Assignment 3 - Association Rule Mining

Instructions:

For the Third Assignment, we are going to be performing association rule mining with the Apriori and FP-Growth algorithms.

There are a set of questions associated with each task for Apriori and FP-Growth, respectively. The dataset used for the entire assignment is the same.

Task A: 5 points

- Print details on the data.
- Check for unique items and modify them if any misfit names are found.

Task B: 7 points

- Create a list of lists by grouping the data based on member number and transaction date, and store it in a dataframe.

Task C: 7 points

- Transform the data to have unique products as column names and indications of the product in each transaction as 0 or 1.

Task D: 7 points

- Apply the Apriori algorithm with a support threshold of 0.001.

Task E: 5 points

- Print the top 5 association rules based on the lift metric and a threshold of 1, according to the support value.

Task F: 3 points

- Store the association rules from top to bottom, sorted by the lift parameter, and store them in a variable.

Task G: 10 points

- Define a function that takes the sorted association rules and a product name as inputs.
- The function should return the top recommendation list for that product name.

- The recommended items are the consequents that match the antecedents of the product being searched for.

Task H: 5 points

- Print the top 5 recommended items for the product 'sausage'.

Task I: 5 points

- Print the top 5 recommended items for the product 'sausage' with a support threshold of 0.005.

Task J: 10 points

- Perform tasks D through I, except for task G, using the FP-Growth algorithm.

Task K: 8 points

- Print the top 3 recommendations for the following products using the Apriori algorithm with a support threshold of 0.0015:
    - citrus fruit
    - whole milk
    - soda

Task L: 8 points

- Print the top 3 recommendations for the following products using the FP-Growth algorithm with a support threshold of 0.0015:
    - canned beer
    - yogurt
    - pastry

Task M: 20 points

- Submit a document showing your solution for the given problem statement.

Programming Assignment Details:

- You can use the following libraries: Pandas, Apriori, FP-Growth.
- Explain each task and its actions in Markdown format.
- Comment your code.
- If using any external resources (books, internet), cite them within the cell.
- Do not rename the dataset files.

Submission Details:

- Fill in the name and ID of each group member in the Jupyter notebook in the provided format.
- Name your submission files using the following format: yourLastName_Last4digitsofyourID.ipynb (e.g., smith_1234_assignment3.ipynb).
- Only one team member should submit the fi