

The Python Scraper application is a Scrapy based web scrapy that retrieves an Items MSRP, name, and price at each store. The entire project only took a few hours however I had repeated issues with VS Code activating the environment Scrapy was installed in.

Day 1 (4hrs): To start, I researched different web scrapers to use for the project and settled on scrapy due to the speed of the API. After creating a new environment, installing scrapy, and sorting out the installation issues; I auto generate a scrapy project, adding a spider, and a main class. I wrote the main class to get an items UPC code from the user, and then construct and start a spider before printing the results. Then I constructed the spider; Having the first parse function find the item from a search page of the UPC, and the second collect the data and write it to a file.

Day 2 (3hrs): On the second day I added the ability to choose what website to scrape for information, between Walmart and Home Depot. To do this I wrote a function that returned a class definition object from the selected store and UPC that is passed to a CreateProcess object to start scraping.

In conclusion the project took about 7 hours to learn scrapy and code a scraper for these websites, and showcases my ability to quickly learn new libraries.