UFAZ

FRENCH-AZERBAIJANI UNIVERSITY

Data Analysis and Statistical Thinking

Association Rules and Apriori Analysis: Summary Report and Business Recommendations

Author:

Javidan Hajiyev

Date: February 19, 2025

Abstract

This report presents an analysis of grocery shopping transactions using association rule mining with the Apriori algorithm. It summarizes key findings from the analysis and provides practical business recommendations based on the discovered rules. Additionally, a bonus analysis applies the algorithm to a real dataset and visualizes the rules using a network graph.

Contents

1 Introduction 3

2 Methodology 3

2.1 Data Preparation 3

2.2 Frequent Itemset Mining 3

2.3 Rule Generation 3

3 Key Findings 3

4 Limitations 3

5 Business Recommendations 4

5.1 Product Placement and Store Layout 4

5.2 Promotional Bundles 4

5.3 Inventory Management 4

5.4 Targeted Marketing and Cross-Selling 4

6 Bonus Analysis: Real Dataset and Visualization 4

6.1 Application to a Real Dataset 4

6.2 Network Graph Visualization 4

7 Conclusion 4

1 Introduction

This project examines grocery shopping transactions to identify patterns in items bought together. By applying the Apriori algorithm, we uncovered frequent item combinations and generated association rules. These insights help in understanding customer behavior and can inform strategies for store layout and promotions.

2 Methodology

2.1 Data Preparation

Transactions were grouped by customer and purchase date to create a dataset where each row represents a single transaction. A one-hot encoded table was then constructed to indicate whether each item was present in the transaction.

2.2 Frequent Itemset Mining

The Apriori algorithm was applied to the one-hot encoded data to identify itemsets that occur frequently. A minimum support threshold was set to filter out infrequent combinations, highlighting the most common item groups.

2.3 Rule Generation

Association rules were generated from the frequent itemsets. The key metrics used include:

- **Support**: The fraction of transactions that contain the itemset.
- **Confidence**: The probability that the consequent is purchased when the antecedent is purchased.
- **Lift**: A measure of the strength of the association, showing how much more likely the consequent is purchased with the antecedent than by random chance.

3 Key Findings

- **Popular Items:** Items such as *whole milk*, *other vegetables*, and *rolls/buns* appeared frequently across transactions.
- **Interesting Pairings:** Rules like *whole milk* → *other vegetables* indicate that these items are often bought together.
- **Threshold Sensitivity:** Adjusting the support and confidence thresholds greatly affects the number and type of rules generated, emphasizing the need for careful parameter tuning.

4 Limitations

- The findings are specific to the dataset's time period and store location.
- Small changes in support or confidence thresholds can lead to different sets of rules.
- Some associations may be seasonal or influenced by specific shopping behaviors.

5 Business Recommendations

5.1 Product Placement and Store Layout

If items like *whole milk* and *other vegetables* are frequently purchased together, consider placing them near each other in the store. This can facilitate convenient shopping and potentially increase sales.

5.2 Promotional Bundles

For item pairs with a high lift value, such as *yogurt* and *rolls/buns*, offering them as a bundled deal or at a discounted rate when bought together could encourage higher basket sizes.

5.3 Inventory Management

Ensure that items commonly purchased together are stocked simultaneously. Coordinated inventory management can help prevent situations where one item is out of stock, affecting the sales of its associated product.

5.4 Targeted Marketing and Cross-Selling

Utilize the association rules to develop targeted promotions. For example, if customers who buy *soda* often also purchase *tropical fruit*, offering a coupon for *tropical fruit* when a customer buys *soda* can enhance cross-selling opportunities.

6 Bonus Analysis: Real Dataset and Visualization

6.1 Application to a Real Dataset

For the bonus part of this project, the Apriori algorithm was applied to a real-world grocery dataset (e.g., the Groceries dataset from Kaggle). The same data preparation, frequent itemset mining, and rule generation processes were followed. This real dataset provided additional insights into customer purchasing patterns in a more diverse setting.

6.2 Network Graph Visualization

The association rules generated from the real dataset were visualized using a network graph created with `networkx` and `matplotlib`. In this graph:

- **Nodes** represent individual items.
- **Edges** represent association rules between items, with labels showing the confidence values.

Figure 1 shows the resulting network graph.

7 Conclusion

The Apriori algorithm provided valuable insights into customer purchasing patterns through association rule mining. The findings can be used to optimize store layout, refine promotional strategies, and improve inventory management. The bonus analysis further demonstrates the algorithm's utility on a real dataset and highlights the benefits of visualizing association rules for a more intuitive understanding of the data. Future work could explore larger datasets or seasonal trends to further enhance business strategies.

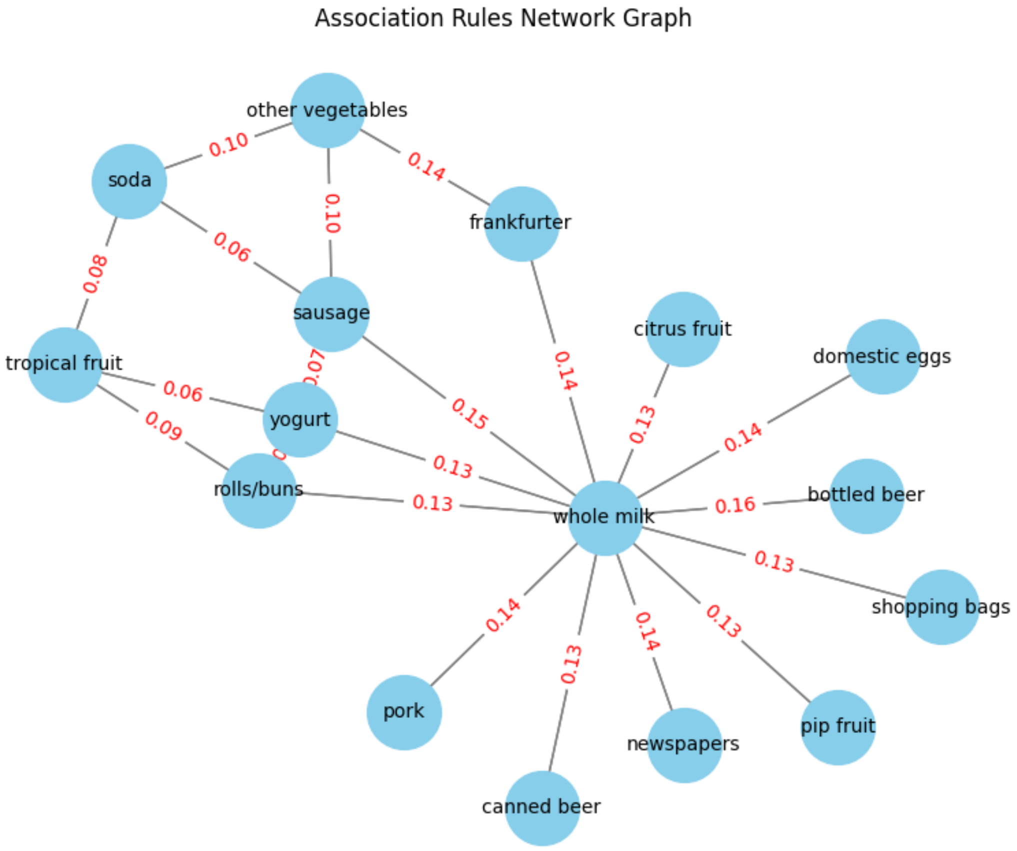

Figure 1: Association Rules Network Graph