1.1)

Sec-downloader was used to get the URL for the required files and the content was extracted from the URL using Beautiful soup

1.2)

I wanted to give a summary of the positives and negatives about the company and also give a chart of sentiment scores of the company over the years. The score ranges from -1 to 1.

Since 10-k filings are huge, I feel a summary of positives and negatives about the company derived from 10-k would make investors work easy. We also calculate and plot the sentiment across years, so that investors would get an idea about the historical sentiment of the company.

I use pipelines from transformers to create summariser and analyser. For summariser, facebook/bart-large-cnn model was used and ProsusAI/finbert was used as the sentiment analyser.

The analyser had restrictions input restrictions of only 512 tokens, So each yearly data was split into chunks of the required size and overlapping. This is done using regex, which extracts words from the content, thereby also removing unwanted text. If analysis of the chunk shows that the probability of being neutral is <0.5, then the chunk was considered to be valuable, hence summarised and stored. Such summaries from different years were combined to get an overall summary. To make the sentiment score range between -1 to 1, $\{-1*(prob\_negative)+1(prob\_positive)\}$ was calculated as the sentiment score. Since we deal with large files, most of it would tend to be neutral, So rather than taking an average of sentiment scores across all chunks in a year, the average was only taken across chunks that have prob_neutral<0.5

The combined summary was further analysed for sentiment by splitting into chunks, and based on the sentiment, chunks were separated into positive and negative segments. The chunks were further summarised and combined

2)

I built a dashboard using streamlit library and used npx localtunnel to host it on the local server, the dashboard will have a dropdown to select a ticker and 3 buttons (positives, negatives and sentiment over the years) which upon clicking would open a pop-up window, showing the respective content.