# Mining Customer Purchase Patterns Using Association Rule Mining

* Student Name: Vaishanavi M
* Register Number: 613523106052
* Institution: Government College of Engineering, Dharmapuri
* Department: Electronics and Communication Engineering
* Date of Submission: 09-05-2025
* GitHub Repository Link: (https://github.com/vaishu200419/customer\_purchasing)

## 1. Problem Statement

In modern retail, understanding customer purchasing behavior is critical for maximizing profit and improving customer satisfaction. With massive volumes of transactional data, manually identifying patterns is impractical. This project focuses on uncovering hidden patterns through association rule mining. By using the Apriori algorithm, the goal is to generate frequent itemsets and extract meaningful rules that support decisions around shelf placement, promotions, bundling, and cross-selling strategies. This descriptive analytics project aims to deliver actionable insights from raw transaction data for retail optimization.

## 2. Project Objectives

* Guiding Questions:
* Identify frequently purchased itemsets using transactional data.
* Generate and evaluate association rules based on support, confidence, and lift.
* Derive insights for product bundling and strategic layout improvements.
* Use visualization to aid interpretation and decision-making.
* Present findings in an interactive format (if extended to deployment).

Guiding Questions:

* Which item combinations occur most frequently?
* What are the strongest association rules based on support, confidence, and lift?
* Can these rules help improve product placement and promotions?

Expected Deliverables:

* Clean and prepared transactional dataset.
* Association rules mined using Apriori algorithm.
* Visualizations (network graphs, heatmaps) to represent relationships.
* Strategic insights derived from the rules.

## 3. Flowchart of the Project Workflow

Data Collection → Data Cleaning → Exploratory Data Analysis → Transaction Matrix Creation → Apriori Rule Mining → Rule Evaluation → Visualization & Interpretation → Documentation

## 4. Data Description

* Dataset Name: Online Retail Dataset
* Source: Kaggle
* Type: Public, structured
* Size: Over 500,000 transactions
* Nature: Static (downloaded once)

Key Attributes:

* Invoice No. – Unique identifier for each transaction
* Stock Code – Product identifier
* Description – Name of the product
* Quantity – Number of units sold
* Invoice Date – Transaction timestamp
* Unit Price – Price per unit
* Customer ID – Buyer ID
* Country – Country of transaction

## 5. Data Preprocessing

* Handling Missing Values: Removed rows with null customer IDs.
* Removing Duplicates: Eliminated repeated records.
* Data Cleaning: Corrected inconsistencies in product descriptions.
* Transformation: Converted dataset into a one-hot encoded matrix for Apriori.
* Filtering: Removed negative quantities (returns) and non-relevant transactions.

## 6. Exploratory Data Analysis (EDA)

Univariate Analysis:

* Frequency distribution of top-selling items
* Bar plots showing most common products

Bivariate/Co-occurrence Analysis:

* Heatmaps to show co-occurrence of item pairs
* Correlation matrix for frequently bought items

Insights:

* Certain items were consistently purchased together.
* Some items act as triggers for larger baskets.
* Visualization revealed high-value bundles for cross-selling.

## 7. Tools and Technologies Used

**Language: Python**

**IDE/Environment: Google Colab / Jupyter Notebook**

**Libraries and Packages:**

* Data Handling: pandas, numpy
* Visualization: matplotlib, seaborn, networkx, plotly
* Association Rule Mining: mlxtend

## 8. Team Members and Contributions

* Vaishanavi M served as the Project Lead, overseeing data collection and ensuring the final integration of all components of the project.
* Kaviya V was responsible for data cleaning and preprocessing, including handling missing values and preparing the dataset for analysis.
* Bavana A focused on exploratory data analysis (EDA) and data visualization, identifying trends and patterns in the transaction data.
* Risha K implemented the Apriori algorithm for association rule mining and interpreted the resulting rules to extract meaningful insights.