

GROCERY STORE DATA

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.

PART B

Exploratory Analysis --> Exploratory Analysis of data & an executive summary (in PPT) of your top findings, supported by graphs. --> Are there trends across months/years/quarters/days etc. that you are able to notice?

- The overall summary of the data set in terms of info, describe, identification of null values (if any) we made using Jupiter notebook
- EDA for Grocery store was done using Tableau tool.
- Detailed analysis in terms of yearly, quarterly, monthly and daily analysis are made to provide in depth information's of the data.

```
In [4]: df.head()
```

```
Out[4]:
```

	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

```
In [5]: df.tail()
```

```
Out[5]:
```

	Date	Order_id	Product
20636	25-02-2020	1138	soda
20637	25-02-2020	1138	paper towels
20638	26-02-2020	1139	soda
20639	26-02-2020	1139	laundry detergent
20640	26-02-2020	1139	shampoo

```
In [7]: df.shape
```

```
Out[7]: (20641, 3)
```

Data Set

- The data set was loaded perfectly without any issues.
- There were 20641 rows and 3 columns.

```
<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
```

```
In [9]: df.describe()
```

```
Out[9]:
```

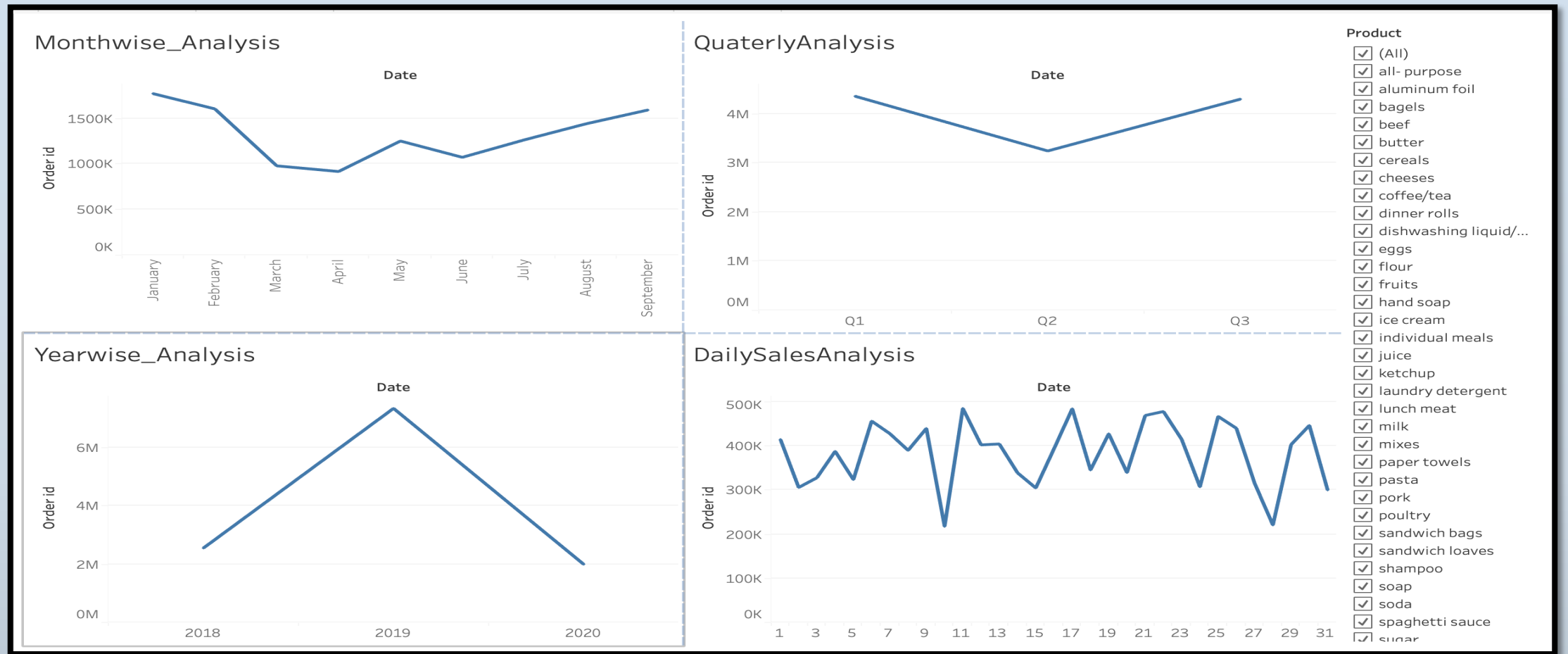
	Order_id
count	20641.000000
mean	575.986289
std	328.557078
min	1.000000
25%	292.000000
50%	581.000000
75%	862.000000
max	1139.000000

```
In [11]: df.isnull().sum()
```

```
Out[11]: Date        0
Order_id    0
Product     0
dtype: int64
```

