

```
!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) (2025.11.3)
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 spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.0.15)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.13)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.12/dist-packages (from spacy) (8.3.10)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.1.3)
Requirement already satisfied: srslly<3.0.0,>=2.4.3 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.5.2)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (0.4.3)
Requirement already satisfied: typer-slim<1.0.0,>=0.3.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (0.20.0)
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.32.4)
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) (2.1.2)
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)
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4)
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4)
Requirement already satisfied: typing-extensions>=4.14.1 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4)
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4)
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.1.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.1.2)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.1.2)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.1.2)
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->spacy) (3.1.2)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.12/dist-packages (from thinc<8.4.0,>=8.3.4->spacy) (3.1.2)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-packages (from weasel<0.5.0,>=0.4.2)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/python3.12/dist-packages (from weasel<0.5.0,>=0.4.2)
Requirement already satisfied: MarkupSafe=>2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2->spacy) (3.0.3)
Requirement already satisfied: wrapt in /usr/local/lib/python3.12/dist-packages (from smart-open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.4.2)
```

writing a sample paragraph

```
sample_paragraph = """Natural Language Processing (NLP) is a field of artificial intelligence that focuses on enabling computers to understand, process, and generate human language. It combines computational linguistics—rule-based models and machine learning approaches to analyze, interpret, and generate human language. NLP has numerous applications, such as speech recognition, text summarization, sentiment analysis, and machine translation. One of the key challenges in NLP is dealing with the vast amount of unstructured text data and the need for domain-specific knowledge. Another challenge is handling multiple languages and dialects, which requires large amounts of annotated data for training and testing. Despite these challenges, NLP has made significant progress in recent years, thanks to advances in deep learning and large-scale datasets. As a result, NLP is now an essential tool for many industries, including finance, healthcare, and e-commerce. In this lab, we will explore some basic concepts of NLP and learn how to build a simple NLP pipeline using Python and its libraries like NLTK and Spacy. We will also see how NLP can be used to extract meaningful insights from text data and how it can help us solve real-world problems. So let's get started!"""
```

```
print(sample_paragraph)
```

```
Natural Language Processing (NLP) is a field of artificial intelligence that focuses on enabling computers to understand, process, and generate human language. It combines computational linguistics—rule-based models and machine learning approaches to analyze, interpret, and generate human language. NLP has numerous applications, such as speech recognition, text summarization, sentiment analysis, and machine translation. One of the key challenges in NLP is dealing with the vast amount of unstructured text data and the need for domain-specific knowledge. Another challenge is handling multiple languages and dialects, which requires large amounts of annotated data for training and testing. Despite these challenges, NLP has made significant progress in recent years, thanks to advances in deep learning and large-scale datasets. As a result, NLP is now an essential tool for many industries, including finance, healthcare, and e-commerce. In this lab, we will explore some basic concepts of NLP and learn how to build a simple NLP pipeline using Python and its libraries like NLTK and Spacy. We will also see how NLP can be used to extract meaningful insights from text data and how it can help us solve real-world problems. So let's get started!
```

Count the number of words

```
words = sample_paragraph.split()
word_count = len(words)
print(f"The paragraph contains {word_count} words.")
```

```
The paragraph contains 50 words.
```

Convert the entire text to lowercase

```
lowercase_paragraph = sample_paragraph.lower()
print(lowercase_paragraph)
```

```
on enabling computers to understand, process, and generate human language. It combines computational linguistics—rule-based models and machine learning approaches to analyze, interpret, and generate human language. NLP has numerous applications, such as speech recognition, text summarization, sentiment analysis, and machine translation. One of the key challenges in NLP is dealing with the vast amount of unstructured text data and the need for domain-specific knowledge. Another challenge is handling multiple languages and dialects, which requires large amounts of annotated data for training and testing. Despite these challenges, NLP has made significant progress in recent years, thanks to advances in deep learning and large-scale datasets. As a result, NLP is now an essential tool for many industries, including finance, healthcare, and e-commerce. In this lab, we will explore some basic concepts of NLP and learn how to build a simple NLP pipeline using Python and its libraries like NLTK and Spacy. We will also see how NLP can be used to extract meaningful insights from text data and how it can help us solve real-world problems. So let's get started!
```

```
import pandas as pd
```

```
print(len(df))
```

```
50
```

```
print(df)
```

	<b>id</b>		<b>text</b>	<b>label</b>
0	1	The weather is sunny and perfect for a picnic.	weather	
1	2	Government announced a new policy for renewable...	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	
5	6	Heavy rains caused traffic jams across the city.	weather	
6	7	The new mall opened with many international br...	news	
7	8	Students participated in a coding hackathon.	education	
8	9	The restaurant had delicious food but slow ser...	reviews	
9	10	Researchers introduced a new deep learning arc...	technology	
10	11	Cloudy skies are expected tomorrow.	weather	
11	12	Parliament discussed the national budget today.	news	
12	13	Teachers are introducing online learning tools.	education	
13	14	The hotel room was clean and comfortable.	reviews	
14	15	Robotics is changing manufacturing industries.	technology	
15	16	Strong winds damaged several trees.	weather	
16	17	Local elections are scheduled next month.	news	
17	18	Students submitted assignments through the por...	education	
18	19	The mobile app had too many bugs.	reviews	
19	20	Blockchain applications are increasing rapidly.	technology	
20	21	It rained heavily last night.	weather	
21	22	The city hosted a cultural festival.	news	
22	23	The professor explained transformers clearly.	education	
23	24	The laptop battery drains very quickly.	reviews	
24	25	Self-driving cars are being tested worldwide.	technology	
25	26	The temperature dropped significantly this week.	weather	
26	27	A new hospital was inaugurated yesterday.	news	
27	28	Students prepared presentations on AI ethics.	education	
28	29	The headphones broke after one week.	reviews	
29	30	Cybersecurity attacks are becoming common.	technology	
30	31	Morning fog reduced visibility on highways.	weather	
31	32	Scientists discovered a new species of bird.	news	
32	33	The online course helped me understand NLP bas...	education	
33	34	The service at the bank was very slow.	reviews	
34	35	Smartwatches can monitor heart rate and steps.	technology	
35	36	The heat wave affected many regions.	weather	
36	37	International summit focused on climate change.	news	
37	38	Students learned tokenization and stemming.	education	
38	39	The phone camera quality is excellent.	reviews	
39	40	Quantum computing may solve complex problems.	technology	
40	41	Snowfall covered the mountains overnight.	weather	
41	42	The mayor launched a city-cleaning mission.	news	
42	43	The workshop explained text classification.	education	
43	44	The smartwatch screen cracked easily.	reviews	
44	45	Startups are building AI-powered assistants.	technology	
45	46	Humidity levels were high throughout the day.	weather	
46	47	A new airport terminal became operational.	news	
47	48	Students practiced summarizing long documents.	education	
48	49	The coffee shop had great ambiance but was cos...	reviews	
49	50	Edge computing reduces latency in smart devices.	technology	

## ◆ Gemini

```
df['text'] = df['text'].str.lower()
display(df.head())
```

	<b>id</b>		<b>text</b>	<b>label</b>	
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		

