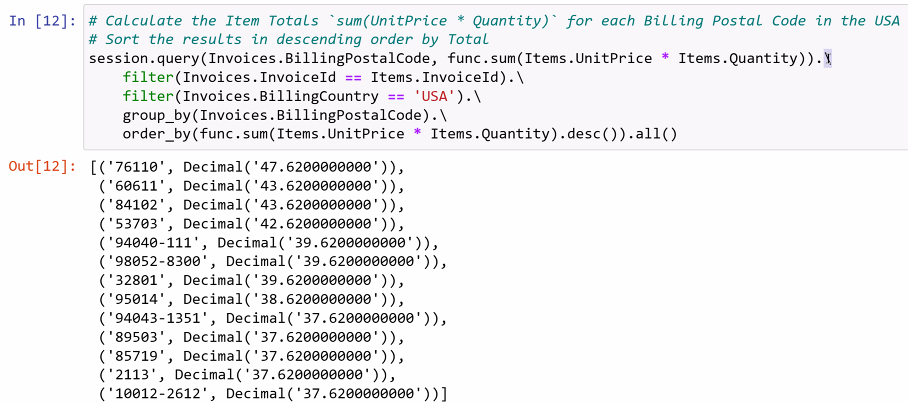
**11-21 notes**

Scraping HTML

Goals:

* Identify HTML components in a website.
* Create a basic HTML document.
* Scrape data from a website by using BeautifulSoup.
* Style HTML elements by using CSS.

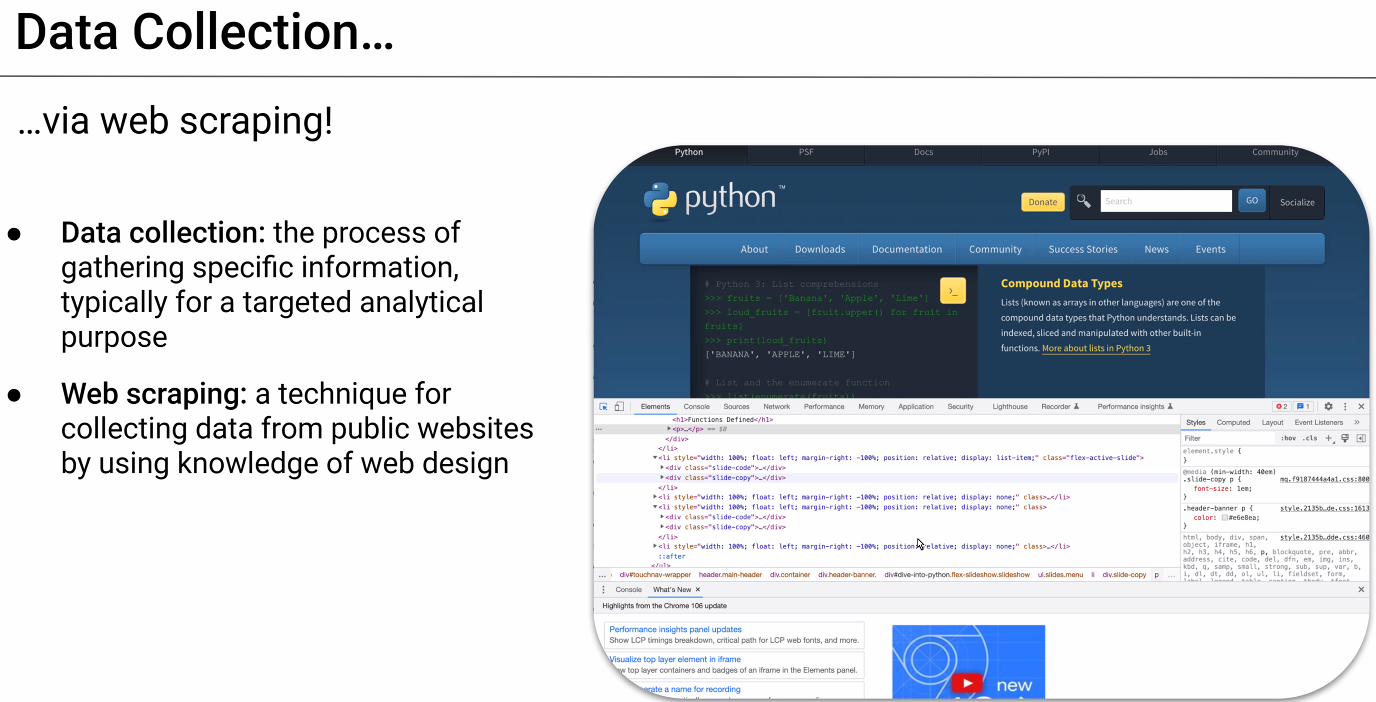
Reviewed module 10, day 3, last activity



