Business Report

Index- **Page no.**

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?

- Head of the data looks 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

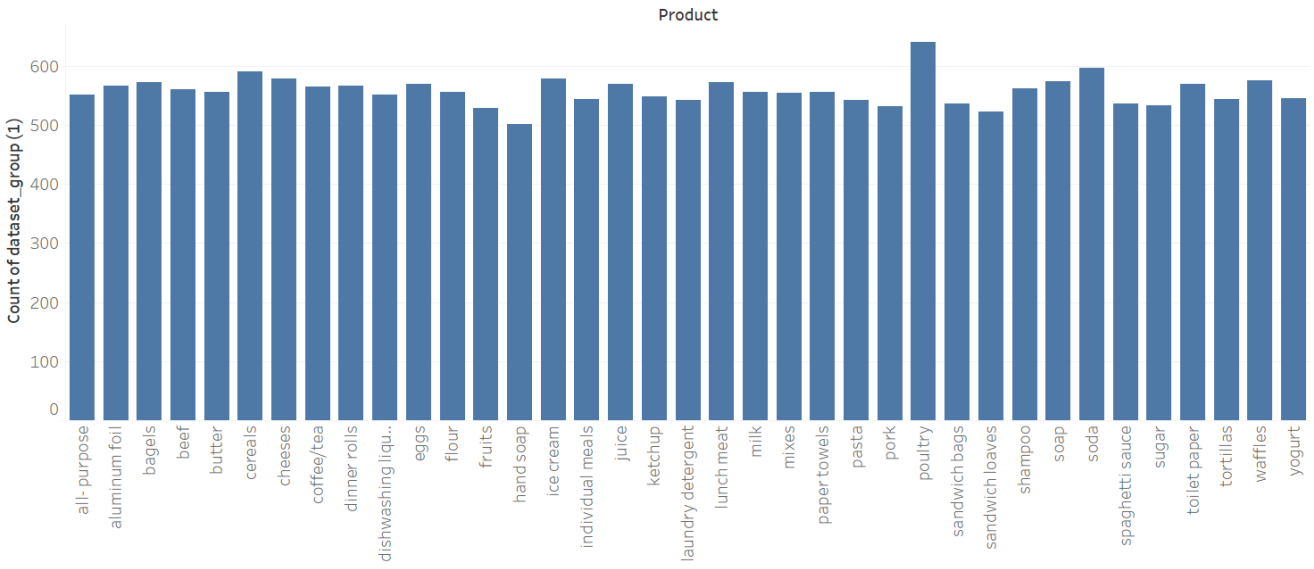
- We have 20641 rows and 3 columns in this data set.

```
 #   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)
```

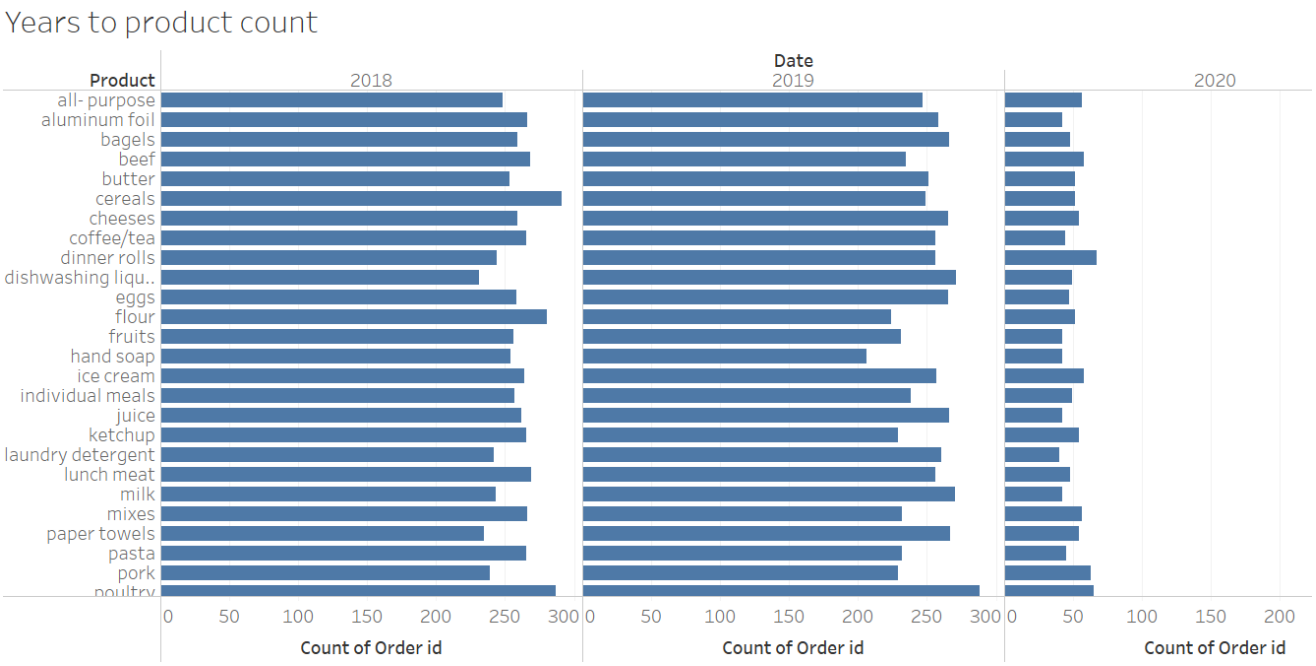
- No null values are there
- We have to convert Date column to datetime format

	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

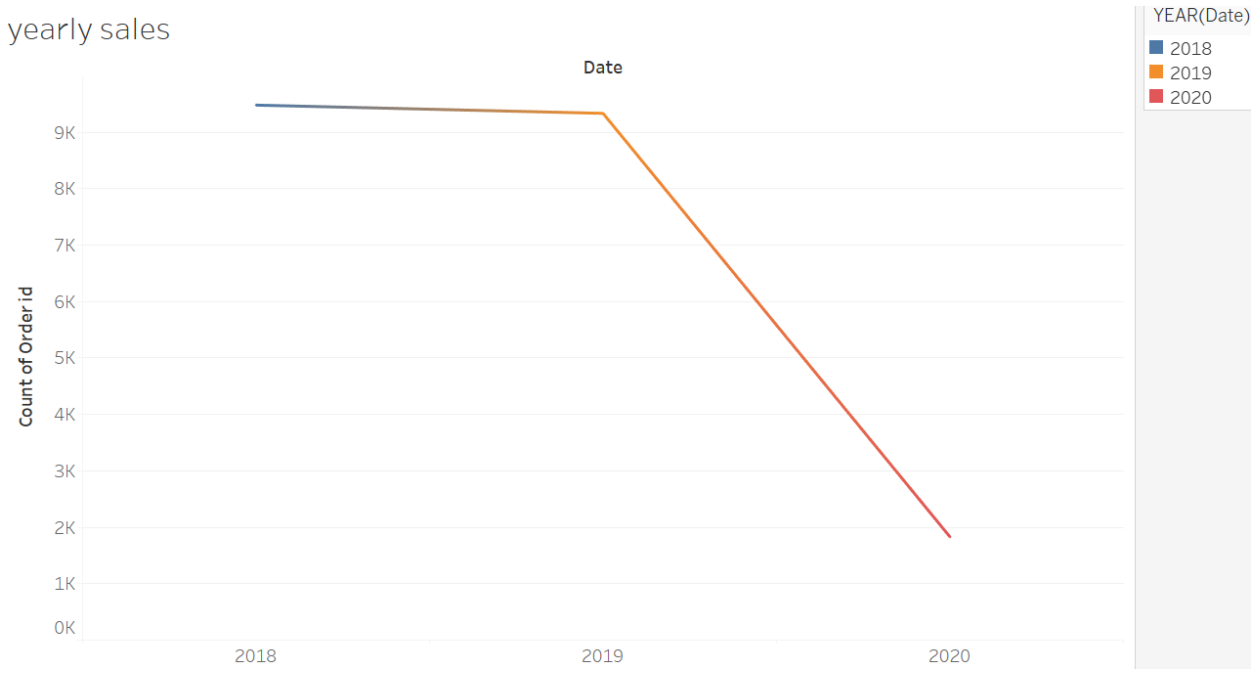
- Only int column we have is order ID, we have 1139 order Id in total.
- We have data of 1066 days starting from 2018-01-01 to 2020-12-02
- We have total of 37 unique products



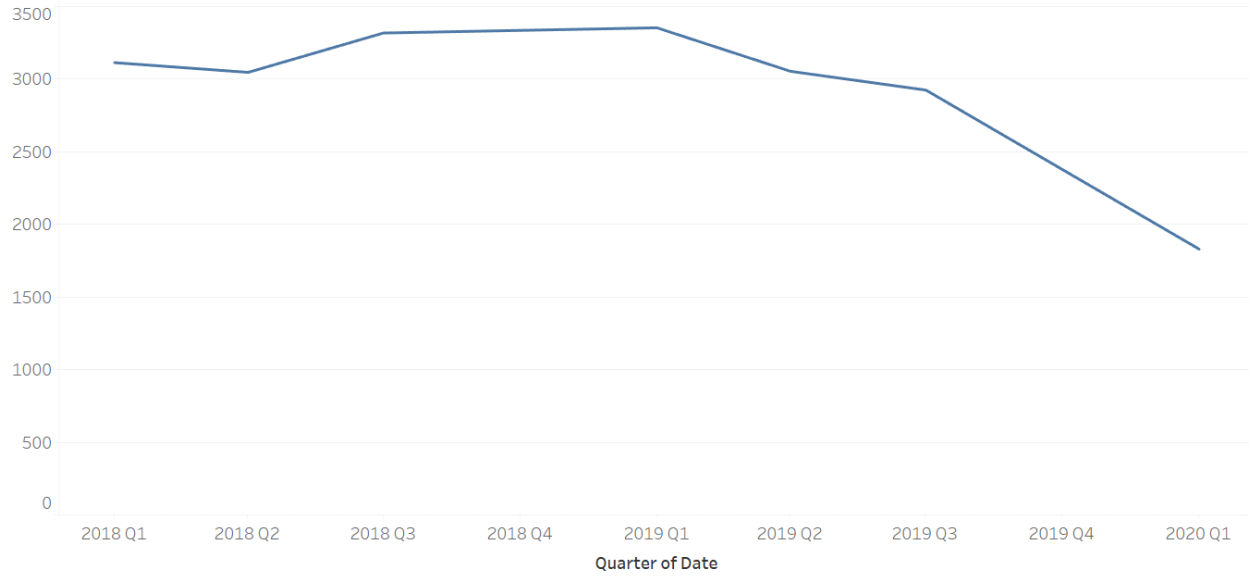
- Frequency of all the products is almost the same.



- The frequency of sales of each product along different years is also quite consistent and all products have sort of the same frequency.

