```
!pip install nltk

Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-
packages (3.8.1)
Requirement already satisfied: click in
/usr/local/lib/python3.10/dist-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in
/usr/local/lib/python3.10/dist-packages (from nltk) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in
/usr/local/lib/python3.10/dist-packages (from nltk) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-
packages (from nltk) (4.66.2)

import nltk

nltk.download('punkt')

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!

True

nltk.download('stopwords')

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!

True

from nltk import sent_tokenize, word_tokenize

from nltk.corpus import stopwords

stopwords = stopwords.words('english')

sent = "I will walk 500 miles and I would walk 500 more, just to be
the man who walks a thousand miles to fall down at your door"
```

Change all the characters to lower case for simplicity or as a necessity

```
sentences = sent_tokenize(sent)
print(sentences)

['I will walk 500 miles and I would walk 500 more, just to be the man
who walks a thousand miles to fall down at your door']

words1= word_tokenize(sent)
print(words1)

['I', 'will', 'walk', '500', 'miles', 'and', 'I', 'would', 'walk',
'500', 'more', ',', 'just', 'to', 'be', 'the', 'man', 'who', 'walks',
'a', 'thousand', 'miles', 'to', 'fall', 'down', 'at', 'your', 'door']
```

```python
words = [word.lower() for word in words1]
print(words)
```

```
['i', 'will', 'walk', '500', 'miles', 'and', 'i', 'would', 'walk',
'500', 'more', ',', 'just', 'to', 'be', 'the', 'man', 'who', 'walks',
'a', 'thousand', 'miles', 'to', 'fall', 'down', 'at', 'your', 'door']
```

```python
print(stopwords)
```

```
['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', 'them',
'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom',
'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was',
'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do',
'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or',
'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with',
'about', 'against', 'between', 'into', 'through', 'during', 'before',
'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out',
'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once',
'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both',
'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor',
'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't',
'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now',
'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't",
'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn',
"hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma',
'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan',
"shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren',
"weren't", 'won', "won't", 'wouldn', "wouldn't"]
```

```python
for word in words:
  if word in stopwords:
    words.remove(word)
print(words)
```

```
['will', 'walk', '500', 'miles', 'i', 'would', 'walk', '500', ',',
'the', 'man', 'walks', 'thousand', 'miles', 'to', 'fall', 'at',
'door']
```

```python
from nltk import PorterStemmer

stemmer = PorterStemmer()

stemmed_words = [stemmer.stem(word) for word in words]
print(stemmed_words)
```

```
['will', 'walk', '500', 'mile', 'i', 'would', 'walk', '500', ',',
'the', 'man', 'walk', 'thousand', 'mile', 'to', 'fall', 'at', 'door']
```

```python
from collections import Counter
from nltk import pos_tag

nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
```

```
True
```

```python
count_dict= {}
```

```python
for word in stemmed_words:
  if word in count_dict:
    count_dict[word]+=1
  else:
    count_dict[word]=1;
```

```python
print(count_dict)
```

```
{'will': 1, 'walk': 3, '500': 2, 'mile': 2, 'i': 1, 'would': 1, ',':
1, 'the': 1, 'man': 1, 'thousand': 1, 'to': 1, 'fall': 1, 'at': 1,
'door': 1}
```

```python
pos_tagged = pos_tag(stemmed_words)
print(pos_tagged)
```

```
[('will', 'MD'), ('walk', 'VB'), ('500', 'CD'), ('mile', 'NN'), ('i',
'NN'), ('would', 'MD'), ('walk', 'VB'), ('500', 'CD'), (',', ','),
('the', 'DT'), ('man', 'NN'), ('walk', 'NN'), ('thousand', 'VBP'),
('mile', 'NN'), ('to', 'TO'), ('fall', 'VB'), ('at', 'IN'), ('door',
'NN')]
```

```python
count = Counter(tag for _ , tag in pos_tagged)
print(count)
```

```
Counter({'NN': 6, 'VB': 3, 'MD': 2, 'CD': 2, ',': 1, 'DT': 1, 'VBP':
1, 'TO': 1, 'IN': 1})
```