# SEC\_financialReports

1. This is a data science assignment which deals with sentiment analysis of SEC/EDGAR to different companies.
2. There are two python codes available here. One is to scrape the webdata, named sec\_report\_py and other is to perform text analysis, named SEC\_financialReports.ipynb.
3. To run both programs, we need any of the available IDE for web scraping and google collab for Text Analysis.
4. We need all the requirements which are mentioned in the file requirement.txt to run the web scraping code sec\_report.py on any IDE.
5. Here I used selenium to scrape the web page , as it suits best for these kinds of data.
6. Here I used pycharm IDE for web scraping, and uploaded the output excel sheet of web scraping manually to the required file path on google drive.
7. For google collab required project libraries and modules are directly imported through the code, no need to import separately .
8. All the required files , i.e., sec\_with\_reports.xlsx ( obtained from web scraping using selenium), StopWords\_Generic.txt, LoughranMcDonald\_MasterDictionary\_2020 (1).xlsx, uncertainty\_dictionary.xlsx, constraining\_dictionary.xlsx and all the guidelines under one repository, i.e., BlackCoffer-assignment.
9. **Note**: Please check the path where the drive is mounted on collab, and the path needs to be checked and changed according to the relevant path on your drive.
10. Output file is also redirected to the same path, BlackCoffer-assignment , named it as “final\_output.xlsx”

**Note** : 'percentage\_of\_complex\_words' is calculated as per the given formula in the Text Analysis.