

```
In [3]: pip install matplotlib
Requirement already satisfied: matplotlib in c:\users\devat\anaconda3\lib\site-packages (3.8.0)
Requirement already satisfied: contourpy<1.8.1 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler<=3.10 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools<=4.25.0 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver<=1.0.1 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy<2,>=1.21 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib) (1.26.4)
Requirement already satisfied: packaging<=20.9 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib) (21.1)
Requirement already satisfied: pillow<=6.2.0 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib) (8.2.0)
Requirement already satisfied: pyparsing<=3.1 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil<=2.7 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six<=1.5 in c:\users\devat\anaconda3\lib\site-packages (from python-dateutil<=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

In [4]: pip install pandas
Requirement already satisfied: pandas in c:\users\devat\anaconda3\lib\site-packages (2.1.4)
Requirement already satisfied: numpy<2,>=1.23.2 in c:\users\devat\anaconda3\lib\site-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil<=2.8.2 in c:\users\devat\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz<=2020.1 in c:\users\devat\anaconda3\lib\site-packages (from pandas) (2023.3.post1)
Requirement already satisfied: tzdata<=2022.1 in c:\users\devat\anaconda3\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: six<=1.5 in c:\users\devat\anaconda3\lib\site-packages (from python-dateutil<=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

In [5]: pip install seaborn
Requirement already satisfied: seaborn in c:\users\devat\anaconda3\lib\site-packages (0.12.2)
Requirement already satisfied: numpy<1.24.0,>=1.17 in c:\users\devat\anaconda3\lib\site-packages (from seaborn) (1.26.4)
Requirement already satisfied: pandas<=2.0.5 in c:\users\devat\anaconda3\lib\site-packages (from seaborn) (2.1.4)
Requirement already satisfied: matplotlib<=3.6.1,>=3.1 in c:\users\devat\anaconda3\lib\site-packages (from seaborn) (3.8.0)
Requirement already satisfied: contourpy<1.8.1 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib<=3.6.1,>=3.1->seaborn) (1.2.0)
Requirement already satisfied: cycler<=0.10 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib<=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools<=4.22.0 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib<=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: kiwisolver<=1.0.1 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib<=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: packaging<=20.9 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib<=3.6.1,>=3.1->seaborn) (21.1)
Requirement already satisfied: pillow<=6.2.0 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib<=3.6.1,>=3.1->seaborn) (10.2.0)
Requirement already satisfied: pyparsing<=3.1 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib<=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil<=2.7 in c:\users\devat\anaconda3\lib\site-packages (from matplotlib<=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz<=2020.1 in c:\users\devat\anaconda3\lib\site-packages (from pandas<=2.0.5->seaborn) (2023.3.post1)
Requirement already satisfied: tzdata<=2022.1 in c:\users\devat\anaconda3\lib\site-packages (from pandas<=2.0.5->seaborn) (2023.3)
Requirement already satisfied: six<=1.5 in c:\users\devat\anaconda3\lib\site-packages (from python-dateutil<=2.7->matplotlib<=3.6.1,>=3.1->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
In [6]: pip install numpy
Requirement already satisfied: numpy in c:\users\devat\anaconda3\lib\site-packages (1.26.4)
Note: you may need to restart the kernel to use updated packages.
```

```
In [ ]:
In [ ]:
In [ ]:
```

```
In [7]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import zipfile
```

```
In [8]: csv_file_path = r'C:\Users\devat\OneDrive\Desktop\Tecno Hack\Task 2 Data set.csv'
column_names = ['ID', 'entity', 'sentiment', 'comment']
df = pd.read_csv(csv_file_path, header=0, names=column_names)
```

```
In [43]: df.head()
```

	ID	entity	sentiment	comment
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 [44]: df.shape
Out[44]: (74881, 4)
```

