```python
In [1]:   import pandas as pd
```

```python
In [2]:   from datasets import load_dataset
          dataset = load_dataset('go_emotions','simplified')
```

      0%| | 0/3 [00:00<?, ?it/s]

```python
In [3]:   df_train = pd.DataFrame(dataset['train'])
          df_test = pd.DataFrame(dataset['test'])
          df_val = pd.DataFrame(dataset['validation'])
```

```python
In [4]:   df = pd.concat([df_train,df_test,df_val])
          df.reset_index(inplace=True, drop=True)
```

```python
In [5]:   df
```

Out[5]:

	text	labels	id
0	My favourite food is anything I didn't have to...	[27]	eebbqej
1	Now if he does off himself, everyone will thin...	[27]	ed00q6i
2	WHY THE FUCK IS BAYLESS ISOING	[2]	eezlygj
3	To make her feel threatened	[14]	ed7ypvh
4	Dirty Southern Wankers	[3]	ed0bdzj
...	...	...	...
54258	It's pretty dangerous when the state decides w...	[14]	edyrazk
54259	I filed for divorce this morning. Hoping he mo...	[20]	edi2z3y
54260	The last time it happened I just said, "No" an...	[10]	eewbqtx
54261	I can't stand this arrogant prick he's no bett...	[3]	eefx57m
54262	::but I like baby bangs:: /tiny voice	[18]	ed5h3jh

54263 rows × 3 columns

```python
In [6]:   for array in df['labels']:
              while len(array) > 1:
                  array.pop()
```

```python
In [7]:   label_names = {'0': 'admiration',
                         '1': 'amusement',
                         '2': 'anger',
                         '3': 'annoyance',
                         '4': 'approval',
                         '5': 'caring',
                         '6': 'confusion',
                         '7': 'curiosity',
                         '8': 'desire',
                         '9': 'disappointment',
                         '10': 'disapproval',
                         '11': 'disgust',
                         '12': 'embarrassment',
                         '13': 'excitement',
                         '14': 'fear',
```

```
                '15': 'gratitude',
                '16': 'grief',
                '17': 'joy',
                '18': 'love',
                '19': 'nervousness',
                '20': 'optimism',
                '21': 'pride',
                '22': 'realisation',
                '23': 'relief',
                '24': 'remorse',
                '25': 'sadness',
                '26': 'surprise',
                '27': 'neutral'}
```

In [8]:
```python
for i in range(len(df['labels'])):
    df['labels'][i] = label_names[str(df['labels'][i][0])]
```

In [9]:
```python
df = df.replace({'labels': {'neutral': 1, 'confusion': 2 , 'curiosity':2 ,'realisation'
                            ,'sadness' :4, 'grief':4 , 'approval' :5 , 'pride
                             'fear':7 , 'nervousness' :7 , 'embarrassment' :
                            'joy':9, 'amusement':9 , 'remorse' : 10 , 'disapp
                            'desire':11 , 'excitement' :11 , 'optimism' :11 ,
                            'disapproval' :13 , 'disgust' :13 , 'love' :14 ,
```

In [10]:
```python
df = df.drop('id', axis=1 )
```

In [11]:
```python
#removing duplicated text
index = df[df['text'].duplicated() == True].index
df.drop(index, axis = 0, inplace = True)
df.reset_index(inplace=True, drop = True)
```

In [12]:
```python
df.head()
```

Out[12]:

	text	labels
0	My favourite food is anything I didn't have to...	1
1	Now if he does off himself, everyone will thin...	1
2	WHY THE FUCK IS BAYLESS ISOING	12
3	To make her feel threatened	7
4	Dirty Southern Wankers	12

In [13]:
```python
len(df)
```

Out[13]:
```
53994
```

In [14]:
```python
len(dataset)
```

Out[14]:
```
3
```

In [15]:
```python
# Save the dataset to a CSV file
df.to_csv("emotions_preprocessed.csv", index=False)
```

In [ ]:

In [ ]:

In [ ]:

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```