601.315 Databases, Spring 2022
Project Phase A: Domain Proposal

Due: Thu, 17 Feb at 11pm. Use of late days *is NOT* permitted.

The long-term project in this course will give you the opportunity to work in pairs to design and implement a moderately large database system in a target domain. You and your partner (registered as in Piazza @38) may select a domain particularly relevant to your outside interests, such as an investment portfolio database, an astronomy database, a medical database, a sports statistics database, etc.

For a future phase of this project, you'll need to locate real-world data related to your domain to use in your system. Some sources for inspiration as you select your domain are listed below. This is by no means an exhaustive list; you are encouraged to search for additional sources relevant to your interests.

- World Heath Organization Data - https://www.who.int/data/gho
- United Nations Population Data - https://www.un.org/development/desa/pd/
- Pew Research Center Datasets - https://www.pewresearch.org/internet/datasets/
- Amazon Web Services Open Data - https://registry.opendata.aws/
- United States Government Open Data - https://www.data.gov/
- Baltimore City Open GIS Data - https://data.baltimorecity.gov/
- World Bank Open Data - https://data.worldbank.org/
- Stanford Open Policing Project - https://openpolicing.stanford.edu/
- Kaggle.com Datasets - https://www.kaggle.com/datasets
- IMDb Movie Data - https://www.imdb.com/interfaces/
- Sports Reference Data https://www.sports-reference.com/
- Fivethirtyeight.com - https://data.fivethirtyeight.com/
- JHU CSSE COVID-19 Case Data https://github.com/CSSEGISandData/COVID-19

Work with your partner to decide on a possible domain. Make sure the data available in your selected domain is rich enough to allow you to answer complex questions about it. *Consider creating a system that combines data that may not seem directly related; perhaps you'll uncover unexpected connections this way.* Then complete the sections described below in a single document:

1. **Partners.** List the full names and JHEDs of each partner working on this project.
2. **Domain.** In one short paragraph, briefly describe the proposed target domain for your project.
3. **Questions.** List a minimum of 20 questions you might like to ask of a database system in your proposed domain. Aim to build a comprehensive set of questions; the more complex the questions, the better. (If you can't think of interesting questions, re-consider your domain choice.) Questions must be expressed in English, not relational algebra or SQL. For example, *"What is the mean literacy rate for countries with a per capita income of $400 per year, grouped by continent?"* This list will help uncover basic objectives to focus your later design choices. In a later phase of the project, you'll formulate many of these as queries in SQL to be executed on your system.
4. **Data sources.** Give URLS for 2-3 potential sources of real-world data you could use to populate your proposed database. At this time, you don't have to know every source you'll use, but show that you've investigated enough to know that relevant data is accessible to you. In a later phase of the project, you'll download and format all the data you'll need to populate your system.

Submit your work as a PDF via Gradescope. One partner will submit the work as a team submission upload, and will indicate both partner names. Therefore, only one partner should submit.