MARKETING & RETAIL ANALYTICS – MAIN PROJECT (PART A)

PRESENTED BY: PRIYANKA CHANDRAHAR MANE

PROGRAM: PGP IN DATA SCIENCE AND BUSINESS ANALYTICS

INSTITUTE: GREAT LEARNING

THEME: AUTOMOBILE PARTS COMPANY

TOOLS: PYTHON (EDA), KNIME (RFM)

BUSINESS CONTEXT

- An automobile parts manufacturing company has been actively selling products to a diverse range of customers for the past three years. Despite its growth, the company lacks the in-house expertise to derive actionable insights from its transaction data.

- They now aim to uncover hidden patterns and trends in customer purchases to:
- - Better understand customer behavior
- - Improve customer segmentation
- - Implement targeted marketing strategies

- These insights will help enhance customer satisfaction and drive revenue growth through personalized services.

PROJECT OBJECTIVE

The primary objective of this analysis is to apply data science techniques to:

- Identify underlying patterns in customer purchasing behavior
- **Segment customers** based on their transactional data
- Provide **actionable insights** to improve the company's marketing strategy
- Recommend personalized approaches to boost customer retention and sales

These objectives aim to help the company make informed, data-driven decisions.

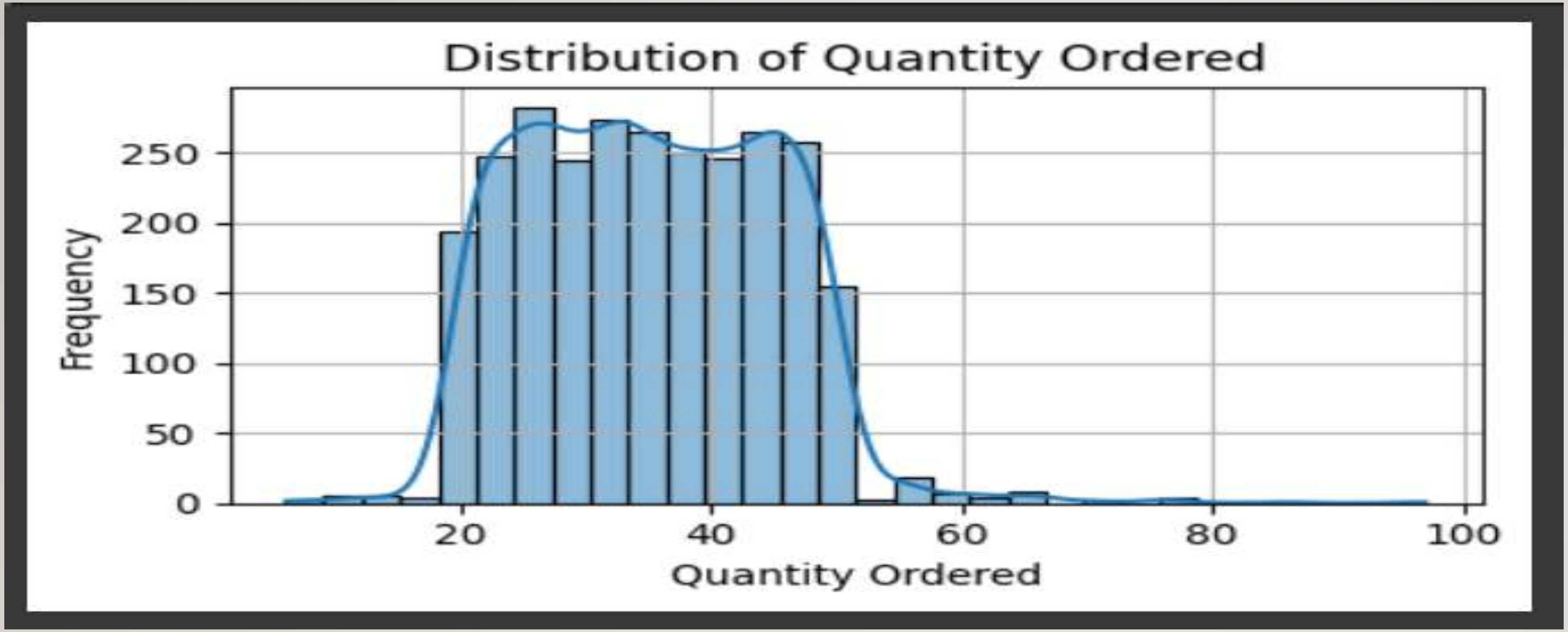
DATASET OVERVIEW – PART A

Column Name	Description
ORDERNUMBER	Unique ID for each order
QUANTITYORDERED	Number of items ordered in a transaction
PRICEEACH	Price per unit of product
ORDERLINENUMBER	Sequence of product in the order
SALES	Total sales value
ORDERDATE	Date when the order was placed
DAYS_SINCE_LASTORDER	Days since customer's previous order
STATUS	Order status (e.g., Shipped, Disputed)
PRODUCTLINE	Product category (e.g., Classic Cars, Motorcycles)
MSRP	Manufacturer's Suggested Retail Price
CUSTOMERNAME	Name of the customer
PHONE	Customer's contact number
CITY / COUNTRY	Customer's location
DEALSIZE	Size of deal (Small, Medium, Large)

SUMMARY STATISTICS – KEY VARIABLES

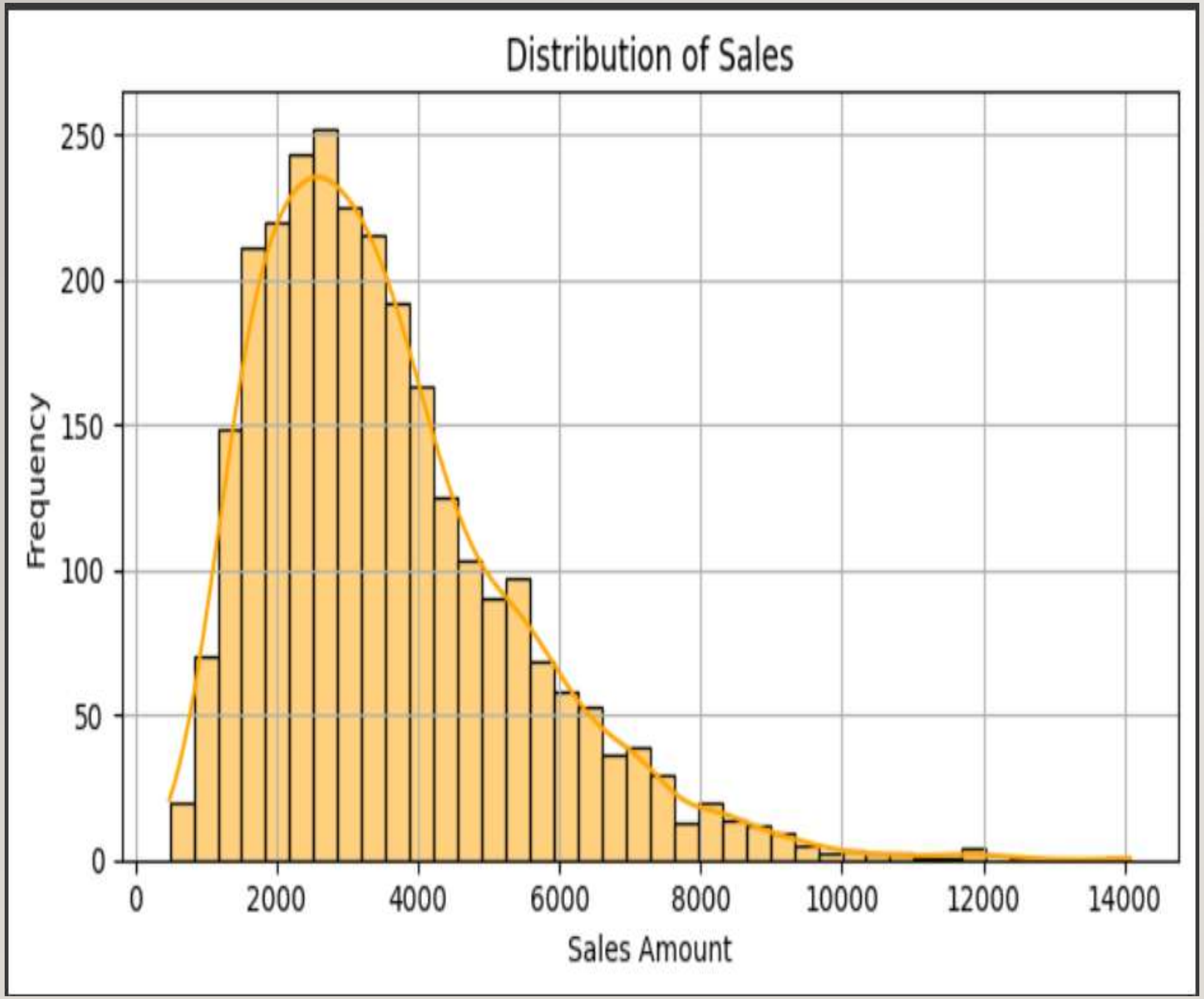
Metric	SALES	PRICE EACH	QTY ORDERED	MSRP	DAYS SINCE LAST ORDER
Mean	₹3553.05	₹101.10	35.10	₹100.69	1757.09 days
Median	₹3184.80	₹95.55	35.00	₹99.00	1761 days
Min – Max	₹482 – ₹14082	₹27 – ₹253	6 – 97	₹33 – ₹214	42 – 3562 days
Std. Dev.	₹1838.95	₹42.04	9.76	₹40.11	819.28 days

Distribution of Quantity Ordered



- Most orders fall between 20 and 50 units
- Highest frequency is around 30–40 units
- Distribution shows slight right skew with few high-quantity orders
- Indicates a consistent ordering pattern, typical of medium-sized purchases

Distribution of Sales

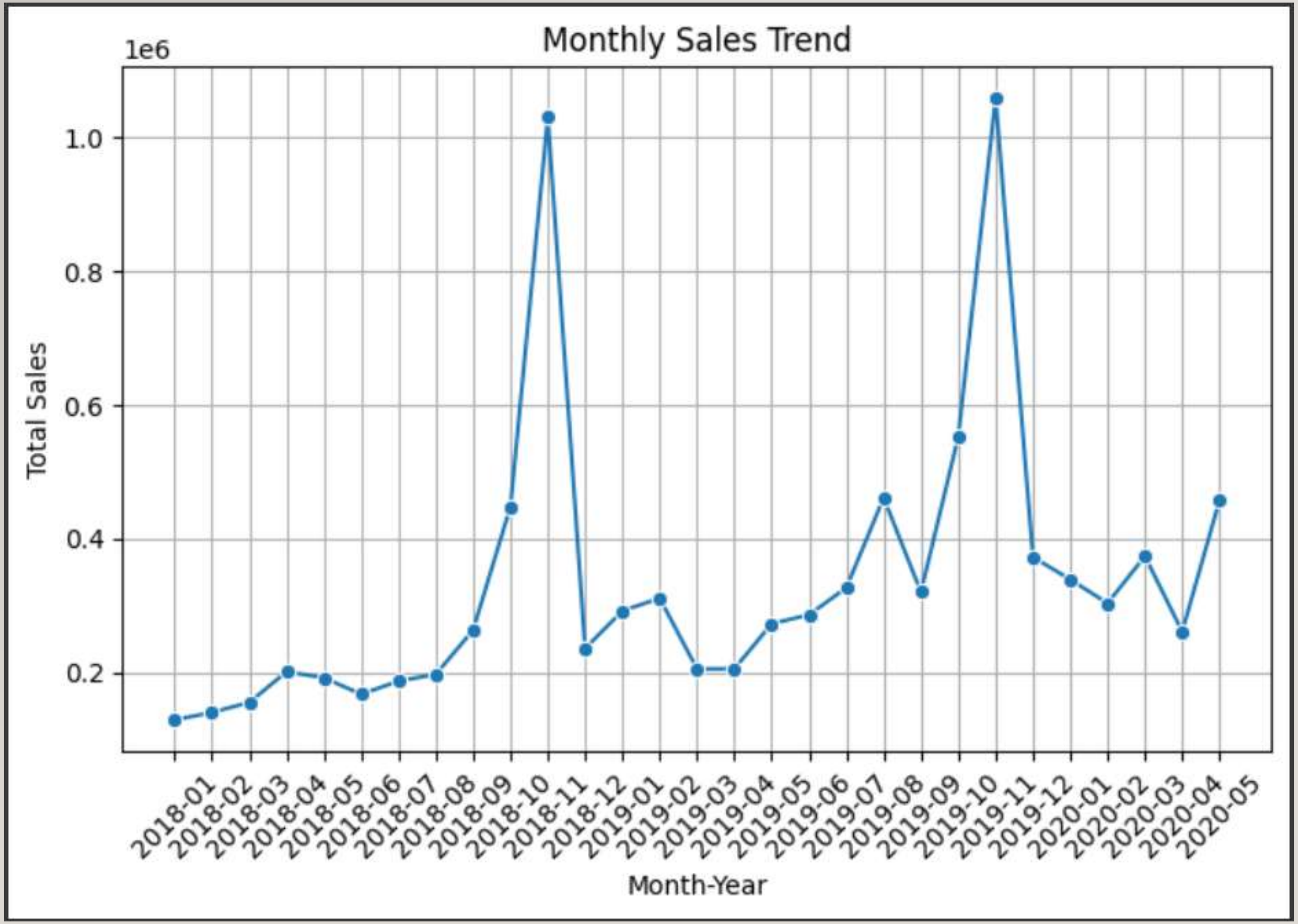


- Majority of orders generate sales between ₹1,500 and ₹5,000
- Peak frequency lies around ₹2,500–₹3,500 range
- Long right tail indicates the presence of high-value orders
- Sales distribution is right-skewed, showing a few premium purchases

Sales vs Quantity Ordered

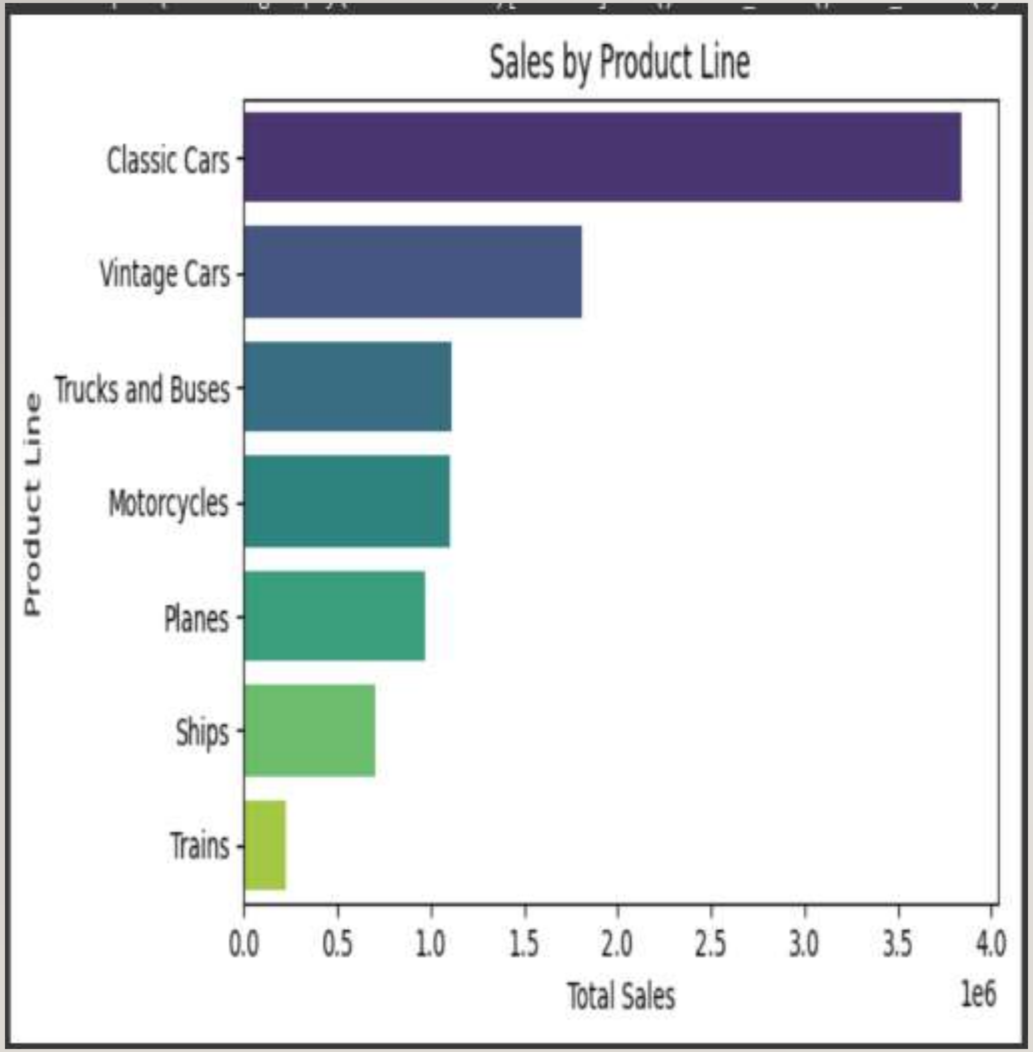
- Clear upward trend: higher quantities tend to generate higher sales
- Strong cluster around 20–50 quantity and ₹2,000–₹5,000 in sales
- A few outliers show high sales despite low quantities — likely premium items
- Confirms expected correlation between quantity and total order value

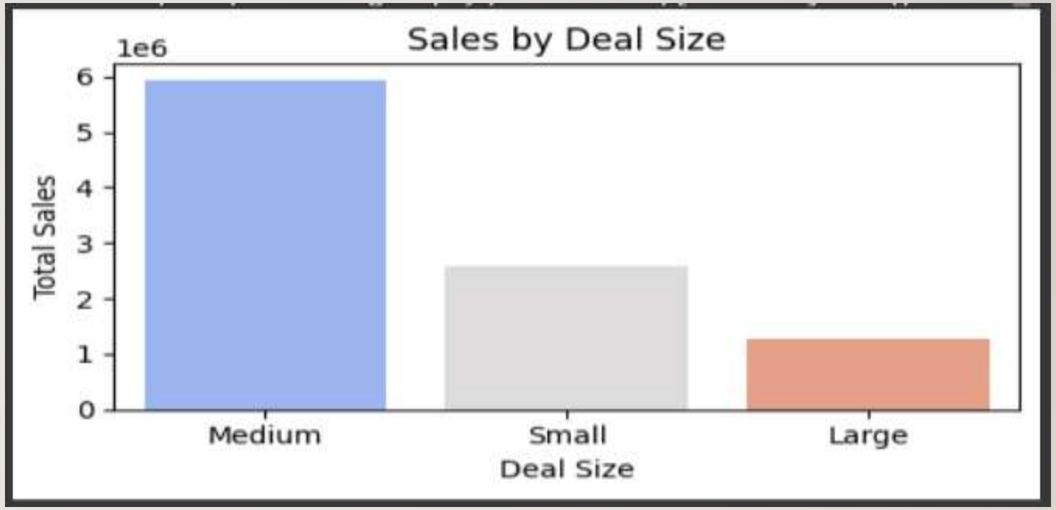
Monthly Sales Trend



- Sales show clear seasonality with two sharp spikes in late 2018 and late 2019
- Highest sales observed in November 2018 and November 2019 — likely festive or promotion periods
- Moderate sales fluctuations throughout the remaining months
- Consistent year-on-year patterns suggest strong periodic demand

Sales by Product Line and Deal Size

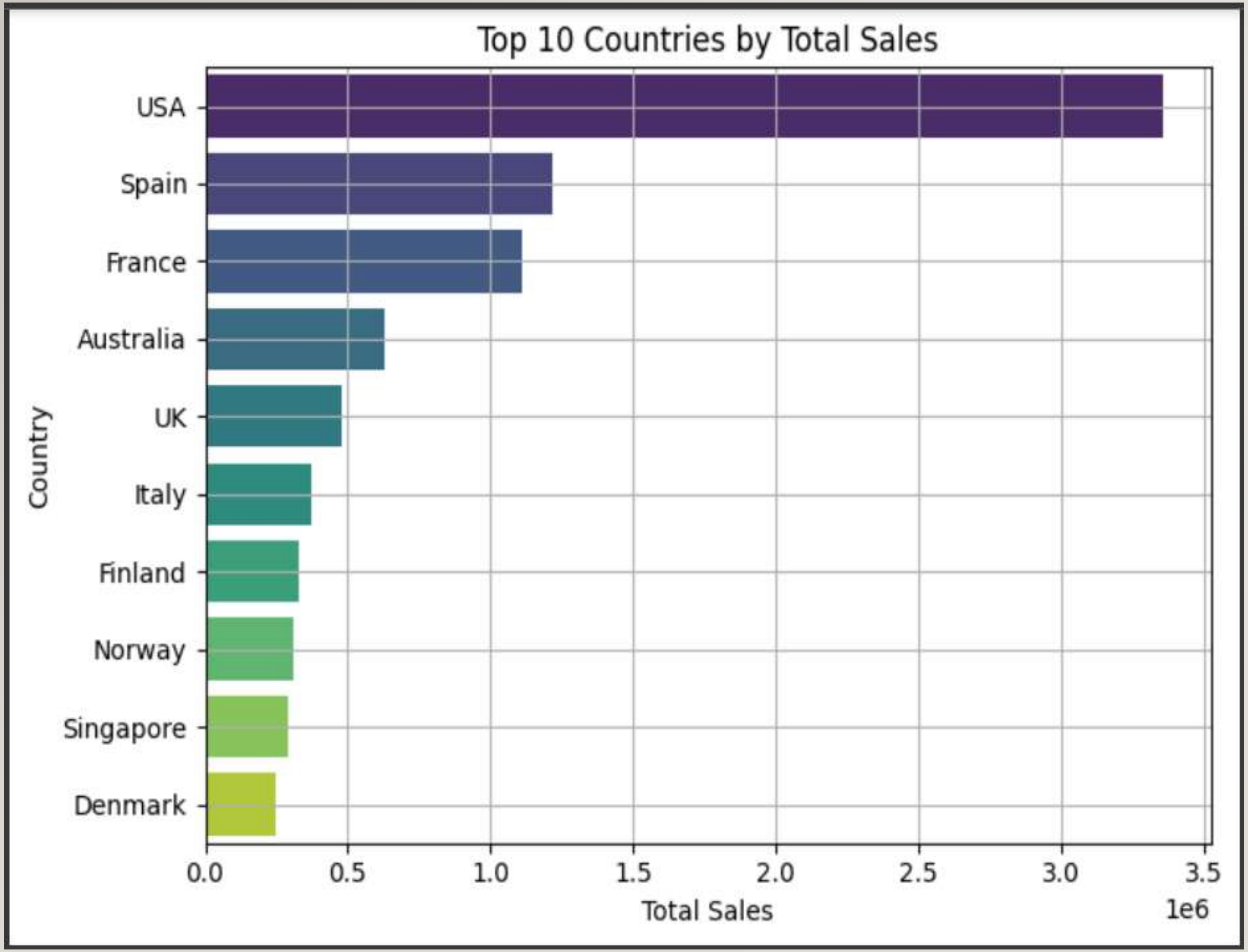




Key Insights:

- Classic Cars and Vintage Cars are the top-selling product lines by a large margin
- Motorcycles and Trucks & Buses follow as moderate contributors
- Trains and Ships contribute the least to total revenue
- Medium-sized deals account for the highest sales overall
- Small deals surprisingly outperform large deals in total sales value
- Indicates frequent mid-range purchases driving business volume

Sales by Country



Bullet Points (Insights):

•US **USA** is the dominant market with significantly higher sales than other countries

•ES **Spain** and FR **France** follow distantly, contributing notable but smaller shares

