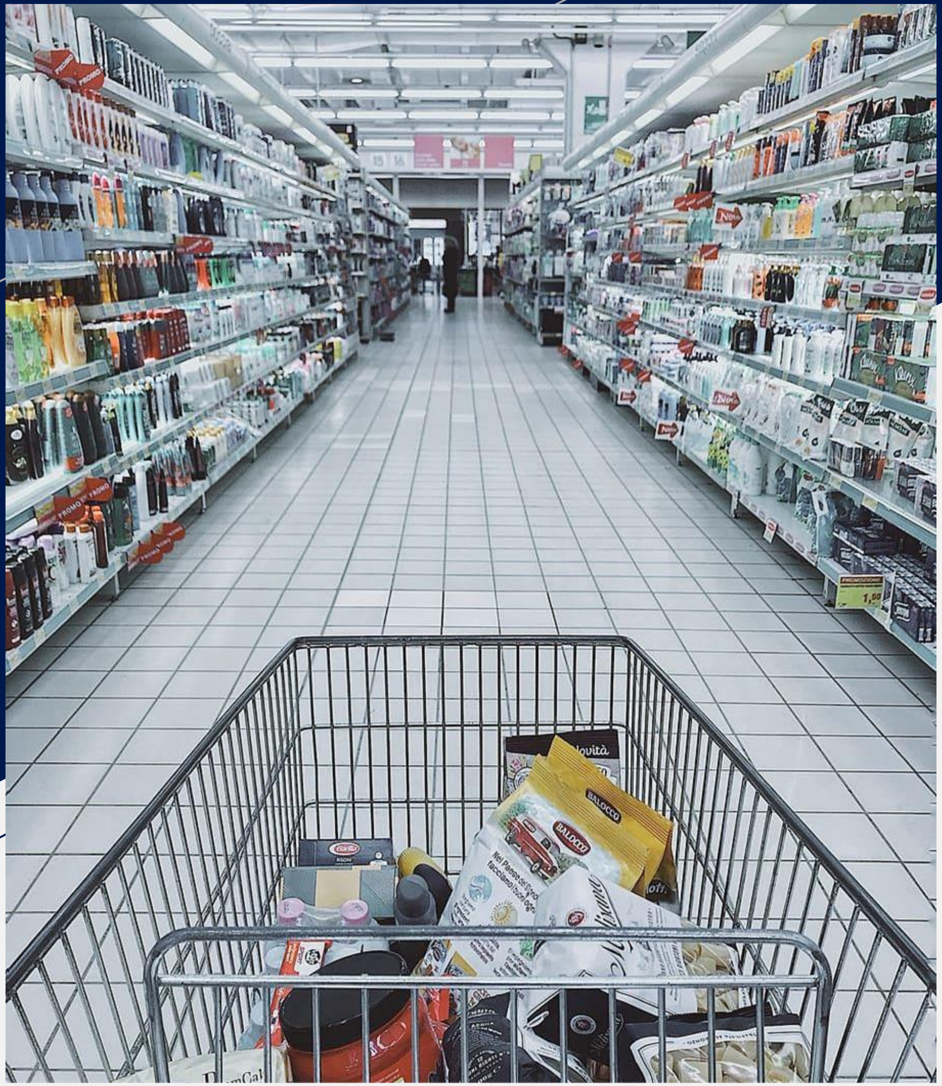
MRA



MARKETING & RETAIL ANALYSIS

MILESTONE 2

HARSH PANDYA

PGP DSBA FEB_A 2021

MRA

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.

MRA

Contents

- Problem Statement
- Data Summary
- EDA
- Trends
- Market Basket Analysis
- MRA in Python
- Association Identified
- Recommendation
- Offers
- Tableau Link

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?

Grocery Store Data:

	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 Summary

MRA

- 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.

```
Shape of the Data is (20641, 3)
Number of Rows= 20641
Number of Calumns= 3
```

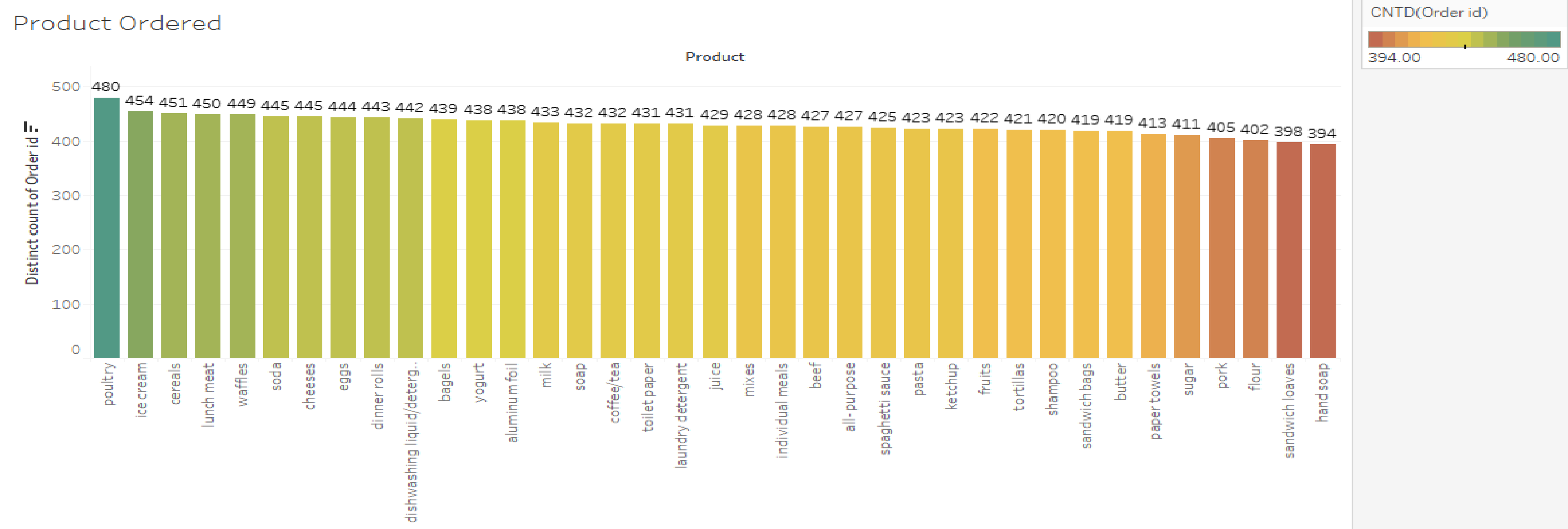
```
Date          603
Order_id     1139
Product        37
dtype: int64
```

```
Total Number of Duplicates are 4730
```

```
<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
```