```
In [45]: # Count of unique entities
entity_count = df['entity'].value_counts()
print(entity_count)

entity
TomClancysRainbowSix      2490
MaddenNFL                 2480
Microsoft                 2460
LeagueOfLegends           2394
CallOfDuty                2394
Verizon                   2392
CallOfDutyBlackopsColdWar 2376
ApexLegends               2376
Facebook                  2370
WorldOfCraft              2364
Dota2                     2364
NBA2K                     2352
TomClancysGhostRecon      2346
Battlefield               2346
FIFA                      2340
Xbox(Xseries)             2334
Overwatch                 2334
JohnsonsJohnson          2328
Amazon                   2316
PlayStations(PSS)         2310
Walmart                  2310
Cyberpunk2077             2304
CS:GO                    2304
GrandTheftAuto(GTA)       2294
Hearthstone               2298
Nvidia                   2298
Google                   2298
Borderlands               2285
PlayerUnknownsBattleGrounds(PUBG) 2274
Fortnite                  2274
RedDeadRedemption(RDR)    2262
AssassinsCreed            2244
Name: count, dtype: int64
```

```
In [46]: df.info
```

```
Out[46]: <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
...
74676  9208  Nvidia      Positive
74677  9208  Nvidia      Positive
74678  9208  Nvidia      Positive
74679  9208  Nvidia      Positive
74680  9208  Nvidia      Positive

comment
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...
...
74676  Just realized that the Windows partition of my...
74677  Just realized that my Mac window partition is ...
74678  Just realized the windows partition of my Mac ...
74679  Just realized between the windows partition of...
74680  Just like the windows partition of my Mac is l...

[74881 rows x 4 columns]>
```

```
In [47]: duplicates = df.duplicated()
duplicated_rows = df[duplicates]
duplicated_rows.count()
```

```
Out[47]: ID      2760
entity    2760
sentiment 2760
comment   2340
dtype: int64
```

```
In [48]: # Dropping Duplicates
df.drop_duplicates(inplace=True)
```

```
In [49]: df.isnull().sum()
```

```
Out[49]: ID      0
entity    0
sentiment  0
comment    0
dtype: int64
```

```
In [50]: # Dropping missing value [ Using Dropna]
df = df.dropna()
```

```
In [51]: df.isnull().sum()
```

```
Out[51]: ID      0
entity    0
sentiment  0
comment    0
dtype: int64
```

```
In [52]: # Number of Unique Values
df.nunique()
```

```
Out[52]: ID      12447
entity      32
sentiment    4
comment     69498
dtype: int64
```

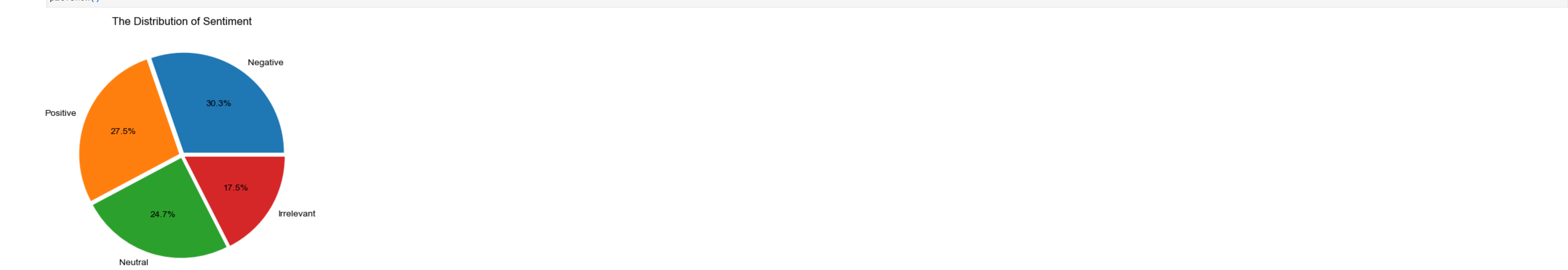
```
In [64]: # Define the number of rows you want to print
num_rows_to_print = 5

# Iterate over the DataFrame for the specified number of rows
for i in range(min(num_rows_to_print, len(df))):
    print(f"[{i+1}]: {df['comment'][i]} -> {df['sentiment'][i]}")
1: I am coming to the borders and i will kill you all, -> Positive
2: im getting on borderlands and i will kill you all, -> Positive
3: im coming on borderlands and i will murder you all, -> Positive
4: im getting on borderlands 2 and i will murder you all, -> Positive
5: im getting into borderlands and i can murder you all, -> Positive
```

```
In [57]: #SENTIMENT ANALYSIS
df['sentiment'].value_counts()
```

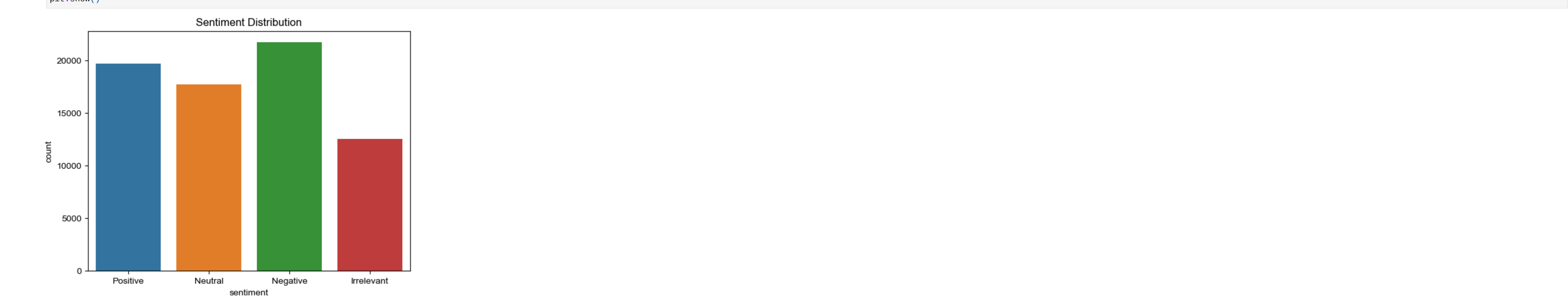
```
Out[57]: sentiment
Negative  21698
Positive  19722
Neutral   17788
Irrelevant 22537
Name: count, dtype: int64
```

```
In [58]: plt.figure(figsize=(10,5))
plt.pie(x=df['sentiment'].value_counts().values,
labels=df['sentiment'].value_counts().index,
autopct='%1.1%', explode=(0.03, 0.03, 0.03, 0.03))
plt.title('The Distribution of Sentiment')
plt.show()
```

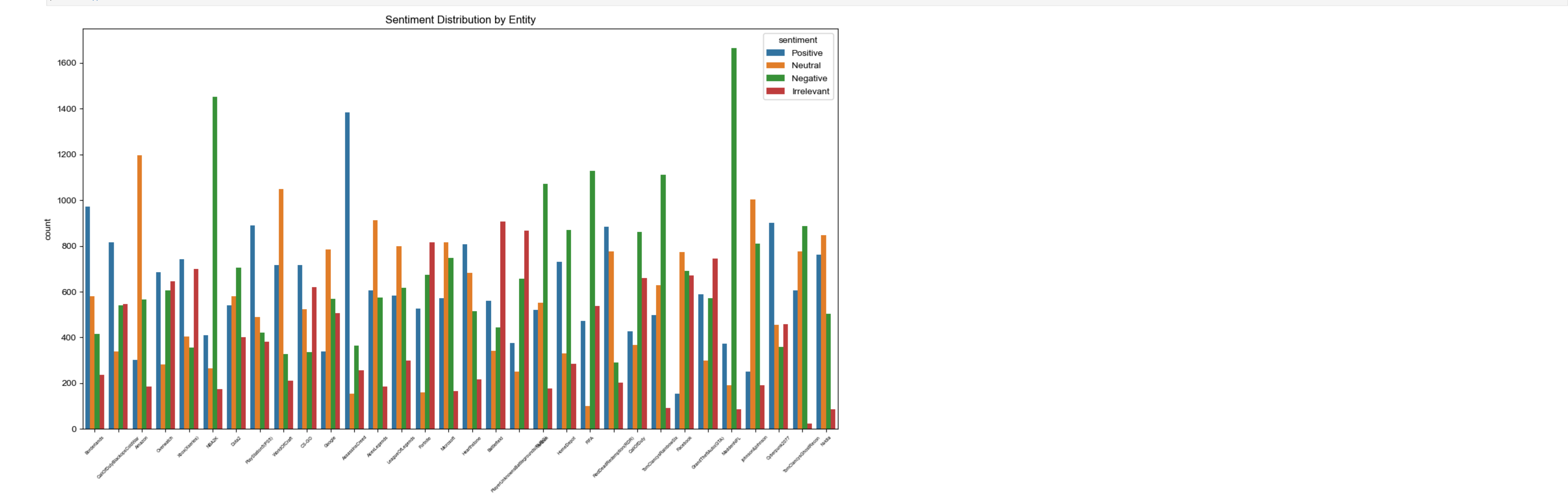


