MARKETING & RETAIL ANALYTICS MILESTONE 2

PROJECT BY :- AKASH KAMBLE, JUNE - A BATCH 2022

Contents

Contents

EDA & Executive Summary of the Data

Use of Market Basket Analysis

Associations Identification

Recommendations & 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.
- Our job is to suggest combo meals to the store.

Executive Summary & Introduction

❖ Executive Summary :-

- The intend of the study 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 using Market Basket Analysis.

❖ Introduction :-

- In this study, we will be using different tools for the analysis such as Python, Tableau and KNIME.
- We will go-through EDA, sales trends, association rules, highly associated products using Market Basket Analysis and eventually drawing inferences based upon the analysis performed.

Data Summary

- **Data Description :-**

Data	Description
Date	Data of Purchase of Products
Ooder_id	Order ID Number
Products	Name of Products Purchased

- **Data Stats :-**

	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

1. There are 603 days for which data/transaction was recorded.
2. There are total 37 unique products in the store.

- **Data Info :-**

1. There are total 20641 records and 3 number of columns in the dataset.
2. Data consists of datetime64 and object datatypes.
3. There are no missing values.
4. There are 4730 duplicate rows in the data which need not to be treated as there could be record of products bought more than 1 quantity but recorded separately. Such products have same date of purchase and same order ID and thus results in duplicate records

```
 #   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)
```