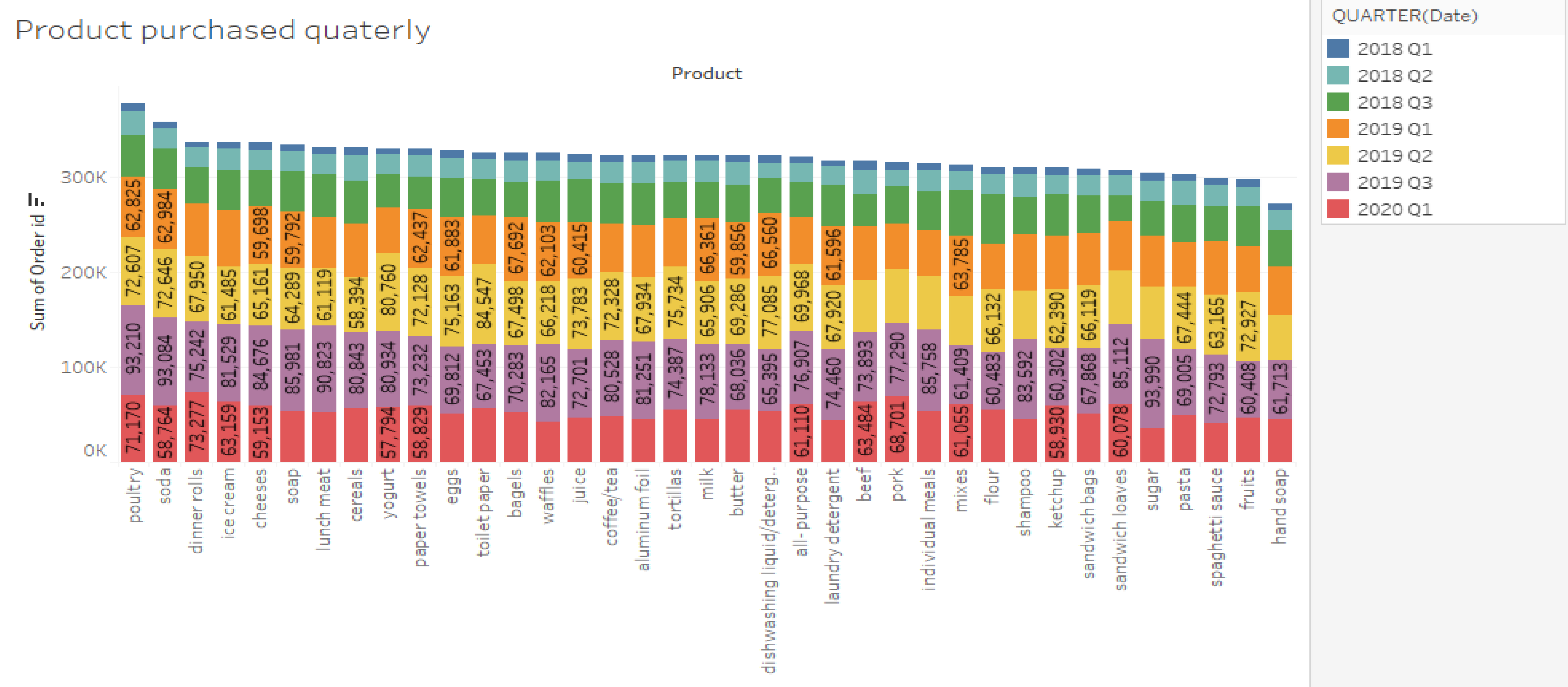
EXPLORATORY DATA ANALYSIS

PRODUCT ORDERED



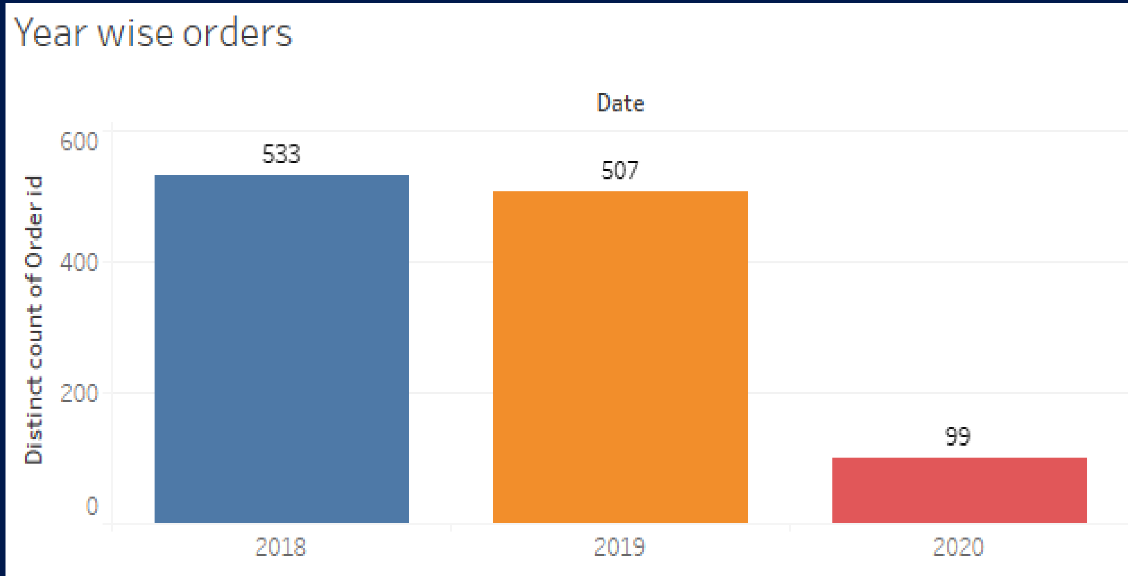
- It is evident that Poultry has been ordered the highest with 480 followed by ice cream with 454 orders.
- The lowest is hand soap 394 orders and sandwich loaves with 398 orders.
- The milk, soap, coffee/tea, soda, cheese are more or less holds the same amount of orders.

