# Marketing & Retail Analysis

Milestone 2 - Garima Gangwar

# 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.

# Executive Summary of the Data-

We have received the 2 years and 2 months data of a Grocery store. Consisting 20641 entries with 3 variable details regarding the demography of the transaction and item information.

# Contents-

- Problem Statement
- Data Summary
- EDA
- Trends
- Market Basket Analysis
- Knime and Workflow
- Association Identified
- Recommendation
- Offers

# Problem Statement-

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.

Grocery Store Data: Dataset_group.csv

| | 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 Summary-

- The data is about an Grocery store. They have provided the data collected of transactions for 2 years and 2 months.
- Pre-processing of the data is done in Jupyter notebook using the following libraries: Pandas, NumPy, Seaborn, Matplotlib
- The data has 20641 entries (0 To 20640) of rows and 3 columns. The data has 2 object data type and 1 integer data type.
- The dataset does not have null values.
- There are duplicates in the data. This may be a reason that customers are repurchasing the same product multiple times.
- In total there are 4730 duplicate values in the data.
- The data contains 603 unique values under Date column, 1139 under Order id and 37 under products.
- This means that 1139 orders were placed in total and the customers bought 37 unique items.

```
df.duplicated().sum()

/usr/local/lib/python
  and should_run_asyn
4730
```

```
df.isna().sum()

/usr/local/lib/pyt
  and should_run_a
Date        0
Order_id    0
Product     0
dtype: int64
```

```
df.nunique()

/usr/local/lib/pyt
  and should_run_a
Date         603
Order_id    1139
Product       37
dtype: int64
```

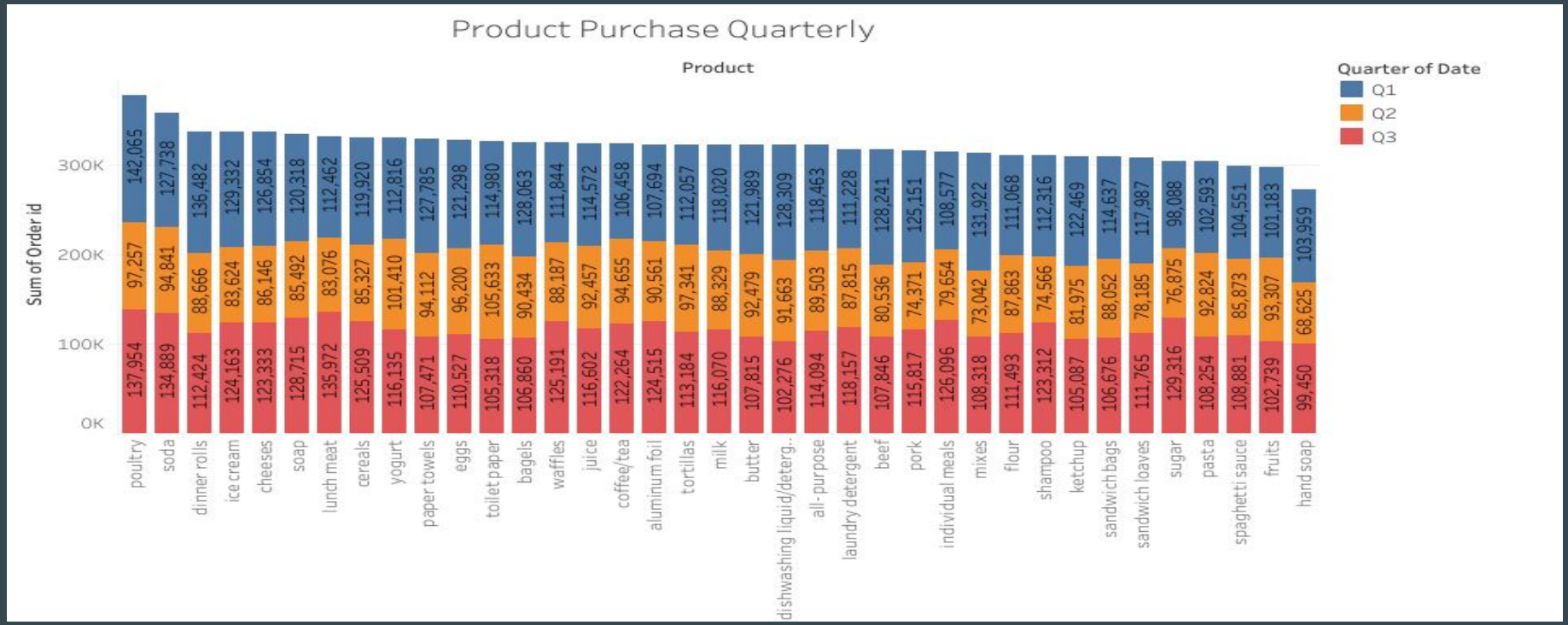
```
df.info()

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

# Exploratory Data Analysis-

Product Ordered-

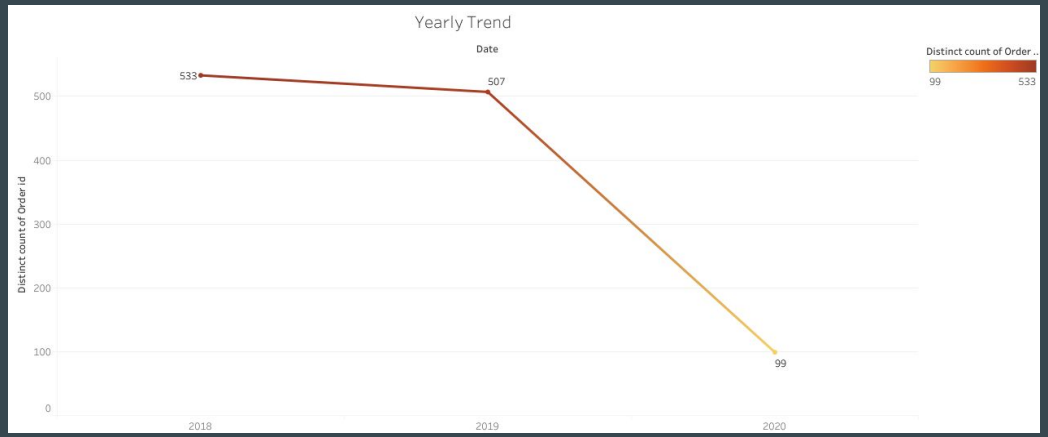

- It is evident that Poultry has been ordered the highest with 480 followed by ice cream with 454 orders.
- The lowest is hand soap 394 orders  and sandwich loaves with 398 orders.
- The milk, soap, coffee/tea, soda, cheese are more or less holds the same amount of orders.

# Product Purchased Quarterly for 2018, 2019 and 2020-



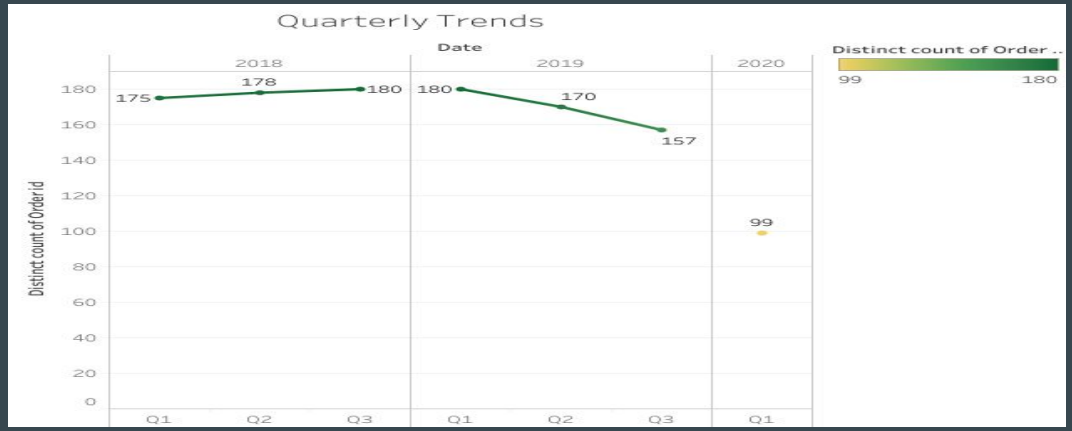

- The transaction report doesn't have data of 4th Quarter for each year .Otherwise the most transact year is 2019 Q3 followed by 2019 Q2.

