Assignment 1

Association Mining

Requirements:

* Using the programming language you prefer, write a program to implements one of the association algorithms (Apriori, FP-Growth or vertical data format) for one of the following problems.
* If the student solves both problems, we will consider the higher mark.
* Your program should include a graphical user interface.
* Your program should be implemented under the following specifications:
  1. Inputs:
  2. A file with a set of transactions (Excel, text, etc…). (Hint. The file attached).
  3. The percentage of the data needed to be read from the input file.
  4. Minimum support count.
  5. Minimum confidence (percentage value).
  6. Outputs

1. Generate all frequent item sets which can be mined from the transactions.
2. The strong association rules that satisfy both the minimum support and minimum confidence threshold.
   1. General Requirements:
3. The program should enable the user to select the file needed to be analyzed,
4. The interface should enable user to select the percentage of the data needed to be read from the input file e.g. if the file contains 100 records, and the user needs to read 70% of the file then the analysis should be done on 70 records only.
   1. Instructions:
5. Assignment should be done individually; copies will be graded to -5.
6. Total grade is 5 marks.
7. No late submissions are allowed.
8. Discussion will be held with Eng. Dina Amr during the office hours:

* Monday 4, November @ 2.30-4
* Tuesday 5, November @ 2.30-4

1. If you have any ambiguity, please do not hesitate to ask your TA.

**Problem 1. Online Retail Problem:**

This dataset is transformed from the Online Retail real dataset. It contains 541,909 transactions and 2603 items. The meaning of each item is given in the file OnlineRetailAttributes.xlsx. Each row represents a transaction for one customer.



**Problem 2. Special Colored Cars Problem:**

A company decided to produce two types of cars (Family and Sports) with new customized colors. The file contains transactions of the best mix between colors and car type according to one-year sales. The program will be then used to detect the best color for each type to be able to produce the new model. The file contains 245,057 transactions and 11 items. The meaning of each item is given in the file ColorAttributes.xlsx.

