**EXPLANATION HOW I APPROACHED THE SOLUTION :**

Step 1: Problem Understanding

1. Read and Understand the Requirements:

- Carefully read and understand the problem statement.

- Identify key tasks: web scraping, text analysis (sentiment and readability), and data processing.

Step 2: Setting Up the Environment

2. Install Necessary Libraries:

- Install libraries required for web scraping (`requests` and `beautifulsoup4`), text analysis (`nltk`), and data processing (`pandas`).

3. Create a Virtual Environment (Optional):

- Optionally, set up a virtual environment to manage dependencies and ensure a clean environment for the project.

Step 3: Web Scraping

4. Implement `extract\_text\_from\_url` Function:

- Design a function to extract article text from a given URL using web scraping techniques.

- Utilize `requests` to fetch the HTML content and `BeautifulSoup` for parsing.

Step 4: Text Analysis

5. Prepare Stopwords:

- Identify and combine stopwords from multiple files into a single file (`StopWords.txt`).

- Use the NLTK library to download required resources (`punkt` and `stopwords`).

6. Implement Sentiment Analysis:

- Create a function (`sentiment\_analysis`) to analyze sentiment in the text.

- Use positive and negative word lists for scoring.

7. Implement Readability Analysis:

- Design a function (`readability\_analysis`) to calculate various readability metrics.

- Include metrics such as average sentence length, complex word percentage, Fog Index, etc.

Step 5: Data Analysis and CSV Output

8. Load Input Data:

- Read input data from the Excel file (`Input.xlsx`) using `pandas`.

9. Iterate Through Rows:

- Iterate through each row in the input data.

10. Perform Text Analysis

- For each URL, apply the web scraping function to get the article text.

- Apply sentiment and readability analysis functions to the text.

11. Create DataFrame:

- Combine the results into a Pandas DataFrame.

12. Save Results to CSV:

- Save the DataFrame to a CSV file (`analysis\_results.csv`).  
  
  
  
  
**Steps to Run the Script:**

1. Download the Required Files:

- Ensure you have the following files in the same directory:

- `analysis.py` (the Python script)

- `Input.xlsx` (the input Excel file with URLs)

2. Run the Script:

- Open a terminal or command prompt.

- Navigate to the directory containing the script and input file.

bash

cd /path/to/directory

- Run the script:

bash

python analysis.py

3. Check the Output:

- The script will generate an output CSV file named `analysis\_results.csv`.

- Verify that the CSV file is created in the same directory.

### **Dependencies:**

**Requests:**

Used for making HTTP requests to fetch web page content.

pip install requests

**Beautiful Soup:**

A library for pulling data out of HTML and XML files.

pip install beautifulsoup4

**Pandas:**

A powerful data manipulation and analysis library.

pip install pandas

**NLTK (Natural Language Toolkit):**

A library for working with human language data.

pip install nltk

### **Installation Commands:**

pip install requests beautifulsoup4 pandas nltk