**Project Approach Overview:**

I thoroughly reviewed the objective document multiple times to make sure I understood it completely. Then, I looked at the textual analysis document to understand the methodology we need to follow. After that, I examined the input and output files. Finally, I took a detailed look at the contents in the master dictionary folder and the stopwords folder. This step-by-step approach helped me to become familiar with the project requirements and resources before starting the implementation.

**1. \*\*Data Extraction:\*\***

* Applied web scraping techniques to extract text content from provided URLs.
* Conducted manual inspections to identify the HTML tags containing titles and text on various websites. Iteratively updated the code to accommodate different HTML structures for accurate text extraction.
* Detected incomplete text extraction through manual checks on specific websites.
* Adjusted the code by identifying and incorporating additional HTML tags to capture missing text.

**2. \*\*Text Analysis:\*\***

* Leveraged Natural Language Processing (NLP) tools for sentiment analysis, readability, and other essential metrics.
* Complied with the guidelines outlined in the provided document to compute variables such as positive and negative scores.
* Ensured a meticulous analysis, addressing variations in text structures across different websites.

**3. \*\*Data Integration:\*\***

* Merged the extracted data with the text analysis results to form a cohesive and structured dataset.
* The integration process facilitated a unified view of the collected information for further analysis.

**\*\*How to Run the Python file:\*\***

1. Ensure you have Jupyter Notebook installed.

2. Open a terminal or command prompt.

3. Navigate to the directory containing the Jupyter Notebook file.

4. Launch Jupyter Notebook using the command: `jupyter notebook`

6. Locate and open the notebook file (e.g., `your\_notebook.ipynb`).

7. Run the notebook cells sequentially to execute the code.

8. Examine the output, visualizations, and insights directly within the notebook.

**Dependencies:**

* Python (version 3.10)
* Libraries: NLTK, Pandas, Beautiful Soup, request, os.