NEXTERA ENERGY
Stock Information Scraping

Web Scraping and Social Media Scraping

Master in Data Science and Business Analytics - University of Warsaw

Alberto Delgado López
Serena Bozheku
Nahid Gulaliyev

Introduction

The finance sector uses a lot of datas. Financial data is useful in many ways as it helps investors analyze a company's performance and reliability. Similarly, it helps a company in analyzing its position and where it stands in terms of finances.

In this project a scrap of the stock information from NEXTERA ENERGY webpage will be done with the purpose of giving a review to an hypothetical investor. NEXTERA ENERGY company is a leading clean energy company which generates clean, emissions-free electricity from seven commercial nuclear power units in Florida, New Hampshire and Wisconsin.

The focus will be on the Investor Relations portal, created by Scott Jehl, Paul Irish, Nicholas Zakas. The data scrapped will contain general information about the company and will give an overview of the company's stock price and its changes during the periods that will help them to analyze their profit expectations.

Scraper mechanics

The Investor Relations portal is a static website, so the three scraper methods known as Beautiful Soup, Scrapy and Selenium will be used to extract the same data for company review, stock table and stock information.

- Beautiful Soup:
  Beautiful Soup is a python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup). It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.

  Firstly, I scraped company overview text from investor relations portal using Initial url. Extracting that information wasn't difficult because I only had to define class for it. Then creating dataframe and save it as a csv file. The second part of coding was kind of difficult because I had to define all tags related to the table and the only way was using span id. Span ids were unique compared to classes.

I created empty list for storing quotes then with append method I stored information there while used find method with specific tags for getting exact data from html elements. Then changed it to a dataframe and saved it as a csv file. The execution time for soup is 0 second.

- Scrapy:

  Scrapy method works using spiders, which are classes that define how the website will be scraped. Since we are scraping data from two different webpages of the same website, we're using two spiders, crawler and stock, to each assigned one of the links of the webpages. After telling each spider which url to enter, it's necessary to use xpaths that will direct the spiders to the needed elements of the page. The data scraped from each spider will be saved in a csv file.

  The webpage of the stock information uses iframes, so the second spider, stock, will be assigned to the url of the iframe. Then, using two xpaths for each column of the table containing the information needed we're able to create a table in the output file. However, in order to get a readable table, it's necessary to remove first all the empty spaces that were being returned and second to append together the text that was being separated such as *ISIN / Symbol* and the dollar signs before the values.

  This method is quite fast as it takes around 2 seconds to run and return the output file.

- Selenium:

  Selenium method opens the browser and scraps the information from an specified XPath. For the first scraped data it was quite simple since the data was in the main page from the initial url path. But for the second scraped data, it requires to automatically change the page to the 'Stock Information' section. It was done introducing the XPath of the section button and clicking it.

