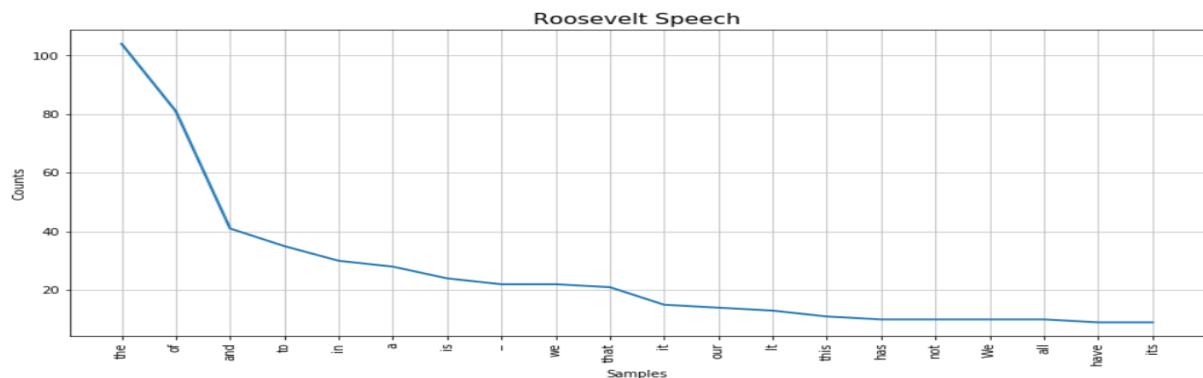# Problem 2 Machine Learning

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:

1.President Franklin D. Roosevelt in 1941
2.President John F. Kennedy in 1961
3.President Richard Nixon in 1973

2.1. Find the number of characters, words and sentences for the mentioned documents. (Hint: use .words(), .raw(), .sent() for extracting counts)
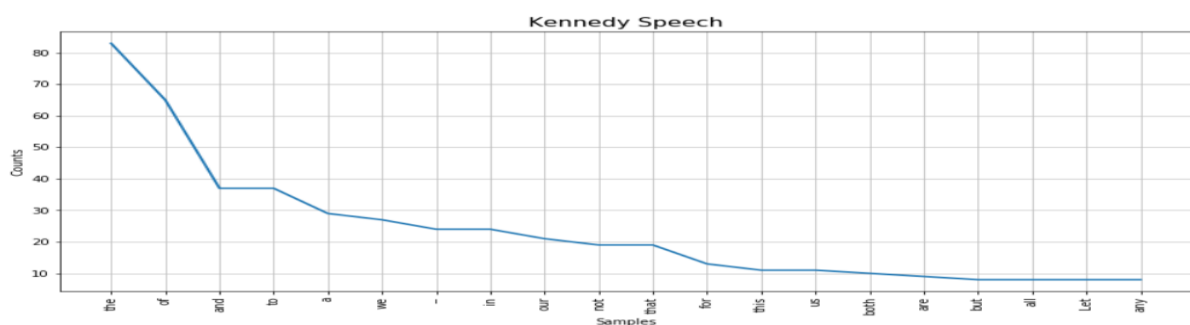
Roosevelt Speech:

'On each national day of inauguration since 1789, the people have renewed their sense of dedication to the United States.\n\nIn Washington\'s day the task of the people was to create and weld together a nation.\n\nIn Lincoln\'s day the task of the people was to preserve that Nation from disruption from within.\n\nIn this day the task of the people is to save that Nation and its institutions from disruption from without...'

