Requirements: see requirements.txt

File 1: Extract Data from Text
File Name: fetching_dataset1.py fetching_dataset2.py
Description: This Python script extracts structured information from unstructured text data. It uses regular expressions to search for patterns and extract data such as API numbers, well names, well numbers, latitude, longitude, address, and stimulation data. The extracted data is stored in a Python dictionary.

File 2: Combine and Clean Data
File Name: combine_clean_data.ipynb
Description: This Jupyter Notebook file combines and cleans data from multiple CSV files. It reads data from several CSV files, combines them into a single DataFrame, and performs data cleaning tasks such as filling missing values in the "WELL NAME" column with "WELL NUMBER." The final cleaned data is saved to a CSV file.

File 3: Scrape Data from Web
File Name: drainedge.ipynb
Description: This Jupyter Notebook file scrapes information from the "drillingedge.com" website using Selenium and BeautifulSoup. It extracts data such as well status, well type, closest city, and well values (303 and 2k) for a list of API numbers and well names. The scraped data is stored in a CSV file.

File 4: Drop Columns and Rename Columns
File Name: drop_clean_data.ipynb
Description: This Jupyter Notebook file loads a combined CSV file, drops certain columns, and renames columns for clarity. It also changes column names to conform to a specific naming convention. The updated data is saved to the same CSV file.

Challenges
Data Extraction Challenges
Regular Expression Complexity:
Web Scraping Challenges
Dependency Management
File Management and I/O
Data Transformation and Renaming
Documentation and Code Organization