1. I did not encounter any difficulties finding all 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 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.

I think your test suite is less effective. 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 from main code, and then I had to run and analyze all of the tests manually. 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. One other tip is perhaps you can write/adjust some tests that take inputs of dummy data, such as a matrix of all zeroes, and test that some code (perhaps your code to calculate daily returns) is accurate (with all zeroes, I would gander that the daily return would also be zero, right? That should make testing easy).

1. The only issue I had in creating your figures was in cell 20; the plot (attached) has overlapping and unreadable labels and legends. I’m not sure if this is intentional or a reproducibility issue. Otherwise, the other plots look decent, though without your midterm report, it’s hard to say if they’ve been reproduced. Some plots are missing labels for one or both axes, and most plots do not have a main title. Adding those annotations in addition to captions will help the figures to be more readable. I would also number the figures so that they can be easily referenced.

Now some very specific comments (the number is the cell number). 11: company labels are hard to see. 14: with all the points, it’s hard to see the overall patterns in the data. 19: are there multiple colored lines? It looks so—what do they mean? 20: see attached. 21: many of the plots look very similar, especially when you have to scroll and can’t juxtapose any next to each other. 26&40: y axis limits seem too large. 58: there are no labels here, what are you plotting?

1. Which parts did you find especially good and/or useful?

You displayed a wide-variety of statistical techniques, which is both very exciting to see and could require some finesse with explanation (see 5). Your code is very well organized, which makes reading it quite easy, and as I mentioned earlier, the headings that you do have there help to annotate the outline of your project.

1. I would say focus your remaining work on two areas. First, clean up your figures a bit by adding labels and captions; I would also venture to say that perhaps you don’t need all of those plots as many of them look very similar (particularly in the lasso densities). Second, make sure that your 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 any text). But make sure to justify all of your methods. Why did you choose to use the variety of 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 so everyone knows. I would put a table right after your introductory paragraphs that links abbreviation to company name. Overall, you have done a TON of work here, and you should be very proud of it. Clean things up and work on packaging and explanation to tie a bow around your work, and you will have something quite impressive.

Figure Created by Cell 20

