

Problem 2 : Speech Analysis

***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 1973 (Hint: use .words(), .raw(), .sent() for extracting counts)***

***Questions:***

2.1 Find the number of characters, words, and sentences for the mentioned documents. – 3 Marks  
2.1.1 : Number of characters, words, and sentences for Franklin D. Roosevelt  
2.1.2 : Number of characters, words, and sentences for John F. Kennedy  
2.1.3 : Number of characters, words, and sentences for Richard Nixon

2.2 Remove all the stopwords from all three speeches. – 3 Marks  
2.2.1 : Remove all the stopwords from Speech Franklin D. Roosevelt  
2.2.2 : Remove all the stopwords from Speech John F. Kennedy  
2.2.3 : Remove all the stopwords from Speech Richard Nixon

2.3 Which word occurs the most number of times in his inaugural address for each president? Mention the top three words.(after removing the stopwords) – 3 Marks  
2.3.1 : Word occurs the most in speech Franklin D. Roosevelt  
2.3.2 : Word occurs the most in speech John F. Kennedy  
2.3.3 : Word occurs the most in speech Richard Nixon

2.4 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 ]  
2.4.1 : WordCloud for Franklin D. Roosevelt  
2.4.2 : WordCloud for John F. Kennedy  
2.4.3 : WordCloud for Richard Nixon

Code Snippet to extract the three speeches:

" import nltk nltk.download('inaugural') from nltk.corpus import inaugural inaugural.fileids() inaugural.raw('1941-Roosevelt.txt') inaugural.raw('1961-Kennedy.txt') inaugural.raw('1973-Nixon.txt') "

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

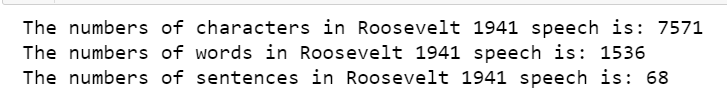
2.1.1 : Number of characters, words, and sentences for Franklin D. Roosevelt.  
2.1.2 : Number of characters, words, and sentences for John F. Kennedy  
2.1.3 : Number of characters, words, and sentences for Richard Nixon

**2.1.1 : Number of characters, words, and sentences for Franklin D. Roosevelt.**

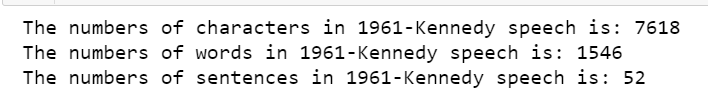
First download the inaugural and import the text files “1941-Roosevelt.txt”,

“1961-Kennedy.txt”.” 1973-Nixon.txt”.

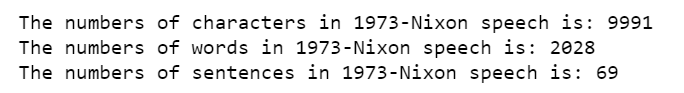
And do the analysis.



**2.1.2 : Number of characters, words, and sentences for John F. Kennedy**



**2.1.3 : Number of characters, words, and sentences for Richard Nixon**



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

2.2.1 : Remove all the stopwords from Speech Franklin D. Roosevelt  
2.2.2 : Remove all the stopwords from Speech John F. Kennedy  
2.2.3 : Remove all the stopwords from Speech Richard Nixon

**2.2.1 : Remove all the stopwords from Speech Franklin D. Roosevelt**

***Download stopwoprds from nltk.***

***After removing all the stopwords from Speech Franklin D. Roosevelt***

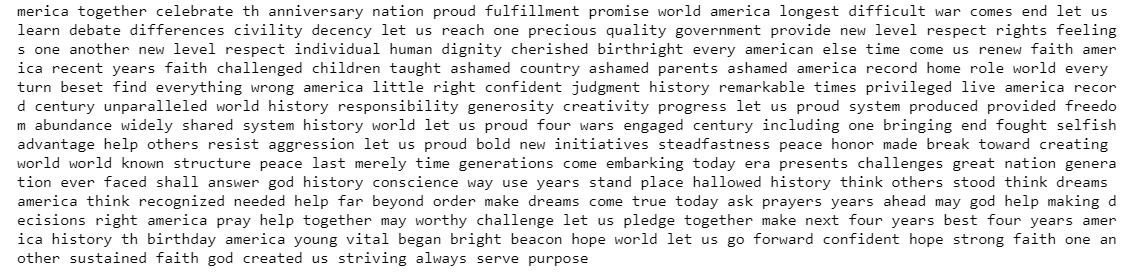


**2.2.2 : Remove all the stopwords from Speech John F. Kennedy**



**2.2.3 : Remove all the stopwords from Speech Richard Nixon.**



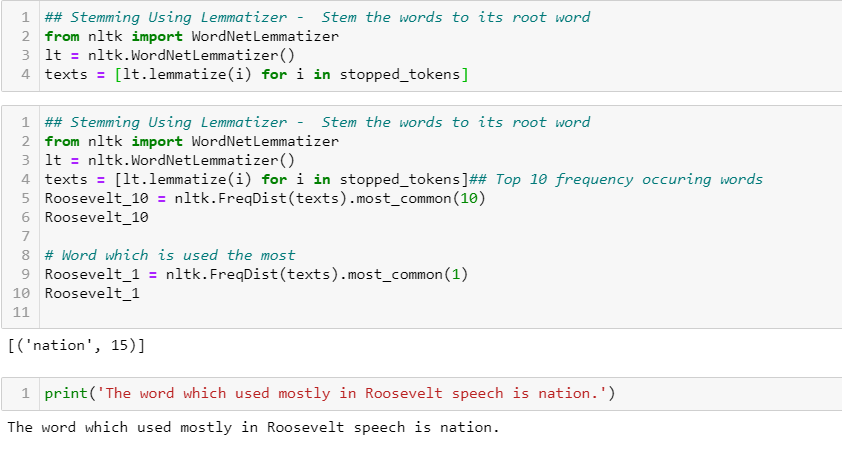


**2.3 Which word occurs the most number of times in his inaugural address for each president? Mention the top three words.(after removing the stopwords)**

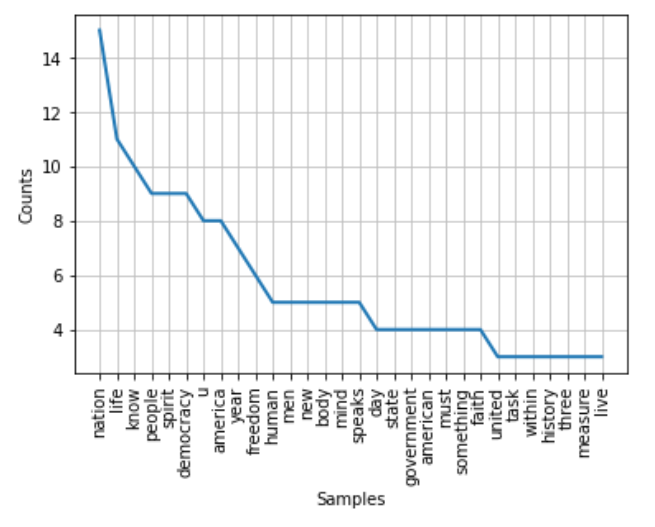
2.3.1 : Word occurs the most in speech Franklin D. Roosevelt  
2.3.2 : Word occurs the most in speech John F. Kennedy  
2.3.3 : Word occurs the most in speech Richard Nixon

