**11-28 notes**

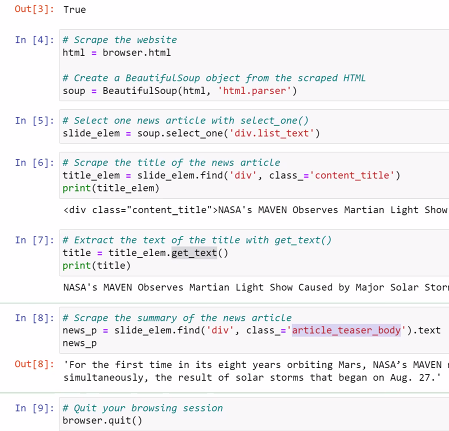
Automated Browsing

Goals:

* Use Splinter to perform automated browser actions.
* Automate the web scraping process by using Splinter and Beautiful Soup.
* Organize scraped information into a Python data structure.

Leftover from prior class session:

**Mars News Solution:**

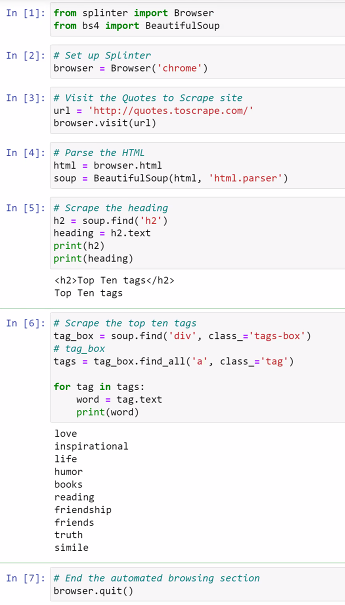


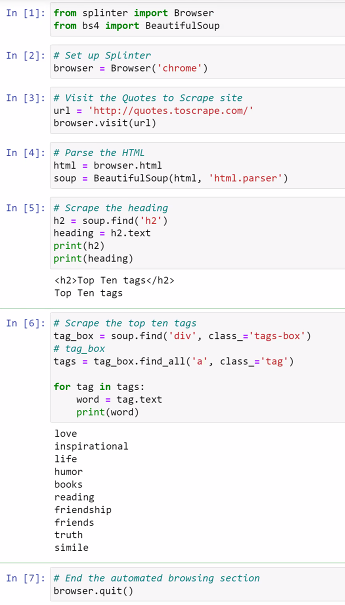
Always read a website’s terms of use or terms of service before scraping the website.

Look for sections that prohibit automated scraping, or directions on using their data.

Research recent legal cases involving web scraping, since legal stances are evolving.

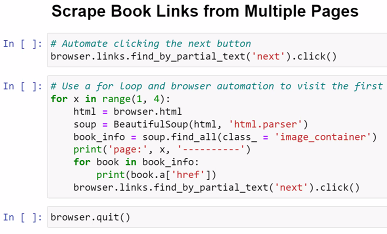
**Automated Web Scrape solution:**





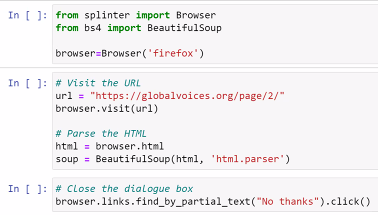
Scraping Books Solution:

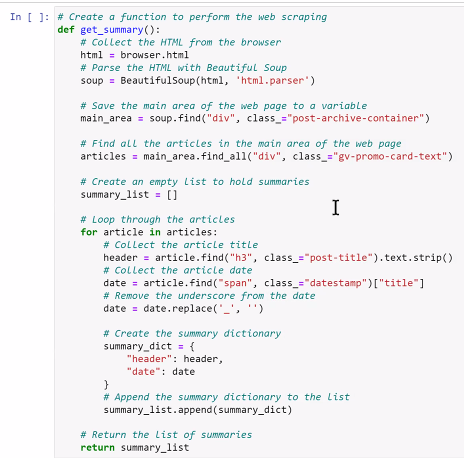


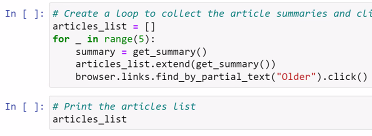


**News Solution: (This one brings it all together in a Pandas DataFrame)**

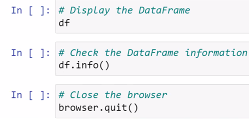
Very useful for Module 11 challenge. (recording- 8:20-8:40pm)



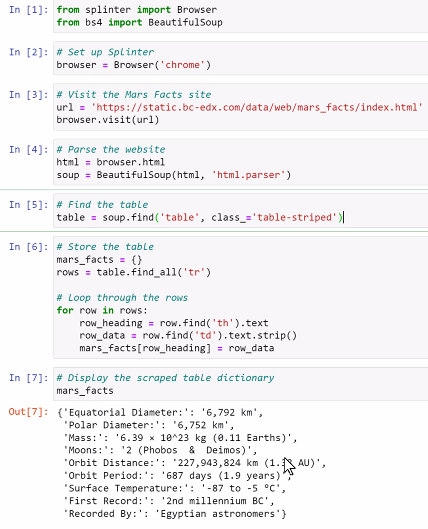


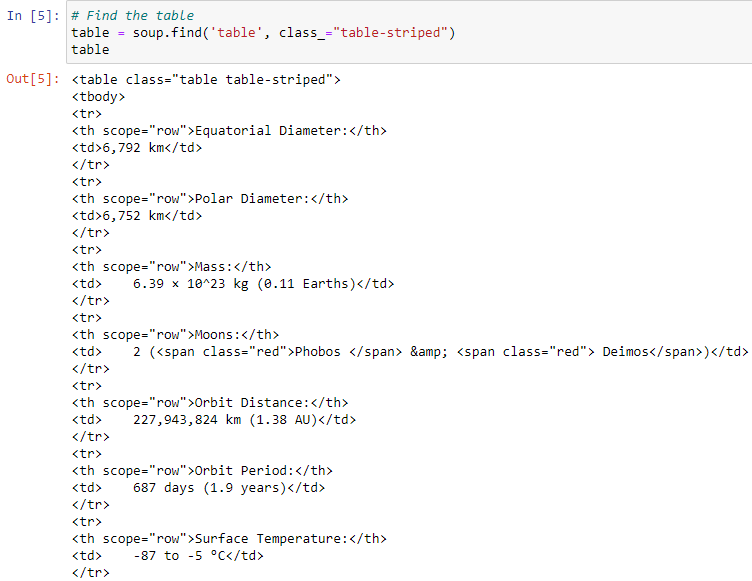






**Mars Fact Scrape Solution:**

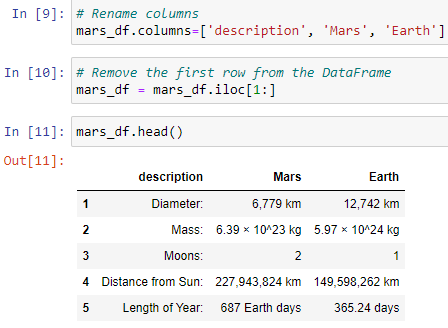




^ In [5] view of the table. TR stands for Table Row, TH for Table Header, TD for Table Data Cell

**Scrape Pandas Table Solution: Review 8:40-8:54 in the recording**





^ This is useful for the homework also.

Review 9:00-9:04pm in recording

Using **NoSQL** is using json format or \_\_\_\_\_\_\_\_

MongoDB stores data as a collection, not a table.

