Esteban Lopez

Dr. Glen Mules

INFM 203

April 17, 2022

Week 4 Report

This week, besides the lectures and readings, I started playing around with Jupyter notebooks because I saw some tutorials for how to get information from Yahoo Finance. This is for my mini project. It is easy enough to get stock and cryptocurrency information. The difficult part for me so far is I just want a list of all the different market indices. There are tutorials for how to retrieve the companies that make up the S&P 500, but I did find a tutorial to get a list of indices from Yahoo Finance. It did not work for me. The primary API library I have been playing around with is yfinance, but I found another library called yahoo_fin that has some functions I can use for at least retrieving market index components.

2 great sources for me to find tutorials are Medium.com (really their subpages called Towards Data Science and Analytics Vidhya) and since I am trying to retrieve stock information algotrading101.com has been useful. Of course, now that I am playing with Jupyter notebooks, Kaggle and GitHub are even more valuable because they have examples and projects. Another option I am seeing is to use Google Sheets to get information from Google Finance. I haven't tried that yet. I might in order to save information in a CSV file that I can then access with Python. I did follow one tutorial to scrape S&P 500 information from Wikipedia.

In selecting my primary data set, it will be downloading information mainly from Yahoo Finance. I want to see how index funds and ETFs compare to the market indices they claim to follow. Maybe even compare index funds and ETF holdings that follow the same market index, just to see how different they might be.

I did play with the labs this week; it was straightforward playing with some commands. Instead, I will show some screen shots from the Jupyter notebook work I have been playing with. I guess I am guilty of what Godsey calls data snooping (2017, p.98).

In the screen shot Figure 1, I was able to plot a graph very easily. There are many tutorials on how to customize the graph, but this was just out of the box using the information grabbed from Yahoo Finance using pandas, I had not discovered the yfinance API at this point. It was not all roses, but it was all fun and games. Figure 2 shows me following a tutorial exactly and the process is a failure (except I learned and that is a win). Most sources say that the Yahoo Finance API is not official and can fail at any time; but it is free so that is a risk I am willing to take. Figure 3 shows a nice scraping output of S&P 500 information from Wikipedia. It is nice and informative when compared to the Figure 4 which is just a list retrieved using the yahoo_fin API. I feel like the yahoo_fin method is immediately usable to retrieve the stock information

using the tickers. Though I was able to create a CSV of just the tickers using scraping of Wikipedia method. So there are many options, since I know Yahoo Finance isn't guaranteed to work I should probably make CSVs as soon as I retrieve information.

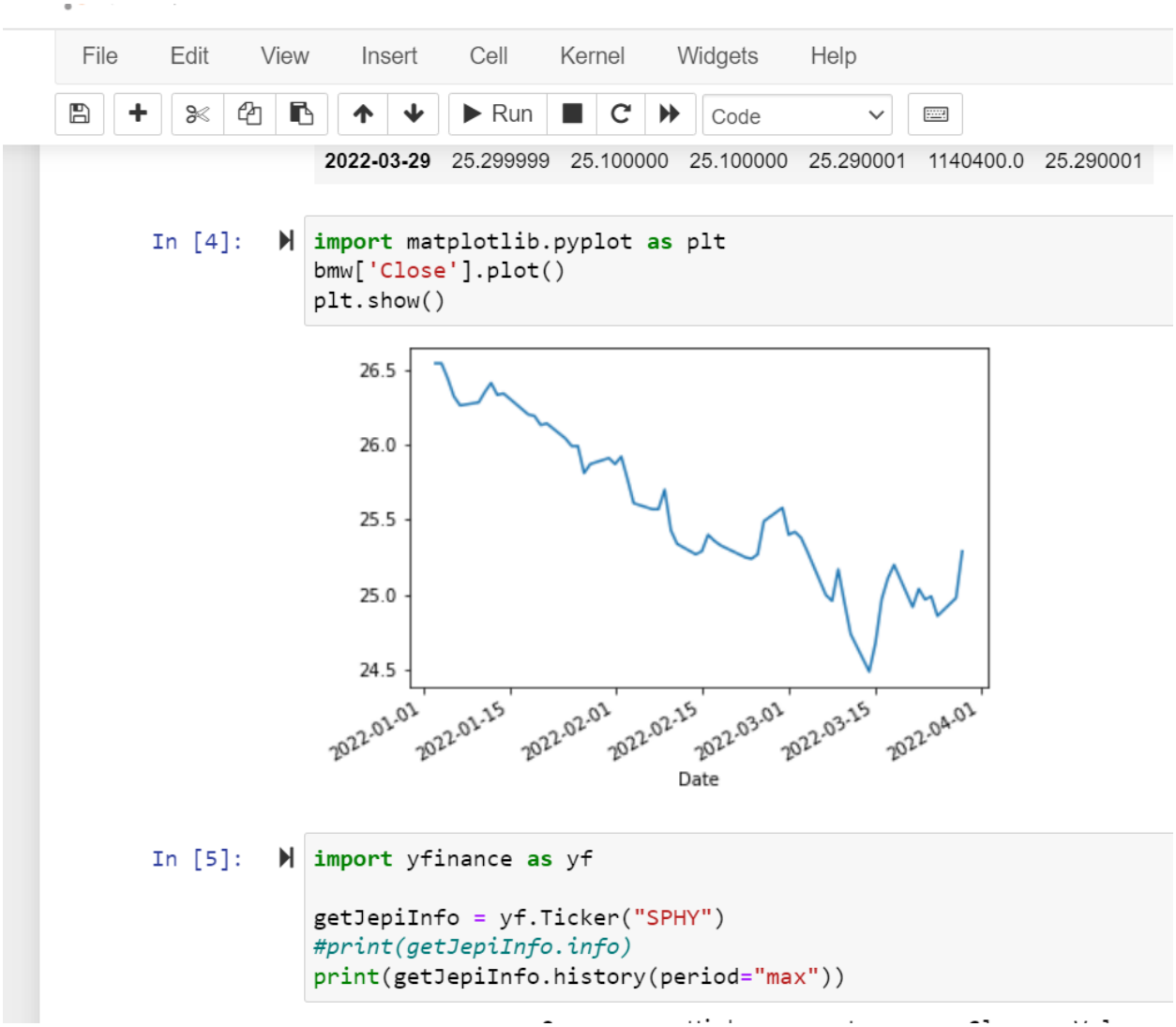


Figure 1: Basic Graphing

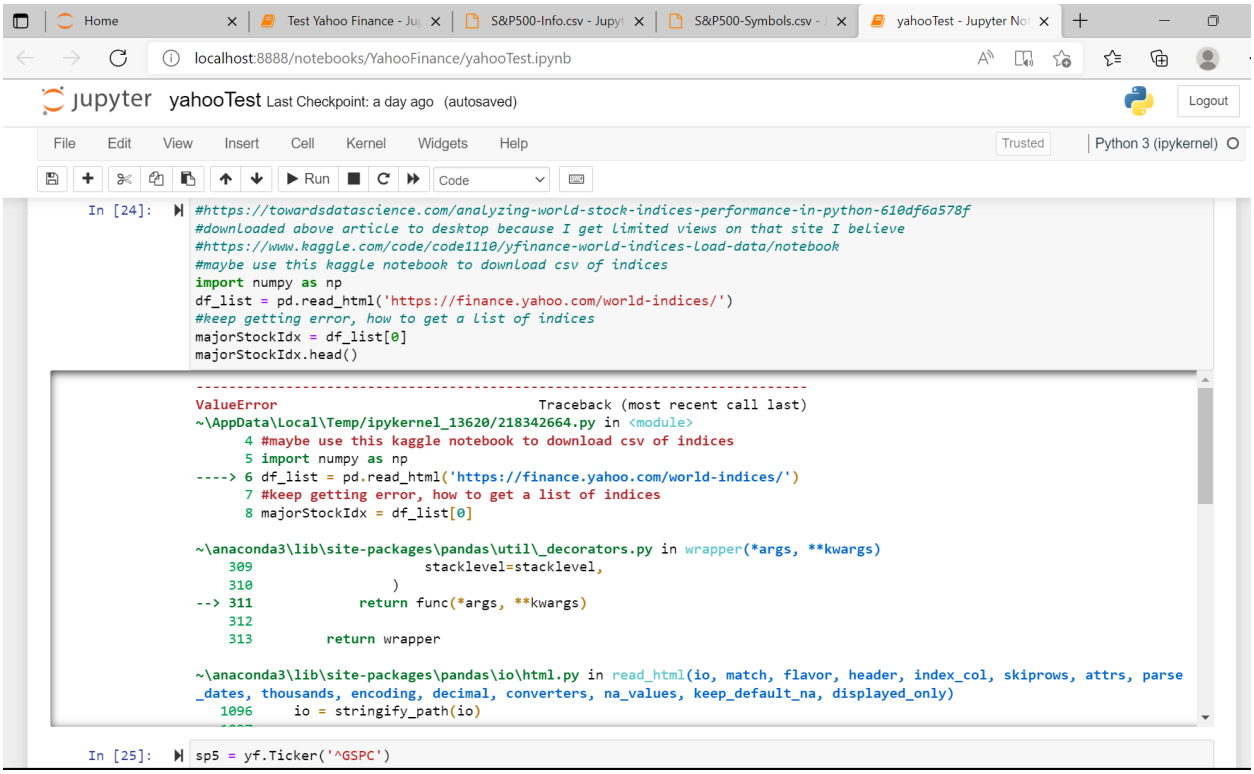
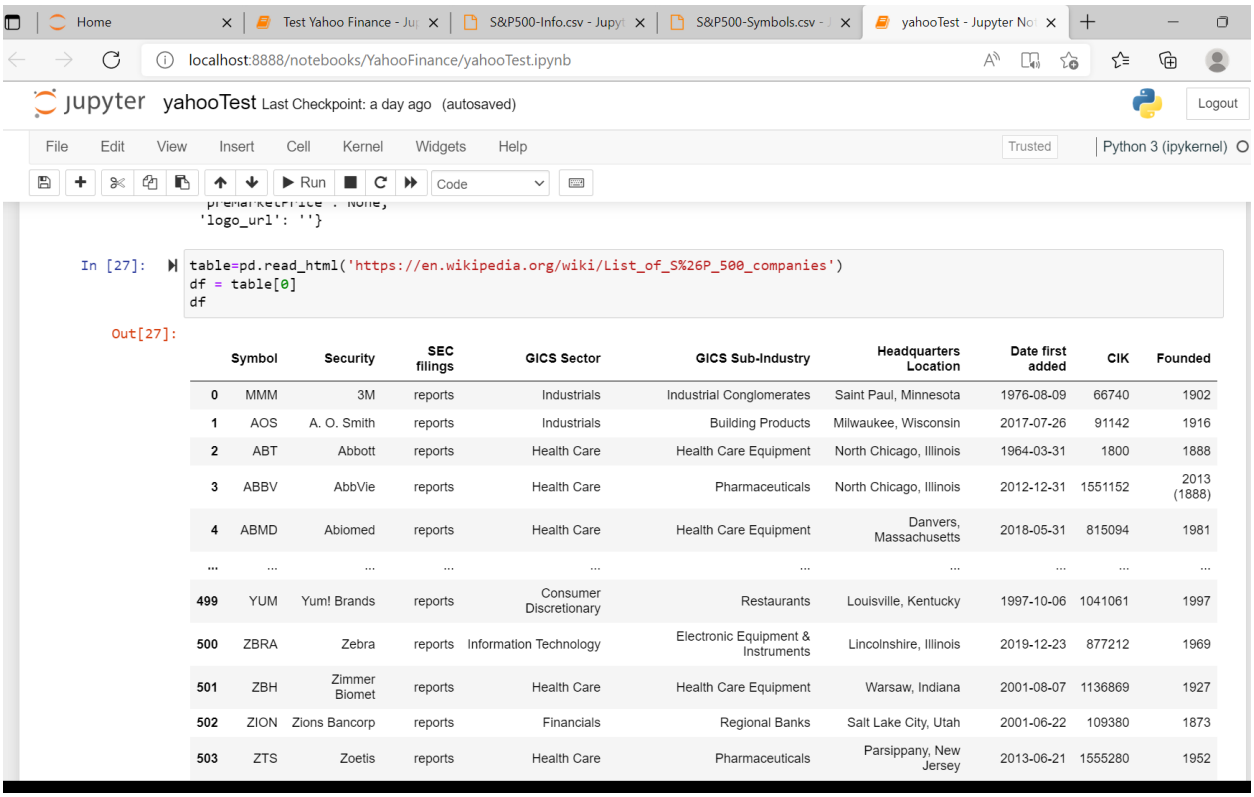


Figure 2: Failure



	Symbol	Security	SEC filings	GICS Sector	GICS Sub-Industry	Headquarters Location	Date first added	CIK	Founded
0	MMM	3M	reports	Industrials	Industrial Conglomerates	Saint Paul, Minnesota	1976-08-09	66740	1902
1	AOS	A. O. Smith	reports	Industrials	Building Products	Milwaukee, Wisconsin	2017-07-26	91142	1916
2	ABT	Abbott	reports	Health Care	Health Care Equipment	North Chicago, Illinois	1964-03-31	1800	1888
3	ABBV	AbbVie	reports	Health Care	Pharmaceuticals	North Chicago, Illinois	2012-12-31	1551152	2013 (1888)
4	ABMD	Abiomed	reports	Health Care	Health Care Equipment	Danvers, Massachusetts	2018-05-31	815094	1981
...	...	...	...	...	...	...	...	...	...
499	YUM	Yum! Brands	reports	Consumer Discretionary	Restaurants	Louisville, Kentucky	1997-10-06	1041061	1997
500	ZBRA	Zebra	reports	Information Technology	Electronic Equipment & Instruments	Lincolnshire, Illinois	2019-12-23	877212	1969
501	ZBH	Zimmer Biomet	reports	Health Care	Health Care Equipment	Warsaw, Indiana	2001-08-07	1136869	1927
502	ZION	Zions Bancorp	reports	Financials	Regional Banks	Salt Lake City, Utah	2001-06-22	109380	1873
503	ZTS	Zoetis	reports	Health Care	Pharmaceuticals	Parsippany, New Jersey	2013-06-21	1555280	1952

Figure 3: Retrieving S&P 500 from Wikipedia

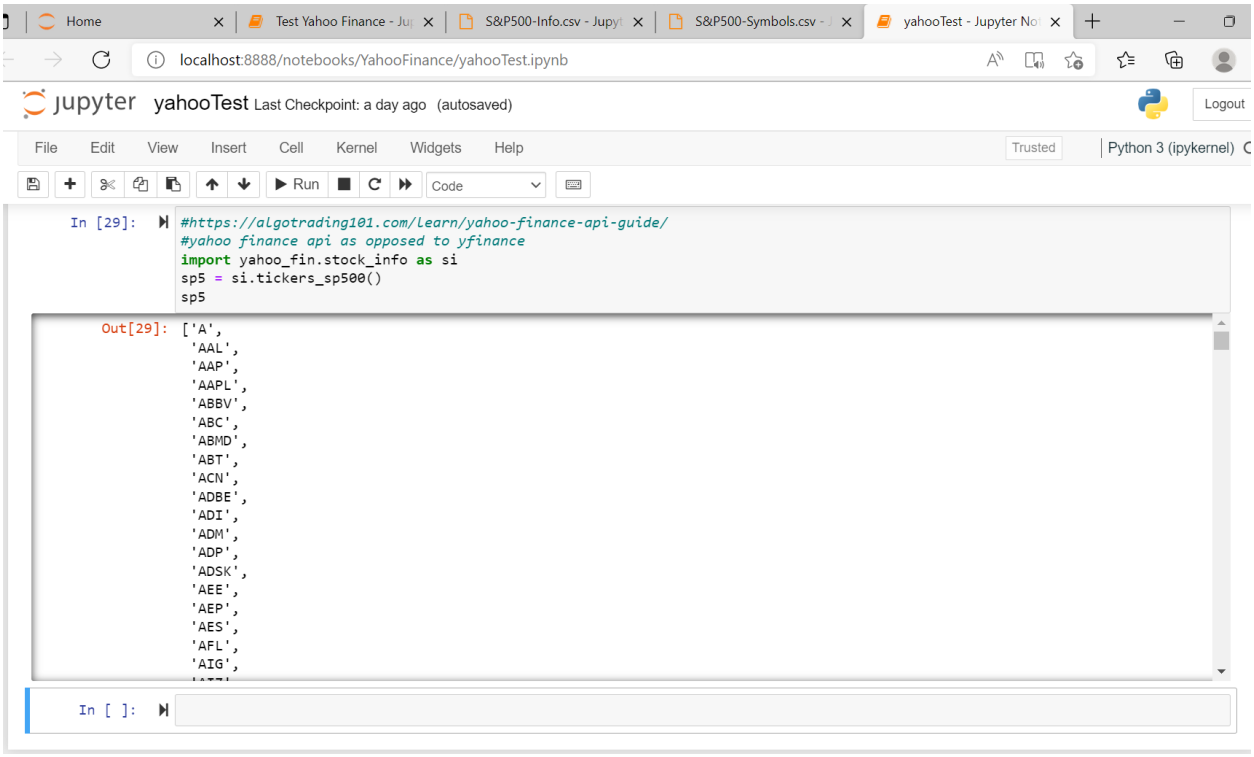


Figure 4: S&P 500 List Using yahoo_fin API

Reference

Godsey, B. (2017). *Think like a data scientist: Tackle the data science process step-by-step*. Manning Publications.