1) Which tasks have been completed?
- Data Processing: We built a web scraper with requests_html package to do opinion mining - particularly extract reviews from 300 different Amazon products. We also included 5k review comments from an already-curated dataset in Kaggle to add more data points.
- Data Mining: We obtained an overall distribution of how the data are categorized based on our custom thresholds. We also did a bit of clean-up and formatting as well as splitting them into sets suitable for the next tasks.
- Model research: We did research on models suitable for opinion analysis and made a decision on what kind of tokenization we would go with.
- Data Analysis: Sentiment Analysis was run with the selected models and for each we ran the confusion matrix as well as F1 score.
- Analysis Discussion: Together as a group we discussed these results and planned a few tuning steps to see if there are any more improvements we can make.

2) Which tasks are pending?
- UI Creation: Creating a UI that allows for the model to be accessed in an easier way
    - We are considering using either a Python library or React based on the speed of the model
- Creating API endpoints that allow for the model to be called upon button clicks and display results
    - We will be using Flask to create the endpoints

3) Are you facing any challenges?
- We faced an initial challenge of scraping Amazon reviews as there are many protections around web scraping when trying to scrape more than one page in a product review. To work around these permissions, we scraped only the first page/landing page of several hundred products (breadth rather than depth).
- Another challenge we faced was a lack of data points for the sentiment model. We initially scraped 3k reviews and weren't confident about our model results. We included 5k more reviews to reduce our uncertainties.
- The selection of models also posed a bit of challenge as some are a bit hard to understand at first so it was more time-consuming than expected. After doing a few test runs, the models' results were quite close to each other so it is questionable which one is the best.
- A challenge we faced was the ability to run the web scraping program on all our machines. Therefore, we decided the UI will display a text box that allows uses to submit reviews and output the sentiment prediction rather than scraping more review data from an Amazon URL