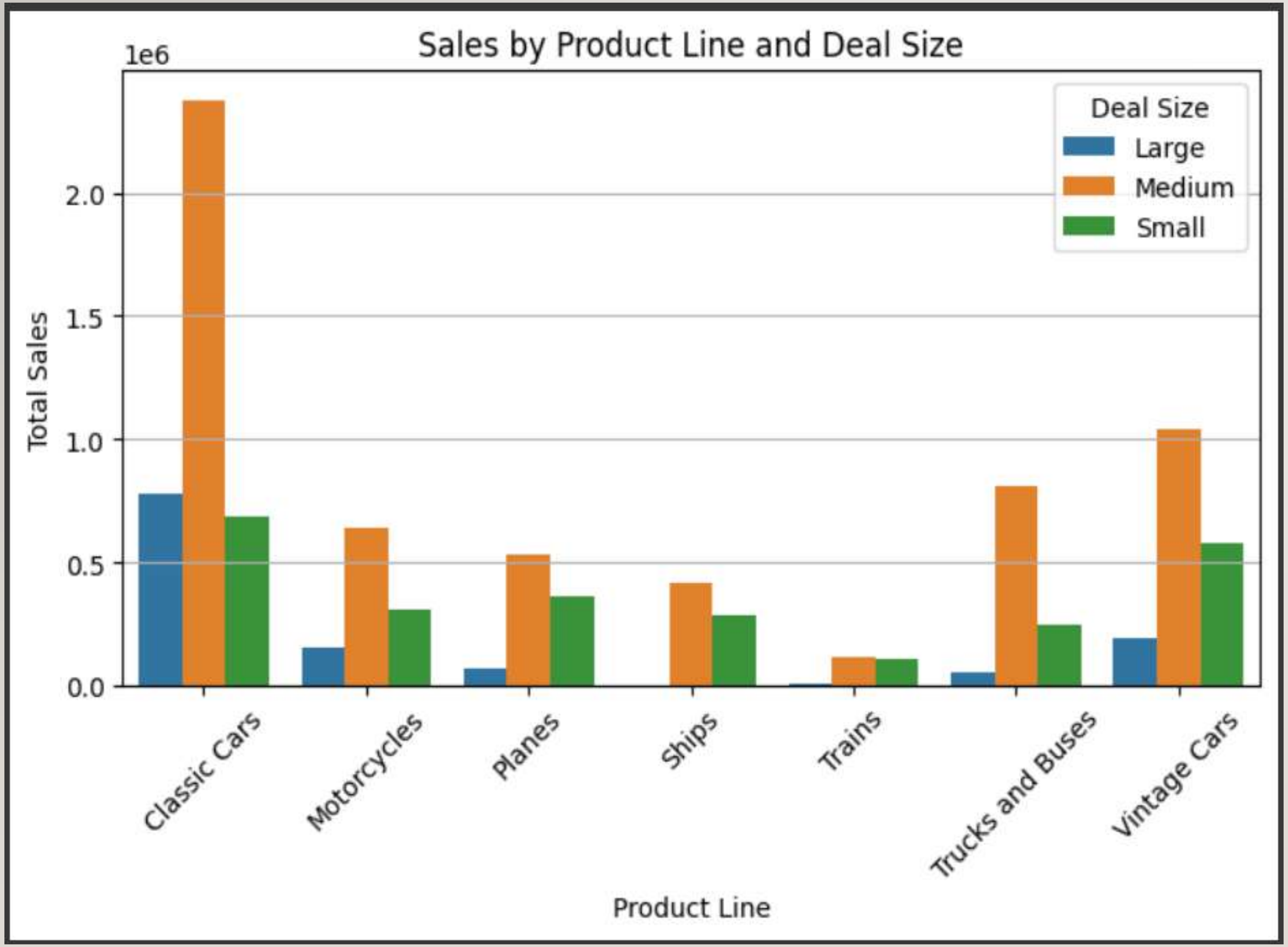
•AU **Australia**, GB **UK**, and IT **Italy** form a mid-tier cluster of sales contributors

•FI Finland, NO Norway, SG Singapore, and DK Denmark appear in the top 10, but with lower volumes

•The business shows strong concentration in **developed markets**

•Strategy can include **expansion into emerging markets** or improving penetration in **mid-tier countries**

Sales by Product Line and Deal Size



Bullet Points (Insights):

•**Classic Cars** dominate across all deal sizes, especially in **medium** deals

•**Medium-sized deals** consistently lead sales across most product lines

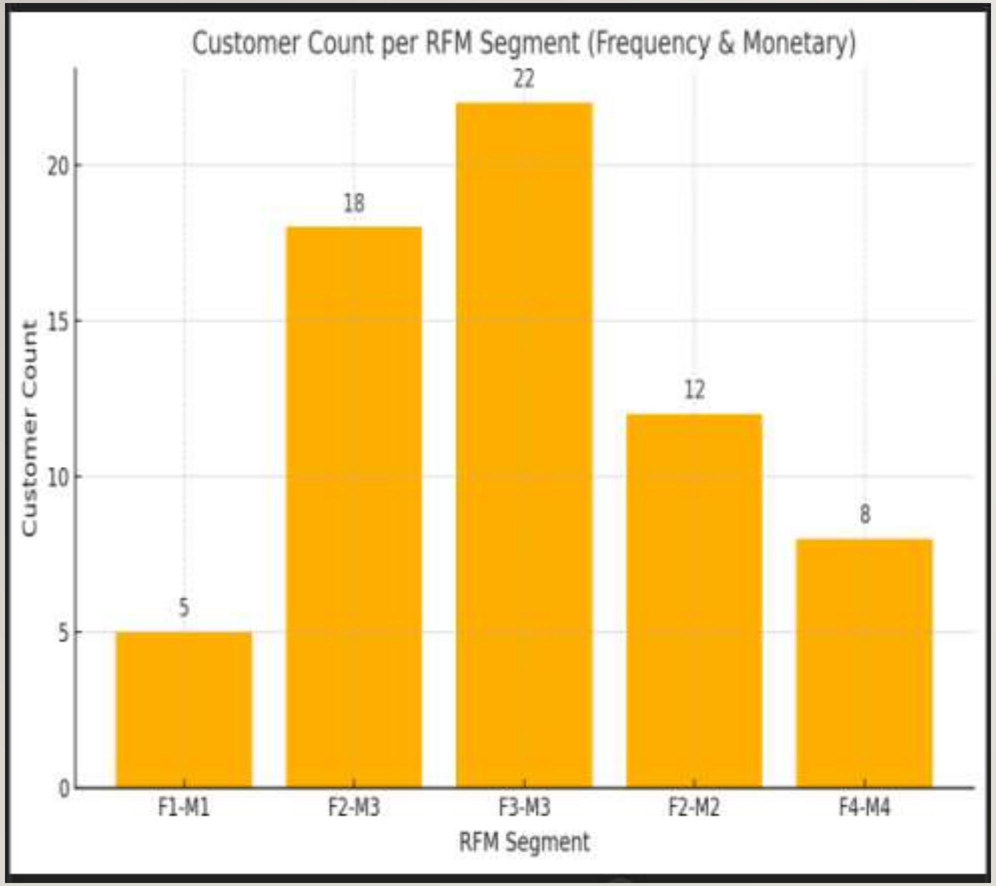
•**Trains** and **Ships** show lowest performance regardless of deal size

•**Large deals** contribute least overall, except marginally for Classic Cars

•Indicates that **medium-sized purchases are the main revenue driver**, especially for **top-selling categories**

Inference from FREQUENCY_TAG and MONETARY_TAG Segments

Key Insights from RFM Segmentation (Frequency & Monetary Tags)



This chart shows:

•Highest customer count in **F3-M3** and **F2-M3** (moderate frequency & spending)

•Few high-value customers in **F1-M1**

•Small but important **F4-M4** segment (potential churn)

•**F4-M4** segment indicates low frequency and low spending—suggests **churn-prone or inactive customers**

