

PRODIGY INFOTECH - DATASCIENCE INTERNSHIP

TASK - 4

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

```
In [31]: #Import required libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [32]: #Load the data
Data=pd.read_csv("C:/Users/steph/OneDrive/Documents/twitter.csv")
print(Data)
```

```
          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]

```
In [33]: #Create a dataframe
df=pd.DataFrame(Data)
df
```

Out[33]:

	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 ...
...
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 ...
74680	9200	Nvidia	Positive	Just realized between the windows partition of...
74681	9200	Nvidia	Positive	Just like the windows partition of my Mac is l...

74682 rows × 4 columns

In [34]:

```
#First 5 rows of the dataset  
df.head()
```

Out[34]:

	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 [35]:

```
#Last 5 rows of the dataset  
df.tail
```

```
Out[35]: <bound method NDFrame.tail of
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]>

```
In [36]: #Check for the information ,i.e, dtype and null value for each column
df.info
```

```
Out[36]: <bound method DataFrame.info of
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]>

```
In [37]: #Check for Statistical analysis  
df.describe()
```

```
Out[37]: 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 [38]: #Check for unique values  
df.nunique()
```

```
Out[38]: ID      12447  
Entity    32  
Sentiment 4  
Content   69489  
dtype: int64
```

```
In [39]: #Check for the no.of rows and columns  
df.shape
```

```
Out[39]: (74682, 4)
```

```
In [40]: #Check for the data type  
df.dtypes
```

```
Out[40]: ID      int64  
Entity    object  
Sentiment  object  
Content   object  
dtype: object
```

```
In [41]: #Checking for null values  
df.isnull().sum()
```

```
Out[41]: ID      0  
Entity    0  
Sentiment 0  
Content   686  
dtype: int64
```

```
In [42]: #dropping irrelevant columns  
df.dropna(subset=['Content'], inplace=True)
```

```
In [43]: #Check for null values  
df.isnull().sum()
```

```
Out[43]: ID      0  
Entity    0  
Sentiment 0  
Content   0  
dtype: int64
```

```
In [44]: #Check for the unique values in the column 'Sentiment'  
df.Sentiment.unique()
```

```
Out[44]: array(['Positive', 'Neutral', 'Negative', 'Irrelevant'], dtype=object)
```

```
In [45]: #Replacing 'Irrelevant' by 'Neutral' in the 'Sentiment' column  
df.Sentiment=df.Sentiment.replace('Irrelevant','Neutral')  
df.Sentiment
```

```
Out[45]: 0      Positive  
1      Positive  
2      Positive  
3      Positive  
4      Positive  
...  
74677  Positive  
74678  Positive  
74679  Positive  
74680  Positive  
74681  Positive  
Name: Sentiment, Length: 73996, dtype: object
```

```
In [46]: #Check for the unique values of the column 'Sentiment'  
df.Sentiment.unique()
```

```
Out[46]: array(['Positive', 'Neutral', 'Negative'], dtype=object)
```

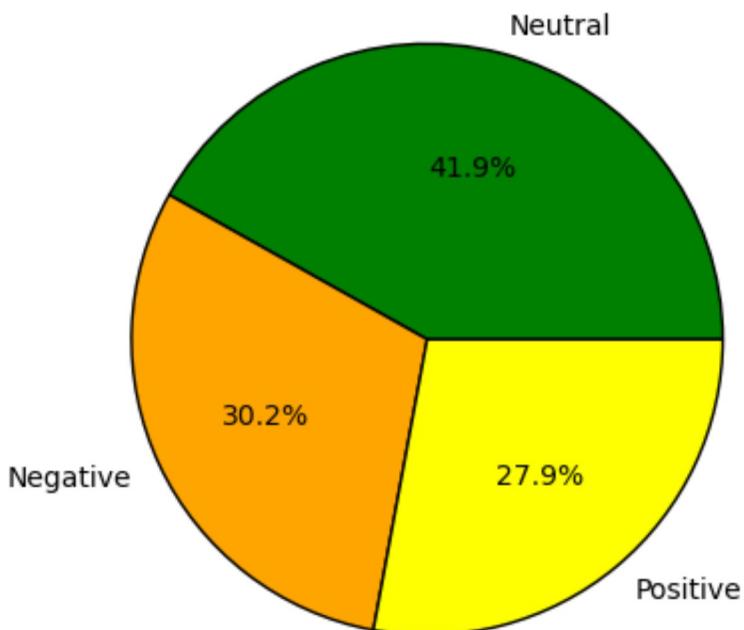
```
In [47]: #Counting the occurrences of each value in the 'Sentiment' column  
Sentiment_count=df.Sentiment.value_counts()  
Sentiment_count
```

