Phase 3 Problem Statement — RetailSmart Advanced Analytics

Business Context

RetailSmart is an omni-channel e-commerce company that has already built strong capabilities in data cleaning, exploratory analysis, and churn prediction.
While the predictive model (from Phase 2) helps the company identify which customers are likely to churn, senior management now wants to deepen analytical intelligence in three critical areas:

1. **Customer Segmentation** — to personalize campaigns and retention offers
2. **Demand Forecasting** — to plan production, procurement, and logistics
3. **Cross-Sell Recommendations** — to increase average order value through bundled sales

The goal of this phase is to use **unsupervised learning**, **time-series analysis**, and **association rule mining** to discover patterns and trends that cannot be easily captured through supervised modeling.

Analytical Objectives

1. **Customer Segmentation (Unsupervised Learning)**
    - Group customers into meaningful segments using behavioral and marketing features (e.g., Recency, Frequency, Monetary, Avg Spend, Response Rate).
    - Identify cluster profiles such as *High-Value Loyalists*, *Price-Sensitive Frequent Buyers*, or *At-Risk Customers*.
    - Visualize segment separation using PCA and summarize behavioral differences between clusters.
    - Generate and save a summary of customer clusters as data_outputs/cluster_summary.csv.
2. **Demand Forecasting (Time Series Analysis)**
    - Analyze historical order and revenue trends.
    - Build a forecasting model to predict order volumes and total revenue for the next 6 months.
    - Provide insights on seasonality and growth trends for operational planning.
    - Save actual vs. forecasted monthly order volumes as data_outputs/forecast_results.csv.
    - Save the top product-pair associations (support, confidence, lift) as data_outputs/association_rules.csv.

Key Questions to Answer

Customer Segmentation

- How many distinct customer groups exist in the data?

- What behavioral traits differentiate high-value customers from low-value ones?
- How can these insights help design targeted marketing campaigns?

Demand Forecasting

- What is the monthly pattern of orders and revenue?
- Which months show seasonal peaks or dips?
- How accurately can we forecast demand for the next six months?

Market Basket Analysis

- Which product categories are most frequently purchased together?
- Which associations have the highest confidence and lift?
- How can RetailSmart use these insights to create combo offers or product recommendations?

Datasets Used

You will use the same cleaned datasets prepared in earlier phases:

Dataset	**Description**	**Key Columns**
customers_cleaned.csv	Customer profile and churn info	customer_id, city, state, total_orders, total_spent, churn_flag
sales_cleaned.csv	Transaction-level data	order_id, customer_id, product_id, category_english, total_price, order_purchase_timestamp
marketing_cleaned.csv	Campaign performance data	campaign_id, customer_id, channel, spend, conversions, response_rate
products_cleaned.csv	Product catalog	product_id, category_english, product_description_length, product_photos_qty