STUDENT’S NAME: EBENEZER ESSILFIE WILSON

COURSE INSTRUCTOR: DR. DANIEL SMITH

COURSE TITLE: Data Analytics Graduate Capstone - D214(TASK 3)

JULY 4, 2025

A.

**Executive Summary**

* Statement of the Problem and Hypotheses

This project sought to investigate the effectiveness of association rule mining as a method for identifying statistically significant product combinations that could inform cross-selling strategies in a grocery retail context. The central aim was to determine whether discernible patterns of item co-occurrence exist within the dataset that could be leveraged for promotional and marketing decisions.

To guide this inquiry, the following hypotheses were formulated:

Null hypothesis(H₀) - There is no statistically significant relationship between products purchased together in a grocery store; association rule mining does not reveal effective product combinations for cross-selling.

Alternate Hypothesis(H₁) - There is a statistically significant relationship between products purchased together in a grocery store; association rule mining reveals effective product combinations for cross-selling.

* Summary of the Data Analysis Process

The analysis was conducted on the Grocery\_Dataset.csv, which consisted of individual product listings, including variables such as Price, Discount, and Product Description. Notably, the dataset lacked explicit transactions or customer identifiers, which are typically essential for market basket analysis. To address this limitation, a synthetic transaction structure was created by grouping records by the 'Price' attribute. Each unique price group was treated as a transaction, and associated 'Discount' values were considered as items within these pseudo-baskets.

The analytical process comprised the following steps:

Data Transformation: Grouping of discount descriptions by price.

One-Hot Encoding: Binary matrix creation using the TransactionEncoder from the mlxtend Python library.

Frequent Itemset Mining: Application of the Apriori algorithm with a minimum support threshold of 0.02 (2%).

Association Rule Generation: Derivation of rules using lift as the primary metric, with a minimum threshold of 1.0 to ensure meaningful relationships.

* Outline of Findings

The analysis revealed several association rules with notable statistical significance. Among these, three rules demonstrated both acceptable support levels and strong lift values:

Antecedents Consequents Support Confidence Lift

(After $3 OFF)(No Discount) 2.17% 100.00% 1.108

(No Discount) (After $3 OFF) 2.17% 2.41% 1.108

(No Discount) (After $4 OFF) 4.35% 4.82% 1.108

These findings indicate that certain discount types tend to co-occur more frequently than would be expected by random chance. The lift values exceeding 1.0 suggest a statistically significant positive relationship, thereby supporting the alternative hypothesis.

* Limitations of Techniques and Tools

A primary limitation of the analysis was the absence of true transactional data, which necessitated the use of synthetic groupings based on price. While this enabled the application of association rule mining, it may have introduced artificial correlations that do not accurately reflect consumer purchasing behavior.

Additionally, the Apriori algorithm, though effective for uncovering frequent item sets, can be computationally intensive and sensitive to parameter tuning. There is also the risk of excluding meaningful but less frequent patterns due to rigid support thresholds.

* Summary of Recommended Actions

Based on the analysis, the following recommendations are proposed:

Pilot Bundled Promotions: Implement cross-sell campaigns involving items with the strongest association metrics (e.g., “After $3 OFF” with “No Discount”).

Improve Data Collection Practices: Integrate transaction-level data (e.g., receipt or customer session IDs) in future datasets to enable more accurate and granular analysis.

Monitor Key Performance Indicators: Evaluate the effectiveness of pilot promotions by tracking metrics such as conversion rate, average transaction value, and revenue uplift.

* Expected Benefits

Quantifiable Revenue Impact: The implementation of targeted promotions based on high-lift item combinations is projected to improve cross-sell conversion rates by 2%–5%.

Operational Insight: Understanding co-occurring discount types can guide inventory management, shelf placement, and marketing strategies.

Enhanced Analytical Capability: Future availability of transaction-level data would enable more advanced techniques, including sequential pattern mining, customer segmentation, and predictive modeling.