Project on Marketing & Retail Analytics

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. Can you suggest the best combo meals?

Contents:

- Basic Information of the data
- Results of Exploratory Data Analysis
- Results of Market Basket Analysis
- Finding best basket combinations along with its suggested item.

Summary:

Exploratory Data Analysis (EDA) was done on the dataset to understand whether data had any discrepancies. EDA helped in drawing useful insights & inferences from the data to understand business & customer patterns. Market Basket Analysis was done to understand and find best and most likely combination of items to be seen in an order. This helped in finding the best possible product combinations, by which customers can be analyzed easily and new strategies for further business growth can also be devised.

Basic Information of the dataset:

```
<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  datetime64[ns]
 1   Order_id  20641 non-null  object
 2   Product   20641 non-null  object
dtypes: datetime64[ns](1), object(2)
memory usage: 483.9+ KB
```

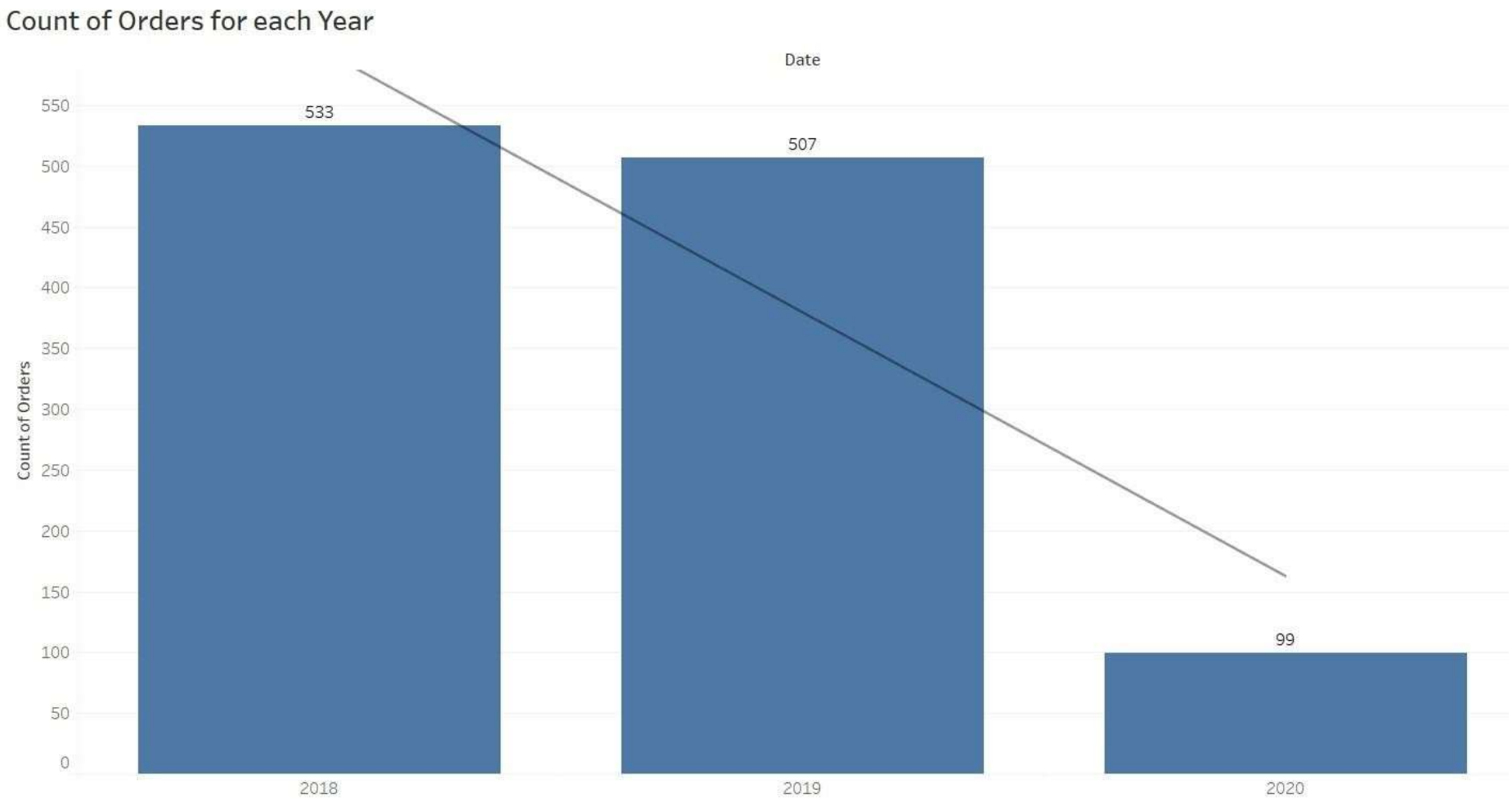
Summary of the dataset:

	count	unique	top	freq	first	last
Date	20641	603	2019-08-02 00:00:00	183	2018-01-01	2020-12-02
Order_id	20641.0	1139.0	226.0	34.0	NaT	NaT
Product	20641	37	poultry	640	NaT	NaT

Insights:

- There are no null & missing values in the dataset.
- There are 2 object datatype variables (namely Order-Id & Product) and 1 DateTime datatype variable (namely Date).
- There are 20641 rows and 3 columns.
- Summary table includes most frequent and unique occurrences of data points.

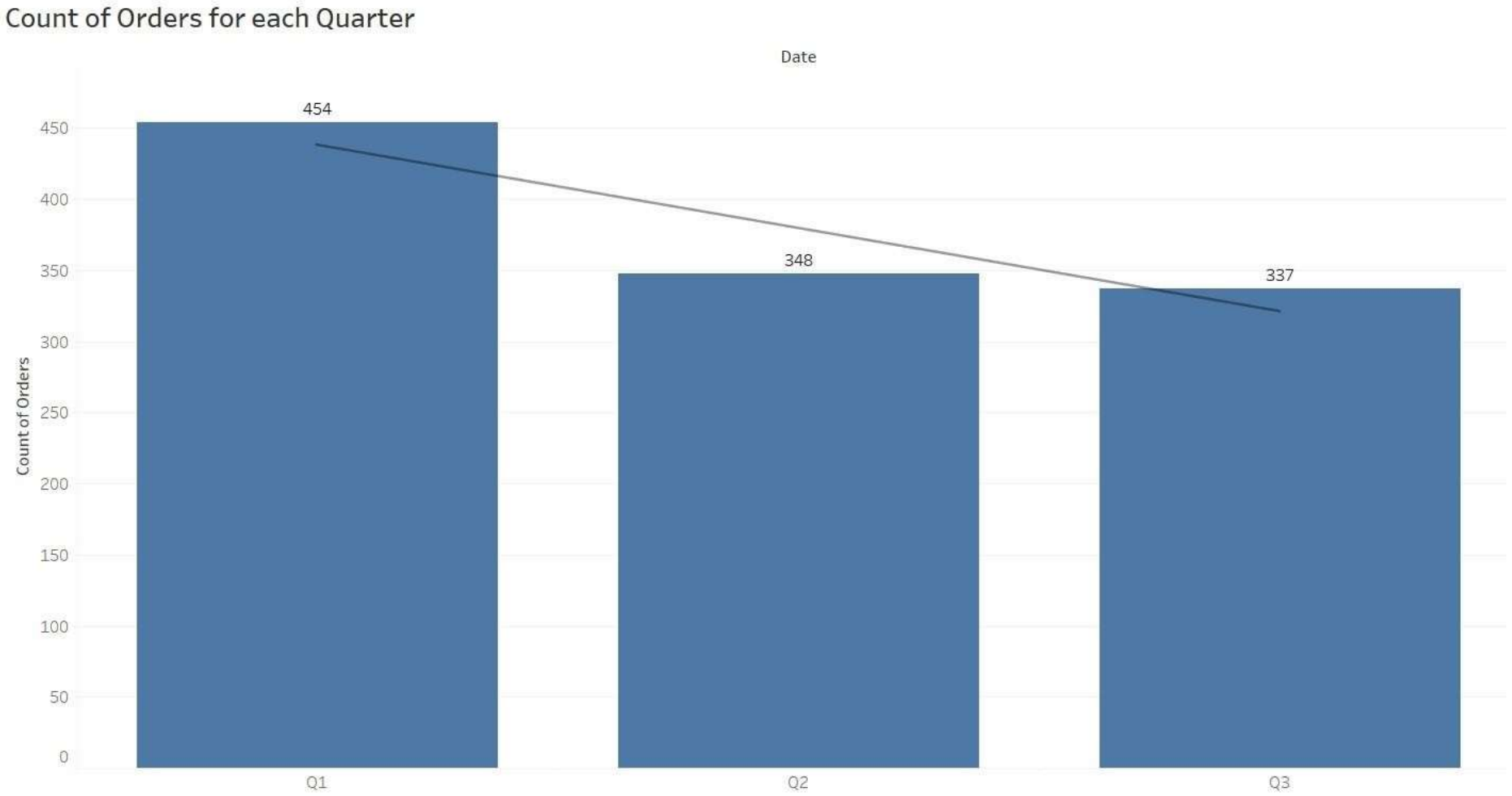
Plot for Count of Orders for each Year



Insights:

From the plot above, we can say that, the year 2018 had the highest number of orders (533). However, the trend line shows decrease in the number of orders over the years.

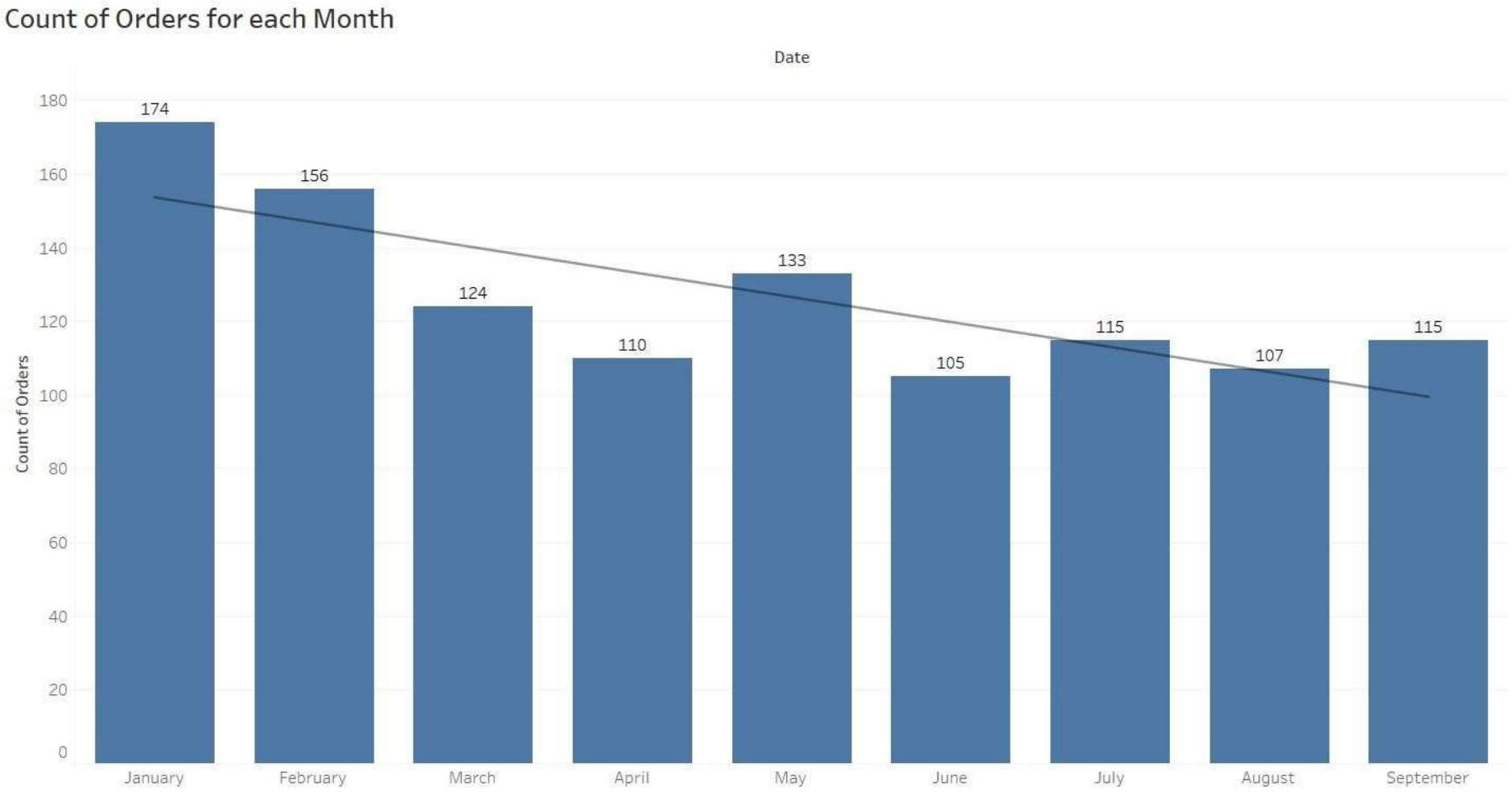
Plot for Count of Orders for each Quarter



Insights:

From the plot above, we can say that, the Quarter-1 had the highest number of orders (454). However, the trend line shows decrease in the number of orders over the quarters of a year.

