

```
!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) (8.1.6)
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) (2023.6.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.1)
```

```
import nltk
```

```
nltk.download()
```

```

🔗 NLTK Downloader
-----
    d) Download  l) List  u) Update  c) Config  h) Help  q) Quit
-----
Downloader> q
True
```

Sentence Tokenization

```
from nltk.tokenize import sent_tokenize
```

```
text = '''Martha is at the grocery store, getting ready for a house party. She has a list of what she needs with her as she goes along. The first section she comes has produce. Martha sees apples, bananas, cherries, grapes, and strawberries.'''
```

```

Martha is at the grocery store, getting ready for a house party. She has a list of what she needs with her as she goes along. The first section she comes has produce. Martha sees apples, bananas, cherries, grapes, and strawberries.
```

```
sentences = sent_tokenize (text)
```

```
sentences
```

```

['Martha is at the grocery store, getting ready for a house party.',
 'She has a list of what she needs with her as she goes along.',
 'The first section she comes has produce.',
 'Martha sees apples, bananas, cherries, grapes, and strawberries.']
```

Word Tokenization

```
from nltk.tokenize import word_tokenize
```

```
words = word_tokenize (text)
```

```
words
```

```

['Martha',
 'is',
 'at',
 'the',
 'grocery',
 'store',
 ',',
 'getting',
 'ready',
 'for',
 'a',
 'house',
 'party',
 '.',
 'She',
 'has',
 'a',
 'list',
 'of',
 'what',
 'she',
 'needs',
 'with',
 'her',
 'as',
```

```
'she',
'goes',
'along',
'.',
'The',
'first',
'section',
'she',
'comes',
'has',
'produce',
'.',
'Martha',
'sees',
'apples',
',',
'bananas',
',',
'cherries',
',',
'grapes',
',',
'and',
'strawberries',
'.']
```

```
for w in words:
    print (w)
```

```
Martha
is
at
the
grocery
store
,
getting
ready
for
a
house
party
.
She
has
a
list
of
what
she
needs
with
her
as
she
goes
along
.
The
first
section
she
comes
has
produce
.
Martha
sees
apples
,
bananas
,
cherries
,
grapes
,
and
strawberries
.
```

Levels of Sentences Tokenization using Comprehension

```
sent_tokenize (text)
```

```
['Martha is at the grocery store, getting ready for a house party.',
 'She has a list of what she needs with her as she goes along.',
 'The first section she comes has produce.',
 'Martha sees apples, bananas, cherries, grapes, and strawberries.']
```

```
[word_tokenize (text) for t in sent_tokenize(text)]
```

```
[['Martha',
 'is',
 'at',
 'the',
 'grocery',
 'store',
 ',',
 'getting',
 'ready',
 'for',
 'a',
 'house',
 'party',
 '.',
 'She',
 'has',
 'a',
 'list',
 'of',
 'what',
 'she',
 'needs',
 'with',
 'her',
 'as',
 'she',
 'goes',
 'along',
 '.',
 'The',
 'first',
 'section',
 'she',
 'comes',
 'has',
 'produce',
 '.',
 'Martha',
 'sees',
 'apples',
 ',',
 'bananas',
 ',',
 'cherries',
 ',',
 'grapes',
 ',',
 'and',
 'strawberries',
 '.'],
 ['Martha',
 'is',
 'at',
 'the',
 'grocery',
 'store',
 ',',
 'getting',
```

```
from nltk.tokenize import wordpunct_tokenize
```

```
wordpunct_tokenize (text)
```

Filteration of Text by converting into lower case

```
text.lower()
```

```
'martha is at the grocery store, getting ready for a house party. she has a list
of what she needs with her as she goes along. the first section she comes has pro
duce martha sees apples bananas cherries grapes and strawberries '
```

```
text.upper()
```

```
'MARTHA IS AT THE GROCERY STORE, GETTING READY FOR A HOUSE PARTY. SHE HAS A LIST  
OF WHAT SHE NEEDS WITH HER AS SHE GOES ALONG. THE FIRST SECTION SHE COMES HAS PRO  
DUCE MARTHA SEES APPLES BANANAS CHERRIES GRAPES AND STRAWBERRIES '
```

[Colab paid products](#) - [Cancel contracts here](#)

