GROCERY STORE POINT OF SALE

MRA PROJECT 2

BY GIRIDHARAN VELMURUGAN

GREAT LEARNING

PROBLEM STATEMENT

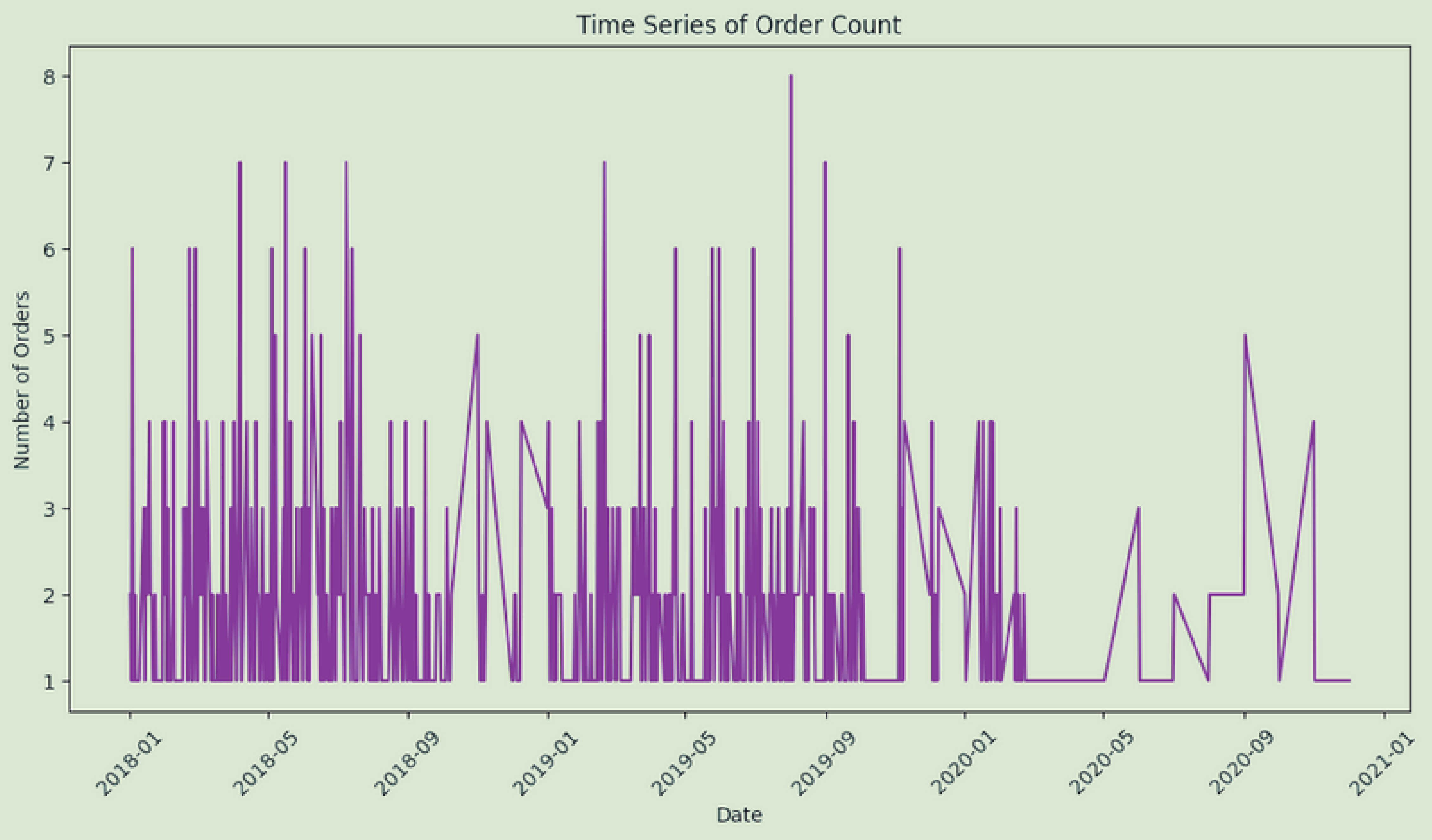
A grocery store shared the transactional data with you. Your job is to conduct a thorough analysis of Point of Sale (POS) data, identify the most commonly occurring sets of items in the customer orders, and provide recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.

