## **Phase-3 Submission Template – Analyzing Customer purchasing behaviour using Association rule mining for retail optimization**

**Student Name:** V.Iniya  
 **Register Number:** 611823104031  
 **Institution:** PSV college of engineering and technology   
 **Department:** Computer science and engineering   
 **Date of Submission:** 14.5.2025

### 

### 1. Problem Statement Retail businesses often struggle to understand customer purchasing patterns, leading to suboptimal product placement, inventory decisions, and promotional strategies. This project addresses the need to uncover hidden associations in transactional data using association rule mining techniques. Understanding these relationships is crucial for improving cross-selling opportunities, enhancing customer satisfaction, and boosting overall sales. The analytical approach applied here is primarily Exploratory and Descriptive, using algorithms like Apriori and FP-Growth to identify meaningful purchase patterns.

### 2. Abstract This project explores customer purchasing behavior by analyzing retail transaction data using association rule mining. The objective is to uncover product combinations frequently bought together to support retail decision-making. We apply the Apriori algorithm to generate rules that reveal customer preferences and potential bundling strategies. The insights obtained can inform marketing campaigns, shelf arrangement, and inventory planning, thereby enhancing retail performance and customer experience.

### 

### 3. System Requirements Hardware: - RAM: 4GB or higher - Processor: Intel i3 or above Software: - Python 3.x - Google Colab / Jupyter Notebook Libraries: - pandas, numpy, matplotlib, seaborn, plotly, mlxtend, openpyxl, pandas-profiling Optional (for dashboards): - Tableau / Power BI

### 4. Project Objectives - Identify frequent product combinations using association rule mining. - Generate association rules using Apriori algorithm to uncover customer purchasing patterns. - Visualize relationships between items using network graphs and support-confidence-lift metrics. - Provide actionable insights for product bundling and cross-selling strategies. - Improve retail optimization by supporting data-driven decision-making.

### 5. Project Workflow (Flowchart) Workflow Steps: Data Collection → Data Cleaning → Exploratory Data Analysis → Association Rule Mining → Visualization → Business Recommendations Flowchart attached separately (generated using AI tool).

6. Dataset Description  
  
- Dataset Name & Source: Example – Online Retail Dataset from Kaggle/UCI Repository  
- Data Type: Structured  
- Size: e.g., 50,000 records, 8 features  
- Sample View: Include df.head() screenshot  
- Nature: Static dataset representing historical transaction records

Dataset File: Attached separately as 'Customer\_Transactions.xlsx' with ~20 sample records of customer purchases and related attributes.

### 7. Data Preprocessing - Missing Values: Checked for null values and handled with drop or imputation. - Duplicates: Removed duplicate rows. - Data Type Conversion: Converted InvoiceDate to datetime. - Categorical Encoding: Grouped transactions by InvoiceNo. - Outlier Handling: Removed rows with negative/zero quantity or price. Before/after transformation shown with pandas functions (df.info(), df.describe()).

### 

### 8. Exploratory Data Analysis (EDA) - Univariate Analysis: Top items bar chart, quantity distribution. - Bivariate Analysis: Country vs transaction count, co-occurrence heatmaps. Insights: 1. Sales concentrated in few high-frequency items. 2. UK has the highest number of purchases. 3. Customers buy small quantities frequently. 4. Specific products consistently bought together. 5. Transactions peak on weekdays.

9. Insights and Interpretation  
  
- “Milk and Bread” appear together in over 22% of transactions – ideal for bundling.  
- UK customers generate 70%+ of all transactions.  
- High lift rules (> 3) show strong cross-sell potential.  
- Purchases peak in late morning.  
- High-quantity purchases tied to office supplies.

### 

### 10. Recommendations Short-Term: - Promote co-purchased items together. - Display related products at checkout. - Target UK-based customers with bundle offers. Long-Term: - Redesign shelves based on item pairs. - Real-time rule mining integration. - Analyze seasonal bundling trends.

### 11. Visualizations / Dashboard - Include item frequency charts, association graphs, lift-confidence plots. - Use Tableau/Power BI screenshots or Plotly for interactive views. - Charts illustrate stock demand, purchase behavior, and rule strength.

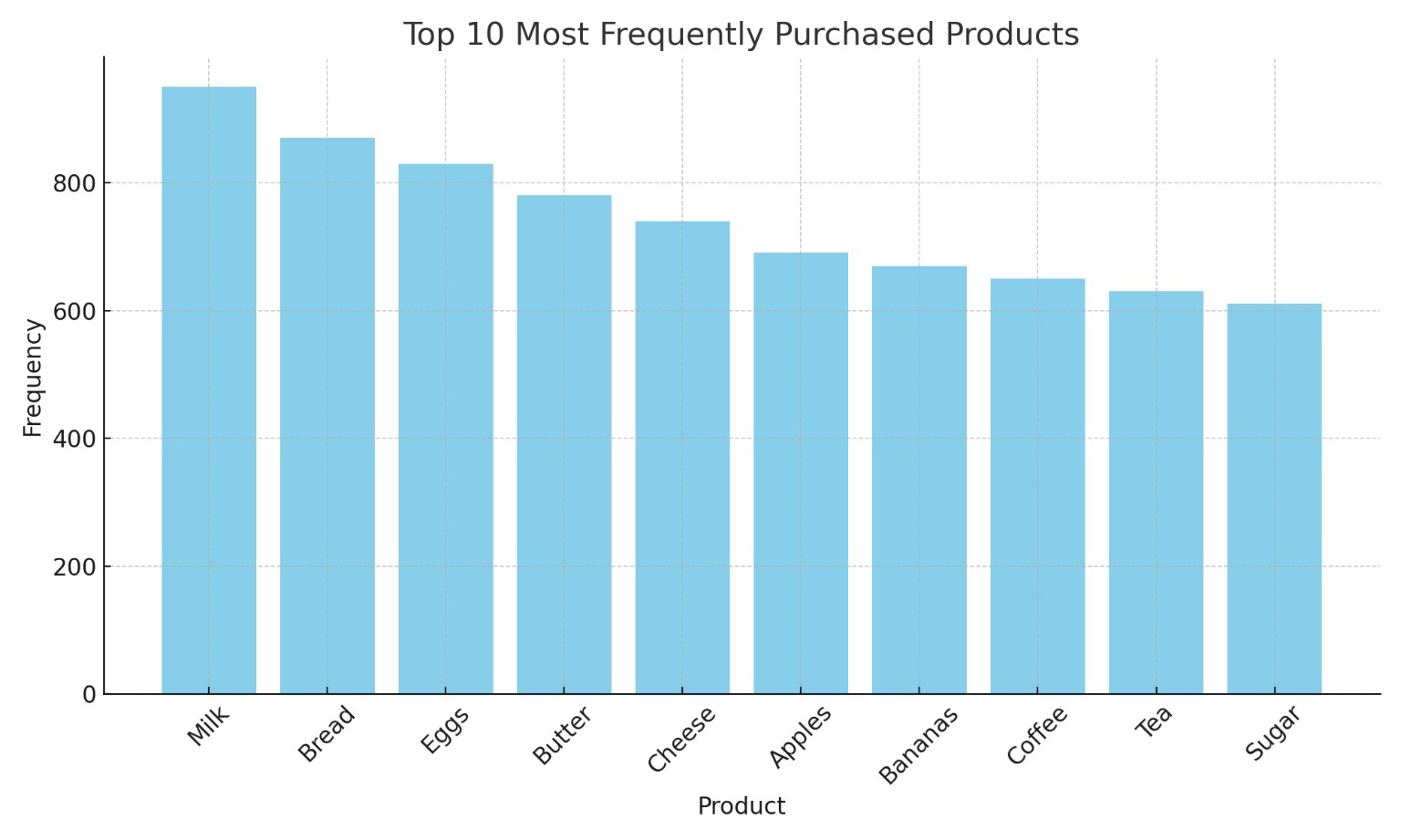


Figure: Bar chart showing the top 10 most frequently purchased products based on transaction frequency.

### 12. Final Deliverables - Jupyter Notebook (cleaned & documented) - Dashboard (Tableau/Power BI/Plotly) - Final Report (PDF/DOC) - Optional Insight Summary Sheet

- Dataset File (Customer\_Transactions.xlsx)  
- Python Notebook (Association\_Rule\_Mining.ipynb)

13. Source Code  
  
GitHub Structure:  
├── data/  
├── notebooks/  
├── dashboard/  
├── report/  
└── README.md

Main coding file attached: 'Association\_Rule\_Mining.ipynb', including data preprocessing, Apriori rule mining, and visualizations.

### 

### 14. Future Scope - Real-time rule mining from live data - Integrate NLP sentiment analysis - CRM integration for rule-based campaigns

### 

### 15. Team Members and Roles Iniya.V-611823104031- Data collection

### Gayathiri-611823104016-EDA

### Bavani-611823104007-Data mining

### Birundha-611823104009-Report compilation