

```
In [1]: import pandas as pd

text = ['Sarah lives in a hut in the village.',
        'She has an apple tree in her backyard.',
        'The apples are red in colour.'].

df = pd.DataFrame(text, columns=['Sentence'])

df
```

Out[1]:

	Sentence
0	Sarah lives in a hut in the village.
1	She has an apple tree in her backyard.
2	The apples are red in colour.

In [9]: `!pip install spacy`

Requirement already satisfied: spacy in c:\users\pooja reddy\anaconda3\lib\site-packages (3.8.4)

Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (3.0.12)

Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (1.0.5)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (1.0.12)

Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (2.0.11)

Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (3.0.9)

Requirement already satisfied: thinc<8.4.0,>=8.3.4 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (8.3.4)

Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (1.1.3)

Requirement already satisfied: srsly<3.0.0,>=2.4.3 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (2.5.1)

Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (2.0.10)

Requirement already satisfied: weasel<0.5.0,>=0.1.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (0.4.1)

Requirement already satisfied: typer<1.0.0,>=0.3.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (0.15.1)

Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (4.65.0)

Requirement already satisfied: numpy>=1.19.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (1.24.3)

Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (2.31.0)

Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (2.10.6)

Requirement already satisfied: jinja2 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (3.1.2)

Requirement already satisfied: setuptools in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (68.0.0)

Requirement already satisfied: packaging>=20.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (23.0)

Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from spacy) (3.5.0)

Requirement already satisfied: language-data>=1.2 in c:\users\pooja reddy\anaconda3\lib\site-packages (from langcodes<4.0.0,>=3.2.0->spacy) (1.3.0)

Requirement already satisfied: annotated-types>=0.6.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (0.7.0)

Requirement already satisfied: pydantic-core==2.27.2 in c:\users\pooja reddy\anaconda3\lib\site-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (2.27.2)

Requirement already satisfied: typing-extensions>=4.12.2 in c:\users\pooja reddy\anaconda3\lib\site-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (4.12.2)

Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\pooja reddy\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in c:\users\pooja reddy\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.4)

Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\pooja reddy\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (1.26.16)

Requirement already satisfied: certifi>=2017.4.17 in c:\users\pooja reddy\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2023.

```

7.22)
Requirement already satisfied: blis<1.3.0,>=1.2.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from thinc<8.4.0,>=8.3.4->spacy) (1.2.0)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in c:\users\pooja reddy\anaconda3\lib\site-packages (from thinc<8.4.0,>=8.3.4->spacy) (0.1.5)
Requirement already satisfied: colorama in c:\users\pooja reddy\anaconda3\lib\site-packages (from tqdm<5.0.0,>=4.38.0->spacy) (0.4.6)
Requirement already satisfied: click>=8.0.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from typer<1.0.0,>=0.3.0->spacy) (8.0.4)
Requirement already satisfied: shellingham>=1.3.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from typer<1.0.0,>=0.3.0->spacy) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from typer<1.0.0,>=0.3.0->spacy) (13.9.4)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from weasel<0.5.0,>=0.1.0->spacy) (0.20.0)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in c:\users\pooja reddy\anaconda3\lib\site-packages (from weasel<0.5.0,>=0.1.0->spacy) (5.2.1)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from jinja2->spacy) (2.1.1)
Requirement already satisfied: marisa-trie>=1.1.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from language-data>=1.2->langcodes<4.0.0,>=3.2.0->spacy) (1.2.1)
Requirement already satisfied: markdown-it-py>=2.2.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (2.2.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (2.15.1)
Requirement already satisfied: mdurl~=0.1 in c:\users\pooja reddy\anaconda3\lib\site-packages (from markdown-it-py>=2.2.0->rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (0.1.0)

```

In [13]: `!python -m spacy download en_core_web_sm`

```

Collecting en-core-web-sm==3.8.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.8.0/en_core_web_sm-3.8.0-py3-none-any.whl (https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.8.0/en_core_web_sm-3.8.0-py3-none-any.whl) (12.8 MB)
----- 0.0/12.8 MB ? eta -:-:-
----- 0.0/12.8 MB 1.4 MB/s eta
0:00:10
----- 0.1/12.8 MB 1.1 MB/s eta
0:00:13
----- 0.1/12.8 MB 1.1 MB/s eta
0:00:12
----- 0.2/12.8 MB 1.3 MB/s eta
0:00:10
----- 0.5/12.8 MB 2.2 MB/s eta
0:00:06
----- 0.5/12.8 MB 2.2 MB/s eta
0:00:06
----- 0.5/12.8 MB 2.2 MB/s eta
0:00:06

```

```
In [16]: import spacy
nlp = spacy.load("en_core_web_sm")
token = []
pos = []
for sent in nlp.pipe(df['Sentence']):
    if sent.has_annotation('DEP'):
        token.append([word.text for word in sent])
        pos.append([word.pos_ for word in sent])
```

```
In [17]: token
```

```
Out[17]: [['Sarah', 'lives', 'in', 'a', 'hut', 'in', 'the', 'village', '.'],
          ['She', 'has', 'an', 'apple', 'tree', 'in', 'her', 'backyard', '.'],
          ['The', 'apples', 'are', 'red', 'in', 'colour', '.']]
```

```
In [18]: pos
```

```
Out[18]: [['PROPN', 'VERB', 'ADP', 'DET', 'NOUN', 'ADP', 'DET', 'NOUN', 'PUNCT'],
          ['PRON', 'VERB', 'DET', 'NOUN', 'NOUN', 'ADP', 'PRON', 'NOUN', 'PUNCT'],
          ['DET', 'NOUN', 'AUX', 'ADJ', 'ADP', 'NOUN', 'PUNCT']]
```

```
In [20]: df['token'] = token
df['pos'] = pos
```

```
In [21]: df.head()
```

```
Out[21]:
```

	Sentence	token	pos
0	Sarah lives in a hut in the village.	[Sarah, lives, in, a, hut, in, the, village, .]	[PROPN, VERB, ADP, DET, NOUN, ADP, DET, NOUN, ...]
1	She has an apple tree in her backyard.	[She, has, an, apple, tree, in, her, backyard, .]	[PRON, VERB, DET, NOUN, NOUN, ADP, PRON, NOUN, ...]
2	The apples are red in colour.	[The, apples, are, red, in, colour, .]	[DET, NOUN, AUX, ADJ, ADP, NOUN, PUNCT]

```
In [22]: text = 'I love NLP and I will learn NLP in 2 months'
```

```
In [23]: import nltk
nltk.download('averaged_perceptron_tagger_eng')
```

```
[nltk_data] Downloading package averaged_perceptron_tagger_eng to
[nltk_data] C:\Users\Pooja Reddy\AppData\Roaming\nltk_data...
[nltk_data] Unzipping taggers\averaged_perceptron_tagger_eng.zip.
```

```
Out[23]: True
```

```
In [26]: import nltk  
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to C:\Users\Pooja  
[nltk_data]   Reddy\AppData\Roaming\nltk_data...  
[nltk_data]   Unzipping corpora\stopwords.zip.
```

```
Out[26]: True
```

```
In [28]: import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize, sent_tokenize
stop_words = set(stopwords.words('english'))
tokens = sent_tokenize(text)
for i in tokens:
    words = nltk.word_tokenize(i)
    words = [W for W in words if not W in stop_words]
    tags = nltk.pos_tag(words)
tags
```

```

-----
-
LookupError                                Traceback (most recent call las
t)
Cell In[28], line 5
      3 from nltk.tokenize import word_tokenize, sent_tokenize
      4 stop_words = set(stopwords.words('english'))
----> 5 tokens = sent_tokenize(text)
      6 for i in tokens:
      7     words = nltk.word_tokenize(i)

File ~\anaconda3\Lib\site-packages\nltk\tokenize\__init__.py:106, in sent_
tokenize(text, language)
    96 def sent_tokenize(text, language="english"):
    97     """
    98     Return a sentence-tokenized copy of *text*,
    99     using NLTK's recommended sentence tokenizer
    (...)
   104     :param language: the model name in the Punkt corpus
   105     """
--> 106     tokenizer = load(f"tokenizers/punkt/{language}.pickle")
   107     return tokenizer.tokenize(text)

File ~\anaconda3\Lib\site-packages\nltk\data.py:750, in load(resource_url,
format, cache, verbose, logic_parser, fstruct_reader, encoding)
    747     print(f"<<Loading {resource_url}>>")
    749 # Load the resource.
--> 750 opened_resource = _open(resource_url)
    752 if format == "raw":
    753     resource_val = opened_resource.read()

File ~\anaconda3\Lib\site-packages\nltk\data.py:876, in _open(resource_ur
l)
    873 protocol, path_ = split_resource_url(resource_url)
    875 if protocol is None or protocol.lower() == "nltk":
--> 876     return find(path_, path + []).open()
    877 elif protocol.lower() == "file":
    878     # urllib might not use mode='rb', so handle this one ourselves
s:
    879     return find(path_, []).open()

File ~\anaconda3\Lib\site-packages\nltk\data.py:583, in find(resource_nam
e, paths)
    581 sep = "*" * 70
    582 resource_not_found = f"\n{sep}\n{msg}\n{sep}\n"
--> 583 raise LookupError(resource_not_found)

```

LookupError:

Resource punkt not found.

Please use the NLTK Downloader to obtain the resource:

```

>>> import nltk
>>> nltk.download('punkt')

```

For more information see: <https://www.nltk.org/data.html> (<https://www.nltk.org/data.html>)

Attempted to load tokenizers/punkt/english.pickle

Searched in:


```
- 'C:\\Users\\Pooja Reddy\\nltk_data'  
- 'C:\\Users\\Pooja Reddy\\anaconda3\\nltk_data'  
- 'C:\\Users\\Pooja Reddy\\anaconda3\\share\\nltk_data'  
- 'C:\\Users\\Pooja Reddy\\anaconda3\\lib\\nltk_data'  
- 'C:\\Users\\Pooja Reddy\\AppData\\Roaming\\nltk_data'  
- 'C:\\nltk_data'  
- 'D:\\nltk_data'  
- 'E:\\nltk_data'  
- ''
```

In []: