

Date	Work	Type	Time
3/24/24	Selected the data source from the website Software Repository for Accounting and Finance - https://sraf.nd.edu/sec-edgar-data/cleaned-10x-files/ . Will start with year 2024 for the data collection into a csv file then later on we can continue to loop the yearly filings from the inception. Data collection will be carried out by Naresh as he had already worked on the EDGAR files. Furqan will work on converting the file information into a csv format. Sai Harshit will work on creating the loop after the successful collection of the data from one year. We will create a github to work on the project. We have already created a teams group to store and track the progress. We have downloaded the 10X filings of the year 2024, as it has all the 10X filings we have to sort the filings and collect only 10K.	Group	60
3/30/24	After the data collection and extracting the 10K files, we have decided to only move forward with the last four years of data, as the data was too large. For the initial phase we have decided to try with the FinBERT pretrained model.	Group	60
4/3/25	As the results from the FiBERT were not optimal, we had to find another way to find the sentiment for large documents. After doing some research we found lexicon based methods would be a better option and started working on the lexicon based method.	Group	60
21/4/25	Discussed about the progress of where we are and the possibilities of what can be done. As we have automated the download process of the index files along with the 10k files, we just have to make the comparison between the sentiment at different levels using different methods. For now we are planning to use FinBERT and compare it with the LM lexicon method. Then using both models on complete document level, segment level. An additional part could be the summarization and using sentiment models on the summarized text to find the relation in these methods.	Group	60
24/4/25	Checked with the final possible analysis, as the summarization part takes a lot of computation and time we thought of skipping it for now and focusing on two level sentiment analysis by using both models.	Group	60
27/4/25	Discussed final presentation part. Divided the sections based on the work related to individuals. Completed the code cleaning and also making sure code is well labelled and well organized.	Group	60