

DSCI 560 - Data Science Professional Practicum
Lab 5 Part 1

Team name: Cleanup Crew

Team members: Jonathan Kusuma (ID:4897950942), James Susanto (ID:3466104463)

Minutes:

February 20: We researched the pdfs and did research on how to effectively create regular expressions to filter out the important data.

February 21: We created the PDF extraction code and implemented the regex. We also optimized the regex to capture all the data and filter out the wrong data from the OCR and available text in the PyPDF.

February 22: We implemented selenium to scrape the web data in drilling edge. We used the API numbers and selected the first result to obtain the necessary data.

February 23: We optimized the selenium and formatted the data scraped from the internet correctly. E.g. 2.7 k of gas is equal to 2700.

February 24: We finalized our code and created the demo video.