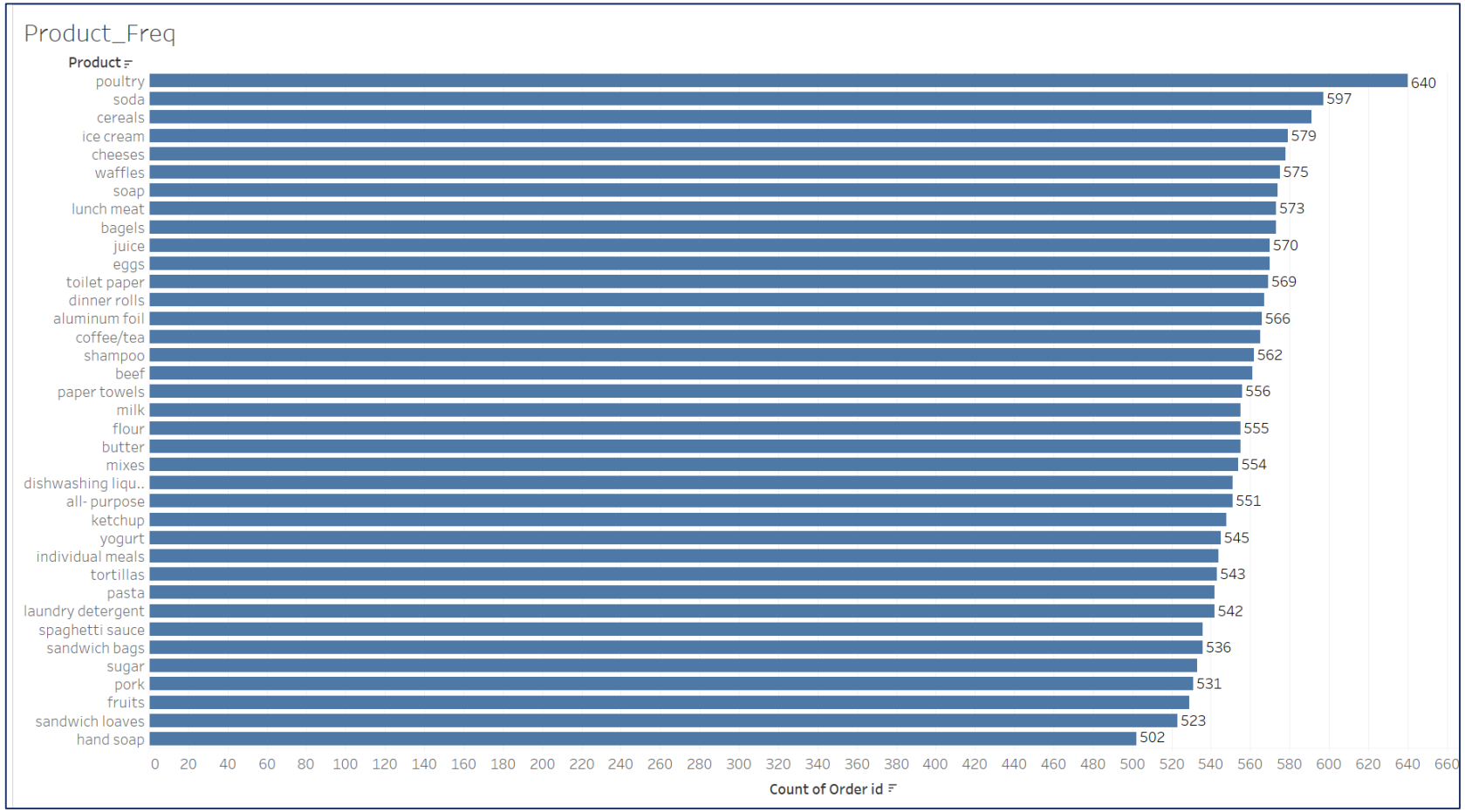
EDA_Jupyter Notebook

Exploratory Analysis & Inferences

- **Univariate Analysis :-**

1. 'Poultry' products appear to be most frequent content in the 640 Order IDs. Followed by 'Soda' and 'Cereals'.
2. Whereas, 'Hand Soap' appear to be least frequent content with a count of 502.

