Practice Tasks

Practice Task 1: Text Preprocessing

- **Objective**: Clean and preprocess a provided text.
- **Instructions**:
    1. Import necessary libraries (**nltk**, **re**).
    2. Define a new text paragraph.
    3. Remove punctuation and convert the text to lowercase.
    4. Tokenize the cleaned text into words.
    5. Print the cleaned and tokenized words.

Practice Task 2: Custom Stop Words Removal

- **Objective**: Implement a custom list of stop words.
- **Instructions**:
    1. Use the same text from Task 1.
    2. Create a custom list of stop words (e.g., **["the", "is", "and", "to"]**).
    3. Tokenize the text and remove both NLTK's stop words and your custom stop words.
    4. Print the remaining words after stop words removal.

Practice Task 3: Word Frequency Analysis

- **Objective**: Perform frequency analysis on a new text.
- **Instructions**:
    1. Define a new text paragraph.
    2. Tokenize the text into words and create a frequency distribution.
    3. Print the frequency of each word.
    4. Identify and print the three least common words in the text.

Practice Task 4: Visualization of Word Frequencies

- **Objective**: Visualize the frequency distribution of words.
- **Instructions**:

1. Use the frequency distribution created in Task 3.
2. Create a bar chart showing the top 10 most common words.
3. Label the axes and give the chart a title.

Practice Task 5: Sentence Length Analysis

- **Objective**: Analyze the length of sentences in a text.
- **Instructions**:
    1. Define a new text paragraph.
    2. Tokenize the text into sentences.
    3. Calculate the length (number of words) of each sentence.
    4. Print the lengths of all sentences and their average length.

Practice Task 6: Part-of-Speech Tagging

- **Objective**: Perform part-of-speech (POS) tagging on a text.
- **Instructions**:
    1. Define a new text paragraph.
    2. Tokenize the text into sentences.
    3. For each sentence, tokenize it into words.
    4. Use NLTK's **pos_tag** function to tag each word with its part of speech.
    5. Print the original sentences along with their corresponding POS tags.