**Instructions** :

* Just need all the files given to us during submission to be present as they are .
* Place the python file in the folder .
* Execute it after checking all the paths remain the same (1st point).
* The file will automatically create output.xlsx file which will contain all the data .
* The outputs are printed on the terminal .
* The code was not able to access 5 links in the dataset .
* And the time required for SYLLABLE PER WORD was too long it was more than 15 mins for each link hence I have commented the code and have shown the output as “taking too long”

**Modules required :**

Nltk

BeautifulSoup

syllables

pandas

numpy

Approach of the solution :

I used beautiful soup for scrapping the data .

All the websites had similar structure especially naming convention of all the elements in the html .

I took the div element which had the information on the article for scrapping.

I looped through the urls, scrapping the data and processing it using the common div element in them .