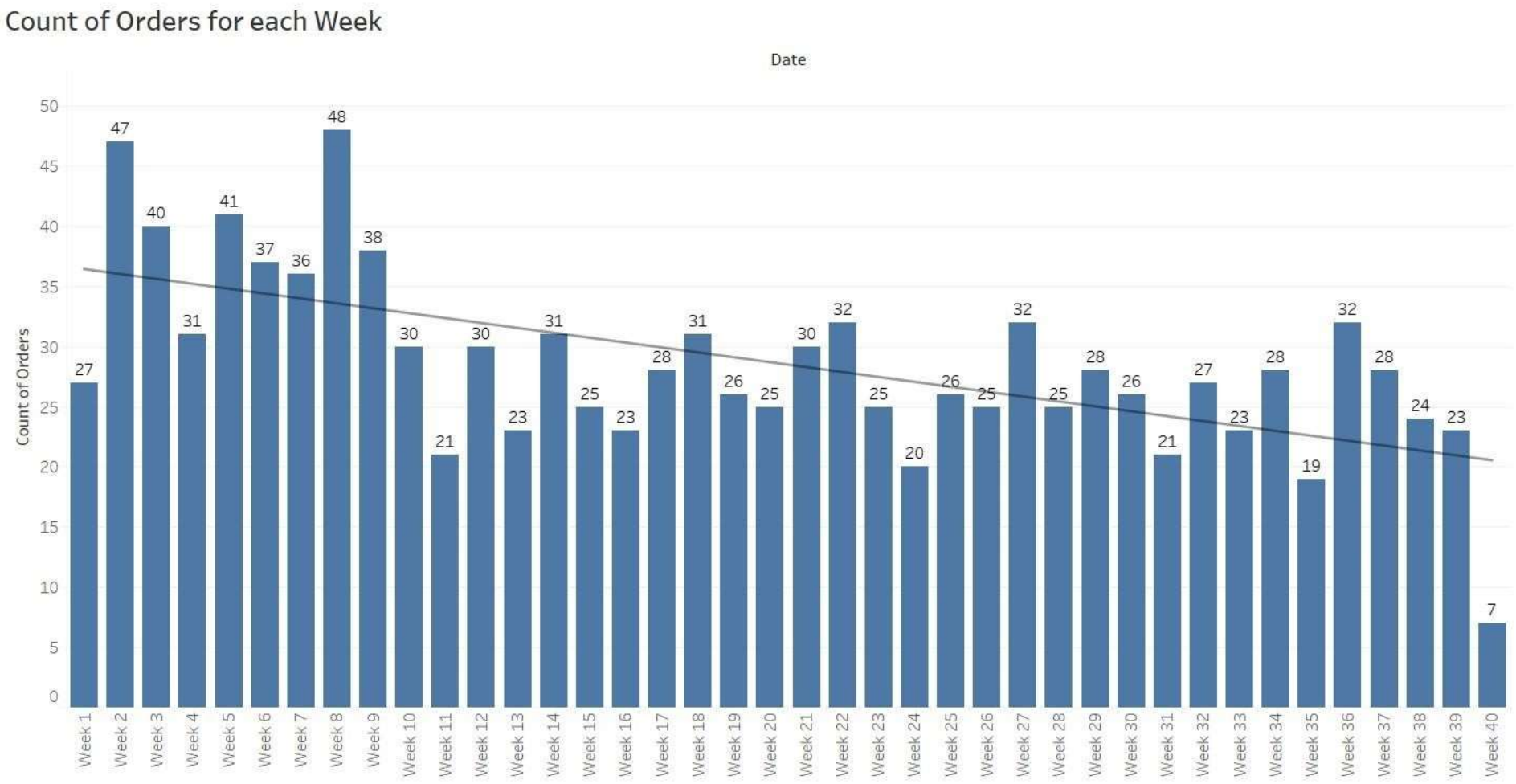
Plot for Count of Orders for each Month



Insights:

From the plot above, we can say that, the month of January had the highest number of orders (174). However, the trend line shows decrease in the number of orders over the months of a year.

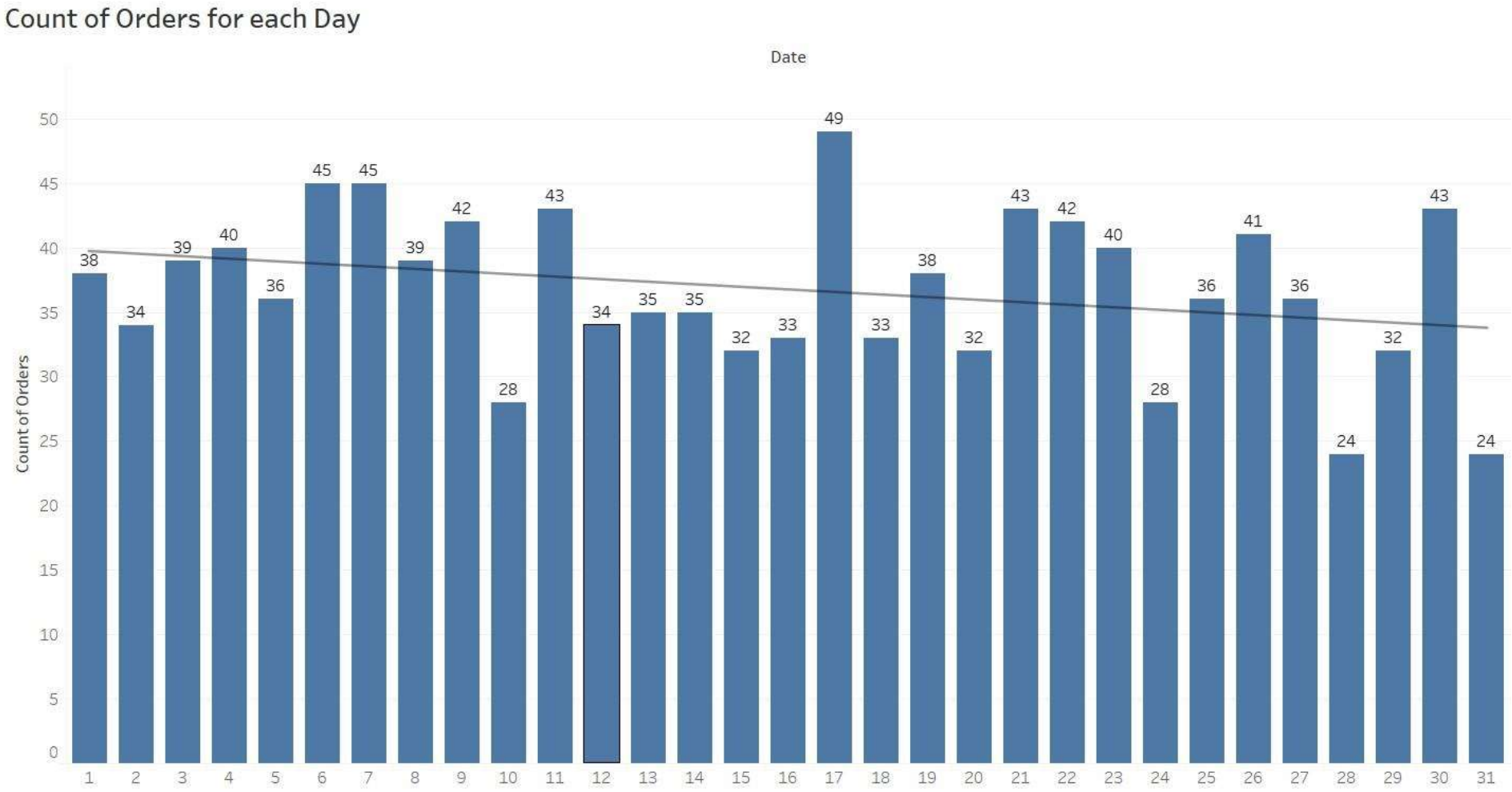
Plot for Count of Orders for each Week



Insights:

From the plot above, we can say that, the week-8 had the highest number of orders (48). However, the trend line shows decrease in the number of orders over the weeks of a year.

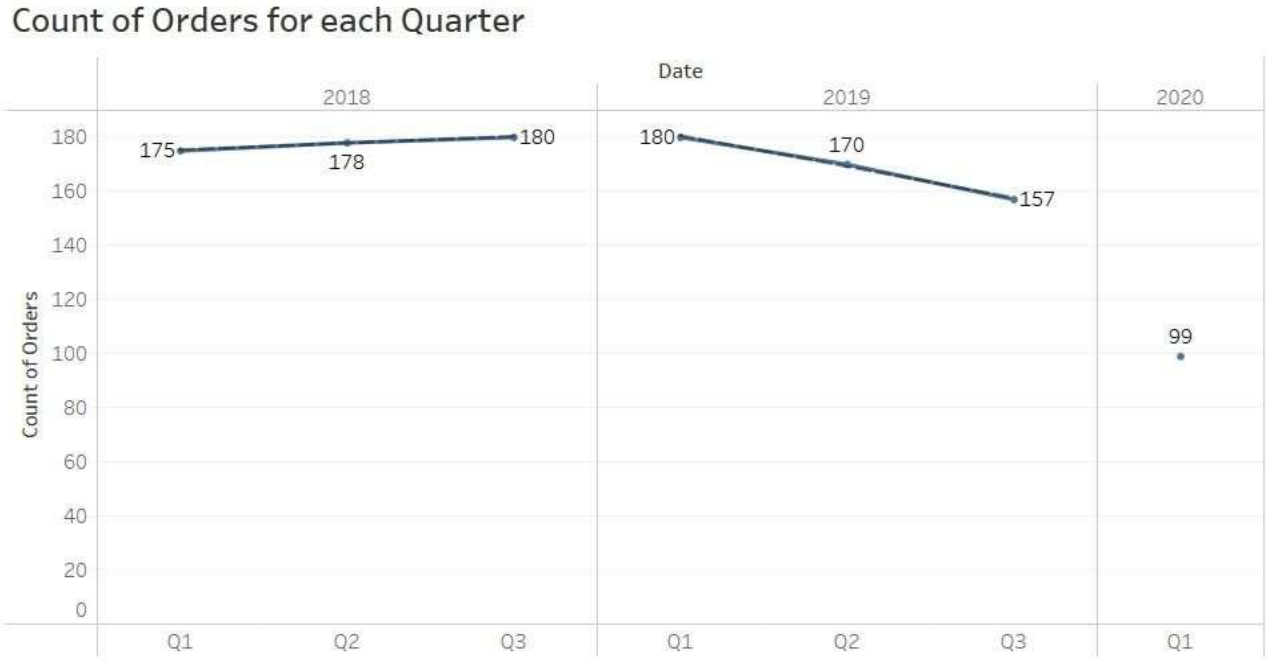
Plot for Count of Orders for each Day

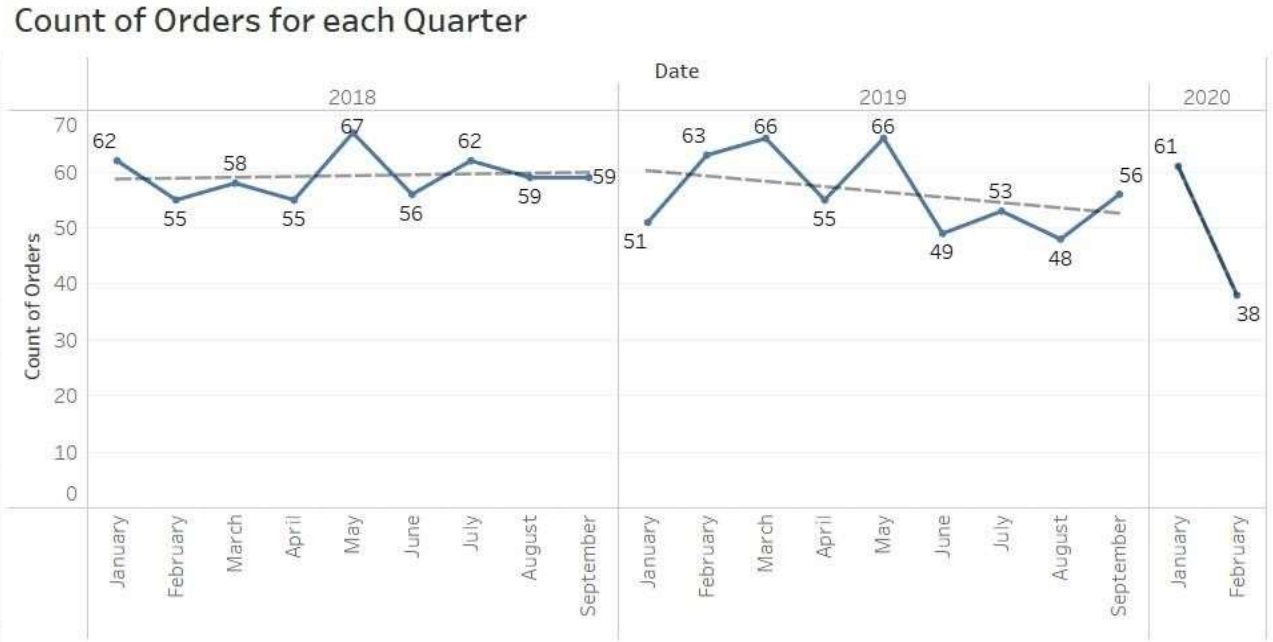


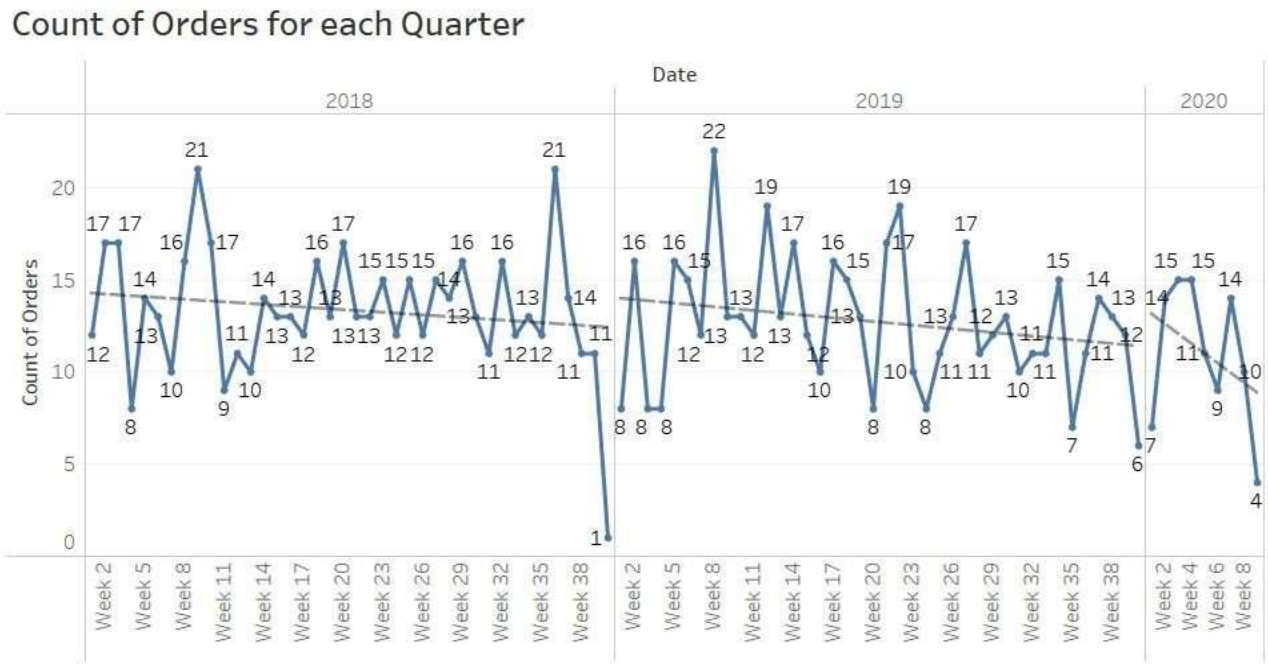
Insights:

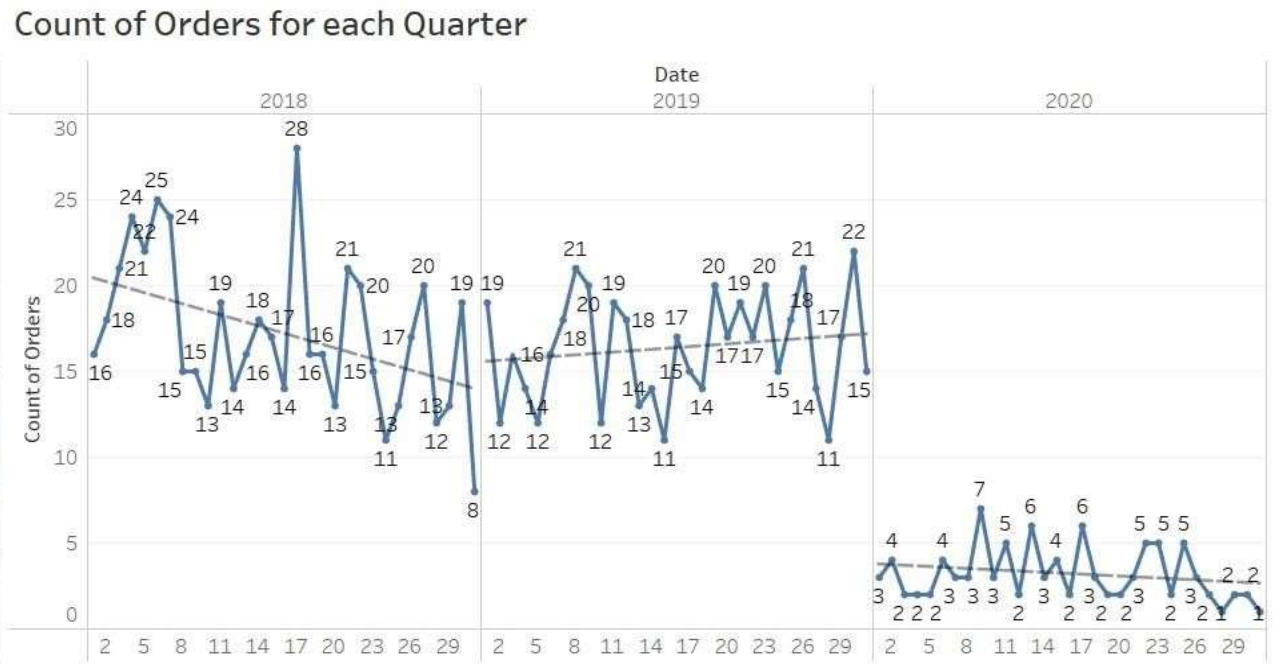
From the plot above, we can say that, the date 17 of all months had the highest number of orders (49). However, the trend line shows decrease in the number of orders over the days of the month of a year.

