

**Frequent Item set Mining**

**ASSIGNMENT NO: 2**

**DATA MINING**

**MEMBERS NAME \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**MEMBER REGISTRATION NO \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**TITLE: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Please read the following and below instructions:

1. Please provide a complete solution.
2. Plagiarism or Cheating is not allowed.  **Even if you cheat in only one part, you will be marked zero in the *entire* assignment.**
3. Maximum member allowed in group is 5 and minimum 1
4. Please do this assignment by yourself and participation of each member is compulsory .As this is a research based assignment try to research and do this assignment for learning not just for marks. If you face any difficulty regarding understanding or doing assignment you may email me your problem.
5. viva and 1 quiz will be expected from this assignment
6. Submit your assignment on LMS session 10.
7. If you are new to python I have compile and attached a pdf file contain all python commands with their syntax for your help.

**QUESTION NO 1:**

Your task for this assignment is to implement and evaluate the Apriori-based algorithm for frequent item sets mining.

1.      Implement the Apriori algorithm for frequent item sets mining. You can also find the pseudocode and its related procedures from the lecture slides and textbook.  You are encouraged to use any existing or your own optimization techniques for the Apriori algorithm.  If you do, explain and discuss the techniques you have used and/or provide the appropriate references in the report.

You can use any programming language that you are familiar with prefer python.

The program should be executable with 3 parameters: the name of the input dataset file, the threshold of minimum support count, and the name of the output file.  The minimum support count should be an integer.  An item set is frequent if its support count is larger or equal to this threshold.

The program should output a file that contains all the frequent item sets together with their support.  The output file ([sample output](http://www.mathcs.emory.edu/~lxiong/cs570/assignment1/test.out)) should have the following format: each line contains a single frequent itemset as a list of items separated by whitespace.  At the end of the line, its support is printed between a pair of parenthesis.  For example: 1 2 3 (6) represents an itemset containing items 1, 2 and 3 with a support count of 6.

2.      Test your implementation on the dataset of your own choice and measure execution time as well as number of frequent item sets with various minimum support values.    You can also try your program with various other [frequent item set mining datasets](http://fimi.cs.helsinki.fi/data/).

3.      Write a summarize report in PDF presenting your results on the test dataset and other datasets if you have tried.  Explain and discuss, if any, the algorithmic optimizations you have used in your implementation.  Discuss the experiences and lessons you have learned from the implementation.

4.      You can work as a team of up to four or you work on your own.  If you work as a team of more than one person, please explain the contribution of each team member in your report.

5.      Your submission should be a zip or tar file that contains the PDF report as well as the program deliverables including your source files, the executable, a readme file explaining how to compile/run your program.

**DECLARATION:**

I am aware of and understand the University’s policy on plagiarism and I certify that this assignment is my own work**,**except where indicated by referencing, and that I have followed the good academic practices noted above

Signed \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Find single frequent item grater than min support

Find double frequent items

Find three than so on