DATA OVERVIEW

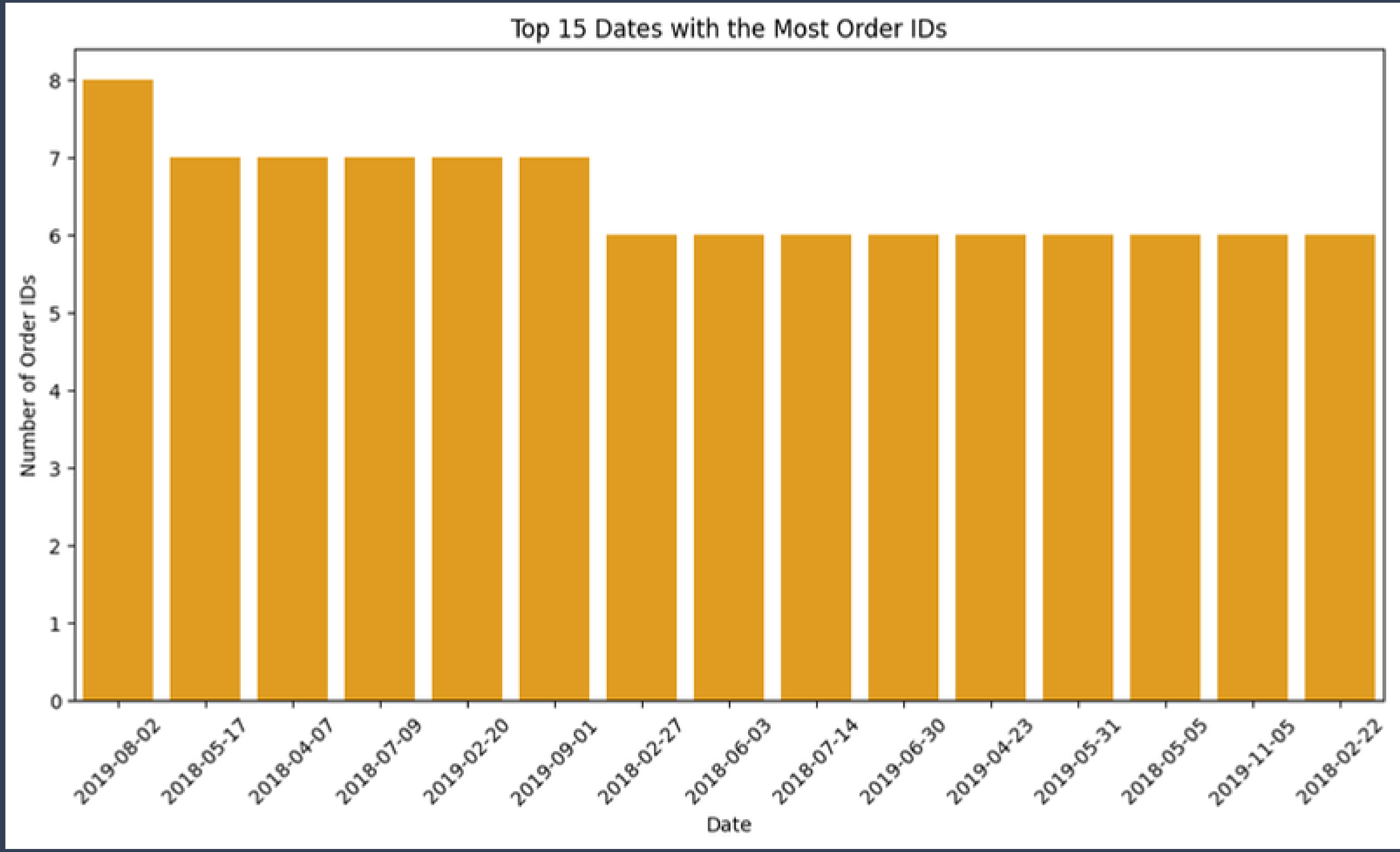
- Dataset has only int(64) and objet datatypes.
- We do have some duplicates (around 20% of the data) in its raw form.
- There are no null values within the features, all the columns have data and there are a total of 20641 instances per feature
- There are a total of 3 features(columns)
- There are multiple products from from meat to veggies and some more, all kitchen and home related stuff.
- The rate of each product occurring is very close, the vairance is at most + or - 75 counts.
- we have only 603 unique date values, meaning multiple orders on a single day.

EXPLORATORY DATA ANALYSIS

- The image on the right is a time-series showing sales that has spanned over the years.
- We can see variations throughout, what is clear that order quantity is going downhill.
- From January start of the year till end of August we can see a huge number of orders then it dips.
- The year 2020 and post is where we see a critical dwindling in quantity.
- We have multiple peaks, the best was 8 orders on August 2019.
- There are quite a number of sales days with only 1 order.



Dates with the most orders.

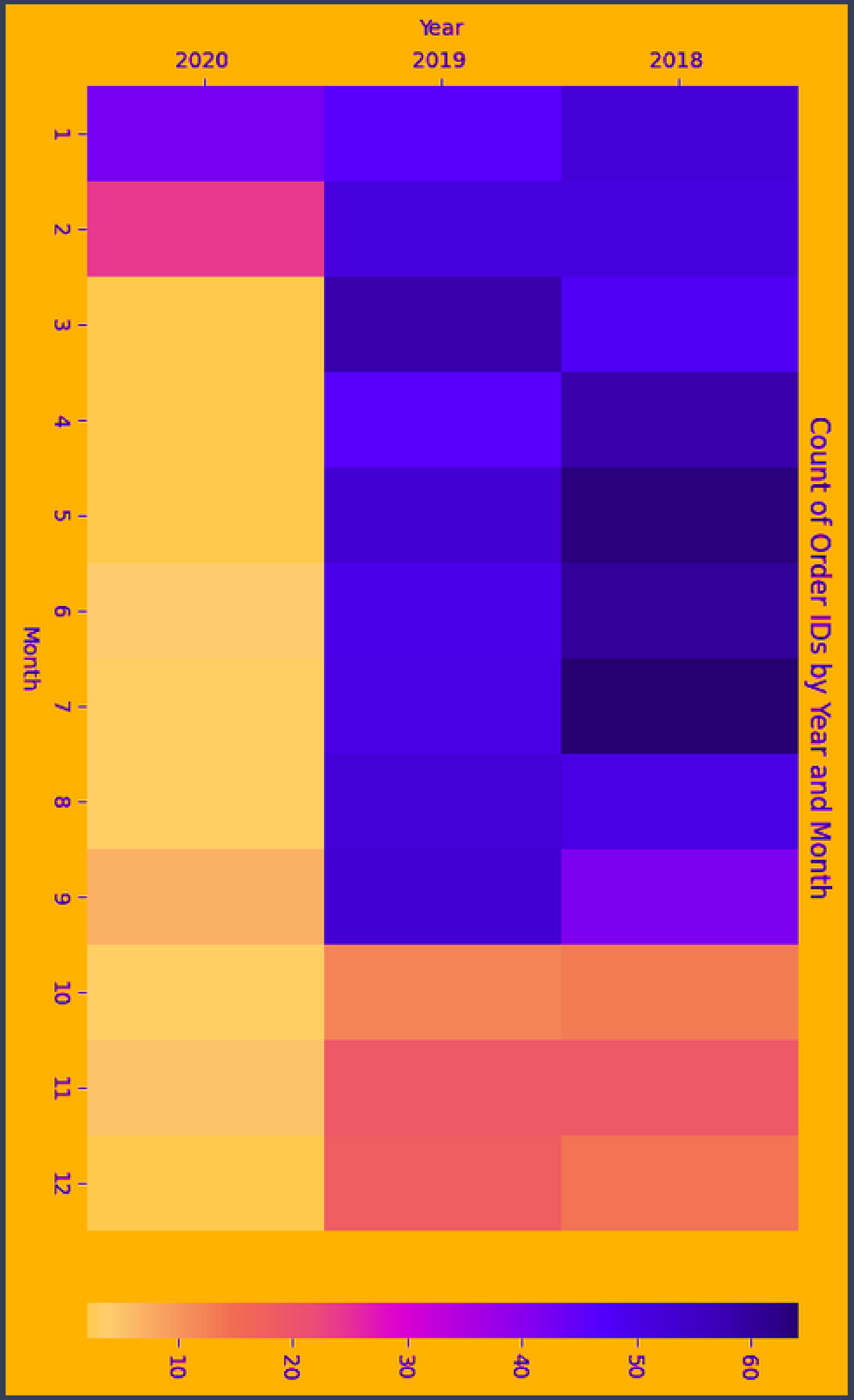


- The Date with the highest count of orders was on 2nd of August 2019. Raking up to 8 orders.
- Next in line we have 4 different dates with 7 orders.
- Next is a series of dates with 6 orders.
- We do not have any date with zero orders, so at the minimum we are making atleast 1 order.
- The trend line is reflecting here as well, most of the dates provided here are either on or before August, the exception being 5th November 2019.

