PRODIGY_DS_04

September 20, 2024

1 TASK 4

Analyze and visualize sentiment patterns in social media data to understand public opinion and attitudes towards specific topics or brands.

[9]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

[11]:
```
pip install textblob
```

```
Defaulting to user installation because normal site-packages is not writeable
Looking in links: /usr/share/pip-wheels
Collecting textblob
  Obtaining dependency information for textblob from https://files.pythonhosted.
org/packages/02/07/5fd2945356dd839974d3a25de8a142dc37293c21315729a41e775b5f3569/
textblob-0.18.0.post0-py3-none-any.whl.metadata
  Downloading textblob-0.18.0.post0-py3-none-any.whl.metadata (4.5 kB)
Requirement already satisfied: nltk>=3.8 in /opt/conda/envs/anaconda-
panel-2023.05-py310/lib/python3.11/site-packages (from textblob) (3.8.1)
Requirement already satisfied: click in /opt/conda/envs/anaconda-
panel-2023.05-py310/lib/python3.11/site-packages (from nltk>=3.8->textblob)
(8.0.4)
Requirement already satisfied: joblib in /opt/conda/envs/anaconda-
panel-2023.05-py310/lib/python3.11/site-packages (from nltk>=3.8->textblob)
(1.2.0)
Requirement already satisfied: regex>=2021.8.3 in /opt/conda/envs/anaconda-
panel-2023.05-py310/lib/python3.11/site-packages (from nltk>=3.8->textblob)
(2022.7.9)
Requirement already satisfied: tqdm in /opt/conda/envs/anaconda-
panel-2023.05-py310/lib/python3.11/site-packages (from nltk>=3.8->textblob)
(4.65.0)
Downloading textblob-0.18.0.post0-py3-none-any.whl (626 kB)
                           626.3/626.3
kB 13.8 MB/s eta 0:00:000:00:01
Installing collected packages: textblob
Successfully installed textblob-0.18.0.post0
Note: you may need to restart the kernel to use updated packages.
```

[15]:
```
data= pd.read_csv('twitter_training.csv')
```

[17]:
```
data.head()
```

[17]:
```
    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…
```

[19]:
```
col_names=['ID','Entity','Sentiments','Contest']
df=pd.read_csv('twitter_training.csv', names=col_names)
```

[21]:
```
df.head()
```

[21]:
```
      ID       Entity Sentiments  \
0  2401  Borderlands   Positive
1  2401  Borderlands   Positive
2  2401  Borderlands   Positive
3  2401  Borderlands   Positive
4  2401  Borderlands   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 …
```

[23]:
```
df.shape
```

[23]:
```
(74682, 4)
```

[25]:
```
df.describe
```

[25]:
```
<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]>
```

[27]:
```
df.isnull().sum()
```

[27]:
```
ID              0
Entity          0
Sentiments      0
Contest       686
dtype: int64
```

[29]:
```
df.dropna(axis=0,inplace=True)
```

[31]:
```
df.isnull().sum()
```

[31]:
```
ID            0
Entity        0
Sentiments    0
Contest       0
dtype: int64
```

[33]:
```
df.duplicated().sum()
```

[33]:
```
2340
```

[35]:
```
df.drop_duplicates(inplace=True)
df.duplicated().sum()
```

[35]:
```
0
```

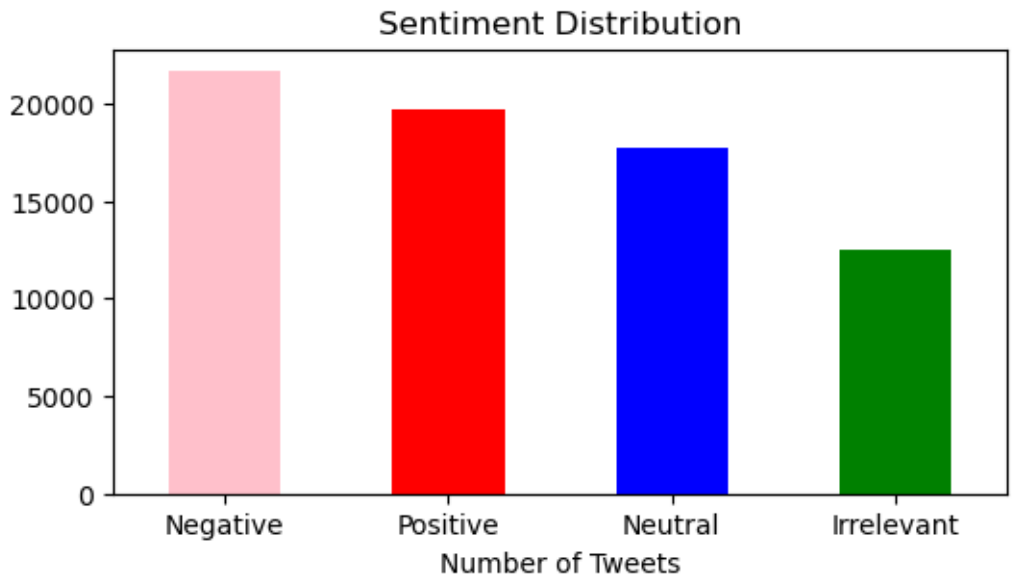
[37]:
```
df.shape
```

[37]:
```
(71656, 4)
```

[39]:
```
sentiment_counts=df['Sentiments'].value_counts()
sentiment_counts
```

[39]:
```
Sentiments
Negative      21698
Positive      19713
Neutral       17708
Irrelevant    12537
Name: count, dtype: int64
```

[41]:
```
plt.figure(figsize=(6,3))
sentiment_counts.plot(kind='bar',color=['pink','red','blue','green'])
plt.title('Sentiment Distribution')
plt.xlabel('Number of Tweets')
plt.xticks(rotation=0)
plt.show()
```



[43]:
```
brand_data=df[df['Entity'].str.contains('Microsoft',case=False)]
brand_sentiment_counts=brand_data['Sentiments'].value_counts()
brand_sentiment_counts
```

[43]:
```
Sentiments
Neutral       816
Negative      748
Positive      573
Irrelevant    167
Name: count, dtype: int64
```

[45]:
```
import matplotlib.pyplot as plt
plt.figure(figsize=(6, 6))
colors = ['#ff9999','#66b3ff','#99ff99','#ffcc99']
plt.pie(brand_sentiment_counts, labels=brand_sentiment_counts.index,␣
 ↪autopct='%1.1f%%', startangle=140, colors=colors)
plt.show()
```