

1. I did not encounter any difficulties finding the objects I needed to replicate your work. Scraping the data directly from the Internet is a fantastic way to streamline all of your work and make everything easily reproducible—there's no fighting with CSV files. However, what happens if these websites change? It might be a good idea to also save off local versions of the data.
2. Your IPython Notebook is generally easy to follow—I found the small headings in Markdown very helpful for giving me an idea of the general trajectory of your project. I would suggest that you include additional comments into your code cells; if someone needed to edit your code, then the lacking of more detailed commentary explaining certain steps could pose a problem.

Generally I like your tests, but why did you choose to include your tests in your main IPython Notebook and not to use Nose? Having everything jumbled together makes it hard to differentiate test code from main code, and then I had to manually run and analyze each test one at a time. All of your tests passed without error, which is certainly good, and I thought the test to randomly pull a stock and make sure that it matched the dataframe was particularly useful to check the “black box” process of scraping Internet data.

3. I created all of your figures, though without your midterm report, it's hard to say if they've been accurately reproduced. Some plots are missing labels for one or both axes, and most plots do not have a main title. Adding those annotations will help the figures to be more readable. I would also number the figures so that they can be easily referenced within the text of your paper.

Now for some specific comments (the number is the cell number). 11: The company labels are hard to see. 14: It's hard to see overall patterns in the data with so many points. 19: Are there colored lines? What do they mean? 20: Plots, legends, and labels overlap; it's hard to see what's happening. 21: Many plots look similar, especially when I can't juxtapose plots to compare. 26&40: Y-axis limits seem too large. 58: There are no labels; what are you plotting?

4. I think one of the greatest strengths here is simply the breadth of your project. You do a lot, both in your application of a very wide variety of statistical techniques and in an extremely large quantity of analyses, and it's quite impressive to see what you've done. Your code is well organized and the headings help to annotate the outline of your project, which makes reading it quite easy. Finally, I was impressed by your process of directly scraping the data from the Internet—while it could pose issues if the websites were to change, reproducing your work was extremely streamlined as I didn't have to worry about any datasets at all.
5. First, I would clean up your figures a bit by adding labels and captions; also, perhaps you don't need all of the plots because many of them look very similar (like in cell 21, see #3 above). Second, make sure that the text that you add does a good job to explain your process—quite honestly, it's a little hard to decipher your research questions, goals, and project impact from your notebook in its current form (which is to be expected without text). Make sure to justify your methods. Why did you choose to use the techniques that you used, and how do they complement each other to get to your ultimate goal? Lastly, make sure to define all of the company abbreviations that you use throughout your analysis. I would just put a table linking abbreviation to company name right after your introductory paragraphs.

Overall, you have done a large amount of work here, and you should be very proud of it. Best of luck as you move forwards in this project.