Applied A.I. Solutions
Foundations of Data Management

Lab Exercises 1, 2, 3

Read the Project first. The Project represents the final work objective. It provides a framework and the context you need to answer the exercises correctly.

Lab Exercise 1 - Due Friday, February 10th, 2023, at 10:00 PM (10 points max)

1. **Analysis:** Basic analysis of **Sample Superstore** spreadsheet. Identify and document the level of completeness, inconsistencies, redundancies, duplicates. Sample Superstore is the data source you will use to generate your reports (available on Blackboard under Course Materials)

2. **Target Audience**: use assumptions to define your target audience for both, the operational and executive reports. Explain the intended use of each report (monitor and control, decision-making, performance improvement, research analysis), what would be achievable by using these reports.

3. **Context and additional Assumptions**: describe the context in which your reports will be created, maintained, and used, and any additional assumptions you need to develop the operational report and the executive report.

4. **Operational and Executive Reports**: **define** and document the information each report will display (sales, geography, customers, revenue, custom KPIs, etc.). If required, you may incorporate additional data including ad' hoc calculations (formulas/algorithms). Document and explain each formula, KPIs and ratio that you will use in the reports. Produce a document with this information.

5. Once you have identified the reports' objectives and the data that each report must show, proceed to **design empty templates,** one for the Operational Report and one for the Executive Report, indicating titles, column titles, date/period, units (examples below).

Submission A: documents produced above under points #1,2,3,4,5 - submit a single document as:

Group#_1_A.pdf

Important: the data source contains controlled errors. In exercise 1, DO NOT MODIFY THE SPREADSHEET – You will clean and change data in exercise 2.

Lab Exercise 2 - Due Wednesday, February 15th, 2023, at 10:00 PM (10 points max)

1. **Analysis:** Advanced analysis of **Sample Superstore** spreadsheet. Identify and document entities, attributes, domains, referential integrity.

2. **Logical-level ERD:** create an ERD indicating entities, attributes, relationships, primary and foreign keys, and Master Data tables. Include a Metadata table, an optionally a Reference Data table. **Note:** you can create new entities and attributes if needed. Use standard notation (**crowfoot**) and any tools such as Lucidchart, Visio, SQL Workbench

3. **Data Flow:** create a data flow indicating data in/out to/from the corresponding processes and the connections among them. Use any tools such as Visio, Lucidchart

4. **Data Cleansing**: clean the data source for further use and data analysis. If needed, create new entities and attributes. (Use Excel). Document every change (field change, update, deletion, override, etc.) performed to the spreadsheet.

5. **Installation**: install **MySQL Workbench** in your computer as per the instruction indicated in this project document

6. **Database Creation**: using your ERD as foundation to create a physical MySQL database "**GBC_Superstore**", create tables and fields and add the primary and foreign keys. From MySQL Workbench, generate a printout of the corresponding database schema.

7. **ETL:** Extract, Transform and Load data from the already cleaned **GBC_Superstore.xlsx to MySQL GBC_Superstore** tables:
   a. generate and export CSV file(s) into MySQL database/tables using SQL scripts and/or use Python to ETL Data from Excel to MySQL tables.
   b. document the ETL process used for the generation of the MySQL GBC_Superstore database.
   c. verify data completeness, data integrity and referential integrity.

8. **Submission A:**
   a. printout of the Logical-level ERD
   b. printout of the Data Flow
   c. printout of the database schema (include all tables, fields)
   d. document with any data source changes
   e. document explaining the ETL process.

   Submit one pdf document as **Group#_2_A.pdf**

   Note: you must submit all SQL scripts, CSV files, and/or Python code used for ETL, MySQL dump/export at the end of the project.

Lab Exercise 3 - Due Monday, February 20th, 2023, at 10:00 PM (10 points max)

Produce the Operational and Executive Reports (program and run SQL scripts)

1. **SQL**: write the SQL sentences necessary to generate both, the operational and executive reports

2. **Report Generation**: run the SQL scripts and **produce the Operational and Executive reports**. These reports should show some similarity with the templates designed in Exercise 1.

   Reports must present data coming from a quality-developed MySQL database. Calculations, KPIs, Ratios, etc., must be checked before submission.

3. **Submission A**: **Final Operational and Executive reports**: **Group#_3_A.pdf**

(*) Note:

Students who know SQL can share their knowledge with those who do not. Also, those students who do not have SQL experience may read SQL basics and self-study. We will use MySQL again in the upcoming Data Visualization Techniques course.

MySQL installation

- Install MySQL V 8.X.X Win64 (x86_64) (MYSQL Community Server – GPL)
- Install MySQL Workbench 8.X - V 8.X Community

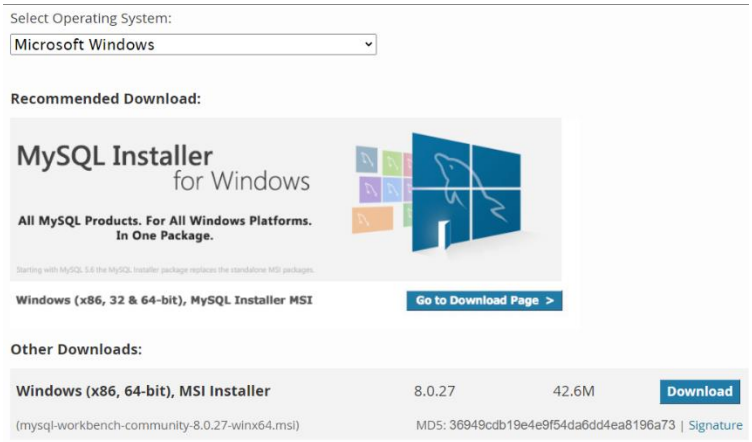
https://dev.mysql.com/downloads/workbench/

or

https://dev.mysql.com/downloads/

MySQL parameters

- Server: localhost
- Username: root
- Password: 1234
- Port 3306



Important

At the end of this project, **DO NOT** uninstall the software from your computers. Write down installation parameters, user IDs, passwords for later use. We will use these tools and deliverables again during the Data Visualization Techniques course.

Excerpt from the College Policy on Academic Dishonesty:

The minimal consequence for submitting a plagiarized, purchased, contracted, or in any manner inappropriately negotiated or falsified assignment, test, essay, project, or any evaluated material will be a grade of zero on that material.

To view George Brown College policies please go to www.georgebrown.ca/policies