PROJECT B | MARKETING & RETAIL ANALYTICS

Market Basket Analysis

SHAISHAV MERCHANT

AGENDA

EXECUTIVE SUMMARY

This report analyzes transaction data using Market Basket Analysis to uncover key product associations, leading to strategic recommendations for targeted promotions, cross-sales, and combo offers to boost sales and customer satisfaction.

EXECUTIVE SUMMARY

- **Objective:** The analysis aimed to identify frequently co-purchased products using Market Basket Analysis to enhance sales strategies through targeted promotions and product bundling.
- **Data Overview:** Transactional data from the grocery store was analyzed, focusing on item sets that appear in at least 5% of transactions to ensure relevance.
- **Association Rules:** We applied thresholds of 5% for support and 60% for confidence, generating 24 actionable rules that highlight strong product associations.
- **Key Findings:** High-confidence rules indicate strong purchasing patterns, such as the frequent combination of "yogurt, toilet paper, aluminium foil, and juice," which presents an opportunity for cross-promotions.
- **Lift Values:** The lift values above 1 in the rules suggest that certain products are significantly more likely to be purchased together, justifying targeted marketing strategies.
- **Suggested Combos & Promotions:** For instance, offering a discount on "juice" when "yogurt, toilet paper, and aluminium foil" are bought together can drive sales of complementary items. Implement mix-and-match deals, like a "Buy Two Get One Free" offer for frequently associated items.
- **Cross-Selling Opportunities:** Consider cross-promotions between food items and household products, like bundling "poultry" with cleaning supplies for a holistic shopping experience.
- **Impact:** These strategies are designed to leverage the identified associations to boost sales, encourage higher spending per transaction, and improve overall customer satisfaction.
- **Next Steps:** Implement these recommendations in a phased manner, track their effectiveness, and refine the strategy based on customer response and further data analysis.

BUSINESS PROBLEM OVERVIEW & SOLUTION APPROACH

The grocery store seeks to optimize sales by identifying frequently co-purchased products through data analysis. Using Market Basket Analysis on transaction data, we identified key product associations, guiding the creation of targeted promotions, combo offers, and cross-selling opportunities to increase customer spending and improve overall revenue.

PROBLEM OVERVIEW

Context

The grocery store aims to optimize its sales by understanding customer purchasing behavior. By analyzing the transactional data from Point of Sale (POS) systems, the store seeks to identify frequently purchased item combinations, which can be used to create attractive combo offers and discounts to boost revenue.

Objective

- Analyze transactional data to understand purchasing patterns.
- Identify commonly occurring sets of items in customer orders.
- Use Market Basket Analysis to discover association rules between products.
- Propose combo offers and discounts to increase customer spending.
- Provide actionable recommendations based on data-driven insights.

SOLUTION APPROACH

To address the business problem, a structured approach will be employed, utilizing Python for an in-depth Exploratory Data Analysis (EDA) and KNIME for Market Basket Analysis. The final business report will compile all inferences, insights, and recommendations, and will be presented as a PDF.

Step-by-Step Solution Strategy:

- Data Overview & Preparation: Use Python to clean and understand the dataset.
- Exploratory Data Analysis (EDA): Identify trends, patterns, and anomalies in the data.
- Market Basket Analysis: Utilize KNIME to perform Market Basket Analysis, uncovering association rules between products.
- Threshold Setting: Define and apply appropriate support, confidence, and lift thresholds.
- Insights & Recommendations: Develop actionable insights and propose combo offers based on analysis results.
- Business Reporting: Compile findings, visualizations, and recommendations into a comprehensive business report.

DATA OVERVIEW

The Data Overview section explores transactional data to uncover patterns in customer purchases. By analyzing key metrics like product frequency and transaction trends, we identify opportunities for strategic promotions and product bundling to enhance sales and customer satisfaction.

DATA OVERVIEW

Field Name	Data Type	Description
Date	Object	The date when the transaction occurred, in the format "DD-MM-YYYY".
Order_id	Int64	A unique identifier assigned to each order.
Product	Object	The name of the product purchased in the order.

Records	Columns	Data type
20,641	3	int64(1), object(2)

Dataset Information:

Key Insights:

- **Shape of the Dataset:** The dataset contains a total of 20,641 records and 3 columns.
- **Order Details:** Dataset has information about 1,139 orders.
- **Products:** 37 different products were sold.
- **Date Range:** The dataset spans transactions from Jan 2018 to Feb 2020.
- **Missing Values:** The dataset does not contain any missing values, ensuring data completeness.
- **Duplicate Values:** There are 4,730 duplicate records in the dataset, could skew the results by inflating the frequency of certain item combinations, leading to misleading association rules. .

Conclusion:

The dataset is well-structured, but shows duplicate values, we will treat them before proceeding with Market Basket Analysis.

EXPLORATORY DATA ANALYSIS (EDA)

The Exploratory Data Analysis (EDA) of the grocery store's transactional dataset provides a comprehensive understanding of customer behavior, product popularity, and purchasing trends over time. The dataset, after cleaning and removing 4,730 duplicate records, was thoroughly analyzed to extract meaningful insights.

EDA INSIGHTS & KEY FINDINGS

Transaction Trends:

- **Yearly/Month Growth:** Significant increase in unique orders from 2018 to 2020, with peaks in Mid and End year, indicating strong seasonal shopping.

Product Popularity:

- **Top Products:** "Poultry," "ice cream," and "cereals" dominate sales, important as staple items.
- **Consistent Demand:** Top products maintain steady sales throughout the year, crucial for inventory and marketing focus.

