
Project Phase 1 - Data Collection and Data Cleaning

In the final project for this course, you will apply the techniques learned in this class to analyze a data sets of interest to you. Your goal should be to create an original project that you would be proud to show off to a potential employer. The project will require you to collect and obtain several data-sets that can be correlated in someways, clean, integrate and apply EDA on the data, perform analysis using ML techniques, and then present your findings by building a dashboard for the intended audience.

The project is divided into several phases. In the first phase (Data Acquisition), we ask you to collect a dataset through web-scraping. You are asked on build a simple web-crawler in Python to collect data of interest. You are also asked to obtain more than one dataset, the additional datasets can either be obtained via web-crawling, APIs, downloading online, etc. The datasets used must be sufficiently large and non-trivial. You can consult with me if you have questions regarding this point.

This submission should identify and describe the datasets that must be obtained. Contain code for web-crawling to gather a dataset, and additional code for obtaining secondary datasets through APIs or other means.

0.1 Submission

Submit your jupyter notebook to iLearn. Note, your notebook file should be named **team-name_proj_phase1**. You should include comments at the beginning of the notebook that includes your name and student id. Submissions will not be accepted via email; you must turn your assignment via iLearn.