MONTHLY ORDERS

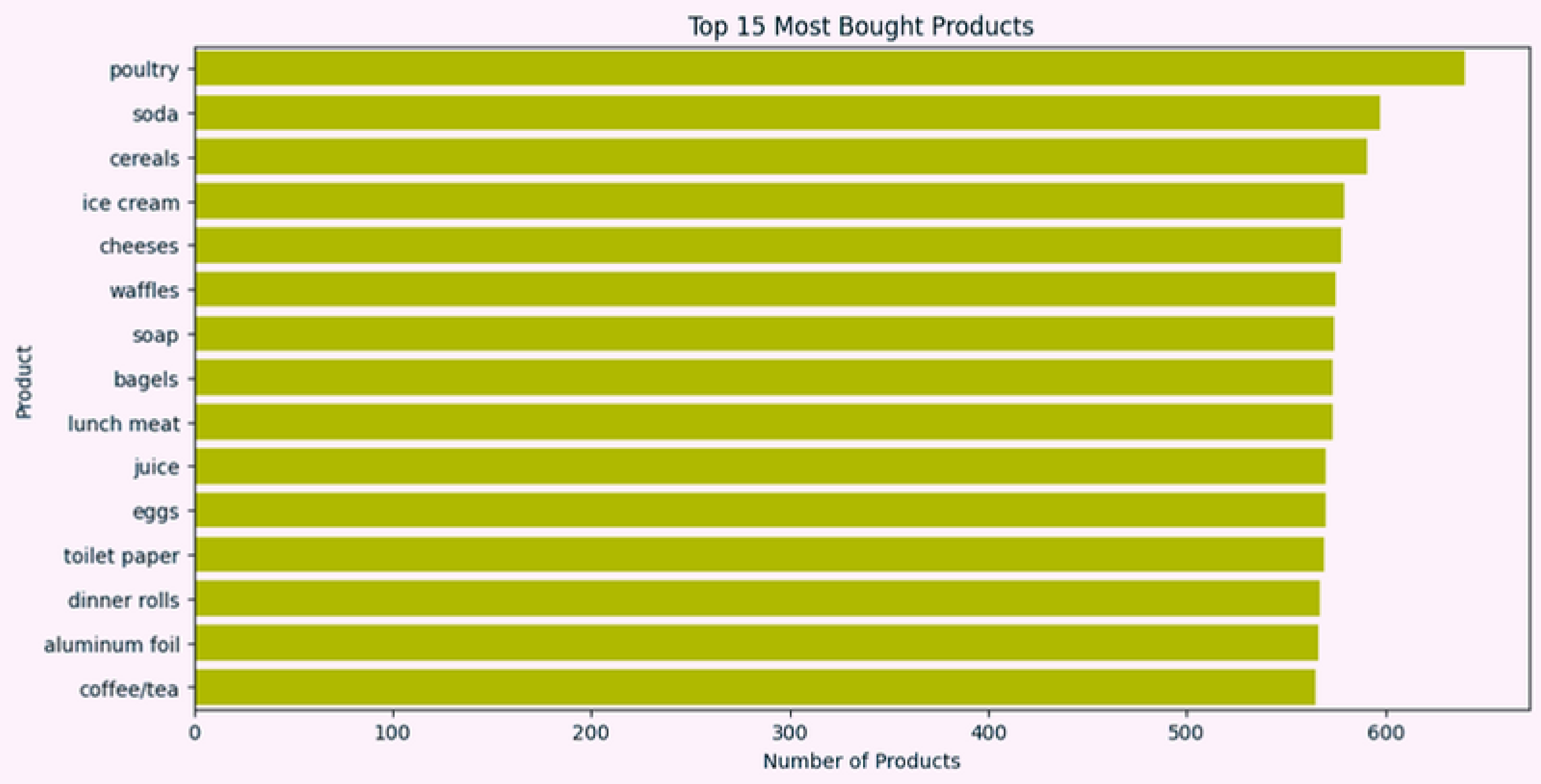
- The image on the right shows the concentration of orders for each month for years.
- the year 2018 has the most densely packed dates with some months reaching more than 60 + orders.
- The latter months of any year from October to December we can see huge dip in order count. This has to be studied further.

