

# Week 6 Assignment



This week, we will learn how to web scrape data, and use it to scrape data and store in a database.

## Written Assignment

Submit up to a 2-page written report addressing the following:

- What are the ethics of web scraping? How do we know when it's allowed vs not allowed? What are possible legal consequences of unauthorized web scraping?
- What are some of the technologies we can use for web scraping?

Useful resources:

- <https://resources.distilnetworks.com/all-blog-posts/is-web-scraping-illegal-depends-on-what-the-meaning-of-the-word-is-is>
- <https://quickleft.com/blog/is-web-scraping-ethical/>

## Demo

### Project Discussion

Scrape news stories from the Denver government news. [Walk Through Video](#) ; also see the `video_demo_denver_news_scrape.ipynb` file and the `Cleaner Denver news example.ipynb` file.

## Advanced Demo

See this [Video](#) and the `Advanced_scraping_example_selenium.ipynb` file.

Scrape MISO [Day-Ahead pricing](#) and put in to SQL database. For MISO, they actually have a file showing the HTML paths to download relevant datasets, so we don't have to use selenium or advanced ways to download the data.

Scrape [Mizuna menu](#).

# Assignment

---

Choose another site to scrape, and save the results in a database. Anything beyond the first site scraped will count as a bonus. You can use either a SQL or NoSQL database to store the data. Ideas for sites to scrape:

- (easiest) --> [CO government news](#)
- [Fed Reserve Site](#)
- [CO DOT Travel Alerts](#)
- A restaurant menu that changes seasonally or more often, like [The Preservary](#), [Mizuna](#), [Tag](#), [Mercantile](#), or a restaurant local to you that you're interested in
- Another MISO dataset (e.g. [historical hourly wind data](#) or [something else](#))

## Bonuses

---

- Extract the data from csv or Excel files (for something like MISO) or PDFs (like restaurant menus) and put these into the database.
- Use a croniob or a loop to check the site daily for new data files.

 Reflect in ePortfolio

 Download

 Print



### Activity Details

Task: View this topic