```
In [60]: import seaborn as sns
import matplotlib.pyplot as plt

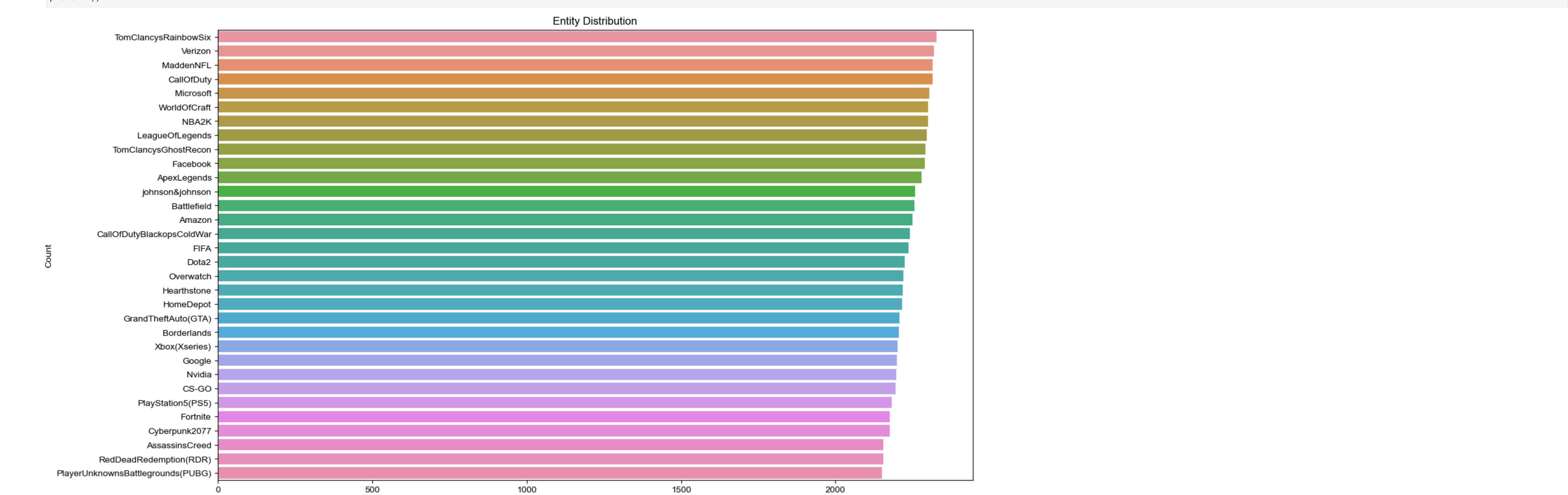
sns.countplot(x='sentiment', data=df)
plt.title('Sentiment Distribution')
plt.show()
```



```
In [61]: plt.figure(figsize=(15, 8))
sns.countplot(x='entity', hue='sentiment', data=df)
plt.title('Sentiment Distribution by Entity')
plt.xticks(rotation=45, fontsize=5)
plt.show()
```



```
In [62]: plt.figure(figsize=(15,9))
sns.barplot(x=df['entity'].value_counts().values,y=df['entity'].value_counts().index)
plt.xlabel('Type')
plt.ylabel('Count')
plt.title('Entity Distribution')
plt.show()
```



```
In [63]: average_sentiment_by_entity = df.groupby('entity')['sentiment'].value_counts(normalize=True).unstack()
average_sentiment_by_entity.plot(kind='bar', stacked=True, figsize=(12, 6))
plt.title('Average Sentiment by Entity')
plt.xticks(rotation=45, fontsize=6)
plt.show()
```

Average Sentiment by Entity

