# Phase-1 Submission Report

## Student Information

Student Name: Kaviya V

Register Number: 613523106012

Institution: Government College of Engineering Dharmapuri

Department: Electronics and Communication Engineering (ECE) Department

Date of Submission: April 23, 2025

## 1. Problem Statement

Retailers often struggle to understand customer purchasing patterns from vast amounts of transactional data. Without proper analysis, it is difficult to make informed decisions about product placement, promotions, and inventory management. The lack of insight can lead to missed cross-selling opportunities and customer dissatisfaction. This project aims to mine association rules from customer purchase data to uncover hidden relationships between items. These rules can help optimize shelf arrangement, personalized promotions, and overall retail strategy.

## 2. Objectives of the Project

- Identify frequent itemsets in customer transactions.  
- Generate strong association rules using the Apriori algorithm.  
- Derive insights for product bundling and cross-selling.  
- Provide actionable recommendations for retail layout optimization.  
- Visualize the findings to aid strategic decision-making.

## 3. Scope of the Project

Scope:  
- Analyze historical transaction data.  
- Apply association rule mining techniques (e.g., Apriori algorithm).  
- Focus on market basket analysis.  
- Use Python and relevant data science libraries.  
  
Limitations:  
- Only applies to transactional datasets (not predictive modeling).  
- Assumes customer data is anonymized and cleaned.  
- Not deployed as a live retail optimization tool (analysis-only project).

## 4. Data Sources

The dataset used is the Online Retail Dataset from Kaggle.  
  
- Type: Public  
- Nature: Static (downloaded once)  
- Description: Contains over 500,000 transactions from a UK-based online retailer, including Invoice No., Stock Code, Description, Quantity, Invoice Date, Unit Price, Customer ID, and Country.

## 5. High-Level Methodology

Data Collection: Dataset downloaded from Kaggle.  
Data Cleaning: Handle missing values, remove duplicates, clean inconsistencies.  
Exploratory Data Analysis (EDA): Use bar plots and heatmaps to explore most purchased items and item co-occurrence.  
Feature Engineering: Convert data into a format suitable for the Apriori algorithm (e.g., one-hot encoded item matrix).  
Model Building: Use Apriori algorithm to extract frequent itemsets and generate association rules based on support, confidence, and lift.  
Model Evaluation: Evaluate rules using thresholds (Min Support: 0.01, Min Confidence: 0.2, Min Lift: 3).  
Visualization & Interpretation: Visualize rules using network graphs and matrix plots.  
Deployment: No deployment planned. Final analysis delivered as a Jupyter Notebook with visual outputs.

## 6. Tools and Technologies

- Programming Language: Python  
- Notebook/IDE: Google Colab or Jupyter Notebook  
- Libraries: pandas, numpy, mlxtend, matplotlib, seaborn, networkx, plotly  
- Optional Tools for Deployment: Streamlit (if needed)

## 7. Team Members and Roles

1. Vaishanavi M – Team Leader & Data Collection Lead  
 - Coordinate project timeline and meetings.  
 - Source and prepare the dataset.  
 - Oversee task delegation and milestone tracking.  
 - Final review and integration of all deliverables.  
  
2. Kaviya V (613523106012) – Data Cleaning & Preprocessing Lead  
 - Handle missing values, remove duplicates, and clean inconsistencies.  
 - Transform dataset into transactional format.  
 - Document all preprocessing steps.  
  
3. Bavana A (613523106003) – EDA & Visualization Specialist  
 - Perform EDA to identify trends and patterns.  
 - Generate graphs, heatmaps, and item frequency plots.  
 - Assist in visualizing association rules.  
  
4. Risha (613523106032) – Modeling & Rule Mining Expert  
 - Implement the Apriori algorithm and generate rules.  
 - Tune support, confidence, and lift thresholds.  
 - Interpret and summarize the most important rules.