```
Out[47]: Neutral     30983  
Negative    22358  
Positive    20655  
Name: Sentiment, dtype: int64
```

```
In [82]: plt.pie(Sentiment_count, labels= ['Neutral','Negative','Positive'],colors=['green',  
                           wedgeprops={'edgecolor':'black'}, autopct='%0.1f%%')  
plt.title('Pie chart of Sentiments')  
plt.tight_layout  
plt.show
```

```
Out[82]: <function matplotlib.pyplot.show(close=None, block=None)>
```

Pie chart of Sentiments



```
In [49]: #Unique values of the column 'Entity'  
df.Entity.unique()
```

```
Out[49]: array(['Borderlands', 'CallOfDutyBlackopsColdWar', 'Amazon', 'Overwatch',  
       'Xbox(Xseries)', 'NBA2K', 'Dota2', 'PlayStation5(PS5)',  
       'WorldOfCraft', 'CS-GO', 'Google', 'AssassinsCreed', 'ApexLegends',  
       'LeagueOfLegends', 'Fortnite', 'Microsoft', 'Hearthstone',  
       'Battlefield', 'PlayerUnknownsBattlegrounds(PUBG)', 'Verizon',  
       'HomeDepot', 'FIFA', 'RedDeadRedemption(RDR)', 'CallOfDuty',  
       'TomClancysRainbowSix', 'Facebook', 'GrandTheftAuto(GTA)',  
       'MaddenNFL', 'johnson&johnson', 'Cyberpunk2077',  
       'TomClancysGhostRecon', 'Nvidia'], dtype=object)
```

```
In [50]: #Counting the occurrences of each value in the 'Entity' column  
Entity_count=df.Entity.value_counts()  
Entity_count
```

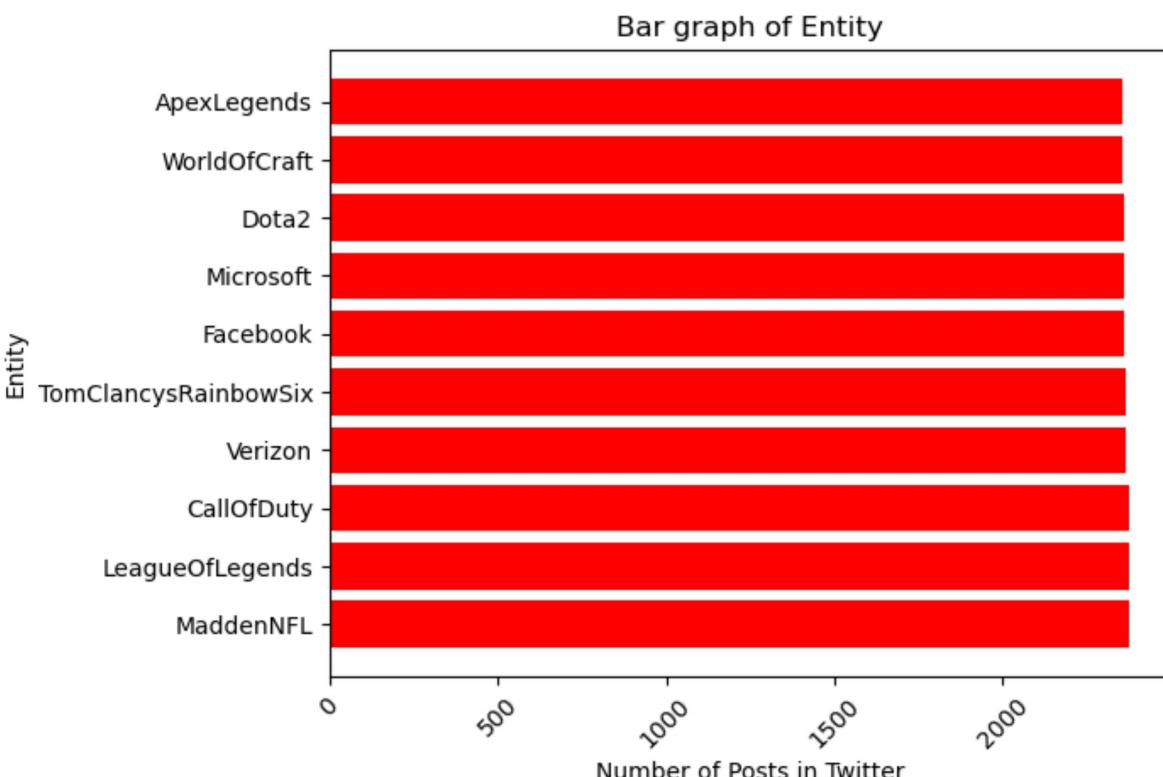
```
Out[50]: MaddenNFL          2377  
LeagueOfLegends           2377  
CallOfDuty                2376  
Verizon                   2365  
TomClancysRainbowSix     2364  
Facebook                  2362  
Microsoft                 2361  
Dota2                     2359  
WorldOfCraft              2357  
ApexLegends               2353  
NBA2K                     2343  
CallOfDutyBlackopsColdWar 2343  
FIFA                      2324  
johnson&johnson          2324  
TomClancysGhostRecon     2321  
Battlefield                2316  
Overwatch                 2316  
GrandTheftAuto(GTA)       2293  
HomeDepot                 2292  
PlayStation5(PS5)         2291  
Hearthstone               2286  
CS-GO                     2284  
Xbox(Xseries)             2283  
Borderlands                2280  
Amazon                    2276  
Google                     2274  
Nvidia                     2271  
Cyberpunk2077              2262  
RedDeadRedemption(RDR)    2249  
Fortnite                   2249  
PlayerUnknownsBattlegrounds(PUBG) 2234  
AssassinsCreed             2234  
Name: Entity, dtype: int64
```

