1. What's the name of your final project? Please describe it as a research question and provide a short description.
    - Name: Comparing Amazon Product Ratings against Amazon Stock Price
    - Description: In this project, I aim to use the Amazon Review Data k-cores data to perform sentiment analysis on various product segments sold on Amazon. Based on the date of the review, I will then perform linear regression to assess whether there is correlation between the measured sentiment and Amazon's historical stock price (found using the yfinance API) during that time period.
2. What data sources are available? Could you find multiple data sources? How are you going to collect them? How many data samples are you going to collect?
    - The data sources I intend to use for this project are the Amazon Review Data k-cores data (https://nijianmo.github.io/amazon/index.html) and Amazon's historical stock data from yfinance (https://pypi.org/project/yfinance/).
    - The Amazon Review Data contains minimized datasets of complete reviews for sparse sets of items in various categories. This reduces the number of data points being measured.
    - Amazon's historical stock data from finance is a variable dataset including information on opening, closing, and adjusted closing prices (among other things). I can access various pieces of data as needed for this project.
    - Other available data sources include the complete Amazon Review Data. However, for the scope of this project I do not need millions of data points. Thousands will suffice.
    - The data is structured and stored as a JSON (Amazon Review Data) or is accessible as a dataframe (Amazon historical stock data). I intend to perform sentiment analysis on the Amazon Review Data and use linear regression to find correlation between the sentiment and the historical stock price.
        i. The bulk of the Amazon Review Data is textual, but much of it can be discarded for the purposes of this project. I will have to clean this dataset heavily.
        ii. The historical stock data is numerical. I will filter this dataset to find data in the appropriate timeframes.
3. What kind of analyses or visualizations do you want to do?
    - First, I would like to perform sentiment analysis on the reviews. Specifically, I'd like to see whether there are words that occur more frequently in reviews based on the number of stars.
        i. I would also like to make a pie chart denoting the most frequently used words across each rating.
    - For each product category, I would like to plot the average review rating against Amazon's stock price in the same time period.
        i. For example, if I have a month's worth of reviews that are generally positive for books, I would plot that against Amazon stock price in the same month to be able to visually see whether there is a strong correlation.
    - In addition to the aforementioned plots, I would also like to measure the statistical significance of the correlation between a category's ratings and the stock price in the same time period. I will use linear regression to do this.