Capstone Project 1 Proposal

The upstream of oil and gas industry is a combination of geophysics, geology and engineering, and the complicated and varied nature of the underground formations makes the industry a mixture of science and art. Therefore people must be careful when applying previous experience to new projects through data science analysis.

Well completion is one of the key considerations when oil and gas companies design a well, and it may determine the success of the whole project since well completion affects the efficiency of oil and gas production in the future. This project will focus on analyzing well completion strategies targeting various formations, comparing the injection, production and incident data, to provide oil and gas companies with useful reference information to make a good well completion design and to enhance oil and gas recovery ratios.

For this capstone project, I'm using the latest version of well and production data collected by the Oil Conservation Division, which regulates oil, gas, and geothermal activity in New Mexico. This dataset is public and free to download.
(ftp://164.64.106.6/Public/OCD/OCD%20Interface%20v1.1/)

My preliminary plan is this:

1. Download the datasets, unzip, and convert these XML files (about 65GB) to flat files (CSVs)
2. Identify the interesting tables and columns for this project, and deal with data wrangling and cleaning
3. Explore the data for interesting correlations between well completion design elements and oil and gas production.
4. Report results in a Jupyter Notebook containing codes, graphics and texts, and a slide deck.