

**First mid-term/Written exam 2/11/2016****First mid-term: deliver solutions to ex. 1, 2 within 3 h****Written exam: deliver solutions to ex. 1, 2, 3, 4 within 4 h****Notice:** use your own SQL Server credentials (the lbi account is disabled)

**Exercise 1 (8 pts).** Consider the `foodmart` database. The *dissimilarity index* of year 1998 is defined as:

$$D = \frac{1}{2} \sum_{i=1}^n \left| \frac{f_i}{F} - \frac{m_i}{M} \right|$$

where  $n$  is the number of stores, and, for a store id  $i$ :

- $f_i$  is the number of distinct female customers who made at least one purchase in the store  $i$  during 1998;
- $m_i$  is the number of distinct male customers who made at least one purchase in the store  $i$  during 1998;
- $F = \sum_{i=1}^n f_i$  is the sum of the  $f_i$ 's;
- $M = \sum_{i=1}^n m_i$  is the sum of the  $m_i$ 's.

Write a Java program `Dissimilarity.java` which outputs such the value  $D$ . The Java program can submit only SQL queries of the form “SELECT \* FROM table”.

**What to deliver:** `Dissimilarity.java`, `myJDBCdef.props` (with only the parameters needed for a test of the program).

**Exercise 2 (8 pts).** Develop a SSIS package that outputs on a CSV file the result of Ex. 1. The usage of GROUP BY / WHERE / ORDER BY clauses in SQL queries to perform computation at server side is not permitted. All the work must be done by the SSIS package.

**What to deliver:** SSDT solution.

**Exercise 3 (8 pts).** Write a SQL query with analytic functions or, at your choice, a MDX query that solves the problem of Ex. 1.

**What to deliver:** text file with SQL/MDX query and with a brief comment about, a screenshot of SQL Management Studio with query result.

**Exercise 4 (8 pts).** Consider the problem of predicting **the number** of customers who will made at least one purchase next year. Model the problem as a classification problem. Use SQL plus Weka Explorer, or Weka Knowledge Flow or Weka API at your choice for experimenting your solution.

**What to deliver:** screenshots of SQL Management Studio plus either a Weka knowledge flow `.kfml` file or a PowerPoint file with screenshots of Weka explorer or a Java program with Weka API calls, and a description of the steps of the designed solution.

**How to deliver:** send an e-mail with a single `<your surname>.zip` file attached to `ruggieri@di.unipi.it`, including your name, surname, student ID, and computer IP address (`http://www.whatismyip.com`).

**Results and oral exam.** Results will be published on-line by this week. Oral exam dates (for the “Appello straordinario”) will be emailed directly to you.