Feedback for Project Number 12

Destinee Garcia

Cyber Breaches and the S&P 500

***What is their topic on?***

* *Is the title consistent with the topic?*
* *In other words does the title make sense for the project?*

Destinee’s project proposal is:

* “Is there a relationship between a business’ industry/size/market value and recent reports of cyber security breaches?”
* “Are larger companies a bigger target for cybersecurity hacks?”
* “What are the most common methods that hackers have been using?”

The title makes sense in relation to these goals. The second bullet point seems to say the same thing as the first. The third is an interesting topic but perhaps doesn’t relate directly to the main research question.

***Are the objectives of the project clearly identifiable?***

* *What are they wanting to study?*
* *What is the motivation of this project?*
* *What does the presenter hope to accomplish with this project?*

Data breaches can cause great reputational harm and loss of trust in a company, which can affect individual companies’ market values, and in aggregate could play havoc with the stock market. Destinee states the objectives clearly. And whether the bigger the company, the more likely that it will be subject to a breach, is true is an interesting question.

***What data are used?***

* *What website(s) are scraped?*
* *Are other data brought in from outside sources?*
* *If so, what are they and how do they add to the project?*

Destinee first looked at using the [Fortune 500 web page](https://fortune.com/fortune500/2020/search) as her primary source material. She found that it would be difficult to use the methods Dr. Martinez taught to us in class to get past the first page of ten results.[[1]](#footnote-1) Destinee was able to find an alternative data source that in the end was more useful to the research question, as well as being a more straghtforward web scraping project. She used the [Value Today](https://www.value.today/stocks/sp-500-index) website to scrape the data for the S&P 500 companies (which really is better for this question, as some Fortune 500 companies are privately held), and Wikipedia for the data breach information.

***What is your overall impression of the project?***

I thought this was a very creative topic, and I can tell that Destinee put a lot of thought and work into it. I think I would have come to a different conclusion. The chart with the counts per company of the major breaches showed that most of the major corporations had just one breach over a four-year period. This seems like insufficient data on which to base a conclusion. I also do not see how it answers the question of whether larger companies are more subject to breaches. A comparison of smaller to larger companies would have provided this answer, although as noted above there is such a small number of breaches that I don’t think this type of comparison is possible.

Notes on the visualizations:

Slide 1: This was interesting general information. I couldn’t see where it added information relevant to the research question, and I would have left it out.

Slide 2: I liked this Tableau slide. The 1400% increase in cyber attacks in 16 years was impactful. (And shame on those large companies that were hacked because they had poor security. They should have known better!)

Slide 3: The time period wasn’t labeled and in the narration it was just “the last four years”. This slide would have benefited from noting the time frame: “2016-2020” or whatever the dates were.

Slide 4: The pie chart and the bar graph were difficult to compare because the bar graph had more categories. I realize this is probably because they come from two different data sources. This would have been a useful note to the viewer, since the two charts can’t be directly compared for every category. Still, I would have made this two pie charts; even with the categories not aligning perfectly, it would still have shown that tech firms comprise 13% of the S&P 500 and 24% of cyber threats. You could make a case for including social media with the tech companies, which would be 28%. Also, this is a little bit picky but the pie chart is pretty small compared to the bar chart; visually I would have tried to make them about the same size.

**Other encouragements/critiques you would like to provide to your colleague in order for them to have a more refined project.**

I noticed a few typos. One I think isn’t a big deal; I think I saw three.

The code seemed very “Pythonic”! I was unable to access the github but I was able to enlarge it in the slide deck.

Overall, I thought Destinee were an effective presenter and her slides were easy to follow. I enjoyed it!

1. I was curious to see if there was anything online where someone had been able to scrape the Fortune 500 page, and I found this article, [“Visualizing Fortune 500 Cos.”](https://nycdatascience.com/blog/student-works/visualizing-fortune-500-cos/) There is a [github](https://github.com/jdsipala/seleniumProj) page with the [code](https://github.com/jdsipala/seleniumProj/blob/master/seleniumProjectNotebook2.ipynb). This is not a critique of the project at all, because I could see it would have been pretty difficult for a first project. I just thought Destinee would be curious to take a look at the code and see how it can be done. There is code that will “click to the next page”. [↑](#footnote-ref-1)