Order Size Distribution:

- **Moderate Purchases:** Most orders contain 10-18 products, with an average of 14 items per order, reflecting typical shopping habits.

Customer Shopping Patterns:

- **Weekend Preference:** Higher transactions on weekends, with a dip on Mondays, suggesting weekend shopping is preferred.
- **Seasonal Peaks:** Increased sales during holiday seasons, pointing to strong seasonal demand.

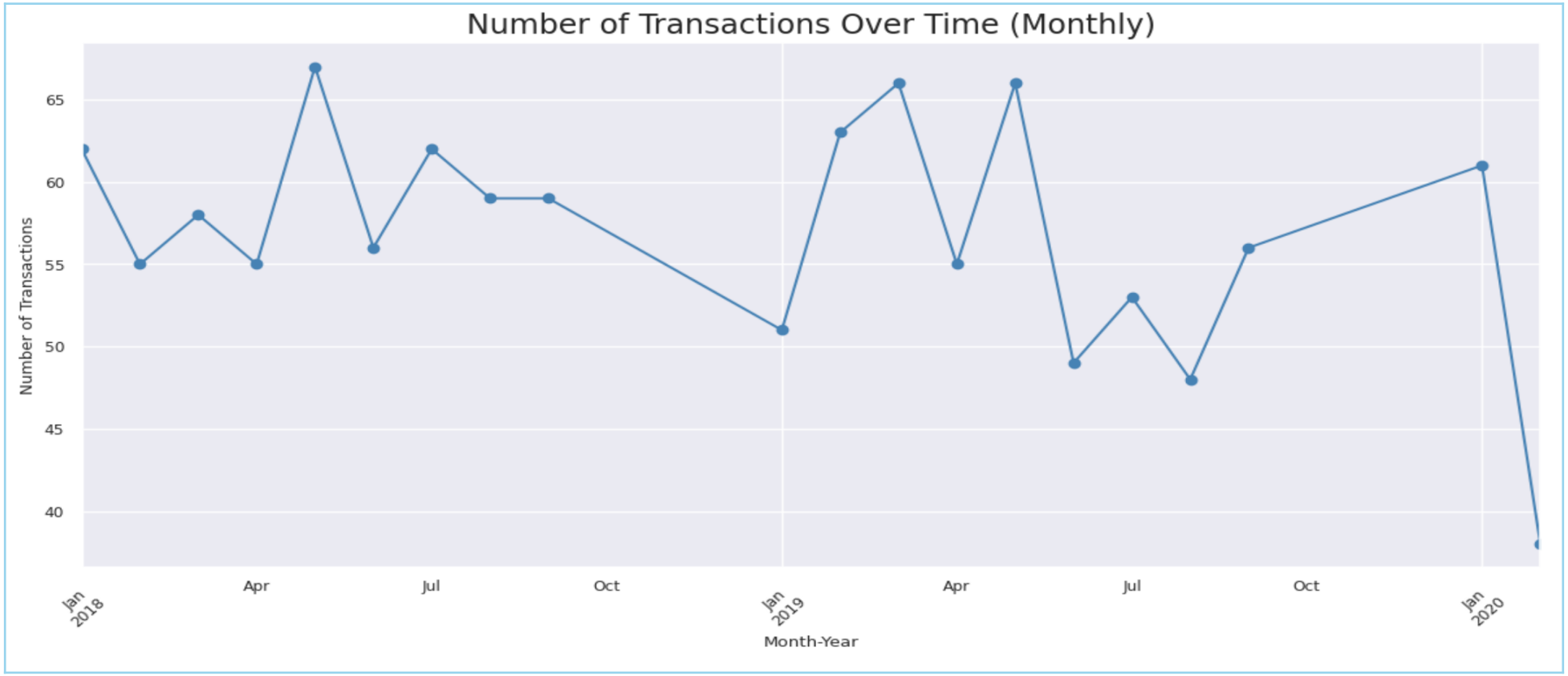
Duplicate Removal:

The dataset initially contained 4,730 duplicate records, which were removed to ensure the accuracy of the analysis.

Conclusion:

The analysis highlights consistent demand for key products and clear seasonal shopping patterns. These insights can guide targeted promotions, efficient inventory management, and strategic planning to enhance store performance.

UNIVARIATE ANALYSIS – ORDER DATE

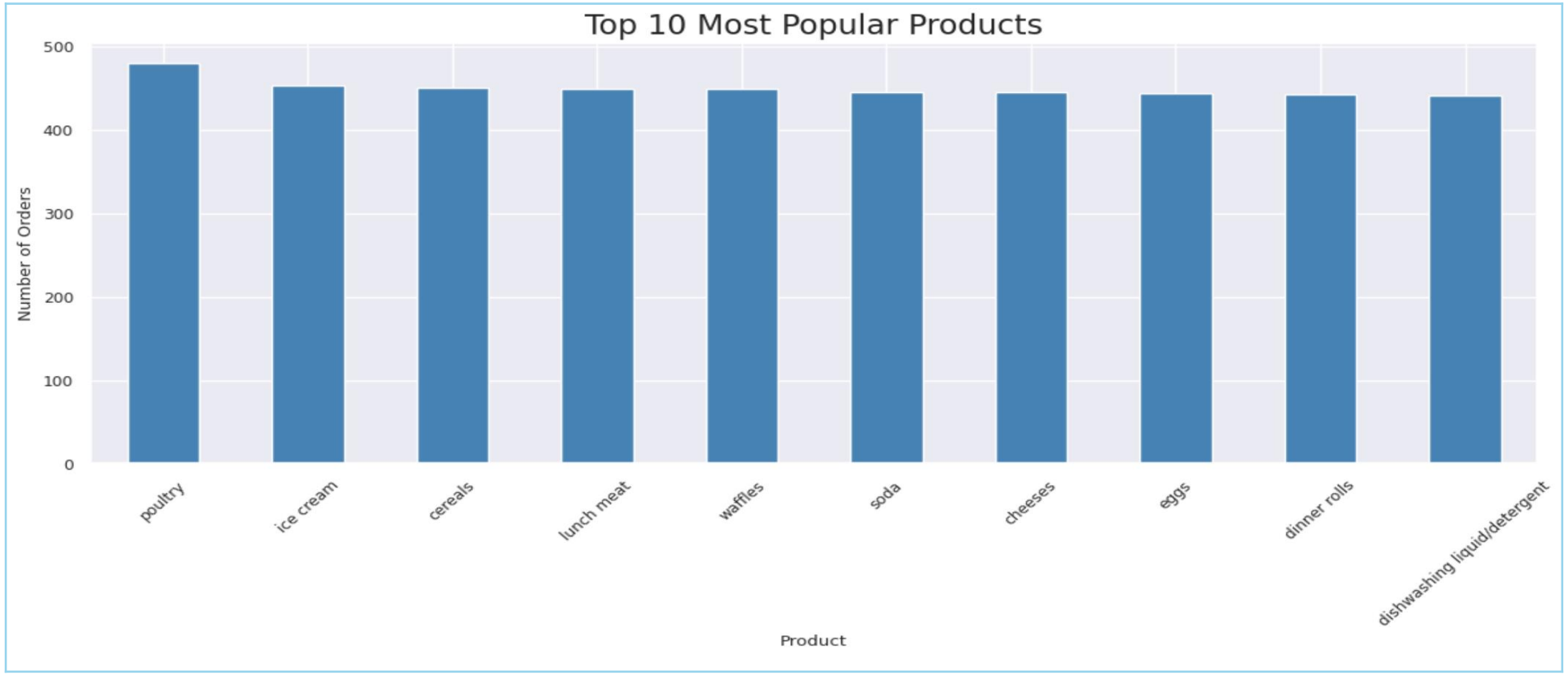


Order Date Analysis

Key Insights:

- The number of transactions shows a steady pattern with notable spikes during certain months, which could be indicative of seasonal promotions or holidays.
- There is a visible peak in transactions during May month, suggesting a potential increase in shopping activity during the period.
- The steady decline in transactions toward the end of the dataset might indicate either a data collection issue or a natural decrease in shopping activity.

UNIVARIATE ANALYSIS – PRODUCT

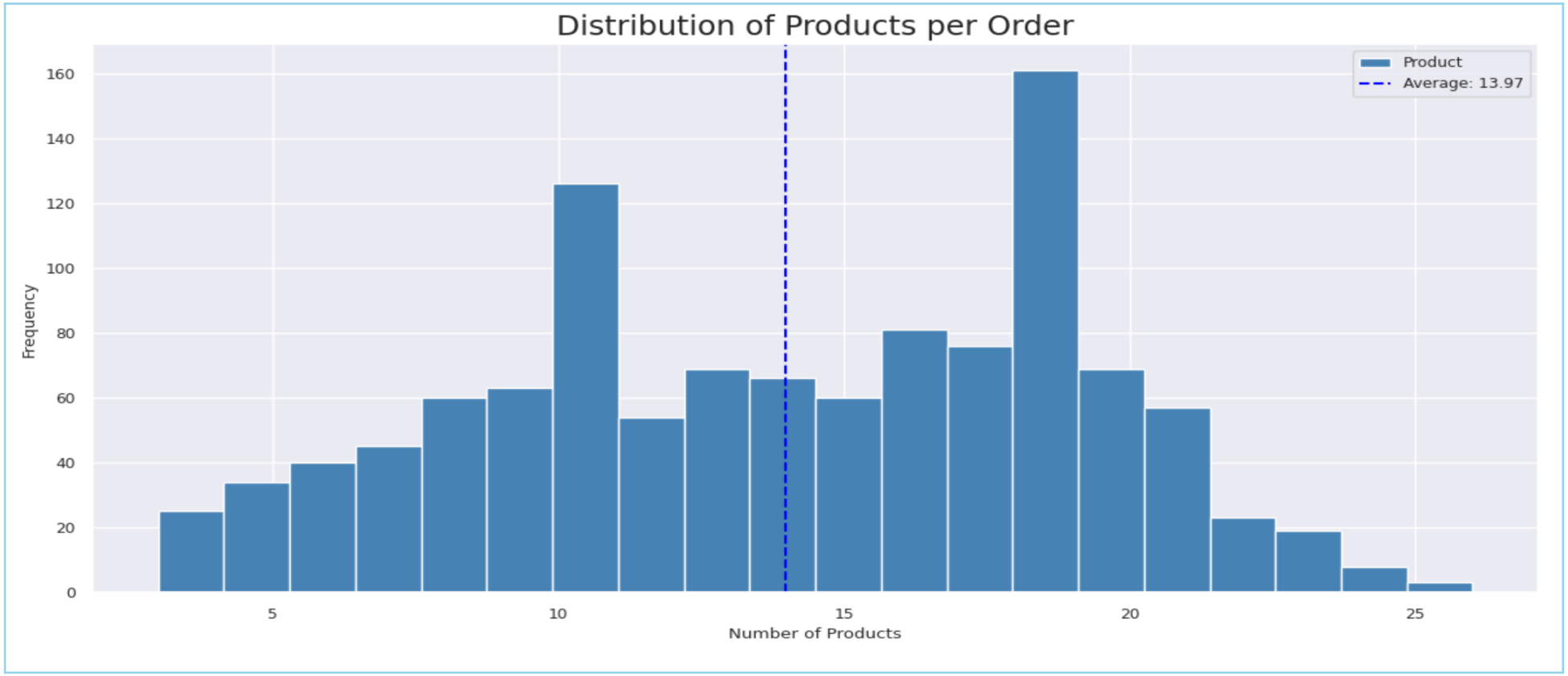


Product Analysis

Key Insights:

- Poultry is the most frequently purchased product, highlighting its importance as a staple item.
- The top 10 products account for a significant portion of sales, emphasizing the concentration of demand on a few key items.

UNIVARIATE ANALYSIS – ORDER SIZE

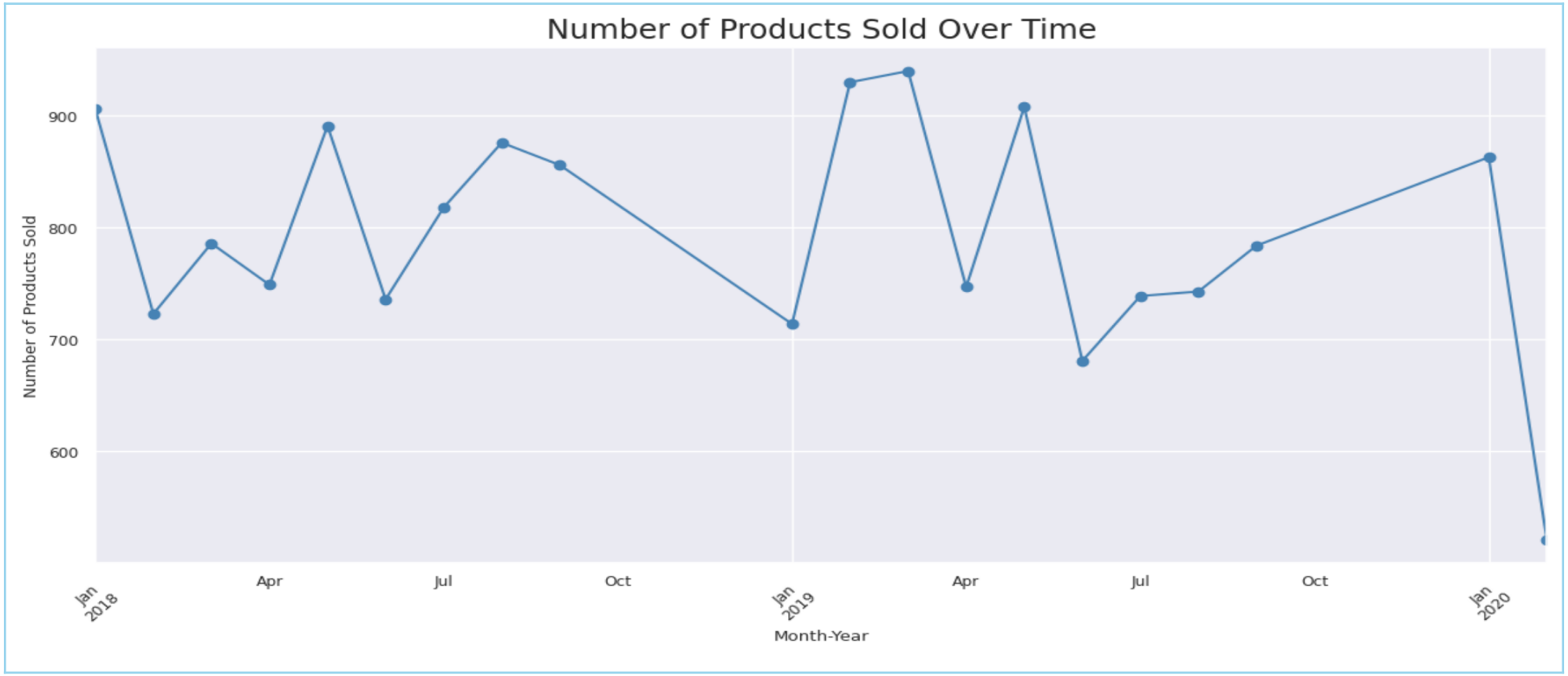


Order Size Distribution

Key Insights:

- The average order contains about 14 products, indicating customers typically make moderate-sized purchases.
- Most orders include 10 to 18 products, suggesting a balanced mix of shopping patterns.

BIVARIATE ANALYSIS – DATE VS PRODUCT

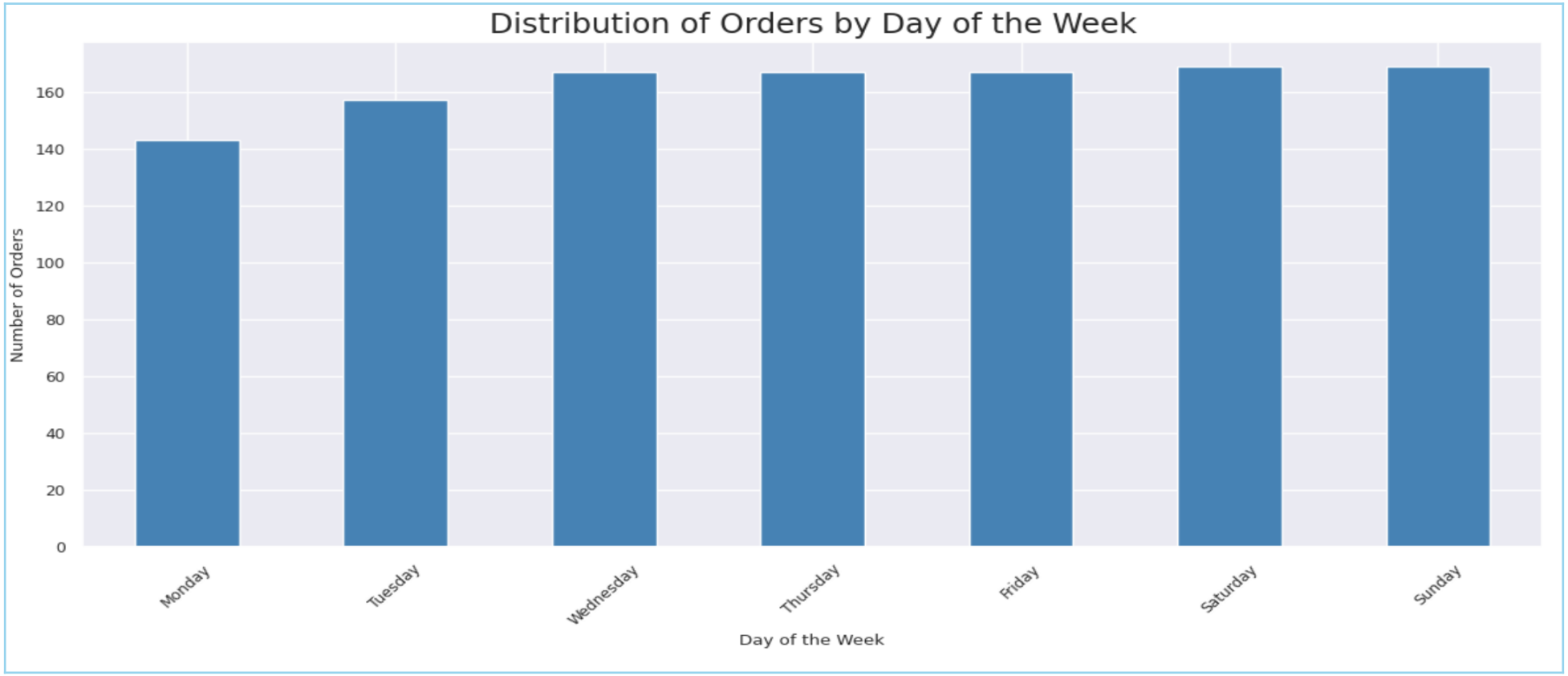


Date vs. Number of Products Sold:

Key Insights:

- Product sales align with transaction trends, peaking during certain months, likely due to promotions or holidays.
- Consistent sales with periodic spikes indicate regular purchasing behavior with occasional influences from external events.

