**Project Overview:**

This project involved extracting textual data from URLs, performing comprehensive text analysis, and presenting the results in a structured format. The assignment required using Python for data extraction and analysis, focusing on a range of textual variables such as sentiment scores, readability indices, and complexity metrics.

**Approach:**

1. **Data Extraction:**
   * **Objective:** Extract article text from given URLs while excluding website headers, footers, and other non-article content.
   * **Tools & Libraries Used:**
     + **Requests:** For fetching the web pages.
     + **BeautifulSoup:** For parsing HTML and extracting relevant content.
   * **Process:**
     + Read URLs from the provided Input.xlsx file.
     + Use the requests library to download web pages.
     + Employ BeautifulSoup to parse HTML and extract article text, ensuring only the article content is captured.
2. **Text Analysis:**
   * **Objective:** Analyze the extracted text to compute various textual metrics.
   * **Tools & Libraries Used:**
     + **TextBlob:** For sentiment analysis to compute polarity and subjectivity scores.
     + **NLTK (Natural Language Toolkit):** For sentence tokenization.
     + **Textstat:** For readability and complexity metrics (e.g., FOG index, percentage of complex words).
   * **Process:**
     + Implement functions to compute:
       - Sentiment Scores: Polarity and subjectivity.
       - Readability Metrics: FOG index, percentage of complex words.
       - Text Statistics: Average sentence length, syllables per word, personal pronouns count, etc.
     + Each function processes the text and returns the required metrics.
3. **Output:**
   * **Objective:** Save the computed metrics in a CSV file following the specified format.
   * **Process:**
     + Collect results for each text file.
     + Save the results in a CSV file named output.csv, including all required variables and metrics as per the provided output structure.

**Dependencies Required:**

* **Python Libraries:**
  + requests
  + beautifulsoup4
  + textblob
  + nltk
  + textstat
  + pandas

**How to Run the Python Script:**

1. Ensure all required Python libraries are installed. You can install them using:

bash

Copy code

pip install requests beautifulsoup4 textblob nltk textstat pandas

1. Place the analyze\_text.py script and the Input.xlsx file in the same directory.
2. Update the directory path in the analyze\_text\_files function to match the location of your text files.
3. Run the script using:

bash

Copy code

python analyze\_text.py

1. The script will generate output.csv in the same directory, containing the results of the text analysis.