**Instructions**

1. Firstly download all three excel files (Input, uncertaining,constraining.) and converted(uncertaining, constraining )to lower case using excel.
2. Then I loaded file named - Input into dataframe appended the link in secfname row.
3. Extract data.
4. Performed Text Preprocessing using nltk including

\*convert to lower text

\*Tokenization(word-tokenize)

\*Stemming

\*lemmitization

\*removing stop word

5) then performed text analysis using vectorization from sklearn library CountVectorizer().

6) using for loop I iterated through each data and calculated

All the variables including

1. All input variables in “cik\_list.xlsx”
2. positive\_score
3. negative\_score
4. polarity\_score
5. average\_sentence\_length
6. percentage\_of\_complex\_words
7. fog\_index
8. complex\_word\_count
9. word\_count
10. uncertainty\_score
11. constraining\_score
12. positive\_word\_proportion
13. negative\_word\_proportion
14. uncertainty\_word\_proportion
15. constraining\_word\_proportion
16. constraining\_words\_whole\_report

as per their formulas given and append them into a original dataframe df and downloaded the file from colab as csv named df.csv(renamed as output).