Market Basket Analysis and Customer Segmentation for Grocery Retail Optimization

Project by Parth Chaudhari

Introduction

In today's competitive retail landscape, understanding customer behavior is key to optimizing sales and operational efficiency. Market Basket Analysis is a powerful tool for identifying patterns in customer transactions. By analyzing these patterns, businesses can identify associations between items, segment customers based on purchasing behaviors, and uncover actionable insights to inform marketing strategies, inventory management, and customer retention efforts.

This analysis focuses on a grocery store dataset, leveraging association rule mining, customer segmentation, and exploratory visualizations to extract meaningful insights. These findings will guide data-driven strategies to boost revenue, improve customer satisfaction, and optimize operations.

Information About the Data

The dataset used for this analysis contains transactional records from a grocery store, with the following key attributes:

- **Member_number**: A unique identifier for each customer.
- **Date**: The date of the transaction.
- **itemDescription**: The items purchased in each transaction.

Characteristics of the Data:

- **Size**: 38,765 records.
- **Structure**: Each row represents an individual item purchased within a transaction.
- **Product Diversity**: Includes staples like whole milk, rolls/buns, and vegetables, as well as complementary items like yogurt, pastries, and tropical fruits.

This dataset provides a foundation for understanding customer purchase behavior, identifying frequent itemsets, and clustering customers based on their shopping patterns.

Methodology

To extract actionable insights, the analysis followed these structured steps:

1. **Data Preprocessing**:
    - Transactions were grouped by customer and date to create a list of items purchased together.
    - The data was converted into a transactional format for analysis, followed by one-hot encoding for association rule mining.

2. **Exploratory Data Analysis (EDA)**:
    - Transaction trends were analyzed across months and days of the week.
    - The distribution of frequent itemsets and item co-occurrence patterns were visualized.
3. **Association Rule Mining**:
    - The Apriori algorithm was used to identify frequent itemsets and generate association rules.
    - Metrics like **support**, **confidence**, and **lift** were calculated to evaluate the strength of associations.
4. **Customer Segmentation**:
    - Customers were clustered based on the diversity of their purchases using the K-Means algorithm.
    - The purchasing patterns and time-based behaviors of each cluster were analyzed.
5. **Visualization**:
    - Bar charts, heatmaps, and scatter plots were created to visually represent trends, correlations, and cluster-specific behaviors.

Tools Used

1. **Programming Language**: Python
2. **Libraries**:
    - **Pandas**: Data manipulation and transaction grouping.
    - **Matplotlib** and **Seaborn**: Visualization.
    - **Mlxtend**: Association rule mining using the Apriori algorithm.
    - **Scikit-Learn**: K-Means clustering for customer segmentation.
3. **Environment**: Google Colab for executing the analysis.

Findings and Insights

1. Association Rule Mining

- **Top Rules**:
    - Yogurt → Whole Milk (Confidence: 12.99%, Lift: 0.82)
    - Rolls/Buns → Whole Milk (Confidence: 12.70%, Lift: 0.80)
    - Other Vegetables → Whole Milk (Confidence: 12.15%, Lift: 0.77)
    - Soda → Whole Milk (Confidence: 11.98%, Lift: 0.76)

Insights:

- **Whole Milk** is a frequently purchased anchor product, weakly associated with complementary items like yogurt, rolls/buns, and soda.
- The low lift values (<1) indicate opportunities to strengthen these associations through marketing strategies.

Actionable Steps:

- Use cross-selling strategies to recommend yogurt, rolls/buns, or soda when whole milk is purchased.
- Design bundled discounts to encourage complementary purchases.

Here's the association rule output formatted as a table:

Antecedents	Consequents	Support	Confidence	Lift	Leverage	Conviction
Yogurt	Whole Milk	0.011161	12.99%	0.82	-0.002401	0.967861
Rolls/Buns	Whole Milk	0.013968	12.70%	0.80	-0.003404	0.964550
Other Vegetables	Whole Milk	0.014837	12.15%	0.77	-0.004446	0.958551
Soda	Whole Milk	0.011629	11.98%	0.76	-0.003707	0.956636

2. Monthly and Day-of-Week Trends

Monthly Trends:

- Transactions peak in **August**, likely due to summer holidays and back-to-school shopping.
- A dip in **February** suggests reduced activity post-holidays.

Day-of-Week Trends:

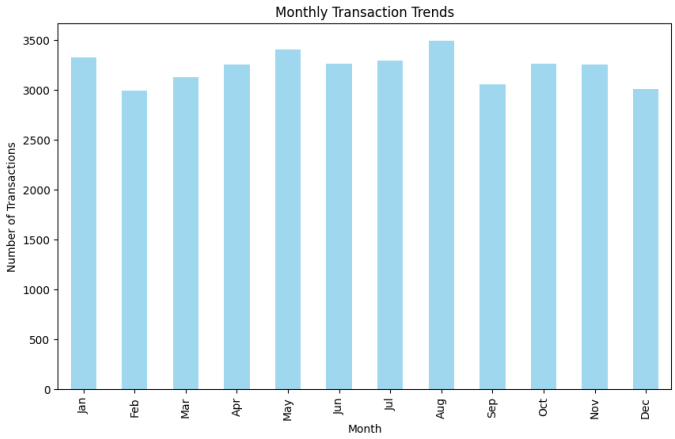
- Transactions are evenly distributed across all days of the week, indicating consistent shopping habits.

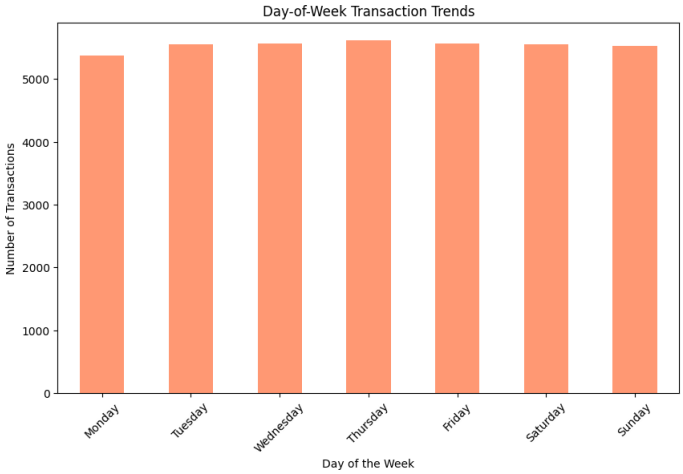
Insights:

- The August peak highlights seasonal demand, while the February dip indicates a need for targeted campaigns.
- The lack of day-specific peaks suggests flexibility in customer shopping patterns.

Actionable Steps:

- Launch back-to-school promotions in August and post-holiday discounts in February.
- Experiment with themed promotions like “Weekend Treats” or “Midweek Essentials.”





---

3. Heatmap of Co-occurrence Between Top Items

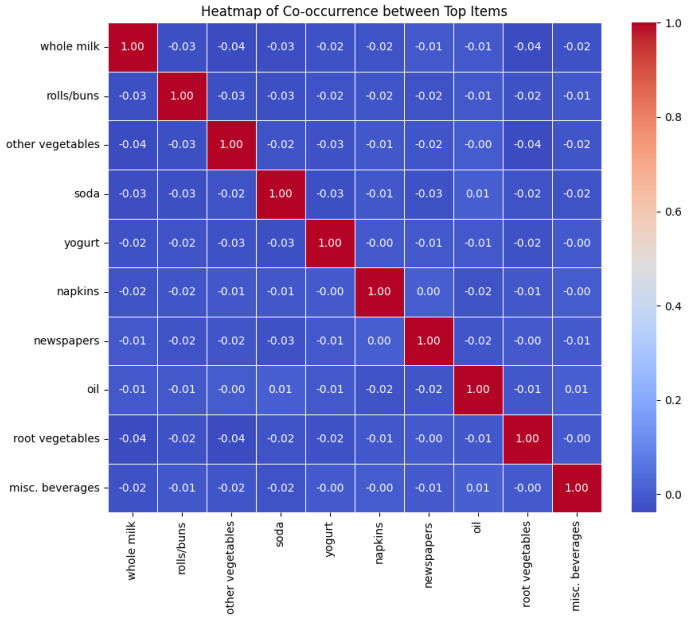
- The heatmap reveals weak correlations between top items, suggesting diverse and individualistic purchasing behaviors.
- Slight positive correlations exist between staples like whole milk and rolls/buns.

Insights:

- The weak co-occurrence patterns reflect highly personalized shopping preferences.
- Staples remain the backbone of most transactions.

Actionable Steps:

- Use targeted cross-sell recommendations based on weak correlations.
- Highlight staple items prominently in-store or online to encourage purchases.



---

4. Customer Segmentation

Clusters Identified:

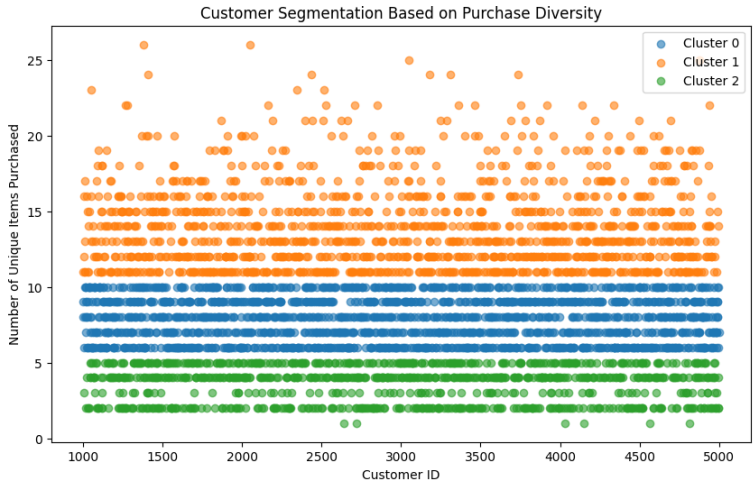
- **Cluster 0 (Low Diversity)**: Purchases focus on staples like milk and bread.
- **Cluster 1 (Medium Diversity)**: Includes complementary items like pastries and sausage.
- **Cluster 2 (High Diversity)**: High-value customers purchasing a wide range of items, including premium products.

Insights:

- Cluster 0 represents price-sensitive customers focused on essentials.
- Cluster 1 offers opportunities for cross-selling complementary products.
- Cluster 2 comprises high-value customers who can be retained through loyalty programs and personalized offers.

Actionable Steps:

- Offer basic bundles for Cluster 0 to retain their loyalty.
- Promote complementary products to Cluster 1 to increase basket size.
- Reward Cluster 2 with exclusive deals and loyalty points to maintain engagement.



Conclusion

This analysis highlights key opportunities to optimize marketing strategies, improve customer satisfaction, and boost revenue.

1. **Anchor Products**:
   - Whole milk serves as an anchor product. Bundling it with complementary items can increase sales.
2. **Seasonal Campaigns**:
   - Focus on back-to-school promotions in August and winter campaigns in February.
3. **Customer Segmentation**:
   - Tailor promotions and engagement strategies for each cluster:
     - Cluster 0: Affordable bundles.
     - Cluster 1: Cross-sell complementary products.
     - Cluster 2: Premium rewards and loyalty programs.
4. **Cross-Selling Opportunities**:
   - Use weak associations to create personalized recommendations and targeted campaigns.
5. **Operational Efficiency**:
   - Align inventory and stock levels with demand trends to prevent shortages or overstock.

Future Scope

1. **Advanced Personalization**:
   - Develop a recommendation engine to offer real-time product suggestions.
2. **Seasonal Analysis**:
   - Deep dive into product-specific trends for better alignment with seasonal demands.
3. **Dynamic Pricing**:
   - Implement pricing strategies tailored to cluster behaviors and seasonal trends.
4. **Loyalty Programs**:
   - Build a robust loyalty program for high-value customers to ensure retention and long-term profitability.
5. **Product Expansion**:
   - Use insights from Cluster 2’s preferences to expand into premium product lines.

This comprehensive analysis lays the groundwork for data-driven decision-making, enabling the business to thrive in a competitive retail environment. Let me know if you'd like further assistance with implementation or expansion!