In [20]:
```python
import pandas as pd
import matplotlib.pyplot as plt
```

In [21]:
```python
data = pd.read_csv('twitter_training.csv')
```

In [22]:
```python
data.head(10)
```

Out[22]:

	2401	Borderlands	Positive	im getting on borderlands and i will murder you all ,
0	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
1	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
2	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
3	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...
4	2401	Borderlands	Positive	im getting into borderlands and i can murder y...
5	2402	Borderlands	Positive	So I spent a few hours making something for fu...
6	2402	Borderlands	Positive	So I spent a couple of hours doing something f...
7	2402	Borderlands	Positive	So I spent a few hours doing something for fun...
8	2402	Borderlands	Positive	So I spent a few hours making something for fu...
9	2402	Borderlands	Positive	2010 So I spent a few hours making something f...

In [23]:
```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 74681 entries, 0 to 74680
Data columns (total 4 columns):
 #   Column                                                Non-Null Count
Dtype
---  ------                                                --------------
-----
 0   2401                                                  74681 non-null
int64
 1   Borderlands                                           74681 non-null
object
 2   Positive                                              74681 non-null
object
 3   im getting on borderlands and i will murder you all ,  73995 non-null
object
dtypes: int64(1), object(3)
memory usage: 2.3+ MB
```

In [24]:
```python
data.describe()
```

Out[24]:

	2401
count	74681.000000
mean	6432.640149
std	3740.423819
min	1.000000
25%	3195.000000
50%	6422.000000
75%	9601.000000
max	13200.000000

In [25]:
```python
col_names = ['ID', 'Entity', 'Sentiment', 'Content']
data = pd.read_csv('twitter_training.csv', names = col_names)
```

In [26]:
```python
data.head()
```

Out[26]:

	ID	Entity	Sentiment	Content
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 ...

In [27]:
```python
data.shape
```

Out[27]: (74682, 4)

In [28]:
```python
data.isna().sum()
```

Out[28]:
```
ID             0
Entity         0
Sentiment      0
Content      686
dtype: int64
```

In [29]:
```python
data.dropna(axis=0 , inplace=True)
```

In [30]: `data`

Out[30]:

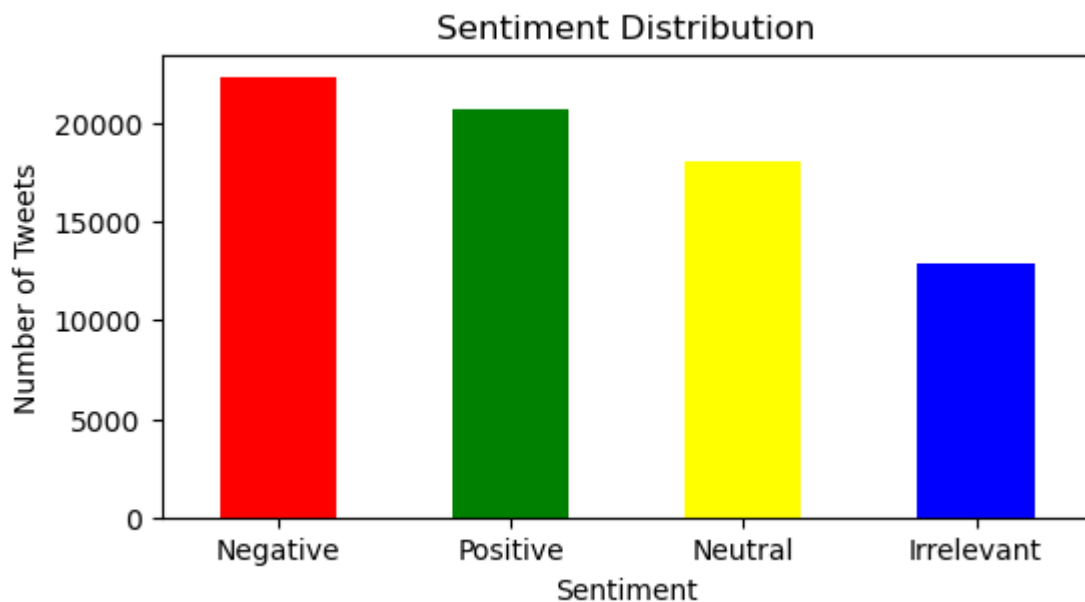|	ID	Entity	Sentiment	Content
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...
...	...	...	...	...
74677	9200	Nvidia	Positive	Just realized that the Windows partition of my...
74678	9200	Nvidia	Positive	Just realized that my Mac window partition is ...
74679	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
74680	9200	Nvidia	Positive	Just realized between the windows partition of...
74681	9200	Nvidia	Positive	Just like the windows partition of my Mac is l...