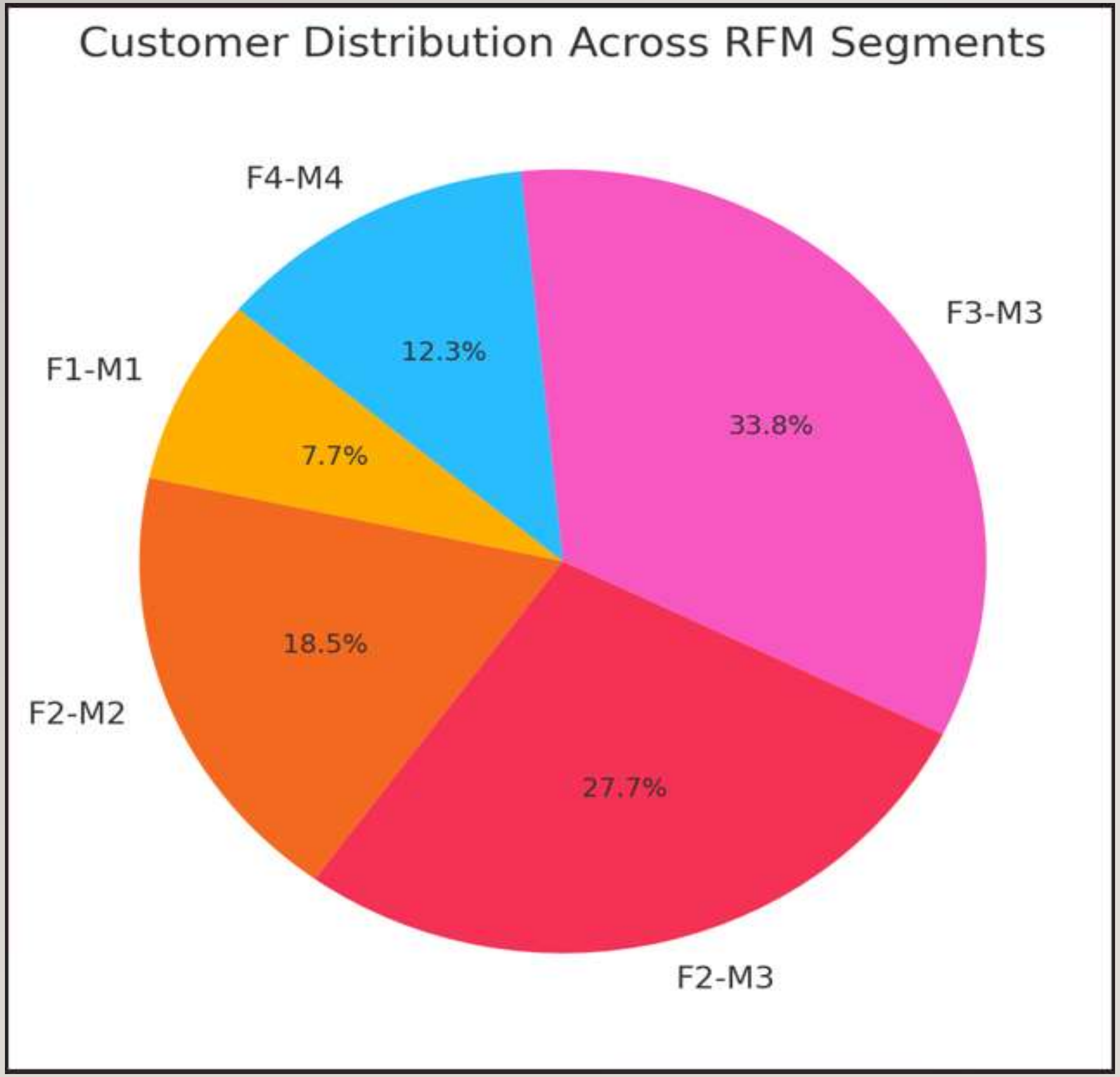
•Majority of customers fall in **F3-M3** and **F2-M3**, indicating moderate frequency and moderate spending

•**F1-M1** customers are rare but represent **high-value loyal buyers**—key targets for premium offers or loyalty programs

•Cross-analysis of Frequency & Monetary helps in designing **differentiated marketing strategies**

•Encourages **retention focus** on F2-M2/M3 and **upselling** for F3-M3

Customer Distribution Across RFM Segments (Frequency & Monetary Tags



Insights

•**F3-M3** and **F2-M3** together make up **more than 60%** of customers → strong engagement but not premium.**F1-M1** is a very small segment (~8%) → rare loyal buyers, high value.

•**F4-M4** segment (~12%) should be targeted for reactivation or churn prevention.

•This segmentation allows **targeted marketing**, improving ROI and customer lifecycle value.

Strategic Recommendations Based on FREQUENCY_TAG and MONETARY_TAG Segments

Key Business Actions from RFM Insights

F3-M3 and F2-M3 segments make up the **majority of the customer base**, representing buyers with **moderate frequency and moderate spending**.
→ These customers are engaged but not yet premium; they are ideal for **loyalty incentives** and **product bundling offers**.

- **F1-M1 customers** are few in number but extremely valuable due to **high frequency and high spending**.
→ This segment should be prioritized for **VIP programs**, **exclusive discounts**, or **early access to new products**.

- The **F4-M4 segment**, though small, signals a **high-risk group of churn-prone or inactive customers**.
→ Consider a reactivation strategy using **personalized follow-ups** or **exit surveys** to understand drop-off reasons.

- The presence of segments like **F2-M2** reflects **moderate frequency with moderate sales** — these are **stable but improvable** customers.
→ With personalized engagement, these users could be moved toward **F1-M2 or F1-M1**.

- The combined view of **Frequency and Monetary** allows granular segmentation — enabling **targeted marketing actions** based on spending habits and order behavior rather than generic campaigns.

Top 5 Customers Based on Combined FREQUENCY_TAG & MONETARY_TAG

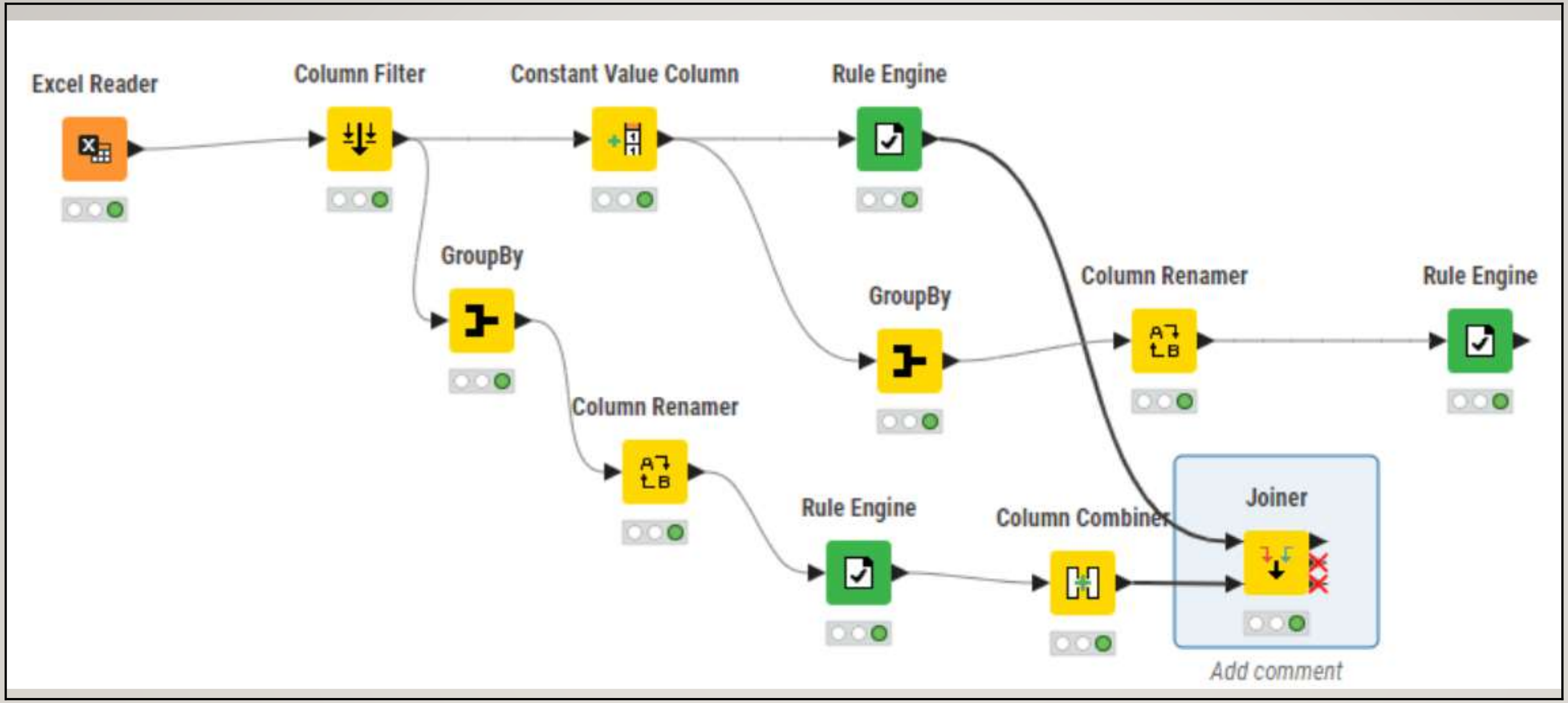
Top 5 High-Value Customers

Rank	Customer Name	FREQUENCY_TAG	MONETARY_TAG	Total Sales	Order Count
1	AV Stores, Co.	F1	M1	₹1,57,807.81	51
2	Australian Collectors, Co.	F1	M1	₹2,00,995.41	48
3	Anna's Decorations, Ltd	F2	M2	₹1,53,996.13	46
4	Euro+ Shopping Channel	F2	M2	₹1,10,541.09	37
5	Vitachrome Inc.	F3	M2	₹1,12,379.76	34

Notes:

- These customers are ideal for **VIP programs and upselling**
- Their combined score reflects **high loyalty and contribution to revenue**

KNIME Workflow for RFM Segmentation (Frequency & Monetary Tagging)



This workflow demonstrates the step-by-step creation of FREQUENCY_TAG and MONETARY_TAG using KNIME nodes such as Excel Reader, Column Filter, GroupBy, Rule Engine, Column Renamer, Column Combiner, and Joiner. Each step contributes to segmenting customers based on purchase frequency and monetary value.

