Start coding or generate with AI.

⌄   1. Install and Import Essential NLP Libraries

```
# Install NLTK and spaCy
!pip install nltk spacy
!python -m spacy download en_core_web_sm
```

Requirement already satisfied: nltk in /usr/local/lib/python3.12/dist-package
Requirement already satisfied: spacy in /usr/local/lib/python3.12/dist-packag
Requirement already satisfied: click in /usr/local/lib/python3.12/dist-packag
Requirement already satisfied: joblib in /usr/local/lib/python3.12/dist-packa
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.12/d
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-package
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/py
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/pyth
Requirement already satisfied: weasel<0.5.0,>=0.4.2 in /usr/local/lib/python3
Requirement already satisfied: typer-slim<1.0.0,>=0.3.0 in /usr/local/lib/pyt
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.12/dis
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/pyth
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/l
Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packa
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-p
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/d
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/pytho
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python
Requirement already satisfied: typing-extensions>=4.14.1 in /usr/local/lib/py
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/pyt
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/pyt
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.1
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.1
Requirement already satisfied: blis<1.4.0,>=1.3.0 in /usr/local/lib/python3.1
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/pyt
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/p
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/pyt
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/d
Requirement already satisfied: wrap in /usr/local/lib/python3.12/dist-packag
Collecting en-core-web-sm==3.8.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_
  ━━ 12.8/12.8 MB 52.0 MB/s eta 0:00
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
⚠ Restart to reload dependencies
If you are in a Jupyter or Colab notebook, you may need to restart Python in
order to load all the package's dependencies. You can do this by selecting th
'Restart kernel' or 'Restart runtime' option.

```
# Import NLTK and spaCy
import nltk
import spacy

# Download necessary NLTK data (e.g., for tokenization if needed later)
nltk.download('punkt')
nltk.download('punkt_tab') # Added to fix the LookupError for word tokenizat

# Load the spaCy English model
nlp = spacy.load('en_core_web_sm')

print("NLTK and spaCy loaded successfully!")
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]    Package punkt is already up-to-date!
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]    Unzipping tokenizers/punkt_tab.zip.
NLTK and spaCy loaded successfully!
```

2. Load a Short Paragraph

You can replace the sample text below with your own paragraph (e.g., a news snippet or a poem).

```
sample_text = "Google Colaboratory, or 'Colab' for short, allows you to writ

print("Loaded paragraph:\n---\n")
print(sample_text)
print("\n---")
```

```
Loaded paragraph:
---

Google Colaboratory, or 'Colab' for short, allows you to write and execute Py

---
```

3. Perform Simple Preprocessing Tasks

```
# Count the number of words in the text
words = nltk.word_tokenize(sample_text)
word_count = len(words)
print(f"Number of words in the text: {word_count}")
```

```
Number of words in the text: 88
```

```python
# Convert the entire text to lowercase
lowercase_text = sample_text.lower()
print("Lowercase text:\n---\n")
print(lowercase_text)
print("\n---")
```

```
Lowercase text:
---

google colaboratory, or 'colab' for short, allows you to write and execute py

---
```