Plot for Understanding Timely Trends

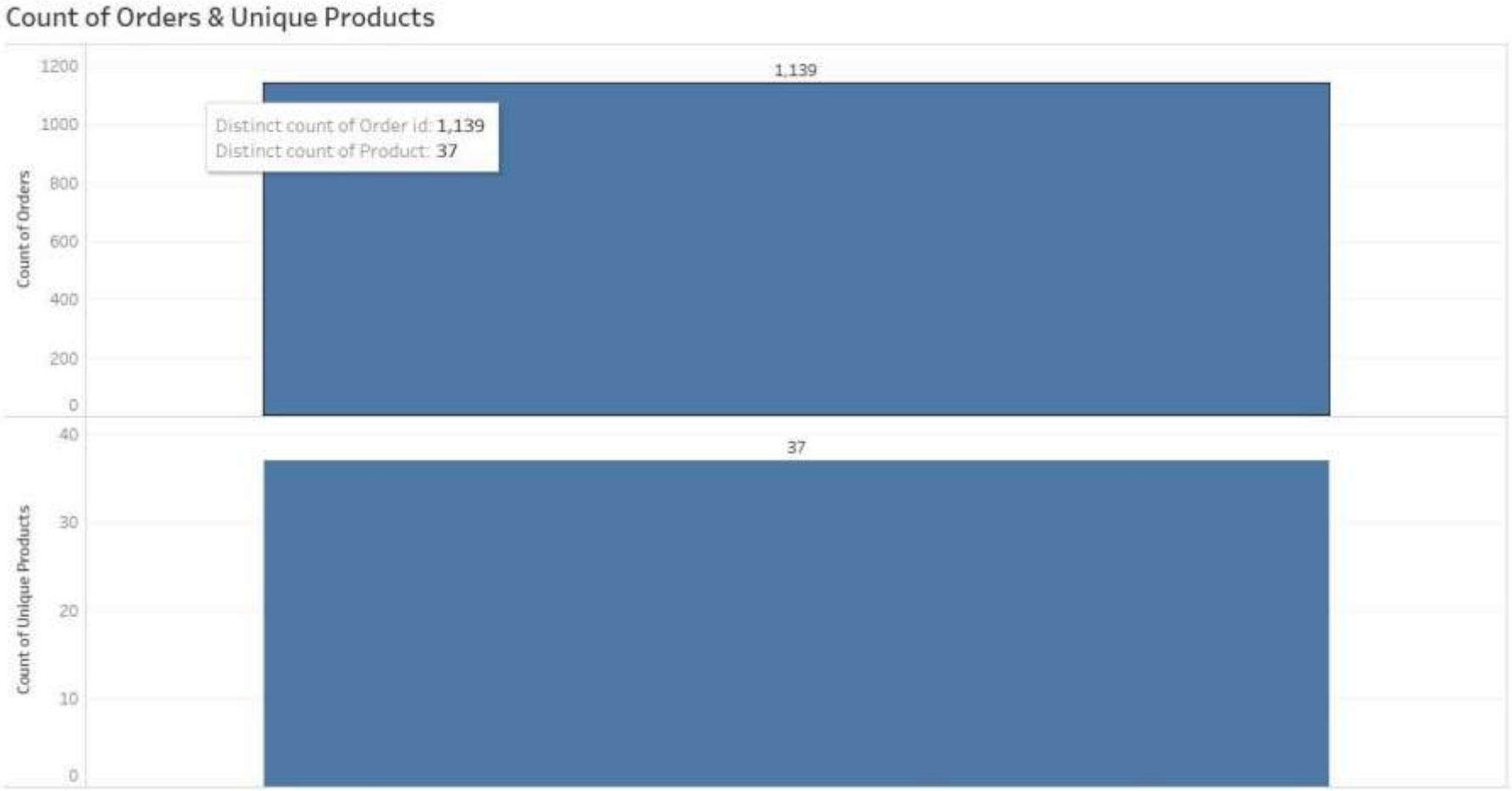








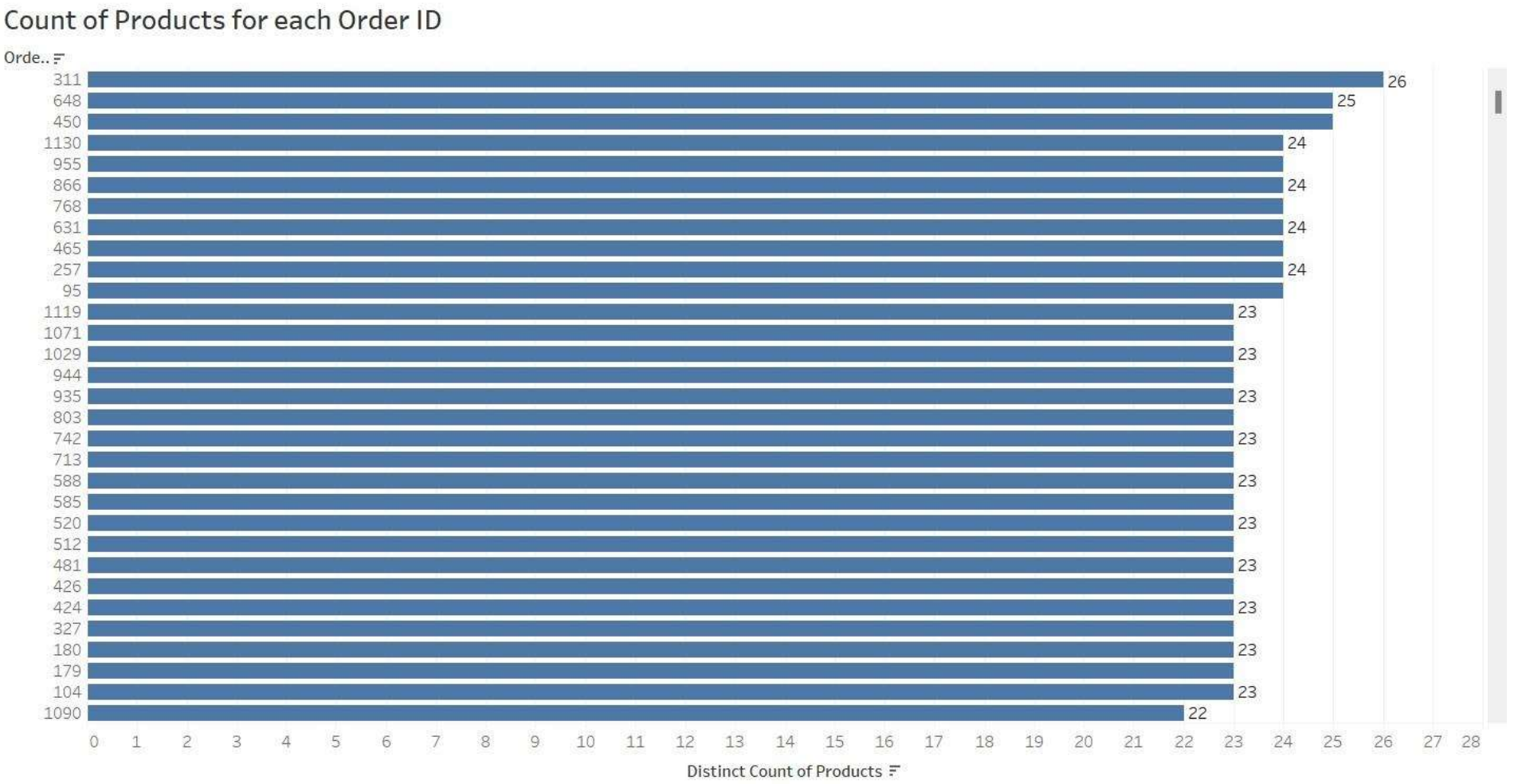
Plot for Count of Orders & Unique Products



Insights:

From the plot above, we can say that, there are 1139 orders consisting of 37 unique products.