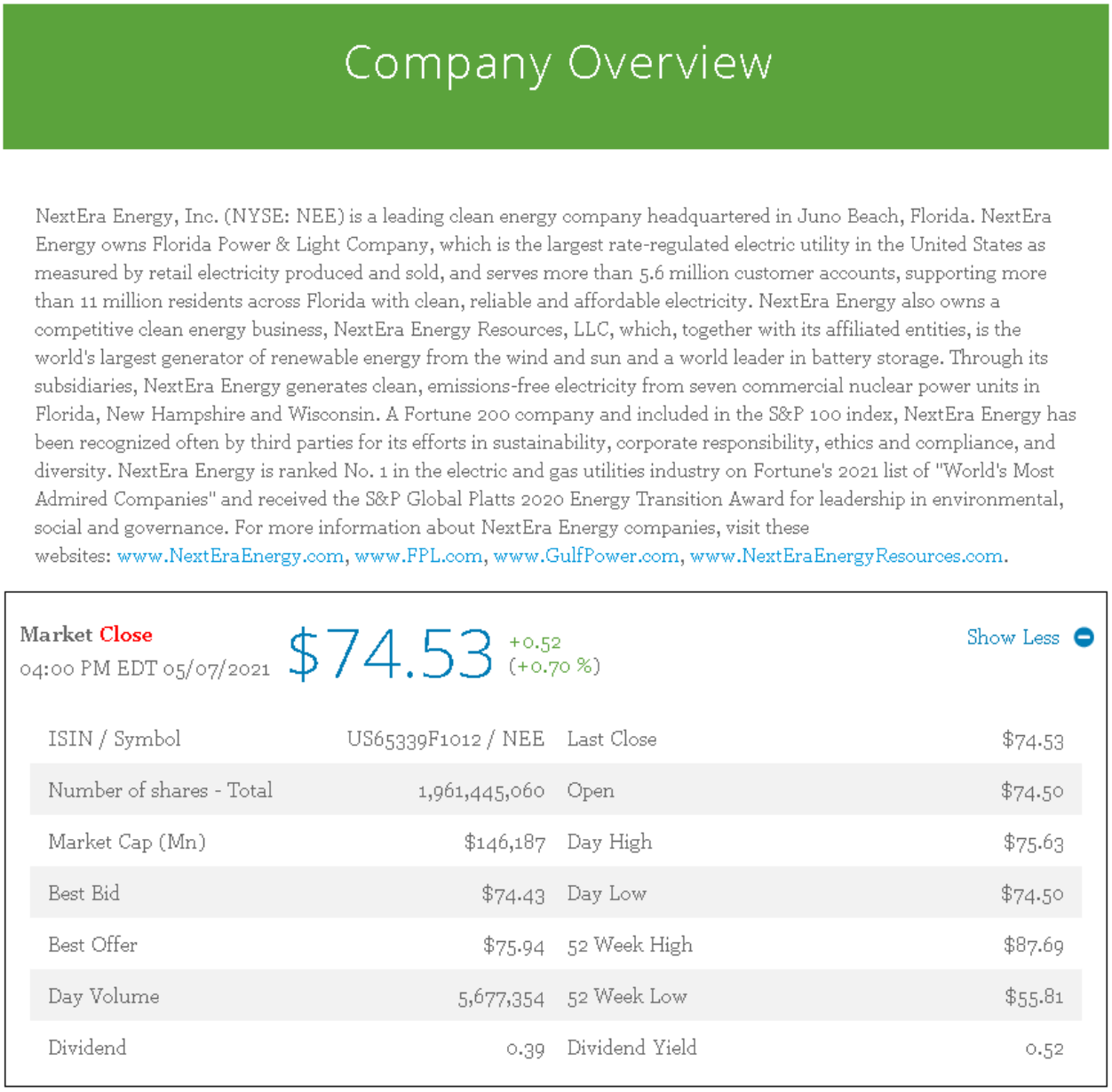
  What was more complicated when scraping this second data was the fact that the table where the data is stored is inside an iframe. It means that the information of the table is stored in another url, and not in the main one, that is also why in the html of the 'Stock Information' site there is no information about this data. So, in order to access the data that we want to scrap, it was necessary to switch the driver to the iframe.

  Then, since the detailed data of the table was hidden, it was also necessary to introduce the XPath of the 'Show more' button and click it. Finally the data from the table became fully accessible, so, using the proper XPaths, the data from the 2 columns was saved in two different vectors, stored in a panda dataframe and printed in a .csv file (very similar way used for the first scraped data).

  It was also very important to add some time.sleep() functions, in order to give time for loading the pages and table to the browser. Ignoring these forced pauses, the execution time for the program is around 7,5s.

Output

These are the data's that we were going to extract.



Company Overview

NextEra Energy, Inc. (NYSE: NEE) is a leading clean energy company headquartered in Juno Beach, Florida. NextEra Energy owns Florida Power & Light Company, which is the largest rate-regulated electric utility in the United States as measured by retail electricity produced and sold, and serves more than 5.6 million customer accounts, supporting more than 11 million residents across Florida with clean, reliable and affordable electricity. NextEra Energy also owns a competitive clean energy business, NextEra Energy Resources, LLC, which, together with its affiliated entities, is the world's largest generator of renewable energy from the wind and sun and a world leader in battery storage. Through its subsidiaries, NextEra Energy generates clean, emissions-free electricity from seven commercial nuclear power units in Florida, New Hampshire and Wisconsin. A Fortune 200 company and included in the S&P 100 index, NextEra Energy has been recognized often by third parties for its efforts in sustainability, corporate responsibility, ethics and compliance, and diversity. NextEra Energy is ranked No. 1 in the electric and gas utilities industry on Fortune's 2021 list of "World's Most Admired Companies" and received the S&P Global Platts 2020 Energy Transition Award for leadership in environmental, social and governance. For more information about NextEra Energy companies, visit these websites: www.NextEraEnergy.com, www.FPL.com, www.GulfPower.com, www.NextEraEnergyResources.com.

Market Close
04:00 PM EDT 05/07/2021

$74.53 +0.52 (+0.70 %)

Show Less

ISIN / Symbol	US65339F1012 / NEE	Last Close	$74.53
Number of shares - Total	1,961,445,060	Open	$74.50
Market Cap (Mn)	$146,187	Day High	$75.63
Best Bid	$74.43	Day Low	$74.50
Best Offer	$75.94	52 Week High	$87.69
Day Volume	5,677,354	52 Week Low	$55.81
Dividend	0.39	Dividend Yield	0.52

Title	Text
Company	NextEra Energy, Inc. (NYSE: NEE) is a leading clean energy company headquartered in Juno Beach, Florida. NextEra Energy owns Florida Power & Light Company, which is the largest rate-regulated electric utility in the United States as measured by retail electricity produced and sold, and serves more than 5.6 million customer accounts, supporting more than 11 million residents across Florida with clean, reliable and affordable electricity. NextEra Energy also owns a competitive clean energy business, NextEra Energy Resources, LLC, which, together with its affiliated entities, is the world's largest generator of renewable energy from the wind and sun and a world leader in battery storage. Through its subsidiaries, NextEra Energy generates clean, emissions-free electricity from seven commercial nuclear power units in Florida, New Hampshire and Wisconsin. A Fortune 200 company and included in the S&P 100 index, NextEra Energy has been recognized often by third parties for its efforts in sustainability, corporate responsibility, ethics and compliance, and diversity. NextEra Energy is ranked No. 1 in the electric and gas utilities industry on Fortune's 2021 list of "World's Most Admired Companies" and received the S&P Global Platts 2020 Energy Transition Award for leadership in environmental, social and governance.

As a result we got exactly the same output as we wanted.

Stock output

	Stock parameter	Value
0	ISIN / Symbol	US65339F1012 / NEE
1	Last Close	$74.53
2	Number of shares - Total	1,961,445,060
3	Open	$74.50
4	Market Cap (Mn)	$146,187
5	Day High	$75.63
6	Best Bid	$74.43
7	Day Low	$74.50
8	Best Offer	$75.94
9	52 Week High	$87.69
10	Day Volume	5,677,354
11	52 Week Low	$55.81
12	Dividend	0.39
13	Dividend Yield	0.52

And this is our stock table information that we wanted to introduce.

Time execution was different for all the methods we used. For BS it was 0 seconds, for Scrapy 2 seconds while for Selenium around 7 seconds. While BS and Scrapy are considered fast methods Selenium is usually slower than them.

Data Analysis

NextEra Energy Stock Analysis

Stock output

	Stock parameter	Value
0	ISIN / Symbol	US65339F1012 / NEE
1	Last Close	$74.53
2	Number of shares - Total	1,961,445,060
3	Open	$74.50
4	Market Cap (Mn)	$146,187
5	Day High	$75.63
6	Best Bid	$74.43
7	Day Low	$74.50
8	Best Offer	$75.94
9	52 Week High	$87.69
10	Day Volume	5,677,354
11	52 Week Low	$55.81
12	Dividend	0.39
13	Dividend Yield	0.52

As it seems from the extracted table, we can write down number of shares in total as 1,961,445,060, market capacity as 146,187 $ million, dividend 0.39 $ annually and dividend yield 0.52 % and daily stock price changes like opening, high, low and also 52 weeks(yearly) high or low prices of stock per share.

Company pays 0.39 $ dividend annually for per share which means its dividend yield is as described 0.52 % (0.39/74.53*100). Dividend yield indicates how much you are earning for each dollar you invested in the company. In this case, we can say 0.52 % earning per dollar is just 5 cents. And for today's stock price as 74.53 $ you earn 3.9 $ per share. Good dividend yield is considered between 4-6 % which means generally investors buy the shares with this range of dividend yield. In addition, from opening, day high and day low you can see that the stock price is stable, it is because NextEra is energy company and its stock price changes differently compared to other kinds of companies such as oil companies. Generally, risk averse investors buy such kind of shares because there is low risk of losing money and also majority of risk averse investors buy huge amounts of shares like in hundred thousands or even in

millions which means their earning seems high as well. This is why, companies like NextEra split their shares more rather than other companies. As mentioned above, it has 1,961,445,060 shares and let's say one of the investors have 1 mln shares which means he or she is going to earn 3,875,560 $ (1000000*74.53*0.52%) yearly. Finally, from the Day Volume we can see that 7 May 5,677,354 shares traded and it shows that company is healthy and strong in terms of its economic impact.

Most common words

Using a function, we can check which are the most common words from the overview output.

```
and 12
in 8
Energy 7
the 7
NextEra 6
is 4
a 3
energy 3
its 3
clean 2
company 2
owns 2
Florida 2
largest 2
electric 2
by 2
electricity 2
more 2
than 2
million 2
```

The most common words are the conjunction 'and' and preposition 'to', and this makes sense because they're used often in sentences.
Next we see that NextEra and Energy are also very often used and the reason why is very intuitive. However the expectation was for the number of recurrences of these words to be higher.

Distribution of the work

Nahid Gulaliyev - BeautifulSoup , NextEra Energy Stock Analysis and project description
Serena Bozheku - Scrapy and project description
Alberto Delgado Lopez - Selenium and project description