

```
!pip install nltk spacy
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.12/dist-packages (3.9.1)
Requirement already satisfied: spacy in /usr/local/lib/python3.12/dist-packages (3.8.11)
Requirement already satisfied: click in /usr/local/lib/python3.12/dist-packages (from nltk) (8.3.1)
Requirement already satisfied: joblib in /usr/local/lib/python3.12/dist-packages (from nltk) (1.5.3)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.12/dist-packages (from nltk) (2021.8.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages (from nltk) (4.67.1)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.12/dist-packages (from nltk) (3.0.11)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from nltk) (1.0.0)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.12/dist-packages (from nltk) (0.28.0)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: weasel<0.5.0,>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: typer-slim<1.0.0,>=0.3.0 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.0)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.1.6)
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from spacy) (75.2.0)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (25.0.3)
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic)
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (from pydantic)
Requirement already satisfied: typing-extensions>=4.14.1 in /usr/local/lib/python3.12/dist-packages (from pydantic)
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from pydantic)
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests)
Requirement already satisfied: blis<1.4.0,>=1.3.0 in /usr/local/lib/python3.12/dist-packages (from thinc<8.4.0,>=8.3.4)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.12/dist-packages (from thinc)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2->spacy)
Requirement already satisfied: wrapt in /usr/local/lib/python3.12/dist-packages (from smart-open<8.0.0,>=5.2.1)
```

▼ Step 1: Import NLTK and spaCy

First, you need to import the libraries you just installed so you can use their functions in your notebook.

```
import nltk
import spacy
```

▼ Step 2: Download NLTK data (if needed)

NLTK often requires downloading specific data packages for its functionalities. The `punkt` tokenizer is commonly used for sentence splitting.

```
nltk.download('punkt')

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

▼ Step 3: Load a short paragraph of text

Now, let's load a short paragraph into a variable. You can replace the example text with any text of your choice.

```
text = """Artificial Intelligence is transforming the world.  
It helps in healthcare, education, and many other fields."""  
print(text)
```

```
Artificial Intelligence is transforming the world.  
It helps in healthcare, education, and many other fields.
```

```
lower_text = text.lower()  
print(lower_text)
```

```
artificial intelligence is transforming the world.  
it helps in healthcare, education, and many other fields.
```

```
words = lower_text.split()  
word_count = len(words)  
  
print("Number of words:", word_count)
```

```
Number of words: 15
```

```
print(words)
```

```
['artificial', 'intelligence', 'is', 'transforming', 'the', 'world.', 'it', 'helps', 'in', 'healthcare', ,
```

```
text = """Natural Language Processing is a major area of Artificial Intelligence that focuses on enabling
```

```
lower_text = text.lower()  
words = lower_text.split()  
print("Number of words:", len(words))
```

```
Number of words: 97
```

```
print(words)
```

```
['natural', 'language', 'processing', 'is', 'a', 'major', 'area', 'of', 'artificial', 'intelligence', 'tha
```

```
import pandas as pd  
file_path = "/content/nlp_text_dataset.csv"  
df = pd.read_csv(file_path)  
  
df.head()
```

	id	text	label
0	1	The weather is sunny and perfect for a picnic.	weather
1	2	Government announced a new policy for renewabl...	news
2	3	I love learning Natural Language Processing.	education
3	4	The movie was boring and too long.	reviews
4	5	AI models are transforming healthcare systems.	technology

Next steps: [Generate code with df](#) [New interactive sheet](#)

```
df.columns
```

```
Index(['id', 'text', 'label'], dtype='object')
```

```
text = " ".join(df['text'].astype(str))
```

```
lower_text = text.lower()
words = lower_text.split()
word_count = len(words)

print("Number of words:", word_count)
```

```
Number of words: 313
```