```
# Install dependencies
!pip install textblob
!pip install wordcloud
```

```
Requirement already satisfied: textblob in /usr/local/lib/python3.11/dist-packages (0.19.0)
Requirement already satisfied: nltk>=3.9 in /usr/local/lib/python3.11/dist-packages (from textblob) (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk>=3.9->textblob) (8.2.1)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk>=3.9->textblob) (1.5.1)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk>=3.9->textblob) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk>=3.9->textblob) (4.67.1)
Requirement already satisfied: wordcloud in /usr/local/lib/python3.11/dist-packages (1.9.4)
Requirement already satisfied: numpy>=1.6.1 in /usr/local/lib/python3.11/dist-packages (from wordcloud) (2.0.2)
Requirement already satisfied: pillow in /usr/local/lib/python3.11/dist-packages (from wordcloud) (11.3.0)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (from wordcloud) (3.10.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (1.3.2)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (4.59.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (1.4.8)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (25.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (2.9.0.post0
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1
```

```
import pandas as pd
import matplotlib.pyplot as plt
from textblob import TextBlob
from wordcloud import WordCloud
import seaborn as sns
```

```
from google.colab import files
uploaded = files.upload()
```

Choose Files twitter_training.csv.zip

- **twitter_training.csv.zip**(application/x-zip-compressed) - 2017400 bytes, last modified: 2/8/2025 - 100% done

```
Saving twitter_training.csv.zip to twitter_training.csv.zip
```

```
import zipfile
import pandas as pd

# Unzip the file
with zipfile.ZipFile("twitter_training.csv.zip", 'r') as zip_ref:
    zip_ref.extractall(".")

# Load the CSV
df = pd.read_csv("twitter_training.csv", header=None)
df.head()
```

	0	1	2	3
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
# Rename columns
df.columns = ['ID', 'Entity', 'Sentiment', 'Tweet']

# Check basic info
print(df.info())
df.head(10)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 74682 entries, 0 to 74681
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   ID         74682 non-null  int64 
 1   Entity     74682 non-null  object
 2   Sentiment  74682 non-null  object
 3   Tweet      73996 non-null  object
dtypes: int64(1), object(3)
memory usage: 2.3+ MB
None
```

	ID	Entity	Sentiment	Tweet
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...
5	2401	Borderlands	Positive	im getting into borderlands and i can murder y...
6	2402	Borderlands	Positive	So I spent a few hours making something for fu...
7	2402	Borderlands	Positive	So I spent a couple of hours doing something f...
8	2402	Borderlands	Positive	So I spent a few hours doing something for fun...
9	2402	Borderlands	Positive	So I spent a few hours making something for fu...

Next steps: Generate code with df | View recommended plots | New interactive sheet

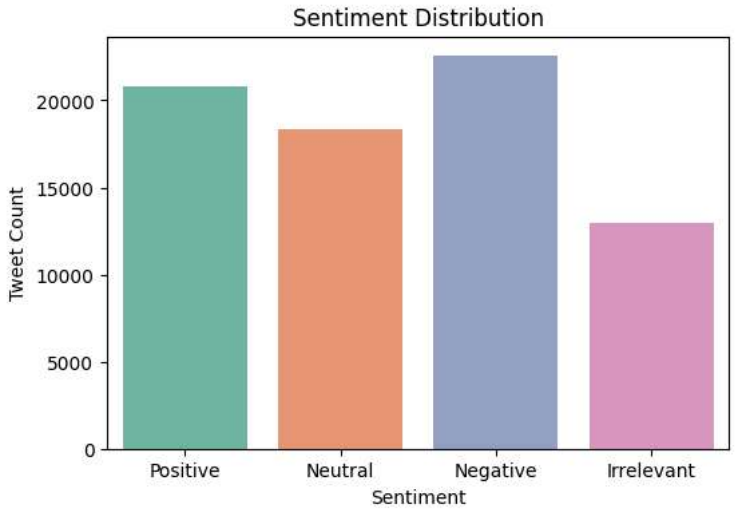
```
import re

def clean_text(text):
    text = str(text).lower()                              # lowercase
    text = re.sub(r"http\S+|www\S+|https\S+", '', text)  # remove URLs
    text = re.sub(r'\@\w+|\#','', text)                  # remove mentions and hashtags
    text = re.sub(r'[^\w\s]', '', text)                  # remove punctuation
    text = re.sub(r'\d+', '', text)                      # remove numbers
    text = re.sub(r'\s+', ' ', text).strip()             # remove extra spaces
    return text

# Apply cleaning
df['Clean_Tweet'] = df['Tweet'].apply(clean_text)

# View cleaned data
df[['Tweet', 'Clean_Tweet']].head(10)
```

	Tweet	Clean_Tweet
0	im getting on borderlands and i will murder yo...	im getting on borderlands and i will murder yo...
1	I am coming to the borders and I will kill you...	i am coming to the borders and i will kill you...
2	im getting on borderlands and i will kill you ...	im getting on borderlands and i will kill you all
3	im coming on borderlands and i will murder you...	im coming on borderlands and i will murder you...
4	im getting on borderlands 2 and i will murder ...	im getting on borderlands and i will murder yo...
5	im getting into borderlands and i can murder y...	im getting into borderlands and i can murder y...
6	So I spent a few hours making something for fu...	so i spent a few hours making something for fu...
7	So I spent a couple of hours doing something f...	so i spent a couple of hours doing something f...
8	So I spent a few hours doing something for fun...	so i spent a few hours doing something for fun...
9	So I spent a few hours making something for fu...	so i spent a few hours making something for fu...

```
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(6,4))
sns.countplot(x='Sentiment', data=df, palette='Set2')
```

```
plt.title("Sentiment Distribution")
plt.xlabel("Sentiment")
plt.ylabel("Tweet Count")
plt.show()
```

```
/tmp/ipython-input-2183413826.py:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend

  sns.countplot(x='Sentiment', data=df, palette='Set2')
```




```
from wordcloud import WordCloud

# Function to plot word cloud for a given sentiment
def plot_wordcloud(sentiment, color):
    text = " ".join(df[df['Sentiment'] == sentiment]['Clean_Tweet'])
    wordcloud = WordCloud(width=800, height=400, background_color='white',
                          colormap=color, collocations=False).generate(text)
    plt.figure(figsize=(10,5))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis("off")
    plt.title(f"Most Common Words in {sentiment} Tweets", fontsize=16)
    plt.show()

# Generate word clouds
plot_wordcloud("Positive", "Greens")
plot_wordcloud("Negative", "Reds")
plot_wordcloud("Neutral", "Blues")
```

Most Common Words in Positive Tweets

Most Common Words in Negative Tweets

Most Common Words in Neutral Tweets