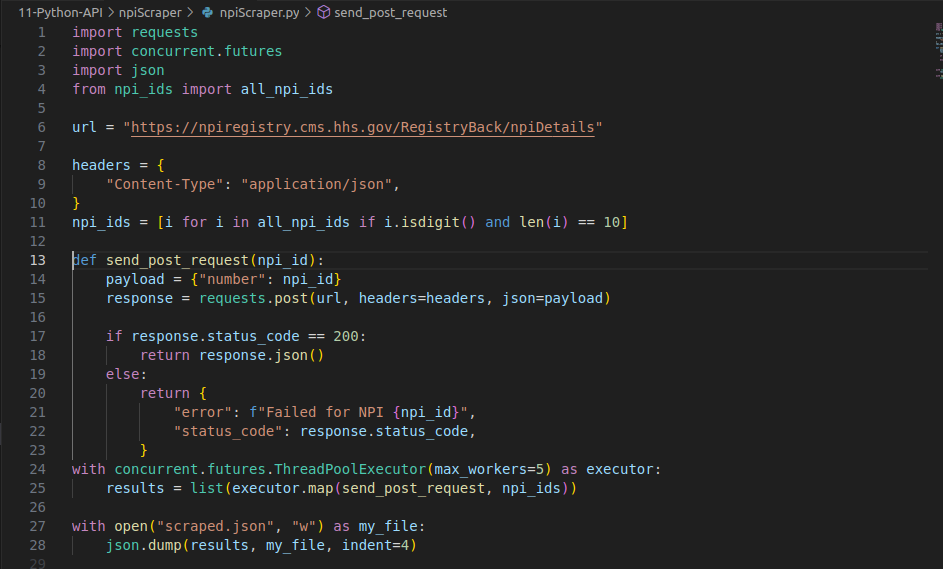
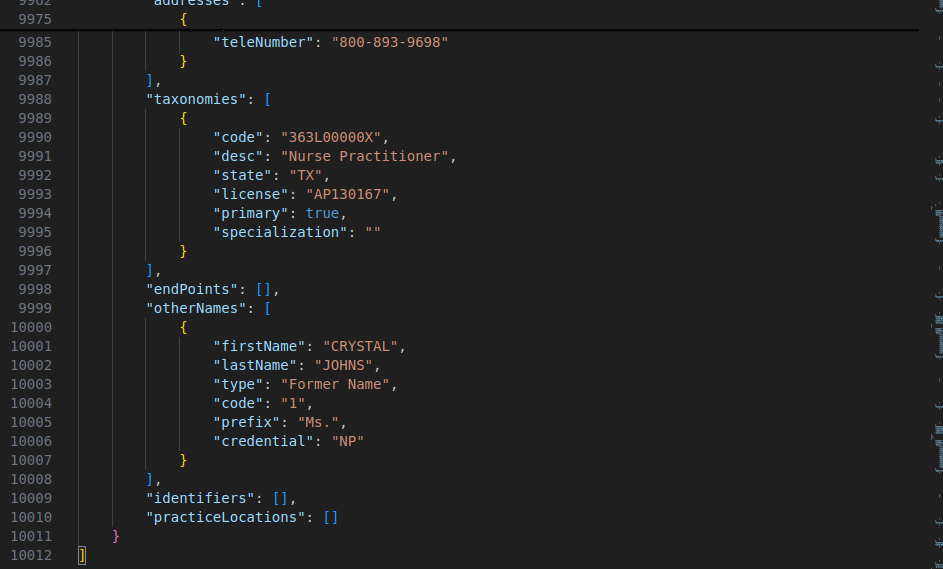
Assignment (Multi-Threading)

1. **Implement multi-threading in NPI Scraper script.**



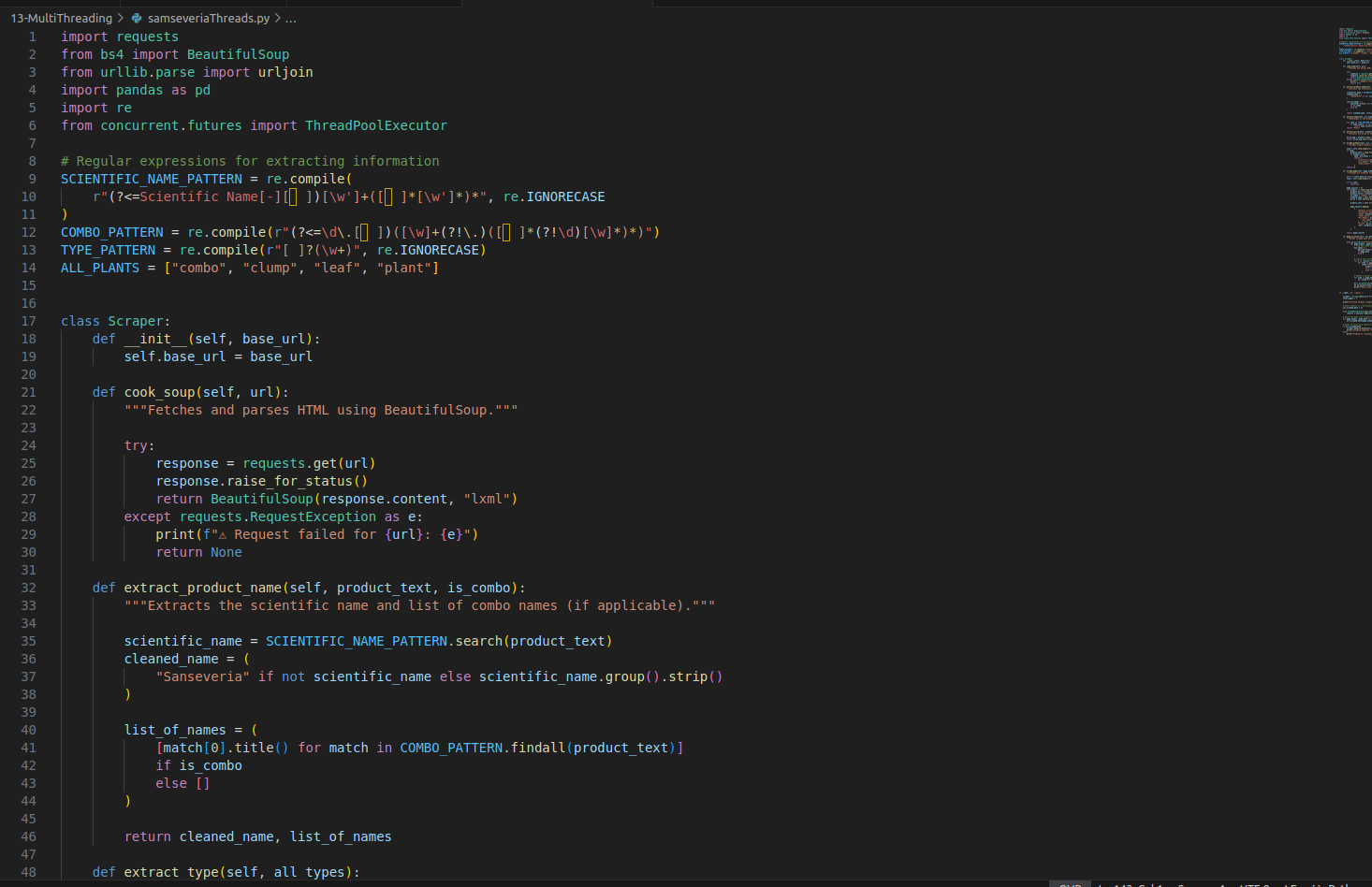
Code 1/1

I used ThreadPoolExecutor to implement multi-threading by taking 5 threads as max, which I later experimented with and finally settled with 25. This made my script more efficient by scraping all the data from 3 minutes to 6 seconds as previously it waited for each id to finish being scraped and being added to the list, here a batch of 25 ids were simultaneously being scraped and as soon as a thread got free, it took the next iterable from npi\_ids list and started sending a post request there.

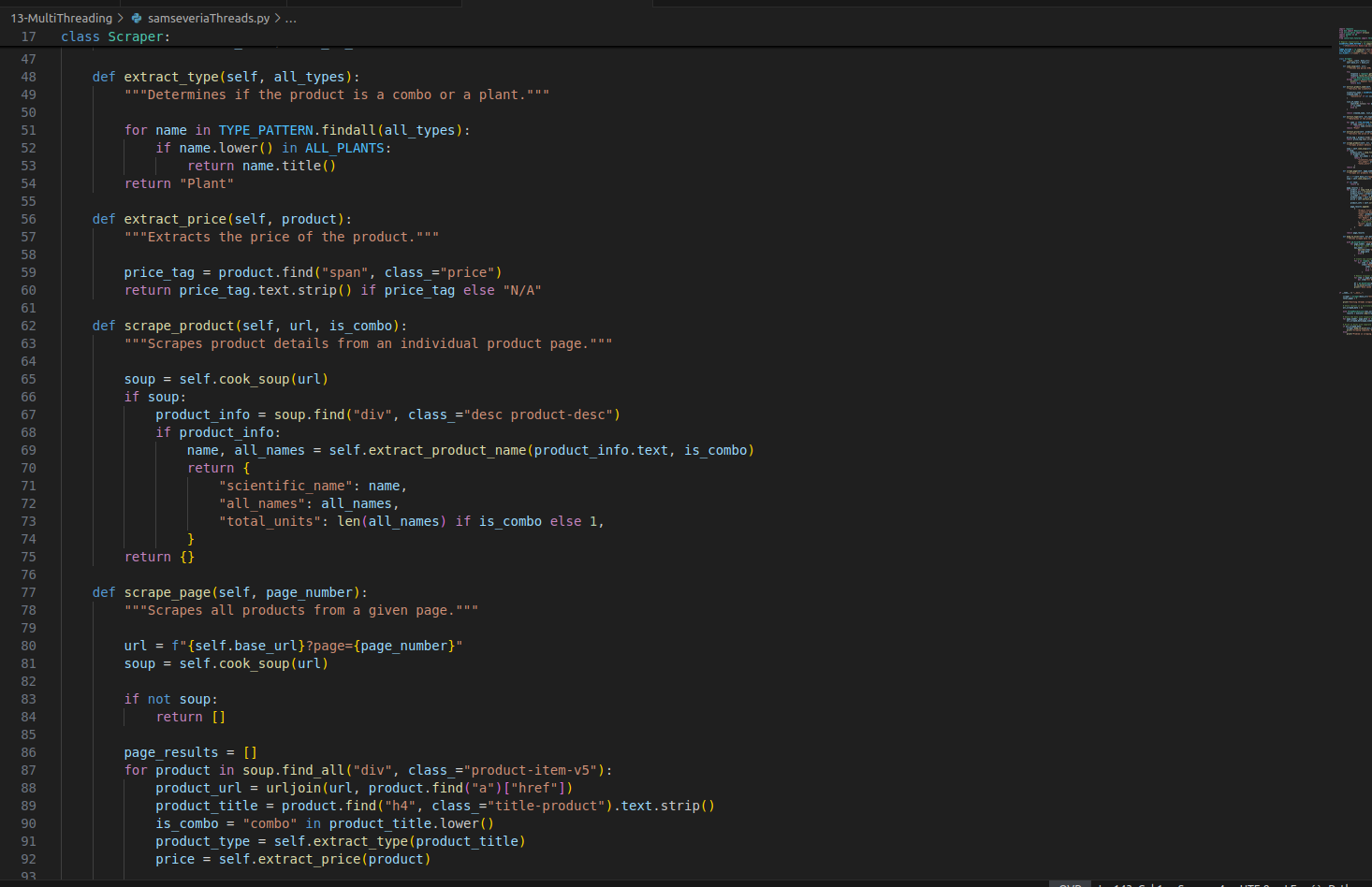


Data was scraped and dumped into a json file with over 10012 lines from the post requests.

