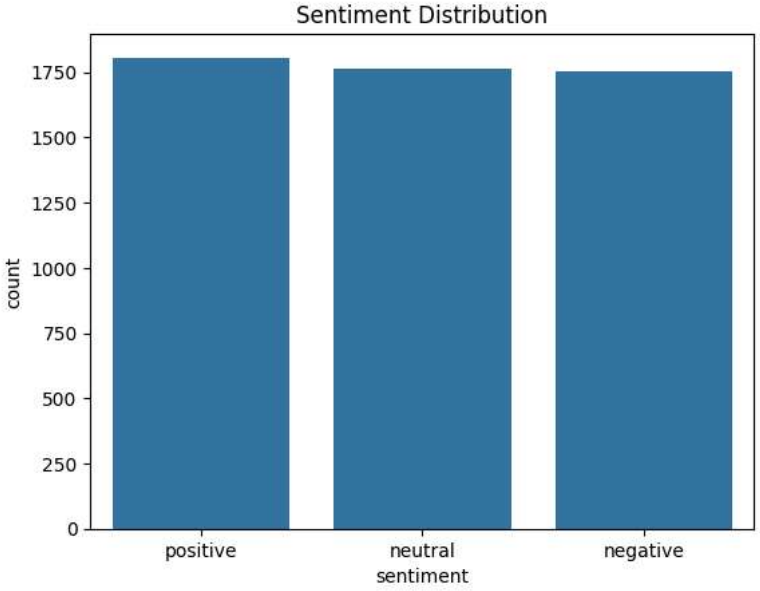
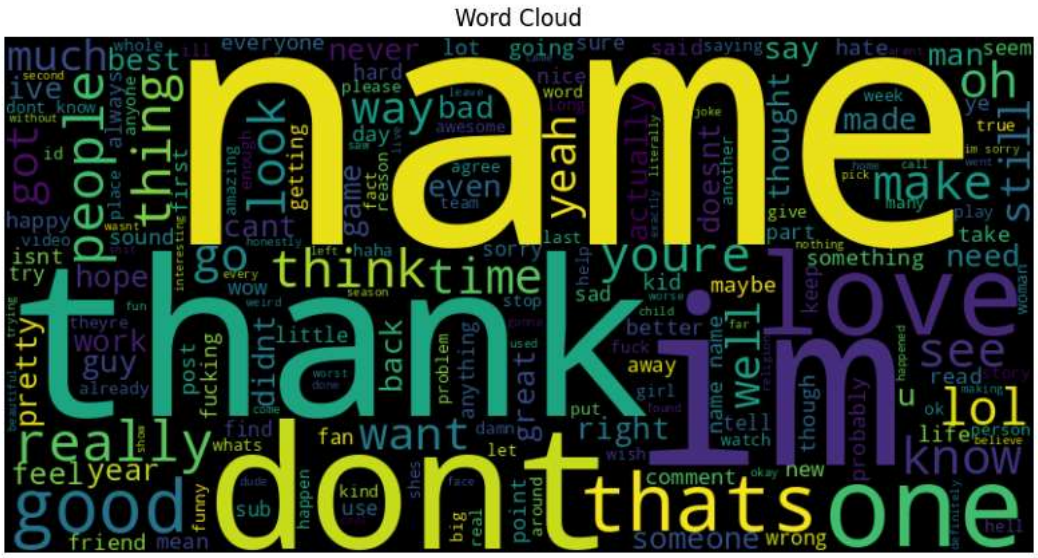
```
# Sentiment Analysis Pipeline for Social Media Data

import pandas as pd
import numpy as np
import re
import string
import seaborn as sns
import matplotlib.pyplot as plt
from wordcloud import WordCloud
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
import nltk

nltk.download('stopwords')
nltk.download('wordnet')

# Load data
file_path = "test.csv"
df = pd.read_csv(file_path, header=None, quoting=1, on_bad_lines='skip')
df.columns = ['text', 'author', 'metadata']

# Drop rows with null text
df = df.dropna(subset=['text'])

# Data Preprocessing
def clean_text(text):
    text = text.lower()
    text = re.sub(r'http\S+|www\S+|https\S+', '', text, flags=re.MULTILINE)  # remove links
    text = re.sub(r'\@\w+|\#','', text)  # remove mentions and hashtags
    text = re.sub(r'[^A-Za-z\s]', '', text)  # remove punctuations
    text = re.sub(r'\s+', ' ', text).strip()  # remove extra whitespace
    return text

df['clean_text'] = df['text'].apply(clean_text)

# Tokenization and Lemmatization
stop_words = set(stopwords.words('english'))
lemmatizer = WordNetLemmatizer()

def preprocess(text):
    tokens = text.split()
    tokens = [lemmatizer.lemmatize(word) for word in tokens if word not in stop_words]
    return " ".join(tokens)

df['processed_text'] = df['clean_text'].apply(preprocess)

# For demonstration: generate synthetic sentiment labels (in real case, these should be provided)
df['sentiment'] = np.random.choice(['positive', 'neutral', 'negative'], size=len(df))

# EDA: Plot label distribution
sns.countplot(data=df, x='sentiment')
plt.title('Sentiment Distribution')
plt.show()

# Word cloud
text_combined = " ".join(df['processed_text'])
wordcloud = WordCloud(width=800, height=400).generate(text_combined)
plt.figure(figsize=(10,5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title('Word Cloud')
plt.show()

# Feature Engineering
vectorizer = TfidfVectorizer(max_features=5000)
X = vectorizer.fit_transform(df['processed_text'])
y = df['sentiment']

# Model Training
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LogisticRegression(max_iter=1000)
model.fit(X_train, y_train)

# Evaluation
y_pred = model.predict(X_test)
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