BIVARIATE ANALYSIS – WEEKDAYS VS ORDER

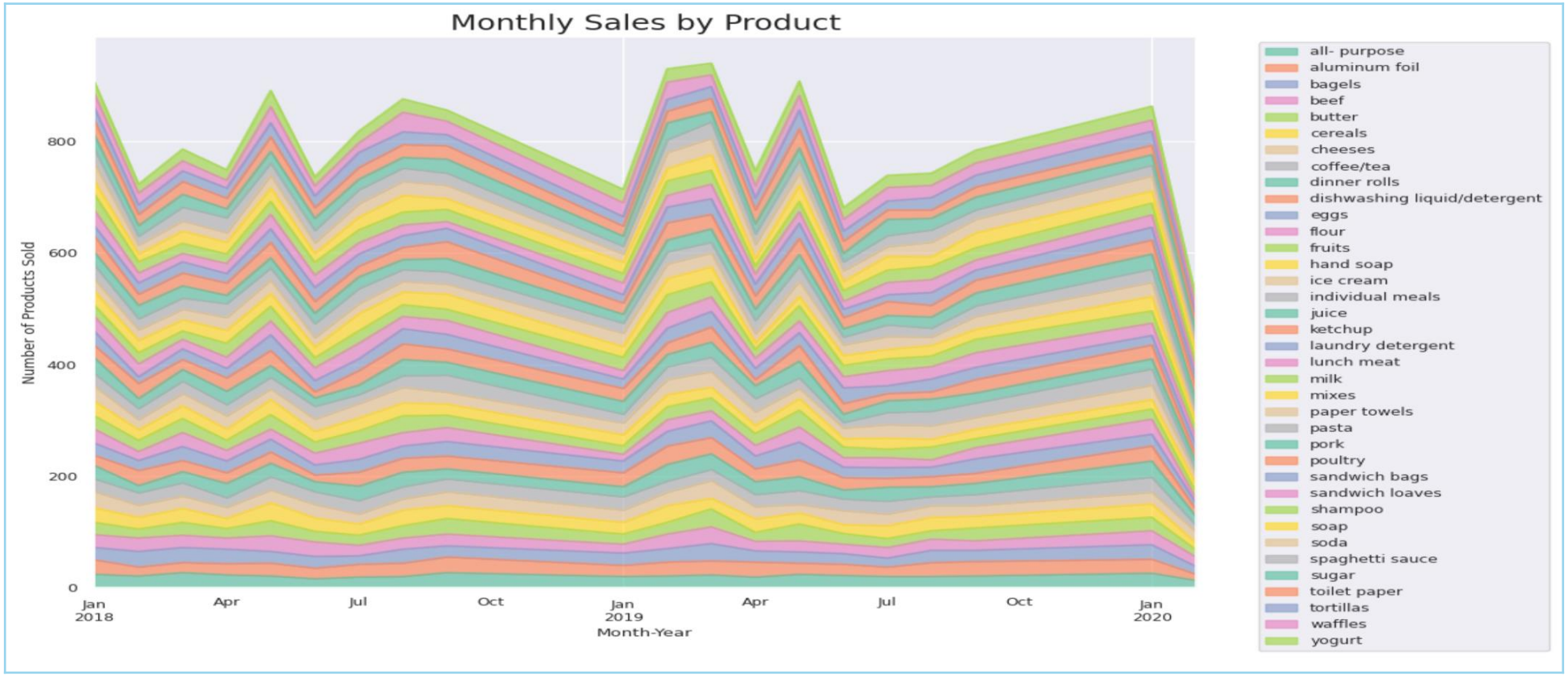


Day of the Week vs. Number of Transactions

Key Insights:

- Transactions peak on weekends, particularly Saturdays, reflecting a preference for shopping during weekends.
- Mondays have the lowest transaction volume, possibly due to customers stocking up over the weekend.

BIVARIATE ANALYSIS – FREQUENCY

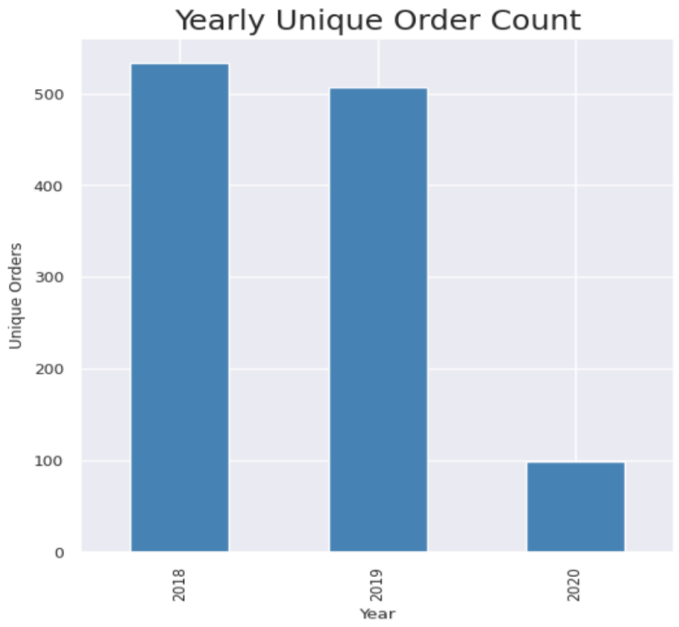


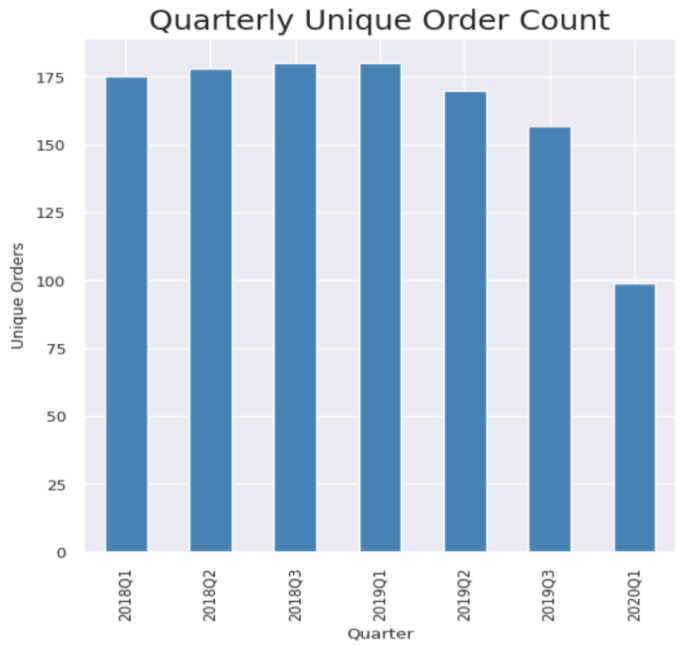
Product vs. Frequency of Purchase

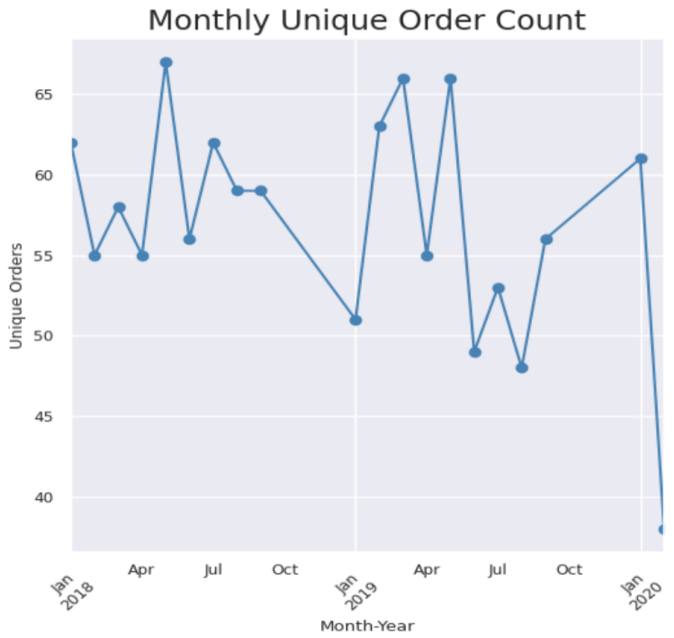
Key Insights:

- Poultry and other top products maintain high sales across all months, indicating stable demand.
- Some products exhibit seasonal variability, suggesting targeted opportunities for promotions.

