Apriori Analysis Report

1. Methods and Approach

The dataset I used for this project has the transactional data with each row listed items purchased together.

- Preprocessing:

I dropped the missing values to keep only the actual items. All items were converted into strings to prevent the encoding errors.

- Encoding:

The transactions were converted into a one-hot encoded format using the TransactionEncoder, each item became a column, and transactions were rows with 1's and 0's.

- Model Selection: I used 2 models:

Apriori Algorithm: to find frequent itemsets based on minimum support.

Association Rules: to generate rules based on the frequent itemsets, using support, confidence, and lift.

- Analysis Tools:

I created visualizations like histograms, bar charts, pie charts, to better understand the dataset and results.

2. Results and Analysis

- Frequent Itemsets:

64 frequent itemsets were identified with a minimum support threshold of 2%. The most common item (3) appeared in over 67% of transactions.

- Association Rules:

Strong associations found between 'whole milk' and other items like (5), with high lift values (1.62), supporting a positive purchasing relationship.

- Visual Insights:

Most transactions only contained only a few items (typically 3), and the scatter plot

showed that the even itemsets with lower support achieve high confidence, inferencing that valuable niche patterns in purchasing behavior exist.

3. Comprehensive Conclusion

Through the Apriori and association rule mining, clear purchasing trends and associations were found. Both models showed that a small set of key items really dominated most of the purchases and that certain products were to be bought together with strong likelihood. After cleaning and encoding the data, the Apriori model found 64 frequent itemsets, with some items (like item 3) being very common in transactions. Using those itemsets, the Association Rules model found interesting item to item relationships, like "whole milk" being bought with item 5, proved by a strong lift score. Overall, the models showed clear buying patterns and connections between popular items, which could be really useful for making smart product bundles or promotions in a real store. The models revealed key buying patterns and helped highlight the interesting purchasing trends.