Lab-13

your name here

4/28/23

Preface

The goal of this assignment is to help you gain familiarity with web scraping. As always, please come to office hours and reach out to your teaching staff if you have any questions.

Data

We will get data for this assignment from the web.

1. Let's warm up by building on our work to scrape stock prices from this week's example. Write code that scrapes Goldman Sach's historical daily share price from https://finance.yahoo.com/quote/GS/history. Clean the data, converting Date to a date and all other variables to doubles. Print the first 10 rows of the resulting data frame. What is the earliest date for which prices are available?

Note: If you don't know where to start, review example-13-solutions on Posit Cloud!

The earliest date for which prices are available is...

2. Now generalize your code from question 1 so that it returns historical daily prices for any ticker assigned to the name `symbol`. First assign a new ticker – "JPM" – to `symbol`. As before, clean the data and convert Date to a date and all other variables to doubles. Print the first 10 rows of the data frame.

Note: example-13-solutions could be helpful here, too!

3. Adapt your code from question 2 to create a variable `ticker` **with the contents of the variable** `symbol`**. Print the first 10 rows of the resulting data frame.**

4. Tickers can be pretty hard to decipher. Modify your code to scrape company names from yahoo finance, and store it in a variable `company` in your data frame. Try to remove the trailing ticker in parentheses and any whitespace to get just the company's name. Assign the resulting dataa frame to the name `symbol_prices`, and print the first 2 rows of the data frame.

5. Use the variable `Open` in the data frame from question 4 to compute the stock's cumulative returns over the period of data you scraped.

Note: Make sure you get the dates in the right order when computing returns!

6. Now let's try to scrape data over a custom date range. Go back to the historical data on yahoo finance and use the header to customize the time period to five years ("5Y") and the frequency to Monthly, then click Apply. Verify that the table updated. Now copy and paste the updated url into the code chunk below. Modify your code from question 4 to create a data frame of these monthly prices. Plot monthly `Open` prices over the period of data you scraped.

7. What if we want to focus on a different time period? In words, explain how you would modify your code from question 6 to depend on not just the variable `symbol` **but also start and end dates that plot the share prices for** `symbol` **during the time you have been at Cornell.**

…

8. OPTIONAL: Try to implement your proposed approach from 7!