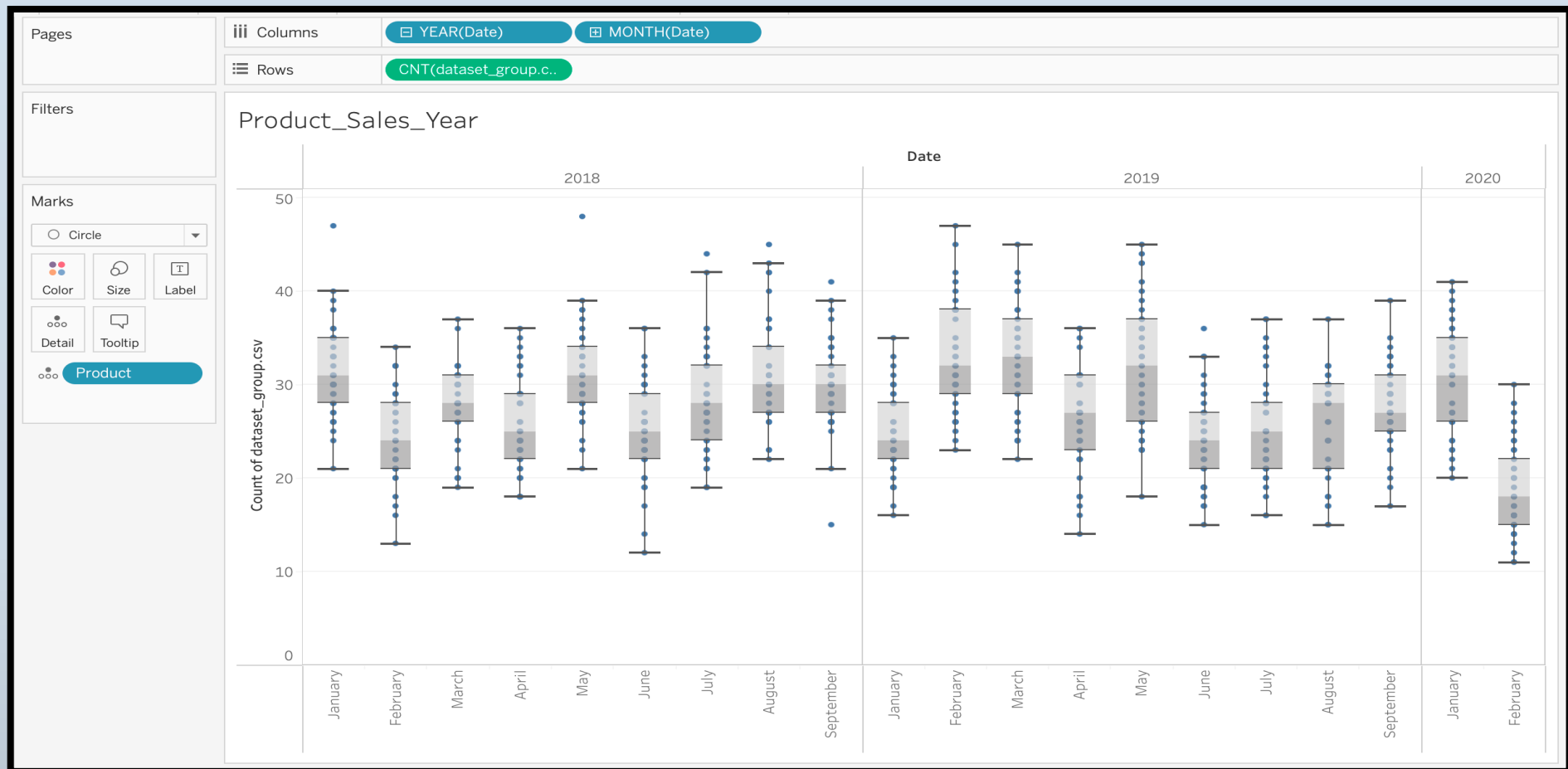
Data Set

- There are no null values present in the data set.
- A max of 1139 orders were made according to this data set.
- Each order can have as many items as possible and there are many listed inside each order id.



