**Scraper for the following website:**

Python version: 3.9  
Selenium version: 4.10 or below  
Website scraped: Surviving Antidepressants

[**https://www.survivingantidepressants.org/**](https://www.survivingantidepressants.org/)

The scraper requires a root directory which has all the links containing discussion threads for a particular area. The scraper is created to record every single URL and iterate through it. It scrapes the content and the date of each comment in discussion threads and adds them to a list. Eventually, this is transformed to a dataframe and added to the main dataframe.

Languages and frameworks used include python, selenium and pandas. Selenium is used to automate the scraping process. Pandas was used to store all the data so that future data analysis could be performed.

The scraper is made up of 3 functions which work together to scrape the entire directory. The generate\_url function scrapes every single URL in the root URL provided. This function simply returns a list containing URLs as strings.

The scraper also makes use of the scrape\_a\_page function which returns a pandas DataFrame object. This function is used in the scrape\_a\_site function multiple times to scrape a whole discussion thread. Since a site is made up of various pages, At the end, it collates all comments, and dates scraped to create a dataframe which gets returned. This dataframe contains all records of a particular discussion post/thread.