Date	Work	Type	Time
3/29/25	Collected the data that was downloaded and understood the file structure for data cleaning process. Because of size constraints, I have only condiered the files from the last four years i.e., 2024,2023,2022,2021.Understood the file structure, YearN --> QTRX(i) --> 10X. Extracted all the 10K files from the 10X files using pyton. For better and easy to use, created a dataframe of all the 10K filings. Saved the dataframe into a .csv file for future uses.	Individual	120
4/2/25	The dataframe contains data of the filings from different companies, did some cleaning, removed HTML tags. Saved this file into .csv format. For the initial trial, I have used the chunking --> sentiment --> aggregation on the cleaned data. To check the performance we have only considered 1000 reports. For the sentiment analysis we have used FinBERT from the huggingface.	Individual	180
4/4/25	As the results from the FinBERT were improper the probability values of certain sentiments were more than 100 which is not an optimum results. So I have pivoted to use the LM lexicon based sentiment analysis. For this approach I have downloaded the LM lexicon dataset from https://sraf.nd.edu/loughranmcdonald-master-dictionary/ . Then applied the text normalization necessary for this method like removing stop words, lowercase and removing numbers. There are 9 different sentiments in the LM dataset but i have considered only 7 as complexity and syllables are not useful in the sentiment analysis as they define how complex the words are and for the syllables are used to check the readability score. The results are better than the FinBERT. Using the results i have applied the dominant score to get the final sentiment of the document	Individual	120
21/4/25	Finalised the end game plan, I have taken the responsibility to run the code with the available filings	Individual	60
24/4/25	Collected the SP500 Data to sort the filings only to those companies based on the CIK. We got the data of 520 companies out of 5202 available over the last 10 years. Applied the section level sentiment analysis using both the methods, i.e., LM and FinBERT. Started working on the paper as an extension personal project for future use.	Individual	120

26/4/25	Ran the final full document level sentiment analysis and saved the data of the results at the respective levels. Used GPUs and batch processing for faster results. Found some errors while running full FinBERT using Transformers on GPUs. Resolved the issue and re applied the sentiment analysis by using a batch size and one process. Could have done multiprocessing but because of time limitation used only single process with a batch size of 32.	Individual	180
27/4/25	Completed the paper with the analysis made on the master sentiment data with all the section level and both models. Used the data visualization from the analysis for the paper and used the references from the papers to make it final draft. Compelted the Github and made a clean file structure and clean code for better readability.	Individual	120