**Instructions**

1. My Approach

The script takes a list of URLs, analyses the corresponding web pages, and extracts various data points. Here's a breakdown of the approach:

1. **Web Scraping:**
   * It uses the requests library to download the webpage content for each URL.
   * Beautiful Soup parses the downloaded HTML to extract the article title and text.
2. **Text Preprocessing**
   * The script cleans the extracted text for further analysis. This involves:
     + Removing punctuation.
     + Converting text to lowercase.
     + Removing stop words (common words like "the", "a", and "an").
3. **Text Analysis:**
   * The script performs various analyses on the cleaned text, including:
     + **Sentiment Analysis:** Using a text blob, it calculates positive and negative sentiment scores to gauge the overall sentiment of the article.
     + **Readability Scores:** It calculates metrics like the Flesch-Kincaid reading ease score to assess the text's difficulty level.
     + **Word Complexity:** It may use nltk to analyze word complexity by functions like counting syllables (count\_syllables).
4. **Data Storage:**
   * The script creates a dictionary containing the extracted data (title, text, analysis results) for each URL.
   * This data is saved to individual text files with a specific naming convention ("blackassign{index}.txt").
   * Finally, it creates a Pandas Data Frame from the list of dictionaries and saves it as a CSV file ("extracted\_data.csv").

Note:  The script tried to access two URLs from Blackcoffer Insights:

* <https://insights.blackcoffer.com/how-neural-networks-can-be-applied-in-various-areas-in-the-future/>
* <https://insights.blackcoffer.com/covid-19-environmental-impact-for-the-future/>

 For both URLs, the script received a "404 Client Error: Not Found" message. This means the server couldn't find the requested web pages at those specific addresses.

* So, I have excluded these 2 articles.

2. Running the Script

 Requirements**:**

* Make sure Python is installed (version 3 recommended).
* Install the required libraries: pip install requests, pandas, beautifulsoup4, textblob, nltk

 Preparation**:**

* Use text files for positive and negative words.

 Execution**:**

* Save the script as a Python file.
* Open a terminal or command prompt and navigate to the directory where the script is saved.
* Run the script using the command: python text\_analysis.py

3. Dependencies

The script relies on several external libraries to function:

* **requests:** Handles making HTTP requests to download webpages.
* **pandas:** Used for data manipulation and creating Data Frames for storing and analyzing extracted data.
* **BeautifulSoup:** A popular library for parsing HTML content and extracting specific elements.
* **textblob:** Provides sentiment analysis functionalities and text processing tools.
* **nltk:** Used for advanced natural language processing tasks like tokenization and word complexity analysis (syllables).

**Note:** Install these libraries using pip install <library\_name> before running the script.