```
In [51]: #1st 10 value of the column 'Entity'  
First_10_Entity=Entity_count.head(10)  
First_10_Entity
```

```
Out[51]: MaddenNFL          2377  
LeagueOfLegends           2377  
CallOfDuty                2376  
Verizon                   2365  
TomClancysRainbowSix     2364  
Facebook                  2362  
Microsoft                 2361  
Dota2                     2359  
WorldOfCraft              2357  
ApexLegends               2353  
Name: Entity, dtype: int64
```

```
In [83]: #Plotting the bar graph of 'Entity'
Entity=['MaddenNFL','LeagueOfLegends','CallOfDuty','Verizon','TomClancysRainbowSix'
        'Facebook','Microsoft','Dota2','WorldOfCraft','ApexLegends']
Id=[2377,2377,2376,2365,2364,2362,2361,2359,2357,2353]

plt.barh(Entity,Id,color='red')
plt.xticks(rotation=45)
plt.ylabel('Entity')
plt.xlabel('Number of Posts in Twitter')
plt.title('Bar graph of Entity')
plt.show()
```



```
In [53]: #1st three values of 'Entity'
First_5_Entity=Entity_count.head(3)
First_5_Entity
```

```
Out[53]: MaddenNFL      2377
LeagueOfLegends    2377
CallOfDuty        2376
Name: Entity, dtype: int64
```

```
In [54]: #Extracting the index values and converting them into list
First_5_Entity_list=First_5_Entity.index.tolist()
First_5_Entity_list
```

```
Out[54]: ['MaddenNFL', 'LeagueOfLegends', 'CallOfDuty']
```

```
In [59]: # Grouping by 'Entity' and 'Sentiments' and counting the occurrences of each unique s
Sentiment_and_Entity=df.loc[df['Entity'].isin(First_5_Entity_list)].groupby('Entity')
Sentiment_and_Entity
```

```
Out[59]: Entity      Sentiment
CallOfDuty    Negative     883
              Neutral    1047
              Positive     446
LeagueOfLegends Negative    632
                     Neutral   1130
                     Positive   615
MaddenNFL     Negative   1694
                     Neutral   287
                     Positive   396
Name: Sentiment, dtype: int64
```

```
In [73]: # Plotting the pie chart of the grouped columns
plt.figure(figsize=(10,5))

Sentiments=['Neutral' , 'Negative' , 'Positive']
color=['red' , 'brown','orange']

plt.subplot(2,3,1)
plt.pie(Sentiment_and_Entity[:3] , labels=Sentiments , autopct='%.1f%%' , colors=color)

plt.subplot(2,3,2)
plt.pie(Sentiment_and_Entity[3:6] , labels=Sentiments , autopct='%.1f%%' , colors=color)

plt.subplot(2,3,3)
plt.pie(Sentiment_and_Entity[6:] , labels=Sentiments , autopct='%.1f%%' , colors=color)
```

```
Out[73]: ([<matplotlib.patches.Wedge at 0x1e1bd7e2b10>,
<matplotlib.patches.Wedge at 0x1e1bd818150>,
<matplotlib.patches.Wedge at 0x1e1bd819d10>],
[Text(-0.6814456002540091, 0.8634997938010486, 'Neutral'),
Text(0.1586389899481369, -1.0885006526723975, 'Negative'),
Text(0.9527489765829378, -0.5497903124102541, 'Positive')],
[Text(-0.3716976001385504, 0.4709998875278446, '71.3%'),
Text(0.0865303581535292, -0.5937276287303985, '12.1%'),
Text(0.5196812599543297, -0.2998856249510477, '16.7%')])
```

