Market and Retail Analytics - 2

Sanjana M
2OA0CSPH75

Index :

- *Understanding the available data*
- *Market Basket Analysis*
- *KNIME Workflow*
- *Associations identified*
- *Recommendations*

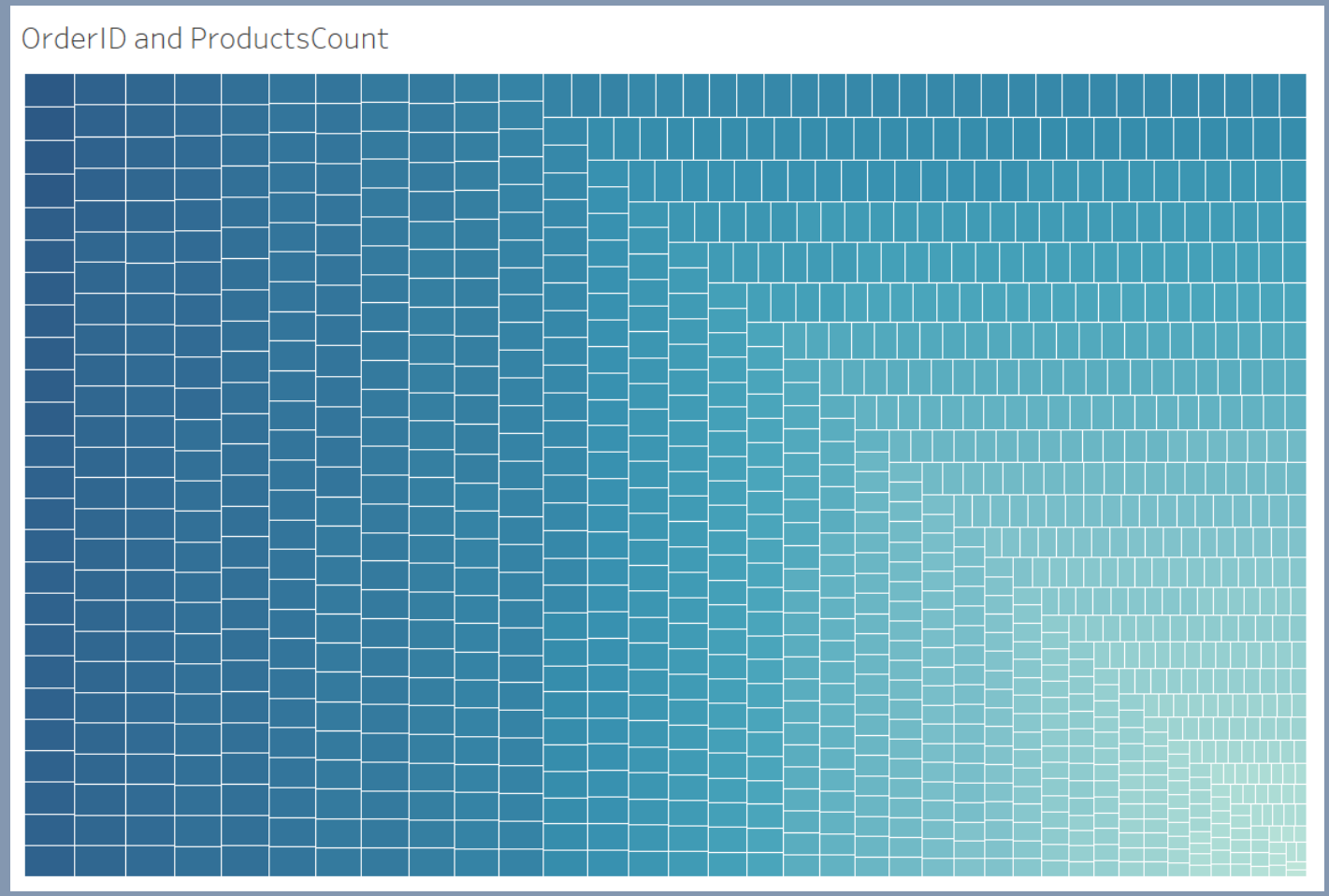
Understanding the available data :

The given data has 3 columns and 20641 rows of entries with no null values.
The parameters are a customer's Date of purchase, OrderID and the Product.
The raw data looks something like this:

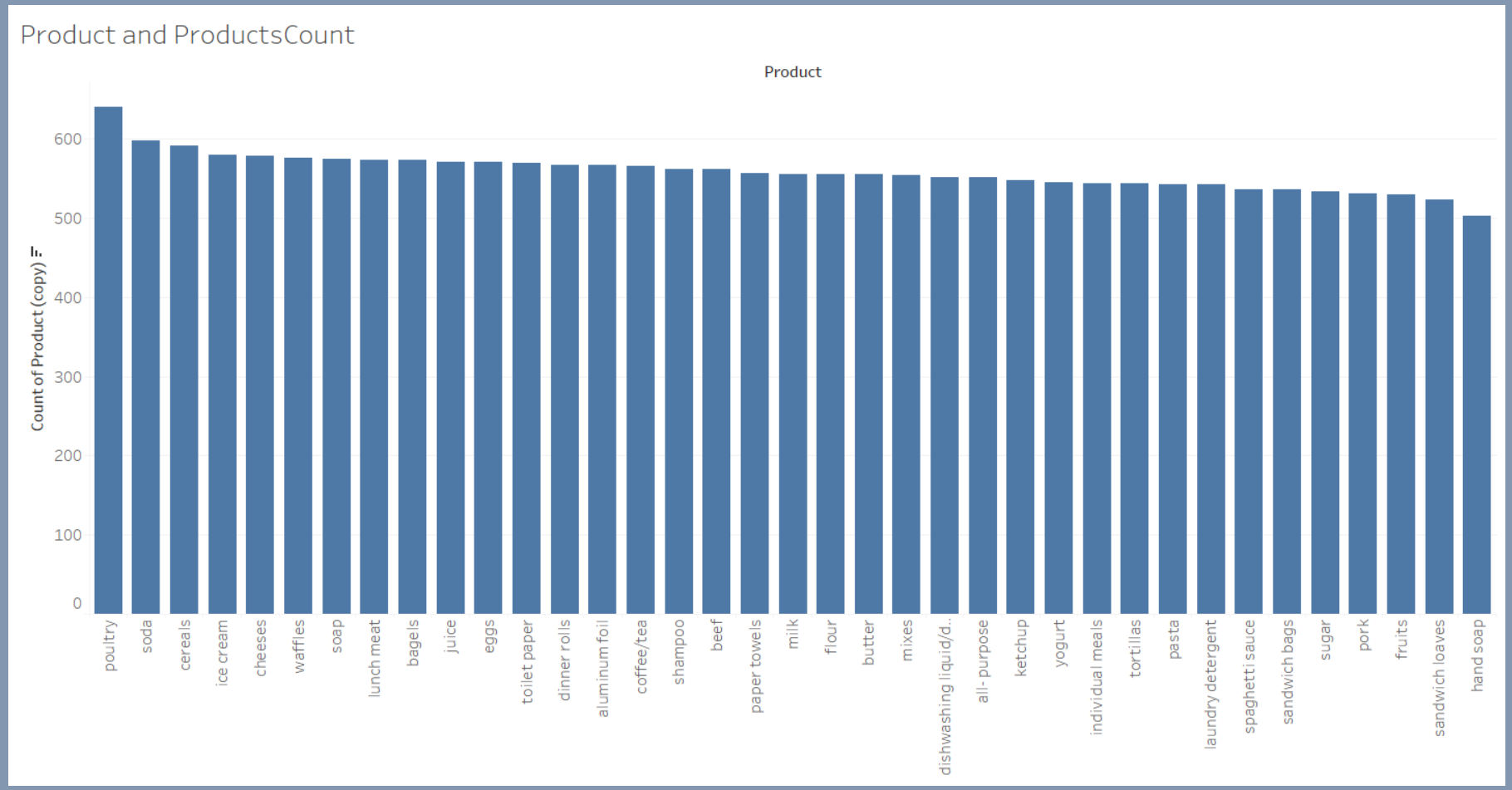
	Date	Order_id	Product
0	2018-01-01	1	yogurt
1	2018-01-01	1	pork
2	2018-01-01	1	sandwich bags
3	2018-01-01	1	lunch meat
4	2018-01-01	1	all- purpose

Data can be grouped by OrderID to get a list of Products purchased are the latest date.
There are a total of 39 Products available to the customer, namely:
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, sugar

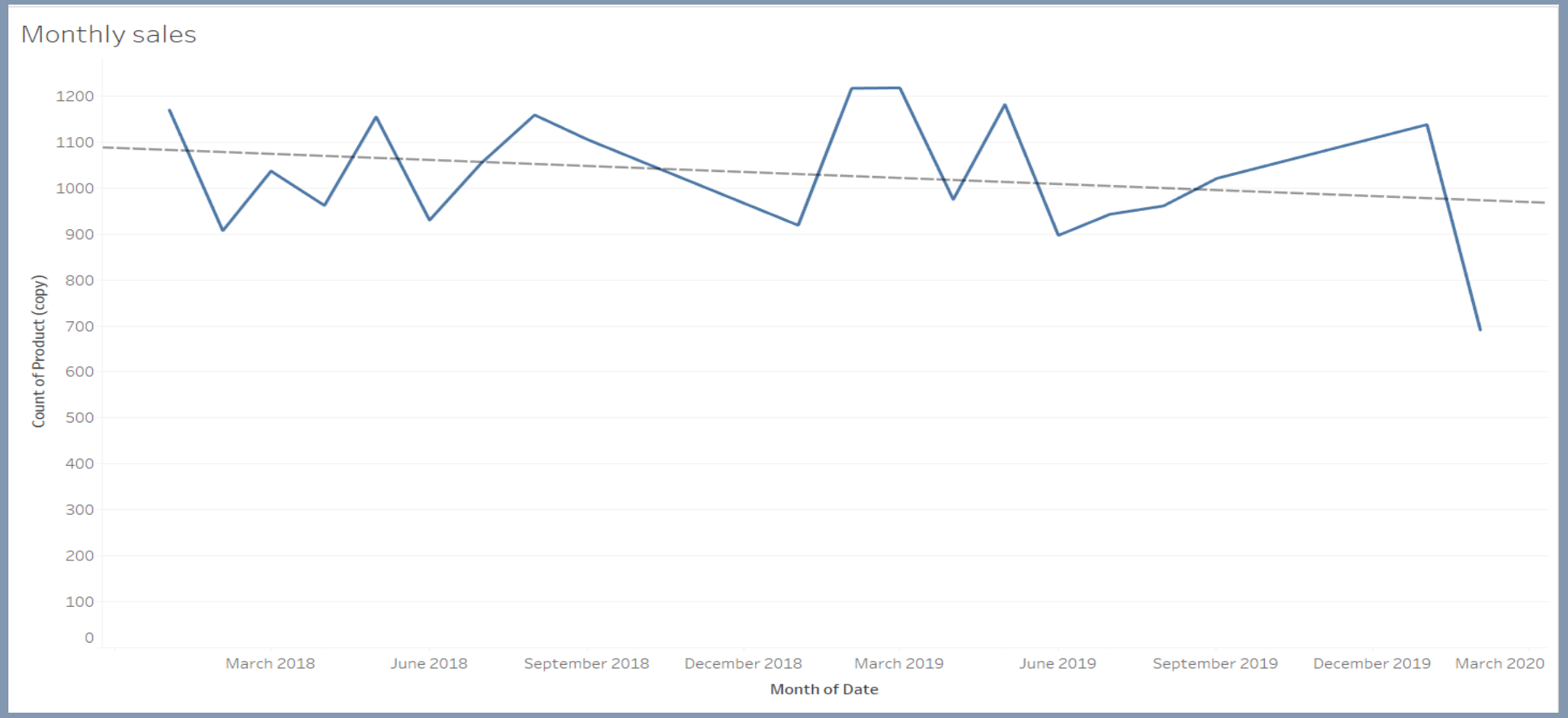
At the higher end, OrderID 226 has total of 34 purchased Products
At the lower end, OrderIDs 408, 1139 have total of 3 purchased Products
The difference is greater than 10 folds of lowest count.
Customers with small basket size can be targeted to improve business.



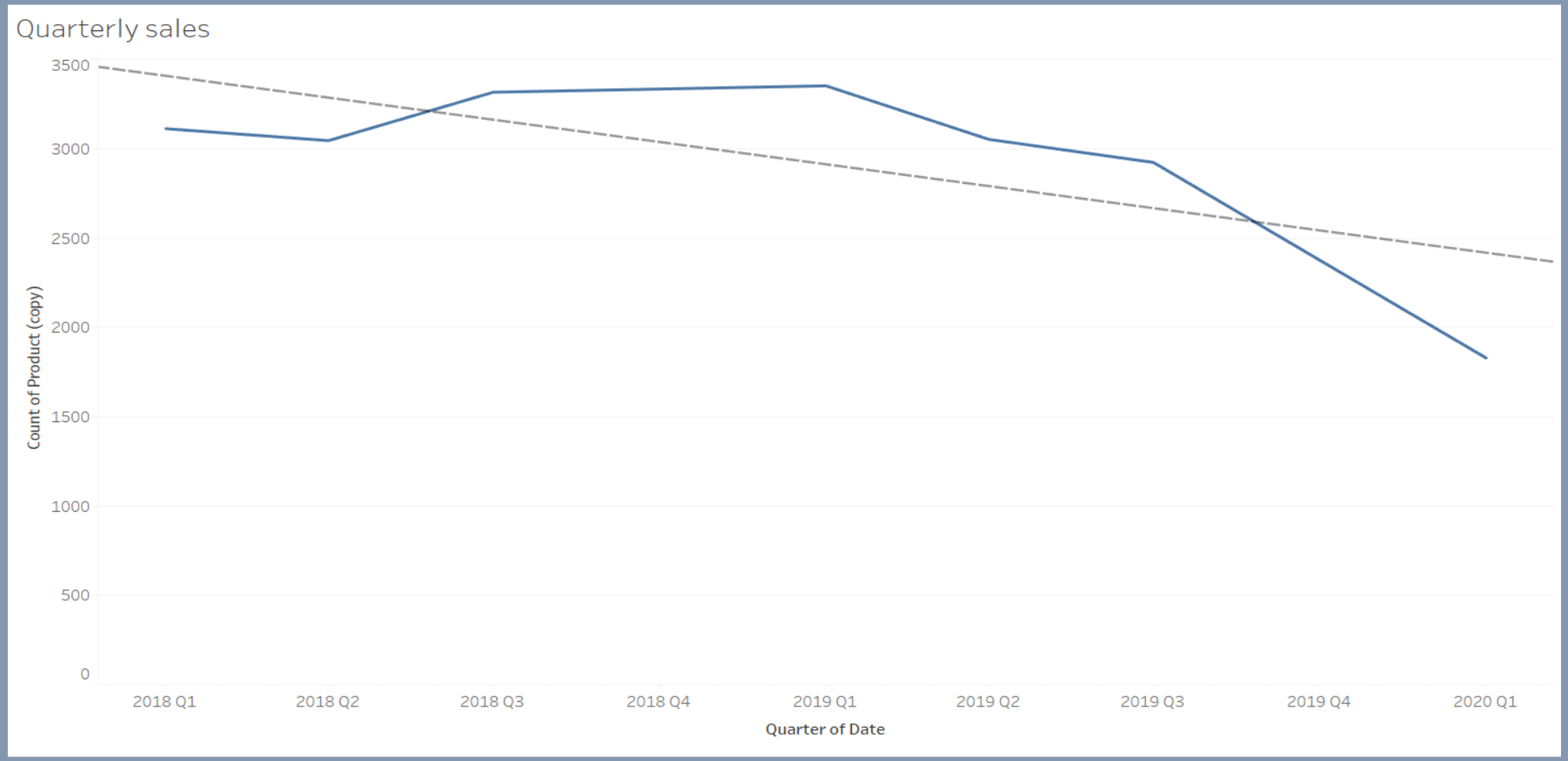
Out of the 39 Product types available at the store,
Poultry is sold highest with a count of 640
Hand Soap is sold the least with a count 502
The drop is not very steep, indicating good choice of items available to the customers.
The second highest selling product is the Soda with count of 597
The drop from highest to second-highest selling product is pretty significant and can be targeted



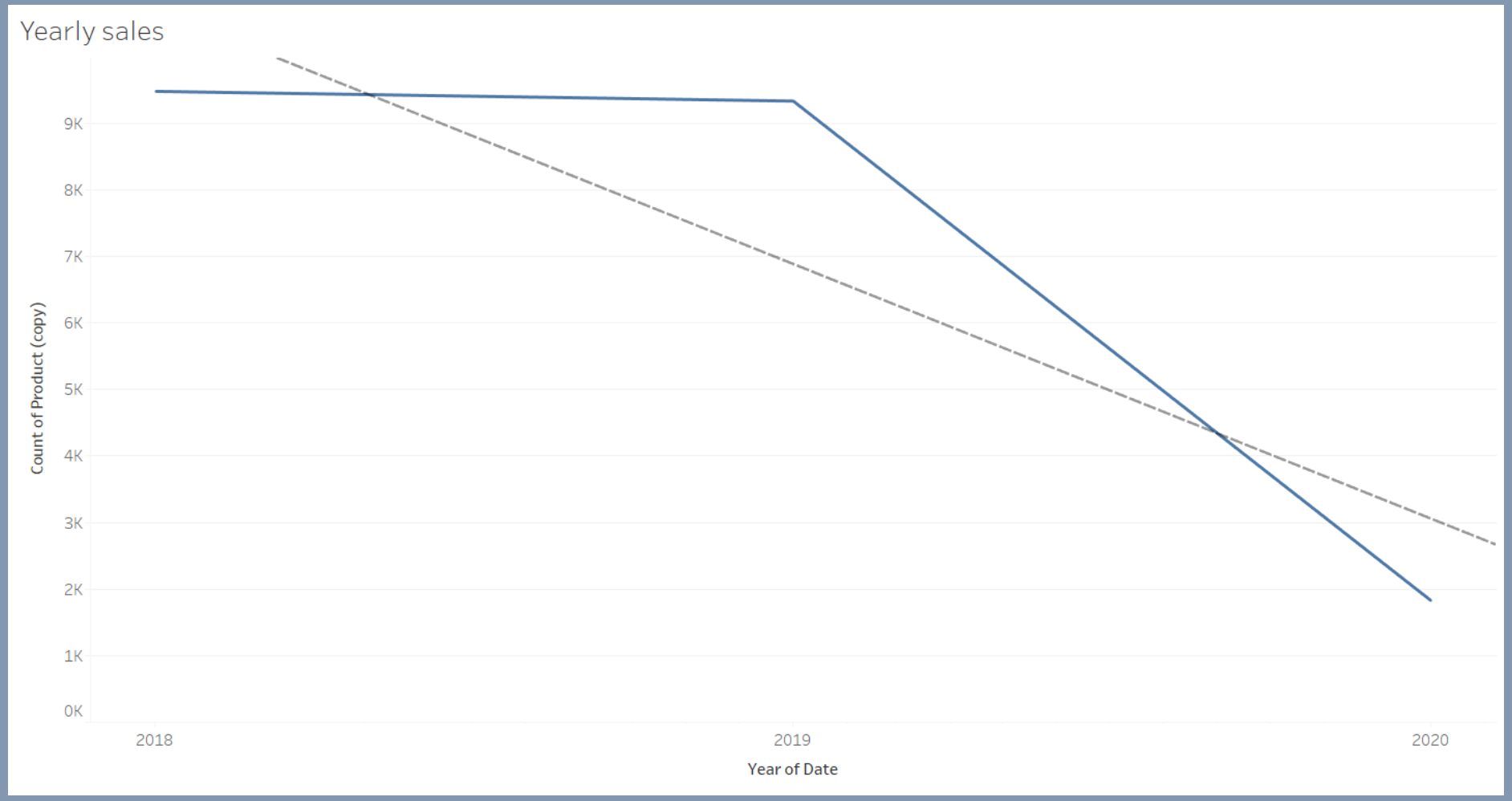
The below graph shows monthly sales
There's an upward trend in sales during the months of March-April, June-Sept
The good sales always drop from May
There's no other significant repetitive pattern to recognise
Over all, monthly sales has a declining trend line