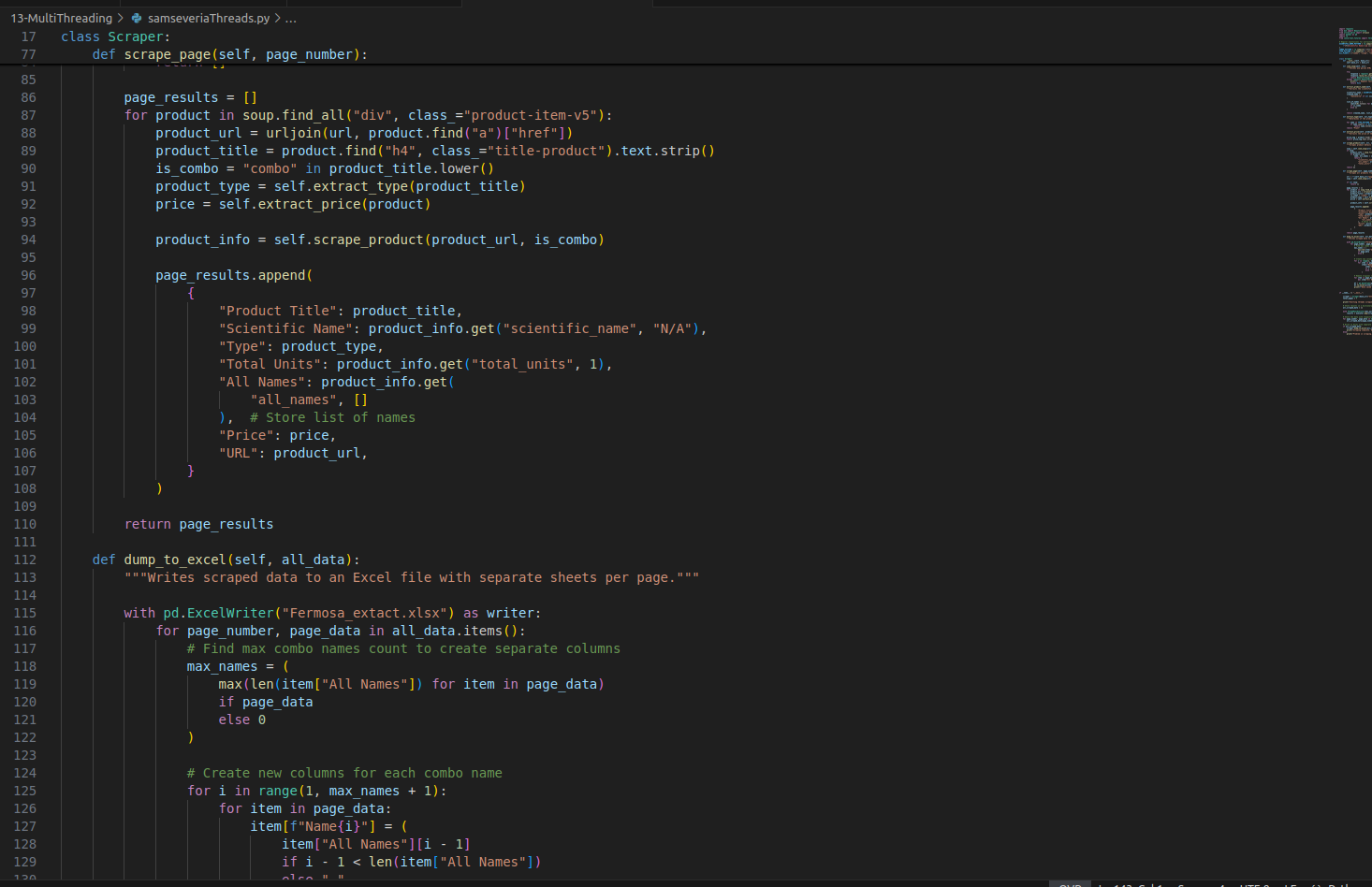
**Q2. Implement Fermosa Scraper with multi-threading**



