CPSC 368: Databases in Data Science
Completing the data analysis

Assignment Description

In this assignment, you and your team will complete the data analysis and visualization portion of your project by using the department provided Oracle database and programming it via Python.

The first step in this process is to populate your Oracle database with your project data. You must transfer the .sql file you created in the previous assignment and run it in Oracle before proceeding. If you are having trouble with this step, refer to the SQL Plus tutorial.

If you are having trouble with connecting to the department database, please refer to this page.

These are some expectations of your project:

1. The project topic and analysis should match the content of the previous assignments. If a change is required, you must discuss this with a member of the teaching team **at least three business days before the assignment deadline**. We want to ensure that your project still meets expectations and will need time to evaluate proposed changes. Please do not make changes prior to receiving approval.

2. The project should fetch data from the department provided Oracle database to use.
    a. For some projects, it may be the case that there is so much data it exceeds the quota limit imposed by the department. Please talk to someone from the teaching team if you are in this position.

3. Code quality
    a. The code should be well documented and follow basic programming principles such as good naming, documentation, general one task per function, etc. The code does not need to strictly follow rules from CPSC 103 or 110; generally speaking, we do **not** want to see unstructured code as that makes it hard to follow and understand your work.
    b. Code that was not written by a member of the team needs to be **clearly indicated and cited**. When looking at your code we should be able to clearly tell which parts were written by a member of the team.

4. Code functionality
    a. You are allowed to use external libraries to perform your data analysis/visualization.
    b. The code should work for any instance of the dataset(s). That is, if we were to replace one or more of your datasets with a file that has the same structure, the code can run and produce a visualization based on the new data.

Deliverables

The teaching team warmly recommends that you look at the rubric on Canvas before submission.

1. A short summary of your project. This will help your TA remember which project they are looking at. You can reuse text from previous assignments.

2. The SQL DDL file that can be run to populate the database.

3. A README file. This is a file that details some things that the teaching team should know about before running your code.
    a. The teaching team may choose to populate their own database with the SQL DDL file or they may choose to use the Oracle account of someone on the team. State which student's account we can use.

4. The data files

5. All the code files used to conduct your data analysis and visualization.