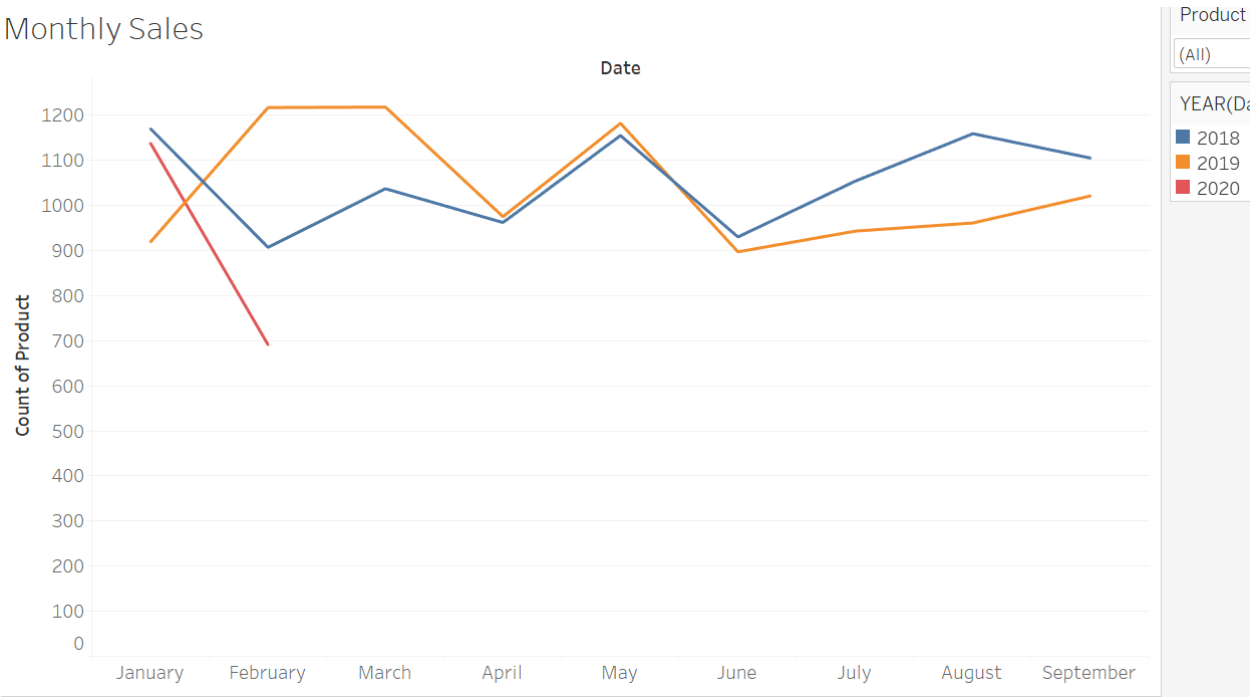


- Number of orders for 2018 and 2019 are almost the same , although they decline in 2020 because of just 2 months of data available for 2020.



- Number of products sold in 2019Q4 are quite low and the starting of Q1 2020 is also not that good.

- We can also see a pattern where more products are being sold in Q1 for both 2018 and 19 compared to Q2 and a repetitive dip is there in the second quarter for both mentioned years.



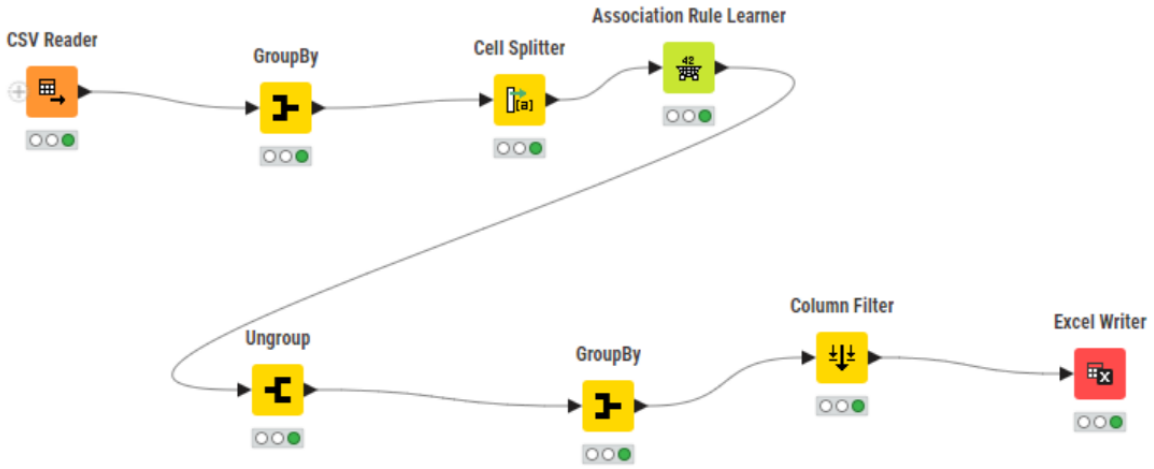
- We do not have any sales in the months of *October, November and December.*
- Sales or number of products sold for February 2020 are lowest.
- May is a consistent month in terms of the number of products sold in both 2018 and 19.
- In other months there is no drastic difference in the number of products sold , there are highs and lows without pattern in the rest of the months.

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

- Association rules are a data mining technique used to discover interesting relationships or patterns within large datasets. These rules aim to identify strong associations between items in a transactional database. In this case, the dataset consists of transactions with the date, order ID, and products purchased.
- The relevance of association rules in this case lies in understanding the purchasing patterns of customers. By analyzing the association between different products, we can gain valuable insights into customer behavior, preferences, and the relationships between items frequently bought together.
- For example, the association rules can reveal which products are frequently purchased together in the same transaction. If yogurt and pork are frequently bought together, it may indicate that these two items are complementary, and offering a discount or promotion on both items can potentially increase sales and encourage customers to purchase them together.
- Association rules can also help identify cross-selling opportunities. For instance, if customers often buy soda and paper towels together, the company can strategically place these items near each other in the store or offer a special bundle deal to encourage customers to purchase both items.
- Additionally, association rules can aid in inventory management and supply chain optimization. By knowing which products are frequently bought together, the company can plan their stock levels and ensure that popular combinations are always available to meet customer demand.

- Overall, association rules provide valuable insights into the relationships between products and customer behavior, helping businesses make data-driven decisions to improve their marketing strategies, optimize their product offerings, and enhance customer satisfaction.

Knime Workflow Images



- We read the file through csv reader
- Then we group-by the data using order_id to get an insight that what products are usually bought together under one order_id.
- Then we used cell splitter to make a set out of these grouped products to remove duplicacy of products in each order_id
- After that we used association rule learner to make rules.
- Once the rules are there then we wanted to import that file to excel but due to the presence of set columns we were unable to export excel file having association rules.
- So we used the ungroup and group-by option to convert that column into a string and we got a file with association rules.

Rows: 1247 | Columns: 5

Table | Statistics

#	RowID	Consequent String	Mean(Suppo... Number (double)	Mean(Confidence) Number (double)	Mean(Lift) ↓ Number (double)	Concatenate(Items (#1)) String
690	Row689	paper towels	0.055	0.649	1.791	eggs, ice cream, pasta
713	Row712	pasta	0.055	0.643	1.731	paper towels, eggs, ice cream
142	Row141	cheeses	0.051	0.674	1.726	bagels, cereals, sandwich bags
518	Row517	juice	0.05	0.64	1.7	yogurt, toilet paper, aluminum foil
685	Row684	mixes	0.051	0.63	1.678	yogurt, poultry, aluminum foil
953	Row952	sandwich bags	0.051	0.611	1.66	cheeses, bagels, cereals
271	Row270	dinner rolls	0.054	0.642	1.651	spaghetti sauce, poultry, laundry detergent
269	Row268	dinner rolls	0.052	0.641	1.649	spaghetti sauce, poultry, ice cream
514	Row513	juice	0.05	0.62	1.645	yogurt, poultry, aluminum foil
786	Row785	poultry	0.052	0.686	1.628	dinner rolls, spaghetti sauce, ice cream
362	Row361	eggs	0.052	0.634	1.627	paper towels, dinner rolls, pasta

Write about threshold values of Support and Confidence

- We have tried different values for support and confidence and we end up having best results at confidence= 0.5 and support as 0.05
- Let's understand their significance-

Support = 0.05 (5%):

- Support refers to the proportion of transactions that contain a specific itemset (basket) out of all the transactions.
- In this case, setting the support threshold to 0.05 means that we are considering only those itemsets (baskets) that appear in at least 5% of the total 1139 baskets.
- Itemsets that are less frequent than this threshold will be excluded from the analysis.

Confidence = 0.5 (50%):

- Confidence measures the strength of association between two itemsets (antecedent and consequent) in a rule.
- In this case, setting the confidence threshold to 0.5 means that we are interested in rules where the likelihood of the consequent itemset being purchased given the antecedent itemset is at least 50%.
- This ensures that only rules with relatively strong associations are considered, where the consequent is purchased in at least half of the transactions where the antecedent is present.

Practical Implications:

- With a support threshold of 0.05, we will focus on finding itemsets that appear frequently enough in the baskets, indicating their popularity among customers.
- With a confidence threshold of 0.5, we will prioritize rules with a strong association between items, indicating a substantial influence of one item on the purchase of another.
- The chosen threshold values ensure that the discovered association rules are both relatively common and have a reasonably strong predictive power, making them more relevant and actionable for business decision-making.

- We have got 1247 association rules from this association-

Rows: 1247 | Columns: 6

Table | Statistics

#	Row...	Support ↓ Number (double)	Confidence Number (double)	Lift Number (double)	Consequent String	implies String	Items Set
1	rule0	0.05	0.576	1.497	yogurt	<---	[toilet paper,juice,aluminum...
2	rule1	0.05	0.594	1.544	aluminum foil	<---	[yogurt,toilet paper,juice]
3	rule2	0.05	0.538	1.421	toilet paper	<---	[yogurt,juice,aluminum foil]
4	rule3	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminu...
5	rule4	0.05	0.523	1.36	yogurt	<---	[poultry,juice,aluminum foil]
6	rule5	0.05	0.564	1.468	aluminum foil	<---	[yogurt,poultry,juice]
7	rule6	0.05	0.538	1.276	poultry	<---	[yogurt,juice,aluminum foil]
8	rule7	0.05	0.62	1.645	juice	<---	[yogurt,poultry,aluminum fo...
9	rule8	0.05	0.588	1.528	yogurt	<---	[cheeses,cereals,coffee/tea]
10	rule9	0.05	0.594	1.52	cheeses	<---	[yogurt,cereals,coffee/tea]
11	rule10	0.05	0.582	1.469	cereals	<---	[yogurt,cheeses,coffee/tea]

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

Initially we are picking up 20 associations with maximum lift.

Antecedent	Consequent	Lift	Support	Confidence
eggs, ice cream, pasta	paper towels	1.791193	0.055312	0.649485
paper towels, eggs, ice cream	pasta	1.731003	0.055312	0.642857
bagels, cereals, sandwich bags	cheeses	1.726209	0.050922	0.674419
yogurt, toilet paper, aluminum foil	juice	1.700401	0.050044	0.640449
yogurt, poultry, aluminum foil	mixes	1.677722	0.050922	0.630435
cheeses, bagels, cereals	sandwich bags	1.659641	0.050922	0.610526
spaghetti sauce, poultry, laundry detergent	dinner rolls	1.650921	0.053556	0.642105
spaghetti sauce, poultry, ice cream	dinner rolls	1.648862	0.051800	0.641304
dinner rolls, spaghetti sauce, ice cream	poultry	1.627931	0.051800	0.686047
paper towels, dinner rolls, pasta	eggs	1.627458	0.051800	0.634409
paper towels, eggs, dinner rolls	pasta	1.621098	0.051800	0.602041
spaghetti sauce, poultry, cereals	dinner rolls	1.620915	0.050922	0.630435
paper towels, ice cream, pasta	eggs	1.616149	0.055312	0.630000
yogurt, cheeses, cereals	coffee/tea	1.615965	0.050044	0.612903
spaghetti sauce, poultry, juice	dinner rolls	1.613779	0.051800	0.627660
dinner rolls, poultry, soda	eggs	1.610145	0.051800	0.627660
dinner rolls, poultry, laundry detergent	spaghetti sauce	1.602745	0.053556	0.598039
poultry, laundry detergent, cereals	milk	1.589251	0.050922	0.604167
dinner rolls, poultry, ice cream	spaghetti sauce	1.581200	0.051800	0.590000
dinner rolls, poultry, juice	spaghetti sauce	1.565545	0.051800	0.584158

Interpretation of the values in the provided data:

- The first row shows that the itemset "eggs, ice cream, pasta" is associated with the consequent "paper towels" with a support of 0.055312. This means that 5.53% of all transactions in the dataset contain the itemset "eggs, ice cream, pasta" and "paper towels" together.
- The confidence for this rule is 0.649485, indicating that 64.95% of the transactions containing "eggs, ice cream, pasta" also contain "paper towels."

- The lift is 1.791193, suggesting that the purchase of "paper towels" is about 1.79 times more likely when "eggs, ice cream, pasta" is already purchased, compared to the case when "paper towels" is purchased independently.

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 the associations discovered from the analysis, here are some suggestions for possible combos with lucrative offers:

Combo Deal 1: "Family Brunch Bonanza"

Items: *Eggs, Ice Cream, Pasta, Paper Towels*

Offer: Buy all four items together and get a **20% discount on the total purchase.**

Combo Deal 2: "Healthy Snack Pack"

Items: *Yogurt, Poultry, Aluminum Foil, Juice*

Offer: Buy all four items together and **get a pack of Juice for free.**

Combo Deal 3: "Italian Pasta Night"

Items: *Spaghetti Sauce, Poultry, Ice Cream, Dinner Rolls*

Offer: Buy Spaghetti Sauce and Poultry together and get a **50% discount on Ice Cream and Dinner Rolls.**

Combo Deal 4: "Breakfast Delight"

Items: *Yogurt, Cheeses, Cereals, Coffee/Tea*

Offer: Buy all four items together and get a **10% discount on the total purchase.**

Combo Deal 5: "Weekend BBQ Party"

Items: *Spaghetti Sauce, Poultry, Soda, Eggs*

Offer: Buy Spaghetti Sauce and Poultry together and get a **6-pack of Soda for free.**

These combo deals are designed to leverage the associations between items to create attractive offers that encourage customers to make larger purchases and increase overall sales.

By bundling items that are frequently purchased together, the company can provide value to customers and enhance their shopping experience. Additionally, offering free or discounted items as part of the combos can entice customers to explore new products and try different combinations.