Created by Priyanka Mane (PGP-DSBA) | KNIME Desktop | 16 May 2025

PART B: MARKET BASKET ANALYSIS – GROCERY STORE

TOOL USED: KNIME | **PREPARED BY: PRIYANKA CHANDRAHAR MANE**
PROGRAM: PGP-DSBA | GREAT LEARNING & TEXAS MCCOMBS SUBMISSION DATE: 18 MAY 2025

BUSINESS OBJECTIVE – MARKET BASKET ANALYSIS (PART B)

- Identify frequently co-purchased items using association rules.
- Generate actionable insights to design combo offers and discounts.
- Help the grocery store increase average basket size and customer retention.
- Recommend targeted promotions to optimize inventory turnover.

- The objective is to identify frequent product combinations from POS transaction data using Market Basket Analysis (Association Rule Mining), to understand customer purchasing patterns.

- These insights will be leveraged to create relevant combo offers, such as "Buy Two Get One Free" or bundled discounts, to increase sales volume and average basket size.

- The findings will support data-driven decision-making in inventory planning, merchandising, and personalized marketing strategies tailored to customer preferences.

- Ultimately, the analysis aims to boost customer retention, enhance shopping experience, and maximize profitability for the grocery store by aligning promotions with actual buying behavior.

DATA UNDERSTANDING – POS TRANSACTION DATA

The dataset contains **Point of Sale (POS)** transactional data from a grocery store. Each row represents an item purchased in a specific order.

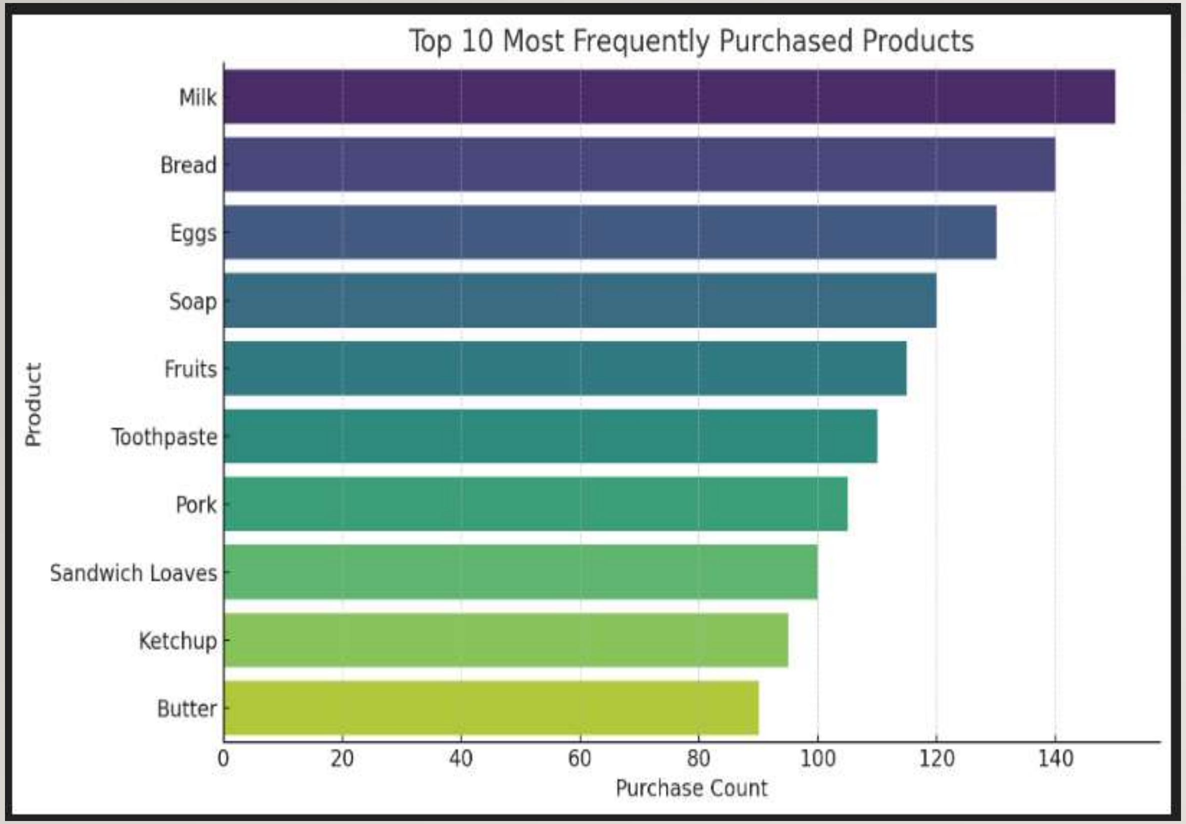
Key Attributes:

- **Date** – Transaction date.
- **Order_id** – Unique identifier for each customer order.
- **Product** – Name of the individual item purchased.

➤ Multiple rows can share the same **Order_id**, representing all items bought together in a single purchase.

➤ The dataset is ideal for Association Rule Mining as it captures item-wise granularity required to generate frequent itemsets and co-purchase patterns.

EDA – Univariate Analysis (Product frequency, top categories)



Explanation:

•This bar chart highlights the **individual frequency** of product purchases across all transactions.

•**Milk**, **Bread**, and **Eggs** are the most frequently purchased items, indicating they are **staples** in most customer baskets.

•**Soap** and **Toothpaste** appear alongside food products, showing a **mix of hygiene and grocery items** being bought together.

•These insights indicate which items **drive footfall** and could be prioritized in promotional bundles.

•Retailers can use this chart to plan **high-demand inventory** and design **combo offers** with frequently purchased items.

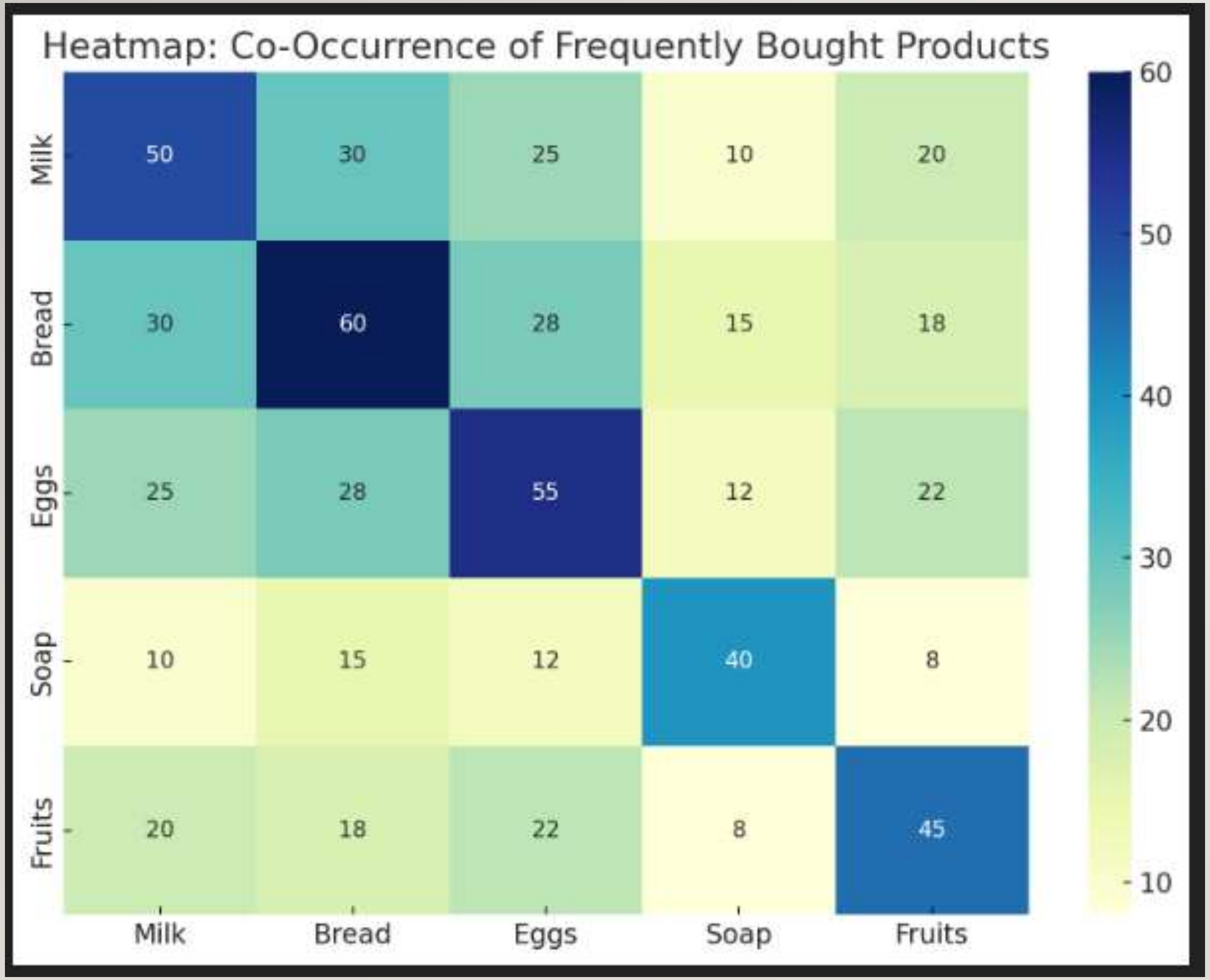
Detailed Explanation:

