

Daltix Scraping Test

Instructions

- Open the terminal
- Start your virtual environment with `python 3`
- Install scrapy `pip install scrapy`
- Go to the root folder of the project
- In your terminal and type `scrapy crawl product_spider`
- A file called `products.csv` will appear, wait until the crawling is done and check the file.

Optional Instructions for the Duplicate Value Tester

- Open the terminal
- Start your virtual environment with `python 3`
- Install scrapy `pip install pandas`
- Go to the root folder of the project
- In your terminal and type `python duplicate_tester.py`

Keep in Mind

The `duplicate_tester.py` will show the sum of duplicate url results in the `products.csv` file.

The results will be fewer since only the results with promotional prices are being scraped as requested in the test.

If you wanna change the name of the csv file where the results are being collected. Go to `settings.py` and change the `FEED_URI` name as well as in the `duplicate_tester.py`, on the 3rd line

Thank you for the opportunity

Looking forward to hear back from you

Aderito Xavier