

Practice Assignment 02

Create a GitHub repository called “st2195_assignment_2” that includes:

1. a README.md file with a short markdown description of this assignment [1 point]
2. a folder called “r_csv” with a R code for scraping the CSV example on cars (we want the table) in the Wikipedia page https://en.wikipedia.org/wiki/Comma-separated_values and saving the resulting output in the local folder (in CSV) [4.5 points]
3. a folder called “python_csv” with a Python version of the code in point 2 [4.5 points]

Note that it is advised to use the packages rvest (R) and BeautifulSoup (Python) for scraping operations. RSelenium (R) and Selenium (Python) can also be used, but they are generally more complicated to setup.

Hint: You may want to have a look at <https://rvest.tidyverse.org/articles/harvesting-the-web.html>

Additional Notes:

- Task clarifications
 - Look for the example cars table from the Wikipedia URL (see below)

Example [\[edit \]](#)

Year	Make	Model	Description	Price
1997	Ford	E350	ac, abs, moon	3000.00
1999	Chevy	Venture "Extended Edition"		4900.00
1999	Chevy	Venture "Extended Edition, Very Large"		5000.00
1996	Jeep	Grand Cherokee	MUST SELL! air, moon roof, loaded	4799.00

- Scrape the table and write to a file in CSV format
- Read the CSV file to a data frame to verify that it was correctly saved

Hints:

- Programming in R
 - While around the table area, right-click and select “Inspect”. Next, hover on the “table class” element and right-click to “Copy” then select “Copy XPath”
 - Paste the **<copied_xpath>** into the R statement `“html_nodes(xpath=“<copied_xpath>”)`
 - You may also refer to this video link on YouTube: [How to find the XPath to scrape tables in rvest - YouTube](#)
- Programming in Python
 - You may use BeautifulSoup
 - Alternatively, using Pandas could also work here