```
pip install praw pandas
```

```
Collecting praw
    Downloading praw-7.7.1-py3-none-any.whl (191 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 191.0/191.0 kB 4.3 MB/s eta 0:00:00
    Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (2.0.3)
    Collecting prawcore<3,>=2.1 (from praw)
    Downloading prawcore-2.4.0-py3-none-any.whl (17 kB)
    Collecting update-checker>=0.18 (from praw)
    Downloading update_checker-0.18.0-py3-none-any.whl (7.0 kB)
    Requirement already satisfied: websocket-client>=0.54.0 in /usr/local/lib/python3.10/dist-packages (from praw) (1.8.0)
    Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
    Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2023.4)
    Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2024.1)
    Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/dist-packages (from pandas) (1.25.2)
    Requirement already satisfied: requests<3.0,>=2.6.0 in /usr/local/lib/python3.10/dist-packages (from prawcore<3,>=2.1->praw) (2.31.0)
    Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
    Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3.0,>=2.6.0->prawcore
    Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0,>=2.6.0->prawcore<3,>=2.1->pr
    Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0,>=2.6.0->prawcore<3,>=2
    Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0,>=2.6.0->prawcore<3,>=2
    Installing collected packages: update-checker, prawcore, praw
    Successfully installed praw-7.7.1 prawcore-2.4.0 update-checker-0.18.0
```

```python
from configparser import SafeConfigParser
import praw
import pandas as pd
import warnings
import re
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from transformers import AutoTokenizer, AutoModelWithLMHead
```

Develop a script to scrape Reddit posts and comments from specific subreddits related to health conditions relevant to the clinical trial.

```python
warnings.filterwarnings("ignore", category=DeprecationWarning)

# Initialize the Reddit instance
reddit = praw.Reddit(
    client_id="***********",
    client_secret="***********",
    password="********",
    user_agent="Comment Extraction (by u/USERNAME)",
    username="Latter_Expert6546",
)
# List of subreddits to search in
subreddits = ['BodyDysmorphia', 'Fibromyalgia', 'rarediseases', 'Celiac', 'clinicaltrials', 'clinicalresearch', 'FindClinTrial', 'health', 'A

# Scrape data
keyword = 'clinical trial'
reddit_text = []
score = []

for subreddit in subreddits:
    for submission in reddit.subreddit(subreddit).search(keyword, limit=None):
        if keyword.lower() in submission.title.lower() or keyword.lower() in submission.selftext.lower():
            reddit_text.append(submission.selftext)
            score.append(submission.score)

            submission.comments.replace_more(limit=None)
            for comment in submission.comments.list():
                reddit_text.append(comment.body)
                score.append(comment.score)

data = {
    'Text': reddit_text,
    'Score': score
}

# Creating a DataFrame from the dictionary
df = pd.DataFrame(data)
# print(df)
# print(len(df))
```

```
# print(len(ur))
```

⮕ **Show hidden output**

Perform sentiment analysis on the scraped data to determine the general attitude and interest levels regarding clinical trials.

```
pip install nltk
```

⮕
```
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2024.5.15)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.4)
```

```
df = pd.read_csv('/content/drive/MyDrive/df.csv')
print(df)
```

⮕
```
                                                   Text  Score  \
0      **Hi everyone, I thought this might be of inte...      4
1          \n\nHi everyone!\n\nI work at [Massachusetts...   8
2      I'm 32, and I've struggled with an eating diso...     17
3      I am so sorry to hear you are going through th...     13
4      He is not a good person. He sounds like a terr...     10
...                                                  ...    ...
6551   Thank you so much for your response!\n\nYou ha...      1
6552   All that certainly fits with NAFLD, which usua...      2
6553                     Any updates on your AIH diagnosis?   1
6554   Not especially. \n\nHave some blood work to do...      1
6555                                            Post em      1

                                             cleaned_text Emotion detected
0      hi everyone thought might interest living melb...              joy
1      hi everyone work massachusetts general hospita...              joy
2      struggled eating disorder almost three quarter...          sadness
3      sorry hear going one husband cheating bad pers...              joy
4      good person sound like terrible person wife ch...          sadness
...                                                  ...              ...
6551   thank much response eased mind much regarding ...            anger
6552   certainly fit nafld usually get way better wei...              joy
6553                                     update aih diagnosis           fear
6554   especially blood work mr ct abdomen shortly li...              joy
6555                                              post em           fear

[6556 rows x 4 columns]
```

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

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

```
def clean_comment(comment):
    if not isinstance(comment, str):
        return ''

    # Remove URLs
    comment = re.sub(r'http\S+|www\S+|https\S+', '', comment, flags=re.MULTILINE)

    # Remove user mentions
    comment = re.sub(r'@\w+', '', comment)

    # Remove special characters, numbers, and punctuation
    comment = re.sub(r'\W', ' ', comment)
    comment = re.sub(r'\d', '', comment)

    # Convert to lowercase
    comment = comment.lower()

    # Tokenize
    tokens = word_tokenize(comment)

    # Remove stopwords
```

```python
    stop_words = set(stopwords.words('english'))
    tokens = [word for word in tokens if word not in stop_words]

    # Lemmatization
    lemmatizer = WordNetLemmatizer()
    tokens = [lemmatizer.lemmatize(token) for token in tokens]

    # Remove extra whitespace
    cleaned_comment = ' '.join(tokens)

    return cleaned_comment

df['cleaned_text'] = df['Text'].apply(clean_comment)
print(df)
```

```
                                               Text   Score  \
0        **Hi everyone, I thought this might be of inte...      4
1         \n\nHi everyone!\n\nI work at [Massachusetts...      8
2        I'm 32, and I've struggled with an eating diso...     17
3        I am so sorry to hear you are going through th...     13
4        He is not a good person. He sounds like a terr...     10
...                                                 ...    ...
6551   Thank you so much for your response!\n\nYou ha...      1
6552   All that certainly fits with NAFLD, which usua...      2
6553                 Any updates on your AIH diagnosis?      1
6554   Not especially. \n\nHave some blood work to do...      1
6555                                          Post em       1

                                     cleaned_text Emotion detected
0        hi everyone thought might interest living melb...            joy
1        hi everyone work massachusetts general hospita...            joy
2        struggled eating disorder almost three quarter...        sadness
3        sorry hear going one husband cheating bad pers...            joy
4        good person sound like terrible person wife ch...        sadness
...                                             ...             ...
6551   thank much response eased mind much regarding ...          anger
6552   certainly fit nafld usually get way better wei...            joy
6553                             update aih diagnosis           fear
6554   especially blood work mr ct abdomen shortly li...            joy
6555                                      post em           fear

[6556 rows x 4 columns]
```

```python
# Emotions classified: anger, fear, joy, love, sadness, surprise

tokenizer = AutoTokenizer.from_pretrained("mrm8488/t5-base-finetuned-emotion")

model = AutoModelWithLMHead.from_pretrained("mrm8488/t5-base-finetuned-emotion")

def get_emotion(text):
  input_ids = tokenizer.encode(text + '</s>', return_tensors='pt')

  output = model.generate(input_ids=input_ids,
               max_length=2)

  dec = [tokenizer.decode(ids) for ids in output]
  label = dec[0]
  clean_label = label.replace('<pad>', '').strip()
  return clean_label

df['Emotion detected'] = df['cleaned_text'].apply(get_emotion)
print(df)
```

```
/usr/local/lib/python3.10/dist-packages/transformers/models/auto/modeling_auto.py:1712: FutureWarning: The class `AutoModelWithLMHead`
  warnings.warn(
Token indices sequence length is longer than the specified maximum sequence length for this model (786 > 512). Running this sequence th
                                               Text   Score  \
0        **Hi everyone, I thought this might be of inte...      4
1         \n\nHi everyone!\n\nI work at [Massachusetts...      8
2        I'm 32, and I've struggled with an eating diso...     17
3        I am so sorry to hear you are going through th...     13
4        He is not a good person. He sounds like a terr...     10
...                                                 ...    ...
6551   Thank you so much for your response!\n\nYou ha...      1
6552   All that certainly fits with NAFLD, which usua...      2
6553                 Any updates on your AIH diagnosis?      1
6554   Not especially. \n\nHave some blood work to do...      1
6555                                          Post em       1
```

```
                       cleaned_text Emotion detected
0      hi everyone thought might interest living melb...          joy
1      hi everyone work massachusetts general hospita...          joy
2      struggled eating disorder almost three quarter...      sadness
3      sorry hear going one husband cheating bad pers...          joy
4      good person sound like terrible person wife ch...      sadness
...                              ...          ...
6551   thank much response eased mind much regarding ...        anger
6552   certainly fit nafld usually get way better wei...          joy
6553                            update aih diagnosis         fear
6554   especially blood work mr ct abdomen shortly li...          joy
6555                                        post em         fear

[6556 rows x 4 columns]
```
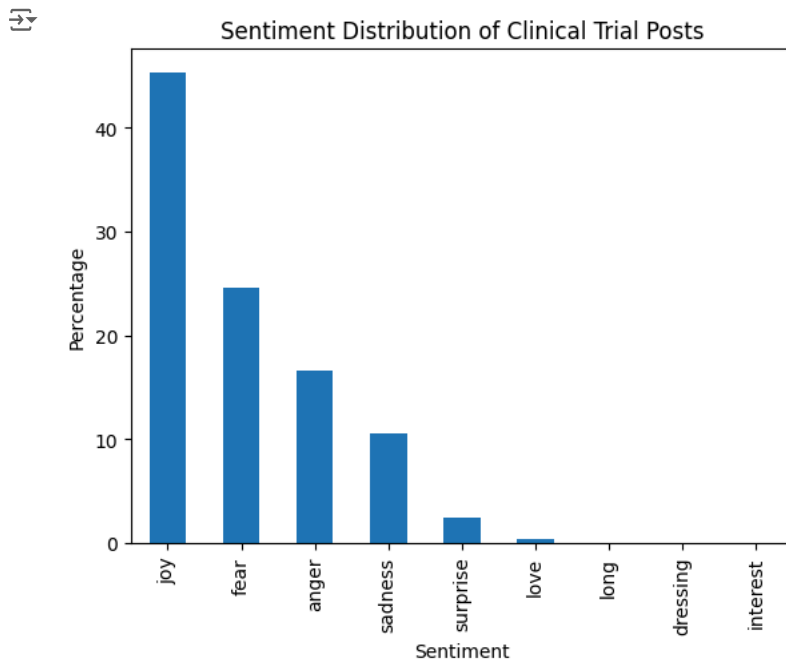
```python
sentiment_distribution = df['Emotion detected'].value_counts(normalize=True) * 100
print(sentiment_distribution)
```

```
Emotion detected
joy          45.302013
fear         24.633923
anger        16.564979
sadness      10.555217
surprise      2.471019
love          0.427090
long          0.015253
dressing      0.015253
interest      0.015253
Name: proportion, dtype: float64
```

```python
import matplotlib.pyplot as plt

# Plot sentiment distribution
sentiment_distribution.plot(kind='bar')
plt.title('Sentiment Distribution of Clinical Trial Posts')
plt.xlabel('Sentiment')
plt.ylabel('Percentage')
plt.show()
```



```python
df.to_csv('/content/drive/My Drive/df.csv', index=False)
```

Use the OpenAI API to generate personalized messages aimed at users who express interest in or could potentially benefit from participating in a clinical trial.

```
pip install openai
```

```
Requirement already satisfied: openai in /usr/local/lib/python3.10/dist-packages (0.28.0)
Requirement already satisfied: requests>=2.20 in /usr/local/lib/python3.10/dist-packages (from openai) (2.31.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from openai) (4.66.4)
Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from openai) (3.9.5)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.20->openai) (3.3.2
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.20->openai) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.20->openai) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.20->openai) (2024.2.2)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.10/dist-packages (from aiohttp->openai) (1.3.1)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->openai) (23.2.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from aiohttp->openai) (1.4.1)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.10/dist-packages (from aiohttp->openai) (6.0.5)
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->openai) (1.9.4)
Requirement already satisfied: async-timeout<5.0,>=4.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->openai) (4.0.3)
```

```python
import openai

openai.api_key = '************************'

def generate_message(text, sentiment):
    messages = [
        {"role": "system", "content": "You are a helpful assistant."},
        {"role": "user", "content": f"User sentiment: {sentiment}\nUser message: {text}\nGenerate a personalized message encouraging the us
    ]
    response = openai.ChatCompletion.create(
        model="gpt-3.5-turbo",
        messages=messages,
        max_tokens=100
    )
    print(response.choices[0].message['content'].strip())
    return response.choices[0].message['content'].strip()

df['personalized_message'] = df.apply(lambda row: generate_message(row['Text'], row['Emotion detected']), axis=1)

print(df)
```

```
---------------------------------------------------------------------------
RateLimitError                            Traceback (most recent call last)
<ipython-input-8-a6e02dc5820a> in <cell line: 20>()
     18
     19 # Generate personalized messages
---> 20 df['personalized_message'] = df.apply(lambda row: generate_message(row['Text'], row['Emotion detected']), axis=1)
     21
     22 # Display the dataframe with personalized messages

                             �’ 10 frames
/usr/local/lib/python3.10/dist-packages/openai/api_requestor.py in _interpret_response_line(self, rbody, rcode, rheaders, stream)
    763             stream_error = stream and "error" in resp.data
    764             if stream_error or not 200 <= rcode < 300:
--> 765                 raise self.handle_error_response(
    766                     rbody, rcode, resp.data, rheaders, stream_error=stream_error
    767                 )

RateLimitError: You exceeded your current quota, please check your plan and billing details. For more information on this error, read
the docs: https://platform.openai.com/docs/guides/error-codes/api-errors.
```