MRA PROJECT – MILESTONE 2

SARAN VIGNESH R

MAY - 2024

rsvignesh616@gmail.com

AGENDA [TABLE OF CONTENT]

1. Problem Statement
2. Data Description
3. Exploratory Data analysis
4. Inferences of Exploratory Data analysis
5. MBA Analysis
6. MBA Analysis: Key Inferences

PROBLEM STATEMENT

Business Context:

In the highly competitive grocery retail industry, understanding customer buying patterns is crucial for enhancing sales, increasing customer satisfaction, and improving profitability. By identifying frequently purchased item combinations, grocery stores can craft effective marketing strategies, optimize inventory management, and tailor promotions to meet customer needs. Leveraging Point of Sale (POS) data can unlock valuable insights that drive customer-centric offerings, such as combo packs, discounts, and targeted promotions, which can increase basket size and improve customer retention. This analysis aligns with business goals by maximizing revenue, reducing operational costs, and boosting customer loyalty.

Objective:

The goal is to analyze the POS transactional data to identify frequently purchased item combinations. Using association rule mining or similar techniques, the aim is to uncover patterns that will help the store create targeted combo offers and discounts, ultimately driving revenue growth by increasing customer purchases and average basket size.

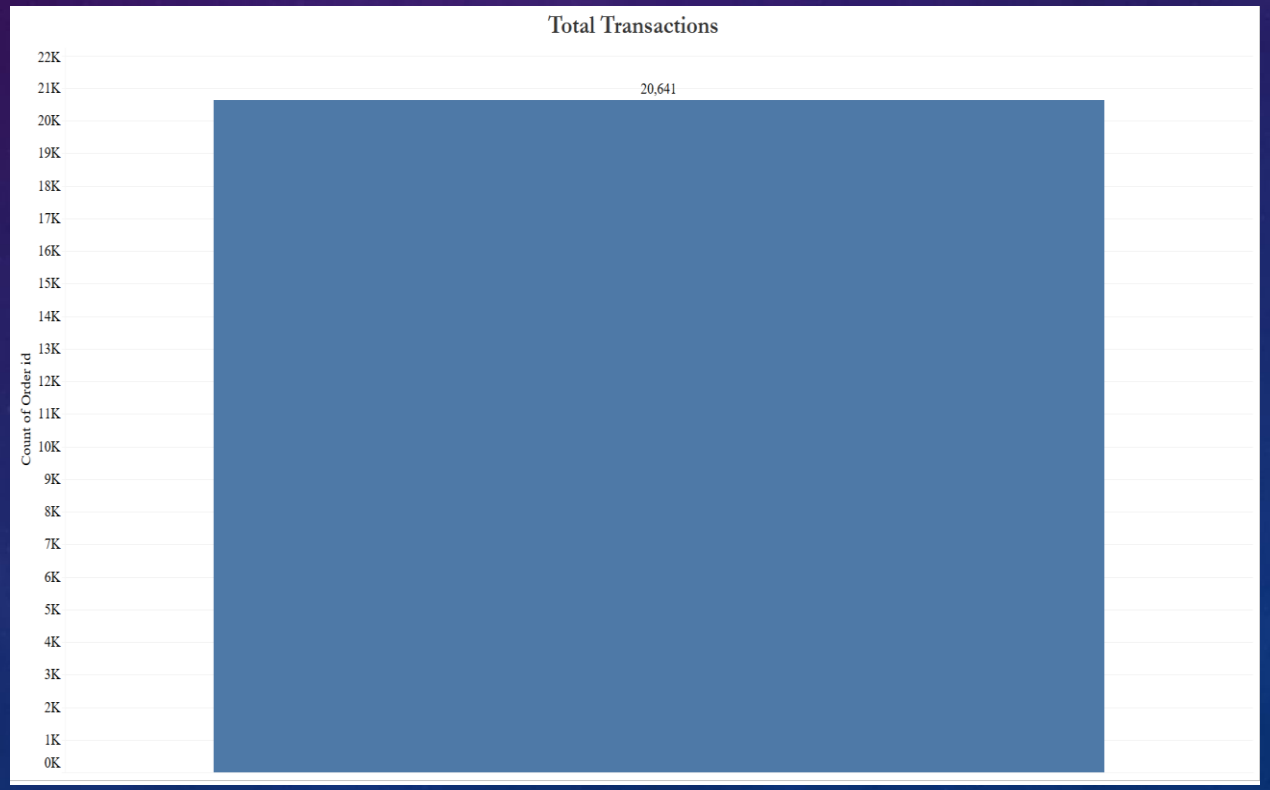
DATA DESCRIPTION:

The dataset consists of transactional data from a grocery store, where each row represents a product purchased in a specific order. The columns in the dataset are as follows:

- Date : The date when the transaction took place.
- Order_id : A unique identifier for each customer order.
- Product : The individual item purchased in the transaction.

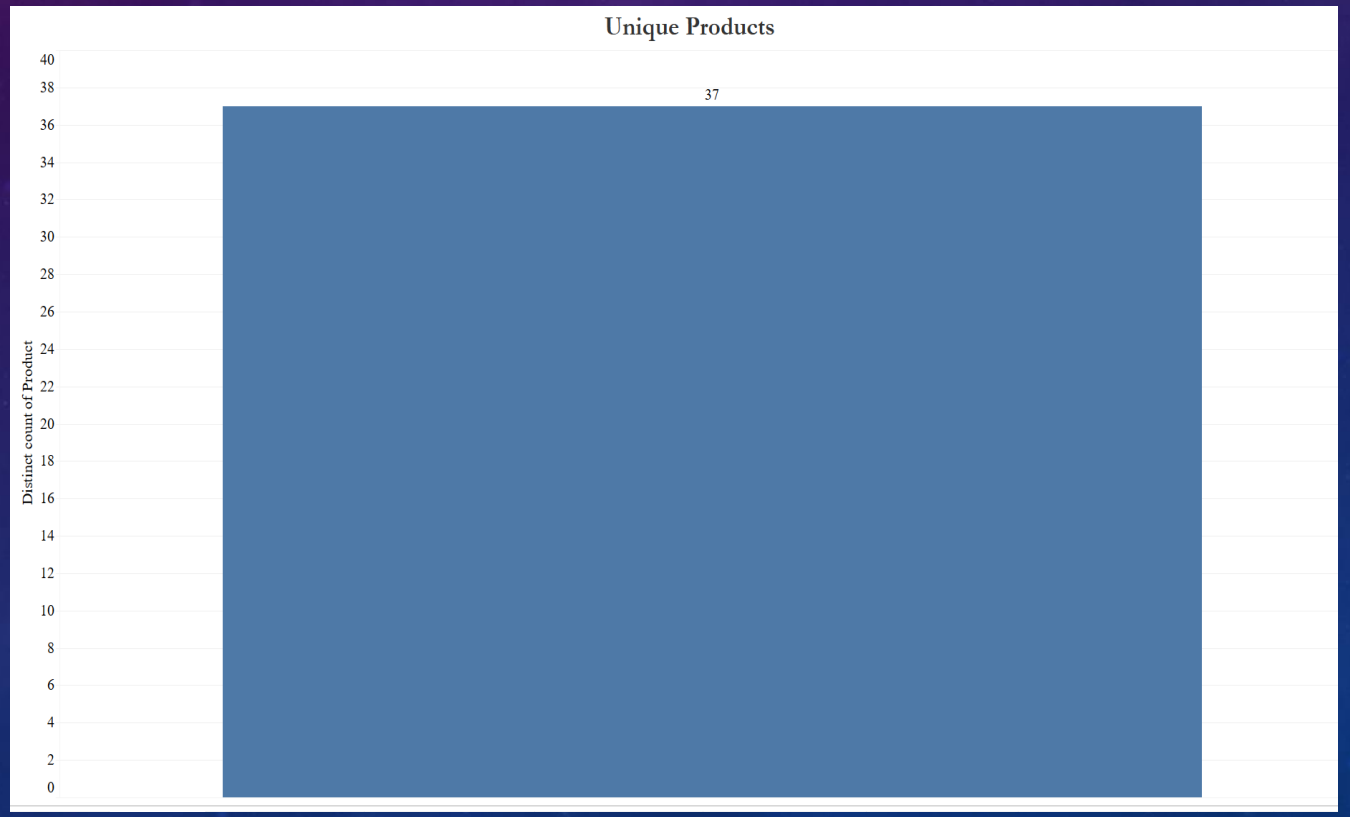
EXPLORATORY DATA ANALYSIS:

- High Transaction Volume: The chart indicates a substantial number of transactions, totaling 20,641. This provides a solid foundation for Market Basket Analysis, as a larger dataset increases the likelihood of identifying meaningful associations.
- Aggregate View: The chart only displays total transactions without segmentation. To gain deeper insights, further analysis is needed to break down these transactions by factors like date, time, customer segment, or product category.
- Potential for Further Analysis: Given the robust transaction volume, exploring patterns within these transactions could reveal valuable opportunities. Market Basket Analysis can help identify frequently purchased item combinations to inform marketing and merchandising strategies.



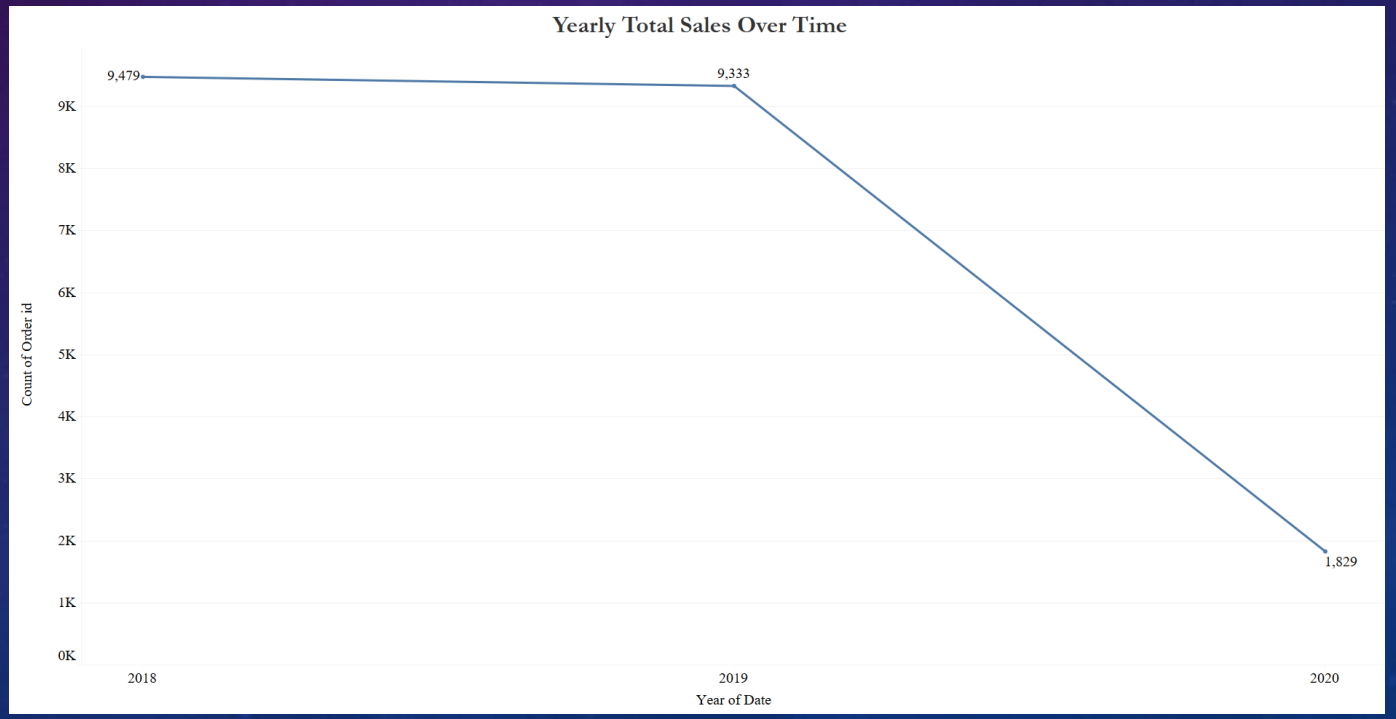
EXPLORATORY DATA ANALYSIS:

- Moderate Product Variety: The image indicates that there are 37 unique products available in the dataset. This suggests a moderately diverse product range, allowing for some meaningful Market Basket Analysis to uncover associations.
- Limited Scope for Variety-Based Strategies: With only 37 unique products, there might be limited scope for highly diverse product bundling strategies. The analysis should focus on the most frequently purchased items and their common combinations.
- Importance of Targeted Analysis: Given the constrained product range, it becomes even more crucial to perform targeted analysis to identify niche product associations and tailor marketing efforts to specific customer segments or seasonal trends.



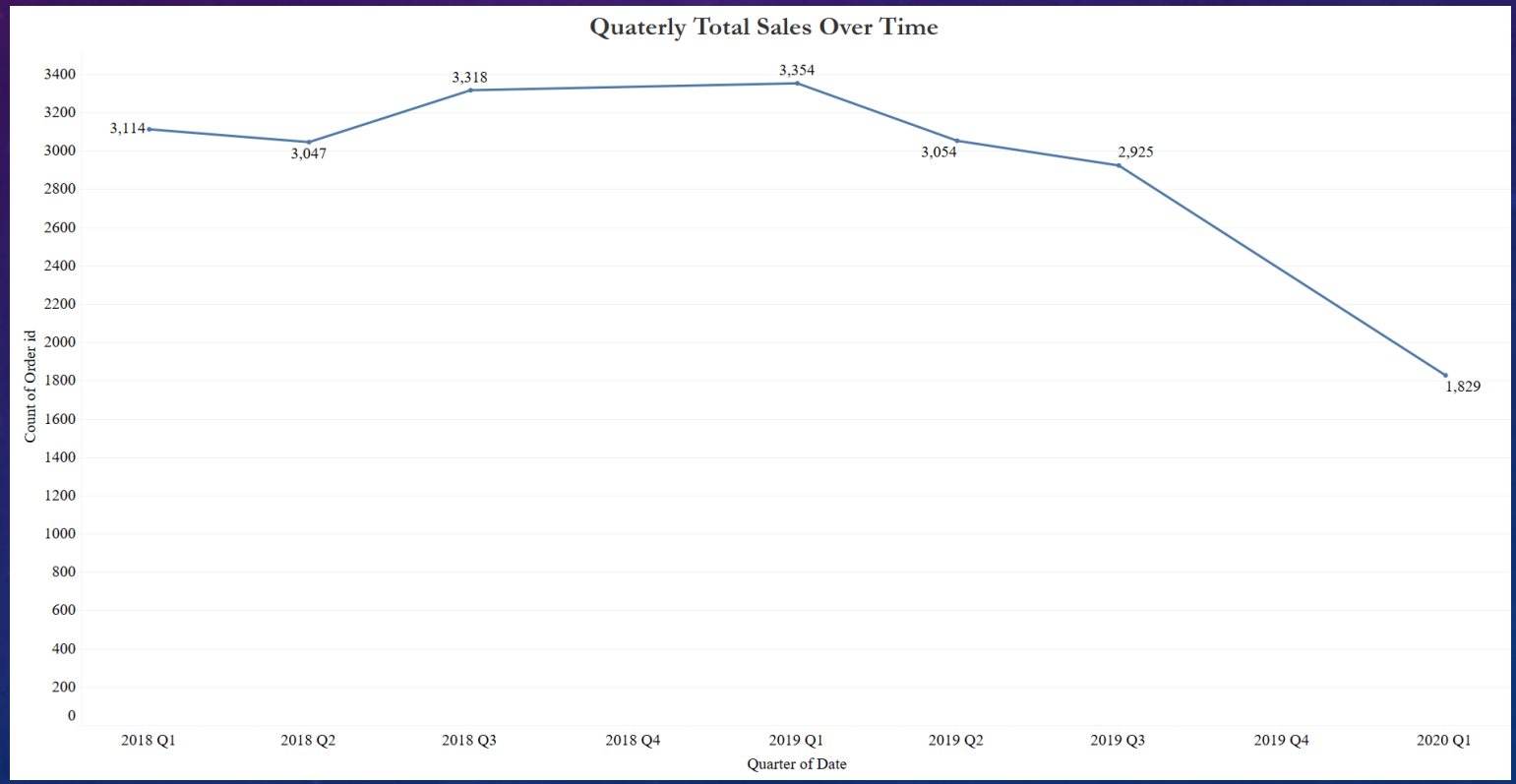
EXPLORATORY DATA ANALYSIS:

- Significant Sales Decline in 2020: There is a dramatic drop in total sales from 2019 (9,333) to 2020 (1,829). This represents a significant decrease in business activity.
- Relatively Stable Performance in 2018-2019: Sales were reasonably consistent between 2018 (9,479) and 2019 (9,333), indicating stability or minor growth during this period.
- Need for Further Investigation: The substantial decrease in sales in 2020 warrants a thorough investigation to understand the underlying causes, which could include economic factors, market shifts, or internal issues. This would necessitate examining factors that would impact the business.



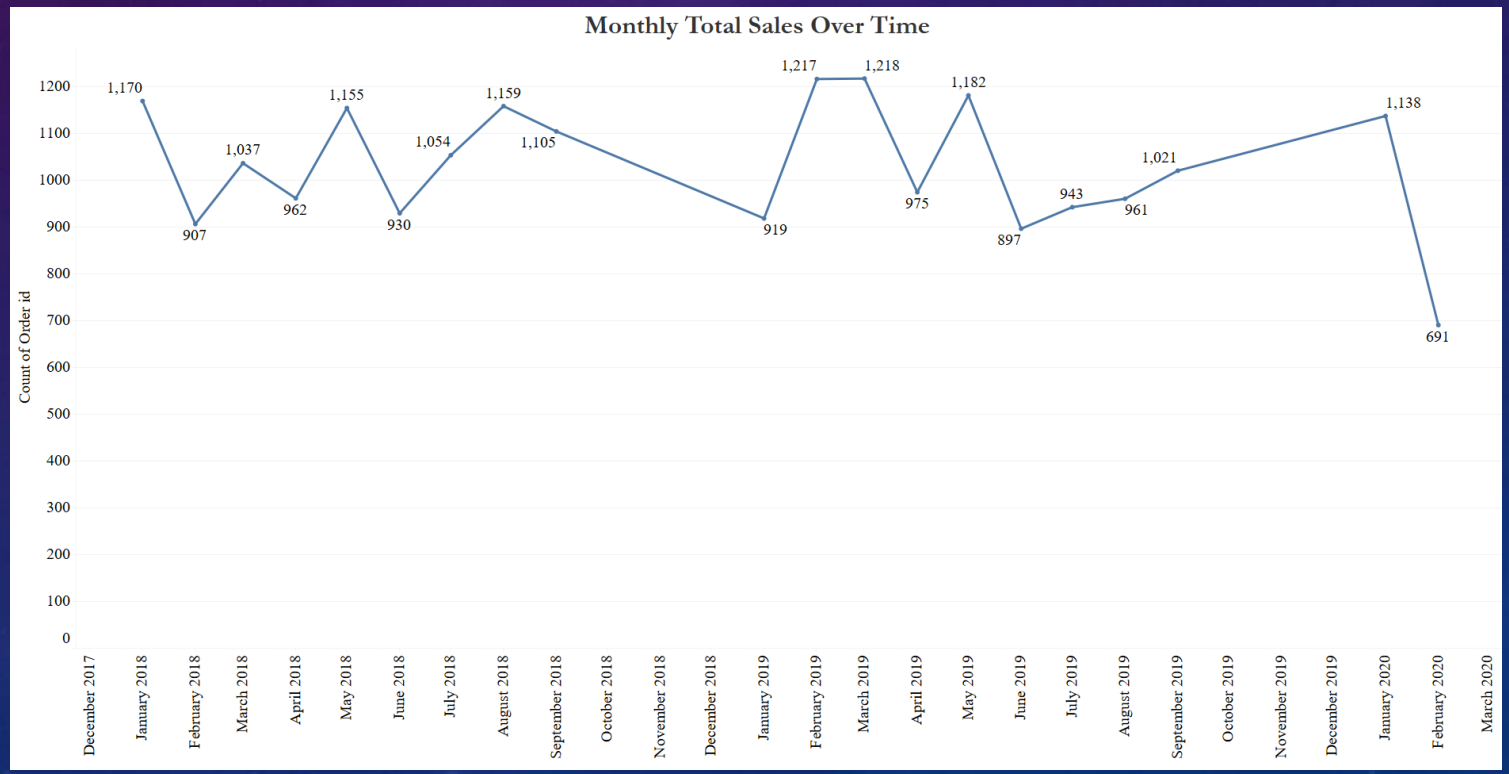
EXPLORATORY DATA ANALYSIS:

- Strong Quarterly Performance Through 2019: The business maintained a strong quarterly sales performance, consistently exceeding 3,000 order IDs, until the end of 2019. This demonstrates a robust market presence and effective sales strategies during this period.
- Peak Sales in Early 2019: The highest sales volume occurred in the first quarter of 2019 (3,354 order IDs), suggesting potential seasonal peaks or successful promotions.
- Dramatic Sales Decline in Early 2020: There was a significant drop in sales in the first quarter of 2020 (1,829 order IDs), indicating a major disruption or challenge affecting the business. This sudden decline requires further investigation to identify the causes and develop appropriate mitigation strategies.



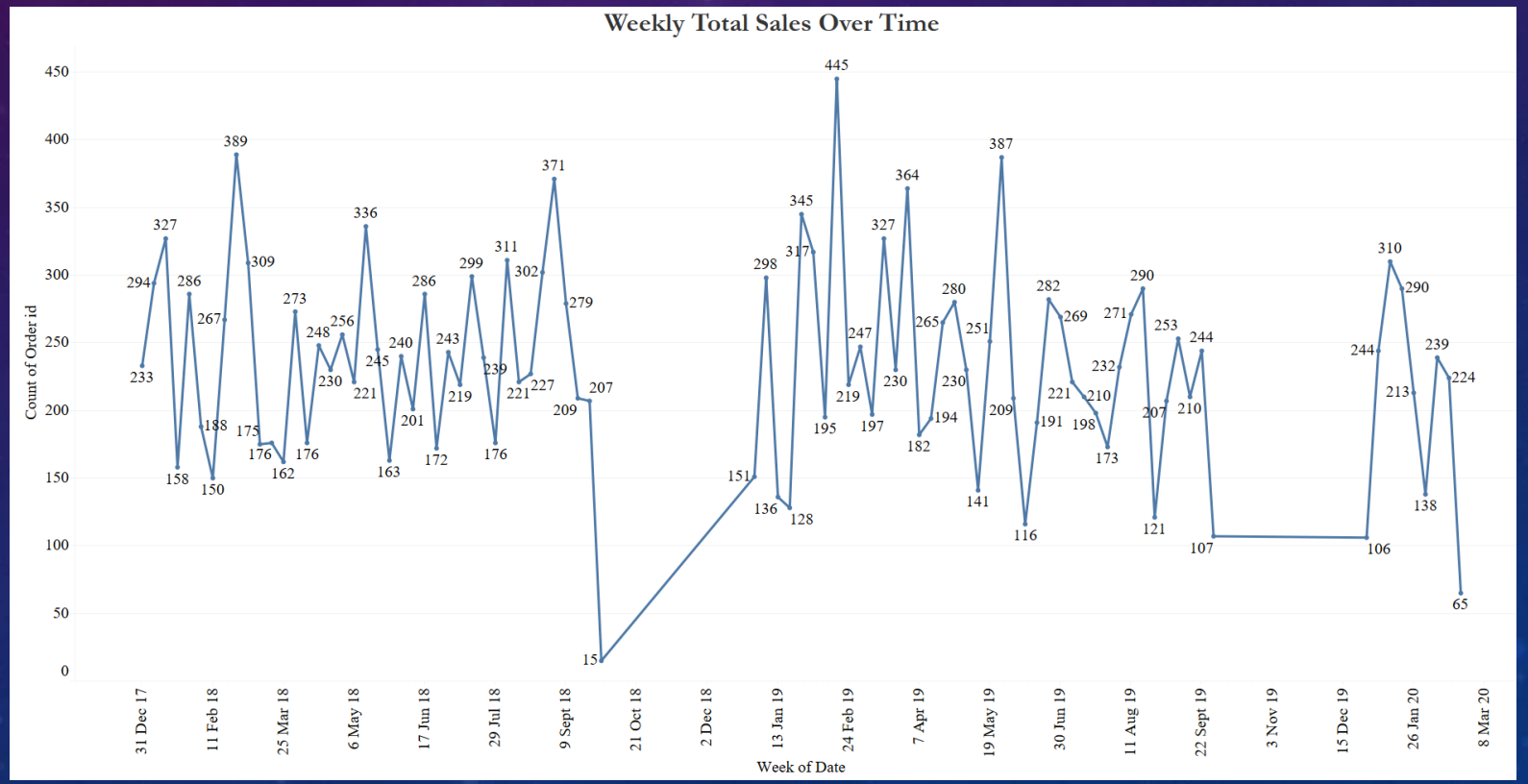
EXPLORATORY DATA ANALYSIS:

- Strong Early Performance (2018-2019): The period from December 2017 through 2019 demonstrates relatively consistent monthly sales, generally fluctuating between 900 and 1200 order IDs. This indicates a healthy market presence and consistent demand during this time.
- Mid-2019 Dip and Recovery: There is a notable dip in sales around July 2019, followed by a recovery toward the end of the year. Understanding the factors behind this dip (e.g., seasonal changes, promotional activity) could inform future strategies.
- Dramatic Decline in Early 2020: Starting in early 2020, specifically with March showing a steep drop to 691 order IDs, there is a significant decline. This abrupt downturn suggests a substantial disruption or external factor impacting the business.



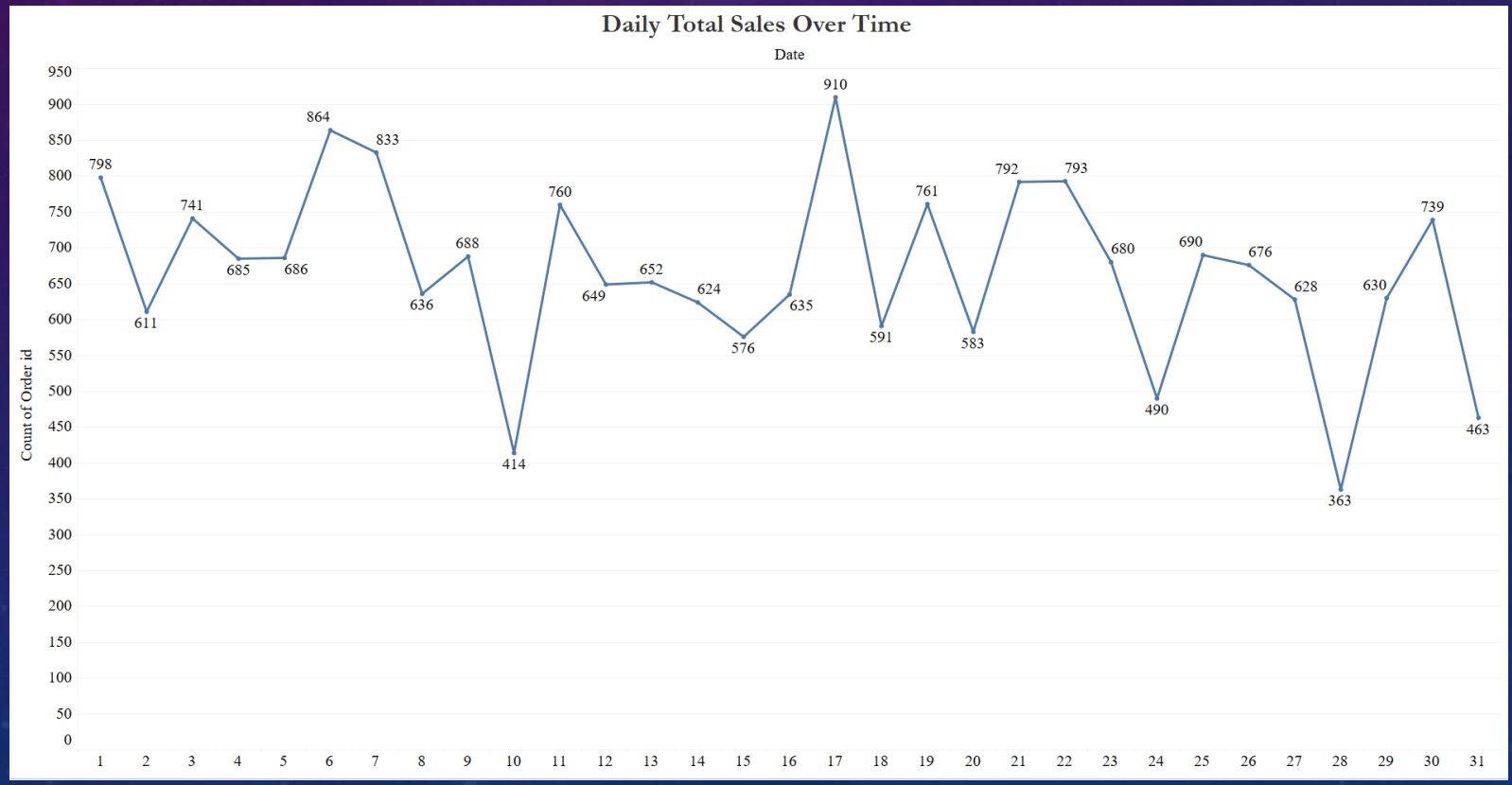
EXPLORATORY DATA ANALYSIS:

- Significant Weekly Sales Volatility: The chart shows substantial fluctuations in weekly sales. This indicates potential seasonal factors, promotional impacts, or external events affecting customer buying patterns. The range of the "Count of Order id" has extreme changes.
- Peak Sales Occur in Late 2018 and Early 2019: The highest sales volume occurs around late 2018 (371 in 9 Sept 18) and early 2019 (445 in 13 Jan 19), suggesting a potential seasonal peak during the holiday season or successful promotions.
- Sharp Decline in Sales in Late 2018: There is a sharp decline in sales around October 2018, dropping to only 15. An investigation of factors contributing to this significant drop should be conducted. There is another significant dip between the weeks of 3 Nov 19 and 8 Mar 20.



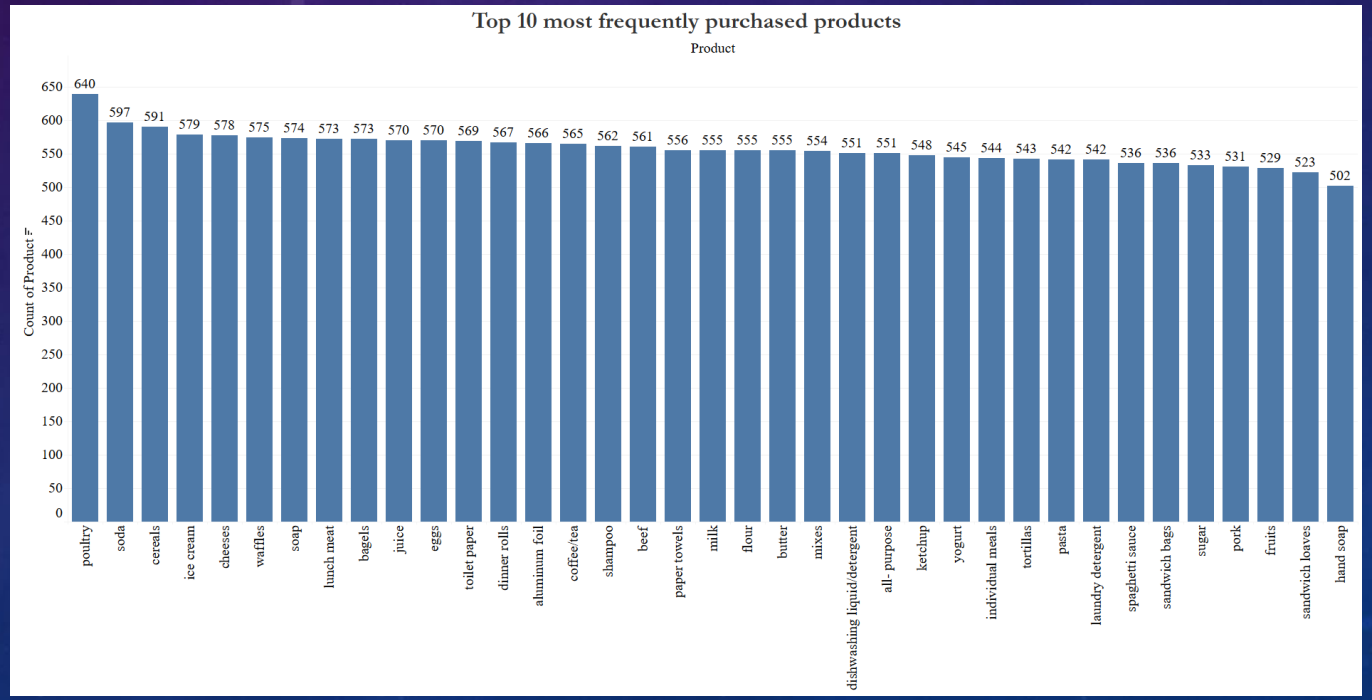
EXPLORATORY DATA ANALYSIS:

- High Daily Sales Variability: The chart displays substantial day-to-day fluctuations in sales. This is a crucial indicator, suggesting that external factors, promotions, or day-of-week effects significantly influence customer purchase behavior.
- Peak Sales Occur on Day 17: The highest sales volume occurs on Day 17 with 910. There needs to be an investigation of factors contributing to this peak.
- Significant Sales Dips: Sales experience significant declines, reaching lows on Day 10 and Day 28. An investigation of factors contributing to these significant drops should be conducted. There is a decline to sales around the days 24-28.



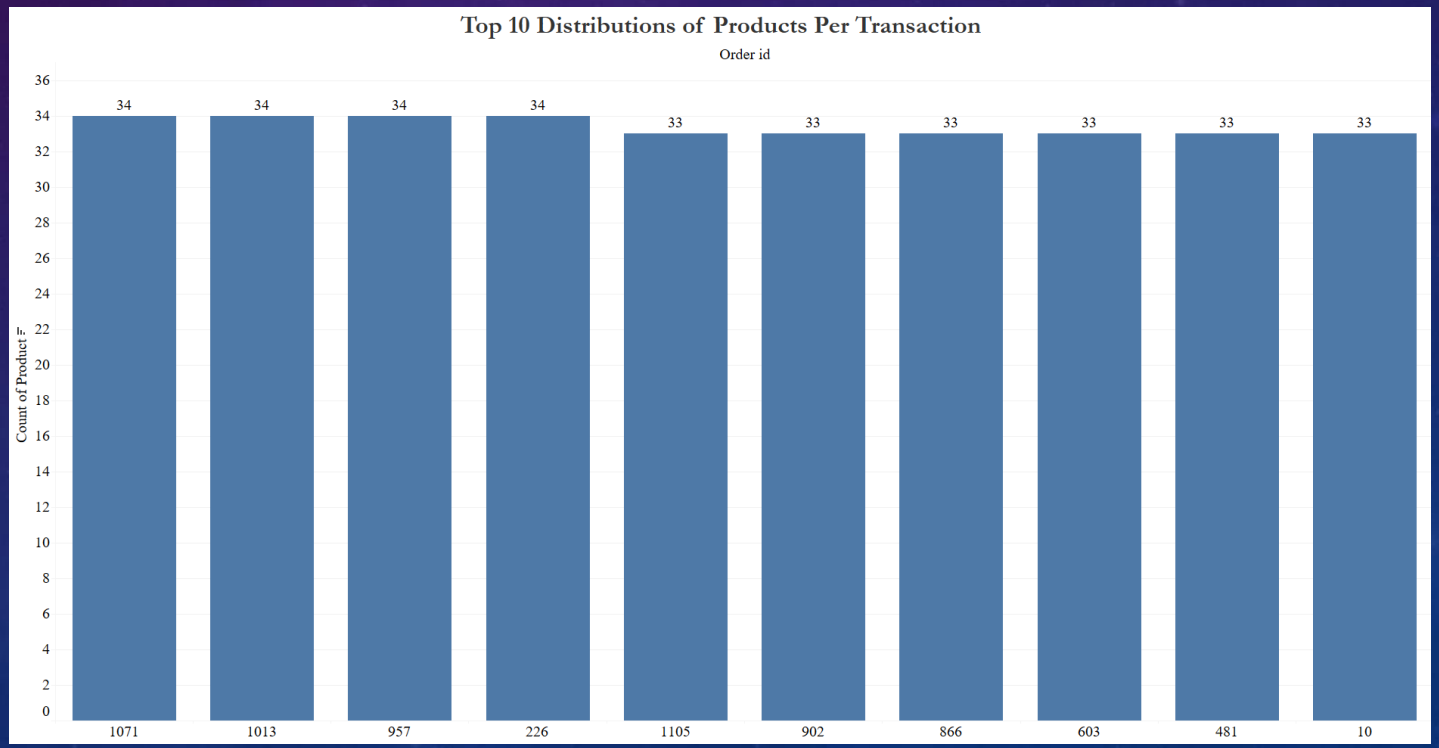
EXPLORATORY DATA ANALYSIS:

- Poultry is the Most Frequently Purchased Product: Poultry has the highest count (640), indicating that it is the most popular item among customers. This suggests that the grocery store should ensure a consistent supply of poultry and consider related promotions or product placements.
- Soda, Cereals, and Ice Cream are Staple Items: Soda (597), Cereals (591), and Ice Cream (579) consistently appear as popular products. Their prominent ranking emphasizes that these goods are staples for many households. They can be strategically positioned in high-traffic areas to maximize sales and profit margins.
- Slightly Declining but Consistent Popularity: While there's a slight decline in frequency from Poultry down to Hand Soap (502), the consistent popularity of all these products suggests a core set of grocery items that customers frequently purchase. There are many product categories that should be placed closer to each other for optimal customer experiences.



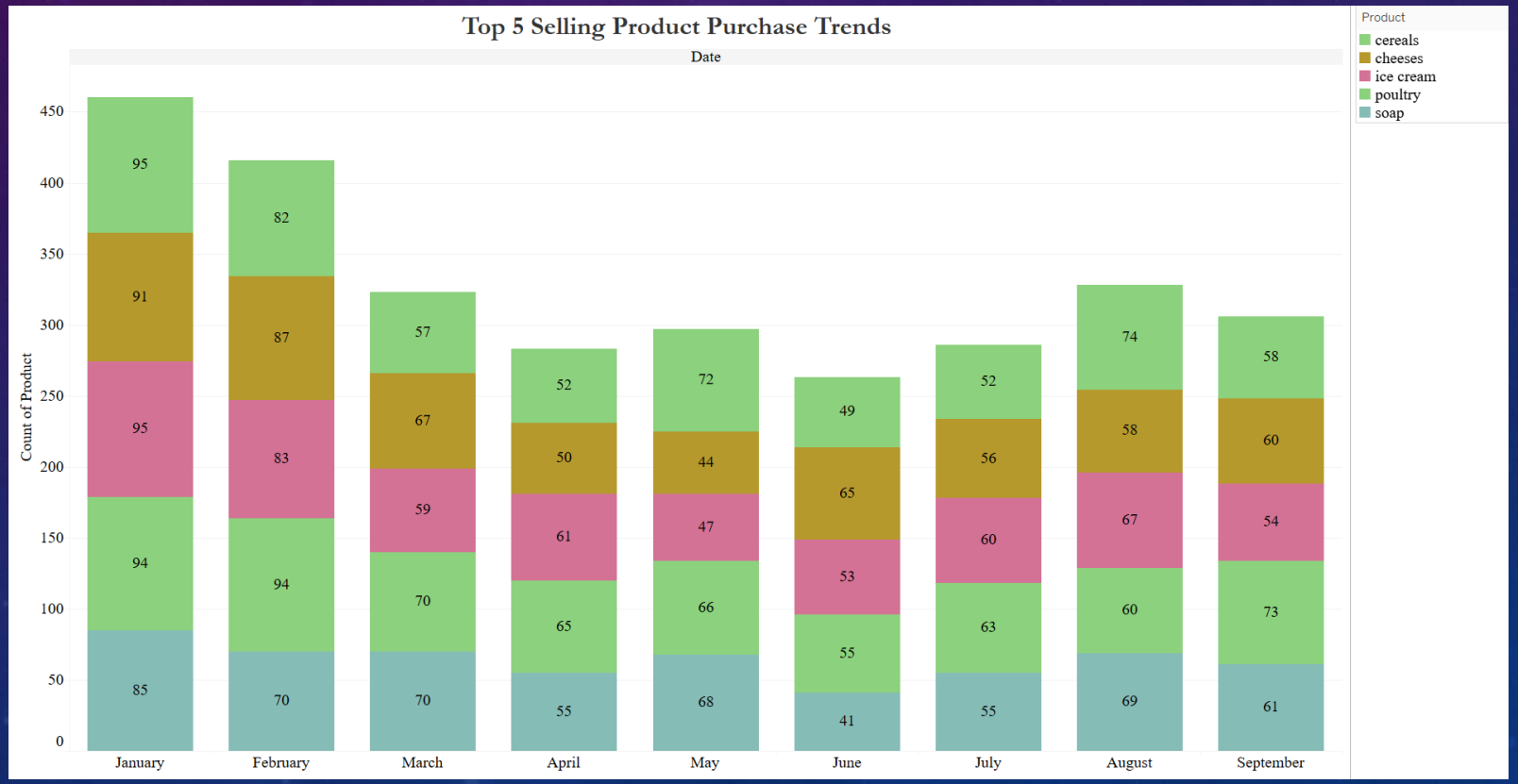
EXPLORATORY DATA ANALYSIS:

- Most Transactions Include a Consistent Number of Products: Transactions tend to have a consistent number of products with order ids 1071, 1013, 957, 226 having 34 products, while order ids 1105, 902, 866, 603, 481, and 10 all have 33 products per transaction. There is only one product difference among all the transactions.
- Opportunity to Investigate and Optimize Smaller Basket Sizes: There is a marginal gap between order ids 1071, 1013, 957, and 226 and order ids 1105, 902, 866, 603, 481, and 10, there should be more efforts focused on investigating the additional products that can be included in the 33 product baskets.
- Very little deviation in purchase behavior: Given that almost all transactions have the same count of items, there is very little deviation in the basket sizes of all transactions. The grocery store seems to have a standard set of products purchased across all customers.



EXPLORATORY DATA ANALYSIS:

- Poultry and Cereals Have Consistent High Demand: Poultry and Cereals consistently rank high in sales volume throughout the months. This confirms that they have strong demand and are regularly purchased by customers.
- Ice Cream Sales are Seasonal, Peaking in Summer: Ice Cream sales peak during the summer months (June, July, August), indicating a clear seasonal trend. Understanding the underlying drivers of this peak is essential for inventory management and promotional planning.
- Soaps & Detergents are Relatively Stable: Sales of soaps and detergents demonstrate less seasonal variation compared to other product categories. This highlights their essential nature as regular household items. Inventory levels should be relatively consistent throughout the year.



INFERENCES OF EXPLORATORY DATA ANALYSIS:

- Strong Customer Engagement with High Transaction Volume: The substantial transaction volume of 20,641 indicates robust customer engagement, providing a solid foundation for Market Basket Analysis to identify purchasing patterns and enhance marketing strategies.

- Impact of Limited Product Variety on Marketing Strategies: With only 37 unique products available, the scope for diverse bundling strategies is constrained. Focusing on frequently purchased items will be crucial for optimizing promotions and inventory management.

- Significant Sales Decline in 2020 Requires Attention: The dramatic drop in sales from 2019 to 2020 highlights the need for a thorough investigation into potential external factors affecting business performance, which could inform future risk mitigation strategies.

- Seasonal Trends and Core Products Drive Purchasing Behavior: Identifying seasonal trends, such as increased ice cream sales during summer, along with the consistent demand for staple items like poultry and cereals, can guide promotional planning and inventory management to align with customer preferences.

- Opportunities for Targeted Marketing and Inventory Optimization: The consistent basket sizes and core product demand suggest opportunities for targeted marketing strategies and optimized inventory management, ensuring essential products are readily available while enhancing customer satisfaction and loyalty.

MBA ANALYSIS:

Process of Using Association Rules

Market Basket Analysis (MBA) employs association rule mining to uncover relationships between products frequently purchased together. In this project, we utilized the Association Rule Learner to analyze transactional data from the grocery store. The analysis helps identify patterns in customer buying behavior, allowing for the development of targeted marketing strategies and optimized product placements.

MBA ANALYSIS:

KNIME Workflow

The KNIME workflow for conducting Market Basket Analysis involved the following steps:

1. Data Input: Import the transactional data from the Point of Sale (POS) system.
2. Preprocessing: Clean and prepare the data, ensuring it is structured appropriately for analysis.
3. Frequent Itemset Generation: Apply the Association Rule Learner to identify frequent itemsets using a minimum support threshold.
4. Rule Generation: Generate association rules from the frequent itemsets, calculating support, confidence, and lift values.
5. Output Presentation: Display the results in a tabular format for easy interpretation.

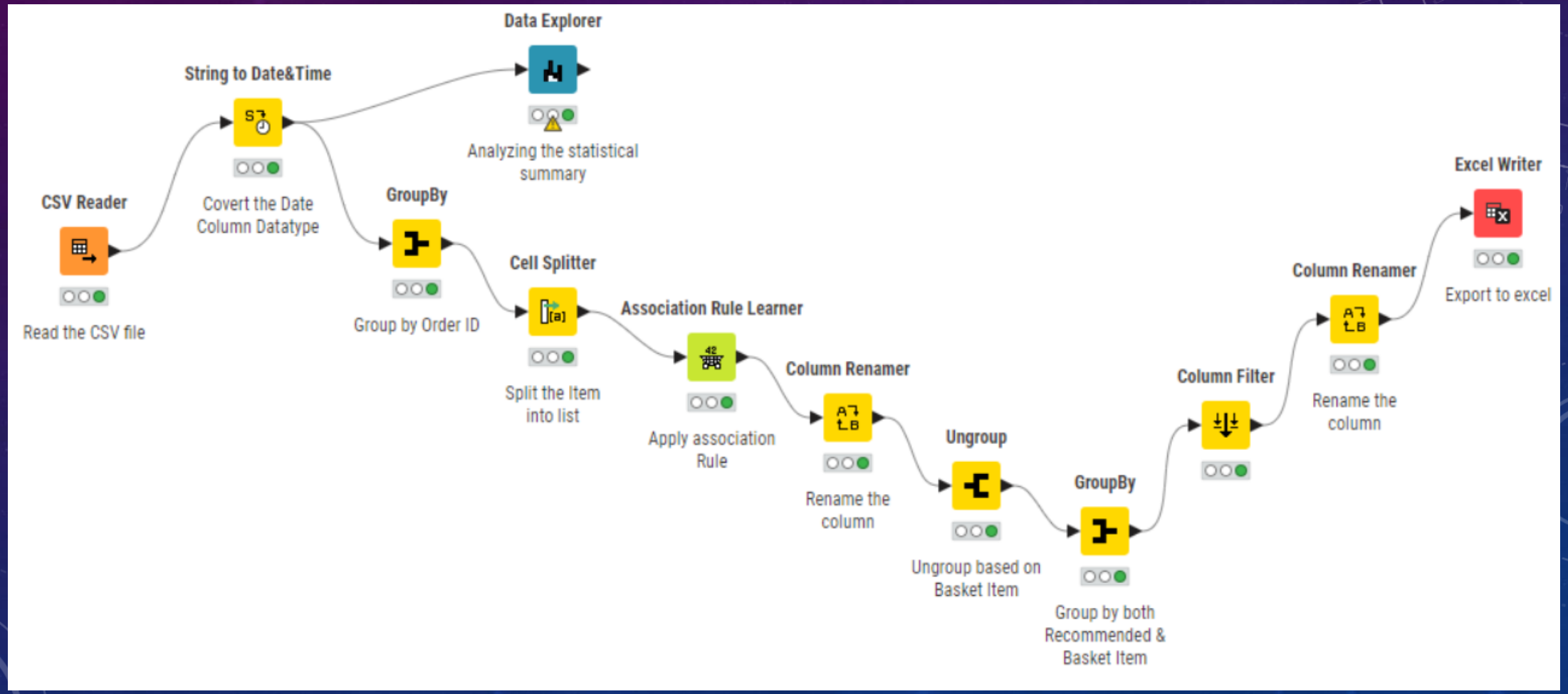
Threshold Values of Support and Confidence

For this analysis, we set the following threshold values:

- Minimum Support: 0.02 (2%) - This indicates that an itemset must appear in at least 2% of all transactions to be considered significant.
- Minimum Confidence: 0.3 (30%) - This means we are looking for rules where at least 30% of the time, when item A is purchased, item B is also purchased.

MBA ANALYSIS:

KNIME Workflow



MBA ANALYSIS:

Support, Confidence & Lift Values Explained

Support : Measures how often the itemset appears in transactions. Higher support → More frequent occurrence.

Confidence : Measures how often the consequent appears given the antecedent. Higher confidence → Stronger relationship.

Lift : Measures how much more likely items are bought together than independently. Lift > 1 → Stronger association.

Output Table Top few rows based on high lift:

Recommended_Item	Support	Confidence	Lift	implies	Basket_Item
paper towels	0.020193152	0.851851852	2.349296027	<---	eggs,dinner rolls,ice cream,pasta,lunch meat
sandwich loaves	0.020193152	0.793103448	2.269710622	<---	yogurt,hand soap,toilet paper,soap
mixes	0.020193152	0.851851852	2.266960886	<---	yogurt,dishwashing liquid/detergent,all- purpose,hand soap
paper towels	0.020193152	0.821428571	2.265392598	<---	eggs,dinner rolls,poultry,ice cream,pasta
ketchup	0.022827041	0.838709677	2.258369557	<---	tortillas,coffee/tea,juice,soap
pasta	0.021949078	0.833333333	2.243892829	<---	paper towels,dishwashing liquid/detergent,eggs,dinner rolls,ice cream
sandwich loaves	0.021949078	0.78125	2.235788317	<---	yogurt,hand soap,aluminum foil,soap
ice cream	0.020193152	0.884615385	2.21933243	<---	paper towels,eggs,dinner rolls,pasta,lunch meat
spaghetti sauce	0.021071115	0.827586207	2.217931034	<---	waffles,laundry detergent,mixes,soap
beef	0.021071115	0.827586207	2.207542599	<---	poultry,fruits,hand soap,sugar
cheeses	0.026338894	0.857142857	2.193900482	<---	paper towels,cereals,sandwich bags,sugar
paper towels	0.030728709	0.795454545	2.19375963	<---	eggs,ice cream,pasta,lunch meat
beef	0.020193152	0.821428571	2.191117431	<---	shampoo,fruits,lunch meat,pork
ketchup	0.022827041	0.8125	2.187795508	<---	toilet paper,mixes,coffee/tea,soap
paper towels	0.020193152	0.793103448	2.187275612	<---	dishwashing liquid/detergent,eggs,ice cream,pasta,lunch meat
sugar	0.022827041	0.787878788	2.183440242	<---	poultry,flour,waffles,beef
soda	0.020193152	0.851851852	2.180357886	<---	bagels,pasta,individual meals,pork
sugar	0.021949078	0.78125	2.165069951	<---	poultry,bagels,lunch meat,coffee/tea
milk	0.020193152	0.821428571	2.160755526	<---	eggs,poultry,beef,sandwich bags
paper towels	0.031606673	0.782608696	2.158332456	<---	eggs,ice cream,pasta,cereals
soap	0.023705004	0.818181818	2.15719697	<---	spaghetti sauce,all- purpose,sandwich bags,ketchup
pork	0.020193152	0.766666667	2.156131687	<---	beef,lunch meat,juice,soap
ketchup	0.021071115	0.8	2.154137116	<---	spaghetti sauce,flour,sandwich bags,soap

MBA ANALYSIS: KEY INFERENCES

Key Insights from Analysis

High-Lift Item Pairs : These represent strong purchase patterns, making them ideal for cross-selling and promotions.

Low-Lift Rules : Some associations have weak dependencies and should not be prioritized.

Frequent Item Combinations : Certain items appear together often, suggesting natural bundles.

Business Recommendations

1. Create Bundles & Offers Based on High-Lift Rules
   Example:
   Dinner rolls ,Lunch meat with a discount.
   Implement "Buy Two, Get One Free" offers on frequently co-bought items.
   Introduce complementary pairing discounts: If a customer buys Bread, offer a discount on Butter.
2. Optimize Store Layout & Online Recommendations
   Place frequently co-bought items closer together in physical stores.
   Implement "Frequently Bought Together" suggestions in online sales channels.
3. Targeted Marketing Campaigns
   Develop personalized promotions based on customer purchase history.
   Utilize email and app notifications to suggest high-confidence products.
4. Testing & Refinement
   Conduct A/B tests on different discount structures to determine effectiveness.
   Monitor customer response rates and fine-tune offers accordingly.