Total Data Quality Specialization Course 1; Week 1 Web Scraping Demonstration
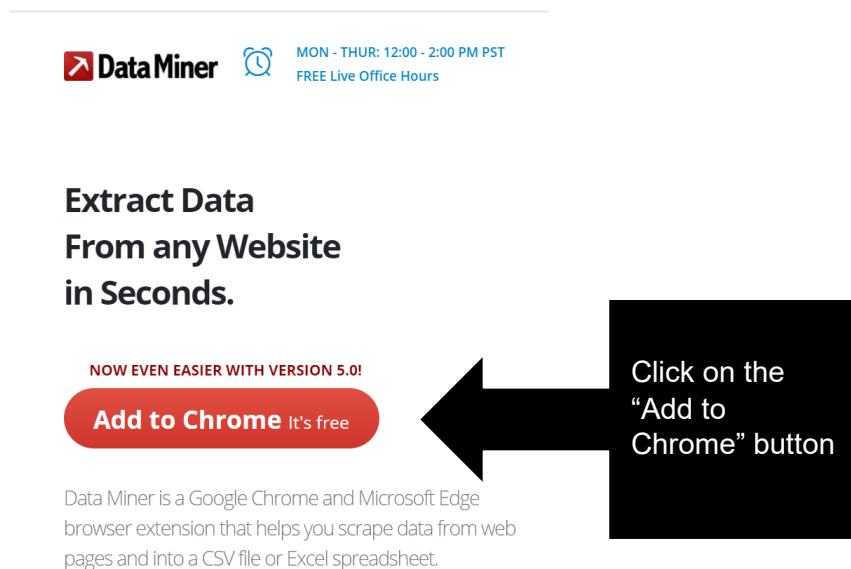
Overview:

In this demonstration we will walk you through a simple and an intermediate web scraping example using the free online tool offered by Data Miner. The tool is free to use for web scraping tasks and allows the user to customize web scraping tasks to a certain extent. There are more advanced platforms such as Octoparse and Import.IO that allow the user to customize web scraping tasks and editing of captured data. For even more advanced web scraping tasks there are also libraries in Python, such as BeautifulSoup or packages in R such as rvest. We will focus on a tool that can be accessed and embedded directly into your web browser for user convenience.

In our demonstration we will specifically deploy the Data Scraper Chrome Web Browser extension offered by Data Miner.   There is also a Data Scraper extension that can be added to Microsoft Edge as well.
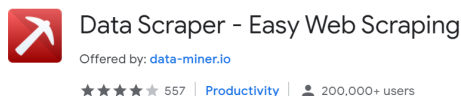
Installing the Data Scraper Extension from Data Miner

1. From your Google Chrome (or Microsoft Edge) browser, navigate to: dataminer.io.
2. Click on the "Add to Chrome" button as illustrated below.

3. The Data Miner web page will take you to a chrome web store (or Microsoft store) page and will prompt you to add the Data Scraper extention to your Browser. Click on "Add to Chrome" as shown below.
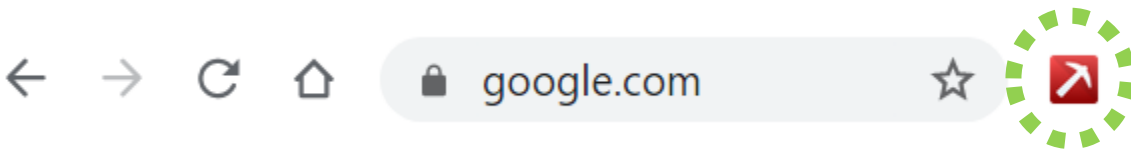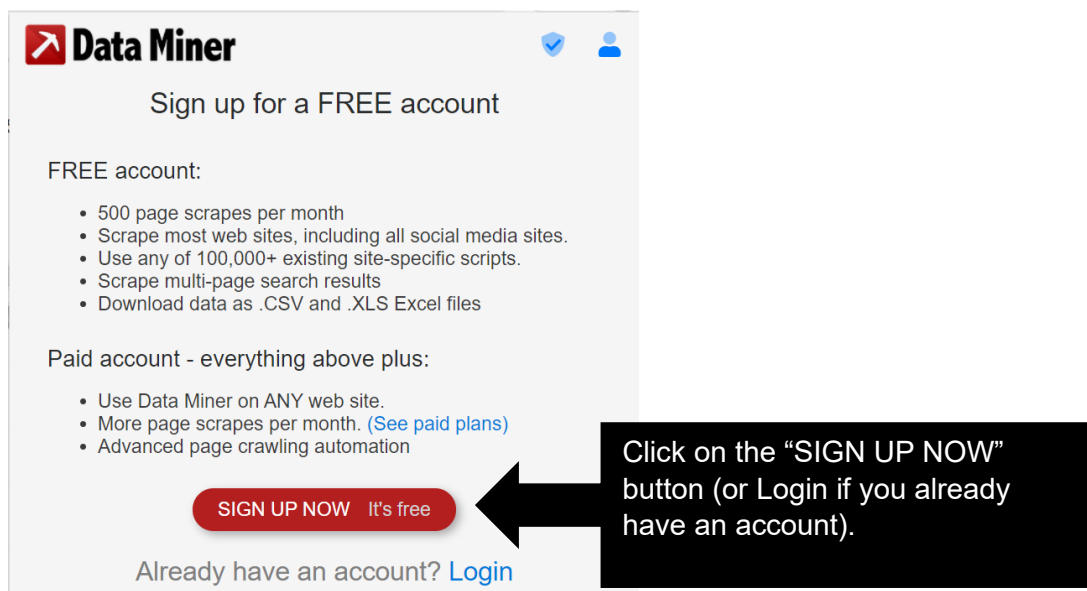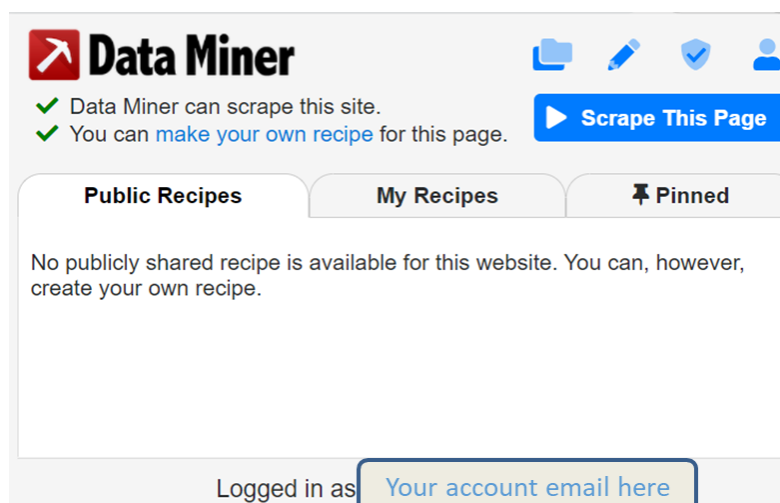
4. Once you have successfully added the extension, you should see the Data Scraper Icon to the right of the search bar in chrome as indicated below.

5. Once you have installed the extension, click on the Data Scraper icon (⬈) in your browser and you will be prompted to log in or to sign up (if you haven't already established an account). It is free to sign up for the account so you should click the "SIGN UP NOW" button as shown below to establish an account that will enable you to begin to use the Data Scraper Extension.

6. Once you have created an account you should see the basic Data Scraper Interface for Data Miner once you click on the Data Scraper icon ⬈ in your browser – if you see a pane like the one shown below things are working properly and you are ready to begin your scraping tasks.

Example websites we will scrape in this Demonstration:

A. Simple table scraping of a list of counties in the state of Ohio
B. Customizing a web scraping task for a list of counties in Ohio

Example websites you can try on your own:

- Scrape a table of counties in the state of Iowa similar to what we did in Example A for Ohio.