CSE 5334 - DATA MINING

Spring 2017 – Assignment 4
Due: 11:59pm Central Time, Thursday March 23, 2017

REQUIREMENTS:

Read the following requirements carefully, and make sure you follow every rule. If you fail to meet any of requirements, marks will be deducted accordingly.

- The project document describes how you tackle the problem
  - No limitation on the number of pages, or the format of the document;
  - The document should clearly describe how you design and implement. **(5 points)**
- Submit your source code and project document
  - ONE zip file contains all source codes and the project document. **(5 points)**
  - You have to implement the whole project by yourself.
  - You can use any open source library to handle retail.dat file.
  - Your source code must pass compilation. Any non-executable submission is not acceptable.
  - Your submission must EXCLUDE the input file. **(2 points)**
- Execution
  - Your program has to be executed using command line as followed: **(8 points)**

```
python your_script_file.py retail.dat min_sup_value min_conf_value
```

THE PROBLEM:

In this assignment, you will experiment with association rule mining using Apriori algorithm.

The data set you will use is based on a retail market basket data obtained from an anonymous Belgian retail store. First 10 transactions in the data set is as follow:

```
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
30 31 32
33 34 35
36 37 38 39 40 41 42 43 44 45 46
38 39 47 48
38 39 48 49 50 51 52 53 54 55 56 57 58
32 41 59 60 61 62
3 39 48
63 64 65 66 67 68
32 69
```

TASKS :

Your tasks in this assignment are the following:

a) Load the retail market basket data into an appropriate data structure. **(10 points)**
b) Run Apriori on the retail data using a user-defined min-support value. **(35 points)**
c) Generate rules from the frequent itemsets discovered in the previous step using a user-defined min-confidence value. **(35 points)**