TIME-SERIES ANALYSIS







Time-Series Analysis with Yearly, Quarterly and Monthly Trends

Yearly Trends:

- There is a significant increase in unique orders from 2018 to 2020, indicating growing customer engagement or effective business expansion strategies.

Quarterly Trends:

- Unique orders peak during the last quarter each year, particularly around the holiday season, suggesting strong seasonal shopping behavior. Mid-year quarters (Q2 and Q3) generally see lower order volumes, possibly reflecting a dip in customer activity during these periods.

Monthly Trends:

- Monthly order counts fluctuate, with clear peaks in November and December, aligning with holiday shopping trends.
- An overall upward trend in monthly orders indicates increasing customer participation over time, possibly due to successful marketing or increased brand loyalty.

MARKET BASKET ANALYSIS

Market Basket Analysis identifies relationships between products by analyzing transaction data. It uncovers patterns of items frequently bought together, helping businesses create targeted promotions, optimize store layouts, and improve product bundling, ultimately enhancing sales and customer satisfaction.

MARKET BASKET ANALYSIS (MBA)

About MBA, Association Rules and Parameters

MBA:
It helps identify which products are often bought together. This insight allows businesses to create combo offers, optimize store layouts, and increase sales by promoting related items.

Association Rules:
Association rules reveal the relationship between products in transactions. For this project, these rules help identify product pairs or groups that frequently appear together, guiding promotional strategies and product placement.

Support and Confidence – Threshold Values:
Support measures how often a product pair appears in transactions, Confidence shows the likelihood of buying one item when another is purchased. Higher thresholds means stronger, more relevant rules.

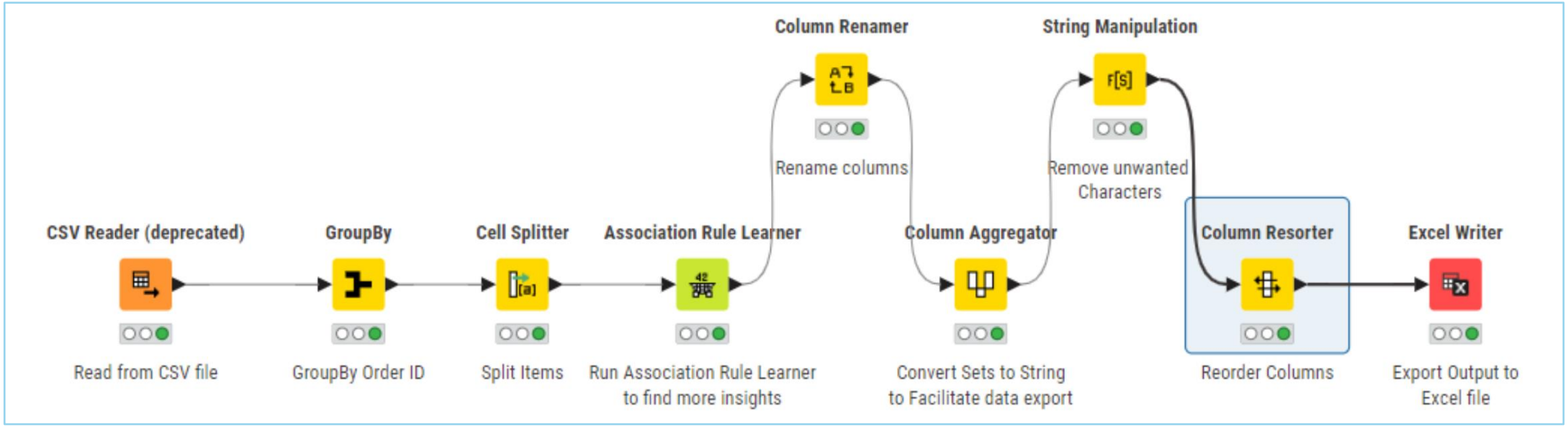
Evaluating Associations

Support:
- Frequency of the itemset in transactions.
- Ensures the rule is relevant and applies to a significant portion of transactions.

Confidence:
- Likelihood of buying the consequent when the antecedent is purchased.
- Indicates the strength and reliability of the association.

Lift:
- Measures the strength of the association beyond chance.
- Highlights the most impactful and significant rules for targeted marketing.

MBA – KNIME WORKFLOW & THRESHOLD



Threshold Values for Association Rule

Support = 5%	Itemset Max Length = 5	Confidence = 60%
• Focuses on Item-sets that appear in at least 5% of transactions. • Ensures the rules are relevant to a significant portion of customers, targeting commonly bought-together items.	• Limits the number of items in each rule to 5. • Captures complex associations without overwhelming the analysis, balancing depth and manageability.	• Requires that 60% of transactions with the antecedent also include the consequent. • Ensures that the rules are strong and reliable, focusing on associations that are likely to predict customer behavior.

Impact on Associations: These thresholds balance relevance, complexity, and reliability, helping to identify actionable rules that reflect significant customer purchasing patterns, thereby improving targeted promotions and product bundling strategies.

MBA – OUTPUT & EXPLANATION

Recommended	implies	Basket Items	Support	Confidence	Lift
paper towels	<---	eggs, ice cream, pasta	0.0553	0.6495	1.7912
pasta	<---	paper towels, eggs, ice cream	0.0553	0.6429	1.7310
cheeses	<---	bagels, cereals, sandwich bags	0.0509	0.6744	1.7262
juice	<---	yogurt, toilet paper, aluminum foil	0.0500	0.6404	1.7004
mixes	<---	yogurt, poultry, aluminum foil	0.0509	0.6304	1.6777
sandwich bags	<---	cheeses, bagels, cereals	0.0509	0.6105	1.6596
dinner rolls	<---	spaghetti sauce, poultry, laundry detergent	0.0536	0.6421	1.6509
dinner rolls	<---	spaghetti sauce, poultry, ice cream	0.0518	0.6413	1.6489
juice	<---	yogurt, poultry, aluminum foil	0.0500	0.6196	1.6450
poultry	<---	dinner rolls, spaghetti sauce, ice cream	0.0518	0.6860	1.6279
eggs	<---	paper towels, dinner rolls, pasta	0.0518	0.6344	1.6275
pasta	<---	paper towels, eggs, dinner rolls	0.0518	0.6020	1.6211
dinner rolls	<---	spaghetti sauce, poultry, cereals	0.0509	0.6304	1.6209
eggs	<---	paper towels, ice cream, pasta	0.0553	0.6300	1.6161
coffee/tea	<---	yogurt, cheeses, cereals	0.0500	0.6129	1.6160
dinner rolls	<---	spaghetti sauce, poultry, juice	0.0518	0.6277	1.6138
eggs	<---	dinner rolls, poultry, soda	0.0518	0.6277	1.6101
milk	<---	poultry, laundry detergent, cereals	0.0509	0.6042	1.5893
ice cream	<---	paper towels, eggs, pasta	0.0553	0.6238	1.5649
cereals	<---	cheeses, bagels, sandwich bags	0.0509	0.6170	1.5583
poultry	<---	dinner rolls, spaghetti sauce, laundry detergent	0.0536	0.6559	1.5564
poultry	<---	dinner rolls, spaghetti sauce, cereals	0.0509	0.6374	1.5124
poultry	<---	dinner rolls, spaghetti sauce, juice	0.0518	0.6020	1.4286
poultry	<---	dishwashing liquid/detergent, laundry detergent, mixes	0.0500	0.6000	1.4238

Explanation of Support, Confidence, and Lift:

Support: (Frequency of a combination of an Item)

- A Support value of 0.05 means the itemset appears in approximately 5% of all transactions.
- Higher Support values suggest the itemset is more common in the dataset.
- Support is used to filter out less frequent item combinations that might not be useful for analysis.

Confidence: (Purchasing item with antecedent item)

- A Confidence of 0.64 means that 64% of the time, when "yogurt, toilet paper, aluminum foil" are bought, "juice" is also purchased.
- Higher Confidence values suggest a stronger association between the items in the rule.
- Confidence helps identify which items are often bought together, guiding promotional and marketing strategies.

Lift: (Measures the strength of the association)

- A Lift value greater than 1 indicates a positive correlation between the items, meaning they are more likely to be bought together than separately.
- For example, a Lift of 1.70 suggests that "juice" is 70% more likely to be purchased when "yogurt, toilet paper, aluminum foil" are in the basket, compared to these items being purchased independently.
- Higher Lift values imply stronger associations between items, making them important for cross-selling opportunities.

SUGGESTIONS & RECOMMENDATIONS

Insights through Market Basket Analysis enables in crafting targeted promotions, smart cross-promotions, and compelling discount offers like Buy-Two-Get-One Free, mix-and-match deals, and discount combos. These strategies are designed to enhance customer experience, boost sales, and maximize value for both shoppers and the business.

MBA – OFFERS, RECOMMENDATIONS AND DISCOUNTS

Possible Combos with Lucrative Offers	Recommendations	Discount Offers or Combos
Yogurt, Toilet Paper, Aluminum Foil, and Juice: • This combo has a good Support and a high Confidence level, making it an ideal candidate for a combo offer. • Consider offering a discount on **juice** if customers buy **yogurt**, **toilet paper**, and **aluminum foil** together.	**Targeted Promotions:** • Use the identified item sets to create targeted promotions. • For example, a discount on **coffee/tea** when customers purchase **cereals** and **cheeses** together to cater to customers who enjoy a complete breakfast set.	**Buy-Two-Get-One Free:** • For items like **yogurt** and **poultry** where multiple associations are found, consider a "Buy-Two-Get-One Free" offer. • This offer can be rotated weekly to maintain customer interest and boost repeat purchases.
Yogurt, Poultry, Aluminum Foil: • This combination can be bundled with a discount on **aluminum foil** for an added incentive. • Additionally, offer a free small pack of **spices** or **mixes** that complement **poultry** dishes.	**Cross-Promotions:** • Given the association of **poultry** with cleaning products (e.g., **dishwashing liquid**), consider offering bundle discounts that combine these products, promoting both food and cleaning supplies. • Additionally, create a meal prep kit that includes **poultry** along with related items like **spices**, **marinades**, and a small bottle of **dishwashing liquid** as a bonus.	**Discount Combos**: • Offer a 10-15% discount when customers purchase **yogurt**, **toilet paper**, and **aluminium foil** together with **juice**, encouraging customers to buy a complete set rather than individual items. **Mix-and-Match Deals**: • Create a mix-and-match offer for customers to choose any three items from the associated item sets (e.g., **yogurt**, **poultry**, **aluminium foil**, **juice**, **mixes**) at a flat discount.

Market Basket Analysis

THANK YOU

SHAISHAV MERCHANT

PROJECT B | MARKETING & RETAIL ANALYTICS