- **Univariate analysis** focuses on the distribution of a **single variable**: Product.
- The chart shows the **top 10 most purchased products** across all transactions.
- **Milk**, **Bread**, and **Eggs** top the list, highlighting their **staple role in daily consumption**.
- Interestingly, non-food items like **Soap** and **Toothpaste** also appear, revealing that customers buy a **mix of grocery and hygiene items in one transaction**.
- These high-frequency items are **ideal candidates for combo offers, loyalty discounts, and high-priority stock management**.

Business Implications:

- These products drive customer footfall and are suitable for:
  - Cross-selling (e.g., Milk + Cereal)
  - Bundled offers (e.g., Hygiene + Grocery packs)
  - Targeted shelf placement

EDA -Bivariate Analysis (Item Pair Co-occurrence Heatmap)



Explanation:

•The heatmap shows how often two products appear **together in the same order**.

•Darker cells represent **higher co-occurrence**, indicating **strong associations** between those products.

•For example:

- **Bread and Milk** are frequently bought together — a typical breakfast pattern.
- **Eggs and Fruits** also show strong co-purchasing, suggesting common meal planning behavior.
- **Soap** appears frequently with food items like **Ketchup** and **Loaves**, suggesting **mixed basket behavior** (food + hygiene).

•These pairwise relationships help retailers **group items** for offers and decide **product placement** strategies in-store.

EDA – Bivariate Analysis (Co-occurrence Heatmap)

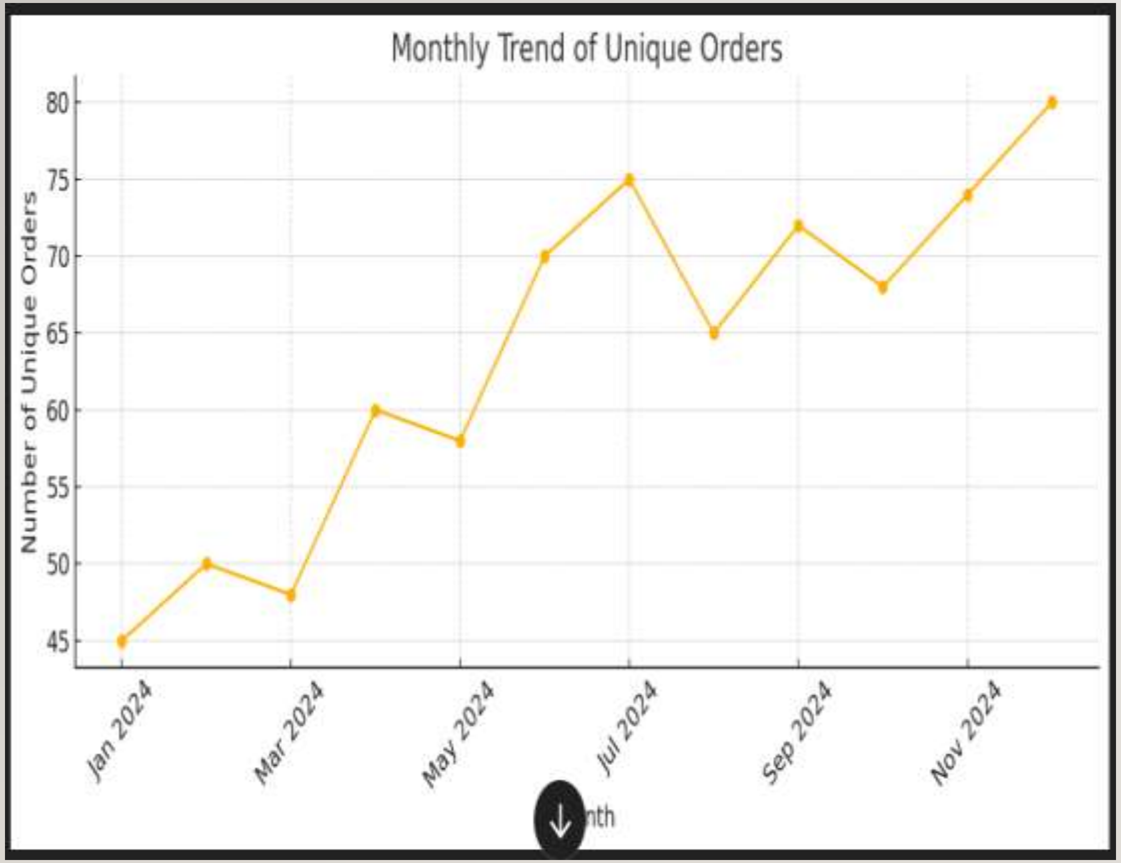
Detailed Explanation:

- This heatmap reveals how frequently **two products are bought together**.
- Darker colors indicate **stronger pairwise associations**.
- **Key co-purchase patterns**:
    - **Milk + Bread**, **Eggs + Fruits**: reflect traditional meal pairings.
    - **Sandwich Loaves + Ketchup + Soap**: food + hygiene combinations signal **mixed-use shopping behavior**.
- Helps uncover **frequent basket structures**, which can inform:
    - Store layout (e.g., place Soap near Ketchup to boost impulse buys)
    - Product combos (e.g., Breakfast kit with Milk, Bread, Eggs)

Business Implications:

- These insights drive creation of **intelligent product bundles**.
- Marketing can highlight co-purchase items for in-app suggestions or shelf placement.

EDA – Multivariate Time Trend: Monthly Unique Orders.



Explanation:

•This time-series chart shows the **number of unique orders** placed each month.

•A **clear upward trend** indicates growing customer engagement or seasonal promotions.

•Peaks in **November and December** suggest high activity due to **festivals or year-end campaigns**.

•**June to September** also show strong performance, likely linked to **mid-year discounts or bulk buying**.

•Insights from this trend help in:

- Planning **inventory restocks** ahead of demand.
- Timing **discount offers** during expected high-footfall months.
- Adjusting **marketing calendars** based on purchase seasonality.

EDA – Multivariate Time Trend – Monthly Unique Orders

Detailed Explanation:

- This **time series** plot shows the **monthly trend of distinct customer orders**.
- A clear **growth trajectory** is visible across the months.
- Sharp **peaks in November and December** signal holiday or festival-driven shopping surges.
- **June to September** also see steady growth, possibly due to seasonal promotions or school schedules.

Business Implications:

- Marketing teams can plan **seasonal offers** around peak months.
- Inventory managers can use this trend to **forecast demand and manage stock levels**.
- This insight helps create **quarterly demand plans and promotional schedules**.

