

Written test 18/06/2019

Deliver Part I (Ex.1 & Ex.2) or Part II (Ex.3 - Ex.5) within 2 h

Deliver solutions for all exercises within 4 h

Notice: use your own SQL Server credentials (the lbi account is disabled)

Exercise 1 (8 pts). Consider the database `foodmart`. Develop a python program `ratio.py` that for each product and store computes:

- a) the number of distinct females that bought the product in that store
- b) the number of distinct males that bought the product in that store
- c) the total store sale generated by the product in that store in January 1998
- d) the total store sale generated by the product in that store in July 1998
- f) the ratio between a) and b)
- g) the ratio between c) and d)

The result must be stored in a `.csv` file. The python program can submit only SQL queries of the form “SELECT * FROM table”. The usage of PANDAS library is not allowed.

What to deliver: `ratio.py` and a `.csv` file with the results.

Exercise 2 (8 pts). Develop a SSIS package solving Exercise 1. No SQL query on data sources is allowed.

What to deliver: SSDT solution.

Exercise 3 (8 pts).

Answer the following business questions using **MDX** over the *Sales* cube of `ruggieri_foodmart` project:

- (a) For every product category and store city, the distinct customers who bought the product category in January 1998, the distinct customers who bought the product category in June 1998 and their ratio
- (b) for every product family and store, the id and profit of the customer with the highest profit for that product family in that store.

What to deliver: (1) Power Point file with the MDX queries and results and with a brief comment about them; (2) text file with the MDX queries.

Exercise 4 (2 pts). Answer the business question of Exercise 3 with **SQL** over the `foodmart` datawarehouse. Use analytic functions as needed.

What to deliver: (1) Power Point file with SQL queries and results and with a brief comment about them; (2) text file with SQL queries.

Exercise 5 (6 pts). Let A be an index describing the tendency of a product to be bought during the summer (July, August, September). How could this index be computed considering the information stored in the database `foodmart`? Compute this index for each product and describe and discuss a data mining approach for the prediction of A . Provide the description of each attribute and its computation for each product.

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 Azure ML workflow and all the python scripts used) 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 **SUBJECT:LDS-June** with a single <your surname>.zip file attached to **roberto.pellungrini@gmail.com** including your name, surname, student ID, and computer IP address (<http://www.whatismyip.com>).

Results and oral exam. Results will be emailed to the students shortly, including the date and time for those who are admitted to the oral exam.