

A thick dark blue vertical bar runs down the left side of the page. To its right, several thin, curved lines in dark blue and light grey sweep upwards from the bottom left corner.

# Business Report

## Natural Language Processing (NLP)

Ayush Sharma

# Table of Contents

A. Find the number of characters, words and sentences for the mentioned documents. (Hint: use <code>.words()</code> , <code>.raw()</code> , <code>.sent()</code> for extracting counts)	2
B. Remove all the stopwords from the three speeches. Show the word count before and after the removal of stopwords. Show a sample sentence after the removal of stopwords.	3
C. Which word occurs the most number of times in his inaugural address for each president? Mention the top three words. (after removing the stopwords)	4
D. Plot the word cloud of each of the three speeches. (after removing the stopwords)	5

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

Ans: Following are the number of characters, words and sentences in all the speeches respectively:

Speech 1:  
Characters: 7571  
Words: 1536  
Sentences: 68

Speech 2:  
Characters: 7618  
Words: 1546  
Sentences: 52

Speech 3:  
Characters: 9991  
Words: 2028  
Sentences: 69

- B. Remove all the stopwords from the three speeches. Show the word count before and after the removal of stopwords. Show a sample sentence after the removal of stopwords.**

Ans:

```
Length of words in document 1 before removing the stopwords: 1536
Length of words in document 2 before removing the stopwords: 1546
Length of words in document 3 before removing the stopwords: 2028
```

```
Length of words in document 1 after removing the stopwords: 627
Length of words in document 2 after removing the stopwords: 692
Length of words in document 3 after removing the stopwords: 834
```

Cleaned words in Speech 1:

```
['national', 'day', 'inauguration', 'since', 'people', 'renewed', 'sense', 'dedication', 'united', 'states', 'washington', 'day', 'task', 'people', 'create']
```

Cleaned words in Speech 2:

```
['vice', 'president', 'johnson', 'mr', 'speaker', 'mr', 'chief', 'justice', 'president', 'eisenhower', 'vice', 'president', 'nixon', 'president', 'truman']
```

Cleaned words in Speech 3:

```
['mr', 'vice', 'president', 'mr', 'speaker', 'mr', 'chief', 'justice', 'senator', 'cook', 'mrs', 'eisenhower', 'fellow', 'citizens', 'great']
```

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

Ans: Following are the words that occur the greatest number of times for each president:

- Speech 1 -> Nation (12 times)
- Speech 2 -> Let (16 times)
- Speech 1 -> Us (26 times)

Most common words for Speech 1:

nation	12 times
know	10 times
spirit	9 times

Most common words for Speech 2:

let	16 times
us	12 times
world	8 times

Most common words for Speech 3:

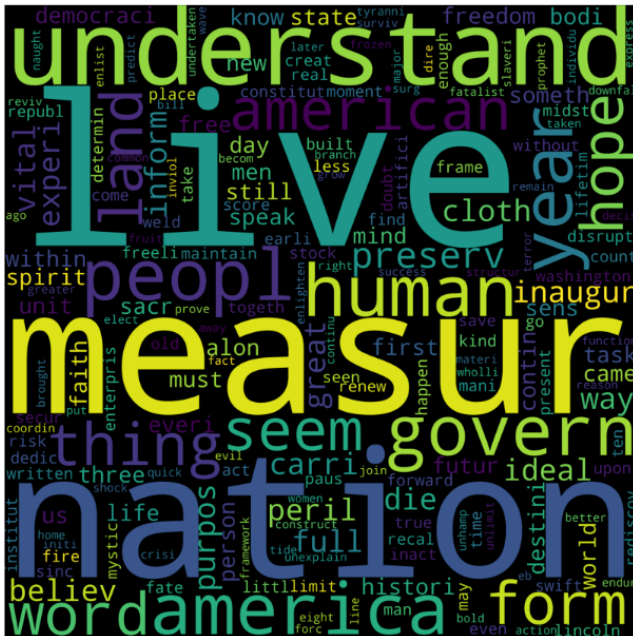
us	26 times
let	22 times
america	21 times

Most common words in all three speeches:

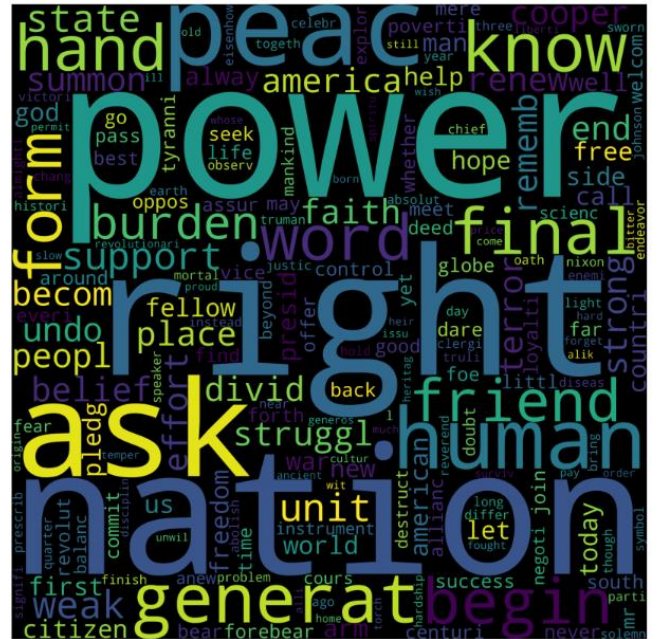
['nation', 'life', 'us', 'people', 'america', 'freedom', 'human', 'new', 'must', 'faith']

D. Plot the word cloud of each of the three speeches. (after removing the stop words)

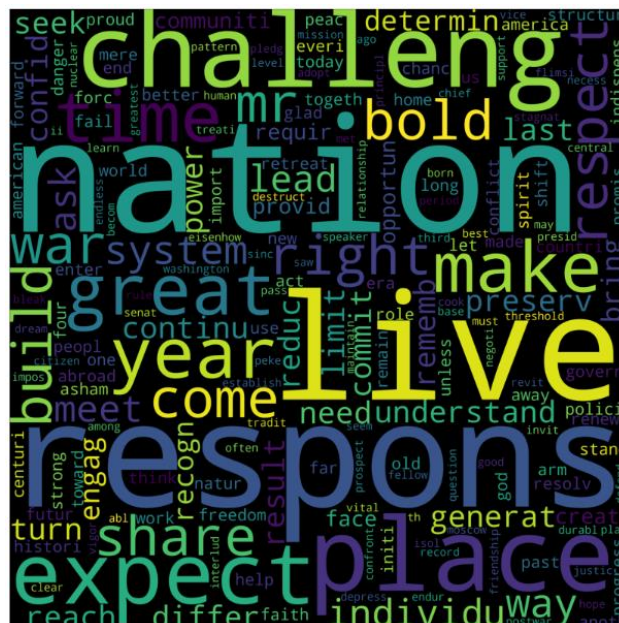
Ans:



Speech 1



Speech 2



Speech 3