

Final Project Report: Market Basket Analysis & Store Optimization (Apriori Model)

1. Problem Statement

In the competitive retail sector, understanding customer purchase behavior is critical for maximizing sales and improving customer satisfaction. However, many retailers face challenges due to poor shelf organization, lack of targeted promotions, and missed cross-selling opportunities. To address these issues, this project leverages Market Basket Analysis (MBA) using the Apriori algorithm to uncover frequent itemsets and strong association rules. By analyzing transaction-level data, we aim to identify patterns in product purchases that can guide decisions related to product bundling, promotional strategies, and inventory planning.

2. Technical Approach

Tools Used: Python, pandas, numpy, seaborn, matplotlib, apyori

Dataset:

- Source: Kaggle Groceries Dataset
- Total Transactions: 38,765
- Columns: Member_number, Date, itemDescription

Steps Executed:

1. Data Loading & Cleaning:

- Removed null entries (none found).
- Converted 'Date' column to datetime and set it as index.

2. Data Transformation:

- Grouped by Member_number and Date to create unique transactions.
- Converted grouped data into transaction lists.

3. Apriori Algorithm:

- Applied the apyori library to find frequent itemsets and association rules.
- Parameters:
 - min_support = 0.0003
 - min_confidence = 0.05
 - min_lift = 2
 - min_length = 2

4. Result Extraction:

- Interpreted results focusing on high-confidence and high-lift rules.

3. Key Findings & Visuals

Most Purchased Items:

Rank	Item	Frequency
1	Whole milk	2502
2	Other vegetables	1898
3	Rolls/Buns	1716
4	Soda	1514

Top Association Rules:

Rule	Confidence	Lift
Whole milk → Sausage	34.2%	2.17
Pastry + Soda → Sausage	13.1%	2.17
Sausage + Curd → Yogurt	20.6%	3.41
Brown Bread + Beer → Sausage	16.7%	2.76

Visuals:

- Bar chart showing top 20 frequently bought items.
- Optional: heatmap of item correlations for further visual exploration.

4. Strategic Recommendations

A. Product Bundling

Based on frequent item combinations:

- Offer combo packs like:
 - "Whole Milk + Sausage + Yogurt"
 - "Pastry + Soda + Coffee"
- Create seasonal bundles for holidays using items like tropical fruits, desserts, and beverages.

B. Shelf & Stock Optimization

- Place highly associated items adjacent to each other in-store.

- Reserve front or eye-level shelves for high-frequency items.
- Promote impulse-purchase pairs like "Rolls/Buns + Soda" near checkout.

C. Pricing & Promotions

- Provide discounts on bundled items (e.g., buy 2 get 1 free).
- Use cross-item promotions: "Buy Whole Milk, Get 10% off on Sausage".
- Promote underperforming but highly associated items like sweet spreads via pairing.

D. Enhanced Customer Experience

- Use signs or QR codes to recommend popular bundles.
- Organize store layout based on buying patterns to reduce customer effort.
- Incorporate association-based product suggestions in online carts.

Conclusion:

The Market Basket Analysis helped identify critical associations in the grocery dataset, revealing insights to optimize layout, pricing, and promotion. These strategies can improve customer satisfaction, drive sales, and support data-driven decision-making across departments like merchandising and marketing.