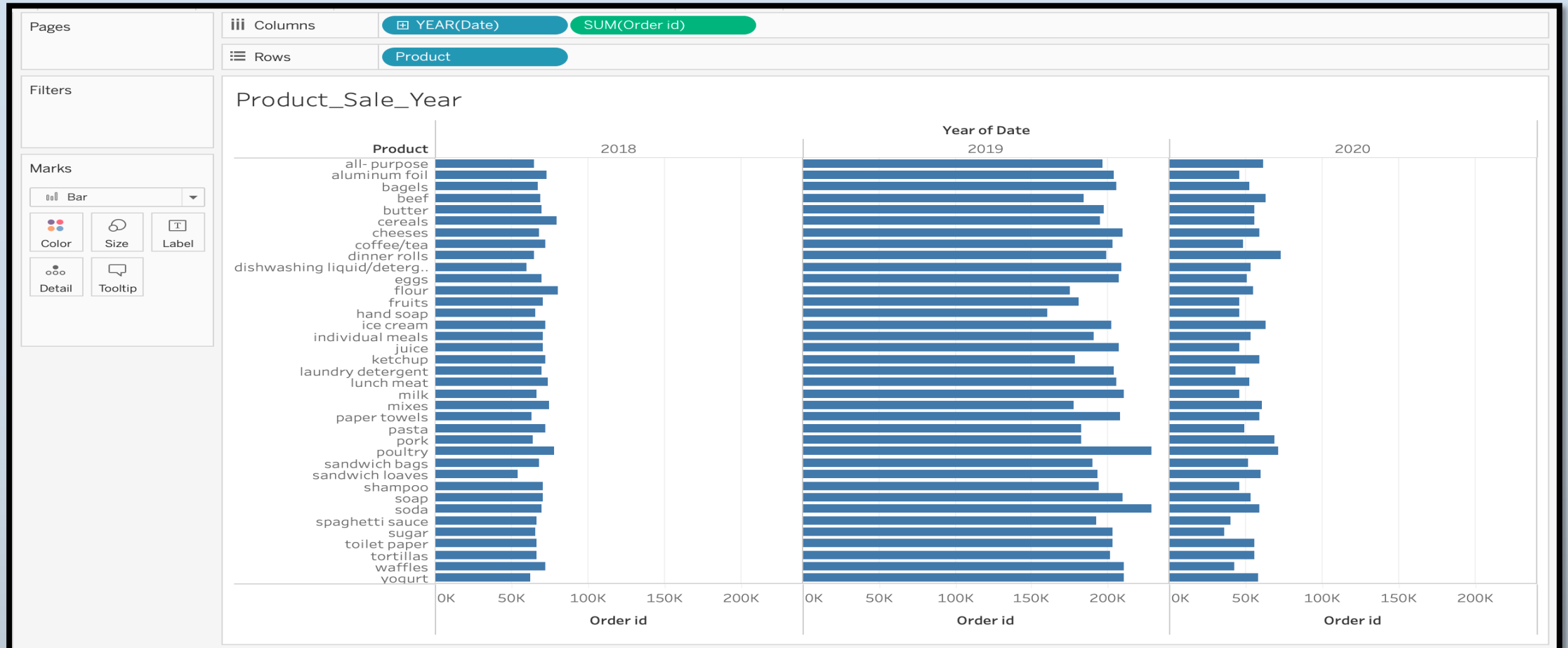
Customer order – Monthwise, Yearwise, Quaterly, Daily

- This chart describe how the orders are been made by the customers in terms of years, quarterly, monthly and daily.
- This data set consists of order details made for 3 different years from 2018 to 2020 and for first nine months of an year.
- Highest orders were made during 2019.
- There is slight dip in orders for Q2.
- Month wise Jan and Feb has maximum orders and there is a slight dip during the mid of the year, which again starts to gradually increase during end of the year.
- This is indication of seasonality present in the data set.