'the' is the most frequent word in the Roosevelt speech and has repeated for **104** times in the speech.
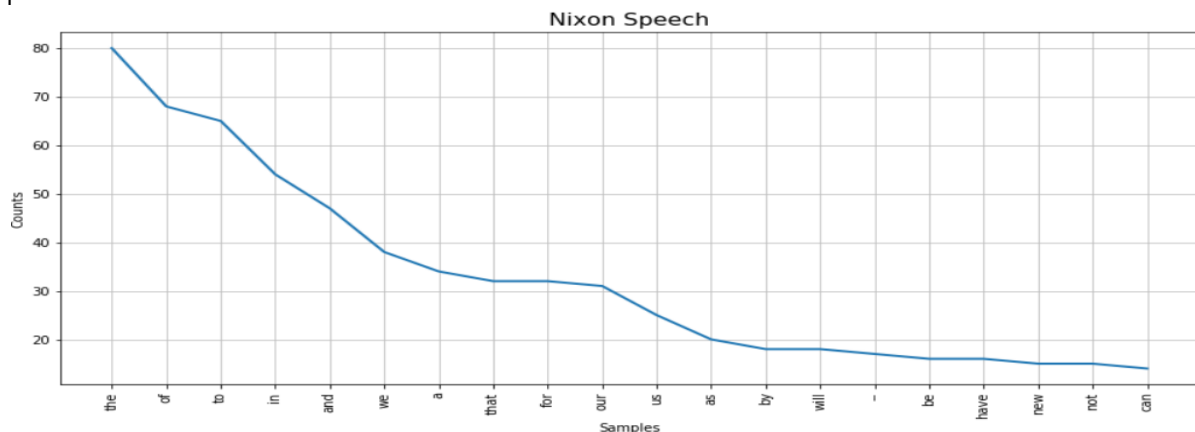
Kennedy Speech :

'Vice President Johnson, Mr. Speaker, Mr. Chief Justice, President Eisenhower, Vice President Nixon, President Truman, reverend clergy, fellow citizens, we observe today not a victory of party, but a celebration of freedom -- symbolizing an end, as well as a beginning -- signifying renewal, as well as change. For I have sworn I before you and Almighty God the same solemn oath our forebears I prescribed nearly a century and three quarters ago...'

'the' is the most frequent word in the Kennedy speech and has repeated for **83** times in the speech.

Nixon Speech :

'Mr. Vice President, Mr. Speaker, Mr. Chief Justice, Senator Cook, Mrs. Eisenhower, and my fellow citizens of this great and good country we share together:\n\nWhen we met here four years ago, America was bleak in spirit, depressed by the prospect of seemingly endless war abroad and of destructive conflict at home.\n\nAs we meet here today, we stand on the threshold of a new era of peace in the world.\n\nThe central question before us is: How shall we use that peace? Let us resolve that this era we are about to enter will not be what other postwar periods..'

Nixon Speech

'the' is the most frequent word in the Kennedy speech and has repeated for **80** times in the speech.

Comparing details of all the three speeches , we can infer following results

Speech		totalwords	word_count	char_count	avg_word	stopwords	upper_letter	numerics	uppercase_words
Roosevelt	On each national day of inauguration since 178...	1360	1323	7571	4.539706	632	119	14	3
Kennedy	Vice President Johnson, Mr. Speaker, Mr. Chief...	1390	1364	7618	4.461871	618	94	7	5
Nixon	Mr. Vice President, Mr. Speaker, Mr. Chief Jus...	1819	1769	9991	4.465091	899	132	10	14

Nixon Speech has highest value in the variables
'totalwords','word_count','char_count','stopwords','upper_letter' and 'uppercase_words'

Roosevelt Speech has highest value in the variables 'avg_word' and 'numerics'

2.2. Remove all the stopwords from the three speeches.

Original Speech before text processing:

	Speech
Roosevelt	On each national day of inauguration since 178...
Kennedy	Vice President Johnson, Mr. Speaker, Mr. Chief...
Nixon	Mr. Vice President, Mr. Speaker, Mr. Chief Jus...

	Speech	stopwords
Roosevelt	On each national day of inauguration since 178...	632
Kennedy	Vice President Johnson, Mr. Speaker, Mr. Chief...	618
Nixon	Mr. Vice President, Mr. Speaker, Mr. Chief Jus...	899

Roosevelt Speech has **632** no of stopwords
Kennedy Speech has **618** no of stopwords
Nixon Speech has **899** no of stopwords

Speech after removal of stopwords:
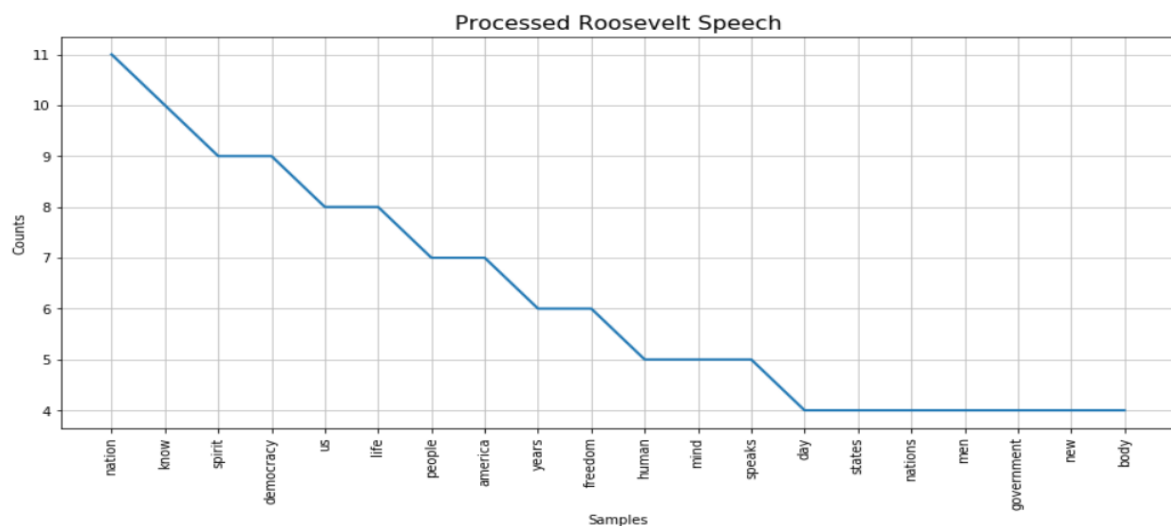
```
Roosevelt      national day inauguration since 1789 people re...
Kennedy        vice president johnson mr speaker mr chief jus...
Nixon          mr vice president mr speaker mr chief justice ...
Name: Processed_Speech, dtype: object
```

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)

After processing text in the inaugural speech of each president,
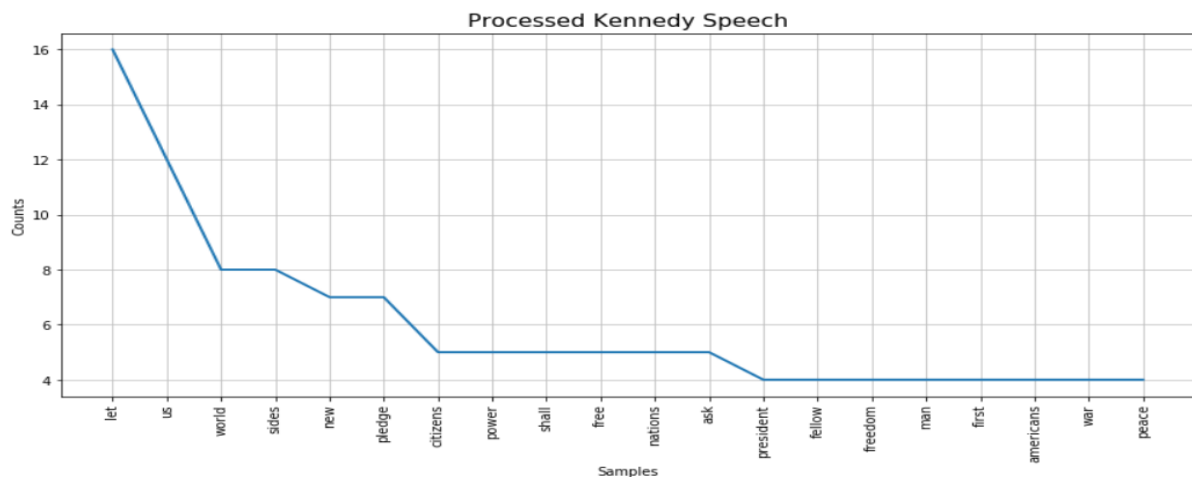**Roosevelt Speech:**

'national day inauguration since 1789 people renewed sense dedication united states washingtons day task people create weld together nation lincolns day task people preserve nation disruption within day task people save nation institutions disruption without us come time midst swift happenings pause moment take stock recall place history rediscover may risk real peril inaction lives nations determined count years lifetime human spirit..'

Processed Roosevelt Speech

'nation','know' and 'spirit' are the three most frequent words in the Roosevelt speech and repeated for **11,10 and 9** times in the speech respectively.
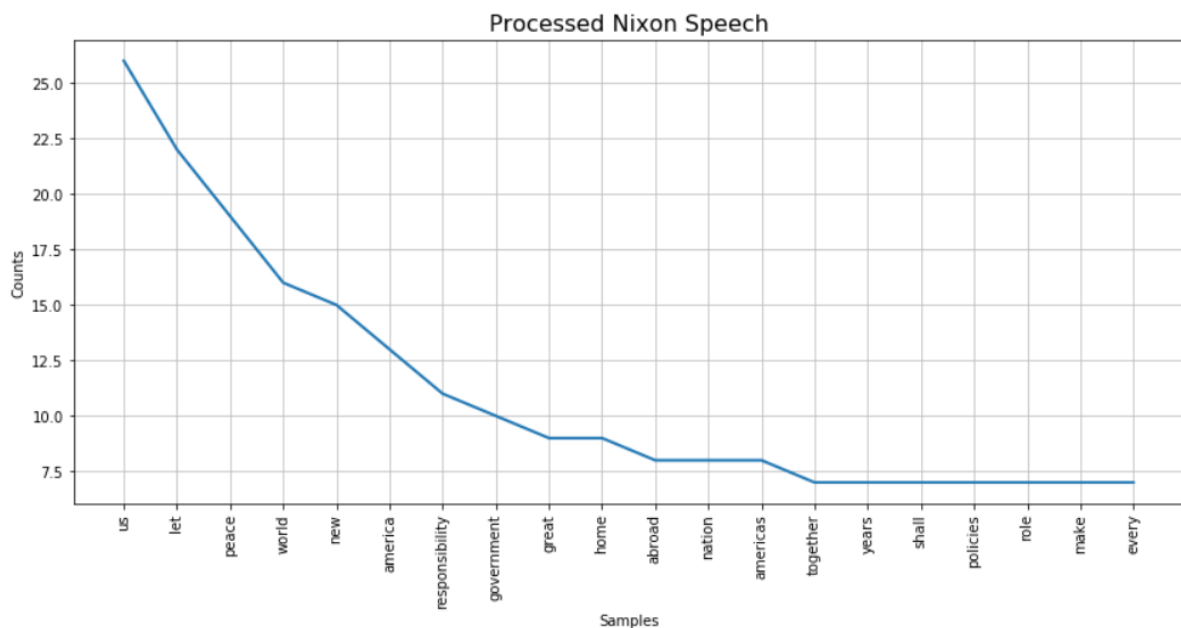
Kennedy Speech :

'vice president johnson mr speaker mr chief justice president eisenhower vice president nixon president truman reverend clergy fellow citizens observe today victory party celebration freedom symbolizing end well beginning signifying renewal well change sworn almighty god solemn oath forebears l prescribed nearly century three quarters ago world different man holds mortal hands power abolish forms human poverty forms human life yet revolutionary beliefs forebears fought still issue around globe belief rights man come generosity state..'

'let','us' and 'world' are the three most frequent words in the Kennedy speech and repeated for **16,12 and 8** times in the speech respectively.

Nixon Speech :

'mr vice president mr speaker mr chief justice senator cook mrs eisenhower fellow citizens great good country share together met four years ago america bleak spirit depressed prospect seemingly endless war abroad destructive conflict home meet today stand threshold new era peace world central question us shall use peace let us resolve era enter postwar periods often time retreat isolation leads stagnation home..'

'us','let' and 'peace' are the three most frequent words in the Nixon speech and repeated for **26,22 and 19** times in the speech respectively.

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

Word Cloud of Roosevelt Speech :

Word Cloud for Roosevelt Speech (after cleaning)!!

nation',’know’ and ‘spirit’ are the three most frequent words in the Roosevelt speech

Word Cloud of Kennedy Speech :

Word Cloud for Kennedy Speech (after cleaning)!!

‘let’,’us’ and ‘world’ are the three most frequent words in the Kennedy speech

Word Cloud of Nixon Speech :

Word Cloud for Nixon Speech (after cleaning)!!

'us','let' and 'peace' are the three most frequent words in the Nixon speech