###### **Data Mining**

**Information Systems Department**

**Faculty of Computers and Information**

###### **Cairo University**

**Assignment 1**

**Association Rules**

**Instructions:**

1. Assignment should be done individually; copies will be graded to zero.
2. You should choose only 1 problem out of the 3 given problems to solve.
3. Total grade is 5 marks.
4. No late submissions are allowed.
5. The submission will be on classroom.
6. Deadline will be on 8/4 until 11:55 pm

**Problem 1**

**Description:**

* Consider the attached file “hotel\_booking.csv”.
* It contains the number of rows collected by a hotel (each record represents a customer reservation per a specific date).
* We need to know the association between customer\_type, the country where he/she is from and his/her reserved\_room\_type.
* For example: customer who are from Egypt “EGY”, and reserves as a “Transient”, will reserve a room of type “B”.

**Requirements:**

* You should get all permutations of the associations between the 3 columns.
* Write a program in any programming language that implements one of the association algorithms (Apriori, FP-Growth or vertical data format) on this dataset.
* Minimum support & minimum confidence should be variable as per user input during runtime.
* Then generate all association rules which can be mined from the transactions.
* The final output of your program should show the frequent item sets and association rules with their confidence.

**Problem 2**

**Description:**

* Consider the sets of transactions for a coffee shop in the attached file "CoffeeShopTransactions.xlsx".
  + The file contains transactions of items for a coffeeshop.
  + Each row represents a transaction at a specific time.
  + Each transaction contains a set of items defined with their names.
  + Each item is in a separate column.

**Requirements:**

* Write a program in any programming language that implements one of the association algorithms (Apriori, FP-Growth or vertical data format) on this set of transactions.
* Minimum support & minimum confidence should be variable as per user input during runtime.
* Then generate allassociation rules which can be mined from the transactions.
* The final output of your program should show the frequent item sets and association rules with their confidence.

**Problem 3**

**Description:**

* Consider the attached file "categories.txt".
* The provided input file consists of the category lists of 77,185 places in the US. Each line corresponds to the category list of one place, where the list consists of a number of category instances (e.g., hotels, restaurants, etc.) that are separated by semicolons.
  + An example line is provided below:

Local Services; IT Services & Computer Repair

In the example above, the corresponding place has two category instances: “Local Services” and “IT Services & Computer Repair”.

* Each line corresponds to the category list of one place.
* Write a program in any programming language that implements one of the association algorithms on this set of transactions.
* Minimum support & minimum confidence should be variable as per user input during runtime.
* Then generate allassociation rules which can be mined from the transactions.
* The final output of your program should show the frequent item sets and association rules with their confidence.