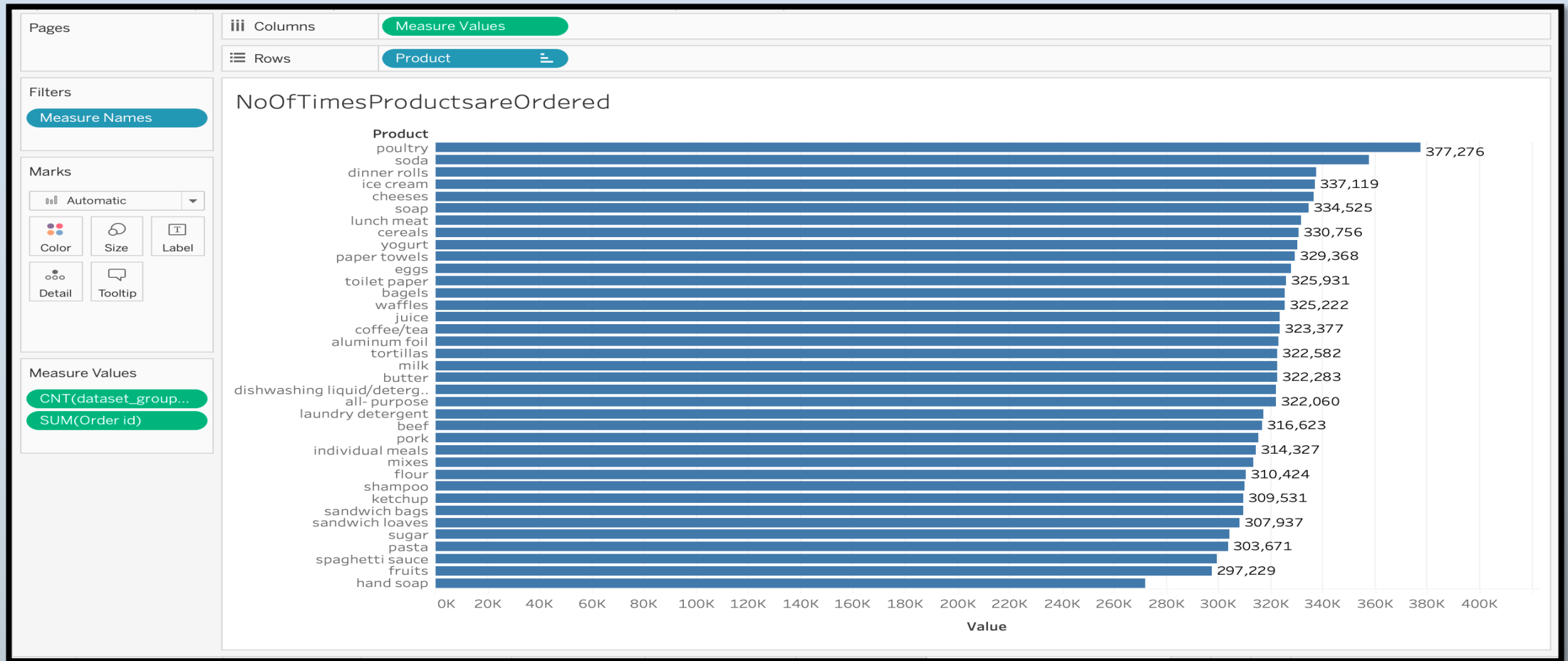
Customer order across years and months

- The box plot suggest the orders made across years and months by the customers.



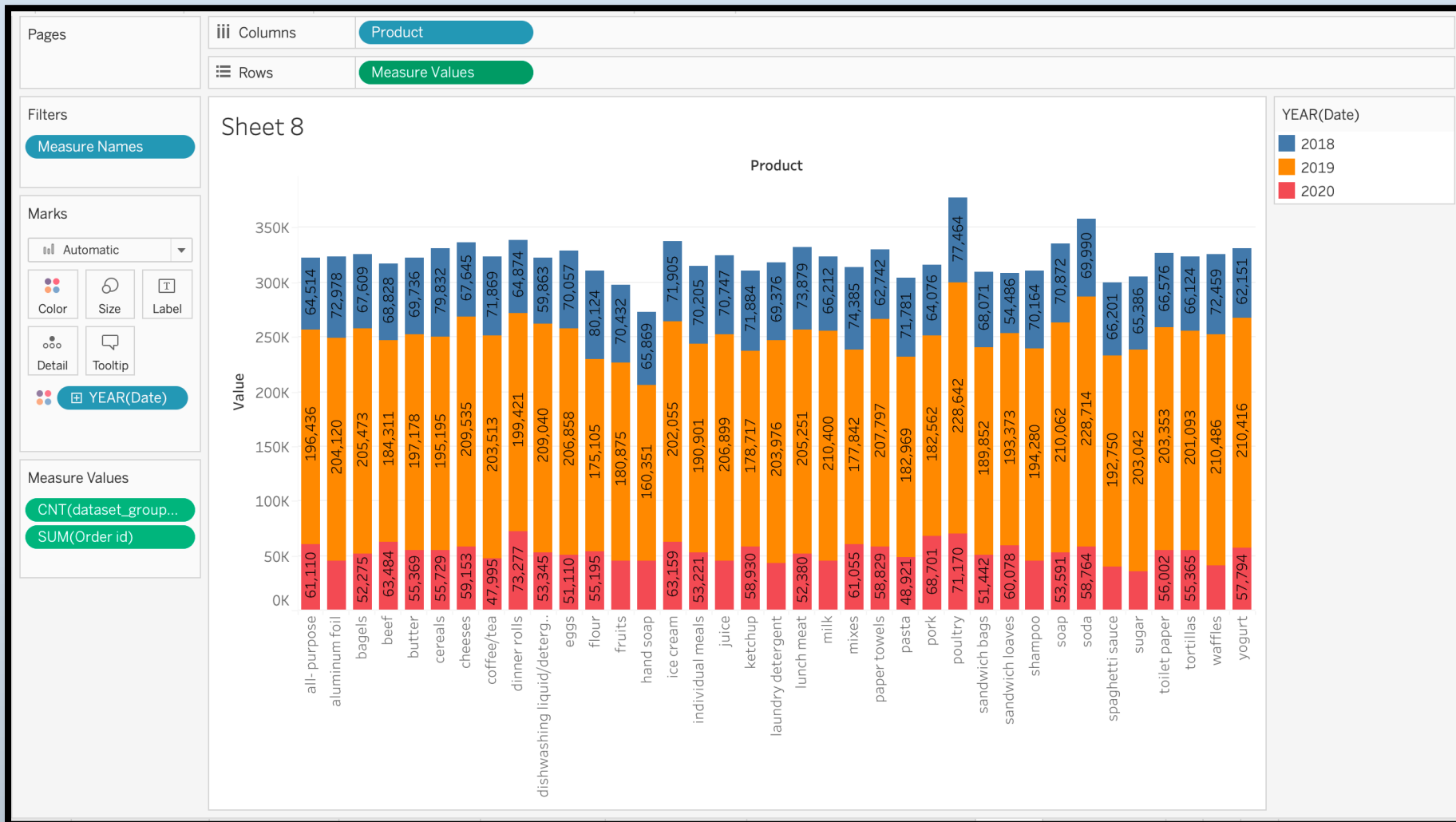
Products ordered by the customers

- Each and every product ordered by the customers.
- This will certainly help to provide suggestions based on the market basket analysis technique.



No of times products are ordered

- Number of times each product has been ordered by the customers.
- In this list, poultry has been ordered for maximum number of time.
- Hence the grocery store can provide a package or bundle like stuff to attract the customers purchase more items.



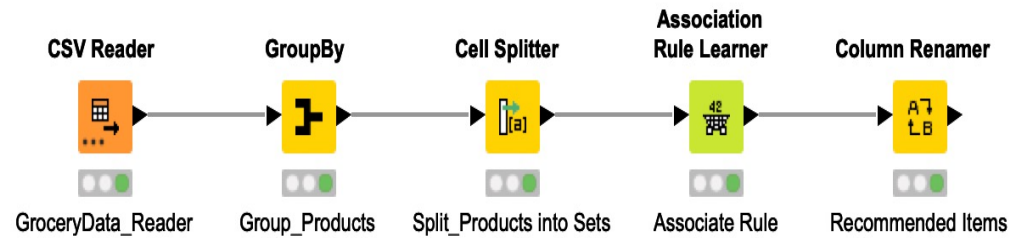
Product ordered every year

- Each product or items been ordered for every available years in the data set.

PART B

Use of Market Basket Analysis (Association Rules) --> Write Something about the association rules and its relevance in this case --> Add KNIME workflow image --> Write about threshold values of Support and Confidence

Knime WorkFlow



Knime Workflow

About MBA technique and Association rules

- MBA – Market basket analysis certainly talks about the how to influence the customers to buy products
- That not initially there in there list.
- In this data set there are list of different products been listed which are been purchased by the customers previously.
- From this list by using the MBA technique we need to suggest to the customers on what is been or can be purchased.
- For instance a customers would have purchased milk, cereal and water bottle previously in the store.
- On his next visit to the same store a list in terms of his previous purchase should be displayed and along with this list there
- Can also be suggestions made.
- For instance customers who buy milk, cereal and water bottle will mostly buy chocolates as well.
- This suggestions will definitely influence the customers to buy products which is not in there list also certainly it will help to
- Increase the sales and revenue of the store.
- There are no certain values which can be given for association rules.
- In this data set if the minimum support exceeds 10% the 'Association Rule Learner' is not getting executed.
- Hence I have given the minimum support to be 0.05% and the minimum confidence to be 0.4%.

PART B

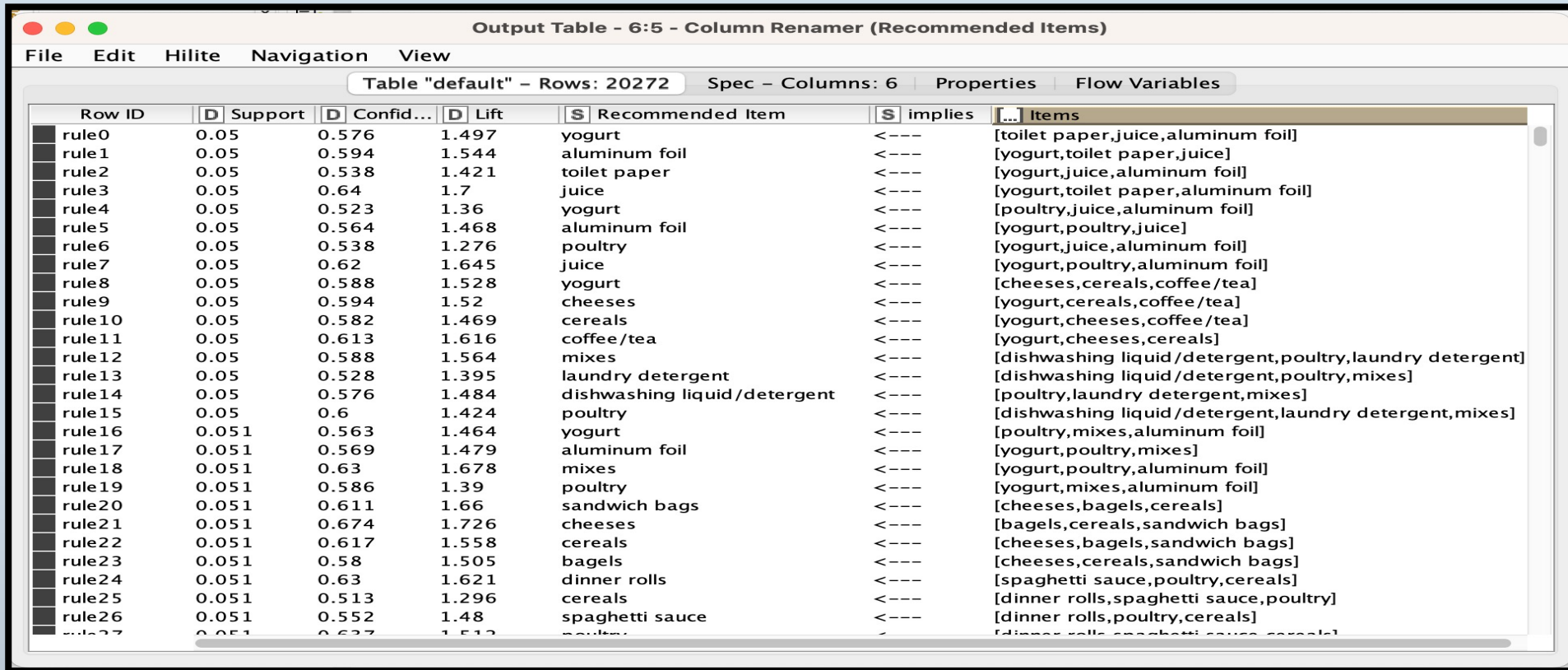
Associations Identified --> Put the associations in a tabular manner -->
Explain about support, confidence, & lift values that are calculated

Recommended list of items along with the confidence, support and Lift values.

Table "default" – Rows: 20272							Spec – Columns: 6	Properties	Flow Variables
Row ID	D Support	D Confid...	D Lift	S Conse...	S implies	Items			
rule211	0.055	0.649	1.791	paper tow...	<---	[eggs,ice cream,pasta]			
rule210	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 ...	<---	[cheeses,bagels,cereals]			
rule108	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce,poultry,laundry detergent]			
rule49	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]			
rule52	0.052	0.686	1.628	poultry	<---	[dinner rolls,spaghetti sauce,ice cream]			
rule58	0.052	0.634	1.627	eggs	<---	[paper towels,dinner rolls,pasta]			
rule59	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]			
rule209	0.055	0.63	1.616	eggs	<---	[paper towels,ice cream,pasta]			
rule11	0.05	0.613	1.616	coffee/tea	<---	[yogurt,cheeses,cereals]			
rule53	0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce,poultry,juice]			
rule44	0.052	0.628	1.61	eggs	<---	[dinner rolls,poultry,soda]			
rule110	0.054	0.598	1.603	spaghetti ...	<---	[dinner rolls,poultry,laundry detergent]			
rule29	0.051	0.604	1.589	milk	<---	[poultry,laundry detergent,cereals]			
rule51	0.052	0.59	1.581	spaghetti ...	<---	[dinner rolls,poultry,ice cream]			
rule54	0.052	0.584	1.566	spaghetti ...	<---	[dinner rolls,poultry,juice]			
rule208	0.055	0.624	1.565	ice cream	<---	[paper towels,eggs,pasta]			
rule60	0.052	0.567	1.565	paper tow...	<---	[eggs,dinner rolls,pasta]			
rule12	0.05	0.588	1.564	mixes	<---	[dishwashing liquid/detergent,poultry,laundry detergent]			
rule22	0.051	0.617	1.558	cereals	<---	[cheeses,bagels,sandwich bags]			
rule111	0.054	0.656	1.556	poultry	<---	[dinner rolls,spaghetti sauce,laundry detergent]			
rule1	0.05	0.594	1.544	aluminum ...	<---	[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 d...	<---	[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]			
rule57	0.052	0.584	1.502	dinner rolls	<---	[paper towels,eggs,pasta]			
rule18611	0.083	0.563	1.498	individual ...	<---	[sandwich loaves,lunch meat]			
rule0	0.05	0.576	1.497	yogurt	<---	[toilet paper,juice,aluminum foil]			
rule19354	0.099	0.579	1.49	dinner rolls	<---	[spaghetti sauce,poultry]			
rule46	0.052	0.578	1.487	dinner rolls	<---	[eggs,pasta,soda]			
rule17292	0.078	0.56	1.486	juice	<---	[shampoo,spaghetti sauce]			
rule14	0.05	0.576	1.484	dishwashi...	<---	[poultry,laundry detergent,mixes]			
rule26	0.051	0.552	1.48	spaghetti ...	<---	[dinner rolls,poultry,cereals]			
rule17	0.051	0.569	1.479	aluminum ...	<---	[yogurt,poultry,mixes]			
rule48	0.052	0.546	1.471	pasta	<---	[eggs,dinner rolls,soda]			
rule18717	0.083	0.514	1.47	sandwich l...	<---	[cheeses,ketchup]			
rule47	0.052	0.573	1.469	eggs	<---	[dinner rolls,pasta,soda]			
rule10	0.05	0.582	1.469	cereals	<---	[yogurt,cheeses,coffee/tea]			
rule5	0.05	0.564	1.468	aluminum ...	<---	[yogurt,poultry,juice]			
rule19127	0.086	0.547	1.467	spaghetti ...	<---	[dinner rolls,juice]			
rule16	0.051	0.563	1.464	yogurt	<---	[poultry,mixes,aluminum foil]			

List of items with confidence, support and lift values

Rename of the column name consequent to 'Recommended Items'



Row ID	Support	Confid...	Lift	Recommended Item	implies	Items
rule0	0.05	0.576	1.497	yogurt	<---	[toilet paper,juice,aluminum foil]
rule1	0.05	0.594	1.544	aluminum foil	<---	[yogurt,toilet paper,juice]
rule2	0.05	0.538	1.421	toilet paper	<---	[yogurt,juice,aluminum foil]
rule3	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum foil]
rule4	0.05	0.523	1.36	yogurt	<---	[poultry,juice,aluminum foil]
rule5	0.05	0.564	1.468	aluminum foil	<---	[yogurt,poultry,juice]
rule6	0.05	0.538	1.276	poultry	<---	[yogurt,juice,aluminum foil]
rule7	0.05	0.62	1.645	juice	<---	[yogurt,poultry,aluminum foil]
rule8	0.05	0.588	1.528	yogurt	<---	[cheeses,cereals,coffee/tea]
rule9	0.05	0.594	1.52	cheeses	<---	[yogurt,cereals,coffee/tea]
rule10	0.05	0.582	1.469	cereals	<---	[yogurt,cheeses,coffee/tea]
rule11	0.05	0.613	1.616	coffee/tea	<---	[yogurt,cheeses,cereals]
rule12	0.05	0.588	1.564	mixes	<---	[dishwashing liquid /detergent,poultry,laundry detergent]
rule13	0.05	0.528	1.395	laundry detergent	<---	[dishwashing liquid /detergent,poultry, mixes]
rule14	0.05	0.576	1.484	dishwashing liquid /detergent	<---	[poultry,laundry detergent, mixes]
rule15	0.05	0.6	1.424	poultry	<---	[dishwashing liquid /detergent,laundry detergent, mixes]
rule16	0.051	0.563	1.464	yogurt	<---	[poultry, mixes, aluminum foil]
rule17	0.051	0.569	1.479	aluminum foil	<---	[yogurt,poultry, mixes]
rule18	0.051	0.63	1.678	mixes	<---	[yogurt,poultry, aluminum foil]
rule19	0.051	0.586	1.39	poultry	<---	[yogurt, mixes, aluminum foil]
rule20	0.051	0.611	1.66	sandwich bags	<---	[cheeses, bagels, cereals]
rule21	0.051	0.674	1.726	cheeses	<---	[bagels, cereals, sandwich bags]
rule22	0.051	0.617	1.558	cereals	<---	[cheeses, bagels, sandwich bags]
rule23	0.051	0.58	1.505	bagels	<---	[cheeses, cereals, sandwich bags]
rule24	0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce, poultry, cereals]
rule25	0.051	0.513	1.296	cereals	<---	[dinner rolls, spaghetti sauce, poultry]
rule26	0.051	0.552	1.48	spaghetti sauce	<---	[dinner rolls, poultry, cereals]
rule27	0.051	0.627	1.512	poultry	<---	[dinner rolls, spaghetti sauce, cereals]

Recommended Items

- Lets compare the items, recommended items and lift from the above screenshot for the first row.
- Here it suggests that customer who buys toilet paper, juice, aluminum foil will also tend to buy yougurt.
- The Lift value is the most important column which tells us that there are certainly 1.497% of chances that customer
- Will tend to buy the suggested items 'yogurt' as well along with the other list of items.
- From the above list of items for the same item list there also can be more than one recommended items.

PART B

PART B: Suggestion of Possible Combos with Lucrative Offers --> Write recommendations --> Make discount offers or combos (or buy two get one free) based on the associations and your experience

Based on these list the store can even go for other options to attract customers such as.

- 1) Providing offers.
- 2) Introducing packages of items.
- 3) In some cases the items that are less soled can be offered for best discount price.
- 4) This will certainly help attract customers to be associated with the store for longer period of time.
- 5) Say for instance the highest selling product according to the store data is 'Poultry' and the least being 'handsoap'
- 6) The store can certainly suggest something like if a customer purchases more than 5 items in the grocery store they can Get an 'handsoap' for free of cost.