Objective:

The objective of this assignment is to extract textual data articles from the given URLs and perform text analysis to compute various variables as specified. The results are then saved in an Excel file following the specified output structure.

Requirements:

Python 3.x

Libraries:

* requests
* beautifulsoup4
* pandas
* openpyxl
* nltk
* Textstat

Setup Instructions

**Step 1: Install Required Libraries**

Run the following command to install the necessary Python libraries:

pip install requests beautifulsoup4 pandas openpyxl nltk textstat

**Step 2: Ensure NLTK Data is Downloaded**

Run the following commands to download the required NLTK data:

import nltk

nltk.download('vader\_lexicon')

nltk.download('punkt')

**Step 3: Place Input Files**

Ensure that Input.xlsx is in the same directory as main.py.

**Running the Script**

To run the script and generate the output, use the following command:

Python textAnalysis\_code.py

**Output**

The results of the analysis will be saved in an Excel file named Outpu.xlsx in the same directory.

**Script Details**

The script performs the following steps:

1. **Data Extraction**:
   * Load URLs from Input.xlsx.
   * For each URL, extract the title and article text using BeautifulSoup.
   * Save the extracted text to a file named with the corresponding URL\_ID.
2. **Text Analysis**:
   * Load the extracted text files.
   * Perform text analysis to compute the following variables:
     + Positive Score
     + Negative Score
     + Polarity Score
     + Subjectivity Score
     + Average Sentence Length
     + Percentage of Complex Words
     + FOG Index
     + Average Number of Words per Sentence
     + Complex Word Count
     + Word Count
     + Syllables per Word
     + Personal Pronouns
     + Average Word Length
3. **Save Results**:

* Save the computed variables along with the input data to Output.xlsx.

After Running the Main Code wait for Atlest 7 to 8 min because it takes that much time to generate the results