4.0 hours

Project 6

Build a Content Scraper

Why can't I submit this project yet?

You must first complete all prior Techdegree content before you can submit this project. Additionally, you cannot submit more than one project at a time.

Submit for Peer Review

Your current activity is [HTML Tables](#).

- [Instructions](#)
- [How you'll be graded](#)

To get a passing grade you'll need to submit a project that meets the following requirements. If your project receives a "Needs Work" rating on any of these tasks, then you'll need to edit your project and re-submit it until it "Meets" or "Exceeds" each task requirement.

	Needs Work	**Meets Expectations**	**Exceeds Expectations**
Command Line Interface and Dependency Management	- Project doesn't include a `package.json` file containing relevant dependencies to run `node scraper.js`. - Running `npm install` does not install relevant dependencies	- Project includes a package.json file containing dependencies to run `node scraper.js` - Running `npm install` installs relevant dependencies	The `scraper.js` file can be run with the `npm start` command
Folder Generation	- Program doesn't check for a 'data' folder and/or fails to create a 'data' folder if one doesn't exist - When the program runs a second time, a duplicate 'data' folder is created or the program throws errors because the 'data' folder already exists	Program creates one `data` folder if that folder doesn't already exist. If the folder does exist, the program does nothing.	N/A

	Needs Work	**Meets Expectations**	**Exceeds Expectations**
Research npm packages	• Chosen scraping and CSV packages don't meet the following requirements on npm: • 1,000 downloads • Updated in the last 6 months	• Chosen scraping and CSV packages meet the following requirements on npm: • 1,000 downloads • Updated in the last 6 months	N/A
Crawling	• 'scraper.js' uses hardcoded URLs to find all the shirts on the page, or finds less than 8 shirts	• The project uses the http://shirts4mike.com/shirts.php URL as an entry point to look through the links on the page to find 8 shirts	• N/A
Scraping and Saving Data	• Project doesn't scrape all of the required information: product title, price, image and url, or information is incorrect or in the wrong place • CSV is not successfully saved to the 'data' folder in this format: 'YYYY-MM-DD.csv', e.g. '2016-12-30.csv' • Column headers are not in this order: Title, Price, ImageURL, URL, Time	• Project scrapes the product title, price, image and url, and all information is correct and in the correct place • A CSV is successfully saved to the 'data' folder in this format: 'YYYY-MM-DD.csv', e.g. '2016-12-30.csv'. • Column headers are in this order: Title, Price, ImageURL, URL, Time	• N/A
Overwriting Data	• If the script is run twice, the program files to overwrite the data. Instead an error occurs, or the script appends data to the existing CSV file.	• If the script is run twice, the program overwrites the data. The file contains the data from the second call.	• N/A

	Needs Work	**Meets Expectations**	**Exceeds Expectations**
Error Handling	• The program does not display an error in the console when it is unable to connect to http://shirts4mike.com	• The program displays a human-friendly error (not just the original error code) when it cannot connect to http://shirts4mike.com	• Program logs errors in a scraper-error.log` file.  • New errors append to the end of the file with a timestamp, e.g. `[Tue Feb 16 2016 13:00:55 GMT-0800 (PST)] <error message>`