Task 1 :

A generic Scraper in python to Collect the below listed data points from this website :
https://www.nyse.com
a. Actual name in listing Inside “Quote” section.
b. Last Trade Time.
Examples

Example Entity Name : Wells Fargo & Company

In the search result page, exact name match after the ticker

Example in this case :

“WFC: Wells Fargo & Company” --> ‘Wells Fargo & Company’ should be matched and clicked to extract the above points.

Example Entity Name : Citigroup INC

In the search result page, the exact name match after the ticker.

Example in this case :

“C: Citigroup INC” --> ‘Citigroup INC’ should be matched and clicked to extract the above points.

Approach :

First and foremost thing, inspect the website thoroughly.
Inspect all the elements which will come into play here in this task we need to interact with the search bar and send some value , click on the nearest match and from quote section extract the actual company name and last trade time.

This process of obtaining information from website is known as web scrapping or web crawling.

As I am web scraping in python, there are few framework which we could use.

1. Beautiful Soup
2. Selenium
3. Scrapy

Here i have used Selenium to get the results. The reason behind choosing this framework is its versatile, easy to use and it works well with javascript.

I have written the code in jupyter notebook and it works perfectly.

- Approach is very simple, first we have to take company name from user and save it to a variable.
- Then find the element (here in this case search bar) by css selector and dump our value .
- The script must perform the click operation for that i have used *click()* function. This will give us the dropdown list . Here we have to wait for few seconds as it takes some time to get the data in form of lists.
- Once again we have to select the list item which is matching to our user input value.
- This again can be done by the same way we have found the search element.
- The very next step is to extract the name and last trade time.
- The extraction of name and time is very simple. Go to the area where they are shown and right click to inspect the element. Here i got the division and class name they are in.
- Once i got them it was matter of time when i used the *find_element_by_css_selector()* function to get the elements and used .text method to get the value in string format.
- The reason why i used the *find_element_by_css_selector because in many cases there was no div id and hence forth i went for the function.*

  Finally save the extracted data to a .txt file in append mode . Data saved is in CSV format.