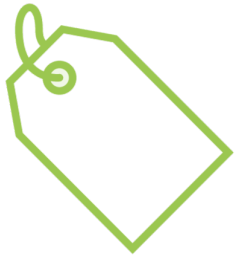
Web Scraping Wrap Up



Clarke Bishop
BIG DATA ENGINEER

@ClarkeBishop www.clarkebishop.com

Web Scraping Applications



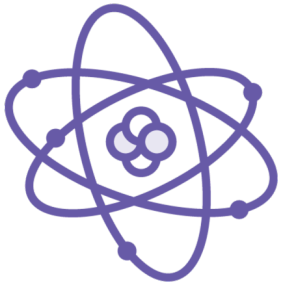
Price Analysis



Lead Generation



Keyword Research



Science



Product Research



Find a Job



Oversight



Stock Analysis



Risk Management



News Gathering

Scraping Up a Tesla



Web Request



Request & Response

HTTP

- URL
- Verb - GET or POST
- User Agent

URL Hacking

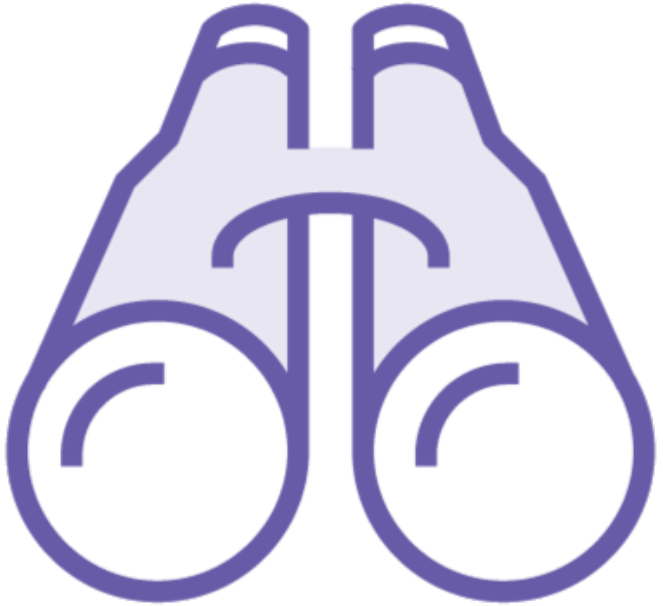
Parse & Extract

SQL for web scraping

CSS & HTML

XPath

Chrome developer tools





Basic Web Scraping



Beautiful Soup

Requests - Download web pages

Beautiful Soup - Extract data

Jupyter lab

Next steps

- Play with Chrome dev tools
- Practice selectors
- Learn requests-html

Scrapy

Libraries versus frameworks

Scrapy CLI

- Setup a scraping project
- Run a crawler
- Refine selectors with scrapy shell

Scrapy - Build spiders

Next steps

- Write a spider
- Explore scrapy-splash



Advanced Web Scraping



Requests-HTML

The JavaScript rendering problem

Requests-HTML

Selenium

Next steps

- Try out Pyppeteer and compare with Selenium
- Learn Requests-HTML or Scrapy

Thank You!

Twitter: @ClarkeBishop

LinkedIn: linkedin.com/in/clarkebishop/

Web: www.clarkebishop.com

