

IS 362 – Project 2

The goal of this assignment is to give you practice in preparing different datasets for downstream analysis work.

Your task is to:

- (1) Choose any three of the “wide” datasets identified in the Week 5 Discussion items. (You may use your own dataset; please don’t use my Sample Post dataset, since that was used in your Project 1 assignment!) **For each of the three chosen datasets:**
 - Create a .CSV file (or optionally, a MySQL database!) that includes all of the information included in the dataset. You’re encouraged to use a “wide” structure similar to how the information appears in the discussion item, so that you can practice tidying and transformations as described below.
 - Read the information from your .CSV file into Python, and use pandas as needed to tidy and transform your data. [Most of your grade will be based on this step!]
 - Perform the analysis requested in the discussion item.
 - Your code should be in a Jupyter Notebook, posted to your GitHub repository, and should include narrative descriptions of your data cleanup work, analysis, and conclusions.
- (2) Please include the URL for the Jupyter Notebook in your homework submission, **for each of the three chosen datasets**. You may create a single Jupyter Notebook, or three separate notebooks.