Association Rule Mining (Market Basket Analysis)

Kwagala Ian	2023-B072-20932
Aijuka Josbert	2023-B071-20758
Babirye Stephania	2023-B071-21283
Nakakande Jennifer W.S	2023-B071-22675
Kyoloobi Edwin	2023-B071-22336

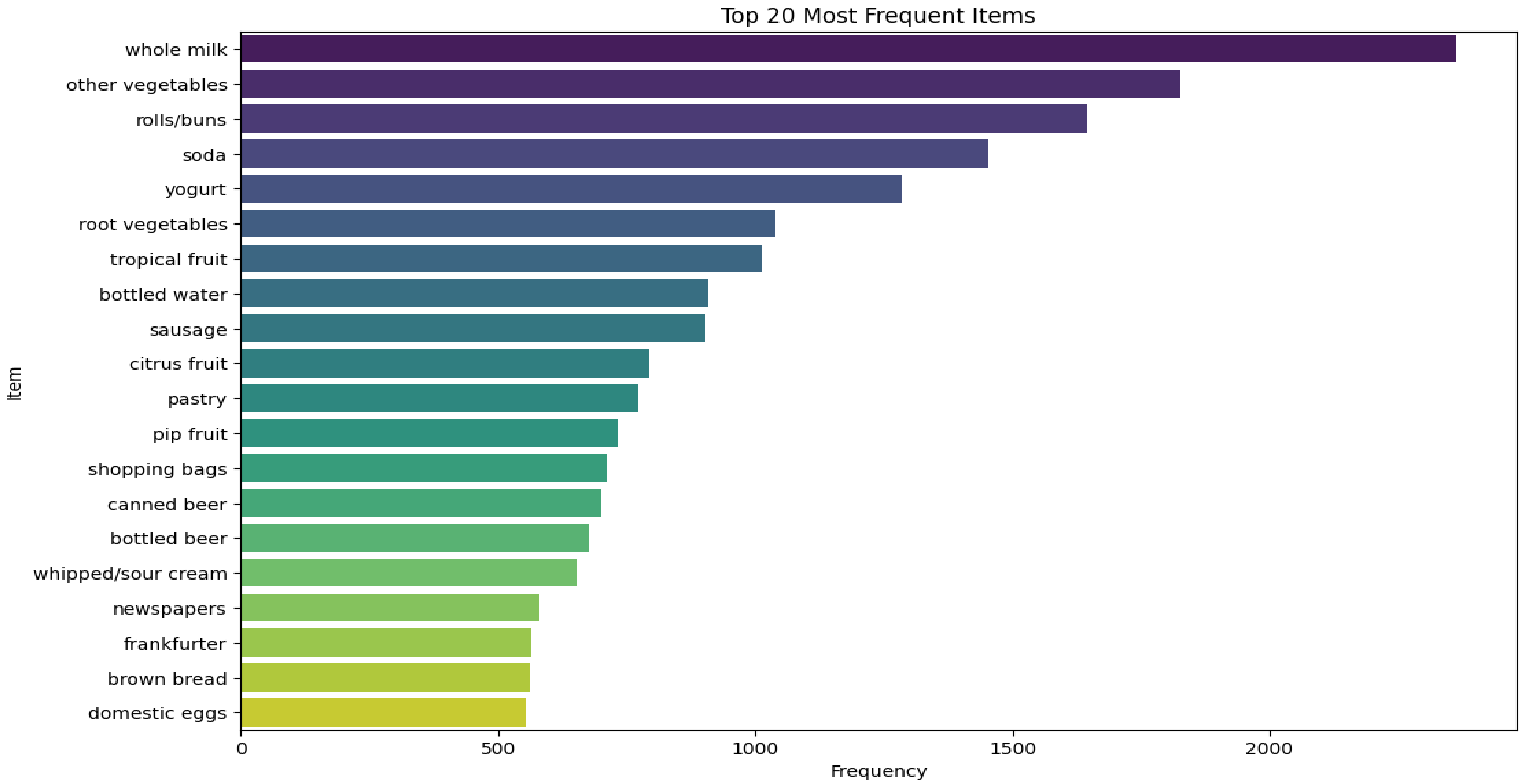
Groceries Dataset Analysis

- **Objective**: To analyze grocery transactions to find frequent co-purchase patterns.
- **Algorithms**: Apriori vs. FP-Growth
- **Dataset**: 9,835 transactions from a grocery store.

The Process: From Raw Data to Actionable Insights

- **Our Workflow**
- **Data Preprocessing**: Loaded the data, removed duplicates, and aggregated items into transactions. Cleaned item names for consistency.
- **Exploratory Data Analysis (EDA)**: Visualized the most popular items and their co-purchase frequencies to understand the data's landscape.
- **Model Building**: Applied Apriori and FP-Growth algorithms to mine frequent itemsets from the data.
- **Rule Generation**: Generated association rules and ranked them by *lift* to find the most interesting patterns.
- **Interpretation**: Translated statistical rules into actionable business strategies.

The bar chart of the Top 20 most frequent items



Algorithm Showdown: Apriori vs. FP-Growth

Which algorithm is more efficient for our dataset?

Algorithm	How it works	Performance
Apriori	Generates and tests candidate item sets at each level	Slower due to multiple datasets scans
FP-Growth	Uses a compact FP-Tree to mine frequent patterns	**Significantly faster**; only two database scans.

Conclusion: FP-Growth was the superior choice for this task, offering the same results as Apriori in a fraction of the time.

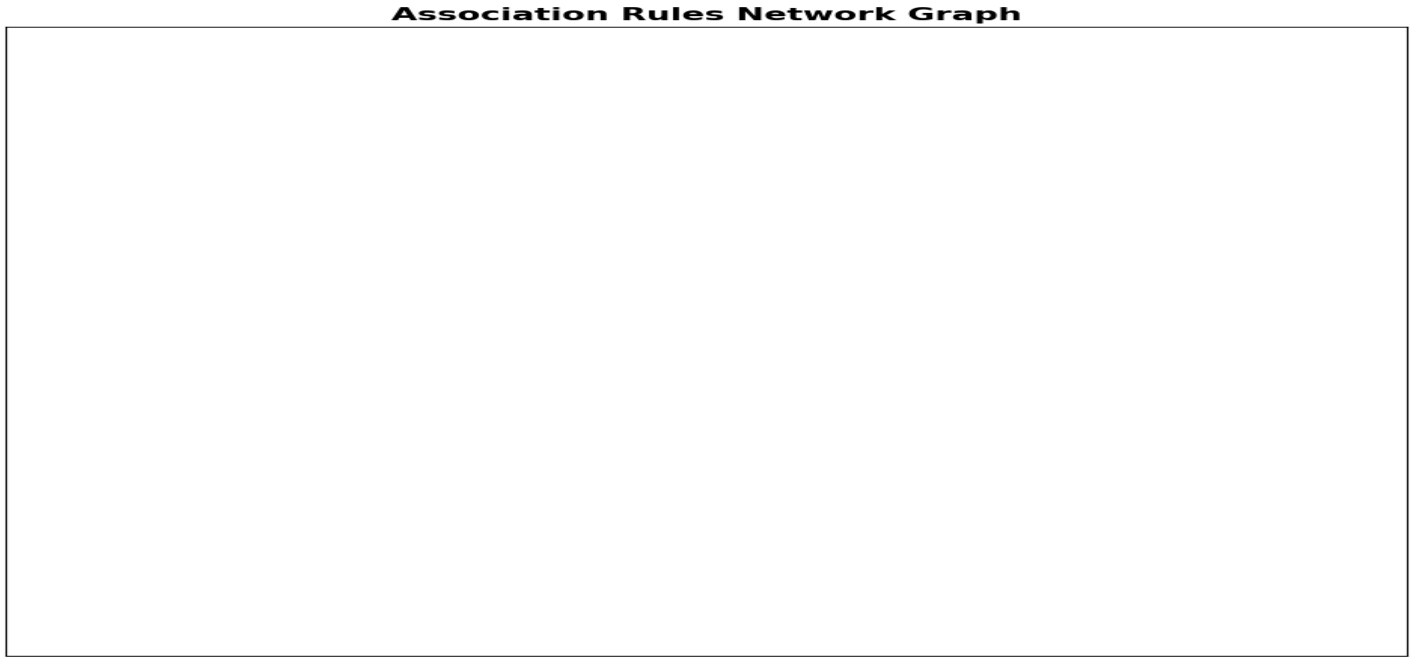
Our Top 3 Most Powerful Rules

We generated association rules with min_support=0.01 and min_confidence=0.4, ranked by **lift**.

Rule	Support	Confidence	Lift	Business Insight
{yogurt, tropical fruit} -> {whole milk}	0.015	0.50	3.23	**"Healthy Breakfast" Basket**: A strong, specific buying pattern.
{other vegetables, tropical fruit} -> {whole milk}	0.021	0.49	3.16	**"Fresh Produce Run"**: Customers buying fresh items often add milk.
{other vegetables, rolls/buns} -> {whole milk}	0.024	0.43	2.79	**"Meal Prep" Basket**: Common ingredients for daily meals.

Key Insight: "Whole milk" is a key connector in many specific purchasing scenarios, not just a popular standalone item.

Visualizing The Rules: A Network Graph



Why Your Rules DataFrame is EmptyYour association rules DataFrame is empty because none of the rules generated meet the minimum confidence threshold of 0.4 that you specified in this line:rules = association_rules(frequent_itemsets_final, metric="confidence", min_threshold=0.4)

- This indicates that in your grocery dataset:The relationships between items are not very strong (at least not at the 0.4 confidence level)
- Your min_support=0.01 combined with min_confidence=0.4 may be too restrictive for this particular dataset

Recommendations & Project Reflections

- **Actionable Strategies**
- **Create "Smart" Bundles**: Offer a "Breakfast Smoothie" deal (yogurt + tropical fruit + milk) to boost sales.
- **Optimize Store Layout**: Place high-association items near each other, such as putting a display of tropical fruit near the dairy aisle.
- **Targeted Promotions**: Use checkout data to offer coupons for items a customer is likely to buy but hasn't picked up yet (e.g., offer a milk discount to someone buying rolls and vegetables).
- **Limitations & Final Thoughts**
- **Support Threshold**: Finding the "right" support level is more of an art than a science.
- **Data Sparsity**: Most potential item combinations never appear, making it hard to find patterns.
- **Future Work**: Incorporating customer demographics could unlock even more personalized marketing strategies.