73996 rows × 4 columns

In [31]: `data.duplicated().sum()`

Out[31]: 2340

In [32]:
```python
sentiment_counts = data['Sentiment'].value_counts()
sentiment_counts
```

Out[32]:
```
Sentiment
Negative     22358
Positive     20655
Neutral      18108
Irrelevant   12875
Name: count, dtype: int64
```
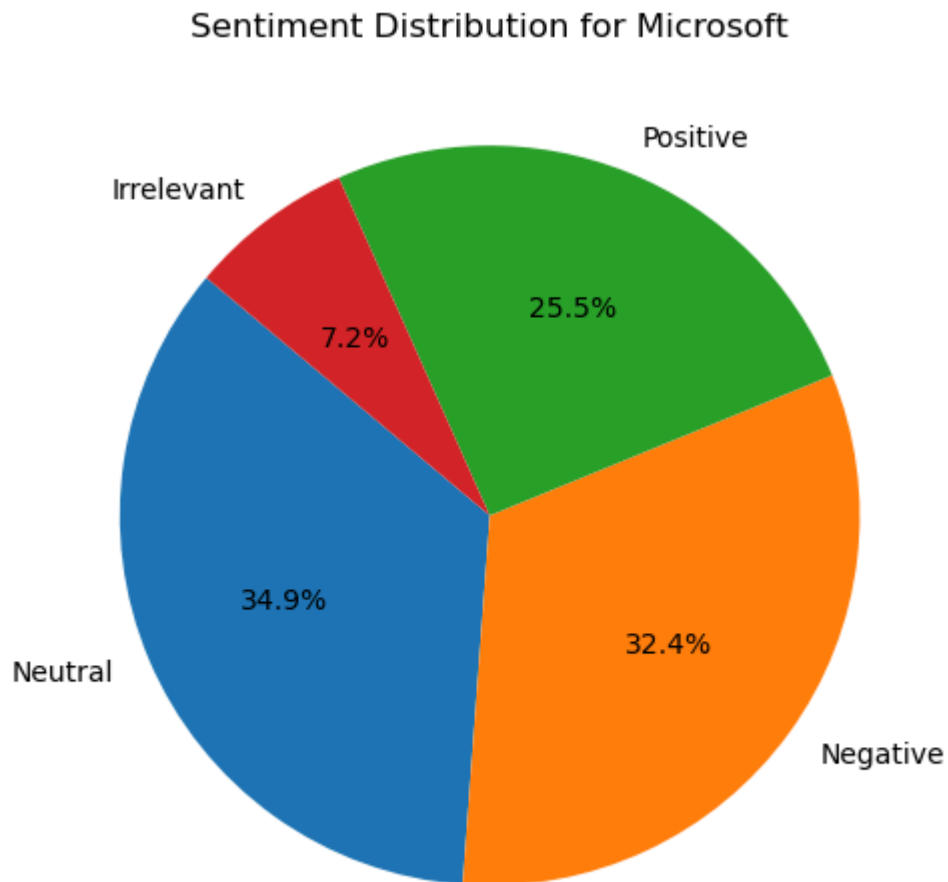
```
In [33]: plt.figure(figsize=(6, 3))
         sentiment_counts.plot(kind='bar', color=['red', 'green', 'yellow', 'blue'])
         plt.title('Sentiment Distribution')
         plt.xlabel('Sentiment')
         plt.ylabel('Number of Tweets')
         plt.xticks(rotation=0)
         plt.show()
```



```
In [34]: brand_data = data[data['Entity'].str.contains('Microsoft', case=False)]
         brand_sentiment_counts = brand_data['Sentiment'].value_counts()
         brand_sentiment_counts
```

```
Out[34]: Sentiment
         Neutral      825
         Negative     764
         Positive     602
         Irrelevant   170
         Name: count, dtype: int64
```

In [35]:
```python
plt.figure(figsize=(6, 6))
plt.pie(brand_sentiment_counts, labels=brand_sentiment_counts.index, autopc
plt.title('Sentiment Distribution for Microsoft')
plt.show()
```

Sentiment Distribution for Microsoft

Positive — 25.5%
Irrelevant — 7.2%
Neutral — 34.9%
Negative — 32.4%

In [ ]:

In [ ]:

In [ ]: