D 351 Mini Project 3

Due: 11/12/2023

Mining Frequent Itemsets

In this project, our aim is to employ the Apriori algorithm and PCY algorithm to identify frequent itemsets within Kaggle Groceries Dataset. This time, every student will perform a Python implementation that faithfully follows the steps described in the text.

Data

You will be working on Groceries dataset which has the data of customers buying groceries with customer id, Date of purchase and the item description. You can find the dataset from this link.

Instructions for python implementation

1.Write a Python program to implement the Apriori algorithm on the Kaggle Groceries dataset using this link. Discuss support, confidence, and lift metrics and their significance. Customize the minimum support threshold to **0.12%** and mine frequent itemsets. Generate association rules with a minimum confidence threshold of **6.0%** and evaluate their importance in market basket analysis. (Example: Support threshold value and confidence threshold values for association rule **rolls/buns + other vegetables -> soda** are **0.11** and **10.76**)

2.Also explain in detail about the key concepts and significance of support, confidence, and lift in association rule mining. How does the Apriori algorithm use these metrics to mine frequent itemsets and generate association rules? Provide insights into how these results are utilized in the context of market basket analysis.

3.(Bonus Question) Implement the PCY algorithm to find frequent itemsets with a minimum support threshold of **0.12%** and a confidence threshold of **6.0%** and compare its performance with the Apriori algorithm.