Assignment #1

CSCI 5408 (Data Management, Warehousing, Analytics)
Faculty of Computer Science, Dalhousie University

Date Given: Sep 22, 2022
Due Date: Oct 03, 2022 at 11:59 pm

Late Submissions are not accepted and will result in scoring "0" in the assignment.

Disclaimer: This assignment requires students to work on various research and open Datasets with appropriate citation. Submissions related to this assignment will not be used for commercial purposes.

Objective:

- The objective of this assignment is to understand industry problems related to data capture, and database design. Create entity relationship model and perform normalization of the database.

Plagiarism Policy:

- This assignment is an individual task. Collaboration of any type amounts to a violation of the academic integrity policy and will be reported to the AIO.
- Content cannot be copied verbatim from any source(s). Please understand the concept and write in your own words. In addition, cite the actual source. Failing to do so will be considered as plagiarism and/or cheating.
- The Dalhousie Academic Integrity policy applies to all material submitted as part of this course. Please understand the policy, which is available at: https://www.dal.ca/dept/university_secretariat/academic-integrity.html

Assignment Rubric

	Excellent (25%)	Proficient (15%)	Marginal (5%)	Unacceptable (0%)	Rubric applied to
Completeness including Citation	All required tasks are completed	Submission highlights tasks completion. However, missed some tasks in between, which created a disconnection	Some tasks are completed, which are disjoint in nature.	Incorrect and irrelevant	Problem #1 Problem #2
Correctness	All parts of the given tasks are correct	Most of the given tasks are correct However, some	Most of the given tasks are incorrect. The	Incorrect and unacceptable	Problem #1 Problem #2

		portions need minor modifications	submission requires major modifications.		
Novelty	The submission contains novel contribution in key segments, which is a clear indication of application knowledge	The submission lacks novel contributions. There are some evidences of novelty, however, it is not significant	The submission does not contain novel contributions. However, there is an evidence of some effort	There is no novelty	Problem #1 Problem #2
Clarity	The written or graphical materials, and developed applications provide a clear picture of the concept, and highlights the clarity	The written or graphical materials and developed applications do not show clear picture of the concept. There is room for improvement	The written or graphical materials, and developed applications fail to prove the clarity. Background knowledge is needed	Failed to prove the clarity. Need proper background knowledge to perform the tasks	Problem #1 Problem #2

Citation:
McKinney, B. (2018). The impact of program-wide discussion board grading rubrics on students' and faculty satisfaction. Online Learning, 22(2), 289-299.

Explanation of the rubric:
Suppose you received different grades in Clarity for the 2 problems
Problem #1: 25% in clarity
Problem #2: 15% in clarity
Then your overall grade for the clarity will be avg of $(25+15)\ \% = 20\%$

Problem #1: Building knowledgebase on Dalhousie University.

Establishing relationships and identifying key entities, attributes are the main requirements of this project. The problem is not well defined; however, the client trusts your expertise, and believe that you will provide an enhanced data modelling for the conceptual phase. In addition, the client wants you to address conceptual design to table structure mapping and normalization/denormalization requirements for the logical design phase. Apart from the design, and the table structure, the client wants your assumptions in the form of one-page written report.

To obtain information on various collected data, you can visit: dal.ca and the subdomains of Dalhousie University. (Please do not perform web data scrapping; you need to visit the website and apply your critical thinking to understand the concepts and identify the entity sets, relationships, and cardinality.)

Submission Expectations: Report, ERD/EERD, Normalization/Denormalization, SQL Dump of Table structure and values.

Problem #2: Format Ocean Tracking Data and Report

Dalhousie Ocean Research wants you to explore the dataset they provided, and perform the following:

- Read the document available at http://oceantrackingnetwork.org/about/#oceanmonitoring
- Write a report on what are the different datasets, and attributes you discovered.
- Clean and transform the dataset, e.g.
    a. remove NULL values
    b. rearrange the columns
    c. transform the data in a column or attribute if required to fit a common format
    d. Is there a possibility of combining some of the tables or attributes without losing information? If yes, please perform the task and report your findings.
    e. Is there a possibility of decomposing some of the tables without losing information? If yes, please perform the task and report your findings.

- Based on the given dataset, create relational schema using MySQL DBMS
- Using MySQL Workbench and reverse engineering create the possible ERD. Your report must contain the ERD produced by the reverse engineering. In addition, you need to add the cardinality.
- Populate the database with clean and transformed dataset

Submission Expectations: Report(s), ERD/EERD, Normalization/Denormalization, SQL Dump of Table structure and values. If you write a script/program to clean/format data, then upload your program code to gitlab (https://git.cs.dal.ca)