# Yearly Orders and Its Trends-


- 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.
- The R-Squared value is - 0.79476
- P-Value is - 0.299317

# Quarterly Orders and Its Trends-

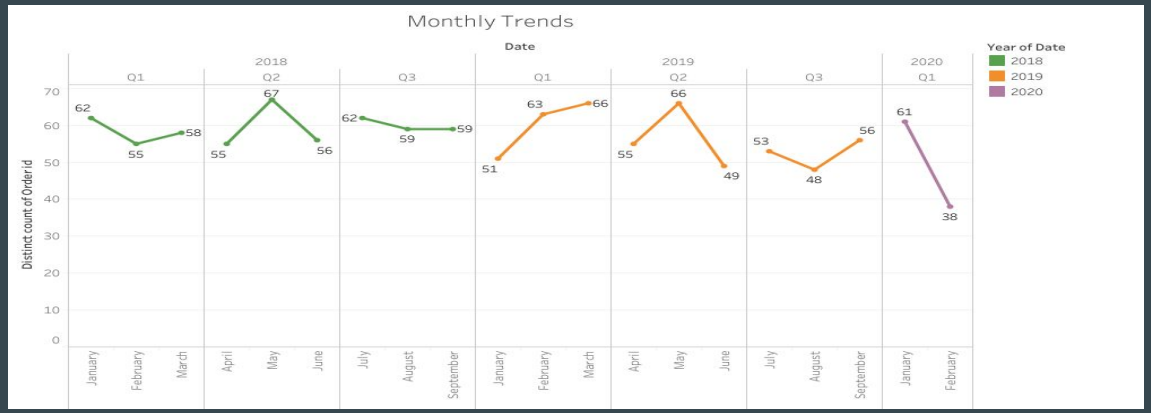

- In 2018, Q3 had the highest number of orders with 180.
- In 2019, Q1 had the highest with 180.
- This started decreasing as quarters passed by. The lowest was recorded in Q1 of 2020.
- Also, it is evident that Q4 of every year doesn't have any sales.
- No proper trend can be analysed in terms of Quarter sales.

➔ 2018-
R-squared – 0.986842
P- Value – 0.0731864

➔ 2019-
R-squared – 0.994361
P- Value – 0.0478513

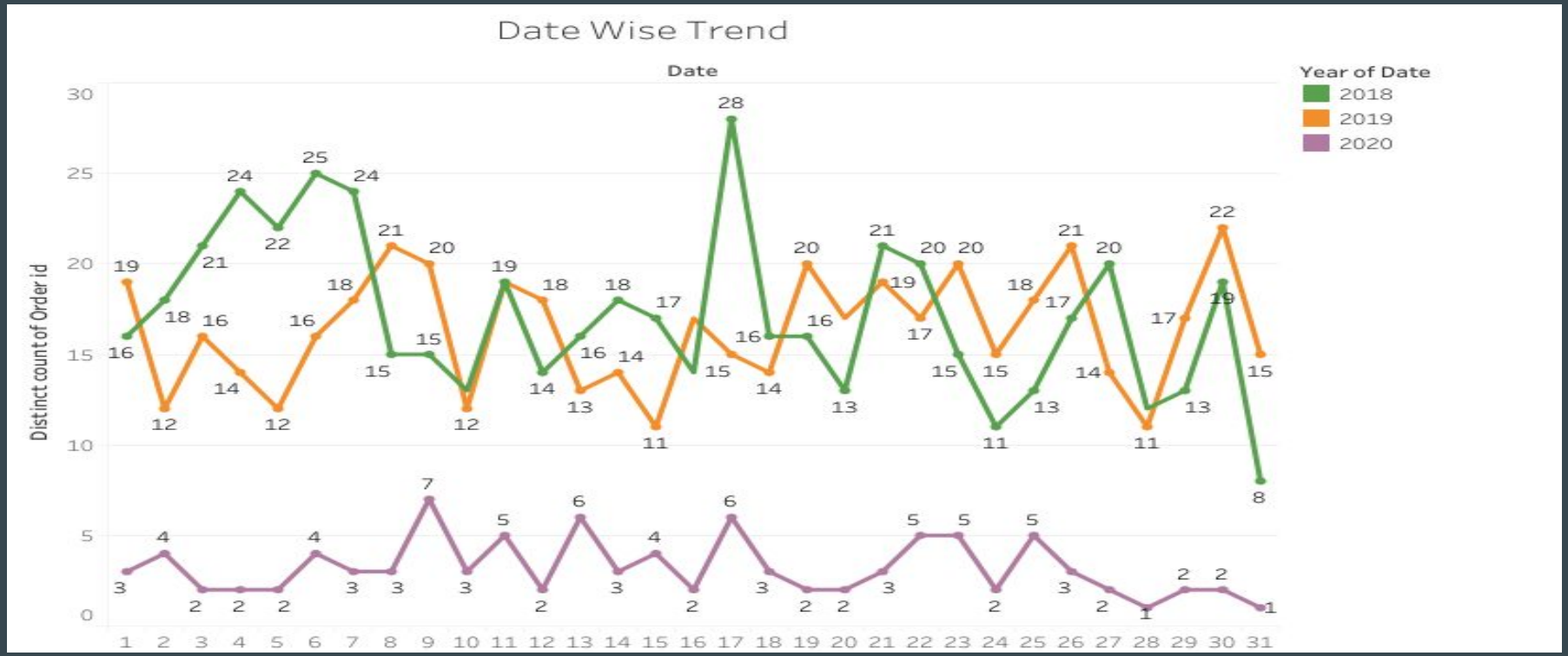
➔ 2020-
R-squared – 1
P- Value – N/A

# Monthly Orders and Its Trends-


- There is no trend and seasonality available in the data provided.
- January, February and May month has the highest numbers of sales.
- There was a drastic decrease in the year 2020 of February month.
- We can not determine the reason behind it as the records shows only the data for two months.

# Day Wise quantity Ordered-




- The highest distinct count of ordered date is 17.
- No proper trend can be analyzed using daily wise orders.

# Market Basket Analysis-

- Using Apriori Algorithm to do Market Basket Analysis of Customers purchasing behaviors. It can predict what the customer is going to buy next by looking at the products he is buying.
- Market Basket Analysis is a modelling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items. For example, you buy a loaf of bread and don't buy a milk, you are more likely to buy crisps (US. chips) or salad at the same time than somebody who didn't buy milk.
- Market Basket Analysis is one of the key techniques used by large relations to show associations between items. It allows retailers to identify relationships between the items that people buy together frequently.

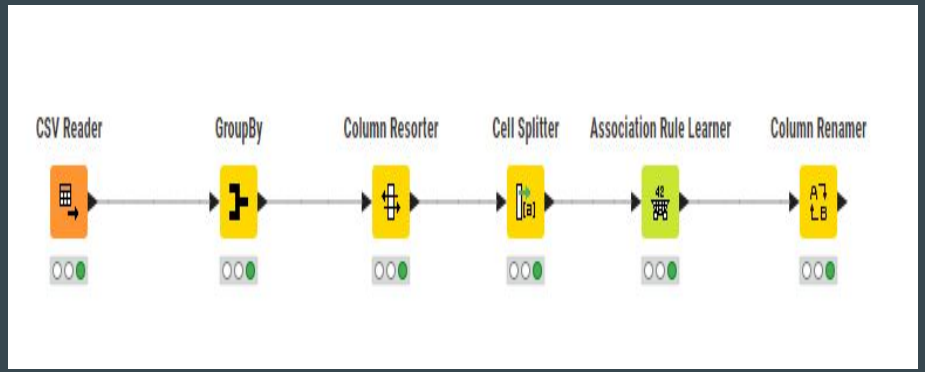
# Association Rules and its Relevance-

- Association Rule Mining is used when you want to find an association between different objects in a set, find frequent patterns in a transaction database, relational databases or any other information repository.
- Support, confidence level, and lift are the three measures used to compute it.
- High-support rules are more likely to apply to a significant number of future transactions.
- You can expect higher return rates if you have more confidence.
- The strength of the relationship between the products on the antecedents and consequents columns of the association is summarised by lift.
- Larger the lift, greater the link between the two products.

★ When you apply Association Rule Mining on a given set of transactions T your goal will be to find all rules with:

- Support greater than or equal to min_support
- Confidence greater than or equal to min_confidence
- There are 20,642 records in the grocery dataset, totalling 1139 orders.
- Customers have purchased numerous goods in a single order, with an average of 18 products per buy.
- As a result, it is critical to investigate the association between two items.
- Once the relationship has been assessed, actions can be taken based on the confidence and lift indicators to boost sales by offering combos and gift packs.

# Knime workflow and output table image-




- KNIME has been used in order to understand different association rules.
- The data has been grouped according to the order ID.
- The tabular representation beside represents the function of cell splitter.

File Edit Hilite Navigation View

Table "default" - Rows: 1139 | Spec - Columns: 4 | Properties | Flow Variables

| Row ID | Order_id | Date | Product | Product_SplitResultSet |
|---|---|---|---|---|
| Row0 | 1 | 01-01-2018 | yogurt, pork, sandwich bags, lunch meat, all- purpose, flour, soda, butter, beef, aluminum foil, all- purpose, din... | [yogurt,pork,sandwich bags,... |
| Row1 | 2 | 01-01-2018 | toilet paper, shampoo, hand soap, waffles, cheeses, mixes, milk, sandwich bags, laundry detergent, dishwashi... | [toilet paper,shampoo,hand s... |
| Row2 | 3 | 02-01-2018 | soda, pork, soap, ice cream, toilet paper, dinner rolls, hand soap, spaghetti sauce, milk, ketchup, sandwich loav... | [soda,pork,soap,...] |
| Row3 | 4 | 02-01-2018 | cereals, juice, lunch meat, soda, toilet paper, all- purpose | [cereals,juice,lunch meat,...] |
| Row4 | 5 | 02-01-2018 | sandwich loaves, pasta, tortillas, mixes, hand soap, toilet paper, paper towels, flour, pork, poultry, eggs, pork,... | [sandwich loaves,pasta,tortill... |
| Row5 | 6 | 02-01-2018 | laundry detergent, toilet paper, eggs, toilet paper, bagels, dishwashing liquid/detergent, cereals, paper towels,... | [laundry detergent,toilet pap... |
| Row6 | 7 | 03-01-2018 | individual meals, paper towels, tortillas, milk, ice cream, juice, dishwashing liquid/detergent, soap, sandwich bag... | [individual meals,paper towel... |
| Row7 | 8 | 04-01-2018 | ice cream, juice, paper towels, waffles, soda, cheeses, poultry, toilet paper | [ice cream,juice,paper towels... |
| Row8 | 9 | 04-01-2018 | juice, poultry, coffee/tea, coffee/tea, dishwashing liquid/detergent | [juice,poultry,coffee/tea,...] |
| Row9 | 10 | 05-01-2018 | ketchup, coffee/tea, toilet paper, pork, flour, milk, soda, dishwashing liquid/detergent, eggs, tortillas, tortillas, ... | [ketchup,coffee/tea,toilet pa... |
| Row10 | 11 | 05-01-2018 | sandwich loaves, ice cream, soda, bagels, dishwashing liquid/detergent, eggs, sugar, waffles, individual meals, ... | [sandwich loaves,ice cream,s... |
| Row11 | 12 | 06-01-2018 | pork, tortillas, pork, shampoo, lunch meat, pasta, juice, bagels, bagels, laundry detergent, yogurt, sugar, waffl... | [pork,tortillas,shampoo,...] |
| Row12 | 13 | 07-01-2018 | sugar, fruits, all- purpose, aluminum foil, laundry detergent, individual meals, flour, pork, shampoo, sugar, alumi... | [sugar,fruits,all- purpose,...] |
| Row13 | 14 | 07-01-2018 | fruits, dinner rolls, individual meals, shampoo, ketchup, cereals, sandwich bags, laundry detergent, sandwich lo... | [fruits,dinner rolls,individual ... |
| Row14 | 15 | 07-01-2018 | individual meals, ice cream, cereals, paper towels, bagels, mixes, lunch meat, juice, toilet paper, cheeses, pape... | [individual meals,ice cream,ce... |
| Row15 | 16 | 08-01-2018 | sugar, sandwich bags, flour, juice, milk, paper towels, cereals, sandwich bags, pasta, soda, dishwashing liquid/... | [sugar,sandwich bags,flour,...] |
| Row16 | 17 | 08-01-2018 | milk, hand soap, pasta, individual meals, spaghetti sauce, cereals, sandwich loaves, hand soap, individual meals | [milk,hand soap,pasta,...] |
| Row17 | 18 | 08-01-2018 | sandwich bags, toilet paper, bagels, shampoo, coffee/tea | [sandwich bags,toilet paper,... |
| Row18 | 19 | 09-01-2018 | individual meals, laundry detergent, coffee/tea, eggs, aluminum foil, beef, juice, flour, sugar, individual meals, ... | [individual meals,laundry det... |
| Row19 | 20 | 10-01-2018 | shampoo, dishwashing liquid/detergent, yogurt, juice, sugar, soap, sandwich loaves, butter, sandwich loaves, ... | [shampoo,dishwashing liquid/... |
| Row20 | 21 | 11-01-2018 | waffles, fruits, all- purpose, pork, juice, bagels, mixes | [waffles,fruits,all- purpose,...] |
| Row21 | 22 | 11-01-2018 | cheeses, cereals, sugar, bagels, soda | [cheeses,cereals,sugar,...] |
| Row22 | 23 | 11-01-2018 | aluminum foil, bagels, shampoo, shampoo, dishwashing liquid/detergent, cereals, cheeses, flour, lunch meat, pa... | [aluminum foil,bagels,shampo... |
| Row23 | 24 | 11-01-2018 | fruits, all- purpose, pasta, cheeses, juice, sandwich bags, sandwich loaves, coffee/tea, juice, lunch meat, spag... | [fruits,all- purpose,pasta,...] |
| Row24 | 25 | 11-01-2018 | bagels, sugar, pork, sandwich loaves, tortillas, ice cream, all- purpose, yogurt, fruits, pasta, lunch meat, dishw... | [bagels,sugar,pork,...] |
| Row25 | 26 | 12-01-2018 | fruits, sandwich loaves, coffee/tea, aluminum foil, shampoo, cereals, dinner rolls, coffee/tea, poultry, butter, j... | [fruits,sandwich loaves,coffe... |
| Row26 | 27 | 13-01-2018 | laundry detergent, pork, pasta, cheeses, fruits, sugar, lunch meat, laundry detergent, paper towels, butter, s... | [laundry detergent,pork,past... |
| Row27 | 28 | 13-01-2018 | pork, bagels, poultry, pasta, butter, all- purpose, pasta, shampoo, sugar, ketchup, eggs, soda, tortillas, soap, ... | [pork,bagels,poultry,...] |
| Row28 | 29 | 13-01-2018 | pasta, butter, sandwich loaves, spaghetti sauce, juice, dinner rolls, all- purpose, pork, yogurt, spaghetti sauce | [pasta,butter,sandwich loave... |
| Row29 | 30 | 14-01-2018 | flour, bagels, cheeses, sandwich loaves, toilet paper | [flour,bagels,cheeses,...] |
| Row30 | 31 | 15-01-2018 | aluminum foil, eggs, ice cream, pasta, juice, waffles, shampoo, dinner rolls, soda, laundry detergent, hand soa... | [aluminum foil,eggs,ice cream... |
| Row31 | 32 | 15-01-2018 | soap, paper towels, individual meals, dinner rolls, lunch meat, sugar, soap, eggs, beef, yogurt, soda, sandwich ... | [soap,paper towels,individual... |
| Row32 | 33 | 15-01-2018 | sandwich loaves, pork, sandwich bags, ketchup, coffee/tea, soda, poultry, pasta, all- purpose, coffee/tea, pap... | [sandwich loaves,pork,sandw... |
| Row33 | 34 | 16-01-2018 | lunch meat, mixes, soap, hand soap, tortillas, coffee/tea, cheeses, tortillas, cheeses, ice cream, tortillas, beef, ... | [lunch meat,mixes,soap,...] |
| Row34 | 35 | 16-01-2018 | aluminum foil, cheeses, cereals, mixes, laundry detergent, juice, pork, bagels, sandwich loaves, yogurt, milk, sa... | [aluminum foil,cheeses,cereal... |
| Row35 | 36 | 17-01-2018 | soap, lunch meat, flour, juice, yogurt, shampoo, spaghetti sauce, ketchup, laundry detergent, dinner rolls, ketc... | [soap,lunch meat,flour,...] |
| Row36 | 37 | 17-01-2018 | fruits, lunch meat, dinner rolls, shampoo, hand soap, mixes, pasta, sugar | [fruits,lunch meat,dinner rolls... |
| Row37 | 38 | 17-01-2018 | eggs, hand soap, dinner rolls, hand soap, lunch meat, laundry detergent, pasta, aluminum foil, toilet paper, cer... | [eggs,hand soap,dinner rolls,... |

# Threshold Values of lift values, support and confidence-

- Support and Confidence measure how interesting the rule is. It is set by the minimum support and minimum confidence thresholds.
- These thresholds set by client help to compare the rule strength according to your own or client's will.
- The closer to threshold the more the rule is of use to the client.
- **Frequent Item sets:**
  Item-sets whose support is greater or equal than minimum support threshold (min_sup).

★ **Strong rules:**
If a rule A=>B[Support, Confidence] satisfies min_sup and min_confidence then it is a strong rule.

★ **Lift:**
Lift gives the correlation between A and B in the rule A=>B.
Correlation shows how one item-set A effects the item-set B.
Lift(A=>B)=Support Supp(A)Supp(B)

★ A rule may appear to have a strong association in a data collection because it appears frequently, but it may emerge much less frequently when applied.

# Lift values, Support and Confidence-

- There is no particular threshold value of support. If the dataset is large then it is advisable to set the value at 10% and increase it accordingly till required number of associations are generated.
- The threshold value of support in this case is 10%. Association rules are not developing if the value goes beyond 10%
- Also, the confidence is kept at 40% which is optimal level.
- The associations with highest lift values are considered to be accepted more.
- In the following case, the association rule is formed highest lift value that is 12.34% at a confidence level of 46% and support of 18%.
- This rule implies that the customers who bought dishwashing also purchased mixes.
- Also the customers who bought soda have high chances of buying eggs.

# Association rule in Tabular-

| Row ID | Support | Confide... | Lift | Conseq... | implies | Items |
|---|---|---|---|---|---|---|
| rule893 | 0.18 | 0.464 | 1.234 | mixes | <--- | [dishwashing liquid/detergent] |
| rule894 | 0.18 | 0.479 | 1.234 | dishwashing ... | <--- | [mixes] |
| rule909 | 0.187 | 0.48 | 1.228 | soda | <--- | [eggs] |
| rule910 | 0.187 | 0.479 | 1.228 | eggs | <--- | [soda] |
| rule787 | 0.17 | 0.452 | 1.226 | shampoo | <--- | [juice] |
| rule788 | 0.17 | 0.462 | 1.226 | juice | <--- | [shampoo] |
| rule825 | 0.172 | 0.457 | 1.224 | spaghetti sa... | <--- | [juice] |
| rule826 | 0.172 | 0.461 | 1.224 | juice | <--- | [spaghetti sauce] |
| rule648 | 0.164 | 0.453 | 1.219 | pasta | <--- | [paper towels] |
| rule649 | 0.164 | 0.442 | 1.219 | paper towels | <--- | [pasta] |
| rule871 | 0.176 | 0.469 | 1.218 | yogurt | <--- | [juice] |
| rule872 | 0.176 | 0.459 | 1.218 | juice | <--- | [yogurt] |
| rule493 | 0.16 | 0.457 | 1.217 | individual meals | <--- | [sandwich loaves] |
| rule494 | 0.16 | 0.425 | 1.217 | sandwich loa... | <--- | [individual meals] |
| rule821 | 0.172 | 0.457 | 1.207 | toilet paper | <--- | [juice] |
| rule822 | 0.172 | 0.455 | 1.207 | juice | <--- | [toilet paper] |
| rule630 | 0.164 | 0.442 | 1.202 | sandwich bags | <--- | [ketchup] |
| rule631 | 0.164 | 0.446 | 1.202 | ketchup | <--- | [sandwich bags] |
