ReadMe

1. All the dependencies and required packages are mentioned in the python file ‘ black.py’ with comments.
2. Run black.py on command line using >>python black.py
3. Read input.xlsx file using openpyxl package to acces url and url\_id from each row
4. Initialize all output data to 0
5. If URL exists
   1. Text file with name url\_id is created and text from the webpage is copied to it using selenium. For this user define function **create\_text\_file(url,output\_text\_file)** is called.
   2. List of stop words is created with by reading keywords from all the files stored in folder stopwords.Function **stop\_words\_list=get\_stopwords(stopwords\_folder\_path)** is called
   3. List of token is created from the text stored in the generated text file. Function **tokens=tokanize\_text(output\_text\_file)** is called
   4. Tokens present in the folder stopwords are removed.
   5. Sentiment analysis is done on filtered tokens. **sentiment\_score=sentiment\_analysis(file\_path\_positive,file\_path\_negative,filtered\_tokens)**
   6. Personal pronuous are computed from the output text file . **personal\_pronoun\_count=personal\_pronouns(output\_text\_file)**
   7. **S**yllable analysis is done **. syllable=sylabble\_analysis(tokens)**
   8. Total cleaned words are calculated from the filtered tokens. **word\_count=count\_words(filtered\_tokens)**
   9. Average word length is computed using characters in token and total number of tokens**. average\_word\_length=calculate\_average\_word\_length(tokens)**
   10. Average sentence length is calculated using total number of words and total number of sentence. **average\_sentence\_length=calculate\_average\_sentence\_length(output\_text\_file)**
   11. An average word per sentence is calculated using total number of words and total number of sentence. **average\_word\_per\_sentence=calculate\_average\_word\_per\_sentence(output\_text\_file)**
6. If Output Data Structure file is not present it will be crated otherwise data structure will **be added in the next row of Output Data Structure file.**