

Written test 06/2/2019

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

Exercise 1 (8 pts). Consider the database `foodmart`. The *Healthy sales delta* of a product is the difference in sales for *low fat* products between year 1997 and 1998, considering only the purchases made by female customers. Produce a CSV file with three columns: *product_id*, *Healthy sales delta*, *mean Healthy sales delta*. We are interested in selecting only the products that have an *Healthy sales delta* greater than the *mean Healthy sales delta* of all products, ordered by descending *Healthy sales delta*. Develop a Python program `Report.py` that solves the problem and produces a CSV file with the results. The python program can submit only SQL queries of the form “SELECT * FROM table”. The usage of PANDAS library is not permitted.

What to deliver: `Report.py` and CSV file.

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, percentage of male customers that generated a profit for that product category over the total number of male customers buying that product category in that country;
- (b) for every product family and store province, the id and profit of the customer with the highest profit for that product family in that province.

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

Exercise 4 (2 pts). Answer the business question of Exercise 1 with **SQL with analytic functions** over the `foodmart` datawarehouse.

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 C be the number of distinct customers that bought a given product category in a given month and store. Design a data mining approach predicting the value of C for a store, year and month number given only information available at the end of previous month.

What to deliver: screenshots of SQL Management Studio plus either a Weka knowledge flow `.kfm` 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 - Feb** with a single `<your surname>.zip` file attached to `annam@di.unipi.it` including your name, surname, student ID, and computer IP address (<http://www.whatismyip.com>).