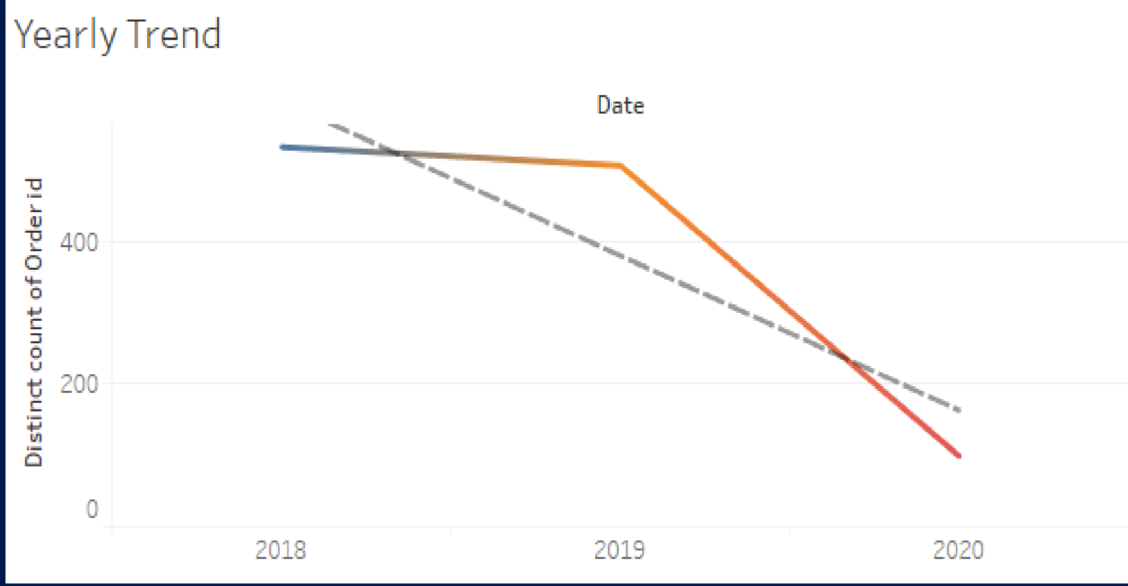
Product Purchased Quarterly for 2018,2019 and 2020



- 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



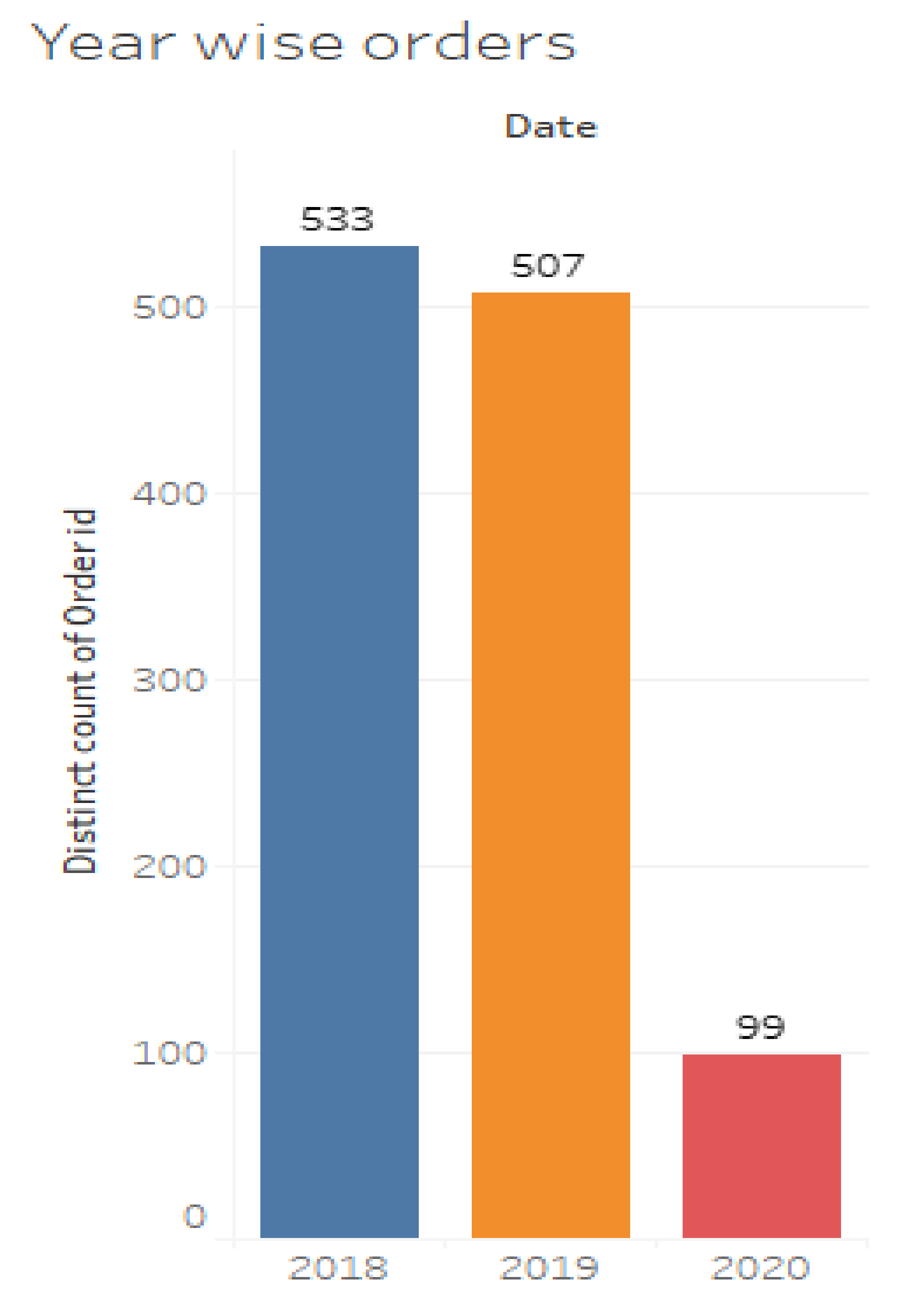


- The total number of orders are highest in the year 2018.
- Followed by 507 orders in 2019.
- 2020 has registered only 99 orders. As only two months data is given in the data set , this might be a reason for low order count.
- Trends shows that the number of orders placed has been decreasing yearly.
- Also it shows a downward trend and no forecast has been generated with this data.
- The

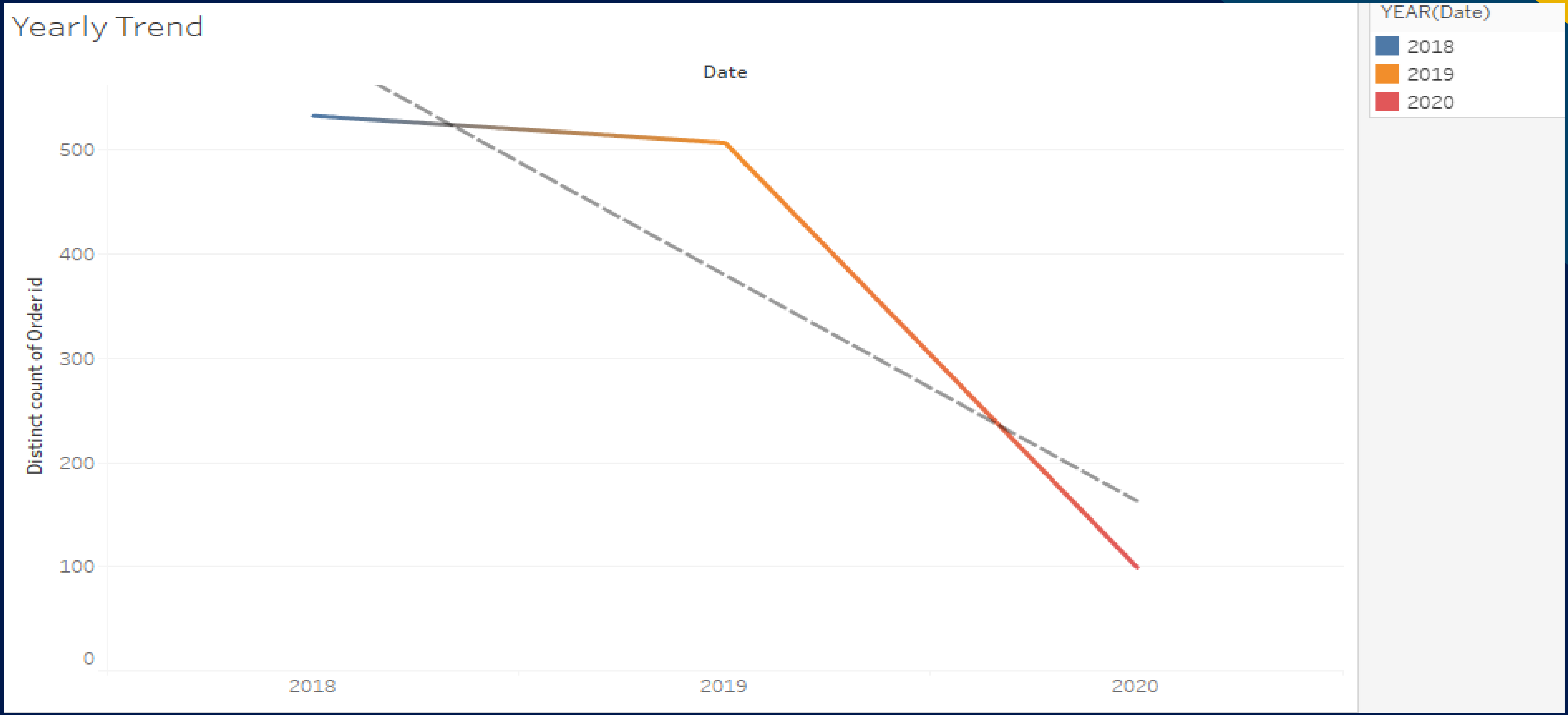
  R-Squared value is- 0.79476

  P-Value – 0.299317

Yearly Orders



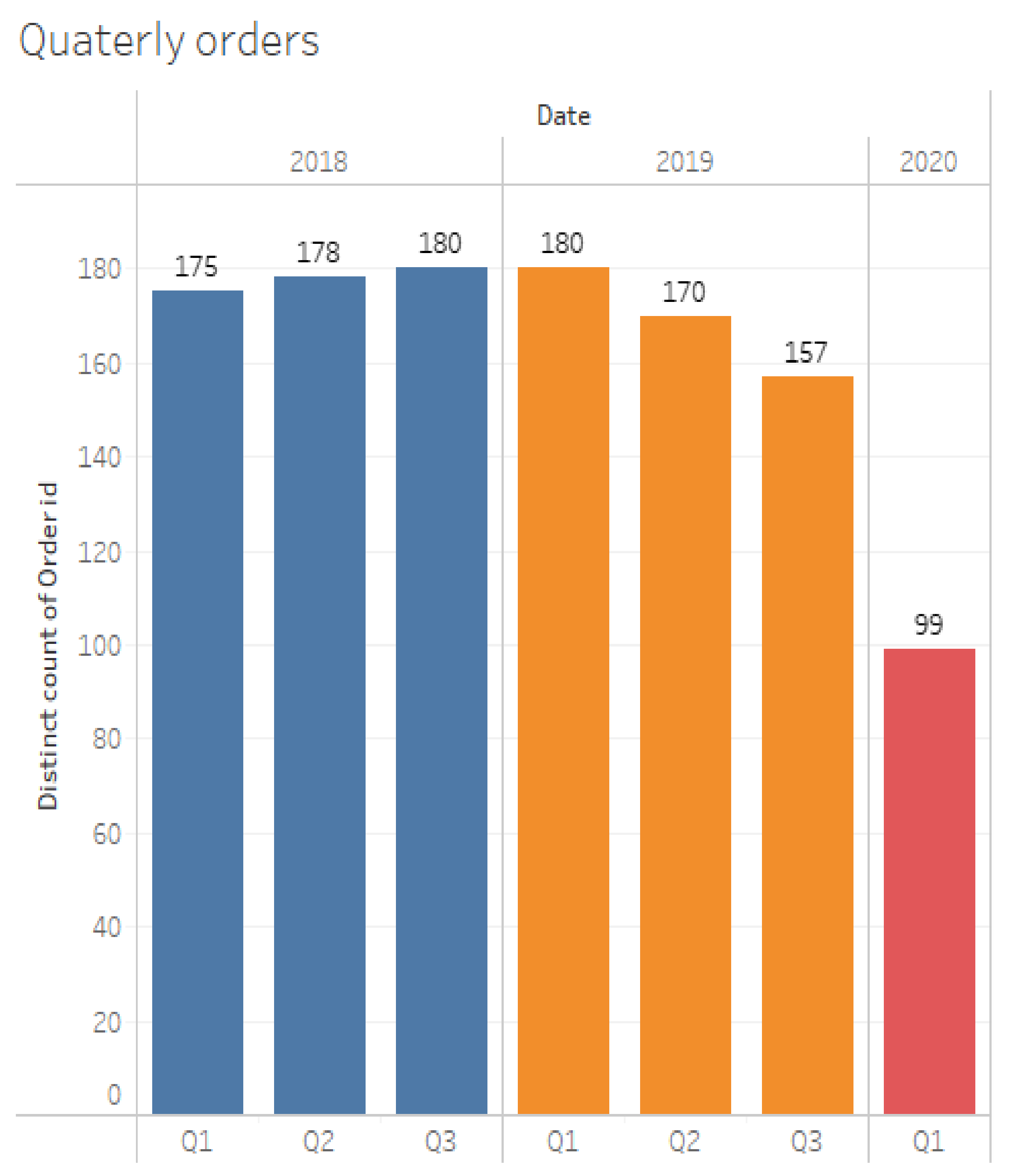
Trends



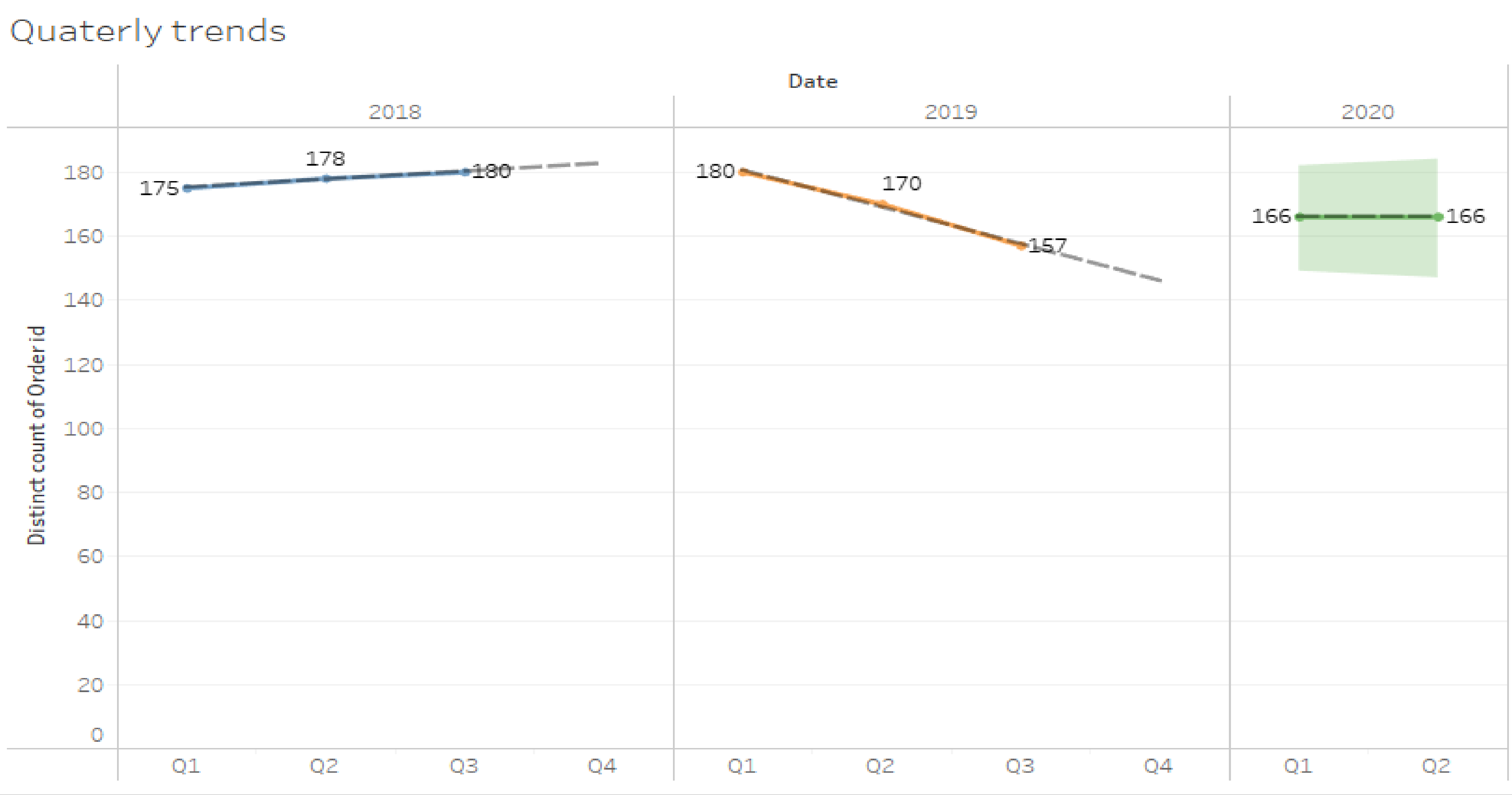
Yearly Trend
Date
YEAR(Date)
2018
2019
2020
Distinct count of Order id
500
400
300
200
100
0
2018
2019
2020

Quarterly orders and Its Trends

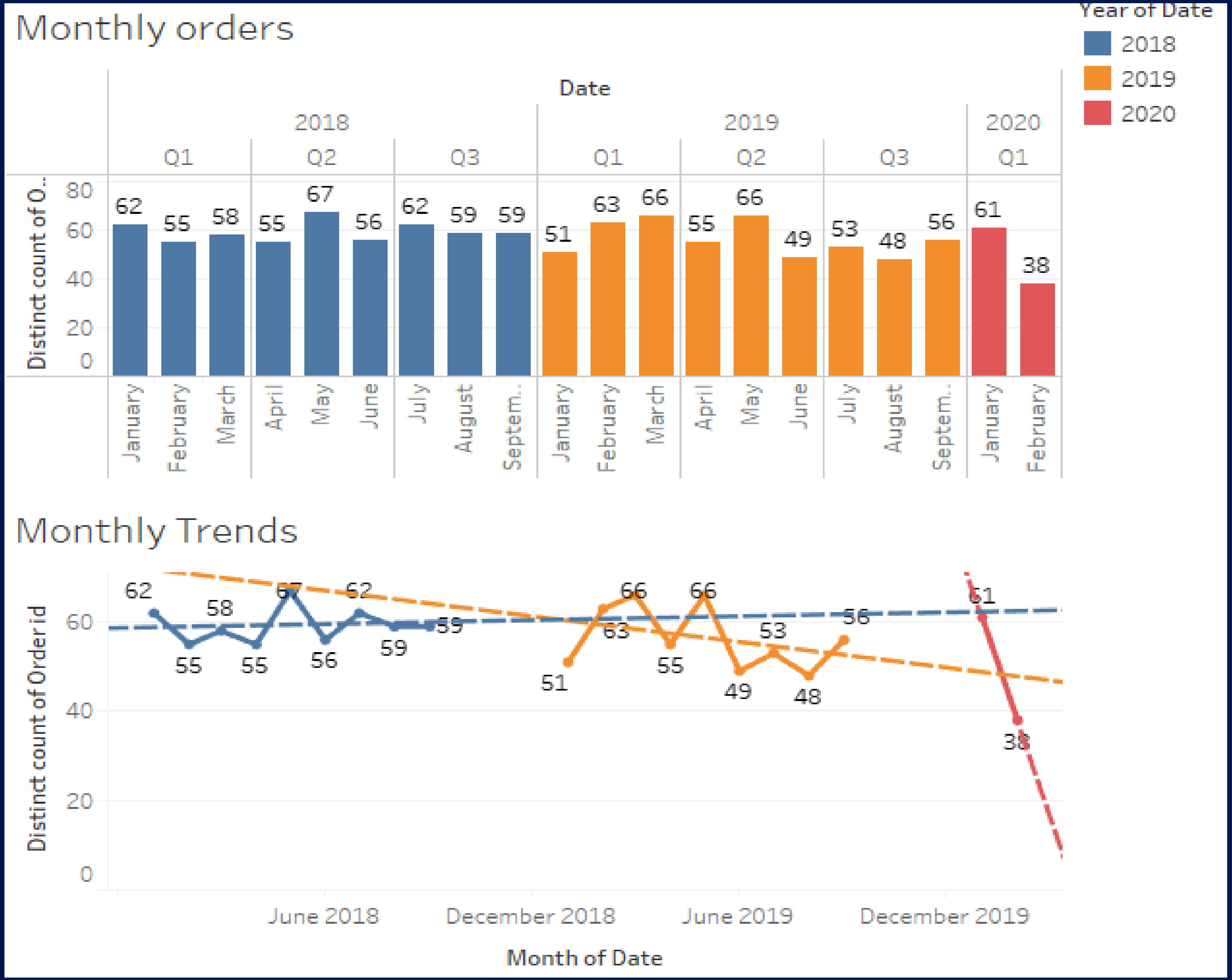
- In 2018, Q3 had the highest number of orders with 180.
- In 2019, Q1 had the highest with 180.
- This started decreasing as quarters passed by. The lowest was recorded in Q1 of 2020.
- Also, it is evident that Q4 of every year doesn't have any sales.
- No proper trend can be analyzed in terms of Quarter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



Quarterly Trends

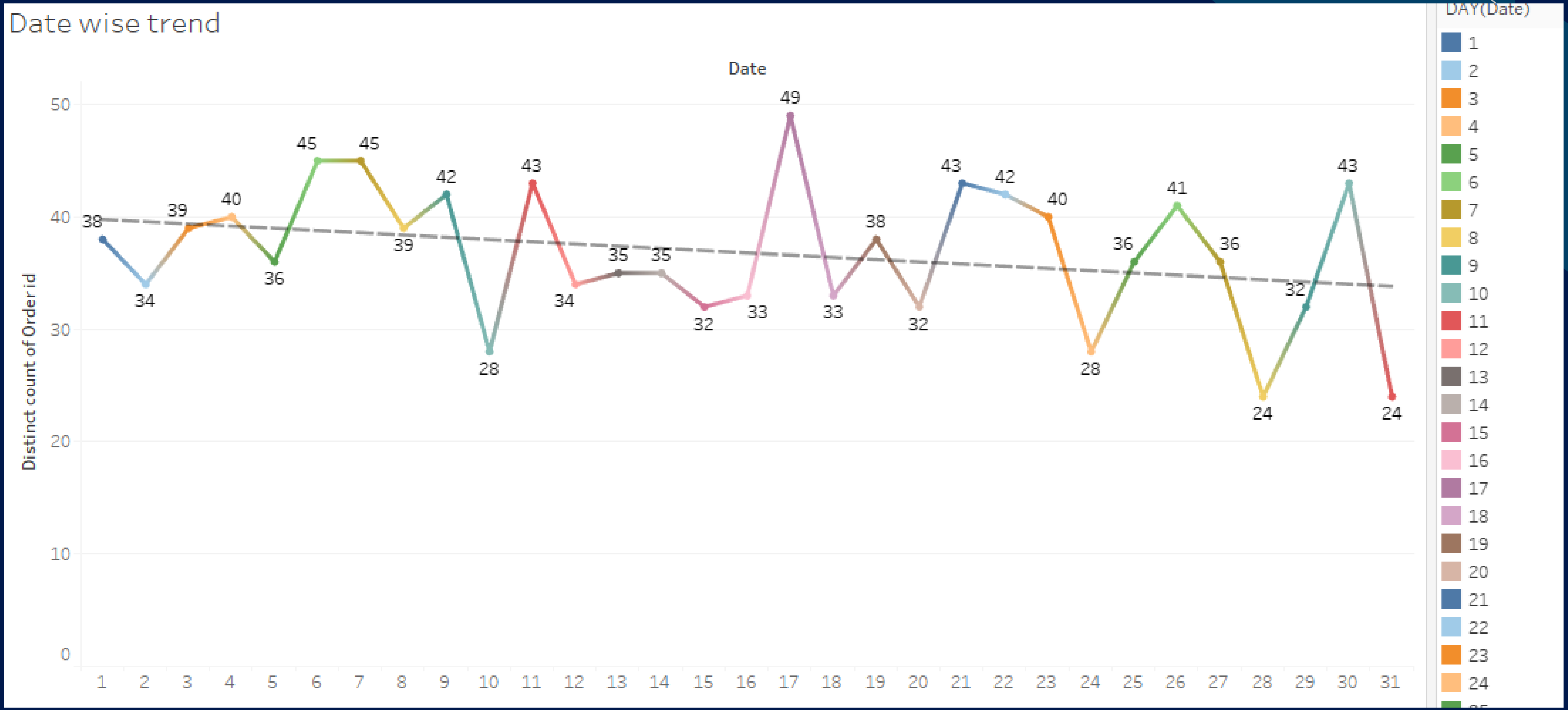


Monthly orders and Its Trends



- There is no trend and seasonality available in the data provided.
- As per the forecast the sales would tend to increase in the mid quarters of the years.
- 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.

Daily Wise Quantity ordered



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

MRA

Market Basket Analysis

- Using A priori 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 modeling 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.

MRA

Association Rule Mining & 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 summarized 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:

1. Support greater than or equal to min_support
2. Confidence greater than or equal to min_confidence

- There are 20,642 records in the grocery dataset, totaling 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 giftpacks.

Python Output table Image for

Product	all-purpose	aluminum foil	bagels	beef	butter	cereals	cheeses	coffee/tea	dinner rolls	dishwashing liquid/detergent	eggs	flour	fruits	hand soap	ice cream	individual meals	juice	ket
Order_id																		
1	3.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	2.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	
2	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	2.0	0.0	2.0	0.0	
3	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	3.0	0.0	0.0	
4	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	
5	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	

- Python 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.

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

	support	itemsets
0	0.374890	(all- purpose)
1	0.384548	(aluminum foil)
2	0.385426	(bagels)
3	0.374890	(beef)
4	0.367867	(butter)
...	...	...
610567	0.010536	(soap, sandwich bags, mixes, soda, ketchup, po...
610568	0.011414	(soap, sandwich bags, mixes, soda, waffles, ke...
610569	0.010536	(soap, shampoo, lunch meat, soda, laundry dete...
610570	0.011414	(mixes, lunch meat, sandwich bags, shampoo, yo...
610571	0.010536	(mixes, lunch meat, sandwich bags, shampoo, to...

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

Association Rule

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(beef, all- purpose, flour, butter)	(aluminum foil)	0.021949	0.384548	0.015803	0.720000	1.872329	0.007363	2.198043
1	(cereals, all- purpose, fruits, aluminum foil)	(beef)	0.026339	0.374890	0.018437	0.700000	1.867213	0.008563	2.083699
2	(beef, all- purpose, waffles, milk)	(aluminum foil)	0.023705	0.384548	0.017559	0.740741	1.926264	0.008444	2.373887
3	(aluminum foil, cereals, all- purpose, butter)	(laundry detergent)	0.026339	0.378402	0.018437	0.700000	1.849884	0.008471	2.071993
4	(cheeses, aluminum foil, all- purpose, butter)	(laundry detergent)	0.024583	0.378402	0.017559	0.714286	1.887637	0.008257	2.175593
...	...	...	...	...	...	...	...	...	...
176192	(tortillas, lunch meat, milk, yogurt, mixes, s...	(shampoo)	0.014047	0.368745	0.010536	0.750000	2.033929	0.005356	2.525022
176193	(tortillas, lunch meat, milk, shampoo, mixes, ...	(yogurt)	0.010536	0.384548	0.010536	1.000000	2.600457	0.006484	inf
176194	(tortillas, lunch meat, yogurt, shampoo, mixes...	(milk)	0.013169	0.380158	0.010536	0.800000	2.104388	0.005529	3.099210
176195	(tortillas, milk, yogurt, shampoo, mixes, sand...	(lunch meat)	0.014047	0.395083	0.010536	0.750000	1.898333	0.004986	2.419666
176196	(lunch meat, milk, yogurt, shampoo, mixes, san...	(tortillas)	0.014047	0.369622	0.010536	0.750000	2.029097	0.005343	2.521510

176197 rows × 9 columns

After export new excel file

MRA

A	B	C	D	E	F	G	H	I	J
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	frozenset({'butter', 'flour', 'all- purpose', 'beef'})	frozenset({'aluminum foil'})	0.021949078	0.384547849	0.015803336	0.72	1.872328767	0.007362865	2.198043396
1	frozenset({'aluminum foil', 'all- purpose', 'cereals', '	frozenset({'beef'})	0.026338894	0.374890255	0.018437226	0.7	1.867213115	0.008563031	2.083699151
2	frozenset({'milk', 'all- purpose', 'beef', 'waffles'})	frozenset({'aluminum foil'})	0.023705004	0.384547849	0.017559263	0.740740741	1.926264164	0.008443554	2.373886868
3	frozenset({'butter', 'aluminum foil', 'all- purpose', 'c	frozenset({'laundry detergent'})	0.026338894	0.378402107	0.018437226	0.7	1.849883991	0.008470533	2.071992976
4	frozenset({'butter', 'all- purpose', 'aluminum foil', 'c	frozenset({'laundry detergent'})	0.024582968	0.378402107	0.017559263	0.714285714	1.887636725	0.008257016	2.175592625
5	frozenset({'butter', 'aluminum foil', 'all- purpose', 'c	frozenset({'mixes'})	0.021949078	0.375768218	0.015803336	0.72	1.916074766	0.00755557	2.229399222
6	frozenset({'butter', 'toilet paper', 'all- purpose', 'ha	frozenset({'aluminum foil'})	0.019315189	0.384547849	0.01404741	0.727272727	1.891241179	0.006619796	2.256657887
7	frozenset({'butter', 'aluminum foil', 'all- purpose', 'j	frozenset({'poultry'})	0.02809482	0.4214223	0.021071115	0.75	1.7796875	0.009231331	2.314310799
8	frozenset({'butter', 'sandwich loaves', 'aluminum fo	frozenset({'ketchup'})	0.022827041	0.371378402	0.016681299	0.730769231	1.967721404	0.008203829	2.334880221
9	frozenset({'butter', 'soap', 'aluminum foil', 'all- pur	frozenset({'ketchup'})	0.027216857	0.371378402	0.020193152	0.741935484	1.997788454	0.010085399	2.435908692
10	frozenset({'butter', 'aluminum foil', 'all- purpose', 'r	frozenset({'poultry'})	0.026338894	0.4214223	0.018437226	0.7	1.661041667	0.007337428	1.928592332
11	frozenset({'butter', 'soap', 'aluminum foil', 'all- pur	frozenset({'paper towels'})	0.027216857	0.362598771	0.020193152	0.741935484	2.046161056	0.010324353	2.469929763
12	frozenset({'butter', 'aluminum foil', 'all- purpose', 's	frozenset({'toilet paper'})	0.027216857	0.378402107	0.020193152	0.741935484	1.960706534	0.009894236	2.408691835
13	frozenset({'all- purpose', 'aluminum foil', 'eggs', 'ce	frozenset({'paper towels'})	0.023705004	0.362598771	0.017559263	0.740740741	2.042866111	0.008963857	2.458547598
14	frozenset({'flour', 'aluminum foil', 'all- purpose', 'ce	frozenset({'yogurt'})	0.025460931	0.384547849	0.018437226	0.724137931	1.883089277	0.00864628	2.231014047
15	frozenset({'hand soap', 'aluminum foil', 'all- purpos	frozenset({'milk'})	0.017559263	0.380158033	0.013169447	0.75	1.972863741	0.006494152	2.479367867
16	frozenset({'hand soap', 'aluminum foil', 'all- purpos	frozenset({'paper towels'})	0.017559263	0.362598771	0.012291484	0.7	1.930508475	0.005924517	2.124670764
17	frozenset({'hand soap', 'aluminum foil', 'all- purpos	frozenset({'poultry'})	0.017559263	0.4214223	0.012291484	0.7	1.661041667	0.004891619	1.928592332
18	frozenset({'aluminum foil', 'all- purpose', 'ketchup',	frozenset({'paper towels'})	0.023705004	0.362598771	0.016681299	0.703703704	1.940722805	0.008085894	2.151229148
19	frozenset({'sandwich bags', 'aluminum foil', 'all- pu	frozenset({'laundry detergent'})	0.021949078	0.378402107	0.015803336	0.72	1.902737819	0.007497759	2.219992475
20	frozenset({'cereals', 'aluminum foil', 'all- purpose', '	frozenset({'paper towels'})	0.024582968	0.362598771	0.017559263	0.714285714	1.969906607	0.008645509	2.230904302
21	frozenset({'all- purpose', 'aluminum foil', 'cheeses',	frozenset({'ice cream'})	0.030728709	0.398595259	0.021949078	0.714285714	1.792007552	0.00970076	2.104916594
22	frozenset({'hand soap', 'all- purpose', 'cheeses', 'mi	frozenset({'aluminum foil'})	0.027216857	0.384547849	0.019315189	0.709677419	1.845485344	0.008849005	2.119890742
23	frozenset({'hand soap', 'all- purpose', 'aluminum fo	frozenset({'milk'})	0.025460931	0.380158033	0.019315189	0.75862069	1.995540336	0.009636011	2.567916719

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:

1. Mixed & Dishwashing liquid
2. Eggs & Soda
3. Juice & Shampoo
4. Juice and Spaghetti sauce
5. Pasta & Paper towels

Tableau Link

Link:- Milestone 2 Harsh Pandya PGP DSBA FEB_A 2021

MRA

THANK YOU