

```
In [20]: '''Lower casing'''  
text = "This is a WONDERFUL Day"  
text.lower()
```

```
Out[20]: 'this is a wonderful day'
```

```
In [21]: '''Sentence Splitter'''  
text = "This is a wonderful day. Today we have SENG 507 class"  
sentences = text.split('.')  
sentences
```

```
Out[21]: ['This is a wonderful day', ' Today we have SENG 507 class']
```

```
In [22]: '''Tokenization'''  
import nltk  
from nltk import word_tokenize  
text = "Please. Tokenize this sentence! can you?!"  
text = nltk.word_tokenize(text)  
text
```

```
Out[22]: ['Please', '.', 'Tokenize', 'this', 'sentence', '!', 'can', 'you', '?', '!']
```

```
In [23]: '''Remove punctuations'''
import re
new_words = []
for word in text:
    new_word = re.sub(r'^\w\s', '', word)
    if new_word != '':
        new_words.append(new_word)
text = new_words
text
```

```
Out[23]: ['Please', 'Tokenize', 'this', 'sentence', 'can', 'you']
```

```
In [25]: """Remove stop words from list of tokenized words"""
from nltk.corpus import stopwords
new_words = []
for word in text:
    if word not in stopwords.words('english'):
        new_words.append(word)
#stopwords.words('english')
text = new_words
text
```

```
Out[25]: ['Please', 'Tokenize', 'sentence']
```

```
In [18]: """Stem words in list of tokenized words"""
from nltk.stem import LancasterStemmer, WordNetLemmatizer
stemmer = LancasterStemmer()
stems = []
for word in text:
    stem = stemmer.stem(word)
    stems.append(stem)
text = stems
text
```

Out[18]: ['pleas', 'tok', 'sent']

```
In [26]: """Lemmatize verbs in list of tokenized words"""
lemmatizer = WordNetLemmatizer()
lemmas = []
for word in text:
    lemma = lemmatizer.lemmatize(word, pos='v')
    lemmas.append(lemma)
text = lemmas
text
```

Out[26]: ['Please', 'Tokenize', 'sentence']

In []: