Generating Metrics for URLs

- Instructions Document Introduction: This document provides instructions on how to run a Python script to extract various metrics for a list of URLs. The metrics include Positive Score, Negative Score, Polarity Score, Subjectivity Score, Average Sentence Length, Percentage of Complex Words, Fog Index, Word Count, and other related metrics.

Approach: The approach taken to generate the metrics involves the following steps: Data Preparation: Load the list of URLs from an Excel file. Load stop words and positive/negative word lists. Metric Calculation Functions: Define functions to calculate each metric based on the text content extracted from the URLs. URL Processing: Iterate through each URL, fetch the content, and calculate the metrics. DataFrame Manipulation: Store the metrics in a DataFrame and extract them into separate columns. Output Generation: Write the DataFrame with the calculated metrics to an Excel file.

Dependencies:

Ensure you have the following dependencies installed: pandas requests nltk beautifulsoup4

Running the Script: <https://colab.research.google.com/drive/1igd9iEMva8euhuEEwJIzwIicayXzerNT?usp=sharing>