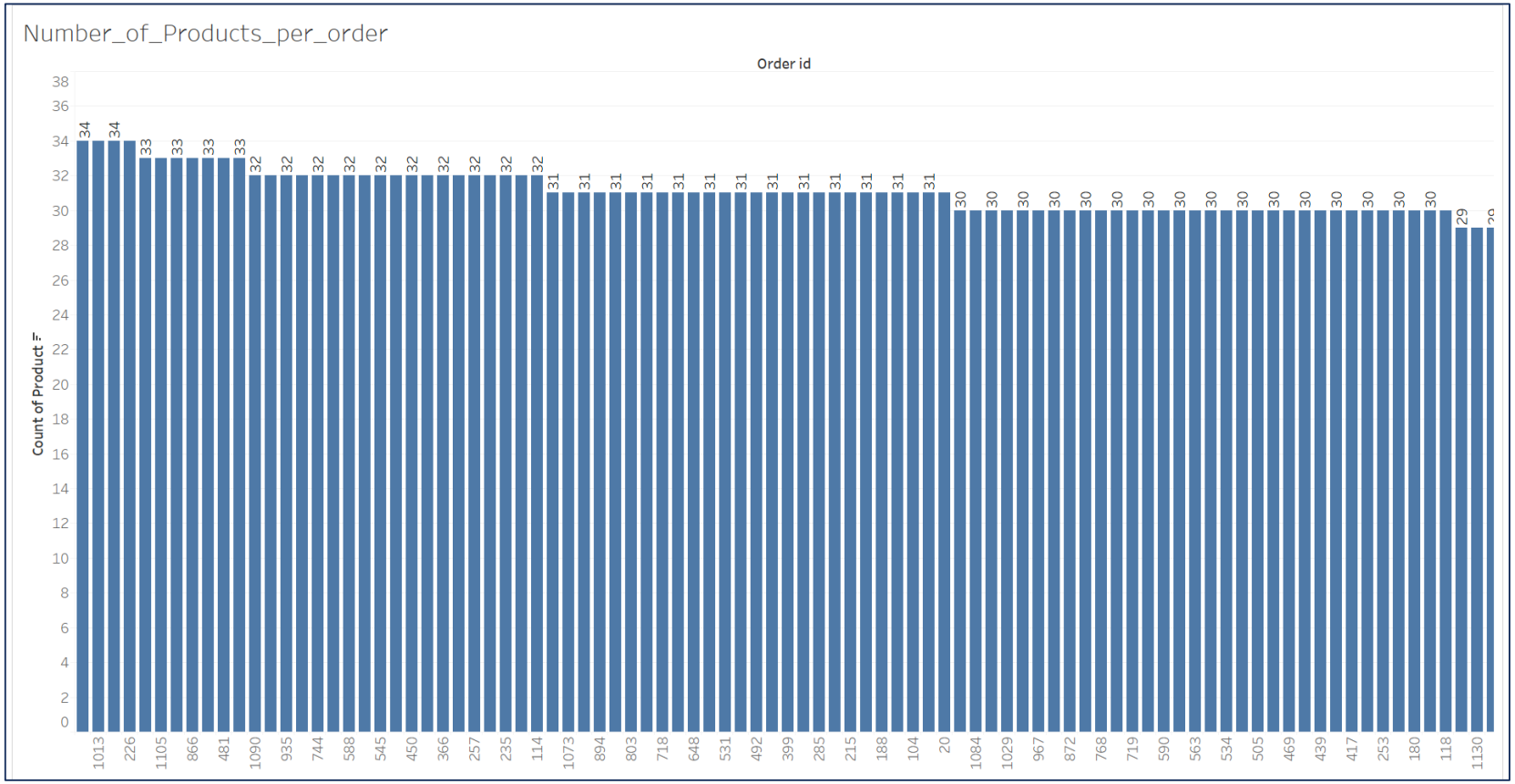
EDA_SalesTrend



Exploratory Analysis & Inferences

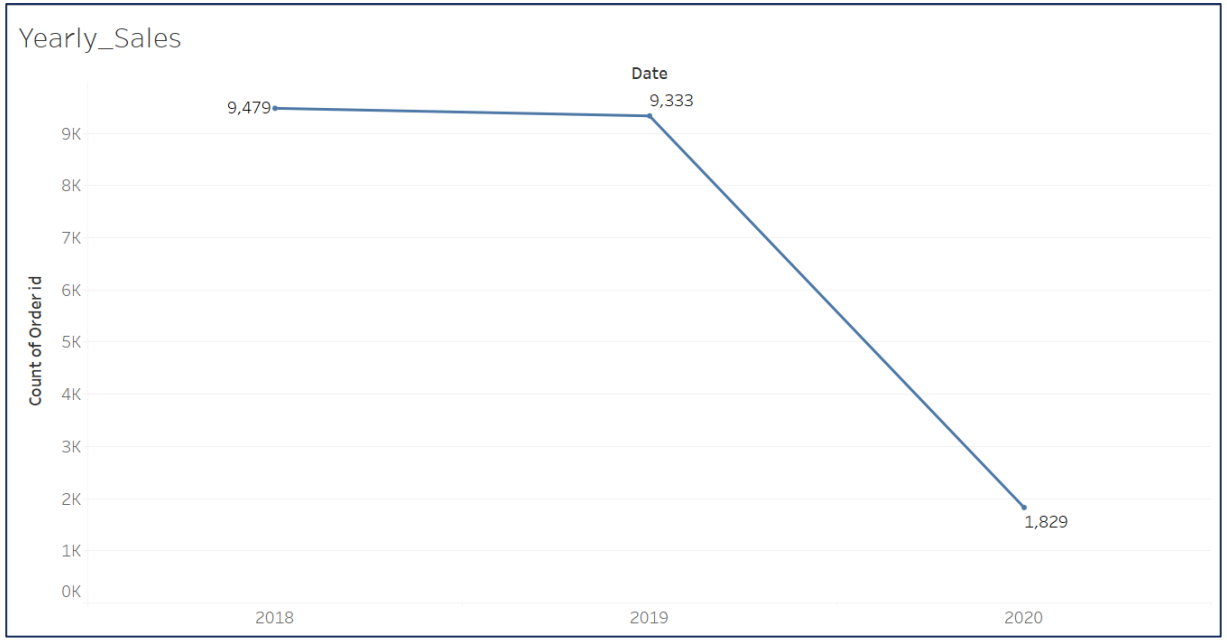
- **Univariate Analysis :-**

1. It seems order ID no. 1071, 1013, 957 and 226 have maximum number of unique products with a count of 34.



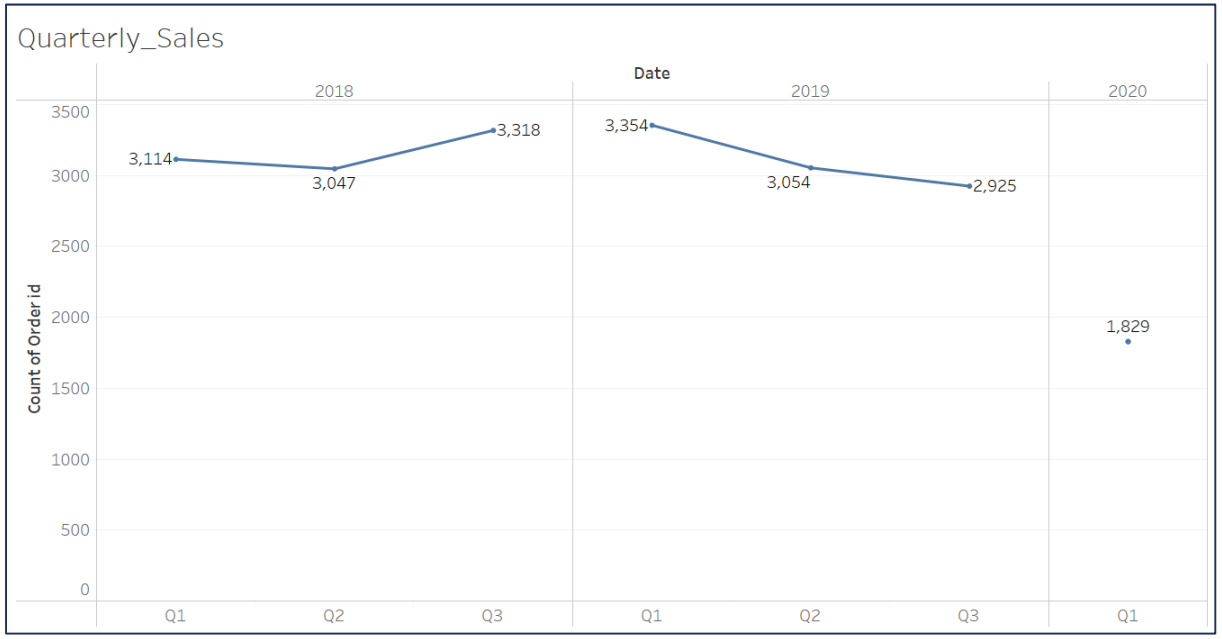
Sales Trends

- **Yearly Sales Trend:-**



1. There data comprises of 3 years ie. 2018, 2019 and 2020. But, if you drill down into the data, the data is not continuous as far as dates are concerned.
2. Year 2018 and 2019 consists only first 3 quarters of data, fourth quarter data is missing. Also, 2020 consists of only 2 months of data ie. Jan'20 and Feb'20.
3. In 2019, orders are slightly reduces in comparison with 2018.