^ The / at the end of the line is an “escape character”. It allows you to continue the line on the next line down without having the program read it as a new line (This way you don’t have a line that is going off the page on the right. It’s for readability).

**HTML- Hypertext Markup Language**

You can scrape a website to gather data if no other way is available (a very last resort).

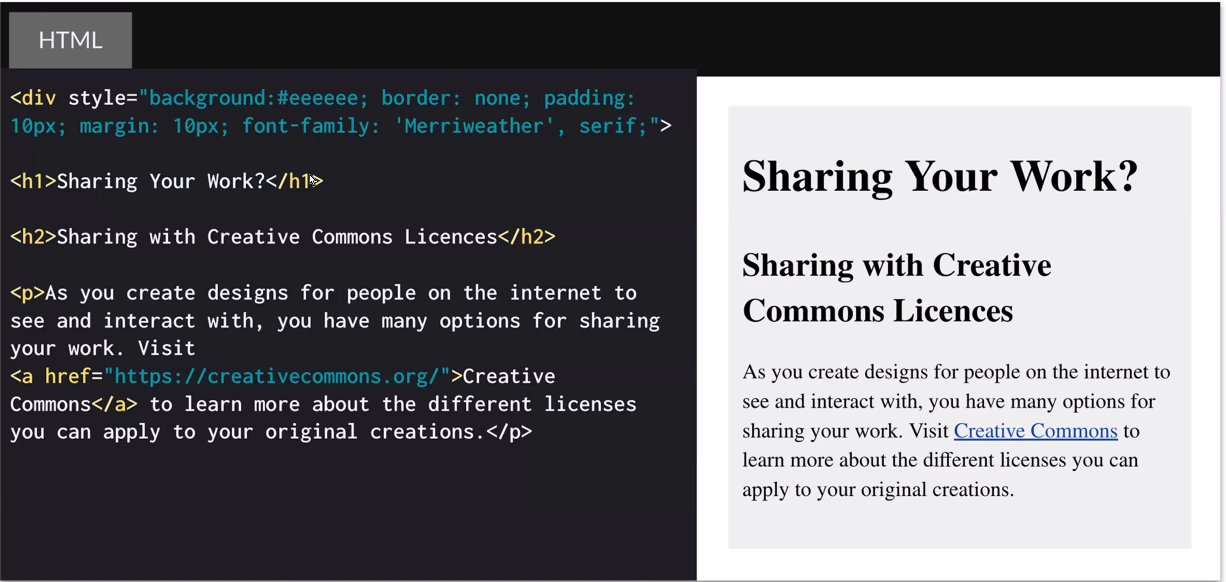




**CSS- Cascading Style Sheet(s)**

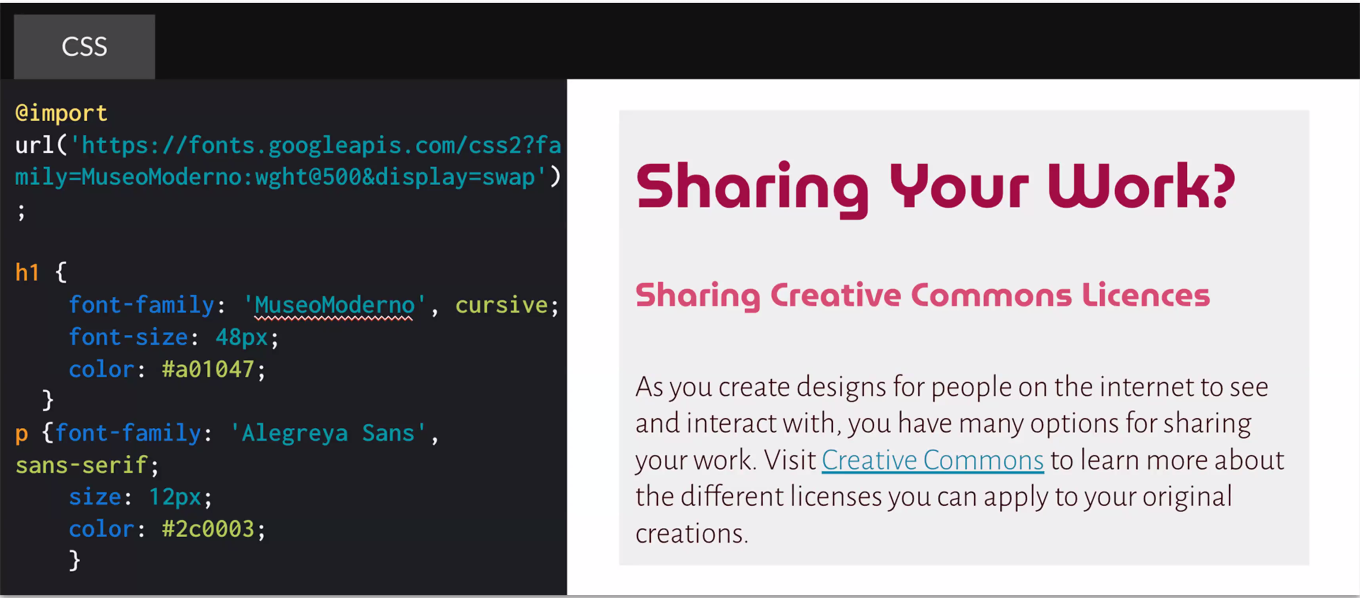
**To do- Review 11/20 recording to point to the chromedriver.exe file.** Right now you’re having to copy/paste it to the same folder as the code.

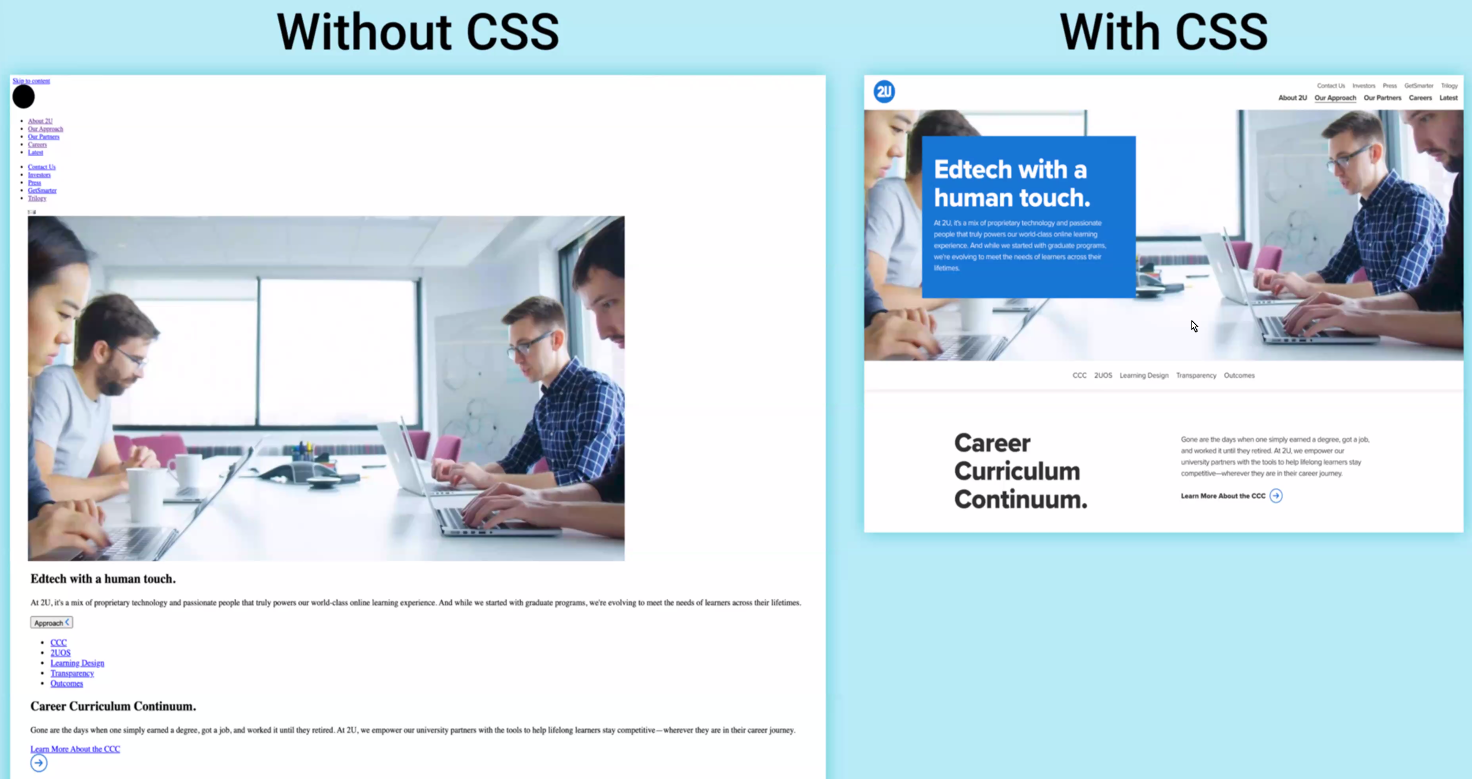
Modern websites are base HTML with CSS overlaid on top.

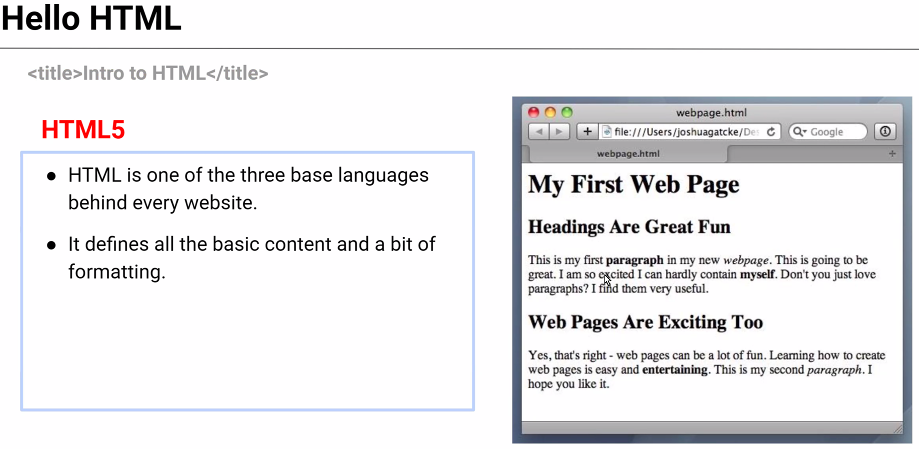


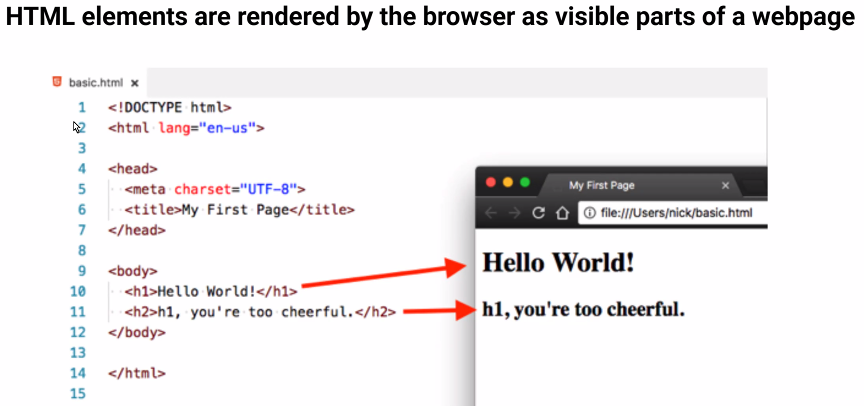
^ Header 1 , Header 2, and a paragraph.

Href means hypertext reference. It tells







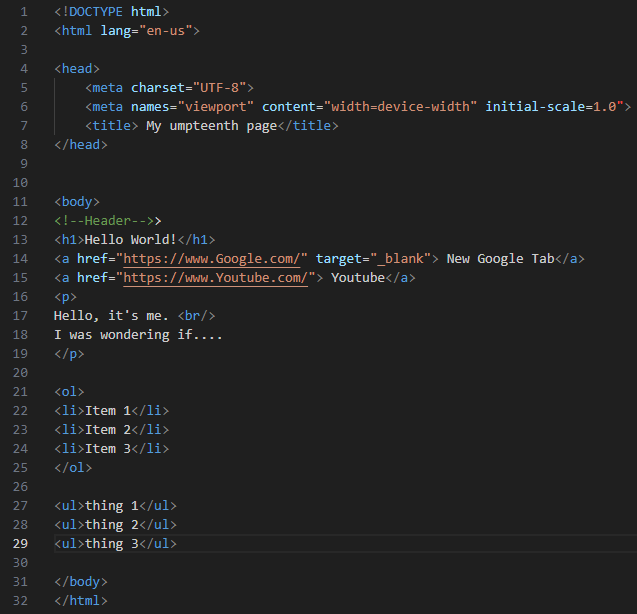


^ Title goes in the tab

Class Demo:



^ Line 11 is a comment. It uses an exclamation point followed by 2 dashes before and after the comment



^ Line 17- <br/> is necessary to tell it to make a new line.

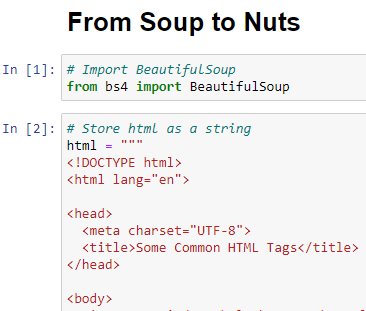
^ Lines 21-25 <ol> is for an ordered list. It’ll put numbers before the text

^ lines 27-29 <ul> is for an unordered list. It’ll make a list of bullet points.

**Beautiful soup:**

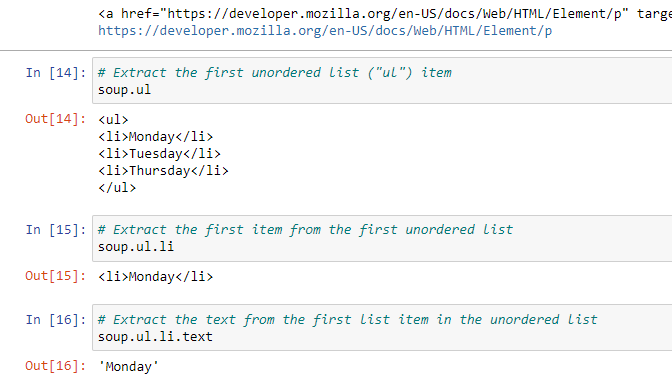
Pip install bs4

You can use it to extract details from the HTML rather than having to scroll through the code manually.



More HTML in IN[2], but skipped over it for space’s sake.







^ IN [18] was manish’s doing from documentation on BeautifulSoup.

You can also download PDFs, convert tables to dataframes, etc….

**Intro to CSS:**



**CSS demo 4 solution:**



^ VS code may not show the colors in the preview. A browser will show the styles applied though.

In example below, using .meat under the styles means it’s a class, and can be applied to an individual item in a list instead of to the whole list as though it was an id

Class uses a period when defining a style.

ID uses a # when defining a style.

They can both be on the list level or on an individual item of a list.

