



Market basket Analysis

Abubacker
Siddhique

Problem Statement

We are provided with the data that contains the details of the items sold from a grocery store. Our role is to conduct a thorough analysis of Point of Sale (POS) data, identify the most commonly occurring sets of items in customer orders, and provide recommendations for a grocery store to increase its revenue through popular combo offers & discounts for customers.



Data Dictionary

- Date-Date of transaction.
- Order ID- Unique ID of each transaction.
- Product- Name of the product.

Content

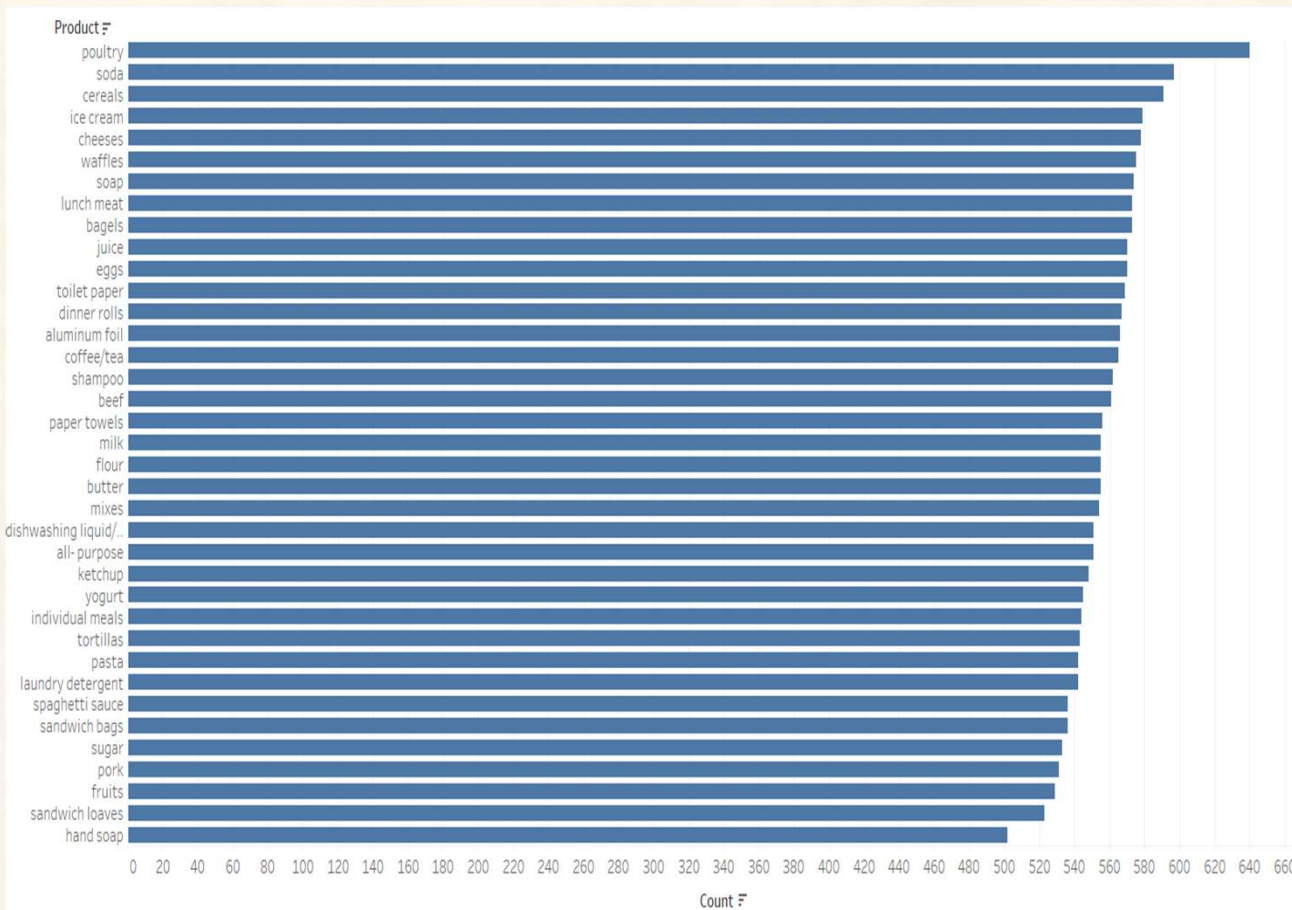
- Exploratory Data Analysis
- Executive Summary of the data
- Market Basket Analysis
- Associations Rules
- Knime workflow
- Threshold Values
- Identified Associations
- Recommendation





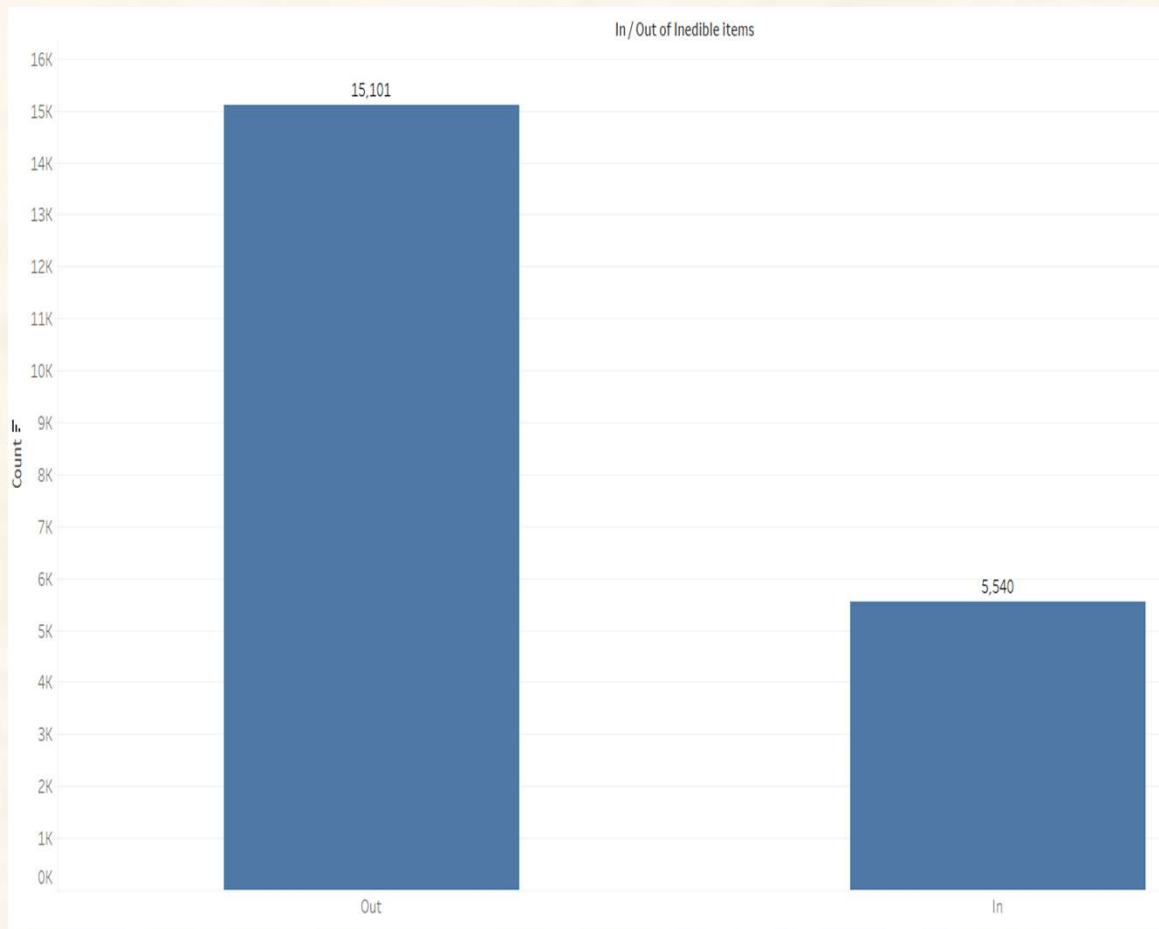
Exploratory Data Analysis

Product Distribution



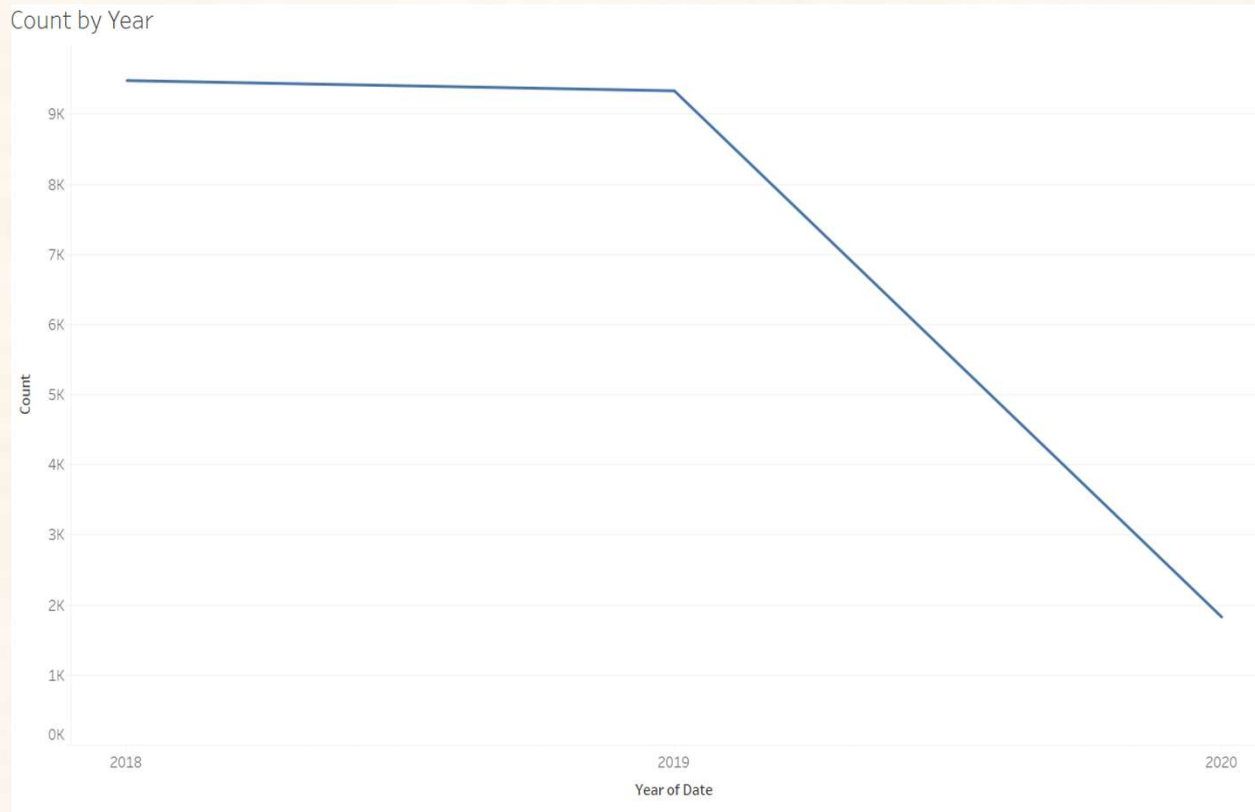
- Poultry and soda are the top 2 most bought items.
- The Majority of items in the dataset are edible.
- Around 500 units of hand soap were sold, making it the least popular item.

Edible Vs Inedible items



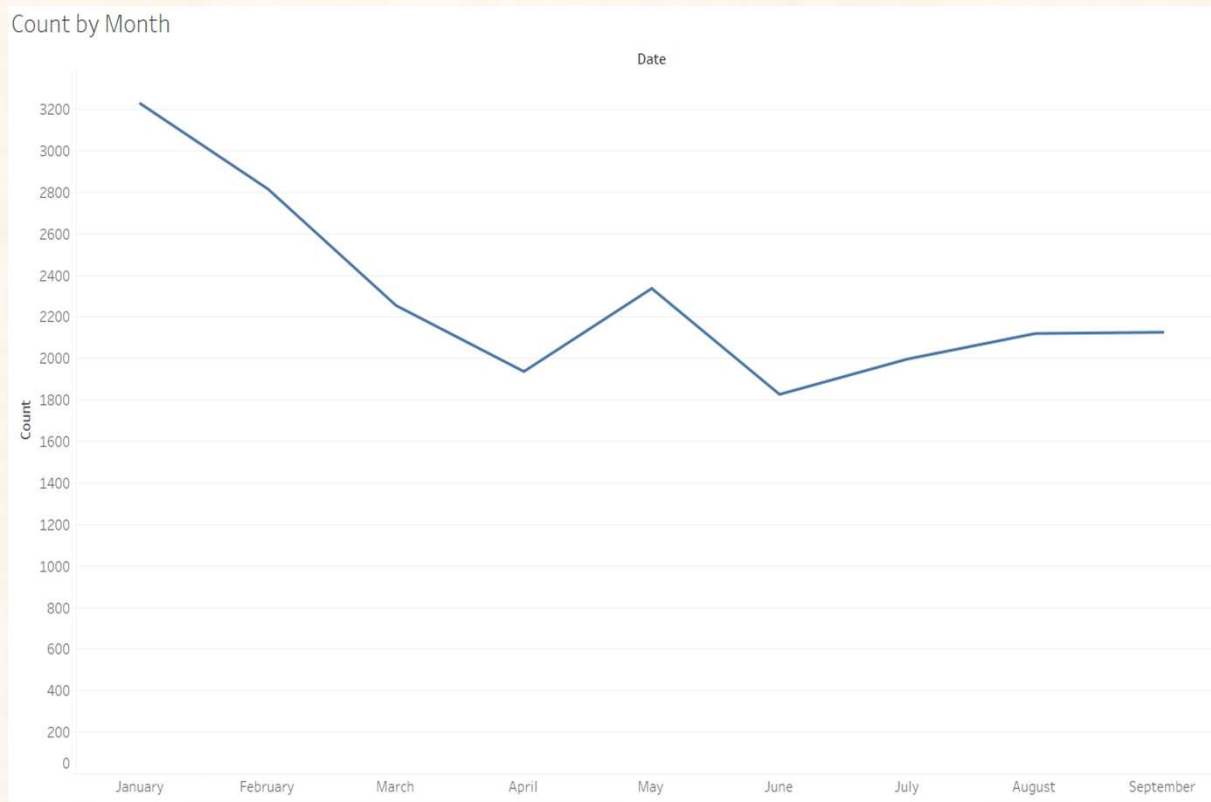
- As we observed in the previous slide, most of the purchased items fall into the categories of either edible goods or beverages.
- The number of edible items sold is nearly three times that of the inedible items

Sales By Year



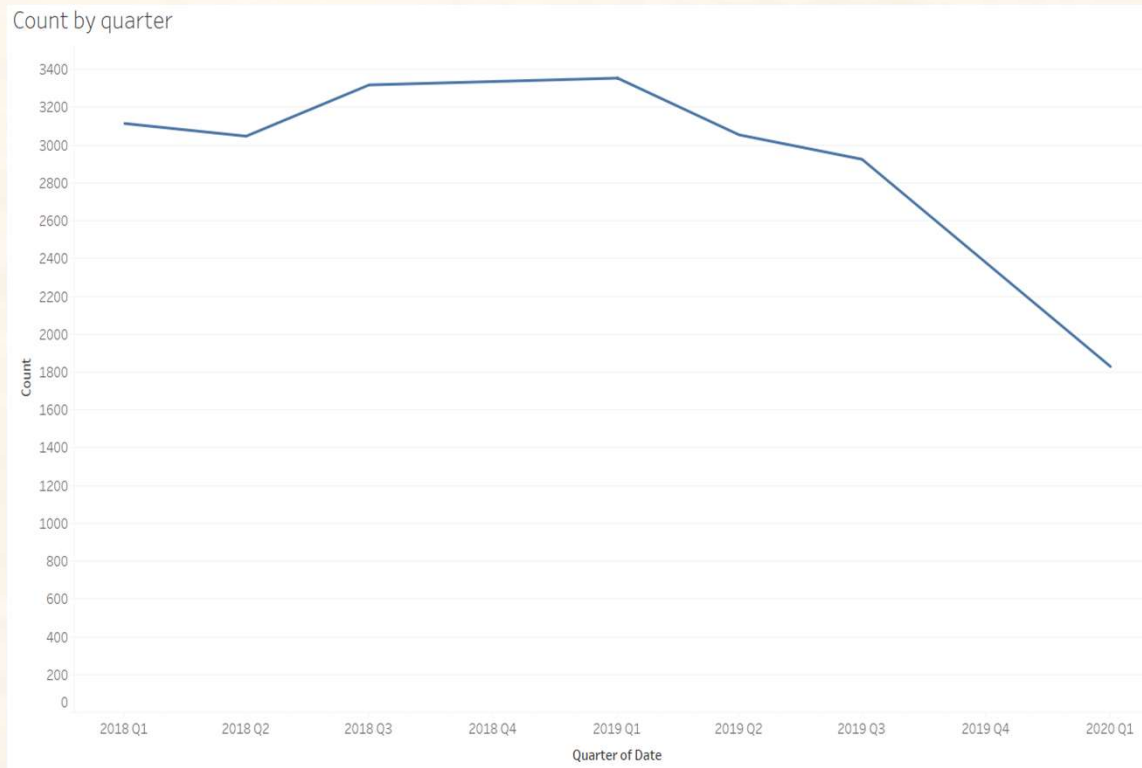
- In terms of items sold, the year 2018 holds the record for the highest sales.
- We can observe a declining trend year over year.
- Notably, there was a sharp dip in quantities sold from 2019 to 2020.
- This could be due to limited data available for 2020, spanning only two months, which could explain the sharp dip in quantities sold during that year.

Sales By Month



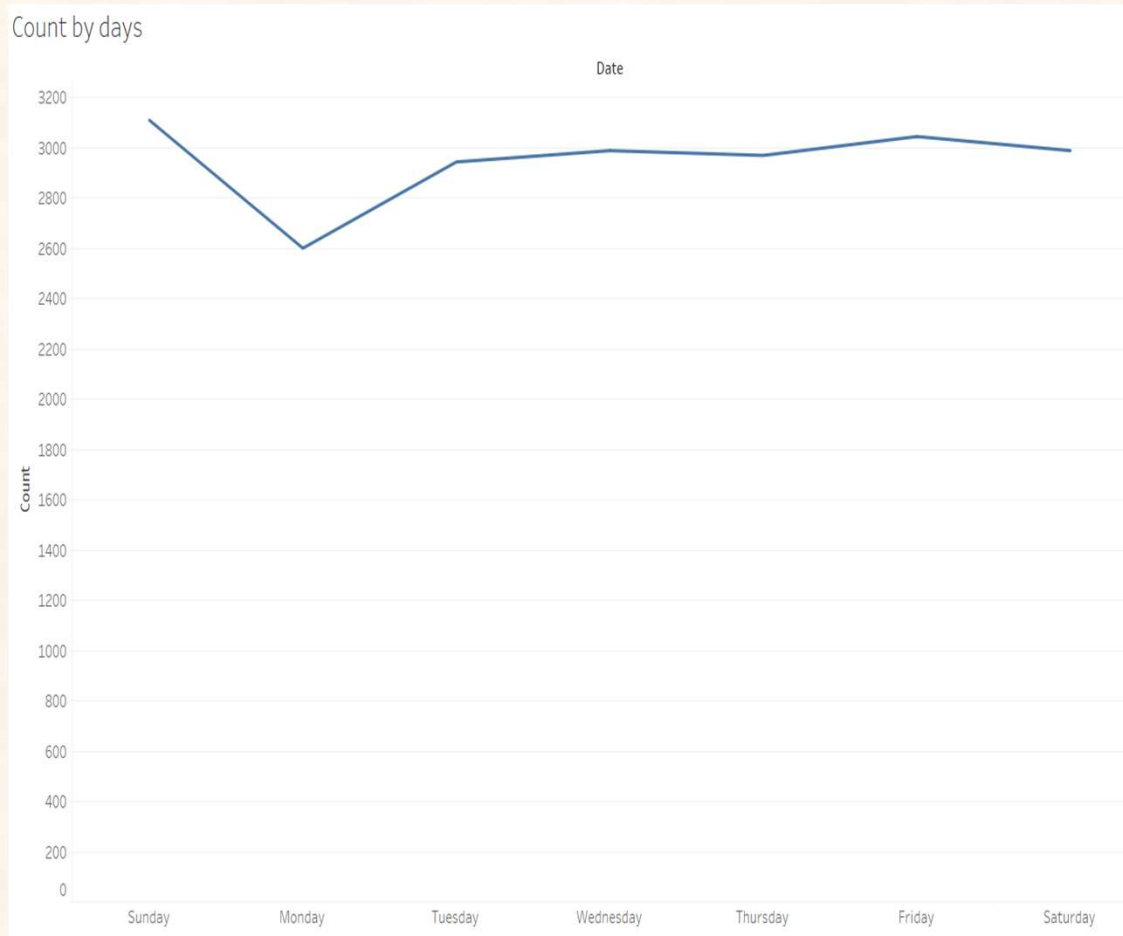
- Sales peak at the beginning of the year and gradually decline toward the end.
- There is a sudden spike during the month of May.
- June has the lowest sales compared to all other months.

Sales By Quarter



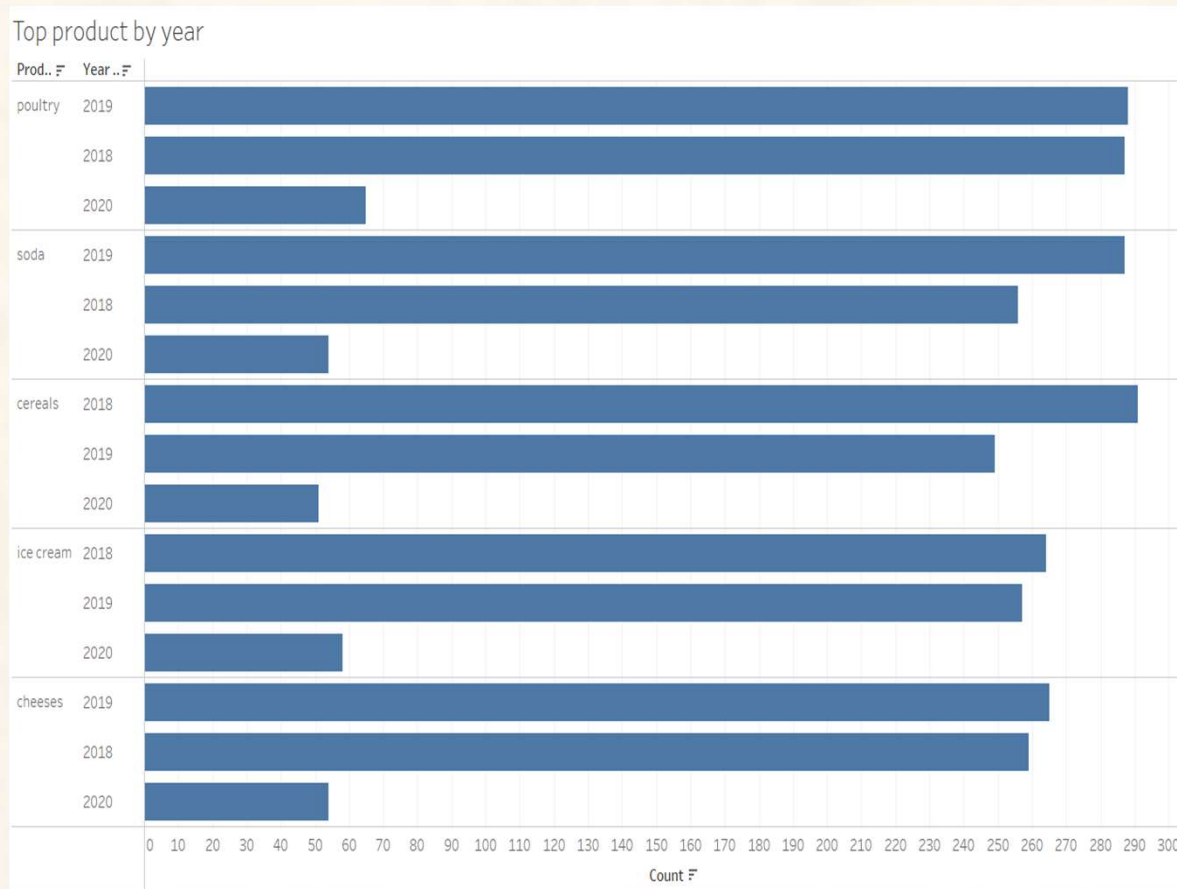
- A similar decreasing trend is observed if we analyze the data by each quarter.
- There is a significant decrease in sales after 2019 Q3.

Sales By Week



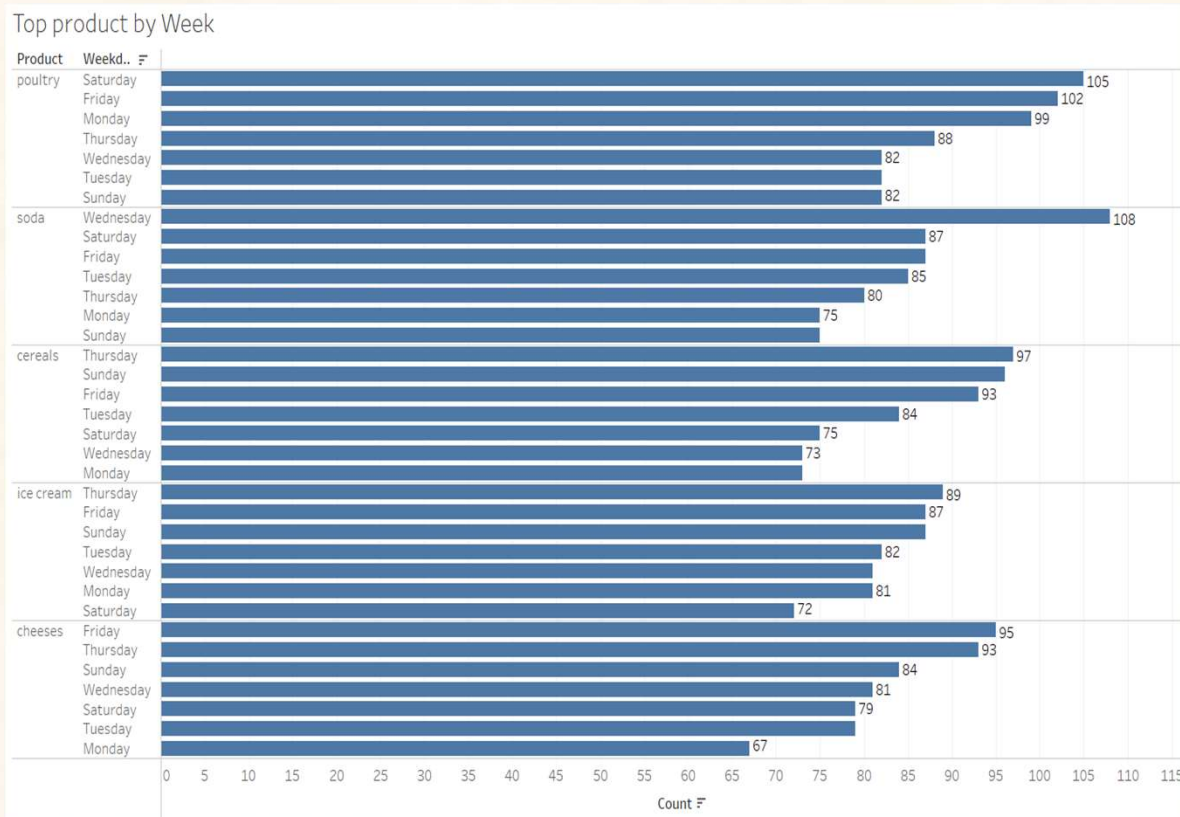
- Sundays consistently show the highest sales.
- This trend could be due to the fact that many people have a day off during the weekend and tend to visit grocery stores.
- Sales remain steady during the middle of the week.
- Mondays contribute the least to overall sales.

Sales of Top 5 Products by Year



- The sales of poultry, ice cream, and cheese have remained consistent over the years.
- There is a significant increase in sales of soda between 2018 & 2019.
- However, cereal sales experienced a sudden dip during those years.

Sales of Top 5 Products by Week



- Although overall sales are highest on Sundays, this pattern does not hold for the top five products.
- Thursdays, Fridays, and Saturdays have the highest overall sales of the top-selling products.
- Interestingly, the maximum soda sales occur on Wednesdays.

Executive Summary

- There are three columns present in the dataset
- The date contains the date of purchase, the order ID represents the unique ID of each transaction and the products represent the sold items.
- There are no null and missing values.
- There are 4730 duplicate records. We are not going to drop these duplicates as there can be multiple items under the same order ID.
- The data is available only till September.
- Each product displays distinct weekly trends.

Market Basket Analysis

Definition:

Market basket analysis is a technique used to find customer transaction patterns. It is used primarily by retailers to identify associations between products frequently purchased together, allowing them to group these items and boost sales for both.

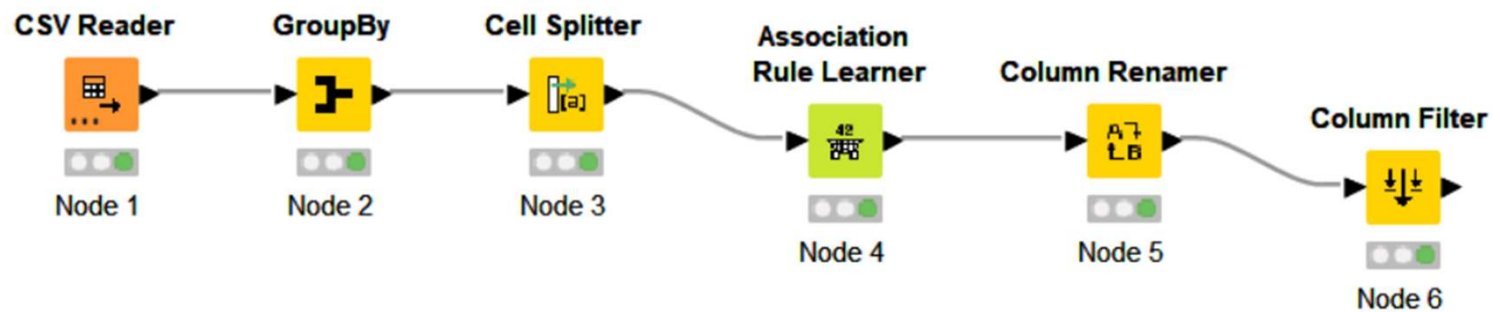
Applications:

- Retailers can organize products in their stores based on the Market basket analysis insights to increase sales and revenue by influencing purchasing decisions.
- It can be used for any new product launch.
- Strategically spacing products apart on the shop floor encourages customers to explore other items in between, potentially intriguing their interest and leading to additional purchases.

Association Rules

- Association rules allow us to discover connections between two products and assess the strength of their relationship.
- It consists of three terms. **Support, Confidence & lift.**
- Let's say you have two products A & B.
- **Support:** Support indicates the frequency of products being sold together.
- **Confidence:** Confidence tells us how often item B appears in transactions containing item A. It is the probability of buying B given the customer already bought A.
- **Lift:** Lift measures the percentage of chance of buying B will increase if a buyer buys A. Lift quantifies how much more likely two items are to be purchased together compared to their individual probabilities.

Knime workflow



Threshold Values

The screenshot shows a software interface with four tabs: 'Options' (selected), 'Flow Variables', 'Job Manager Selection', and 'Memory Policy'. The 'Options' tab is divided into three sections: 'Itemset Mining', 'Output', and 'Association Rules'. In the 'Itemset Mining' section, 'Column containing transactions' is set to 'Concatenate(Product)_SplitResultSet', 'Minimum support (0-1)' is 0.05, and 'Underlying data structure' is 'ARRAY'. In the 'Output' section, 'Itemset type' is 'CLOSED' and 'Maximal itemset length' is 10. In the 'Association Rules' section, 'Output association rules' is checked, and 'Minimum confidence' is 0.601.

Options Flow Variables Job Manager Selection Memory Policy

Itemset Mining

Column containing transactions [...] Concatenate(Product)_SplitResultSet

Minimum support (0-1) 0.05

Underlying data structure: ARRAY

Output

Itemset type CLOSED

Maximal itemset length: 10

Association Rules

☒ Output association rules

Minimum confidence: 0.601

- Setting threshold values is crucial, as we don't want unwanted and redundant rules.
- A support of 0.05 means that, both the items together are present in at least 5% of the total transaction.
- A confidence of 0.60 indicates the probability of buying the recommended items. It measures the strength of the association rules.

Identified Associations

Row ID	D Support	D Confide...	D ▼ Lift	S Recom...	[...] Items bought
rule22	0.055	0.649	1.791	paper towels	[eggs,ice cream,pasta]
rule21	0.055	0.643	1.731	pasta	[paper towels,eggs,ice cream]
rule5	0.051	0.674	1.726	cheeses	[bagels,cereals,sandwich bags]
rule0	0.05	0.64	1.7	juice	[yogurt,toilet paper,aluminum foil]
rule3	0.051	0.63	1.678	mixes	[yogurt,poultry,aluminum foil]
rule4	0.051	0.611	1.66	sandwich bags	[cheeses,bagels,cereals]
rule17	0.054	0.642	1.651	dinner rolls	[spaghetti sauce,poultry,laundry detergent]
rule11	0.052	0.641	1.649	dinner rolls	[spaghetti sauce,poultry,ice cream]
rule1	0.05	0.62	1.645	juice	[yogurt,poultry,aluminum foil]
rule12	0.052	0.686	1.628	poultry	[dinner rolls,spaghetti sauce,ice cream]
rule15	0.052	0.634	1.627	eggs	[paper towels,dinner rolls,pasta]
rule16	0.052	0.602	1.621	pasta	[paper towels,eggs,dinner rolls]
rule7	0.051	0.63	1.621	dinner rolls	[spaghetti sauce,poultry,cereals]
rule20	0.055	0.63	1.616	eggs	[paper towels,ice cream,pasta]
rule2	0.05	0.613	1.616	coffee/tea	[yogurt,cheeses,cereals]
rule13	0.052	0.628	1.614	dinner rolls	[spaghetti sauce,poultry,juice]
rule10	0.052	0.628	1.61	eggs	[dinner rolls,poultry,soda]
rule9	0.051	0.604	1.589	milk	[poultry,laundry detergent,cereals]
rule19	0.055	0.624	1.565	ice cream	[paper towels,eggs,pasta]
rule6	0.051	0.617	1.558	cereals	[cheeses,bagels,sandwich bags]
rule18	0.054	0.656	1.556	poultry	[dinner rolls,spaghetti sauce,laundry detergent]
rule8	0.051	0.637	1.512	poultry	[dinner rolls,spaghetti sauce,cereals]
rule14	0.052	0.602	1.429	poultry	[dinner rolls,spaghetti sauce,juice]

Support, Confidence, & Lift

- It is vital to sort the output of the Market basket Analysis by lift, as it indicates the chances of being purchased together than would be expected if they were statistically independent.
- Rule 23 has the highest lift of 1.79, which means that there is a 1.8 times higher chance of a customer buying paper towel if they have already bought eggs, ice cream & pasta.
- Confidence of 0.649 indicates that there is a 60% probability of buying both items together.
- Support of 0.055 shows that both the items together are present in 5% of the total transaction.
- It is important to analyze these parameters, as it quantifies the association between them.

Summary & Recommendations

- Create a combo that includes eggs, ice cream, and pasta and recommend customers to buy paper towels at a relatively lower price.
- As eggs and ice creams are bought together frequently, introduce a new offer, if customers buy more than 2 ice creams, they get an egg for free.
- Highlight complementary items in ads or promotions to encourage cross-purchases.
- Given that most of the sales are from edible items, consider expanding the floor area dedicated to these products to prevent frequent stockouts.
- During their respective peak hours, consider launching new products for the top-selling items. For example, introduce a new soda product on Wednesdays.
- Provide a discount on cheese, if the customer has already bought bagels, sandwich bags & cereals.
- To optimize sales, refrain from launching new products in April or June, as they historically have the lowest overall sales.

Summary & Recommendations

- Place the top-selling products in high visibility areas, and place the recommended items nearby.
- Create a bundle that includes dinner rolls and paper towels of various sizes, as these are highly recommended items.
- Analyze the unexpected association by sending out questionnaires to customers, understanding the reason behind their purchase. For example, why spaghetti sauce, poultry, and laundry detergent are bought together.
- Offer timely discounts on items during their corresponding meal times. For instance, consider providing discounts on cereals, eggs and milk during breakfast hours.

Thank you

Abubacker
Siddhique

