1. **How I approched the solution?**

I divide the solution into two main parts: Data Extraction and Data Analysis.

Part 1: Data Extraction

**Import Required Libraries:**

* We will use pandas to handle the Excel input file.
* For web scraping, we can use BeautifulSoup to parse HTML content and requests to fetch web pages.

**Read Input Data:**

* Use pandas to read the URLs from the 'Input.xlsx' file.

**Web Scraping:**

* Iterate over each URL in the input data.
* Fetch the web page content using requests.
* Parse the HTML content using BeautifulSoup to extract the article title and body.
* Save the extracted text to a text file named with the corresponding URL\_ID.

**Save Extracted Text:**

* Ensure only the article title and body are saved, excluding headers, footers, and other non-article elements.

**Part 2: Data Analysis**

**Load Required Libraries:**

* Use nltk for tokenization, TextBlob for sentiment analysis, and textstat for text statistics.

**Read Extracted Text Files:**

* Read each text file corresponding to the URL\_ID and perform text analysis.

**Text Analysis:**

* Tokenize the text into sentences and words.
* Compute the required variables: word count, average sentence length, average word length, polarity score, subjectivity score, percentage of complex words, fog index, average number of words per sentence, personal pronouns count, complex word count, and syllable per word.

**Save Results:**

* Store the computed variables in a structured format as specified in the 'Output Data Structure.xlsx'.

2)How to run the .py file to generate output

* **Install required libraries using:**
* pip install pandas requests beautifulsoup4 nltk textblob textstat openpyxl
* **Execute the script in the terminal:**
* Python Blackoffer.py

3)Include all dependencies required:

* pip install pandas requests beautifulsoup4 nltk textblob textstat openpyxl