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 1973

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

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. – 3 Marks

The speeches of the three Presidents were downloaded from the nltk.corpus using inaugural.raw(). Using .words(), .raw(), .sent() the various counts were found.

Franklin D. Roosevelt:

```
The number of words in Roosevelt speech is:  1536
The number of words in Roosevelt speech without punctuation is:  1351
The number of characters in Roosevelt speech is:  7571
The number of sentences in Roosevelt speech is:  68
```

Figure 1: Number of words, characters and sentences in President Roosevelt's speech

Figure 1 shows the number of words, characters and sentences in President Roosevelt's speech.

John F. Kennedy:

```
The number of words in Kennedy speech is:  1546
The number of words in Kennedy speech without punctuation is:  1372
The number of characters in Kennedy speech is:  7618
The number of sentences in Kennedy speech is:  52
```

Figure 2: Number of words, characters and sentences in President Kennedy's speech

Figure 2 shows the number of words, characters and sentences in President Kennedy's speech.

Richard Nixon:

```
The number of words in Nixon speech is:  2028
The number of words in Nixon speech without punctuation is:  1820
The number of characters in Nixon speech is:  9991
The number of sentences in Nixon speech is:  69
```

Figure 3: Number of words, characters and sentences in President Nixon's speech

Figure 3 shows the number of words, characters and sentences in President Nixon's speech.

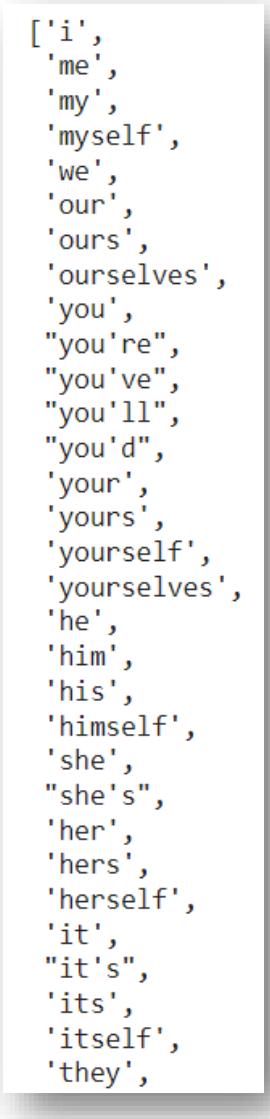
2.2 Remove all the stopwords from all three speeches. – 3 Marks

Stop words are common words that are not useful in providing value or context. They do not add meaning to a sentence. Example of stop words are 'the', 'an', 'in' etc., These words when removed, do not change the meaning of the sentence.

To remove the stop words, the 'stopwords' package is imported from nltk.corpus library.

The stopwords library contains all the stop words like 'and', 'a', 'is', 'to', 'is', '.', 'of', 'to' etc., that do not have any importance in understanding the sentiment or usefulness in machine learning algorithms. These stopwords are already present in the package and it is a set of unirversally accepted stopwords. We can add or include stopwords using the .extend() function or remove them as per our need. The list of words before and after removing the stopwords is then compared.

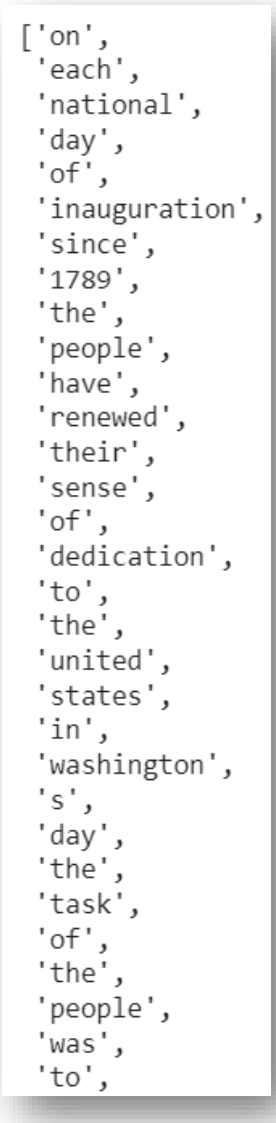
There are many language packages. Therefore, it is necessary to specify the language we are working with. In this case, we will use 'English' as the speeches of the Presidents are in English.



```
['i',
 'me',
 'my',
 'myself',
 'we',
 'our',
 'ours',
 'ourselves',
 'you',
 "you're",
 "you've",
 "you'll",
 "you'd",
 'your',
 'yours',
 'yourself',
 'yourselves',
 'he',
 'him',
 'his',
 'himself',
 'she',
 "she's",
 'her',
 'hers',
 'herself',
 'it',
 "it's",
 'its',
 'itself',
 'they',
```

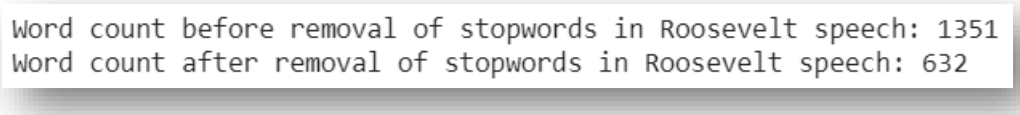
Figure 4: A snippet of stopwords that were removed from the speeches

Franklin D. Roosevelt:



```
['on',
 'each',
 'national',
 'day',
 'of',
 'inauguration',
 'since',
 '1789',
 'the',
 'people',
 'have',
 'renewed',
 'their',
 'sense',
 'of',
 'dedication',
 'to',
 'the',
 'united',
 'states',
 'in',
 'washington',
 's',
 'day',
 'the',
 'task',
 'of',
 'the',
 'people',
 'was',
 'to',
```

Figure 5: A snippet of words without punctuations in President Roosevelt's speech



```
Word count before removal of stopwords in Roosevelt speech: 1351
Word count after removal of stopwords in Roosevelt speech: 632
```

Figure 6: Word count before and after removal of stopwords in President Roosevelt's speech

Sample sentence: national day inauguration since 1789 people renewed sense dedication united states washington day task people create weld together nation Lincoln…

John F. Kennedy:

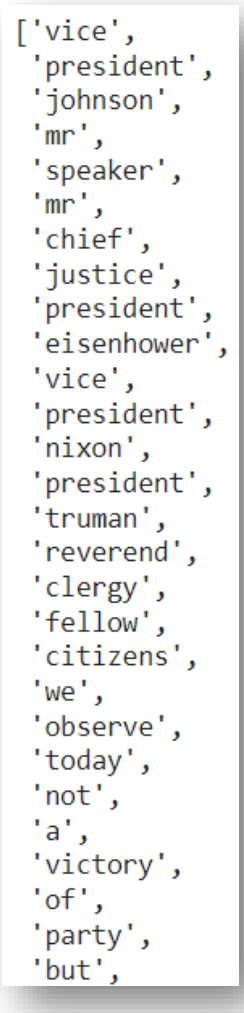


Figure 7: A snippet of words without punctuations in President Kennedy's speech

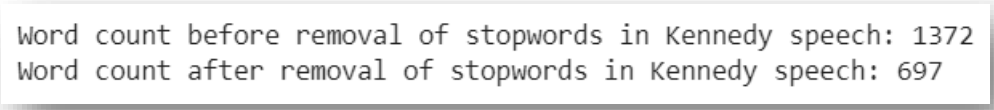
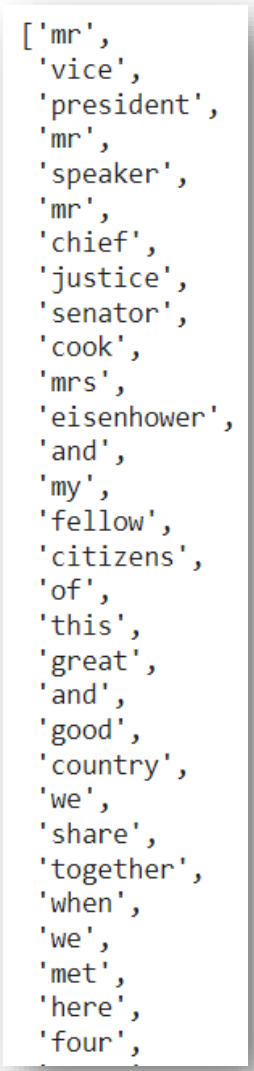


Figure 8: Word count before and after removal of stopwords in President Kennedy's speech

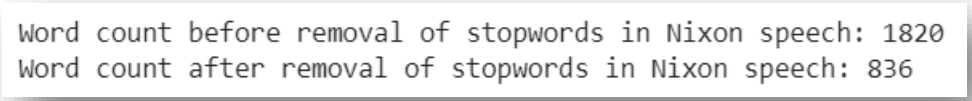
Sample sentence: vice president johnson mr speaker mr chief justice president eisenhower vice president nixon president truman reverend clergy fellow citizens…

Richard Nixon:



```
['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',
 'when',
 'we',
 'met',
 'here',
 'four',
```

Figure 9: A snippet of words without punctuations in President Nixon's speech



```
Word count before removal of stopwords in Nixon speech: 1820
Word count after removal of stopwords in Nixon speech: 836
```

Figure 10: Word count before and after removal of stopwords in President Nixon's speech

Sample sentence: mr vice president mr speaker mr chief justice senator cook mrs eisenhower fellow citizens great good country share together met four years ago…

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

After removing the stop words, the speech of each president is checked for the word that occurs most number of times using nltk.FreqDist() function.

Franklin D. Roosevelt:

```
[('nation', 12),
 ('know', 10),
 ('spirit', 9),
 ('life', 9),
 ('democracy', 9),
 ('us', 8),
 ('people', 7),
 ('america', 7),
 ('years', 6),
 ('freedom', 6)]
```

Figure 11: Frequency distribution of the first 10 words in President Roosevelt's speech

```
The top three words in President Roosevelt speech after removing stop words are:
 [('nation', 12), ('know', 10), ('spirit', 9), ('life', 9), ('democracy', 9)]
```

Figure 12: The top three occurring words in President Roosevelt's speech

Figure 11 shows the frequency distribution of the first 10 most occurring words. Figure 12 shows the top three words that occurs most number of times in the inaugural speech of President Roosevelt. The number within the bracket indicates the number of times the word appears in the speech. Here 'spirit', 'life', 'democracy' are all in the third place as it all occurs the same number of times. The **most occurring** word is **'Nation'**.

John F. Kennedy:

```
[('let', 16),
 ('us', 12),
 ('world', 8),
 ('sides', 8),
 ('new', 7),
 ('pledge', 7),
 ('citizens', 5),
 ('power', 5),
 ('shall', 5),
 ('free', 5)]
```

Figure 13: Frequency distribution of the first 10 words in President Kennedy's speech

```
The top three words in President Kennedy speech after removing stop words are:
 [('let', 16), ('us', 12), ('world', 8)]
```

Figure 14: The top three occurring words in President Kennedy's speech

Figure 13 shows the frequency distribution of the first 10 most occurring words. Figure 14 shows the top three words that occurs most number of times in the inaugural speech of President Kennedy. The number within the bracket indicates the number of times the word appears in the speech. Here **'let', 'us' and 'world'** are top three words. The **most occurring** word is **'let'**.

Richard Nixon:

```
[('us', 26),
 ('let', 22),
 ('america', 21),
 ('peace', 19),
 ('world', 18),
 ('new', 15),
 ('nation', 11),
 ('responsibility', 11),
 ('government', 10),
 ('great', 9)]
```

Figure 15: Frequency distribution of the first 10 words in President Nixon's speech

```
The top three words in President Nixon speech after removing stop words are:
 [('us', 26), ('let', 22), ('america', 21)]
```

Figure 16: The top three occurring words in President Nixon's speech

Figure 15 shows the frequency distribution of the first 10 most occurring words. Figure 16 shows the top three words that occurs most number of times in the inaugural speech of President Nixon. The number within the bracket indicates the number of times the word appears in the speech. Here **'us', 'let' and 'america'** are top three words. The **most occurring** word is **'us'**.

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]

Word clouds or tag clouds are graphical representations of word frequency that give greater prominence to words that appear more frequently in a source text. In other words the size of each word indicates its frequency of occurrence or importance. To create word cloud, the 'word-cloud' package is needed.

Word cloud is generated after removing the stop words. The stop words should be removed before creating the word cloud as removing stop words will remove the unwanted words that have no value or no sentiment analysis.

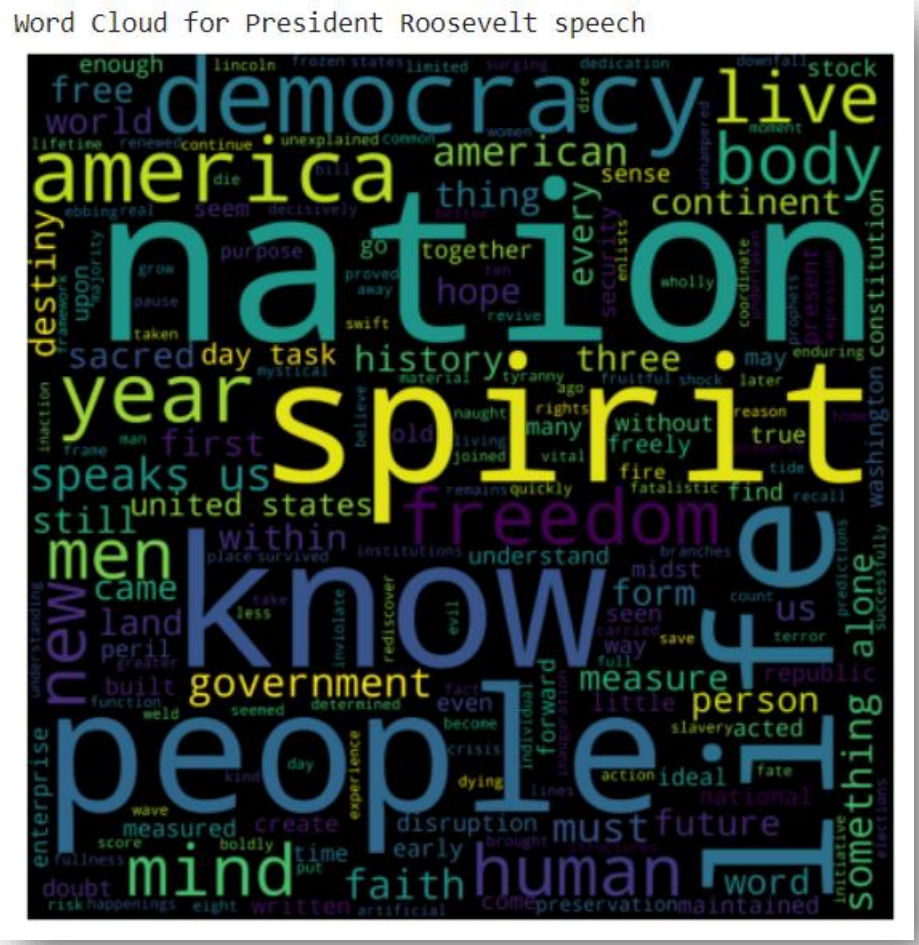


Figure 17: Word Cloud of President Roosevelt's speech

Figure 17 shows the word cloud of President Roosevelt's speech. It is observed that words that had occurred as top words in Figure 11 are seen to be highlighted in Figure 17. The more the frequency of the word, the bigger the size of the word.

Figure 18: Word Cloud of President Kennedy's speech

Figure 18 shows the word cloud of President Kennedy's speech. It is observed that words that had occurred as top words in Figure 13 are seen to be highlighted in Figure 18. The more the frequency of the word, the bigger the size of the word.

Figure 19: Word Cloud of President Nixon's speech

Figure 19 shows the word cloud of President Nixon's speech. It is observed that words that had occurred as top words in Figure 15 are seen to be highlighted in Figure 19. The more the frequency of the word, the bigger the size of the word.

Observations and insights:

The objective of this problem was to analyse the speeches of 3 of the presidents of America. Through the analysis, the strength and sentiment of the speeches was found. Based on the outputs, it was observed that there are few similar words like 'nation', 'us', 'america' that occur in all the three speeches. These words may have inspired the people to vote for the president and hence helped them win the presidential election of the United States of America. It can be noted that among all the three speeches, 'nation' is the word that is frequently occurring in all2.