**Internship\_Blackcoffer\_Assignment.py**

The file contains all the Data Extraction and Data Analysis codes. The objective of this assignment was to extract textual data articles from the given URL and perform text analysis to compute variables. To perform data extraction I used Pandas, Requests and BeautifulSoup(Bs4) library. For each of the articles, given in the input.xlsx file, I extracted the article title and content and then saved the extracted article in a text file with URL\_ID as its file name.

For Data Extraction I used nltk and regular expression (re) libraries. For each of the extracted texts from the article, I performed textual analysis and computed variables given in the output structure excel file. After computing variables I filled out all the columns and save the output in “Output\_Data\_Structure\_File.xlsx” in the exact order as given in the output structure file.

**Output\_Data\_Structure\_File.xlsx**

The output file contains all the computed variables. It includes all input variables in “Input.xlsx” as well as the desired variables; positive score, Negative Score, Polarity Score, Subjectivity Score, Avg. Sentence Length, Percentage Of Complex Words, Fog Index, Avg. Number Of Words Per Sentence, Complex Word Count, Word Count, Syllable Per Word, Personal Pronouns and Avg. Word Length.