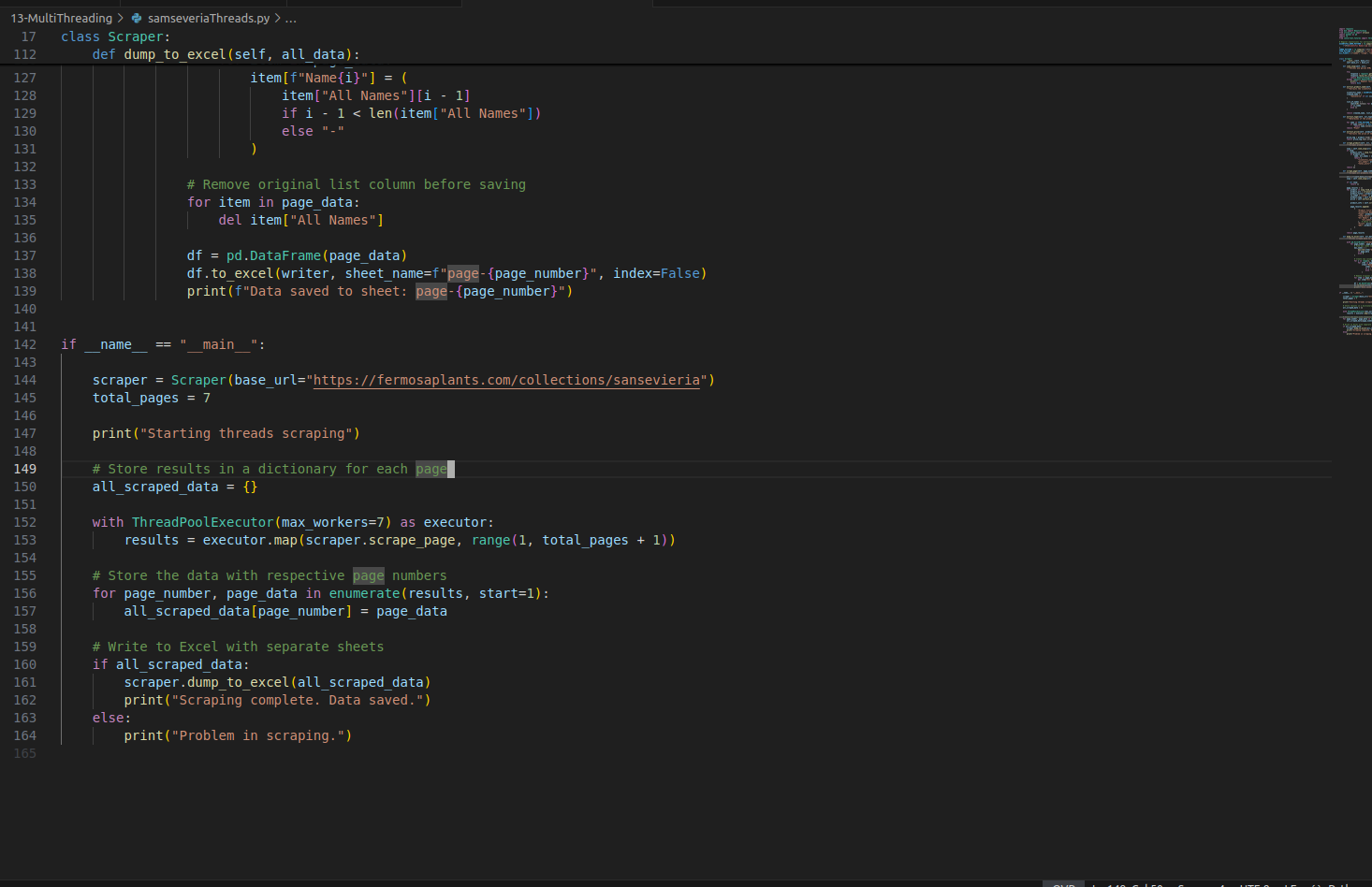
Code 1/4



Code 2/4

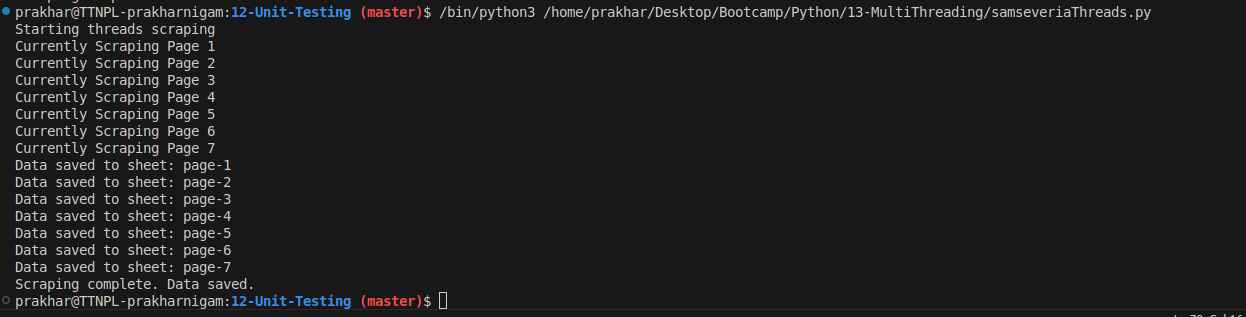


Code 3/4



Code 4/4

I used threadpoolExecutor to create 7 worker as there were 7 pages to scrape, i called scraper\_page method and gave a range in iterator to scrape each page concurrently and store the output in result page and then called the dump\_to\_excel method to dump it. This significantly reduced the time from scraping each page first then dumping it and then going to next one to directly scraping all 7 at once.



Output.