```
print(df)
df.head()
```

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






```
[[111 118 142]
 [111 127 114]
 [103 140 100]]
              precision    recall  f1-score   support

    negative       0.34      0.30      0.32       371
     neutral       0.33      0.36      0.34       352
    positive       0.28      0.29      0.29       343

    accuracy                           0.32      1066
   macro avg       0.32      0.32      0.32      1066
weighted avg       0.32      0.32      0.32      1066
```

```
                                                   text  \
0     I’m really sorry about your situation :( Altho...
1     It's wonderful because it's awful. At not with...
2                                     Kings fan here
3                                  I didn't know that
4     They got bored from haunting earth for thousan...
...                                                 ...
5323  Thanks. I was diagnosed with BP 1 after the ho...
5324                Well that makes sense.\t4\tef9c7s3
5325                  Daddy issues [NAME]\t27\tefbiugo
5326  So glad I discovered that subreddit a couple m...
5327  Had to watch "Elmo in Grouchland" one time too...

                                                 author  \
0                                               Cirilla
1                                                   NaN
2      good luck to you guys! Will be an interesting...
3      thank you for teaching me something today!\t1...
4                                                   NaN
...                                                 ...
5323                                                NaN
5324                                                NaN
```

```
5325                                                NaN
5326                                                NaN
5327                                                NaN

                        metadata  \
0      and Scarlett!\t25\teecwqtt
1                             NaN
2                             NaN
3                             NaN
4                             NaN
...                           ...
5323                          NaN
5324                          NaN
5325                          NaN
5326                          NaN
5327                          NaN

                                           clean_text  \
0     im really sorry about your situation although ...
1      its wonderful because its awful at not with edfd
2                                     kings fan here
3                                  i didnt know that
4     they got bored from haunting earth for thousan...
...                                                ...
5323  thanks i was diagnosed with bp after the hospi...
5324                          well that makes sense efcs
5325                           daddy issues name efbiugo
5326  so glad i discovered that subreddit a couple m...
5327  had to watch elmo in grouchland one time too m...

                                      processed_text sentiment
0     im really sorry situation although love name s...  positive
1                                wonderful awful edfd   neutral
2                                           king fan  positive
3                                         didnt know  positive
4     got bored haunting earth thousand year ultimat...  positive
...                                                 ...       ...
5323   thanks diagnosed bp hospitalization well efeeasc  positive
5324                           well make sense efcs  negative
5325                         daddy issue name efbiugo  positive
5326  glad discovered subreddit couple month ago goo...  positive
5327  watch elmo grouchland one time many kid little...  positive

[5327 rows x 6 columns]
```

	text	author	metadata	clean_text	processed_text	sentiment
0	I'm really sorry about your situation :(Altho...	Cirilla	and Scarlett!\t25\teecwqtt	im really sorry about your situation although ...	im really sorry situation although love name s...	positive
1	It's wonderful because it's awful. At not with...	NaN	NaN	its wonderful because its awful at not with edfd	wonderful awful edfd	neutral
2	Kings fan here	good luck to you guys! Will be an interesting...	NaN	kings fan here	king fan	positive
3	I didn't know that	thank you for teaching me something today!\t1...	NaN	i didnt know that	didnt know	positive
4	They got bored from haunting earth for thousan...	NaN	NaN	they got bored from haunting earth for thousan...	got bored haunting earth thousand year ultimat...	positive

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

```
!pip install gradio==3.38.0
```

```
Collecting gradio==3.38.0
  Downloading gradio-3.38.0-py3-none-any.whl.metadata (17 kB)
Collecting aiofiles<24.0,>=22.0 (from gradio==3.38.0)
  Downloading aiofiles-23.2.1-py3-none-any.whl.metadata (9.7 kB)
Requirement already satisfied: aiohttp~=3.0 in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (3.11.15)
Requirement already satisfied: altair<6.0,>=4.2.0 in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (5.5.0)
Collecting fastapi (from gradio==3.38.0)
  Downloading fastapi-0.115.12-py3-none-any.whl.metadata (27 kB)
Collecting ffmpy (from gradio==3.38.0)
  Downloading ffmpy-0.5.0-py3-none-any.whl.metadata (3.0 kB)
Collecting gradio-client>=0.2.10 (from gradio==3.38.0)
  Downloading gradio_client-1.10.0-py3-none-any.whl.metadata (7.1 kB)
Requirement already satisfied: httpx in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (0.28.1)
Requirement already satisfied: huggingface-hub>=0.14.0 in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (0.31.1)
Requirement already satisfied: jinja2<4.0 in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (3.1.6)
Requirement already satisfied: markdown-it-py>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from markdown-it-py[linkify]>=2.0.(
Collecting markupsafe~=2.0 (from gradio==3.38.0)
  Downloading MarkupSafe-2.1.5-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (3.0 kB)
Requirement already satisfied: matplotlib~=3.0 in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (3.10.0)
Collecting mdit-py-plugins<=0.3.3 (from gradio==3.38.0)
  Downloading mdit_py_plugins-0.3.3-py3-none-any.whl.metadata (2.8 kB)
Collecting numpy~=1.0 (from gradio==3.38.0)
  Downloading numpy-1.26.4-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (61 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 61.0/61.0 kB 5.0 MB/s eta 0:00:00
Requirement already satisfied: orjson~=3.0 in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (3.10.18)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (24.2)
Requirement already satisfied: pandas<3.0,>=1.0 in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (2.2.2)
Collecting pillow<11.0,>=8.0 (from gradio==3.38.0)
  Downloading pillow-10.4.0-cp311-cp311-manylinux_2_28_x86_64.whl.metadata (9.2 kB)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,!=2.0.0,!=2.0.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.11/dist-packages (fron
Collecting pydub (from gradio==3.38.0)
  Downloading pydub-0.25.1-py2.py3-none-any.whl.metadata (1.4 kB)
Collecting python-multipart (from gradio==3.38.0)
  Downloading python_multipart-0.0.20-py3-none-any.whl.metadata (1.8 kB)
Requirement already satisfied: pyyaml<7.0,>=5.0 in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (6.0.2)
Requirement already satisfied: requests~=2.0 in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (2.32.3)
Collecting semantic-version~=2.0 (from gradio==3.38.0)
  Downloading semantic_version-2.10.0-py2.py3-none-any.whl.metadata (9.7 kB)
Requirement already satisfied: typing-extensions~=4.0 in /usr/local/lib/python3.11/dist-packages (from gradio==3.38.0) (4.13.2)
Collecting uvicorn>=0.14.0 (from gradio==3.38.0)
  Downloading uvicorn-0.34.2-py3-none-any.whl.metadata (6.5 kB)
Collecting websockets<12.0,>=10.0 (from gradio==3.38.0)
  Downloading websockets-11.0.3-cp311-cp311-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_64.whl.me
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp~=3.0->gradio==3.38.(
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from aiohttp~=3.0->gradio==3.38.0) (1.3
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp~=3.0->gradio==3.38.0) (25.3.0
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp~=3.0->gradio==3.38.0) (1.(
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp~=3.0->gradio==3.38.0) ((
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp~=3.0->gradio==3.38.0) (0.3
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp~=3.0->gradio==3.38.0) (1.2
Requirement already satisfied: jsonschema>=3.0 in /usr/local/lib/python3.11/dist-packages (from altair<6.0,>=4.2.0->gradio==3.38.0)
Requirement already satisfied: narwhals>=1.14.2 in /usr/local/lib/python3.11/dist-packages (from altair<6.0,>=4.2.0->gradio==3.38.0
Requirement already satisfied: fsspec in /usr/local/lib/python3.11/dist-packages (from gradio-client>=0.2.10->gradio==3.38.0) (2025
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages (from httpx->gradio==3.38.0) (4.9.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-packages (from httpx->gradio==3.38.0) (2025.4.26)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-packages (from httpx->gradio==3.38.0) (1.0.9)
Requirement already satisfied: idna in /usr/local/lib/python3.11/dist-packages (from httpx->gradio==3.38.0) (3.10)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.*->httpx->gradio==3.38.0) (0.1
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.14.0->gradio==3.38.0) (3
Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.14.0->gradio==3.38.(
Requirement already satisfied: hf-xet<2.0.0,>=1.1.0 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.14.0->gradio
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.11/dist-packages (from markdown-it-py>=2.0.0->markdown-it-py[lin
Requirement already satisfied: linkify-it-py<3,>=1 in /usr/local/lib/python3.11/dist-packages (from markdown-it-py[linkify]>=2.0.0-
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib~=3.0->gradio==3.38.0) (1
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib~=3.0->gradio==3.38.0) (0.12
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib~=3.0->gradio==3.38.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib~=3.0->gradio==3.38.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib~=3.0->gradio==3.38.0) (3
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dist-packages (from matplotlib~=3.0->gradio==3.38.(
```