MARKETING AND RETAIL ANALYTICS PROJECT

- MILESTONE 2
- PRAKASH V MAHADOLE

Content
Agenda & Executive Summary of the data
Exploratory Analysis and Inferences
Yearly, Monthly, Quarterly and weekly trends
Market Basket Analysis and Knime workflow
MBA Association Rule
Result table head and inference
Insights and 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?

➢ TOOLS USED -

- Python used for EDA
- Tableau used for visualization :
  - Tableau public link

    https://public.tableau.com/views/MRAmilestone2_16394787556060/Sheet4?:language=en-US&:display_count=n&:origin=viz_share_link

- KNIME Workflow used for MBA Analysis

➢ FACT CHECK ABOUT GIVEN 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
```

```
Out[6]: Date        0
        Order_id    0
        Product     0
        dtype: int64
```

➢ FACT CHECK ABOUT GIVEN DATA -

- Data contains 20641 rows and 3 columns.
- Data don't have any null values.
- It contains datatype :

Object (2) and

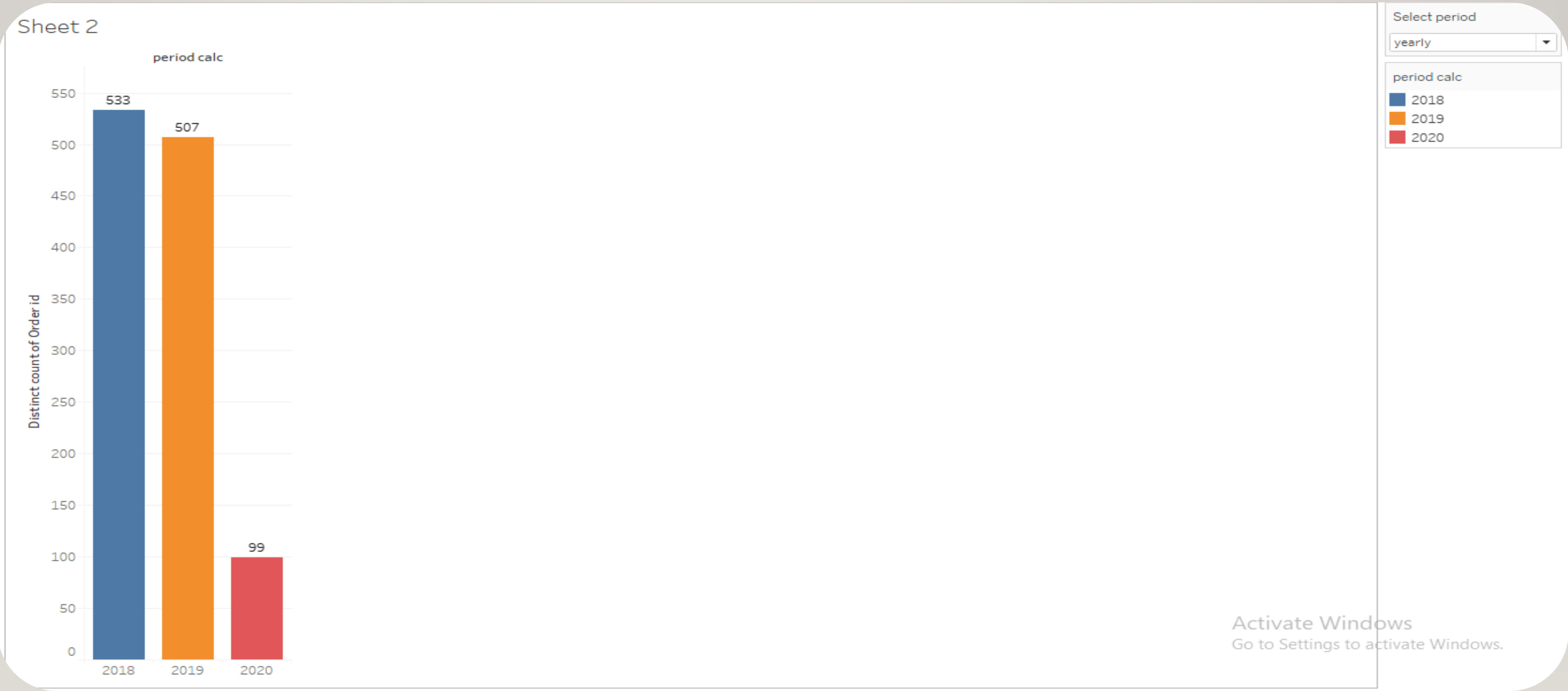
int64 (1)

- No duplicate data found
- Data description is given and data looks good .
- Data provided from Jan to Sep for 2 years (2018, 2019) and 2020 with 2 months(Jan and Feb)

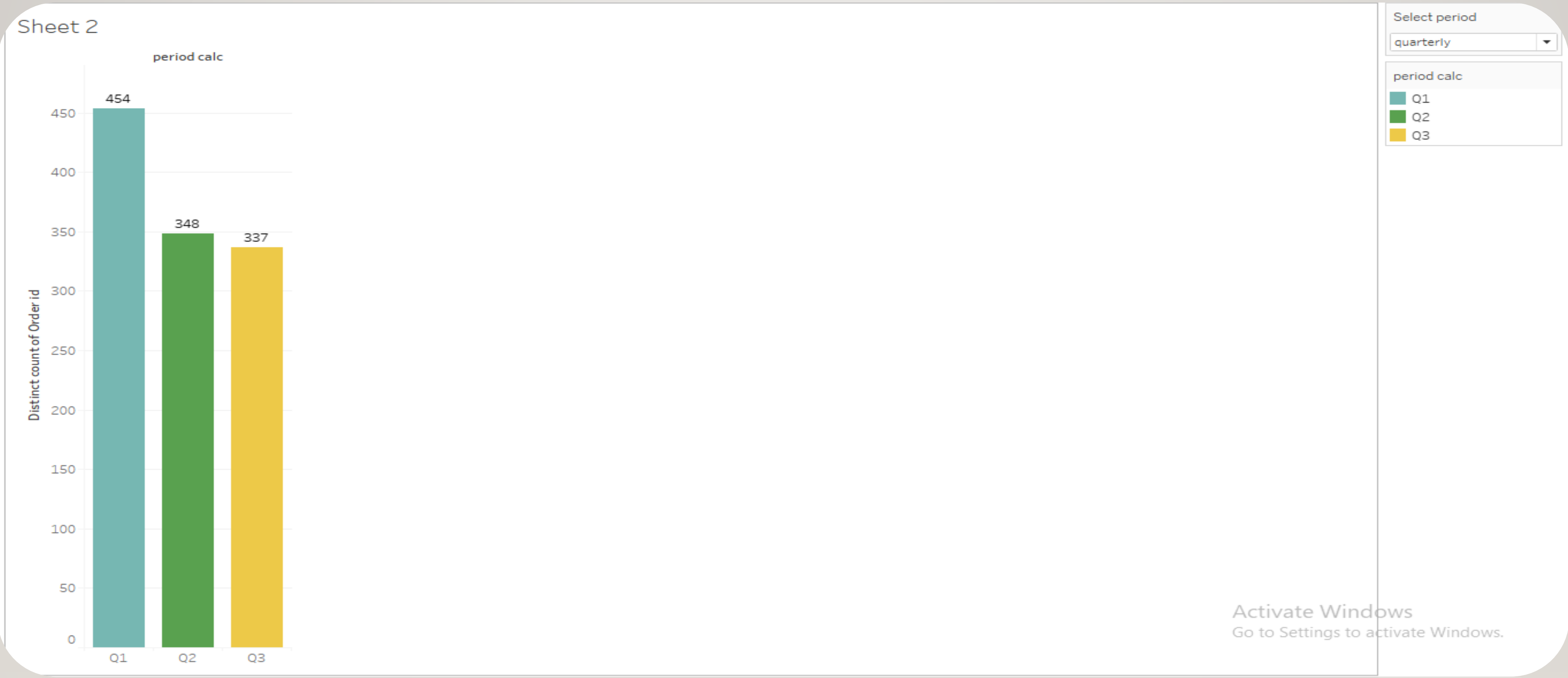
Out[7]:

	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

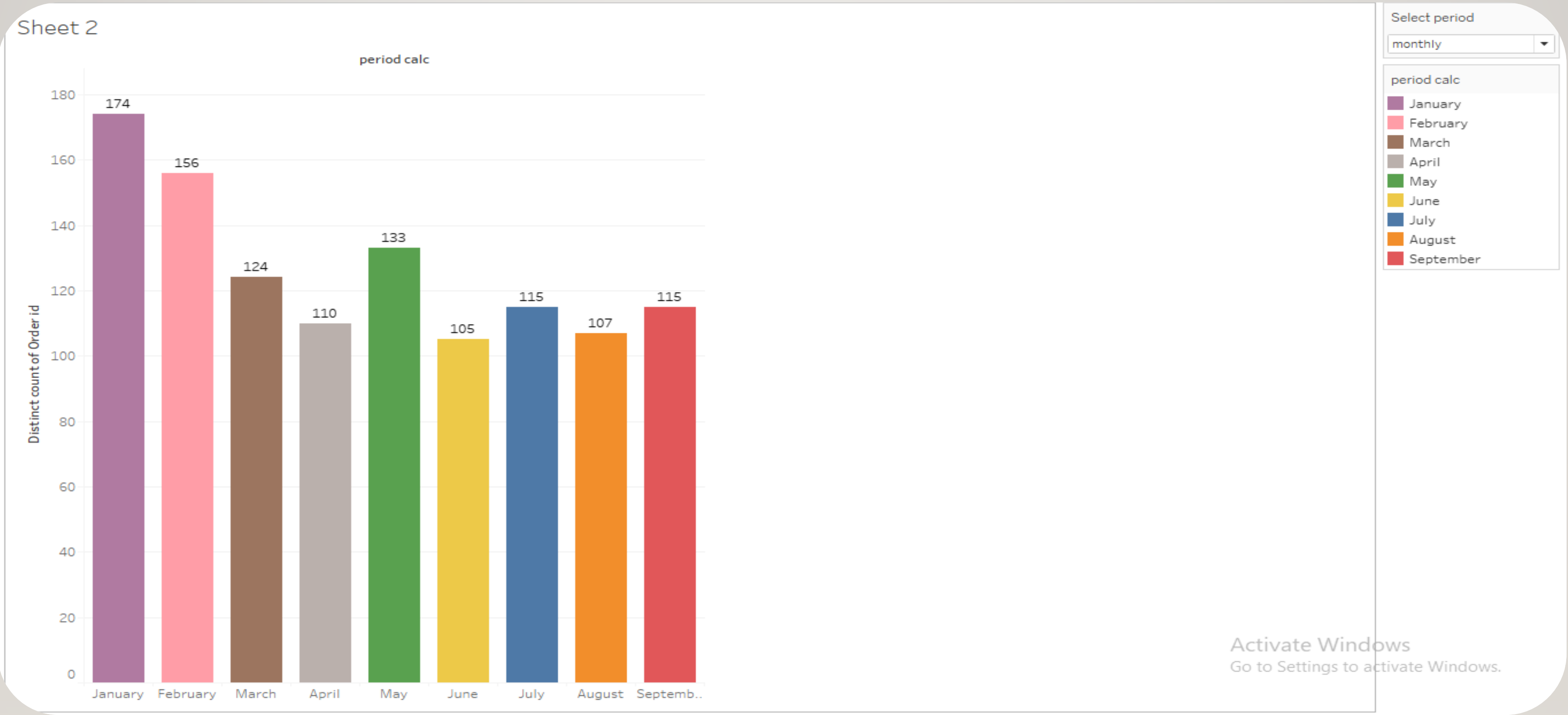
➢ YEARLY TREND :



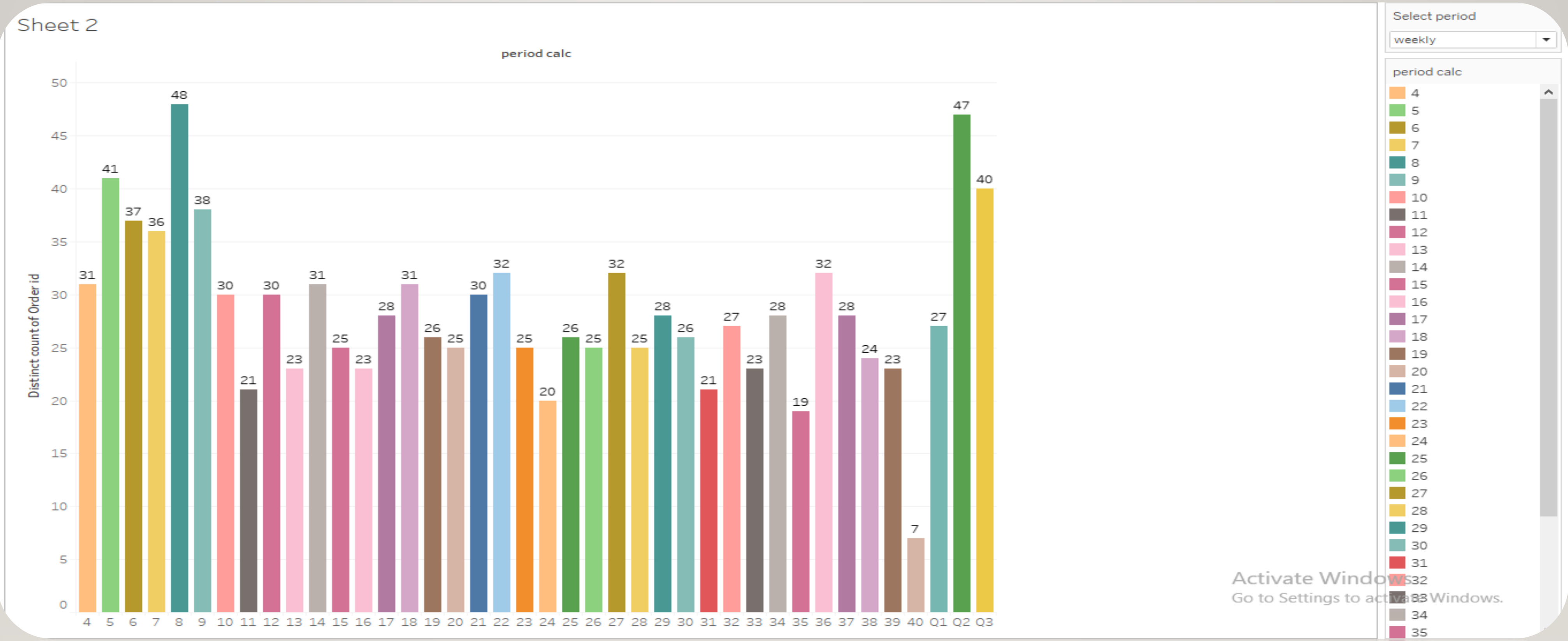
➢ QUARTERLY TREND :



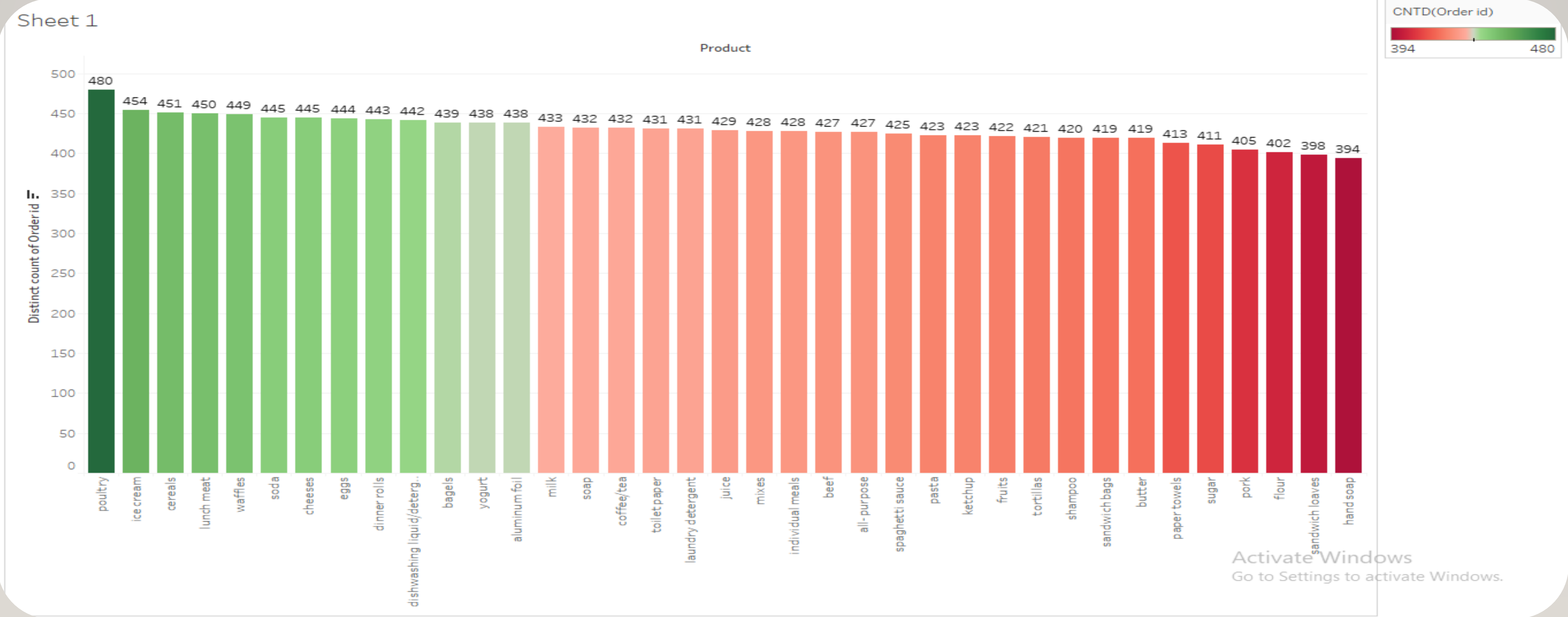
➢ MONTHLY TREND :



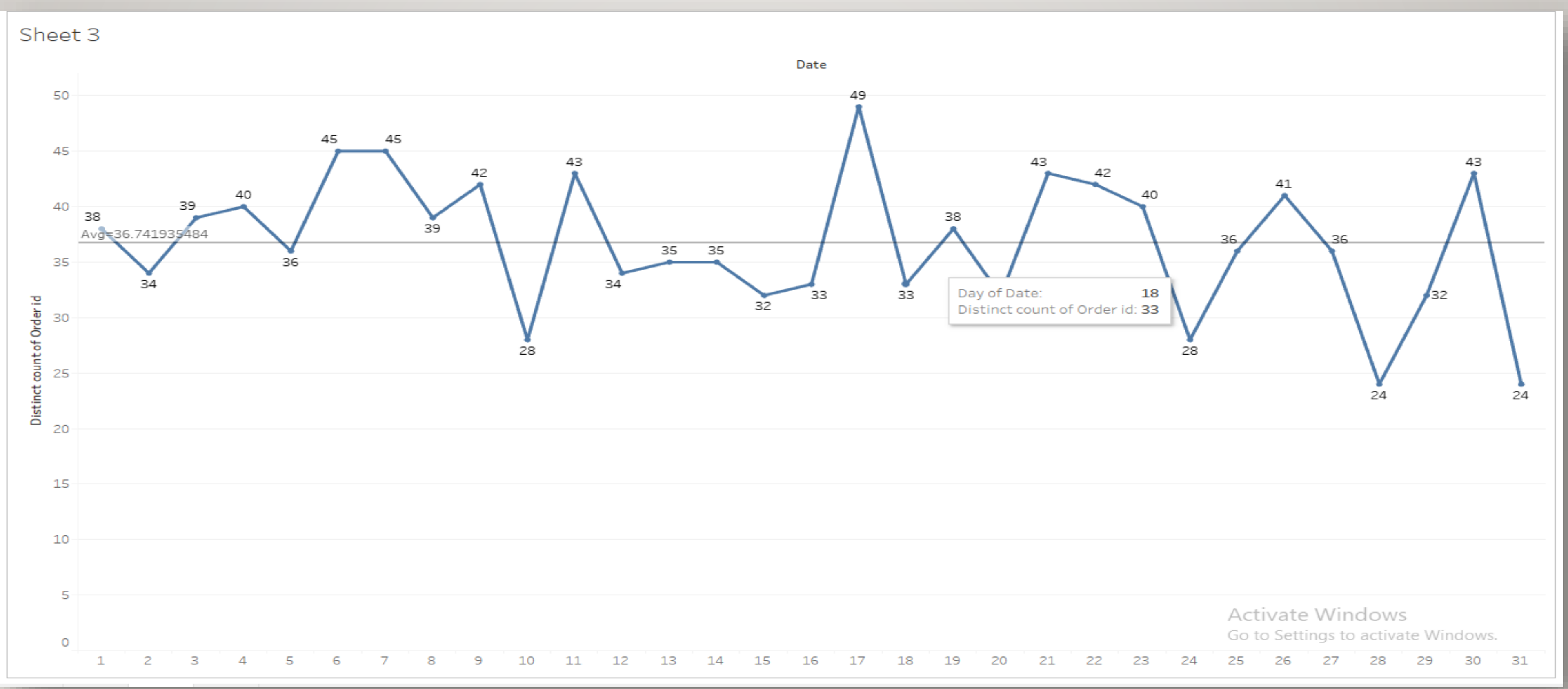
➢ WEEKLY TREND :



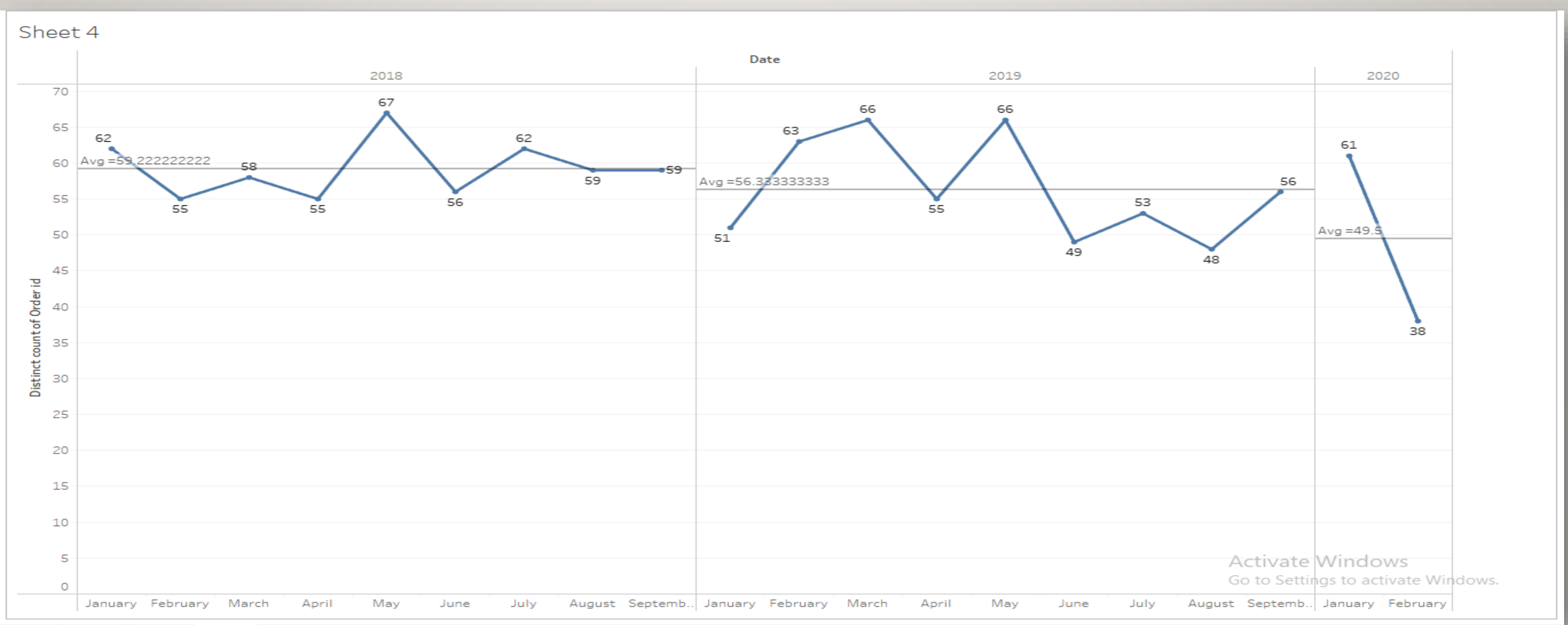
➢ PRODUCTS COUNT :



➢ DAY WISE TREND :



➢ MONTHLY TREND :



➢ INFERENCE FROM THE ABOVE ANALYSIS :

- The year 2018 has the highest no of orders followed by 2019, Since the data in the year 2020 has only 2 months so very low count in orders.
- There is highest no of unique orders in Jan(174) and low number of orders made in June(105).
- There is no trend and seasonality available in the data provided.
- High number of orders made on mid of the month and start of month is low and it reduced at end of month.
- The product poultry is the order highest no of orders and hand soap is the lowest no of orders

➢ MARKET BASKET ANALYSIS

- Market Basket Analysis is a technique which identifies the strength of association between pairs of products purchased together and identify patterns of co-occurrence. A co- occurrence is when two or more things take place together.
- Market Basket Analysis creates If-Then scenario rules, for example, if item A is purchased then item B is likely to be purchased. The rules are probabilistic in nature or, in other words, they are derived from the frequencies of co-occurrence in the observations. Frequency is the proportion of baskets that contain the items of interest. The rules can be used in pricing strategies, product placement, and various types of cross-selling strategies. In order to make it easier to understand, think of Market Basket Analysis in terms of shopping at a supermarket. Market Basket Analysis takes data at transaction level, which lists all items bought by a customer in a single purchase. The technique determines relationships of what products were purchased with which other product(s). These relationships are then used to build profiles containing If-Then rules of the items purchased. The rules could be written as If {A} Then {B} .
- The If part of the rule (the {A} above) is known as the antecedent and the THEN part of the rule is known as the consequent (the {B} above).
- The antecedent is the condition and the consequent is the result. The association rule has three measures that express the degree of confidence in the rule, Support, Confidence, and Lift.

➢ THRESHOLD VALUES :

- Support: Its the default popularity of an item. In mathematical terms, the support of item A is nothing but the ratio of transactions involving A to the total number of transactions.
- Confidence: Likelihood that customer who bought both A and B. Its divides the number of transactions involving both A and B by the number of transactions involving B.
- Lift : Increase in the sale of A when you sell B.

➢ KNIME WORKFLOW :

Node Name	Description
CSV Reader	Read the CSV file
GroupBy	GrpoupBy OrderID
Cell Splitter	Convert the product data to Set
Association Rule Learner	MBA to generate the frequent/item list

➢ MBA ASSOCIATION RULE :

- These rules are actionable in that they can be used to target customers for marketing, or for product placing, or more generally to inform decision making. Examples of areas in which association rules have been used include: Supermarket purchases: common combinations of products can be used to inform product placement on supermarket shelves.
- This is the most important node for our Market Basket Analysis. We have here the three metrics that are Support, Confidence and Lift, we added a value to our Support which is between 0-1. We added value of 0.05 that is 5% sell of a product from overall transactions and we also selected the association rule for the minimum confidence as 0.5. So as you can see the values of confidence and support will help us to get threshold

➢ OUTPUT TABLE HEAD AFTER APPLYING ASSOCIATION RULE :

Row ID	D Support	D Confide...	D Lift	S Consequent	S implies	[...] Items
rule59	0.055	0.649	1.791	paper towels	<---	[eggs,ice cream,pasta]
rule58	0.055	0.643	1.731	pasta	<---	[paper towels,eggs,ice cream]
rule21	0.051	0.674	1.726	cheeses	<---	[bagels,cereals,sandwich bags]
rule3	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum foil]
rule18	0.051	0.63	1.678	mixes	<---	[yogurt,poultry,aluminum foil]
rule20	0.051	0.611	1.66	sandwich bags	<---	[cheeses,bagels,cereals]
rule52	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce,poultry,laundry detergent]
rule40	0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce,poultry,ice cream]
rule7	0.05	0.62	1.645	juice	<---	[yogurt,poultry,aluminum foil]
rule43	0.052	0.686	1.628	poultry	<---	[dinner rolls,spaghetti sauce,ice cream]
rule49	0.052	0.634	1.627	eggs	<---	[paper towels,dinner rolls,pasta]
rule50	0.052	0.602	1.621	pasta	<---	[paper towels,eggs,dinner rolls]
rule24	0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce,poultry,cereals]
rule57	0.055	0.63	1.616	eggs	<---	[paper towels,ice cream,pasta]
rule11	0.05	0.613	1.616	coffee/tea	<---	[yogurt,cheeses,cereals]
rule44	0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce,poultry,juice]
rule35	0.052	0.628	1.61	eggs	<---	[dinner rolls,poultry,soda]
rule54	0.054	0.598	1.603	spaghetti sauce	<---	[dinner rolls,poultry,laundry detergent]
rule29	0.051	0.604	1.589	milk	<---	[poultry,laundry detergent,cereals]
rule42	0.052	0.59	1.581	spaghetti sauce	<---	[dinner rolls,poultry,ice cream]
rule45	0.052	0.584	1.566	spaghetti sauce	<---	[dinner rolls,poultry,juice]
rule56	0.055	0.624	1.565	ice cream	<---	[paper towels,eggs,pasta]
rule51	0.052	0.567	1.565	paper towels	<---	[eggs,dinner rolls,pasta]
rule12	0.05	0.588	1.564	mixes	<---	[dishwashing liquid/detergent,poultry,laun...
rule22	0.051	0.617	1.558	cereals	<---	[cheeses,bagels,sandwich bags]
rule55	0.054	0.656	1.556	poultry	<---	[dinner rolls,spaghetti sauce,laundry dete...
rule1	0.05	0.594	1.544	aluminum foil	<---	[yogurt,toilet paper,juice]
rule8	0.05	0.588	1.528	yogurt	<---	[cheeses,cereals,coffee/tea]
rule9	0.05	0.594	1.52	cheeses	<---	[yogurt,cereals,coffee/tea]
rule28	0.051	0.574	1.518	laundry detergent	<---	[poultry,milk,cereals]
rule27	0.051	0.637	1.512	poultry	<---	[dinner rolls,spaghetti sauce,cereals]
rule23	0.051	0.58	1.505	bagels	<---	[cheeses,cereals,sandwich bags]
rule48	0.052	0.584	1.502	dinner rolls	<---	[paper towels,eggs,pasta]
rule870	0.083	0.563	1.498	individual meals	<---	[sandwich loaves,lunch meat]
rule0	0.05	0.576	1.497	yogurt	<---	[toilet paper,juice,aluminum foil]
rule1243	0.099	0.579	1.49	dinner rolls	<---	[spaghetti sauce,poultry]
rule37	0.052	0.578	1.487	dinner rolls	<---	[eggs,pasta,soda]
rule525	0.078	0.56	1.486	juice	<---	[shampoo,spaghetti sauce]
rule14	0.05	0.576	1.484	dishwashing liquid/detergent	<---	[poultry,laundry detergent,mixes]

➢ INFERENCE :

- As we can see each row contains different rule .
- It has created multiple rules on the basis of threshold limit that we have set in the Association Rule Learner Node and whichever has a higher lift value we recommend that product to the customer .
- Consequent column contains recommended products and we have sorted the lift values from higher to lower for the better recommendations.

➢ INSIGHTS AND RECOMMENDATIONS :

- Generally we recommend the products that are listed in consequent feature which has a higher lift value.
- The product having higher lift value have the higher probability of being purchased by the customer.
- So we can make offers Ex. If a customer buying "cereals, sandwich bags and bagels" we can recommend "cheese" at low cost as an offer which creates higher probability of being purchased by the customer.
- We can make another offer Ex. If customer buying a combo like " cereals, bagels and sandwich bags" in maximum quantity or two so we can offer them free "cheese" Or "Buy two sandwich bags and get one for free ".
- Paper towels, pasta and cheese have a higher lift value which means it has a maximum probability of getting purchased by the customers so creating offers on these products can increase the sells of these products.