1 Instructions

1. You can use any programming language of your choice with the suitable input format for each of the question. Each question should have a readme file and the detailed report.

2. A readme file should precisely tell how to compile and run your program. Give the exact commands with respect to the datasets provided.

3. The marks will be given on the basis of quality of code, use of innovative data structures, scalability, correctness, and completeness of the report.

4. You are supposed to submit the assignment on google classroom no later than 5th **October 2020**. This is a strict deadline and any assignment submitted later will not be consider for evaluation.

Problem 1:

Implement the three frequent itemset mining algorithms: [30 Marks]

1. Apriori Algorithm

2. FP-growth

3. Eclat algorithm

Compare the performance of each algorithm on the below datasets:

1. T10I4D100K (.gz) (http://fimi.uantwerpen.be/data/)

2. T40I10D100K (.gz) (http://fimi.uantwerpen.be/data/)

3. retail.gz (http://fimi.uantwerpen.be/data/)

4. https://www.kaggle.com/irfanasrullah/groceries

Prepare a detailed report containing the following:

- Details of the algorithms implemented. Did you use any optimization techniques on the top of the original algorithm. If yes, explain them.

- Give a brief description of each of the datasets such as how many transactions it had, average width of the transaction, size of maximal frequent itemsets, number of maximal frequent itemsets, total number of items, etc.

- Comparison of the three algorithms in terms of time and space complexity on different datasets.

- Explain why you think a particular algorithm works better in one dataset as compared to other.

.

2 Problem 2:

Implement a program that would generate your own dataset based on varioius parameters such as number of transactions, size of maximal frequent itemset, average width of the transactions, and total number of items m as input. Compare the three algorithms based on different parameters on the dataset generated by you. Write a detailed report explaining the algorithm implemented and the comparison results. [15 Marks]