**2.3.1 : Word occurs the most in speech Franklin D. Roosevelt**

Nation

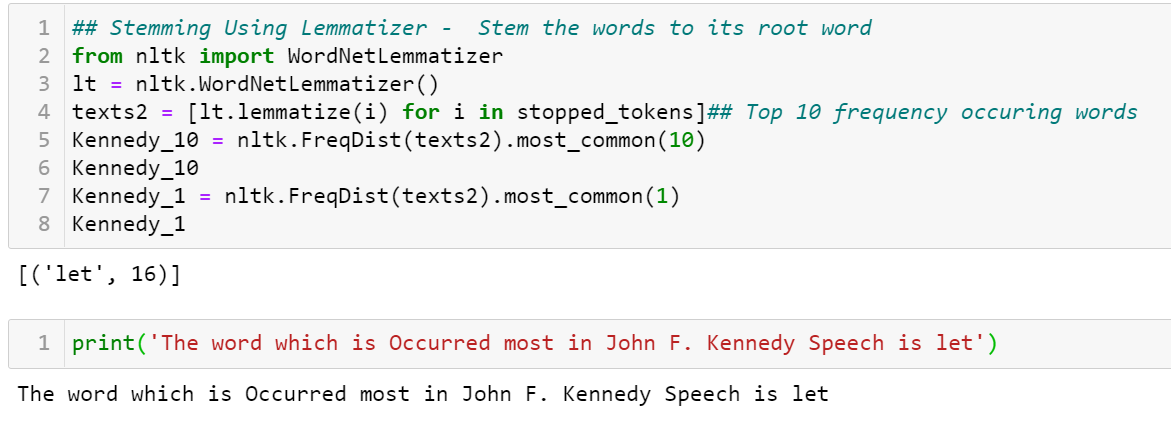


**Frequency plot:**



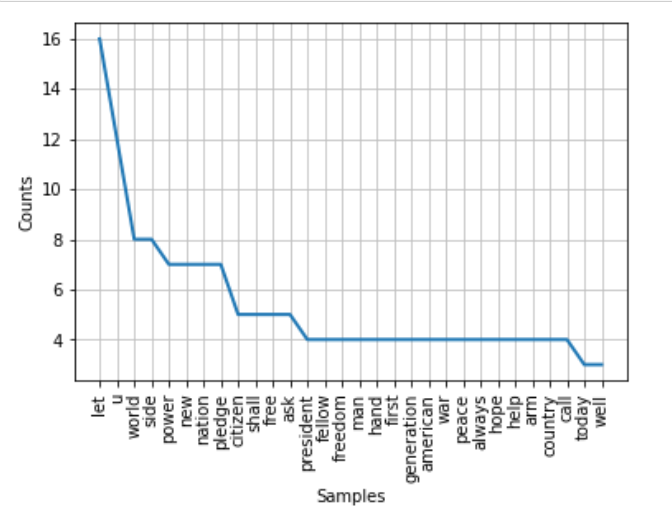
**2.3.2 : Word occurs the most in speech John F. Kennedy**

Let



Here if we additionally remove some words, **world** is the most used words.

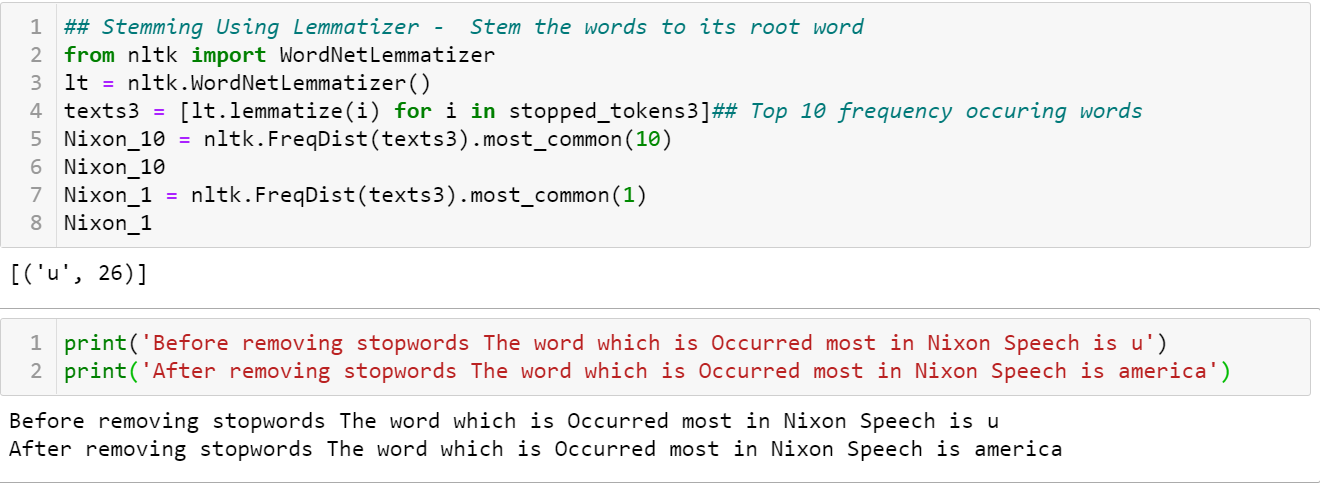
**Frequency plot:**



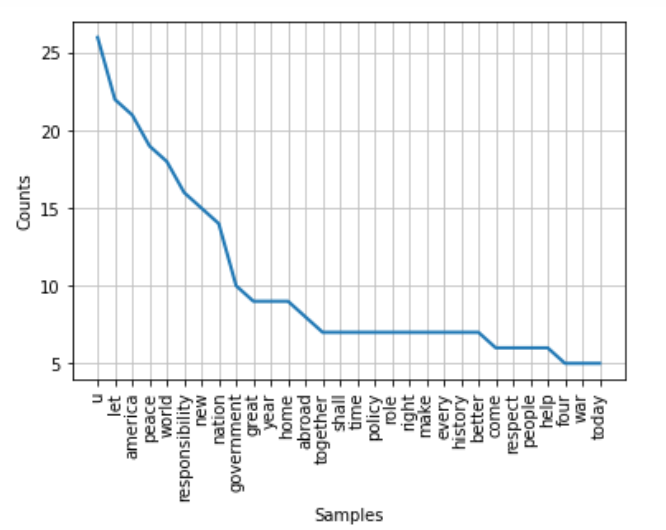
**2.3.3 : Word occurs the most in speech Richard Nixon**

u (Before removing user-defined stopwords like let and u )

america (After removing user-defined stopwords like let and u )



**Frequency plot:**



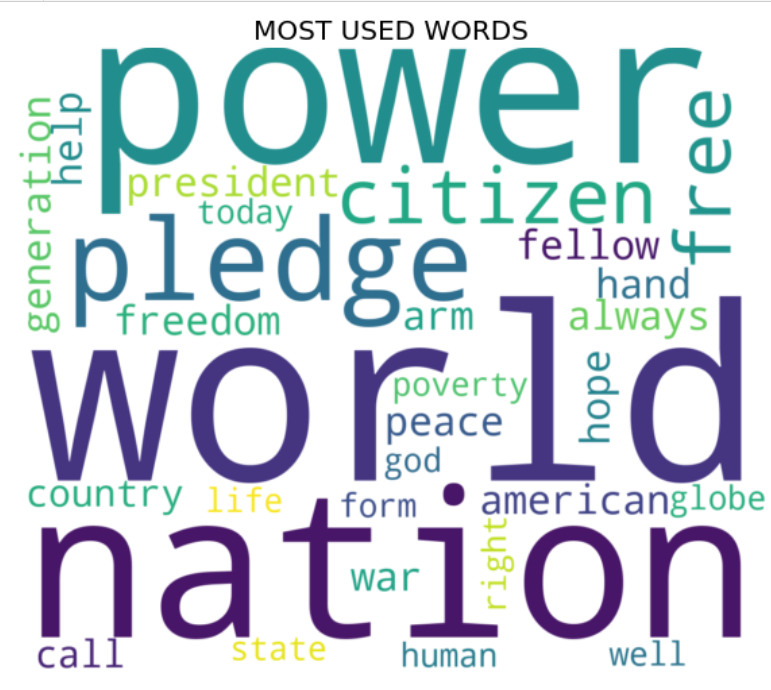
**2.4 Plot the word cloud of each of the speeches of the variable. (after removing the stopwords)**

2.4.1 : WordCloud for Franklin D. Roosevelt  
2.4.2 : WordCloud for John F. Kennedy  
2.4.3 : WordCloud for Richard Nixon

**2.4.1 : WordCloud for Franklin D. Roosevelt**



**2.4.2 : WordCloud for John F. Kennedy**



**2.4.3 : WordCloud for Richard Nixon**

