In [4]:
```
import pandas as pd
from textblob import TextBlob
import matplotlib.pyplot as plt
```

In [2]:
```
data= pd.read_csv('twitter_training.csv')
```

In [3]:
```
data.head()
```

Out[3]:

	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 [5]:
```
col_names=['ID','Entity','Sentiments','Contest']
df=pd.read_csv('twitter_training.csv', names=col_names)
```

In [6]:
```
df.head()
```

Out[6]:

	ID	Entity	Sentiments	Contest
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 [7]:
```
df.shape
```

Out[7]:
```
(74682, 4)
```

In [8]:
```
df.describe
```

Out[8]:
```
<bound method NDFrame.describe of          ID       Entity Sentiments  \
0      2401  Borderlands   Positive
1      2401  Borderlands   Positive
2      2401  Borderlands   Positive
3      2401  Borderlands   Positive
4      2401  Borderlands   Positive
...     ...          ...        ...
74677  9200       Nvidia   Positive
74678  9200       Nvidia   Positive
74679  9200       Nvidia   Positive
74680  9200       Nvidia   Positive
74681  9200       Nvidia   Positive

                                              Contest
0      im getting on borderlands and i will murder yo...
1      I am coming to the borders and I will kill you...
2      im getting on borderlands and i will kill you ...
3      im coming on borderlands and i will murder you...
4      im getting on borderlands 2 and i will murder ...
...                                                  ...
74677  Just realized that the Windows partition of my...
74678  Just realized that my Mac window partition is ...
74679  Just realized the windows partition of my Mac ...
74680  Just realized between the windows partition of...
74681  Just like the windows partition of my Mac is l...

[74682 rows x 4 columns]>
```

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
ID              0
Entity          0
Sentiments      0
Contest       686
dtype: int64
```

In [11]:
```
df.duplicated().sum()
```

Out[11]:
```
np.int64(2700)
```

In [12]:
```
df.drop_duplicates(inplace=True)
df.duplicated().sum()
```

Out[12]:
```
np.int64(0)
```

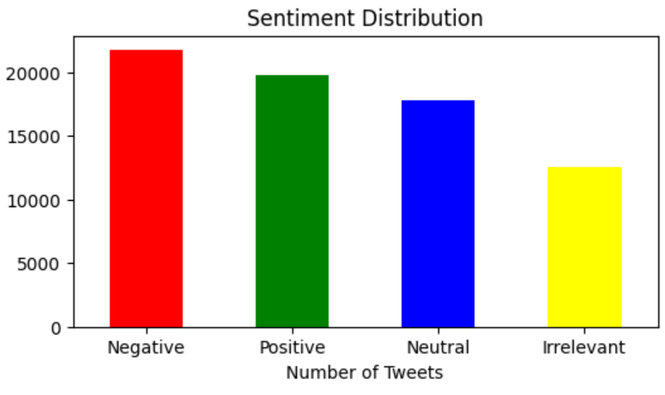
In [13]:
```
df.shape
```

Out[13]:
```
(71982, 4)
```

In [14]:
```
sentiment_counts=df['Sentiments'].value_counts()
sentiment_counts
```

Out[14]:
```
Sentiments
Negative      21787
Positive      19811
Neutral       17800
Irrelevant    12584
Name: count, dtype: int64
```

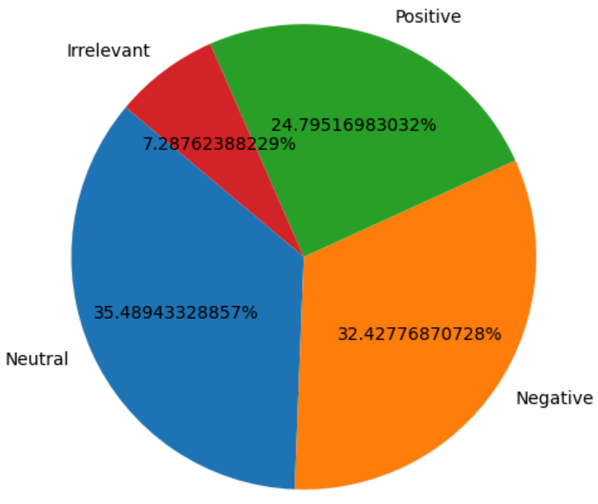
In [15]:
```
plt.figure(figsize=(6,3))
sentiment_counts.plot(kind='bar',color=['red','green','blue','yellow'])
plt.title('Sentiment Distribution')
plt.xlabel('Number of Tweets')
plt.xticks(rotation=0)
plt.show()
```



In [16]:
```
brand_data=df[df['Entity'].str.contains('Microsoft',case=False)]
brand_sentiment_counts=brand_data['Sentiments'].value_counts()
brand_sentiment_counts
```

Out[16]:
```
Sentiments
Neutral       823
Negative      752
Positive      575
Irrelevant    169
Name: count, dtype: int64
```

In [17]:
```
plt.figure(figsize=(6,6))
plt.pie(brand_sentiment_counts,labels=brand_sentiment_counts.index,autopct='%1.11f%%',startangle=140)
plt.show()
```


Positive
Irrelevant
24.79516983032%
7.28762388229%
35.48943328857%
32.42776870728%
Neutral
Negative

In [ ]: