CS486/586 Introduction to Databases
Project - Milestone #1

Every student has to work individually on the project.

Database Project Implementation:

- The goal of this project is to provide real world experience working with database schema design, complex queries built on real world large databases and basic understanding of python connectivity with postgreSQL.
- After completion of this project, you will get hands-on experience with database schema design, PostgreSQL queries, python, dataframes, database connectivity.

There will be 4 milestones as a part of project submission, for this week you only need to work on the first milestone which is the project proposal.

- **Milestone #1: Project proposal (Due date - 2nd week)**
- **Milestone #2: Mid project (Due date - 6th week)**
- **Milestone #3: Progress report (Due date - 7th week)**
- **Milestone #4: Final project (Due date - 8th week)**

Milestone #1, Project proposal (Due date - 2nd week):

Select a subject area on which you wish to build a database. Write approximately one paragraph that gives a general description / background information on that subject area. Describe what source you intend to use for data, and how you intend to ingest the data into your database. You should choose a subject area where you can easily get several hundred rows of data. Your dataset should be diverse so you can build an interesting database on.

Your database should include 6-10 tables.

- Use multiple CSV files to create tables
- Some resources for reference
    - https://data.world/datasets/csv
    - https://www.stats.govt.nz/large-datasets/csv-files-for-download/

What to turn in:

- Description about your topic area.
- Description of your collected datasets.
- A brief description of how you plan to populate your database.
- 1 to 2 pages of documentation in PDF form.

Note: that you are allowed to change small changes after submitting your work here but you need to have done this step prior to moving on to the next steps.

Grading

Submission Milestone #1 is worth 20 points. Grading will be based on:

- Completion of each point described in Submission requirement for this milestone (10 points)
- We are looking for realistic datasets and scenarios; suitability and realistic-ness of data source (5 points)
- The ideal data source demonstrates enough complexity to get 6-10 tables and inclusion of a variety of data types and attributes (5 points)