

Using Scrapy to List Data Analyst Jobs in West Coast Cities

I was naturally interested in consolidating job titles, locations and URL's of jobs in various west coast city centers. The general idea was to write this information into CSV's for further analysis. The best tutorial for this task was found at <https://python.gotrained.com/scrapy-tutorial-web-scraping-craigslist/>. The code was up to date unlike many repositories of web scraping code on Github. The code needed only small updates in order to specialize the job search to Craigslist job listings with either data or analyst in the title or description. I prefer a wide spectrum approach rather than keeping only results that specifically use the phrase "Data Analyst". I will run through the CSV's by hand and analyze the results. There is too high a chance that a good offbeat written job description would be accidentally screened out.

The general idea is to consider the tag as the proper wrapper of the job description and basic information. The class="result-info" gives us the general jobs info wrapper that I will sift into. I loop through this wrapper to parse and save the info on individual jobs. The address of the job is found under the class="result-hood". To extract the URL, I refer to the <a> tag and use the value of the '@href' attribute. This extracts a relative URL. To create an absolute URL, I use the 'urljoin' function to join with the response variable containing the base domain. This data is put into a dictionary and output into a CSV file.

The bar graph below shows the number of responses found in various west coast job centers.

