Analyze and Visualize Sentiment Patterns in Social Media Data

Submission Report

Name: Tarun Chincholia

Internship Domain: Data Science

Task Number: Task 4

Organization: Prodigy InfoTech

Date: 8-08-25

Objective: To analyze and visualize sentiment patterns in social media data to understand public opinion and attitudes towards specific topics or brands. This involves collecting data from social media platforms, performing text preprocessing, applying sentiment analysis techniques, and creating visualizations to identify trends, patterns, and insights.

Dataset Used:

Link : https://github.com/Prodigy-InfoTech/data-science-datasets/tree/main/Task%204

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

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

data.head()
```

```
   2401  Borderlands  Positive  \
0  2401  Borderlands  Positive
1  2401  Borderlands  Positive
2  2401  Borderlands  Positive
3  2401  Borderlands  Positive
4  2401  Borderlands  Positive

  im getting on borderlands and i will murder you all ,
0  I am coming to the borders and I will kill you...
1  im getting on borderlands and i will kill you ...
2  im coming on borderlands and i will murder you...
3  im getting on borderlands 2 and i will murder ...
4  im getting into borderlands and i can murder y...
```

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

df.head()
```

```
     ID       Entity Sentiment  \
0  2401  Borderlands  Positive
1  2401  Borderlands  Positive
2  2401  Borderlands  Positive
3  2401  Borderlands  Positive
4  2401  Borderlands  Positive

                                             Content
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 ...
```

```
df.shape
```

```
(74682, 4)
```

```
df.describe
```

```
<bound method NDFrame.describe of          ID       Entity
Sentiment  \
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

                                                  Content
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]>
```

```
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
```

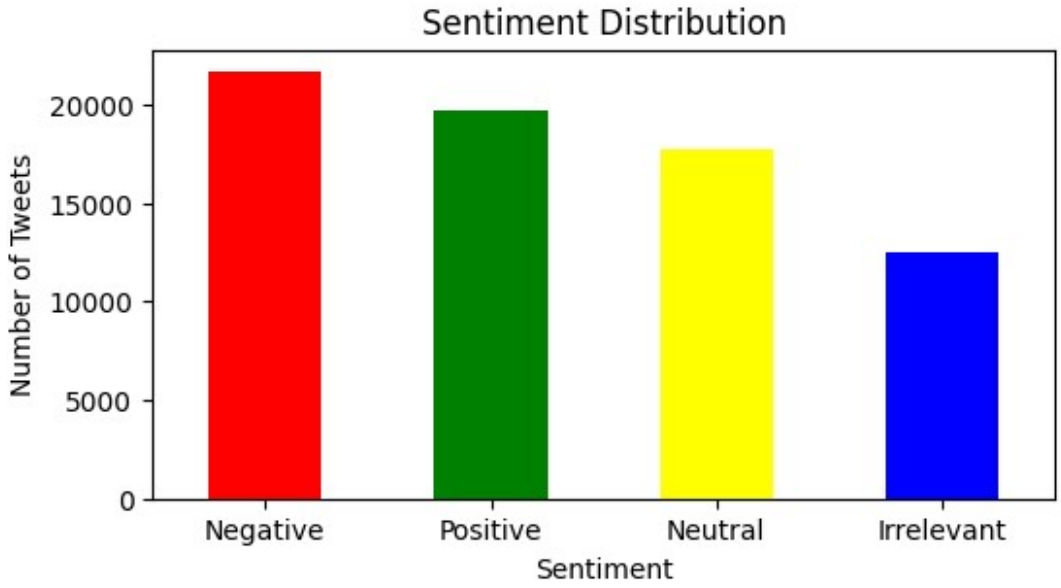
```
df.shape
```

```
(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=['red', 'green', 'yellow',
'blue'])
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       816
Negative      748
Positive      573
Irrelevant    167
Name: Sentiment, dtype: int64
```

```
plt.figure(figsize=(6, 6))
plt.pie(brand_sentiment_counts, labels=brand_sentiment_counts.index,
autopct='%1.1f%%', startangle=140)
plt.title('Sentiment Distribution for Microsoft')
plt.show()
```