Count of Products for each Order ID

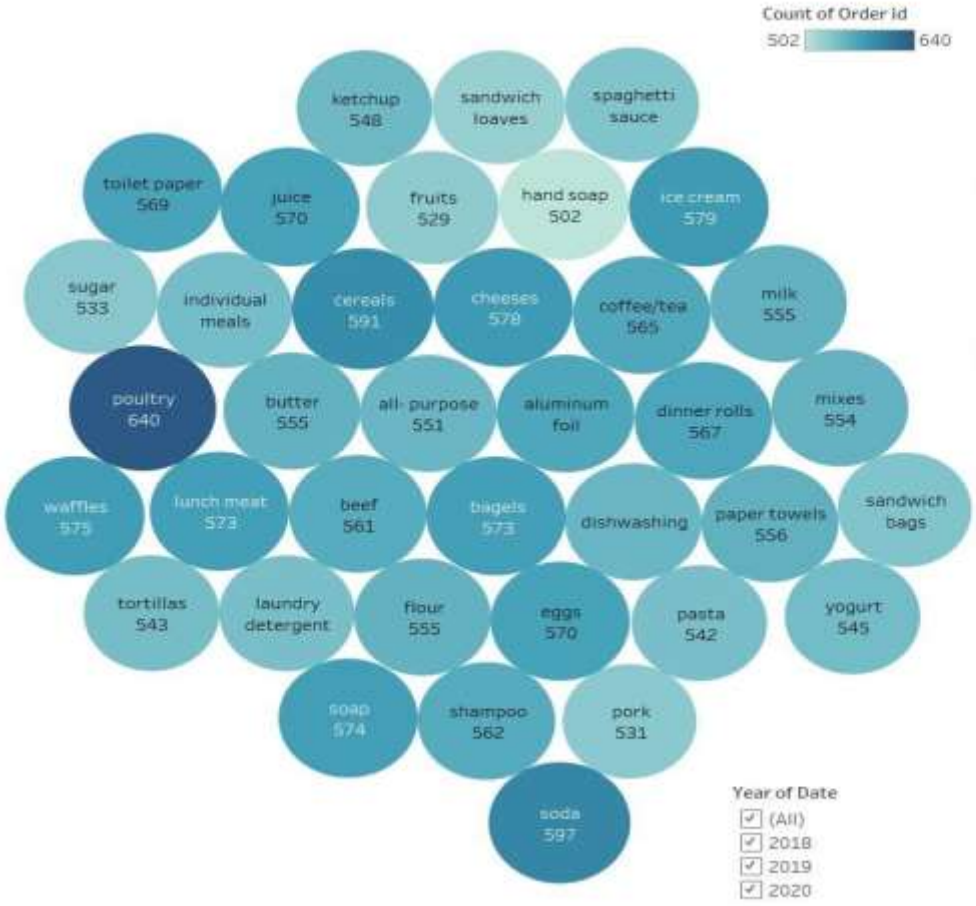


Insights:

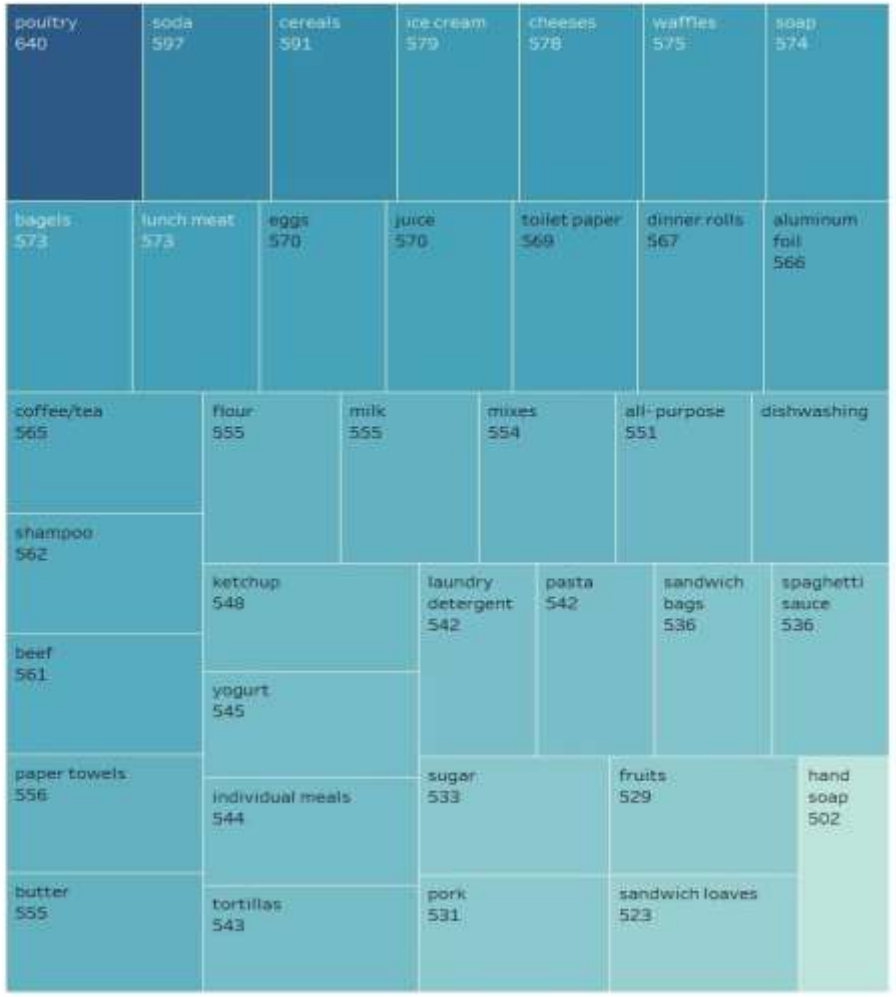
From the plot above, we can say that, the Order ID ‘311’ had the maximum number of items/products.

Plot for Frequency Distribution of Each Product

Bubble Chart for Ordering Frequency of Each Product



Tree Map for Ordering Frequency of Each Product



Insights:

From the plot above, we can say that, the highest ordered product is poultry and the least ordered is hand soap.

What is Market Basket Analysis?

Market Basket Analysis is one of the key techniques used to uncover associations between items. It works by looking for combinations of items that occur together frequently in transactions. To put it another way, it allows us to identify relationships between the items that people buy.

What are Association Rules?

- Association Rules are widely used to analyze 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.
- Generally, Association Rule Mining can be viewed in a two-step process:-

1) **Find all Frequent item-sets:** By definition, each of these item-sets will occur at least as frequently as a pre-established minimum support count, min sup.
2) **Generate Association Rules from the Frequent item-sets:** By definition, these rules must satisfy minimum support and minimum confidence.

- Association Rule Mining is primarily used when we want to identify an association between different items in a set, then find frequent patterns in a transactional database.
- Strength of any rule is decided on 3 metrics: Support, Confidence & Lift.

Metrics of Association rule:

1)Support: The percentage of transactions that contain all of the items in an itemset. The higher the support the more frequently the itemset occurs. Rules with a high support are preferred since they are likely to be applicable to a large number of future transactions.

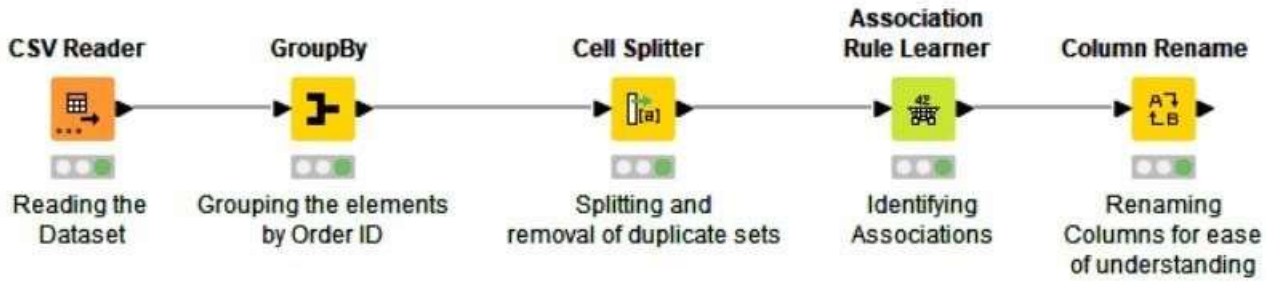
2)Confidence: The probability that a transaction that contains the items on the left hand side of the rule (in our example, pencil and paper) also contains the item on the right hand side (a rubber). The higher the confidence, the greater the likelihood that the item on the right hand side will be purchased or, in other words, the greater the return rate you can expect for a given rule.

3)Lift: The probability of all of the items in a rule occurring together (otherwise known as the support) divided by the product of the probabilities of the items on the left and right hand side occurring as if there was no association between them. Overall, lift summarizes the strength of association between the products on the left and right hand side of the rule; the larger the lift the greater the link between the two products.

The thresholds at which to set the support and confidence are user-specified and are likely to vary between transaction data sets. In this case, the values of,

Support: 0.02
Confidence: 0.4

KNIME Workflow:



Identified Associations: First 20 of 348944 rows

Row ID	Support	Confide...	Lift	Recommended Item	implies	Items
rule0	0.02	0.426	1.108	yogurt	<---	[pasta,pork,soap]
rule1	0.02	0.404	1.135	pork	<---	[yogurt,dishwashing liquid/detergent,hand soap]
rule2	0.02	0.404	1.076	all- purpose	<---	[yogurt,ice cream,hand soap]
rule3	0.02	0.426	1.108	yogurt	<---	[butter,hand soap,beef]
rule4	0.02	0.418	1.115	beef	<---	[yogurt,butter,hand soap]
rule5	0.02	0.418	1.07	cheeses	<---	[yogurt,butter,hand soap]
rule6	0.02	0.418	0.992	poultry	<---	[yogurt,butter,hand soap]
rule7	0.02	0.418	1.103	coffee/tea	<---	[yogurt,butter,hand soap]
rule8	0.02	0.404	1.064	coffee/tea	<---	[yogurt,ice cream,hand soap]
rule9	0.02	0.404	1.019	cereals	<---	[yogurt,dishwashing liquid/detergent,hand soap]
rule10	0.02	0.411	1.155	pork	<---	[sandwich loaves,toilet paper,sandwich bags]
rule11	0.02	0.411	1.116	sandwich bags	<---	[eggs,fruits,pork]
rule12	0.02	0.404	1.035	eggs	<---	[pasta,lunch meat,pork]
rule13	0.02	0.434	1.158	all- purpose	<---	[sandwich loaves,laundry detergent,pork]
rule14	0.02	0.426	1.198	pork	<---	[all- purpose,hand soap,individual meals]
rule15	0.02	0.404	1.135	pork	<---	[all- purpose,hand soap,bagels]
rule16	0.02	0.434	1.23	flour	<---	[sandwich loaves,laundry detergent,pork]
rule17	0.02	0.404	1.143	flour	<---	[fruits,milk,pork]
rule18	0.02	0.411	1.068	aluminum foil	<---	[butter,fruits,pork]
rule19	0.02	0.426	1.158	butter	<---	[shampoo,fruits,pork]
rule20	0.02	0.411	1.114	shampoo	<---	[butter,fruits,pork]

Identified Associations: Best Combos & its Suggestion

Row ID	Support	Confidence	Lift	Recommended Item	implies	Items
rule43600	0.02	0.852	2.349	paper towels	<---	[eggs,dinner rolls,ice cream,...]
rule6699	0.02	0.793	2.27	sandwich loaves	<---	[yogurt,hand soap,toilet paper,...]
rule2891	0.02	0.852	2.267	mixes	<---	[yogurt,dishwashing liquid/detergent,all- purpose,...]
rule43612	0.02	0.821	2.265	paper towels	<---	[eggs,dinner rolls,poultry,...]
rule102758	0.023	0.839	2.258	ketchup	<---	[tortillas,coffee/tea,juice,...]
rule91320	0.022	0.833	2.244	pasta	<---	[paper towels,dishwashing liquid/detergent,eggs,...]
rule75733	0.022	0.781	2.236	sandwich loaves	<---	[yogurt,hand soap,aluminum foil,...]
rule43597	0.02	0.885	2.219	ice cream	<---	[paper towels,eggs,dinner rolls,...]
rule69347	0.021	0.828	2.218	spaghetti sauce	<---	[waffles,laundry detergent,mixes,...]
rule64898	0.021	0.828	2.208	beef	<---	[poultry,fruits,hand soap,...]
rule139952	0.026	0.857	2.194	cheeses	<---	[paper towels,cereals,sandwich bags,...]
rule214210	0.031	0.795	2.194	paper towels	<---	[eggs,ice cream,pasta,...]
rule8442	0.02	0.821	2.191	beef	<---	[shampoo,fruits,lunch meat,...]
rule102364	0.023	0.812	2.188	ketchup	<---	[toilet paper,mixes,coffee/tea,...]
rule43606	0.02	0.793	2.187	paper towels	<---	[dishwashing liquid/detergent,eggs,ice cream,...]
rule98395	0.023	0.788	2.183	sugar	<---	[poultry,flour,waffles,...]
rule9326	0.02	0.852	2.18	soda	<---	[bagels,pasta,individual meals,...]
rule81913	0.022	0.781	2.165	sugar	<---	[poultry,bagels,lunch meat,...]
rule12029	0.02	0.821	2.161	milk	<---	[eggs,poultry,beef,...]
rule230503	0.032	0.783	2.158	paper towels	<---	[eggs,ice cream,pasta,...]
rule108101	0.024	0.818	2.157	soap	<---	[spaghetti sauce,all- purpose,sandwich bags,...]

Insights:

- The best basket or combination of item is (eggs, dinner rolls, ice cream) as it has the highest Lift value as it shows strength of association. And the suggested item along with it is paper towels. Its support value is 0.02 and Confidence value of 0.852.
- Similarly, whenever someone buys items which matches a particular basket as in the above mentioned table, recommended items can be suggested to them.

Business Recommendations:

- We should focus on the best baskets suggested as they are more likely to be purchases from the store again by the customers.
- We should understand the baskets of our leading competitors' as well to understand our performance in the market much better.
- We should devise new dynamic pricing system to determine best prices for products which would maintain balance between store's profits & customer satisfaction. This would lead to increase of store's goodwill amongst the customers.
- There should be some incentives given like discounts or buy some get some free to the customers on recommended items of their purchase basket. This will help retain our existing customers against our competitors and attract new customer's as well.
- We should also account for seasonal item purchases (like Christmas tree & its decorative items on Christmas, etc.)
- The store should not invest much more time & money on poor baskets as their returns might no be that substantial.

THE END