In [1]:
```
import pandas as pd
from textblob import TextBlob
import matplotlib.pyplot as plt

data = pd.read_csv('twitter_training.csv')
data.head()
```

Out[1]:

	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...

In [2]:
```
col_names = ['ID', 'Entity', 'Sentiment', 'Content']
df = pd.read_csv('twitter_training.csv', names=col_names)
df.head()
```

Out[2]:

	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 [3]:
```
df.describe()
```

Out[3]:

	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

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
ID             0
Entity         0
Sentiment      0
Content      686
dtype: int64
```

In [5]:
```
df.dropna(axis=0 , inplace=True)
df.isnull().sum()
```

Out[5]:
```
ID           0
Entity       0
Sentiment    0
Content      0
dtype: int64
```

In [6]:
```
df.duplicated().sum()
```

Out[6]:
```
2340
```

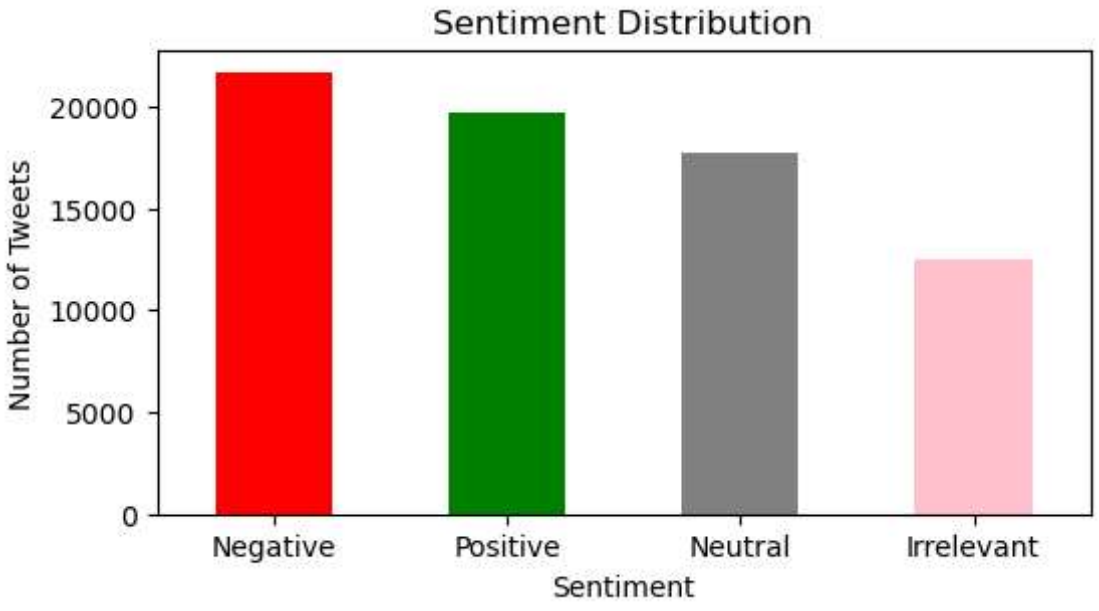
In [7]:
```
df.drop_duplicates(inplace=True)
df.duplicated().sum()
```

Out[7]:
```
0
```

In [8]:
```
sentiment_counts = df['Sentiment'].value_counts()
sentiment_counts
```

Out[8]:
```
Negative      21698
Positive      19713
Neutral       17708
Irrelevant    12537
Name: Sentiment, dtype: int64
```

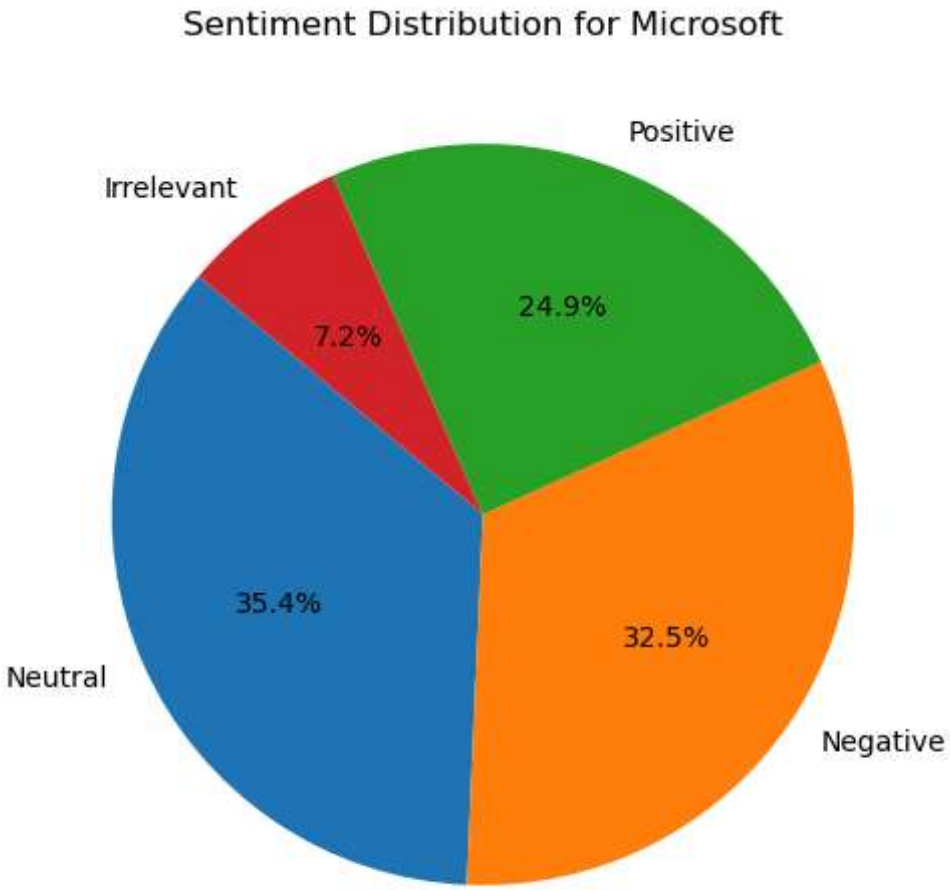
In [9]:
```
plt.figure(figsize=(6, 3))
sentiment_counts.plot(kind='bar', color=['red', 'green', 'grey', 'pink'])
plt.title('Sentiment Distribution')
plt.xlabel('Sentiment')
plt.ylabel('Number of Tweets')
plt.xticks(rotation=0)
plt.show()
```



In [10]:
```
brand_data = df[df['Entity'].str.contains('Microsoft', case=False)]
brand_sentiment_counts = brand_data['Sentiment'].value_counts()
brand_sentiment_counts
```

Out[10]:
```
Neutral       816
Negative      748
Positive      573
Irrelevant    167
Name: Sentiment, dtype: int64
```

In [11]:
```
plt.figure(figsize=(6, 6))
plt.pie(brand_sentiment_counts, labels=brand_sentiment_counts.index, autopct
plt.title('Sentiment Distribution for Microsoft')
plt.show()
```



In [ ]:

In [ ]:

In [ ]: