Mini Project Plan

Research Question: To what extent did the rating sentiments of stock analysts affect the 5 day returns on the closing price of select NASDAQ-100 stocks due to the emergence of OpenAi on November 30th, 2022.

Steps:
1. We first queried the dataset from Google Big Query, limiting our results to select companies. We chose to analyze:
   `'NVDA','META','TSLA','GOOGL','MSFT','AAPL','AMZN','INTC'`
   Our date range was from November 30, 2019 to present, with the last data entry on September 15, 2025. We chose these dates because they were three years before and after November 30, 2022, the day OpenAi launched ChatGPT.
2. Next we did cleaning on the queried dataset. First we removed any neutral sentiments, with keywords like "hold", "neutral", "equal weight", "market perform" and such.
3. After that, we used the no_holds dataframe to clean and organize the data some more. We cleaned so that there was only one instance of a certain stock per day by taking the mean of the sentiments of repeated stocks if they were on the same day. For example, on a certain day, there can be a max of eight stock companies, each with one average sentiment: buy(1), hold(0), and sell(-1). Again, we get rid of all rows that contain 'hold'.
4. Next we used a dataset from Yahoo finance focusing on its closing price. We limited the dataset to only stocks from above as well as only information from September 30th, 2019. We used a 5 day interval to create another column that shows the closing price change 5 days later. Then, we created another column that showed the 5 day return rate by doing (5_day_closing - closing)/closing. If the 5 day return rate is positive, that means the closing price of a stock increased and vice versa.
5. We merge the two datasets together so that we can see analyst sentiments of a stock in the form of 1 and -1 as well as the 5 day return rate representing how the price changed.
6. After that, we created another column to differentiate whether the analysis was made before or after the emergence of OpenAi. Rows before November 30, 2022 were marked with a 0, and rows after November 30, 2022 were marked with a 1. Then, we split the dataset up into two, one with only data before November 30, 2022, and one with only data afterwards.
7. Using these two dataframes, we conducted two separate Welch's T-test to quantify the difference between the means of two datasets given that they had unequal variances. We first conducted the T-test on the Pre-ai dataset, splitting the sentiments up to positive and negative. The result was significant, and showed that there was a difference between the means of the positive sentiment group and the negative sentiment group. We can interpret this as the analyst sentiment did have a significant impact on the change of the closing price of a stock.

8. We conducted the same test on the Post-ai dataset, also focusing on negative and positive sentiments. This time, our results were not significant, showing that there is not necessarily a correlation between the analyst predictions and the actual stock price change.

9. Lastly, we performed an Ordinary Least Squares regression (OLS) to see if analyst ratings mattered after the ai boom. We added another column of data to the combined sentiments and 5 day return dataset which multiplied the ai_status column and the rating sentiments column (both being represented by binary numbers). Using this new column as the Y, we performed the OLS regression using sentiment ratings, ai status, and the multiplied version of those two as the X.

10. The result was a p-value of 0.043, which is less than 0.05, which means we reject the null hypothesis. This means that there is a difference in the amount of weight analysts had on the stock market closing prices as a result of their predictions.