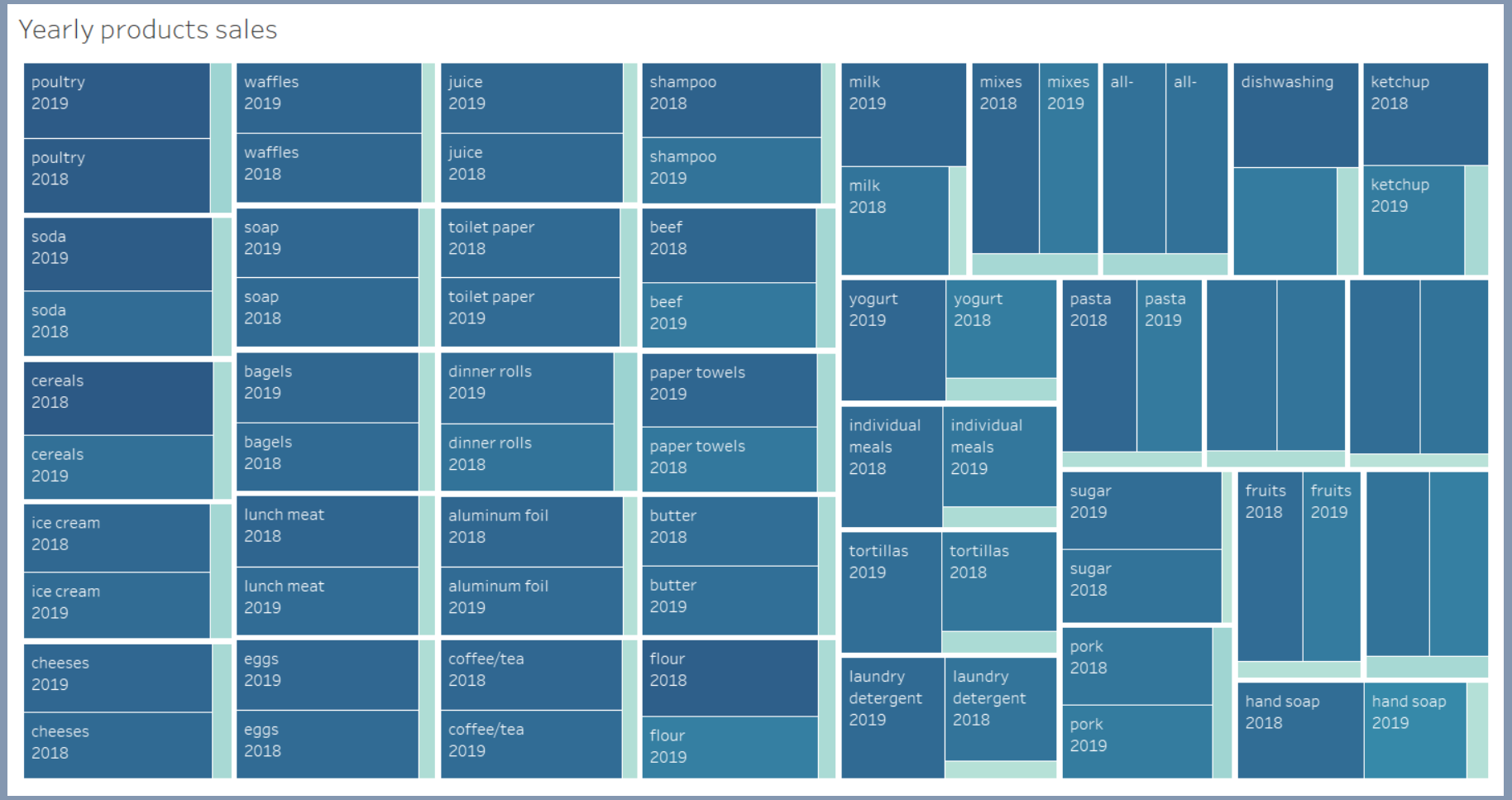
Quarterly sales for the store improved from 2018 Q2 – 2019 Q1 and faced decrement from there
The over all look of quarterly sales in declining
The drop is steeper than monthly sales



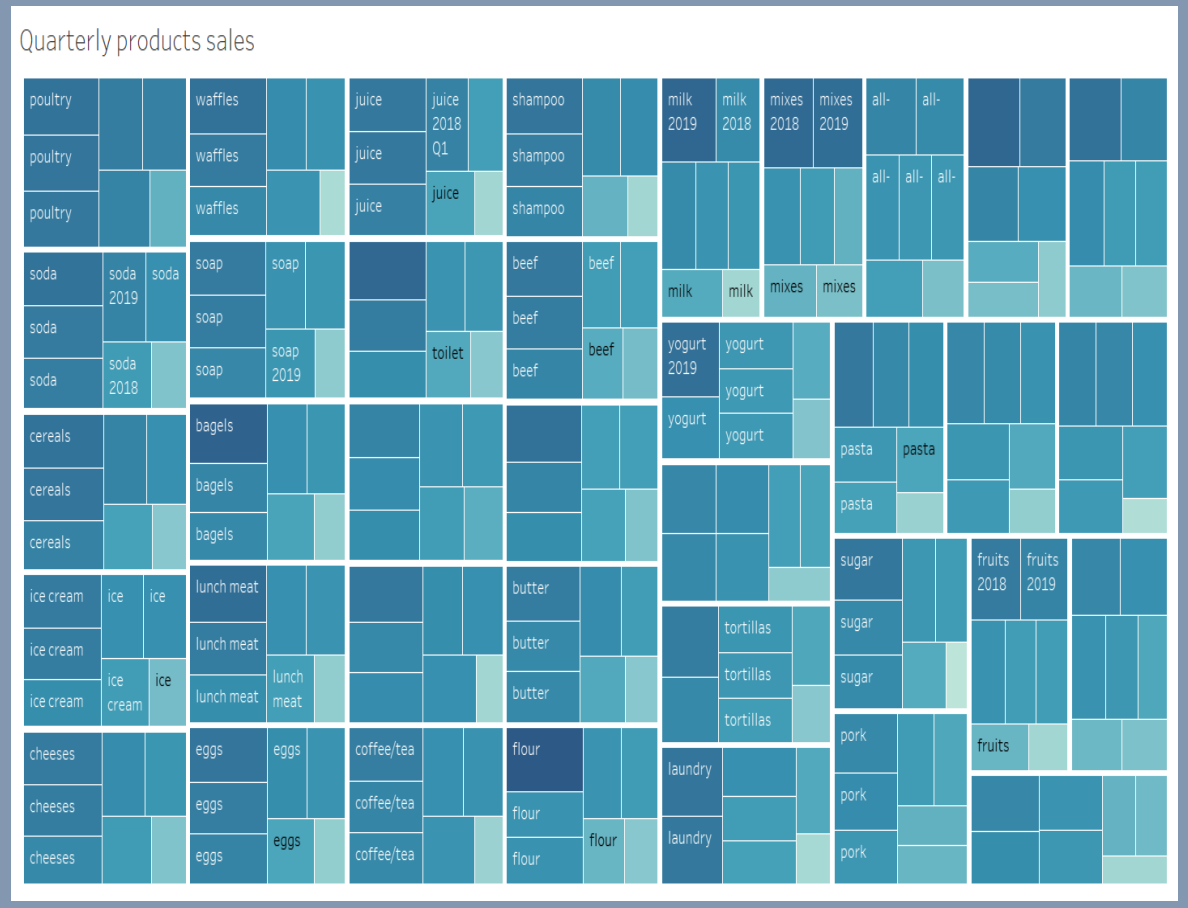
The business has insignificant drop from 2018-2019
But there is a very significant business drop in the year 2019 towards the beginning of 2020.
Overall, the business is not doing well

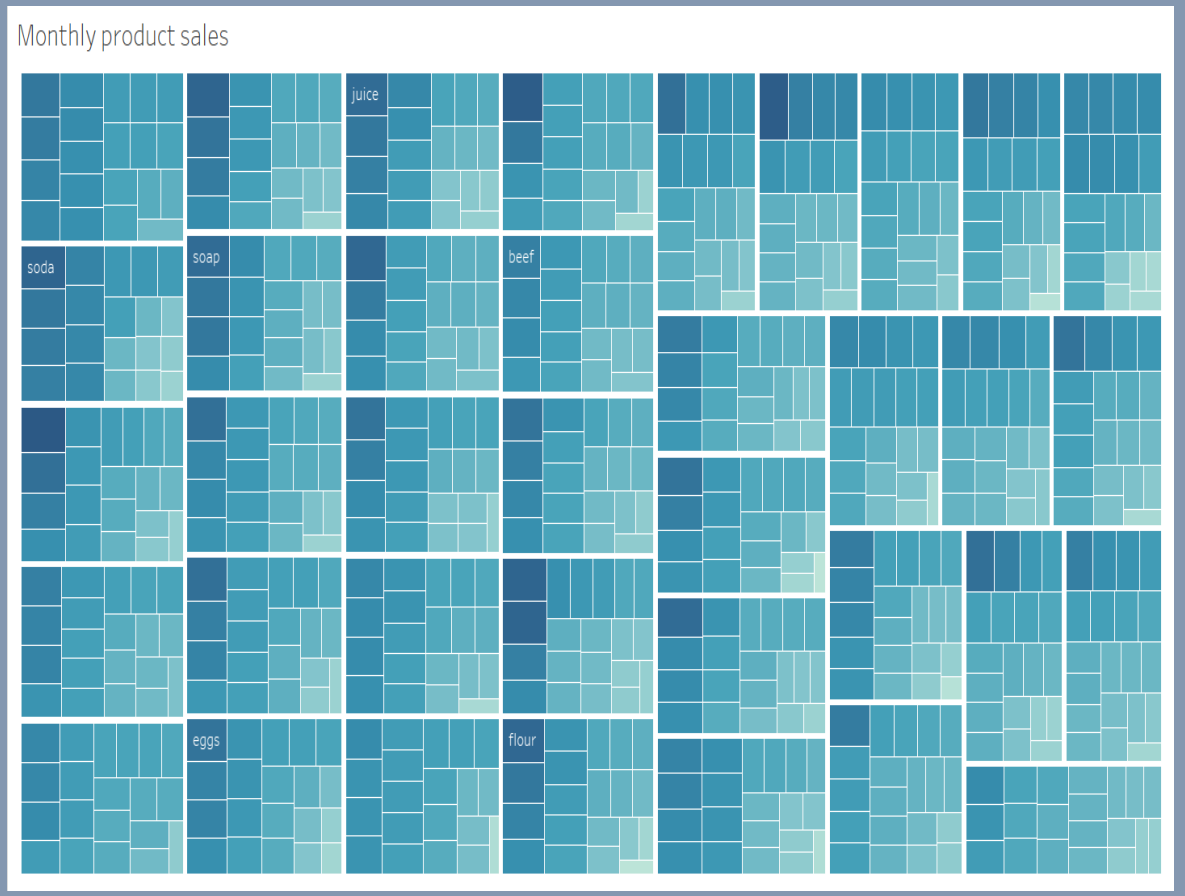


Yearly sales of products is shown below
Poultry has highest sales and hand soap has the least sales through out
Sales of Soda, Cereals, Paper towels, Flour , Beef, Sugar, Fruits, Hand soap and more have declined from 2018 to 2019
Products like Poultry, Ice cream, Cheeses, Coffee/Tea have stable sales
There are no improvement in product sales seen
This is a bad indication of business



Quarterly and Monthly sales of each product is in declining trend.
There are no products with stable or improved sales





Market Basket Analysis :

Market Basket Analysis is a technique used to better understand customer purchasing patterns.
It involves analysing data sets, to reveal product groupings, as well as products that are likely to be purchased together

In market basket analysis, association rules are used to predict the likelihood of products being purchased together. Association rules count the frequency of items that occur together, seeking to find associations that occur far more often than expected.

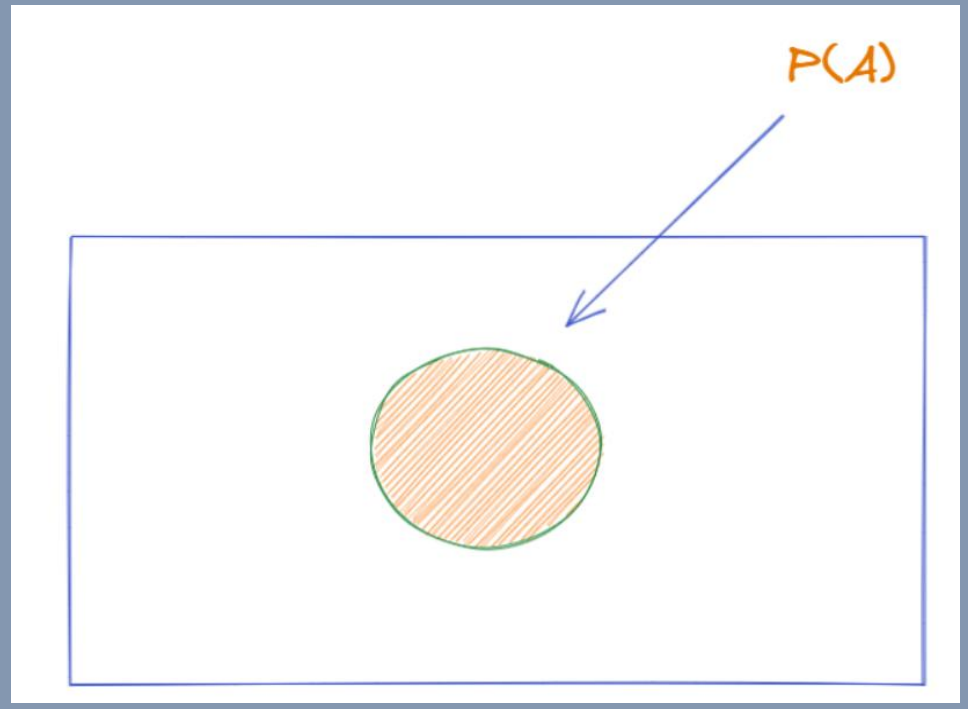
Eg: When a customer purchases All-purpose flour, eggs, butter
there is a high chance of customer also purchasing milk, vanilla essence.

Similarly, in this case, we recommend a product (with/without offers) to the customer based on the products that they have already chosen in the basket.

This analysis is both beneficial to the customer and the business.

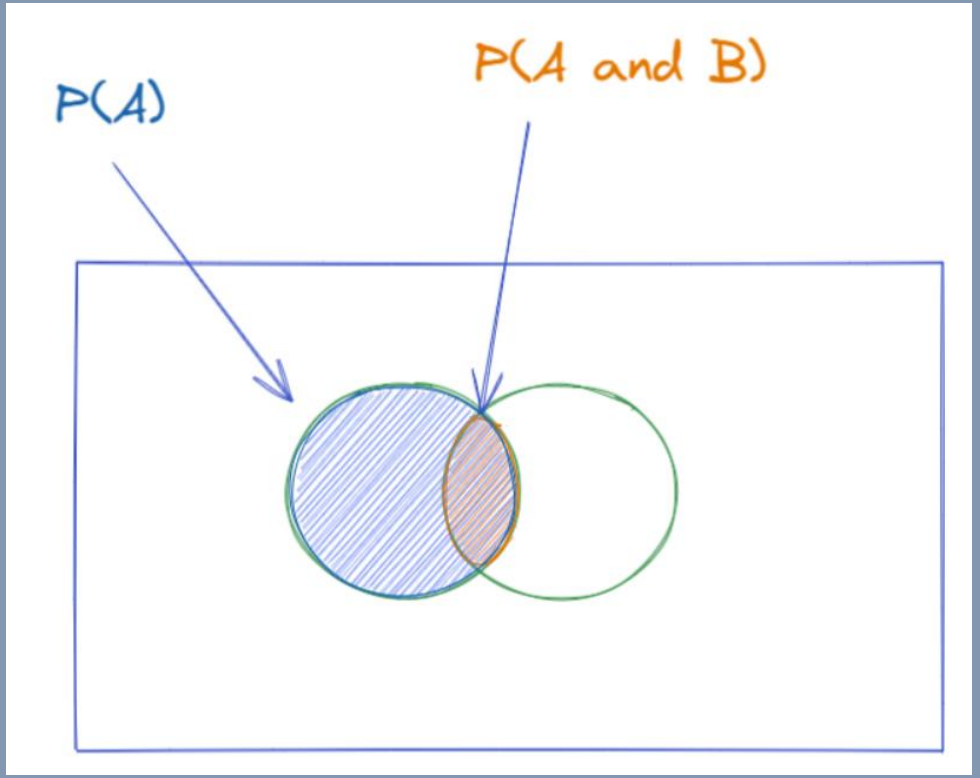
Support:

Within a dataset, i.e. a list of transactions, how many transactions contain **item A**, so it is just the probability of **item A** occurring. Statistically speaking, it is a frequentist's estimate of the probability.



Confidence:

Out of the transactions that contains **item A**, how many also contains **item B**.
The bigger the overlap, the greater the confidence we have that people who are buying item A also buys itemB.
Statistically speaking, it is (estimated) conditional probably of **item B** given **item A**, i.e. **P(B|A)**.



Lift:

The ratio between **Confidence of A** and **Support B**, it is less intuitive with the description, so let's try to visualize it better. First let's see the formula below.

$$\frac{P(A \text{ and } B)}{P(A) \times P(B)}$$

Basket	Product 1	Product 2	Product 3
1	Cereal	Honey	
2	Cereal	Water	Honey
3	Water	Diaper	
4	Cereal	Honey	
5	Water	Diaper	
6	Cereal	Honey	Diaper
7	Cereal	Honey	
8	Honey	Diaper	
9	Honey	Cereal	

Example:

setA : Water+Cereal
setB: Honey

in this data set,
setA occurs 1 time
setB occurs 7 times
setA+setB occurs 1 time
Total baskets are 9

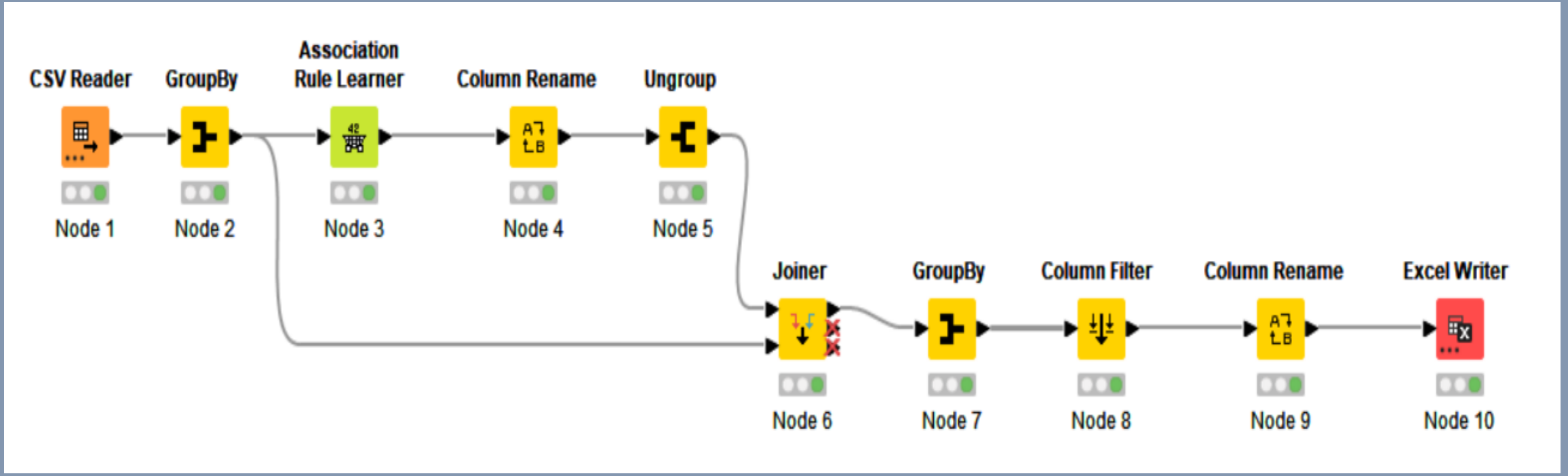
Support: setA/Total baskets = 1/9 = 0.1111
Higher the Support, more popular is the set

Confidence: setA+setB / setA = 1/1 = 1

Support and Confidence used to set the thresholds

Lift: Confidence / (setB/Total) = 1/(7/9) = 1.285
If a customer buys setA, buying setB increases by 1.285%

KNIME Workflow:



Associations identified :

Recommended_Product	Support	Confidence	Lift	Purchased_Products	Order_id
all- purpose	0.0307	0.4217	1.1248	butter, poultry, aluminum foil	243
all- purpose	0.0342	0.5065	1.3510	shampoo, laundry detergent, soda	1139
all- purpose	0.0307	0.4430	1.1818	toilet paper, coffee/tea, aluminum foil	523
aluminum foil	0.0334	0.4935	1.2833	shampoo, laundry detergent, soda	1139
bagels	0.0316	0.4675	1.2130	shampoo, laundry detergent, soda	1139
bagels	0.0351	0.5063	1.3137	toilet paper, coffee/tea, aluminum foil	523
beef	0.0307	0.4217	1.1248	butter, poultry, aluminum foil	243
butter	0.0325	0.4805	1.3062	shampoo, laundry detergent, soda	1139
butter	0.0334	0.4810	1.3076	toilet paper, coffee/tea, aluminum foil	523
cereals	0.0342	0.4699	1.1867	butter, poultry, aluminum foil	243

The probability of a customer buying butter, poultry & aluminium foil is 0.0307%

The probability of a customer buying all-purpose given that the customer has already preferred butter, poultry & aluminium foil is 0.4217%

If a customer has purchased butter, poultry & aluminium foil, we recommend the customer to buy all-purpose too. Probability of this rule being recoganised in 1.1248%

Recommendations :

Customers who have purchased butter, poultry, aluminum foil together, we can recommend all-purpose, cheeses on an offer as a meal-combo.

On purchase of toilet paper, coffee/tea, aluminum foil, we can recommend butter and bagels as a brekfast-combo

Buy-one-get-one-free offer on cereals will increase its sales, which will in-turn increase sales of milk

Year long sale on shampoo, laundry detergent, soda and paper towel recommendation can be given. Since these products have high shelf-life, sales of these products are of low frequency. This will help boost the product movement

There are 3 main sets found:
butter, poultry, aluminum foil
shampoo, laundry detergent, soda
toilet paper, coffee/tea, aluminum foil
On purchase of all these products, a bumper offer on fruits and sandwich loaves can be given
This will push the least-sold short-shelf-life products and also attract customers