- There are no months that have all their dates ticked with orders. This is the reason why we see such dip post 2020 February. This can be attributed to data loss or simply no orders were placed. If it is the latter we need to figure out the root cause
- 2018 > 2019 > 2020 is the hierarchy.
- The reduction in orders is drastic that it mimics non-availability of data.



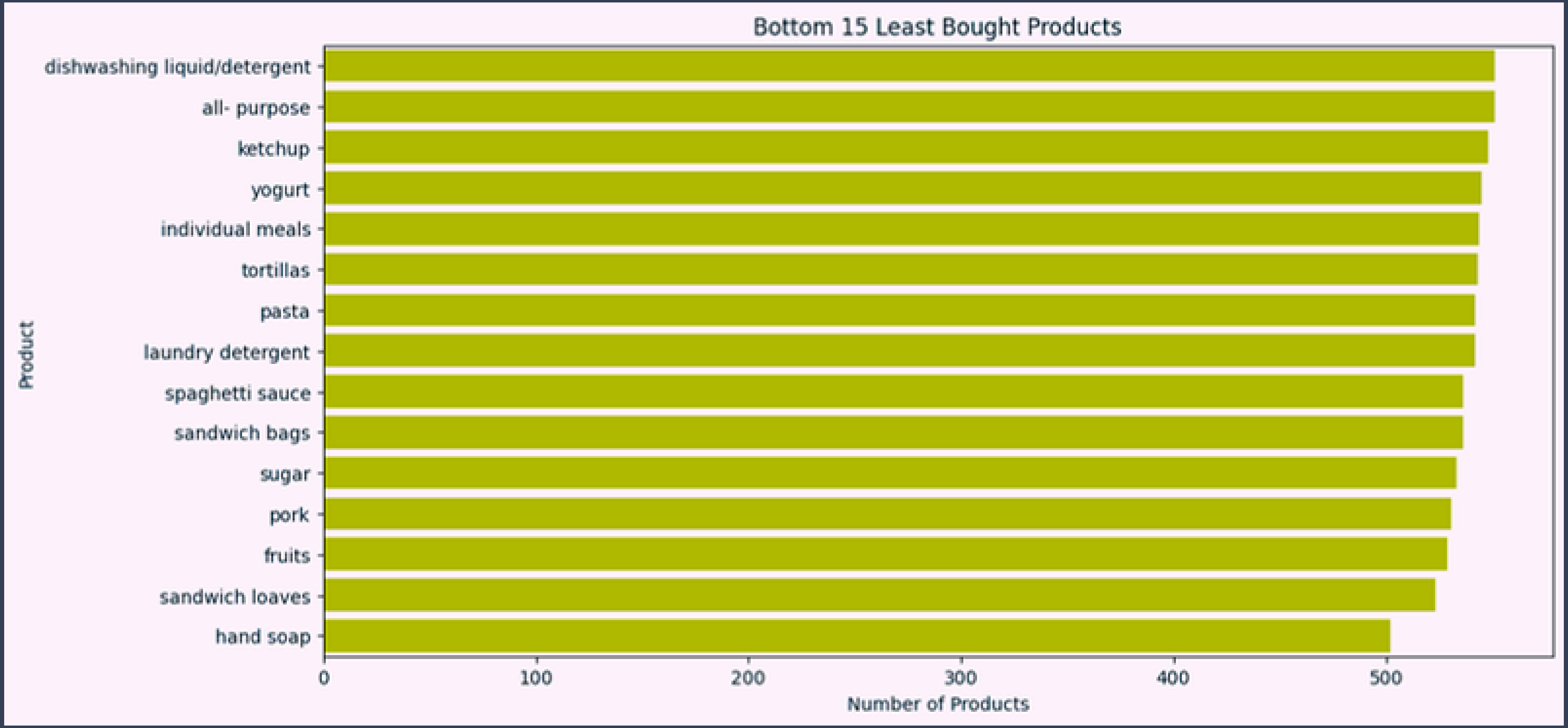
TOP PRODUCTS BOUGHT

- The top product in our list is “poultry” crossing 600 units, it is also the only product to so.
- Our runner up soda, and our second runner up is cereals.
- top 6 is dominated by food items. on the whole 13 of the 15 items seen here are edible or food products. the rest three are housekeeping section.



LEAST BOUGHT PRODUCTS

- The least bought item in our list is “Hand Soap” at 500 units, it is interesting that we have just “soap” that is occupying a spot in the top 10.
- here the ration is slight improved in the household items type, the ration is 10 food items of the bottom 15.
- Food items have dominated the total product space
- There are a total fo 37 unique items in the product space over the years considered.



WHAT IS MARKET BASKET ANALYSIS

SUPPORT

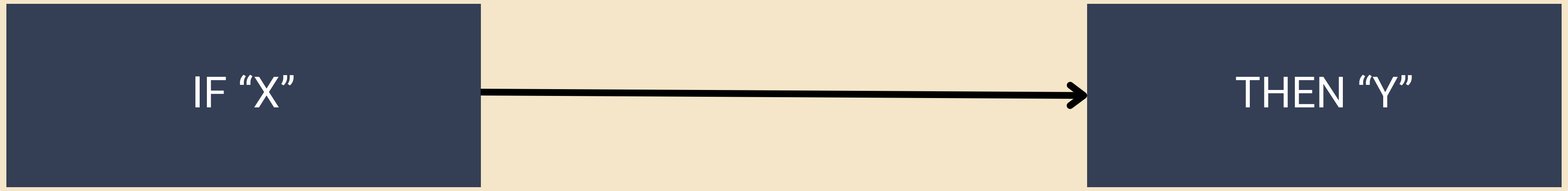
LIFT

CONFIDENCE

Market basket analysis is a technique used by retailers and other data handlers to increase sales by better understanding customer purchasing behavior. The above three metrics determine how we go about this whole process.

ASSOCIATION RULES (USE OF THEM)

In the context of Market Basket Analysis, “Association rules” are used to deduce which items are likely to co-star in a shopping basket based on historical shopping data. We can think of this as guidelines that further helps us to increase sales. These rules are framed based on historical customer purchasing behavior.



The above is an elementary example of an Association rule, substitute “X” with nachos and “Y” with salsa dip. The product list can have anywhere between 20 to 30 items at some scenarios or even more.

RELEVANCE

Association rules are instrumental in understanding customer behavior, driving business decisions, and ultimately increasing sales and profits. Below are some examples.

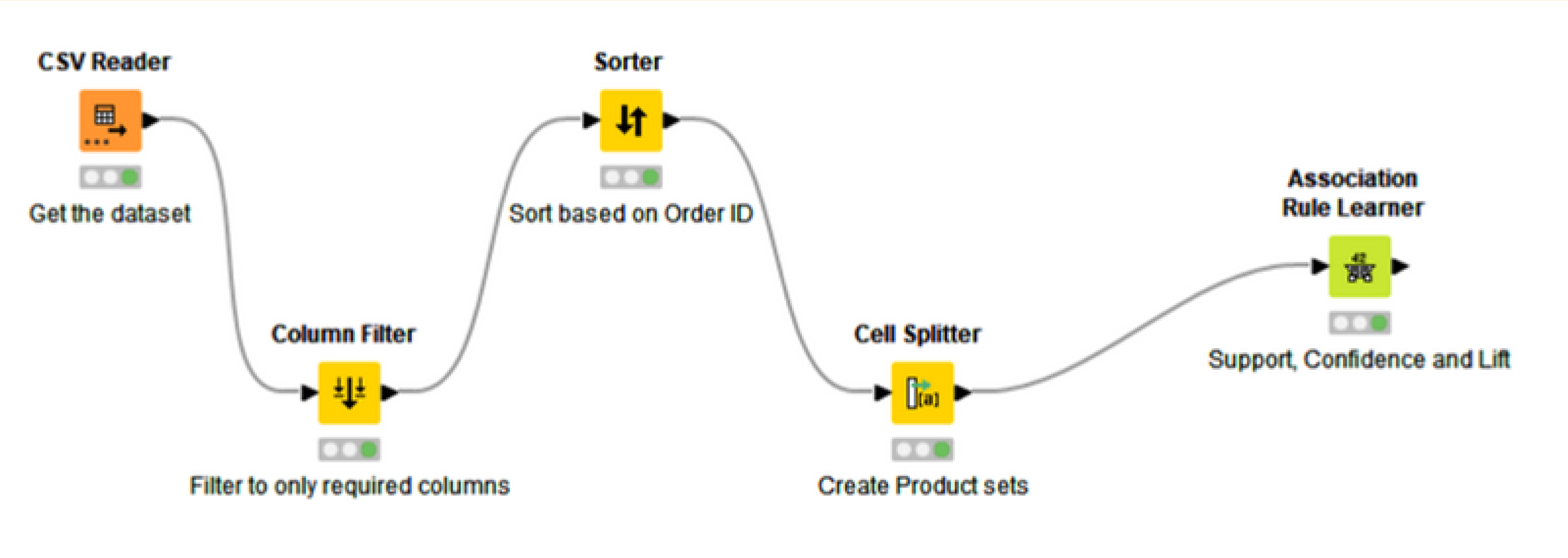
1. ***<u>Product placement:</u>*** Position complementary products (e.g nachos and salsa dip) closer together to encourage impulse purchases.
2. ***<u>Targeted promotions:</u>*** Offer bundled discounts on frequently purchased items, like nachos and dip deals from the same company or sister brands.
3. ***<u>Inventory management:</u>*** Optimize stock levels based on predicted demand derived from association rules.
4. ***<u>Personalized recommendations:</u>*** Suggest relevant products to customers based on their past purchases and similar buying patterns.

SUPPORT AND CONFIDENCE

SUPPORT: --
A popularity contest, Support measures how frequently an item or itemset appears at POS. it is gauging how common it is for people to buy both bread and butter together compared to another itemset.
Determining support value is not something that is pre-determined, but rather an extensive trial and error process. it changes with data and also the context of what is being reviewed.
Too low of support value and start to take in more noise, too high and we might miss out on meaningful relationships that can just be what we need.

CONFIDENCE: --
Confidence is about trustworthiness. here we measure the chances of buying Item A leading to buying Item B. It's the "Given you've bought nachos, how likely are you to buy salsa dip?" metric. A high confidence value means our association rule is backed by high historical similarity, but we have to be cautious that we do not loose other options. As with support the confidence will change based on the data, and our question. Trail and error is the way to go.

MARKET BASKET ANALYSIS USING KNIME



- KNIME is a user friendly, powerful, flexible Data science tool, with aid from KNIME we have performed market basket analysis.
- With KNIME and its Nodes, we have brought in our dataset which holds the required values. The dataframe is primed using a python notebook.
- We need to prepare the products we have into a set, only then can we use the node association rule learner to determine our important metrics. minimum Support was set at “0.02” and minimum confidence was set at “0.4”
- The final result is the output of association rule learner node.
- The following sllide shows a another important pillar of market basket analysis and then we see s snapshot of the table.

“LIFT”

Lift, is the third important measure that goes beyond support and confidence to assess the strength of association between items in a rule. It tells the likely hood for customers to buy both items together compared chance.

Lift = Confidence(A, B) / Support(B)

If lift is greater than “1” it indicates a more positive relationship, meaning chances are more that they will buy product “B” if bought product “A”. If lift is less than “1” then it indicates a negative(red flag) and customers are less likely to buy “B” if bought “A”. The thing to note is if LIFT is exactly “1”, then they are more like coincidences, the net influence is null.

RESULT'S TABLE - SNAP SHOT

Row ID	D Support	D Confide...	D Lift	S Conseq...	S 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]
rule21	0.02	0.411	1.042	waffles	<---	[butter,fruits,pork]
rule22	0.02	0.404	1.097	butter	<---	[fruits,milk,pork]

NOTE - The above results table shows only a tiny portion of rules generated. “Consequences” column is the “y” part of the association rule.

INFERENCES FROM RULE'S

Support	Confidence	Lift	Consequence	Implies	Item set
0.02	0.43	1.11	yogurt	<---	[pasta, pork, soap]
0.02	0.40	1.13	pork	<---	[yogurt, dishwashing liquid/detergent, hand soap]
0.02	0.40	1.08	all- purpose	<---	[yogurt, ice cream, hand soap]
0.02	0.43	1.11	yogurt	<---	[butter, hand soap, beef]
0.02	0.42	1.12	beef	<---	[yogurt, butter, hand soap]
0.02	0.41	0.97	poultry	<---	[butter, fruits, pork]

Here we can see that for the itemset [pasta, pork, soap], the association rule implies that people would also go for "yogurt'. The support and confidence all are close to what we have set. The lift value is at "1.11" meaning the chances of this becoming action are quite good. We also have a rule with lift as "0.97", meaning if customers go for [butters, fruits, pork] it would be good to not to suggest them poultry.

INFERENCES FROM RULE'S

Support	Confidence	Lift	Consequence	Implies	Item set
0.02	0.41	1.07	aluminum foil	<---	[butter, fruits, pork]
0.02	0.41	1.11	shampoo	<---	[butter, fruits, pork]
0.02	0.41	1.04	waffles	<---	[butter, fruits, pork]
0.02	0.41	1.08	milk	<---	[butter, fruits, pork]
0.02	0.41	0.97	poultry	<---	[butter, fruits, pork]
0.02	0.41	1.14	sugar	<---	[butter, fruits, pork]

We can have multiple consequences for the same itemset like we see above. Customers who have [butter, fruits, pork] are likely to buy “aluminum foil”, “shampoo”, “waffles”, “milk” and “sugar”. But as POS what can be suggested is determined by lift, leaving “sugar” as the best choice. It is also important to not suggest “poultry” given the same situation.

INFERENCES FROM RULE'S

Support	Confidence	Lift	Consequence	Implies	Item set
0.19	0.50	1.19	poultry	<---	[dinner rolls]
0.19	0.46	1.19	dinner rolls	<---	[poultry]
0.19	0.44	1.14	dishwashing liquid/detergent	<---	[poultry]
0.19	0.48	1.23	eggs	<---	[soda]
0.19	0.48	1.23	soda	<---	[eggs]
0.19	0.47	1.12	poultry	<---	[lunch meat]

Association rules can also be generated for an item too, the above tables shows the same. we can have counter productive items like eggs to soda. we can also have two "Y" for the same "X", as seen in the third row above. LIFT is the final decider provided all the other features remain in the same ball park.

INFERENCES FROM RULE’S

Support	Confidence	Lift	Consequence	Implies	Item set
0.02	0.85	2.35	paper towels	<---	[eggs, dinner rolls, ice cream, pasta, lunch meat]
0.02	0.88	2.22	ice cream	<---	[paper towels, eggs, dinner rolls, pasta, lunch meat]
0.19	0.46	1.19	dinner rolls	<---	[poultry]

The cells in the above table which are highlighted with a shade of blue are the respective maximums.

A lift of “2.35” for the product set [eggs, dinner rolls, ice cream, pasta, lunch meat] implies people are, to a high level buy paper towel.

Next we have a rule with the maximum confidence value at “0.88”. this means that this combo has been very proficient in those customers in the dataset.

Finally a support of “0.19” means that rule is at the top of the popularity contest.

DEALS THAT CAN BE DEVELOPED FROM RULES

Customers have always bought “Poultry” if they have bought “Dinner rolls”, so we can leave this as it is.

Surprisingly, “Poultry” also seems to be one of the consequences for a lot of rules having life less than “1”, we can come up with strategies that are closely bought together with poultry, like dinner rolls.

We can introduce bundle offers to different consequences of the same itemset, like we saw with [butter, fruits pork]. This can simultaneously improve per order size and impression.

We can attempt A/B testing by providing extra monetary benefits with consequences that have a lift value of “0” to see if we can nudge them to the positive or negative side. Being on the fence is not useful.

For items that have a slight lift, two similar rules can be widened by having buy 1 and get “n” offers where n can be 2, 3, etc.

We can promote differing items from similar item sets that lead to same consequences. This an also be an extension of bundles offers.

Consider seasonal offers and rotate near counter promotions, ice-cream will likely sell more than hot chocolate in the winter.

GREAT LEARNING

THANK YOU

BY GIRIDHARAN VELMURUGAN