DS230 Data-Centric Computing
Assignment 4

Milestone
11/17/21

OBJECTIVES

- Refine experience with programmatic database interactions.

ASSIGNMENT

I. Database Creation for Class Schedule

A. The file "F21 Course Schedule.csv" contains the data associated with the courses that were offered in Fall of 21. Open this file and study its structure.
B. Create some code that creates a database table – SCHEDULE – that is consistent with the data contained in this file
    a. You do not need to infer the structure by reading the file – it is ok for you to have a literal create query.
C. Create some code that is able to read this CSV file and populate your Schedule table.
    a. You may find it beneficial to "pre-process" the CSV file. You are allowed to make any changes that you want/need to this file; however, those changes should be made programmatically. (For example, it may be beneficial to remove some the "Headings" such as "Computer Science,,,,,,," --- do not do this by hand). The idea is that this file will come to you next semester in the same format --- your program should be able to take this file and populate the DB

II. Simple queries with Schedule.

A. Provide functions that are able to ask the following queries:
    a. What courses are available from a given department? (department is provided by an input parameter to your function)
    b. What courses are available at a given time-block?
    c. What courses are available to satisfy my Social Sciences gen Ed requirement (ANT, ECO, POL, PSY, SOC).
    d. What courses are available to satisfy my DCP (ends with 6, 7, 8). You should group the courses according to which category they meet. (Note - you may need to use the SQL 'SUBSTRING' command.)
    e. Which course are offered by my favorite professor (No – you can't use the literal 'Hughes' in your query).

III. Course Registration

A. Expand your database to include a STUDENT table and an ENROLLMENT table.
    a. The Student table should be defined by (<u>STUDENT ID</u>, Student Number, Fname, Lname, ClassYear, Major1, Major2, Minor1, Advisor)
    b. The Enrollment table should be defined by (<u>STUDENT ID</u>, <u>COURSE ID</u>, status); Status is either 'Active' or "WaitList'

These tables should be created programmatically.

B. Create at least 5 students and enroll these students into 4 courses each. (They should not all have the same schedule.

IV. Anonymize and load student data

A. The file "F21 Registration.csv" contains the data associated with the course registrations for Fall of 2021. Open this file and study its structure.
B. This file contains sensitive information that you don't want to become part of your production-level database. Namely, it is associating real Student Numbers with real student names. We need to create aliases for these names and numbers --- however, you should preserve the relationships. This means that if student: (1837273, Beth, Travers) was taking (CS125, MTH125, ARTH 100 and POL 107), her alias (1872822, Carol, Ubben) should also be taking those same courses.
C. After your data has been anonymized, create a function that is capable of adding this enrollment data to the database.

V. (more) Complex queries

TBD