**How I Approached The Problem:**

First comes the **extraction** part,

* I started by first extracting all the articles into text files using Article() class of newspaper module. This would ensure smooth analysis in the code ahead.
* For this purpose, I made a function to extract article title and article content from a particular URL and return a single string file.
* Further, ran all URLs provided in a loop and created a separate text file for each article.

Now, onto the **analysis** part,

* In a separate .py file named “analysis”, imported everything from the “extraction.py” file.
* Started with creating a master dictionary, which would have “positive” and “negative” keys and the positive and negative words as their respective values.
* Proceeded to create a function to count syllables in a word since it would be required later on.
* This will be covered using two functions,
  + **First:** A function which uses nltk’s cmu.dict(), which is an extensive dictionary containing a large number of English words and their respective pronunciations.
  + **Second:** If there comes a word with no respective key in the dictionary, we will calculate the number of syllables it has by using a custom-made function, which loosely generalizes the method of syllable recognition in a word.
* Further, defined a function named “analyse\_text” taking on “filename” as a variable, which would just be the name of the text file we want to analyze. This will return a row containing URL ID, URL, and all the relevant variables with which we can assess and analyze the article. Later, a loop will be created to analyze all files at once by feeding their file names into this function which will be inside the loop.
* This function takes the file name, reads the text inside it and cleans it by removing stop words, then proceeds to create lists of words and sentences. Now, we have everything in place to calculate the required metrics.
* Now, before starting the loop, an empty list is initialized, which we will use to store all the URLs with their respective metrics, in a multi-dimensional array.
* This array is now converted into a Pandas DataFrame and exported as an excel file using Pandas’ “to\_csv()”.

**Dependency Required:**

To run the code, it is important to have the following libraries installed: pandas, newspaper3k, re & nltk. “os” is an inbuilt library in python.

**How To Run The Code:**

Since the code deals with folders and files at times, it is important to load the correct folder into your VS code or any other coding environment. The “Input.xlsx” file should be there in the folder that you have loaded.

Running the “analysis.py” file alone will generate the desired output, and you will find an excel sheet “final\_output.xlsx” in the folder that you have loaded into the environment. This will contain the entire solution.