Practice #5: Data collection (2/2): Web scrapping

Example:

The goal is to predict the stock price using past stock prices and other information available online.

In this practice, you will retrieve stock market prices from 3 companies and make them usable by your python script.

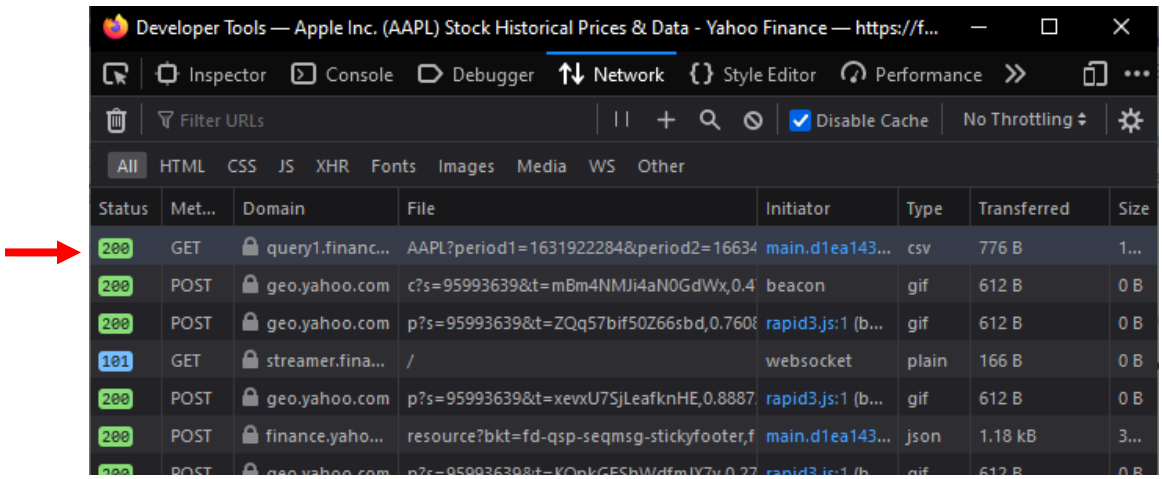
You will first find which link is used when you download your .csv on yahoo finance and use it to automatize .csv files downloading.

The practice can be done with any interpreter (VSCode, Jupyter, Spyder, Pycharm, …).

Steps:

1. **Analyse yahoo finance website to find the request**
   - Open the website https://finance.yahoo.com/
   - Search for your company (or any company)
   - Go to the historical tab
   - Open the inspect tool of your web browser (shortcut: "Q" or right-click => inspect)
   - Go to the network tab of your inspection tool (called "developer tool" in Firefox)
   - Click on the link "download" on the page of your company on yahoo finance
   - Find the GET method which is calling for the .csv file

Example:



2. **Analyse the method to get the website**
   - Click on the previously founded method
   - In the detail of the method, identify:
     - The called URL (it looks like this: https://query1.finance.yahoo.com/v7/finance/download/...)

    - The user agent information (it looks like this: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:104.0) Gecko/20100101 Firefox/104.0)
- Not the URL and the user agent information for later use

3. Analyse the URL to find data

- Use the requests library with the URL and the user agent information to download the .csv file with a python code (see lecture):
    - Import the requests library
    - Create a variable with the previous URL and user-agent information
    - Make a request with the previous variable
    - Store the csv in a dataframe
        - Hint : (where “resp” is the result of your request):
            - Import pandas as pd
            - Import io
            - df = pd.read_csv(io.StringIO(resp.content.decode('utf-8')))
- Display the dataframe

4. Analyse the URL to find data and download another company

- In the previously founded URL, find which part contain:
    - The company’s name
    - The beginning date of the .csv
    - The ending date of the .csv
- Choose two others company and display their data (on the same time period) by modifying the URL