

# Web Scraping Info Sheet

## How do Web Scrapers Work?

Automated web scrapers work in a rather simple but also complex way. After all, websites are built for humans to understand, not machines.

First, the web scraper will be given one or more URLs to load before scraping. The scraper then loads the entire HTML code for the page in question. More advanced scrapers will render the entire website, including CSS and Javascript elements.

Then the scraper will either extract all the data on the page or specific data selected by the user before the project is run.

Ideally, the user will go through the process of selecting the specific data they want from the page. For example, you might want to scrape an Amazon product page for prices and models but are not necessarily interested in product reviews.

Lastly, the web scraper will output all the data that has been collected into a format that is more useful to the user.

Most web scrapers will output data to a CSV or Excel spreadsheet, while more advanced scrapers will support other formats such as JSON which can be used for an API.

## Uses for Web Scraping

### Retail and Manufacturing

- Competitor Price Monitoring
- Monitoring MAP Compliance
- Fetching Images and Product Descriptions
- Monitoring Consumer Sentiment

### Equity and Financial Research

- Aggregated News Articles
- Market Data Aggregation
- Extracting Financial Statement
- Insurance

### Data Science

- Real-Time Analytics
- Predictive Analysis
- Natural Language Processing
- Machine Learning Training Models

## Product, Marketing and Sales

- Data-driven Marketing
- Content Marketing
- Lead Generation
- Competitive Analysis
- SEO Monitoring
- Reputation Monitoring

## Other Industries

- News and Reputation Monitoring
- Academic
- Data Journalism
- Non-Profit
- Employment
- Search engine for classified sites

## Methods of Web Scraping and Resources



### JavaScript

Difficulty: Beginner

Video Tutorial:

[https://www.youtube.com/watch?v=TzZ3YOUhCxo&ab\\_channel=AaronJack](https://www.youtube.com/watch?v=TzZ3YOUhCxo&ab_channel=AaronJack)

Article Tutorial: <https://medium.com/@bretcameron/how-to-build-a-web-scraper-using-javascript-11d7cd9f77f2>

Open Position: <https://www.economicmodeling.com/open-positions>



### Python

Difficulty: Intermediate

Video Tutorial:

[https://www.youtube.com/watch?v=XVv6mJpFOb0&ab\\_channel=freeCodeCamp.org](https://www.youtube.com/watch?v=XVv6mJpFOb0&ab_channel=freeCodeCamp.org)

Article Tutorial: <https://realpython.com/beautiful-soup-web-scraper-python/>

Open Position: <https://mch.applicantstack.com/x/openings>



Difficulty: Advanced

Video Tutorial:

[https://www.youtube.com/watch?v=NwtxrbqE2Gc&ab\\_channel=DataScienceDojo](https://www.youtube.com/watch?v=NwtxrbqE2Gc&ab_channel=DataScienceDojo)

Article Tutorial: <https://www.freecodecamp.org/news/an-introduction-to-web-scraping-using-r-40284110c848/>

Open Position: <https://employment.ku.edu/job-search/staff>