Report on Frequent Itemsets and Association Rules Analysis

Dataset Overview

The dataset represents transaction records of grocery items, where each row corresponds to a transaction and each column represents the presence or absence of an item in that transaction. Items include categories like UHT-milk, beef, yogurt, whisky, white wine, etc. The data consists of 5 rows and 167 columns, with True indicating the presence of an item and False indicating its absence.

Frequent Itemsets Analysis

We applied two algorithms for discovering frequent itemsets from the transaction data: **Apriori** and **FP-growth**. The key objective was to identify which items frequently appear together in transactions. The following summaries provide insights into the results from both algorithms.

1. Apriori Frequent Itemsets

The Apriori algorithm identifies itemsets that occur together in a significant proportion of transactions. Some frequent itemsets identified are as follows:

Support	Itemsets
0.021386	(UHT-milk)
0.033950	(beef)
0.021787	(berries)
0.016574	(beverages)
0.045312	(bottled beer)
...	...

From the Apriori algorithm, itemsets with relatively high support values include:

- **(UHT-milk)** with a support of 0.021386
- **(beef)** with a support of 0.033950
- **(berries)** with a support of 0.021787

2. FP-growth Frequent Itemsets

FP-growth is another method for mining frequent itemsets and often performs faster than Apriori for larger datasets. The results from FP-growth are as follows:

Support	Itemsets
0.157923	(whole milk)

Support	Itemsets
0.051728	(pastry)
0.018780	(salty snack)
0.085879	(yogurt)
0.060349	(sausages)
...	...

Notable frequent itemsets from FP-growth include:

- **(whole milk)** with a support of 0.157923
- **(pastry)** with a support of 0.051728
- **(salty snack)** with a support of 0.018780

The FP-growth algorithm tends to show a higher support for certain items like **whole milk**, which appears significantly more frequently than others.

Association Rules Analysis

After identifying the frequent itemsets, we generated association rules based on the Apriori and FP-growth algorithms. The rules reveal relationships between itemsets, showing how the presence of one item in a transaction can suggest the presence of another item.

1. Apriori Association Rules

For the Apriori algorithm, we considered the "lift" metric to assess the strength of the association between items. Some association rules derived from Apriori are:

Antecedents	Consequents	Support	Confidence	Lift	Leverage	Conviction
(rolls/buns)	(other vegetables)	0.010559	0.095990	0.786	-0.002872	0.971
(whole milk)	(other vegetables)	0.014837	0.093948	0.769	-0.004446	0.968
(rolls/buns)	(whole milk)	0.013968	0.126974	0.804	-0.003404	0.964
(whole milk)	(rolls/buns)	0.013968	0.088447	0.804	-0.003404	0.976

2. FP-growth Association Rules

Association rules derived from the FP-growth algorithm are as follows:

Antecedents	Consequents	Support	Confidence	Lift	Leverage	Conviction
(yogurt)	(whole milk)	0.011161	0.129961	0.823	-0.002401	0.968
(whole milk)	(yogurt)	0.011161	0.070673	0.823	-0.002401	0.984
(soda)	(whole milk)	0.011629	0.119752	0.758	-0.003707	0.957
(whole milk)	(soda)	0.011629	0.073635	0.758	-0.003707	0.975

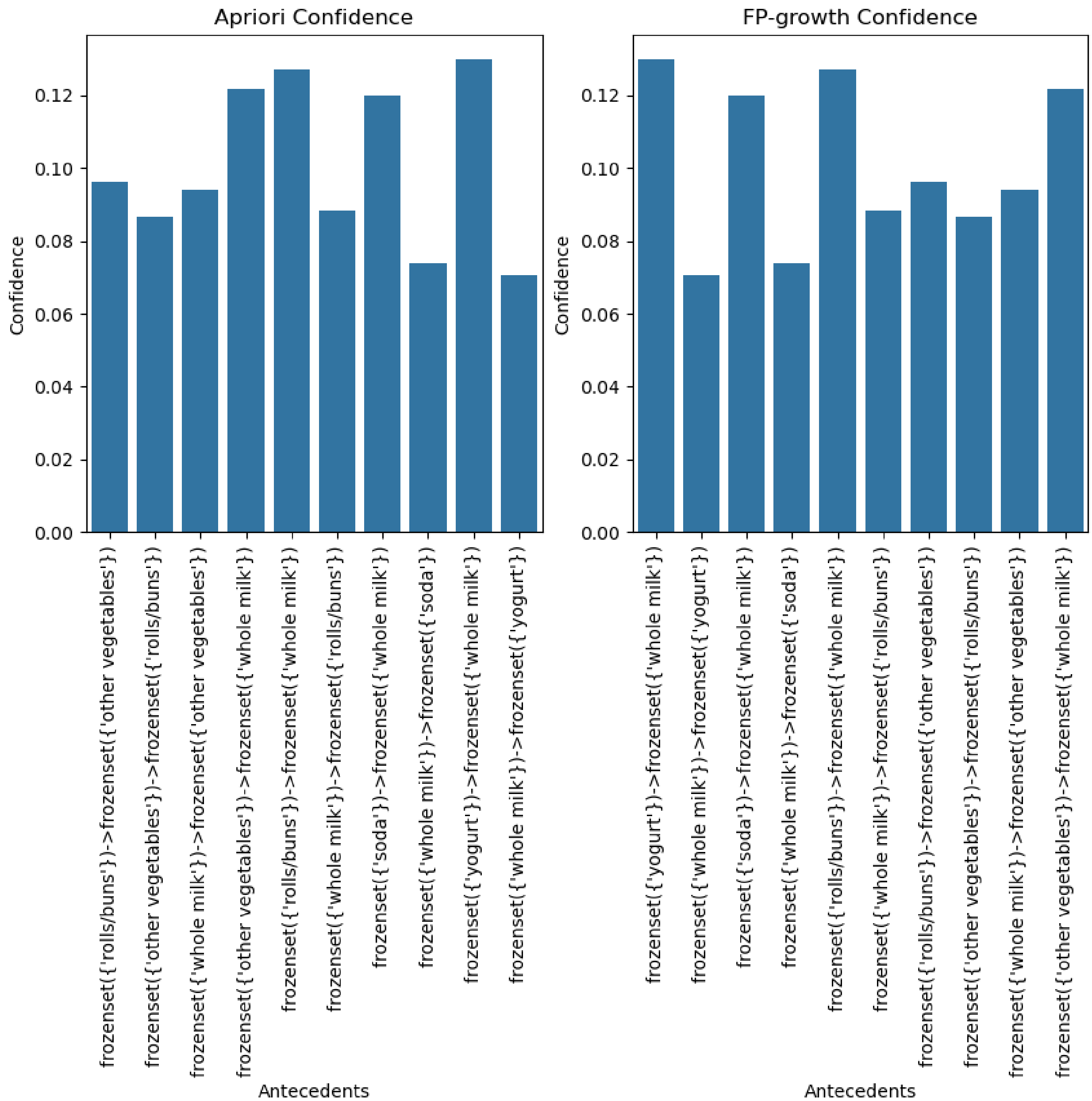
The rules highlight frequent co-occurrences such as:

- **(yogurt) => (whole milk)** with high confidence and lift values.
- **(whole milk) => (yogurt)** with a similar pattern.

Insights & Conclusion

From the analysis of frequent itemsets and association rules, we can conclude:

- **Whole milk** is a dominant item in both the frequent itemsets and association rules, appearing in many strong associations with other products like **yogurt**, **soda**, and **rolls/buns**.
- **Apriori** and **FP-growth** yield slightly different results in terms of itemsets with FP-growth showing a higher support for products like **whole milk**, while Apriori identifies more diverse sets of items.
- The association rules with high lift values suggest significant relationships between products, which could be useful for marketing strategies or product placement decisions.


Apriori Confidence
FP-growth Confidence
Confidence
Antecedents
0.00
0.02
0.04
0.06
0.08
0.10
0.12
frozenset({'rolls/buns'})->frozenset({'other vegetables'})
frozenset({'other vegetables'})->frozenset({'rolls/buns'})
frozenset({'whole milk'})->frozenset({'other vegetables'})
frozenset({'other vegetables'})->frozenset({'whole milk'})
frozenset({'rolls/buns'})->frozenset({'whole milk'})
frozenset({'whole milk'})->frozenset({'rolls/buns'})
frozenset({'soda'})->frozenset({'whole milk'})
frozenset({'whole milk'})->frozenset({'soda'})
frozenset({'yogurt'})->frozenset({'whole milk'})
frozenset({'whole milk'})->frozenset({'yogurt'})