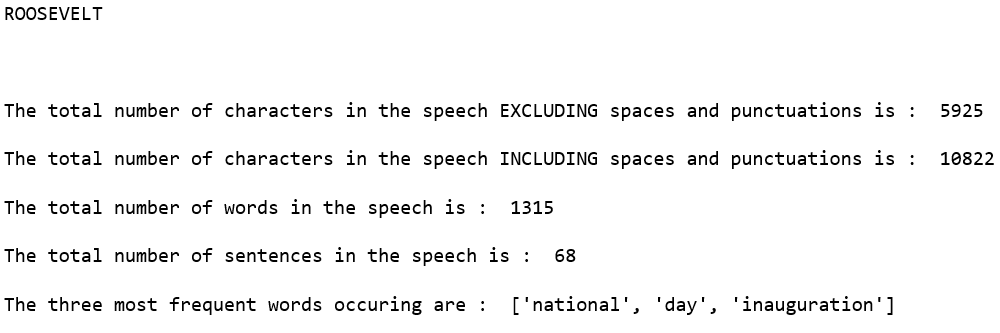
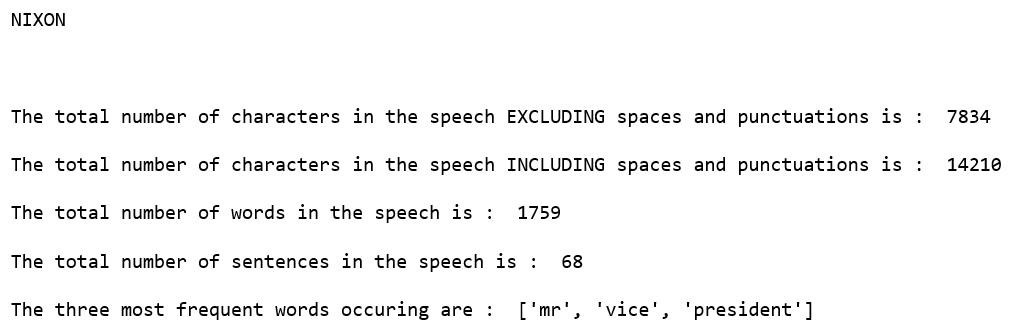
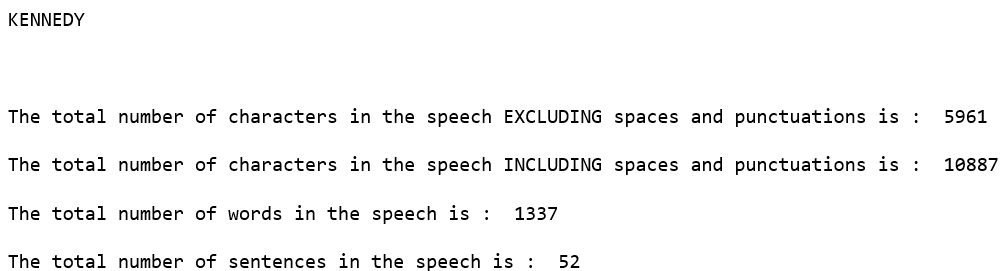
# Problem 2:

# In this particular project, we are going to work on the inaugural corpora from the nltk in Python. We will be looking at the following speeches of the Presidents of the United States of America:

# President Franklin D. Roosevelt in 1941, President John F. Kennedy in 1961, President Richard Nixon in 1963

* **Find the number of characters, words and sentences for the mentioned documents**





* **Remove all the stopwords from all the three speeches.**

This has been done using the corpus from NLTK to get Stopwords in English. The following are a few command snippets used. Please refer to the code for more details.



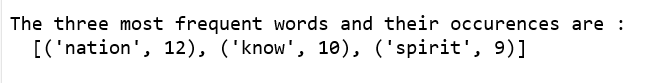




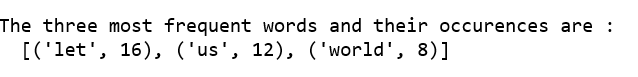
.

* Which word occurs the most number of times in his inaugural address for each president? Mention the top three words. (after removing the stopwords)

ROOSEVELT



NIXON

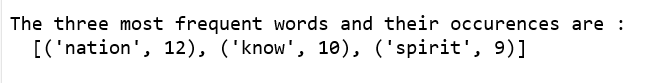


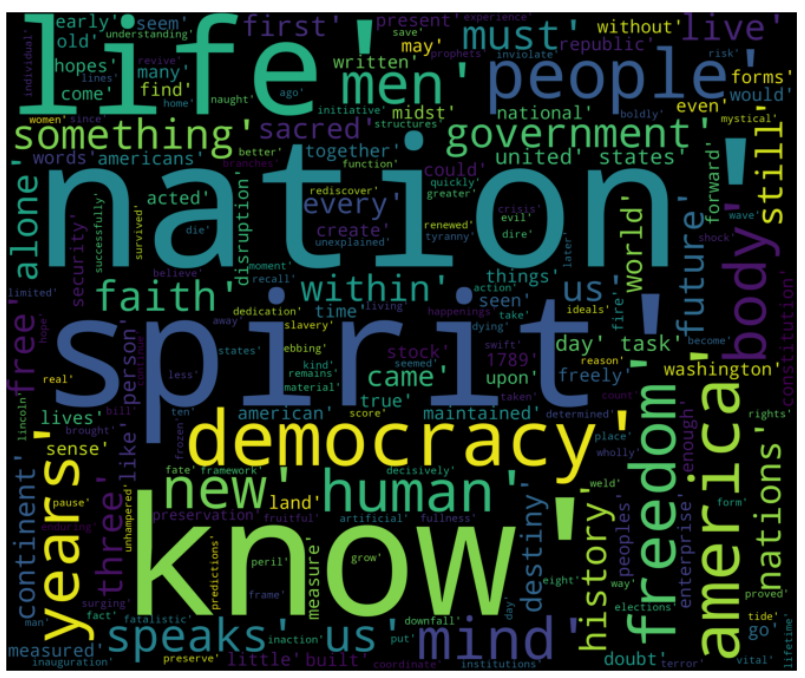
KENNEDY



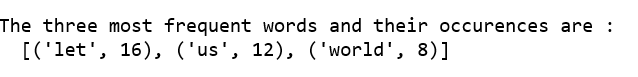
* Plot the word cloud of each of the speeches of the variable. (after removing the stopwords) – 3 Marks [ refer to the End-to-End Case Study done in the Mentored Learning Session ]

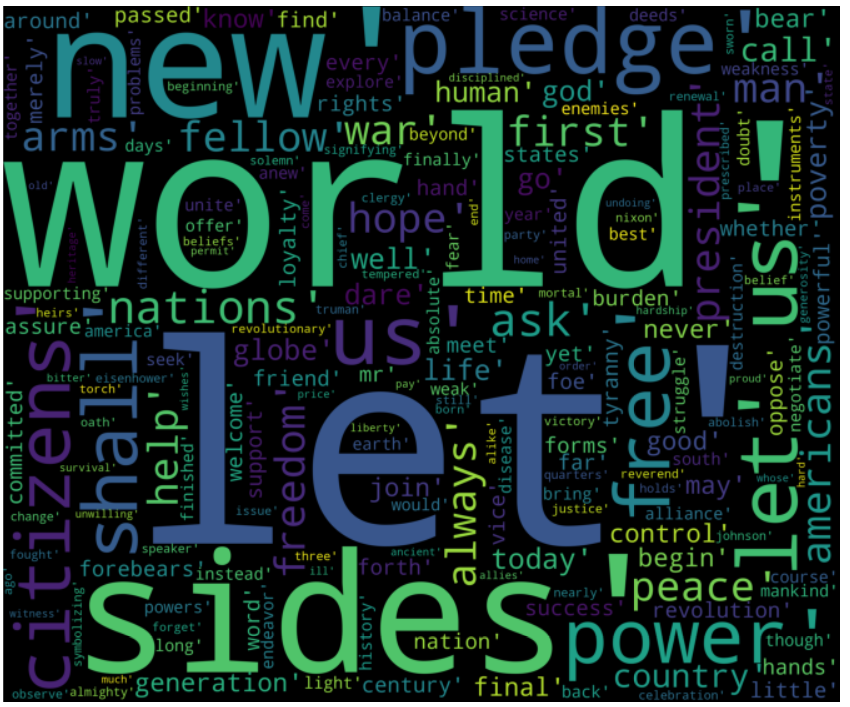
ROOSEVELT





NIXON





KENNEDY



