NYSE Scraper

17th September 2020

OVERVIEW

This is basically an NYSE Scraper. It takes an Entity Name and returns the actual name in the Quotes section and the last_traded_time as the result.

Approach

1. **Selenium is used to scrape the Actual name and the last traded time from the NYSE Website.**

2. **The process or the approach followed is as below**
    a. The first URL (https://www.nyse.com/listings_directory/stock) is visited and using **selenium the text is entered in the filter input box**.
    b. Now, the challenge is that it takes some time to load the results, **hence had to introduce appropriate waiting time, so that the results load**.
    c. Once, the results load an**d the URL to the stock Quote can be scraped**, As it is a **dynamically rendered Website**, it is needed to wait for some time for the entire page to load.
    d. So, I dynamically kept on checking using a while loop, whether the page has loaded, and as soon as the page loads, the three important data points(Actual Name, Last Traded Time, and the Price **are extracted**).
    e. These are returned as a dictionary from the function.
    f. Another edge case handled is when the search stock does not exist, then no URL cannot be extracted and appropriate response with the error message is returned.

3. **Alternate Approaches**:
    a. Alternate approaches to web scraping in Python include beautiful soup, beautiful soup is not a very good approach in this case, as it works well on static pages.
    b. One good approach that could work in this case is using the scrapy package from python.