Python assignments for text pre-processing and creating tf-idf matrix

Assignment-1:

1. Read the "shakespeare-macbeth.txt" text file in python.
2. Apply the necessary pre-processing on the text data.
    a. Remove the stop words and store the tokens in a list.
    b. Removing punctuations.
    c. Applying the required regular expressions to clean the text.
3. How many words are there in the text?
4. How many sentences are there in the text?
5. How many unique words are there in the text?
6. What is the average length of a word in the text?
7. What are the 10 most common unigrams in the processed text?
8. What are the 10 most common bigrams in the processed text?
9. Write code to generate a word cloud

Assignment-2:

1. Read the "ende.json" text file in python.
2. Consider the "Article" column from the data.
3. Apply the necessary pre-processing on the text.
4. Create a tf-idf Matrix.