillas, and coffee/tea are likely to be interested in yogurt as well.
- By offering a discounted price on yogurt when these items are purchased together, business can incentivize customers to buy the complete set and increase sales.



# Recommendations

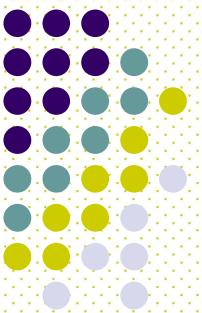
## Offer no 2 : Bundle Deal

Bundle Offer: Purchase paper towels, laundry detergent, soda, and sugar together at a discounted price and get waffles for free.

Explanation:

- The association rule "waffles <--- [paper towels, laundry detergent, soda, sugar]"
- suggests that customers who buy paper towels, laundry detergent, soda, and sugar are more likely to be interested in waffles as well. .
- By offering waffles as a free item when customers purchase the bundle,
- Business can encourage them to buy the entire set of items and create a positive customer experience.

# Recommendations



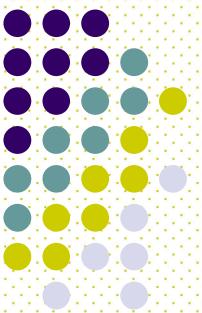
## Offer no 3: Buy Two get One Free

Buy Two, Get One Free Offer: Buy butter, spaghetti sauce, and ice cream together, and get a free pack of lunch meat.

Explanation:

- The association rule "cheeses <--- [butter, spaghetti sauce, ice cream, lunch meat]"
- Indicates a strong relationship between the items butter, spaghetti sauce, ice cream, and lunch meat. Customers purchasing these items are likely to be interested in cheeses as well.
- To capitalize on this association, Business can offer a "Buy Two, Get One Free" promotion, where customers who purchase butter, spaghetti sauce, and ice cream together receive a free pack of lunch meat.

# Recommendations



## Offer no 4:Combo for Meal Preparation

Combo for Meal Preparation Offer: Purchase butter, cheeses, sandwich loaves, and laundry detergent together at a discounted price and receive poultry at a reduced cost.

- Explanation:
  - The association rule "poultry <--- [butter, cheeses, sandwich loaves, laundry detergent]"
  - suggests that customers who buy butter, cheeses, sandwich loaves, and laundry detergent are more likely to be interested in poultry as well.
  - To encourage customers to buy these items together, you can offer a discounted price on poultry when they purchase the specified combination.
  - This can attract customers who are planning meal preparations involving these items and increase overall sales.

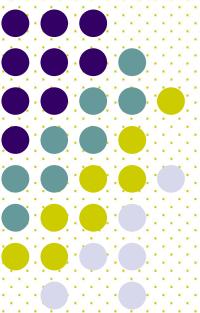
# Conclusion



Market basket analysis and association rules provide valuable insights into customer behavior and product relationships, enabling businesses to make data-driven decisions, improve customer satisfaction, and drive revenue growth

During analyzing of the dataset we should also consider additional factors such as pricing strategies, customer preferences, and market trends to tailor the offers effectively.

(These factors are not in our dataset, the assumption part has been incorporated for suggestions.)



## Tools Used

- Python for basic EDA , Trend, Seasonality, and visualisation (ipynb attached)
- Tableau for Visualisation ([Graded MRA2 | Tableau Public](#))
- KNIME for RFM analysis (Work flow and details mentioned)