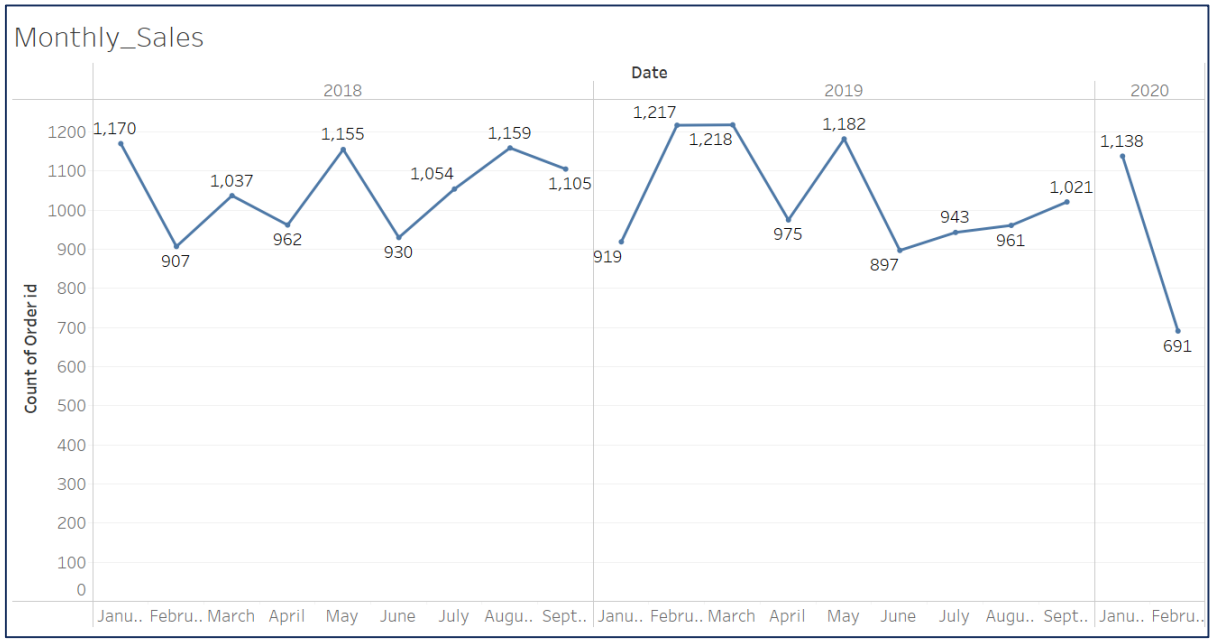
- **Quarterly Sales Trend:-**



1. There seems to be a upward sales trend in 2018 throughout the quarters.
2. There is decrease in the orders in 2019.
3. We can't comment for 2020 as data is not sufficient.

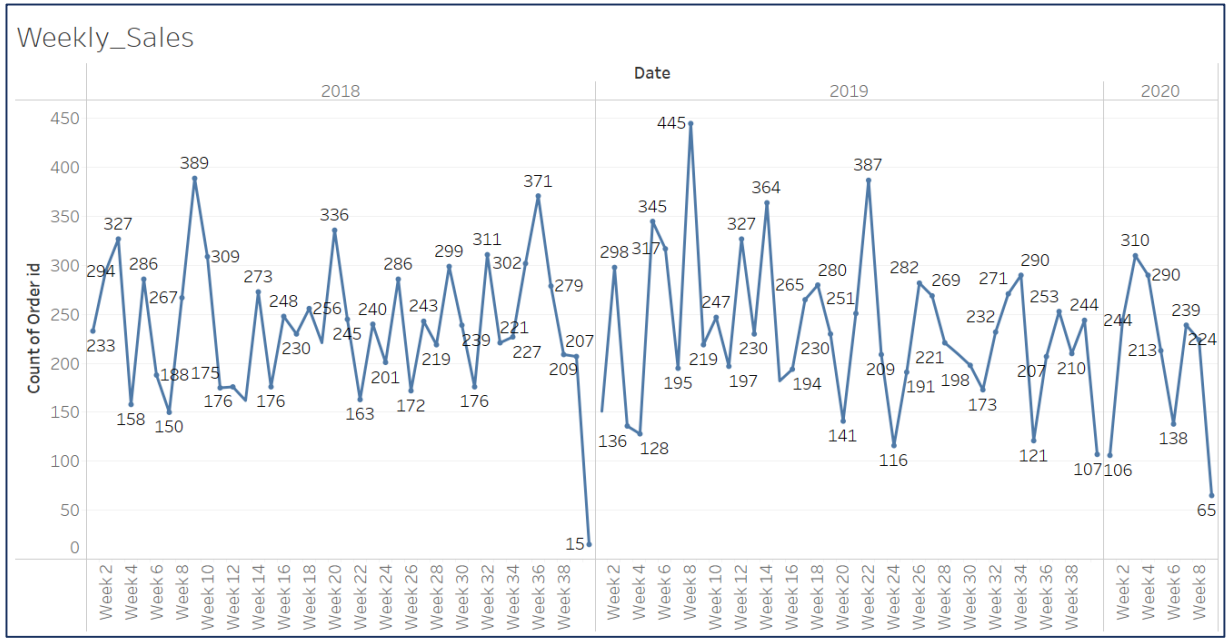
Sales Trends

- **Monthly Sales Trend:-**



1. In 2018, the monthly trend is upward but it is transformed to downtrend in 2019.

- **Weekly Sales Trend:-**



1. The weekly trend seems noisy.
2. Overall trend is most probability a downtrend.

Contents

EDA & Executive Summary of the Data

Use of Market Basket Analysis

Associations Identification

Recommendations & Offers

Use of Market Basket Analysis

- **Association Rules:**

- Association Rules are widely used to analyse retail basket or transaction data, and are intended to identify strong rules discovered in transaction data using measures of interestingness, based on the concept of strong rules.
- In this case given to us, association are found as below:
  - Customer buying spaghetti sauce & poultry products tends to buy dinner rolls also.
  - Customer buying dinner rolls & soda tends to buy eggs as well.
  - Similarly, customer buying eggs & soap also tend to buy soda.
- In the 1st case above, it make sense that a customer who is buying spaghetti sauce and poultry items would also like to buy dinner rolls for the dinner. So, just by positioning the dinner rolls next to poultry and sauce will encourage the customer to buy dinner rolls.

- Thus, it would increase the bucket price/sales.
- To make more sales out of it, we could create such combos and sale the same at some discounts or offers.

- **Following are the parameters which influence the association rules:**

1. Support: It is an indication of how frequent the item appears together in the dataset.

   Support = $P(X \cap Y)$

2. Confidence: It is an indication of how often the rule has been found to be true.

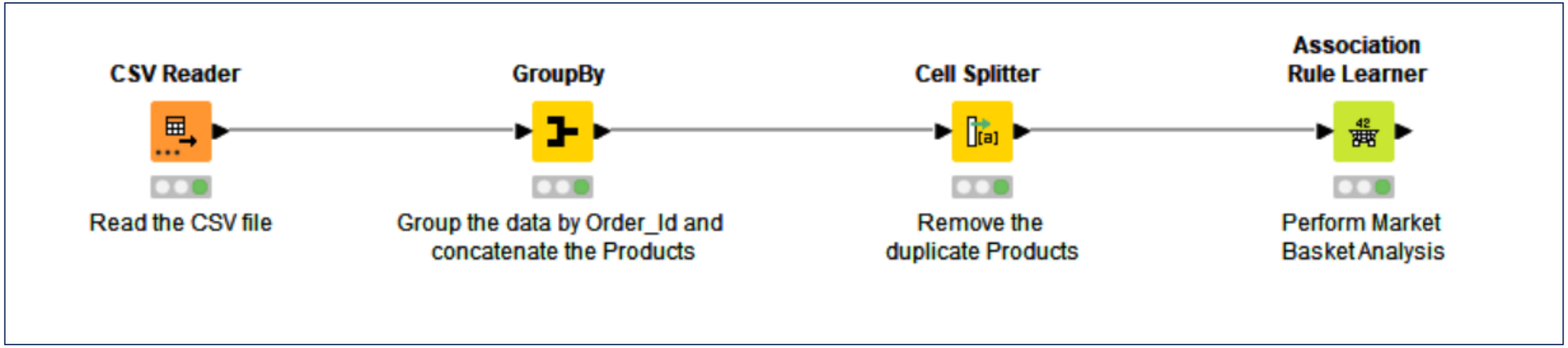
   Confidence = $P(Y \mid X)$

3. Lift: Greater lift values (>1) indicate stronger associations between 'X' & 'Y' items and they depend on one another.

   Lift = $P(Y \mid X) / P(Y)$

Use of Market Basket Analysis

- **KNIME Workflow:**



KNIME Workflow

Use of Market Basket Analysis

Significance of Threshold Values of Support & Confidence:

- The significance of threshold values of support & confidence is to extract a rule with a support value for a product that exceeds the minimum support and confidence value.
- The threshold for support in our analysis is kept at 0.09 whereas the threshold for confidence is kept at 0.55.
- This also means the associations rules are filtered by below rules:
  1. Product should appear together in at-least 9% of the transactions out of total number of transactions.
  2. Association rule must be found true at-least 55% of the times.
- These threshold values can differ from person to person, domain of study, datasets etc.
- In our case these threshold values had given good results and given 7 rules of association.

Parameters	Threshold Value
Support	0.09
Confidence	0.55

Contents

EDA & Executive Summary of the Data

Use of Market Basket Analysis

Associations Identification

Recommendations & Offers

Association Identification

- **Associations Identified in the study:**

Row ID	Support	Confidence	Lift	Consequent	implies	Items
rule5	0.099	0.579	1.49	dinner rolls	<---	[spaghetti sauce,poultry]
rule3	0.095	0.554	1.421	eggs	<---	[dinner rolls,soda]
rule2	0.092	0.553	1.414	soda	<---	[eggs,soap]
rule6	0.099	0.577	1.368	poultry	<---	[dinner rolls,spaghetti sauce]
rule1	0.091	0.562	1.334	poultry	<---	[dinner rolls,lunch meat]
rule0	0.09	0.557	1.321	poultry	<---	[dinner rolls,mixes]
rule4	0.096	0.556	1.32	poultry	<---	[juice,aluminum foil]

1. So, we have identified 7 number of good associations out of the total transactions.
2. Top association is basket of spaghetti sauce & poultry product with dinner rolls with a lift of 1.49
3. In most of the associations, poultry product seems to be most frequently appeared which implies this product would increase the sales if positioned near to the products in 'items' columns.
4. Tough the lift for dinner rolls is maximum, poultry products have high probability of selection by customer as it had appeared frequently (4 times). So it implies both dinner rolls & poultry are good products to promote with great offers and discounts.

Association Identification

- **Significance of support, confidence & lift:**

Row ID	D Support	D Confidence	D Lift	S Consequent	S implies	[...] Items
rule5	0.099	0.579	1.49	dinner rolls	<---	[spaghetti sauce,poultry]

1. So, will take above examples to demonstrate the significance of support, confidence and lift.
2. Support here implies that dinner rolls appeared in the 9.9% of the total transactions recorded.
3. Confidence implies that out of total customers who bought spaghetti & poultry, 57.9% of the total such customer also bought dinner rolls.
4. Lift greater than 1 implies that the presence of the antecedent increases the chances that the consequent will occur in a given transaction.

Contents

EDA & Executive Summary of the Data

Use of Market Basket Analysis

Associations Identification

Recommendations & Offers

Recommendations & Offers

- **Recommendations :-**

1. Grocery store should re-arranged in a fashion where dinner rolls are positioned near the sauce and poultry products.
2. Eggs should be positioned near the dinner rolls and soda.
3. Soap to be kept nearer to soda.
4. So, in above re-arrangement we have addressed to the top 3 association rules.

Recommendations & Offers

- **Offers :-**

1. Give 10% Off on Dinner Rolls when bought with sauce and poultry.
2. Poultry products are frequently associated with many baskets → Give coupon worth 50 bucks when customer buys dinner rolls or lunch meat or mixes or juice, the coupon can be redeemed on poultry product only.
3. Give half a dozen eggs free with dinner rolls and soda.
4. Combos of {Dinner rolls, poultry, sauce} with 15% off.

THANK YOU