```
path='/content/drive/MyDrive/prodigy ds/twitter_training.csv'
import pandas as pd
data=pd.read_csv(path)
data.head()
```

	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...

Next steps: View recommended plots

```
from textblob import TextBlob
import matplotlib.pyplot as plt

col_names = ['ID', 'Entity', 'Sentiment', 'Content']
df = pd.read_csv(path, names=col_names)

df.head()
```

	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 ...
				im coming on borderlands and i will

Next steps: View recommended plots

```
df.describe()
```

	ID
count	74682.000000
mean	6432.586165
std	3740.427870
min	1.000000
25%	3195.000000
50%	6422.000000
75%	9601.000000
max	13200.000000

```
df.tail()
```

	ID	Entity	Sentiment	Content
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 ...
				Just realized between the windows

```
df.shape
```

```
(74682, 4)
```

```
df.isnull().sum()
```

```
ID             0
Entity         0
Sentiment      0
Content      686
dtype: int64
```

```
df.dropna(axis=0 , inplace=True)
```

```
df.isnull().sum()
```

```
ID           0
Entity       0
Sentiment    0
Content      0
dtype: int64
```

```
df.duplicated().sum()
```

```
2340
```

```
df.drop_duplicates(inplace=True)
df.duplicated().sum()
```

```
0
```

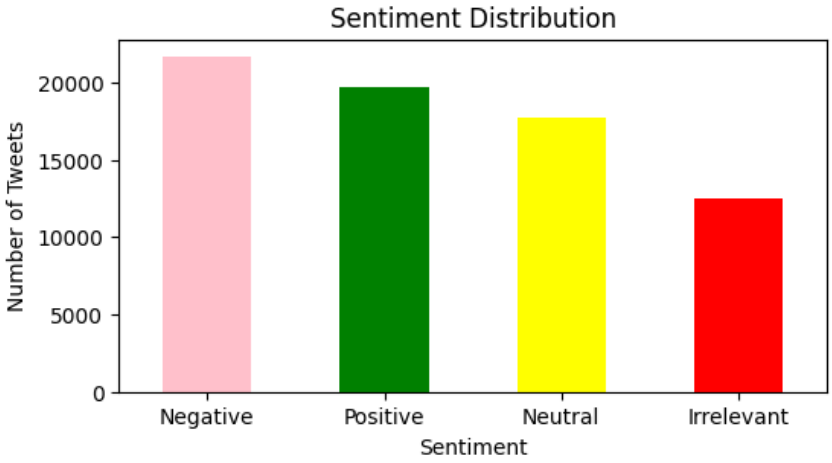
```
df.shape
```

```
(71656, 4)
```

```
sentiment_counts = df['Sentiment'].value_counts()
sentiment_counts
```

```
Negative      21698
Positive      19713
Neutral       17708
Irrelevant    12537
Name: Sentiment, dtype: int64
```

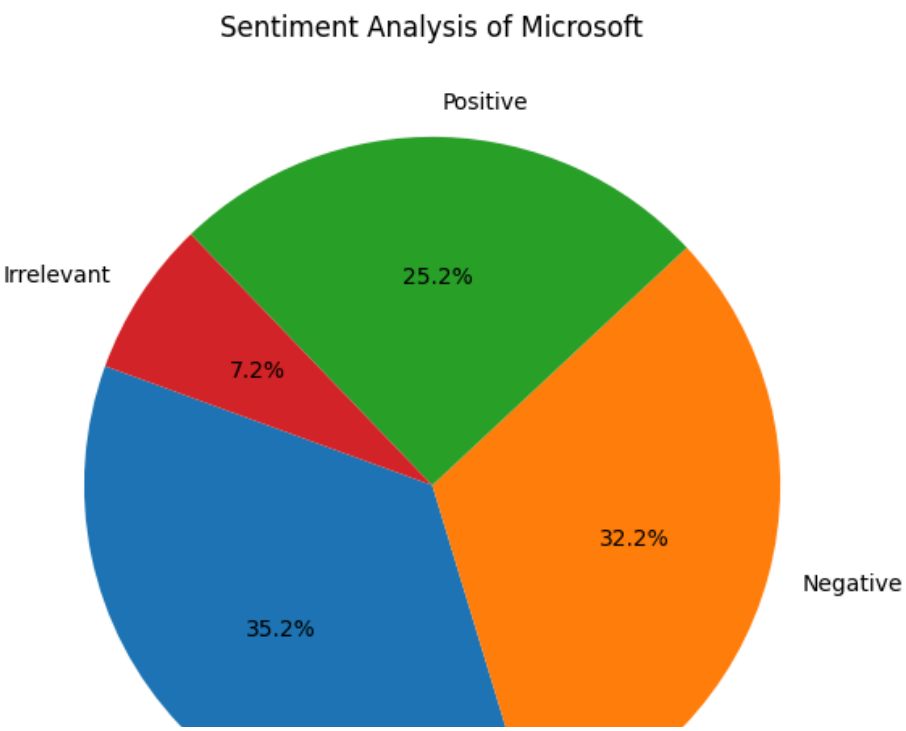
```
plt.figure(figsize=(6, 3))
sentiment_counts.plot(kind='bar', color=['pink', 'green', 'yellow', 'red'])
plt.title('Sentiment Distribution')
plt.xlabel('Sentiment')
plt.ylabel('Number of Tweets')
plt.xticks(rotation=0)
plt.show()
```




```
brand_data = df[df['Entity'].str.contains('Microsoft', case=False)]
brand_sentiment_counts = brand_data['Sentiment'].value_counts()
brand_sentiment_counts
```

```
Neutral       846
Negative      774
Positive      606
Irrelevant    174
Name: Sentiment, dtype: int64
```

```
plt.figure(figsize=(7, 7))
plt.pie(brand_sentiment_counts, labels=brand_sentiment_counts.index, autopct='%
plt.title('Sentiment Analysis of Microsoft')
plt.show()
```


Sentiment Analysis of Microsoft
Positive
Irrelevant
25.2%
7.2%
32.2%
Negative
35.2%