| rule879 | 0.177 | 0.461 | 1.199 | yogurt | <--- | [aluminum foil] |
| rule880 | 0.177 | 0.461 | 1.199 | aluminum foil | <--- | [yogurt] |
| rule831 | 0.173 | 0.466 | 1.197 | dinner rolls | <--- | [pasta] |
| rule832 | 0.173 | 0.445 | 1.197 | pasta | <--- | [dinner rolls] |
| rule680 | 0.166 | 0.448 | 1.195 | beef | <--- | [fruits] |
| rule681 | 0.166 | 0.443 | 1.195 | fruits | <--- | [beef] |
| rule881 | 0.177 | 0.472 | 1.195 | lunch meat | <--- | [individual meals] |
| rule882 | 0.177 | 0.449 | 1.195 | individual meals | <--- | [lunch meat] |
| rule387 | 0.158 | 0.452 | 1.192 | soap | <--- | [sandwich loaves] |
| rule388 | 0.158 | 0.417 | 1.192 | sandwich loa... | <--- | [soap] |
| rule815 | 0.172 | 0.458 | 1.191 | aluminum foil | <--- | [mixes] |
| rule816 | 0.172 | 0.447 | 1.191 | mixes | <--- | [aluminum foil] |
| rule913 | 0.195 | 0.462 | 1.189 | dinner rolls | <--- | [poultry] |
| rule914 | 0.195 | 0.501 | 1.189 | poultry | <--- | [dinner rolls] |
| rule817 | 0.172 | 0.457 | 1.188 | aluminum foil | <--- | [juice] |

- The output has been analysed in the form of descending order of Lift, as higher the list higher will be the association.

# Recommendations-

- From the analysis, one can conclude that the number of orders has been falling drastically over the years.
- The highest number of orders was in 2018 and then followed by 2019.
- The Q4 data has not been provided or can conclude that there are no orders placed in Q4.
- The store can provide discounts in the middle of the month in order to attract more customers.
- Poultry and ice-creams are most frequently ordered items. It is advisable to increase the variety of items in this category so that customers will have a lot to choose.
- The months of mid quarters have been showing a consistent performance over the years.
- Also one can conclude that the numbers of orders are high in starting and end days of the months.
- This can be because most of the customers get their pay either at the starting or at the end of every month.
- Hand soaps and loaves are least purchased items. So the store can invest a bit less on these items.
- In order to increase the sales, Q4 is crucial part for many businesses as it is festive season customers tend to order a lot.
- It is advisable to provide the service in Q4 as well.
- The decrease in the orders over the years can be of many reasons like poor customer service or no proper offers provided. It is recommended to look into the service and provide customers the highest satisfaction in terms of service as well as products.

# Offers-

- Also most of the customers are inclined to buy ketchups along with sandwich bags, the store can provide an offer saying get 1 ketch up free with 2 sandwich bags
- An offer like buy 1 kg of beef and get 1 kg of assorted fruits free can be provided to the customers.
- Since hand soaps are least preferred by customers they can be combined with any of the highest selling product.
- Hand soaps can be combined with detergents to increase the purchase of hand soaps as well.
- Sandwich loaves if also one of the least selling products. It is evident that people who buy loaves also buy individual meals. So the store can provide combo offer like buy three meals and get 5 sandwich loaves free.
- In order the increase the sales, the grocery store can provide combo offers to its customers.
- Top 5 combos according to Market Basket Analysis are:

➔ Mixed & Dishwashing liquid
➔ Eggs & Soda
➔ Juice & Shampoo
➔ Juice and Spaghetti sauce
➔ Pasta & Paper towels

## Tableau Public Link-

Garima_MRA_Milestone2 | Tableau Public

# Thank You