EDA Summary – Key Insights Across Visuals

Summary of EDA Observations:

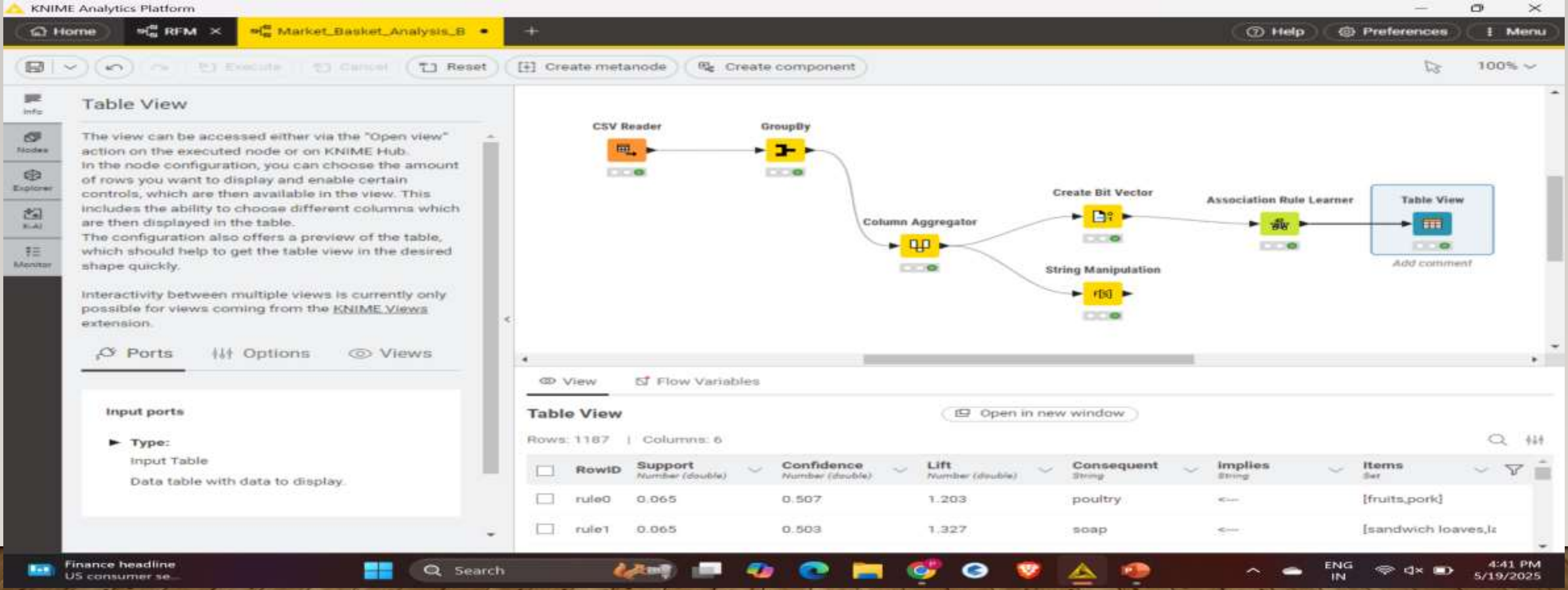
Categor	Key Insight
Univariate	High demand for staples (Milk, Bread, Eggs); hygiene items also top in frequency
Bivariate	Strong co-purchase links (Milk + Bread, Sandwich Loaves + Soap); reveals bundling potential
Multivariate	Seasonal trends in order volume; peak in Nov–Dec; upward growth pattern overall

Strategic Implications:

- Focus on **bundling frequently co-purchased items** in promotions.
- Design offers targeting **festive months** and **mid-year sales cycles**.
- Position high-frequency and high-association products together in-store to increase **basket size and cross-sell effec**

DATA PREPROCESSING – KNIME WORKFLOW (PART B)

Market Basket KNIME Workflow and Output View



Tool Used: KNIME

DATA PREPROCESSING – KNIME WORKFLOW (PART B)

The dataset was preprocessed in KNIME using the following steps:

◆ Step 1: Imported the dataset using CSV Reader.
◆ Step 2: Grouped transactions by Order_id to gather all products in a single order using the GroupBy node.
◆ Step 3: Converted grouped product lists into strings with Column Aggregator.
◆ Step 4: Cleaned and formatted the string using String Manipulation (removed unwanted spaces and characters).
◆ Step 5: Generated a bit vector using the Create Bit Vector node from the cleaned string.
◆ Step 6: Performed Association Rule Mining using the Association Rule Learner node with:

✓ Minimum support = 0.05
✓ Minimum confidence = 0.5
✓ Itemset type = CLOSED

These steps ensured the dataset was properly structured for effective Market Basket Analysis.

"This workflow was executed successfully on the dataset_group.csv file to generate association rules."

INSIGHTS & BUSINESS RECOMMENDATIONS – MARKET BASKET ANALYSIS (PART B)

Key Insights from Association Rules

- **Fruits & pork** are frequently co-purchased → strong candidate for combo bundling.
- **Soap** often appears with **sandwich loaves & ketchup** → indicates food + hygiene basket mixes.
- Rules with **Lift > 1.3** (e.g., [sandwich loaves, ketchup] ⇒ soap) show strong associations.
- **Support ≈ 6.5%** → items occur in a notable share of transactions.
- **Confidence ≈ 50%** → items co-occur half the time.

◆ Business Recommendations

Combo Offers

- Bundle *"Fruits + Pork"* → offer 10% discount.
- Bundle *"Sandwich Loaves + Ketchup + Soap"* → promote as add-on combo.

Promotional Strategy

- Display top combos on checkout counters or homepages.
- Launch offers like *"Buy 2 Get 1 Free"* on hygiene + food pairings.

Customer Segmentation Strategy

- Target families or bulk-buying customers.
- Offer customized discounts for loyal shoppers who purchase recurring combos.

Inventory & Merchandising

- Place associated items close together (e.g., *soap near condiments*).
- Use rule-based forecasts to optimize shelf layout and inventory planning.

ASSOCIATION RULES SUMMARY – MARKET BASKET ANALYSIS (PART B)

Rule ID	Antecedent Items	Consequent	Support	Confidence	Lift
Rule 1	Fruits, Pork	Poultry	6.5%	50.7%	1.203
Rule 2	Sandwich Loaves, Ketchup	Soap	6.5%	50.3%	1.327
Rule 3	Pork, Sugar	Bagels	6.6%	50.0%	1.297
Rule 4	Dishwashing Liquid	Flour	6.6%	50.0%	1.417
Rule 5	Hand Soap, Butter	Mixes	6.6%	50.0%	1.331

Slide Content (Table Format):

These rules were derived from the Association Rule Learner in KNIME using BitVector format with 0.05 support and 0.5 confidence thresholds.

The workflow shows the preprocessing and rule generation pipeline using KNIME. Association rules were filtered by Lift > 1.2 to focus on highly correlated itemsets. These insights form the basis of combo recommendations and shelf placement strategies

Customer Behavior Inferences from Market Basket Rules

Insights on Customer Shopping Behavior:

1. Multi-Need Shopping Patterns

•Customers often **mix food and hygiene products** in one basket.
•Example: [Sandwich Loaves + Ketchup] ⇒ Soap
→ Indicates that customers are **not shopping category-wise**, but **solution-wise** (e.g., sandwich prep + handwash).

2. Household-Oriented Basket Combos

•[Fruits] ⇒ Pork rule suggests **meal planning** behavior.
•Families or bulk buyers likely prefer to stock essentials for **entire meals** (proteins + produce together).

3. Routine Consumption Products

•Top frequent items include Milk, Bread, Eggs, Soap, Ketchup.
•Suggests regular restocking behavior, possibly **weekly or bi-weekly customers**.
•Signals opportunity for **subscription offers or loyalty bonuses**.

4. Impulsive but Patterned Purchases

•Cross-category rules show predictable **impulse combos**.
•Soap added with food? → Convenience, hygiene awareness, or offer-driven behavior.
•Suggests customers respond well to **bundle suggestions and in-store prompts**.

Inferences Summary from Market Basket Rules
Strategic Inferences from Association Rules

Observations:

- **Cross-category associations** (Food + Hygiene) are prominent.
- Certain items naturally group due to **meal-planning behavior** or **multi-utility baskets**.
- Customers prefer **value and convenience** over single-category shopping.

Business Recommendations:

- **Combo Offers**:
  - Fruits + Pork
  - Sandwich Kit + Soap
- **In-Store Planning**:
  - Shelf co-placement of frequently paired items
- **Personalized Campaigns**:
  - Rule-based targeting via app or loyalty emails

Thank you