The dataset that I chose is pretty straight forward. There have been no surprises. I wanted to do a sentiment analysis for this project as it had been a few classes since doing one. The dataset contains information such as the title of the review, the review itself, the age of the reviewee, the type of product being reviewed, and the department the clothing being reviewed is in. Sentiment analysis can be very helpful for businesses. Reviews are important as potential customers can rely on current customers. Word of mouth and recommendations are crucial to businesses as they can keep them afloat.

The dataset was what I thought it was and I think it has more information than what I really need to do a sentiment analysis. I am going to try to perform some other analysis on the data. Since the department of where the product is listed is available, I could do an analysis on what departments are proving to be stronger in sales than others. There is a lot that can be done to this project and I believe that the sentiment analysis is just the start of what can be done. Overall, this dataset has a lot of information that can be worked to do more analysis about the e-commerce company.

I have not had the need to adjust anything on my approach for the sentiment analysis on the reviews. On the dataset overall, I think I can adjust by adding some more research questions and give more of a detailed report on other factors within the e-commerce company. As mentioned earlier, I could see what departments have better reviews than others. I could possibly even narrow down which items have the most positive reviews and which items have the most negative reviews. This could then tell what items a success with the company and which items may need to be reevaluated.

At this point, my method is working. I have been able to remove stop words and punctuation from the reviews. I have also done some data cleaning as I found that there were empty values in the title of the review and the review itself. I want these columns to have some data in it as they are crucial for the sentiment analysis and I would not want the empty values to hurt the prediction score. Sometimes people leave a score but no comment, which does not help and really provide much value when doing a sentiment analysis.

As far as challenges, I would say the most would be time. I have had a project roll out this week at work. The first half of the week was to make sure that everything that needed to be done was done, while the second half was bug fixes and helping the users with the new system. Sometimes I can find some downtime in work to work on homework, but this week that was very thin. I am hoping by after next week when we can fix more issues and less bugs will be found, that work will slow down some and I can get more focus time during the week for school work.