

## Q3

August 12, 2025

```
[1]: import nltk
      from nltk.corpus import stopwords
      from collections import Counter
```

```
[2]: nltk.download("punkt")
      nltk.download("punkt_tab")
      nltk.download("stopwords")
      nltk.download("averaged_perceptron_tagger")
      nltk.download("averaged_perceptron_tagger_eng")
```

```
[nltk_data] Downloading package punkt to /Users/sivakumar/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package punkt_tab to
[nltk_data] /Users/sivakumar/nltk_data...
[nltk_data] Package punkt_tab is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] /Users/sivakumar/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data] /Users/sivakumar/nltk_data...
[nltk_data] Package averaged_perceptron_tagger is already up-to-
[nltk_data] date!
[nltk_data] Downloading package averaged_perceptron_tagger_eng to
[nltk_data] /Users/sivakumar/nltk_data...
[nltk_data] Package averaged_perceptron_tagger_eng is already up-to-
[nltk_data] date!
```

```
[2]: True
```

```
[3]: text= '''President Donald Trump announced he is invoking the District of
      ↪Columbia Home Rule Act to bring the D.C. Metropolitan Police Department
      ↪under direct federal control - a rare and contentious step. The US President
      ↪further added that National Guard troops would be deployed to Washington, D.
      ↪C., to "re-establish law, order, and public safety," stressing they would be
      ↪allowed to carry out their duties without restriction.'''
      text
```

```
[3]: 'President Donald Trump announced he is invoking the District of Columbia Home Rule Act to bring the D.C. Metropolitan Police Department under direct federal control - a rare and contentious step. The US President further added that National Guard troops would be deployed to Washington, D.C., to "re-establish law, order, and public safety," stressing they would be allowed to carry out their duties without restriction.'
```

### Tokenization

```
[4]: tokens= nltk.word_tokenize(text)
tokens
```

```
[4]: ['President',
      'Donald',
      'Trump',
      'announced',
      'he',
      'is',
      'invoking',
      'the',
      'District',
      'of',
      'Columbia',
      'Home',
      'Rule',
      'Act',
      'to',
      'bring',
      'the',
      'D.C.',
      'Metropolitan',
      'Police',
      'Department',
      'under',
      'direct',
      'federal',
      'control',
      '-',
      'a',
      'rare',
      'and',
      'contentious',
      'step',
      '.',
      'The',
      'US',
      'President',
      'further',
```

```

'added',
'that',
'National',
'Guard',
'troops',
'would',
'be',
'deployed',
'to',
'Washington',
',',
'D.C.',
',',
'to',
'',
're-establish',
'law',
',',
'order',
',',
'and',
'public',
'safety',
',',
'',
'stressing',
'they',
'would',
'be',
'allowed',
'to',
'carry',
'out',
'their',
'duties',
'without',
'restriction',
'.']

```

## Remove Stop Words

```

[5]: stop_words= set(stopwords.words('english'))

filtered_tokens= [word for word in tokens if word.lower() not in stop_words and
↳word.isalpha()]

filtered_tokens

```

```
[5]: ['President',  
      'Donald',  
      'Trump',  
      'announced',  
      'invoking',  
      'District',  
      'Columbia',  
      'Home',  
      'Rule',  
      'Act',  
      'bring',  
      'Metropolitan',  
      'Police',  
      'Department',  
      'direct',  
      'federal',  
      'control',  
      'rare',  
      'contentious',  
      'step',  
      'US',  
      'President',  
      'added',  
      'National',  
      'Guard',  
      'troops',  
      'would',  
      'deployed',  
      'Washington',  
      'law',  
      'order',  
      'public',  
      'safety',  
      'stressing',  
      'would',  
      'allowed',  
      'carry',  
      'duties',  
      'without',  
      'restriction']
```

### 0.0.1 POS Tagging

```
[6]: pos_tags= nltk.pos_tag(filtered_tokens)  
pos_tags
```

```
[6]: [('President', 'NNP'),
      ('Donald', 'NNP'),
      ('Trump', 'NNP'),
      ('announced', 'VBD'),
      ('invoking', 'VBG'),
      ('District', 'NNP'),
      ('Columbia', 'NNP'),
      ('Home', 'NNP'),
      ('Rule', 'NNP'),
      ('Act', 'NNP'),
      ('bring', 'VB'),
      ('Metropolitan', 'NNP'),
      ('Police', 'NNP'),
      ('Department', 'NNP'),
      ('direct', 'JJ'),
      ('federal', 'JJ'),
      ('control', 'NN'),
      ('rare', 'JJ'),
      ('contentious', 'JJ'),
      ('step', 'NN'),
      ('US', 'NNP'),
      ('President', 'NNP'),
      ('added', 'VBD'),
      ('National', 'NNP'),
      ('Guard', 'NNP'),
      ('troops', 'NNS'),
      ('would', 'MD'),
      ('deployed', 'VB'),
      ('Washington', 'NNP'),
      ('law', 'NN'),
      ('order', 'NN'),
      ('public', 'JJ'),
      ('safety', 'NN'),
      ('stressing', 'VBG'),
      ('would', 'MD'),
      ('allowed', 'VB'),
      ('carry', 'JJ'),
      ('duties', 'NNS'),
      ('without', 'IN'),
      ('restriction', 'NN')]
```

### 0.0.2 Counts

```
[7]: counts= Counter()

for word, tag in pos_tags:
```

```
if tag.startswith("JJ"):
    counts['Adjectives'] += 1
elif tag.startswith("NN"):
    counts['Noun'] += 1
elif tag.startswith("VB"):
    counts['Verb'] += 1

print(counts)
```

Counter({'Noun': 24, 'Verb': 7, 'Adjectives': 6})

[ ]: