Marketing and Retail Analytics

By,
Akshaya Nallathambi
23rd January 2022

Contents of the Presentation

- Problem Statement
- Aim of the project
- Data Dictionary
- Data Summary
- Inference about the data
- Exploratory Data Analysis (EDA)
- Inference about EDA
- Market Basket Analysis (Association Rules)
- Recommendations

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. Can you suggest the best combos & offers?

Aim of the project

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.

Data Dictionary

Column Name	Description
Date	Order Date
Order_id	Order ID
Product	The product name

Data Summary

	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

This table gives the first 5 rows of sample data.

```
<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
```

This table gives the basic information of the data set. It is clear that there are variables of type int and object with 3 columns and 20641 rows. There are no null values. The memory usage is 483.9+ KB.

	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

This table gives the five point summary of the continuous variables in the data set. This tells how the data is distributed.

Inferences:

- The data is about the transactional data of a Grocery store. They have provided over 3 years of data.
- The given data set has variables of type int and object with 3 columns and 20641 rows. There are no null values. The memory usage is 483.9+ KB.
- This data gives the purchasing behavior of customers in different categories. The grocery store sells different product like 'yogurt', 'pork', 'sandwich bags', 'lunch meat', 'all-purpose', 'flour', 'soda', 'butter', 'beef', 'aluminum foil', 'dinner rolls', 'shampoo', 'mixes', 'soap', 'laundry detergent', 'ice cream', 'toilet paper', 'hand soap', 'waffles', 'cheeses', 'milk', ‘dishwashing liquid/detergent', 'individual meals', 'cereals', 'tortillas', 'spaghetti sauce', 'ketchup', 'sandwich loaves', 'poultry', 'bagels', 'eggs', 'juice', 'pasta', 'paper towels', 'coffee/tea', 'fruits' and 'sugar'.
- Each transaction has an Order ID and for each order number maintained required information like product name and Order Date. Also it can be identified that one Order ID has many entries with different product name.

Exploratory Data Analysis

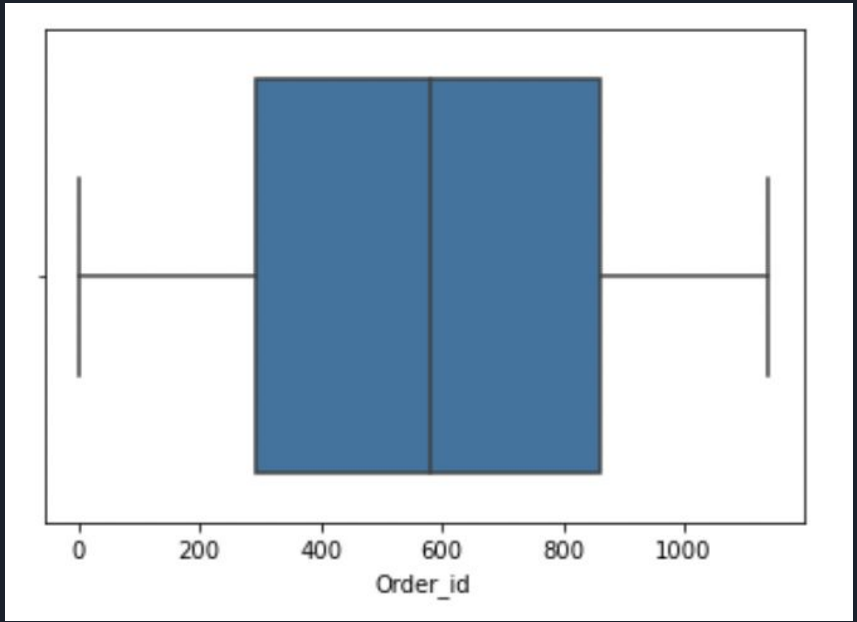
Exploratory Data Analysis (EDA) is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

Box plots help in identifying the outliers in the data

Univariate analysis: It is the simplest form of analyzing data. “Uni” means “one”, so in other words your data has only one variable. The histograms are used for numerical variables and bar plot is used for categorical variable to perform univariate analysis.

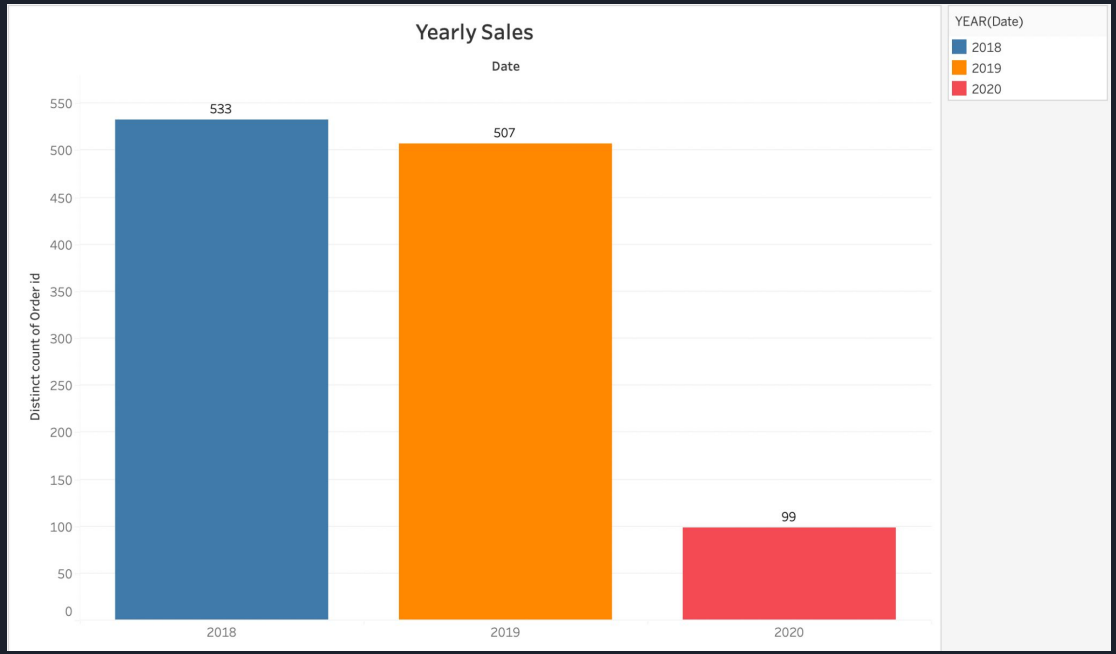
Bivariate analysis: It is the simplest forms of quantitative analysis. It involves the analysis of two variables, for the purpose of determining the empirical relationship between them. The pairplot is generally used for numerical variables to perform bivariate analysis.

Box Plots

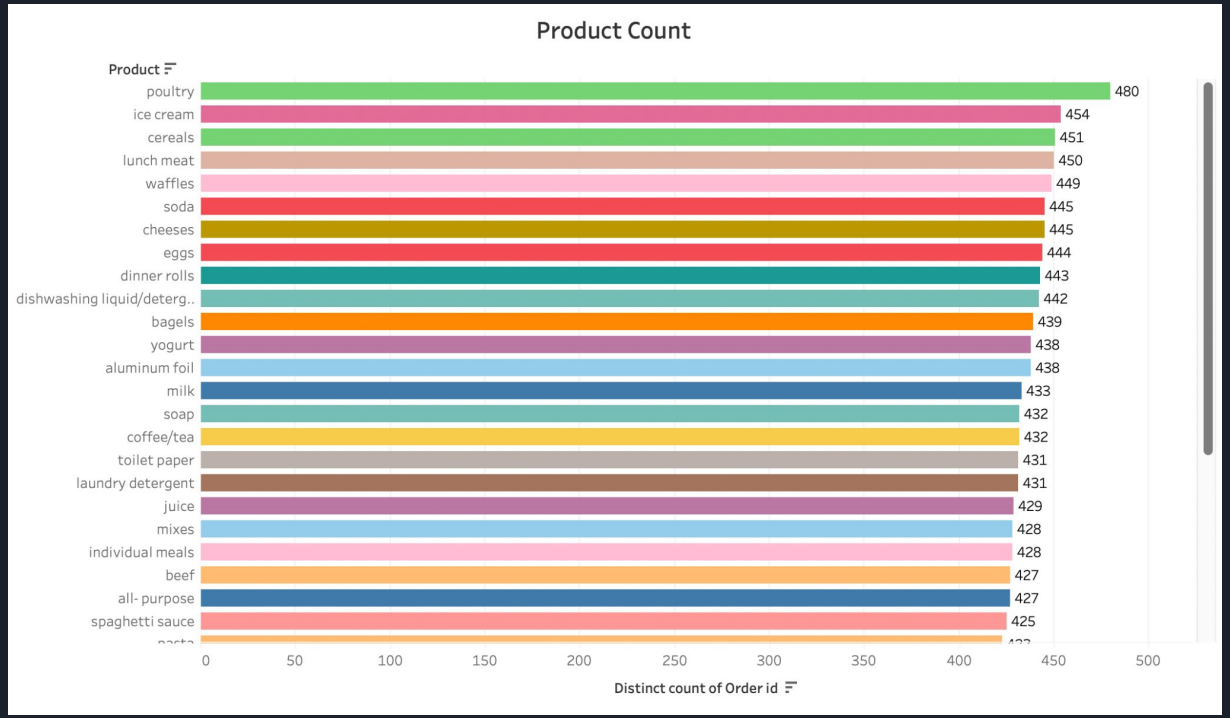


Outliers can be identified using the box plots. It is clear that there are no outliers in the given data set.

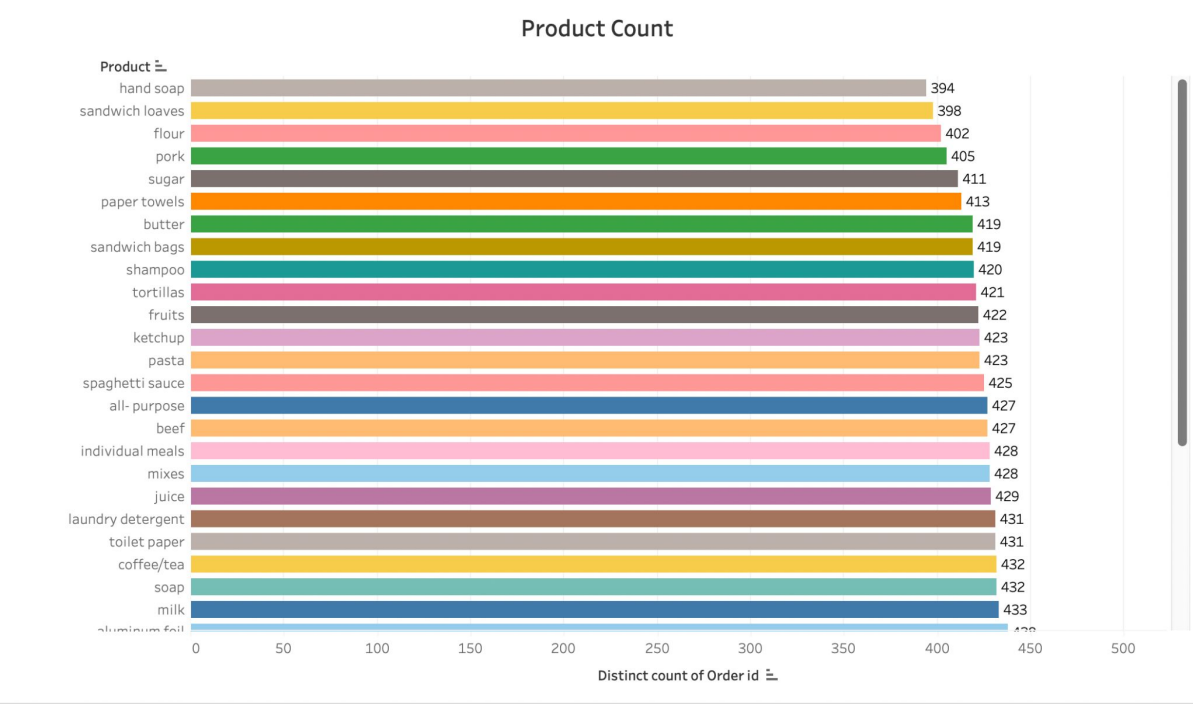
Univariate analysis - Bar Plot



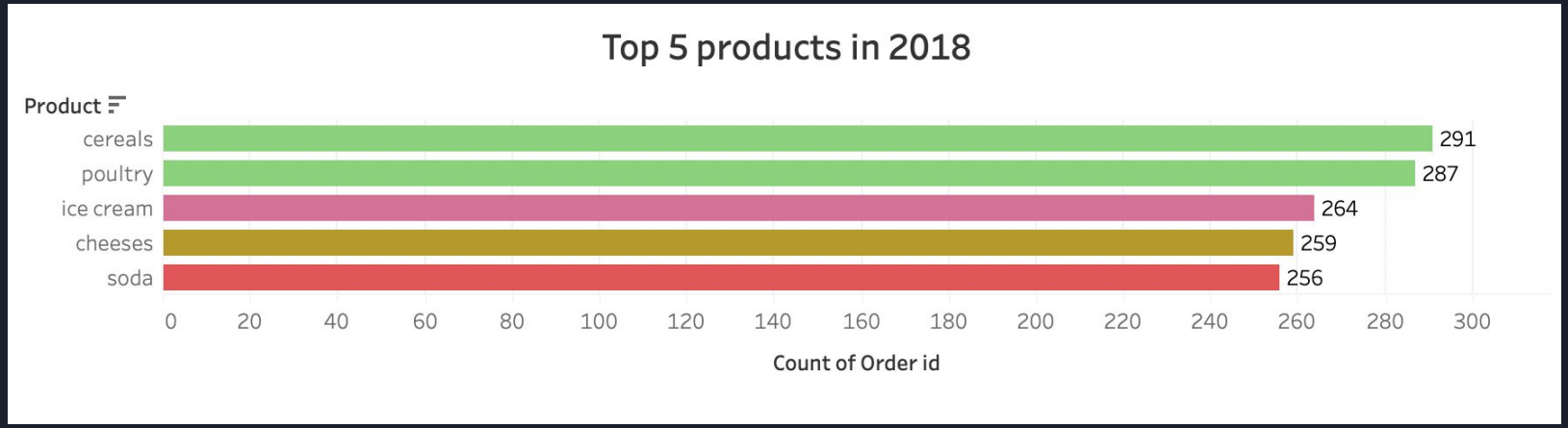
The year 2018 had the most sales. The year 2020 might seem like the least sales but only two months of data is available for this year.



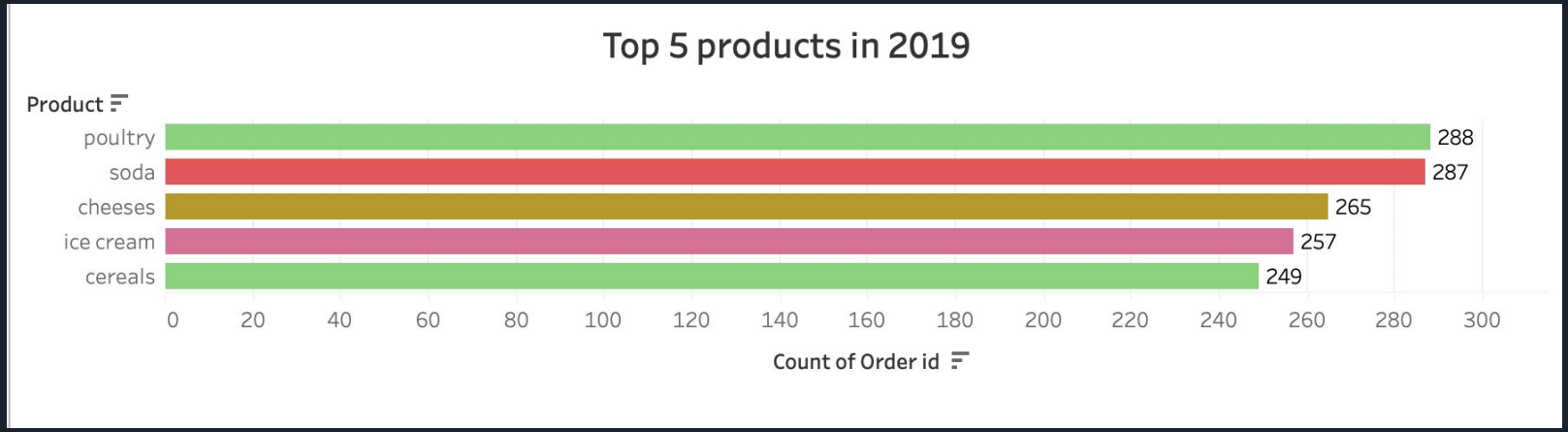
Poultry is the most purchased product followed by ice cream, cereals, etc.,



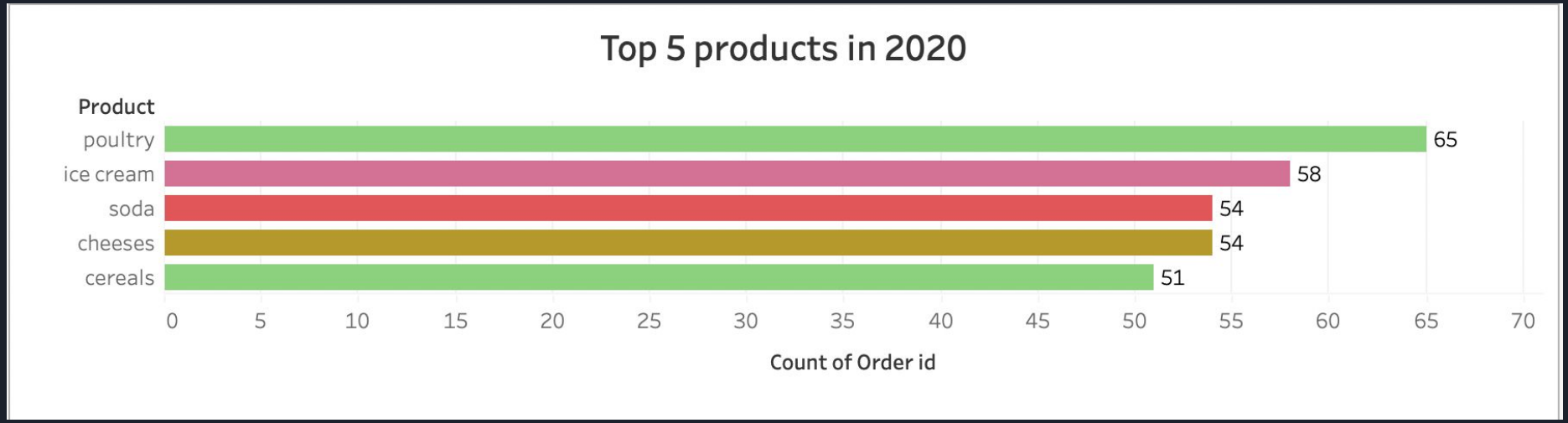
Hand soap is the least purchased product followed by sandwich loaves, flour, etc.,



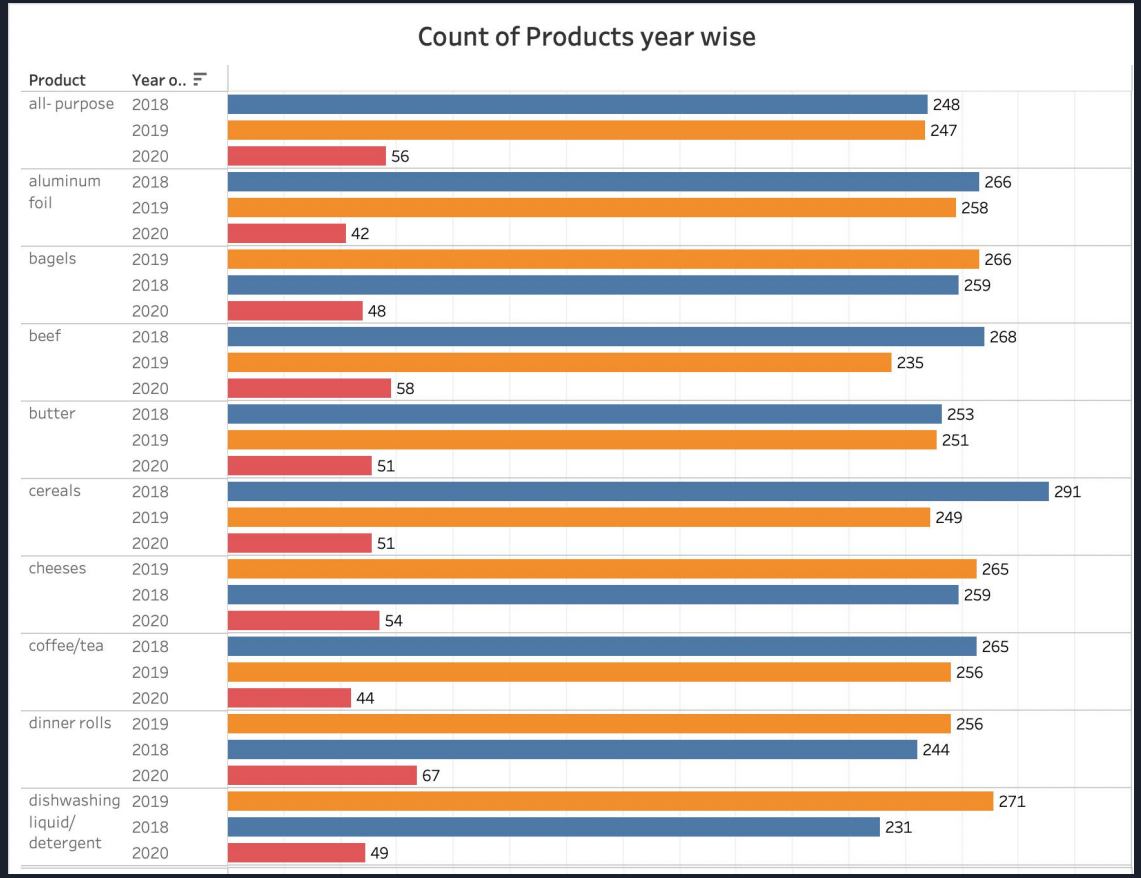
This graph shows the top 5 products purchased by the customers in 2018.



This graph shows the top 5 products purchased by the customers in 2019.

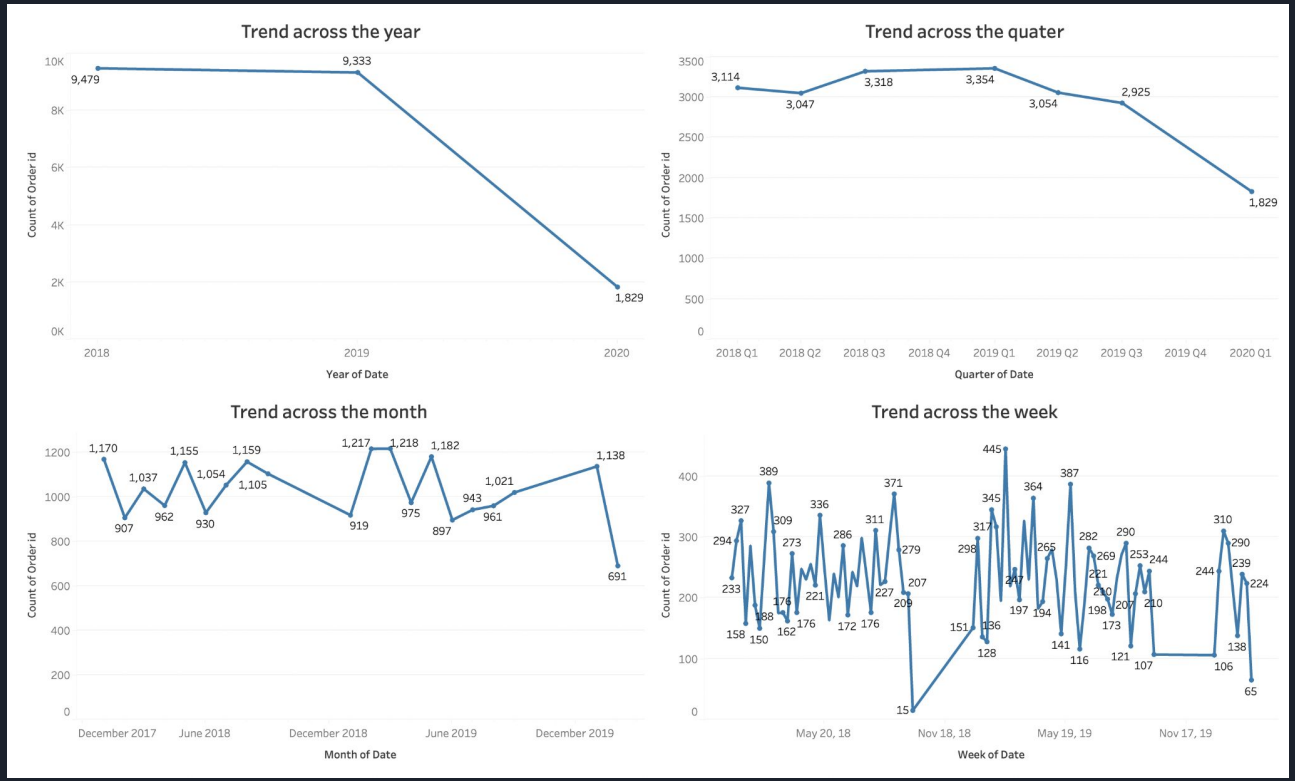


This graph shows the top 5 products purchased by the customers in 2020.



This graph gives an overview of each product purchase year wise.

Time series - Line Chart

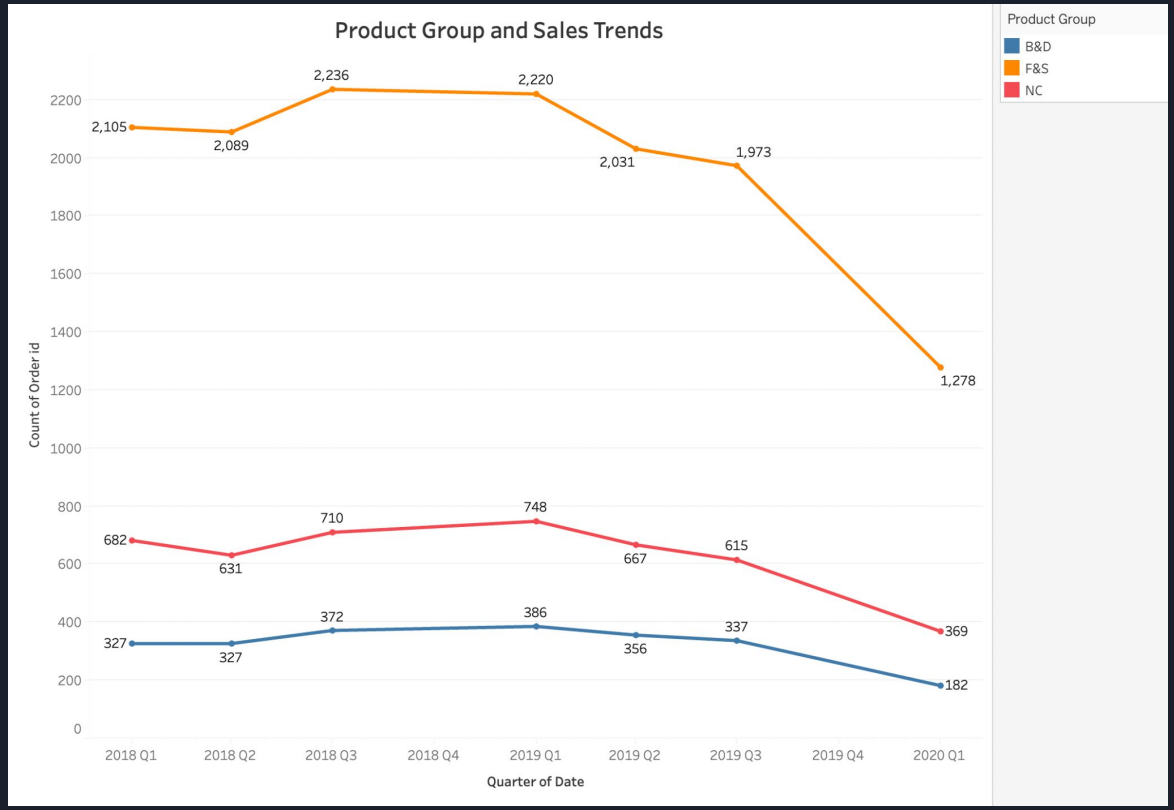


The line chart gives the yearly, quarterly, monthly and weekly trends in the sales

Product and Product Group

The given data set can be categorised into three categories-

B&D	Beverages and Drinks
F&S	Foods and Snacks
NC	Non Consumables



This graph gives the trend of each product group quarter wise.

Inferences:

- The given data has no outliers in them.It is clear from the box plot that there are no outliers.
- The year 2018 had the most sales. The year 2020 might seem like the least sales but only two months of data is available for this year.
- Poultry is the most purchased product followed by ice cream, cereals, etc.,
- Hand soap is the least purchased product followed by sandwich loaves, flour, etc.,
- In 2018, the top 5 products purchased were cereals, poultry, ice cream, cheeses and soda.
- In 2019, the top 5 products purchased were poultry, soda, cheeses ice cream and cereals.
- In 2020, the top 5 products purchased were poultry, ice cream, soda, cheeses and cereals.
- The less frequently sold products can be clubbed with the most frequently sold products and a discount can be offered to increase the sales of such products.
- The Q3 of 2018 and Q1 of 2019 had the most sales.
- The product group Foods and Snacks (F&S) are the most frequently purchased group over the three years.

Market Basket Analysis (Association Rules)

Which tool used?

- Python and Tableau was used for data summary and EDA. I have used KNIME for the Market Basket Analysis (Association Rules)

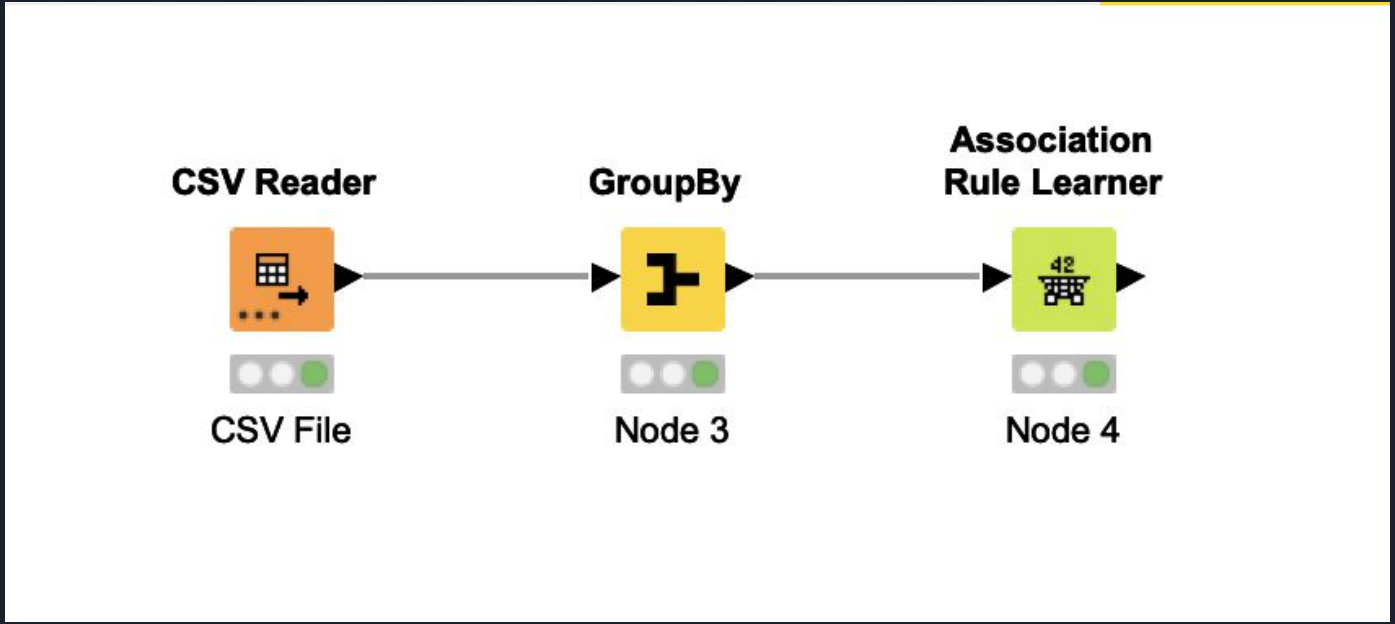
What all parameters used and assumptions made?

- The products are categorised into three categories - Beverages and Drinks, Foods and Snacks and Non Consumables

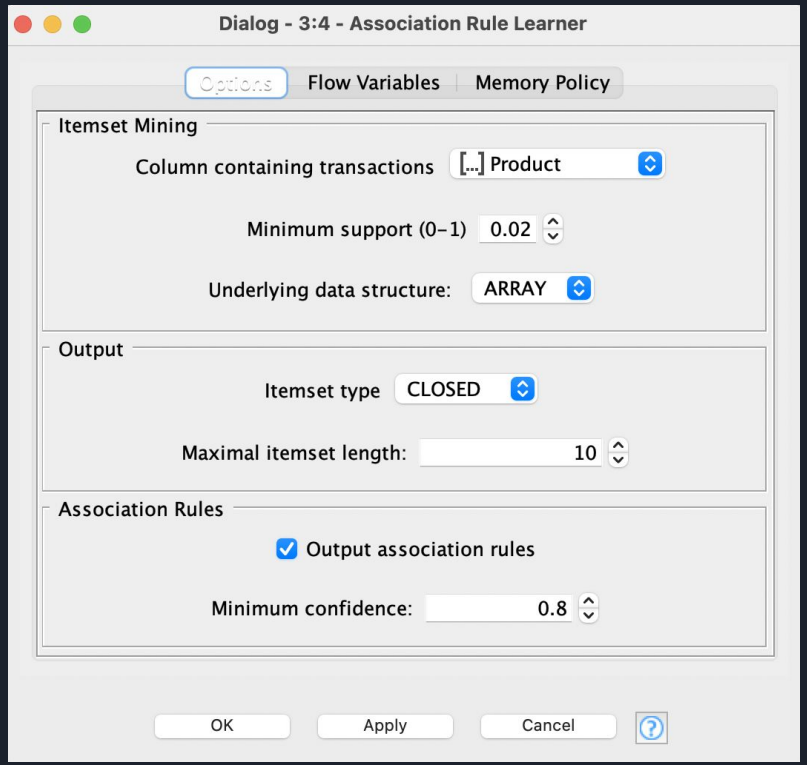
The Association Rules-

- Confidence : measures the percentage of items that an item Y is purchased, given that item X was also purchased
- Support : percentage of orders that contain a set
- Lift : determines the relationship between two items X and Y – >1 shows positive relation and <1 shows negative relation

KNIME workflow



Market Basket Analysis - Association Rule Learner



Sample Data after the Market Basket Analysis (Association Rules)

Row ID	Support	Confidence	Lift	Consequent	implies	Items
rule0	0.02	0.852	2.267	mixes	<---	[yogurt,dishwashing liquid/detergent,all- purpose,...]
rule1	0.02	0.821	2.136	yogurt	<---	[cheeses,all- purpose,tortillas,...]
rule2	0.02	0.821	2.191	beef	<---	[shampoo,fruits,lunch meat,...]
rule3	0.02	0.852	2.18	soda	<---	[bagels,pasta,individual meals,...]
rule4	0.02	0.821	2.161	milk	<---	[eggs,poultry,beef,...]
rule5	0.02	0.821	2.079	lunch meat	<---	[paper towels,milk,individual meals,...]
rule6	0.02	0.885	2.219	ice cream	<---	[paper towels,eggs,dinner rolls,...]
rule7	0.02	0.852	2.349	paper towels	<---	[eggs,dinner rolls,ice cream,...]
rule8	0.02	0.821	2.265	paper towels	<---	[eggs,dinner rolls,poultry,...]
rule9	0.021	0.828	2.152	yogurt	<---	[dishwashing liquid/detergent,eggs,juice,...]
rule10	0.021	0.828	2.208	beef	<---	[poultry,fruits,hand soap,...]
rule11	0.021	0.828	2.218	spaghetti sauce	<---	[waffles,laundry detergent,mixes,...]
rule12	0.021	0.828	2.147	bagels	<---	[sandwich loaves,fruits,toilet paper,...]
rule13	0.021	0.828	1.964	poultry	<---	[eggs,tortillas,coffee/tea,...]
rule14	0.022	0.833	2.133	soda	<---	[ice cream,waffles,milk,...]
rule15	0.022	0.806	1.914	poultry	<---	[butter,cheeses,sandwich loaves,...]
rule16	0.022	0.833	2.138	eggs	<---	[paper towels,dishwashing liquid/detergent,dinner rolls,...]
rule17	0.022	0.833	2.244	pasta	<---	[paper towels,dishwashing liquid/detergent,eggs,...]
rule18	0.023	0.812	2.142	coffee/tea	<---	[yogurt,ice cream,tortillas,...]
rule19	0.023	0.812	2.08	cheeses	<---	[butter,spaghetti sauce,ice cream,...]
rule20	0.023	0.839	1.99	poultry	<---	[dinner rolls,spaghetti sauce,hand soap,...]
rule21	0.023	0.812	2.188	ketchup	<---	[toilet paper,mixes,coffee/tea,...]
rule22	0.023	0.812	1.928	poultry	<---	[spaghetti sauce,laundry detergent,mixes,...]
rule23	0.023	0.839	2.258	ketchup	<---	[tortillas,coffee/tea,juice,...]
rule24	0.024	0.818	2.157	soap	<---	[spaghetti sauce,all- purpose,sandwich bags,...]
rule25	0.024	0.818	2.076	waffles	<---	[paper towels,laundry detergent,soda,...]
rule26	0.025	0.8	2.109	soap	<---	[all- purpose,flour,soda,...]
rule27	0.025	0.8	2.048	soda	<---	[all- purpose,waffles,laundry detergent,...]
rule28	0.025	0.8	1.898	poultry	<---	[dinner rolls,spaghetti sauce,sandwich loaves,...]
rule29	0.025	0.8	1.898	poultry	<---	[dinner rolls,spaghetti sauce,hand soap,...]
rule30	0.025	0.8	2.062	dishwashing li...	<---	[paper towels,spaghetti sauce,milk,...]
rule31	0.025	0.806	2.021	ice cream	<---	[paper towels,yogurt,pasta,...]
rule32	0.025	0.806	1.912	poultry	<---	[dinner rolls,spaghetti sauce,ice cream,...]
rule33	0.025	0.829	2.13	dinner rolls	<---	[spaghetti sauce,poultry,waffles,...]
rule34	0.025	0.829	1.966	poultry	<---	[shampoo,hand soap,juice,...]
rule35	0.026	0.857	2.194	cheeses	<---	[paper towels,cereals,sandwich bags,...]
rule36	0.028	0.842	1.998	poultry	<---	[dinner rolls,spaghetti sauce,beef,...]
rule37	0.028	0.821	1.947	poultry	<---	[dinner rolls,spaghetti sauce,sandwich loaves,...]
rule38	0.029	0.846	2.008	poultry	<---	[dinner rolls,spaghetti sauce,hand soap,...]

Recommendations:

- Milk seems to be everyone's favourite and it is coupled with eggs so that offers can be made by bundling milk and eggs. It can be promoted as a breakfast combo.
- Loyalty points can be provided to make customers return more often.
- Poultry appears to be the most sold product, so it is best to pair it up with some complimenting item that may not be sold as often in order to increase the sales.
- Poultry can also be suggested with other Food and Snacks category like dinner rolls and spaghetti sauce.
- Soda can also be offered in a combo.
- Combination of high selling items with low selling items could help the sale of both products.

Links:https://public.tableau.com/app/profile/akshaya.nallathambi/viz/MRAproject2_16429231298310/ProductGroupandSalesTrends

End of Presentation