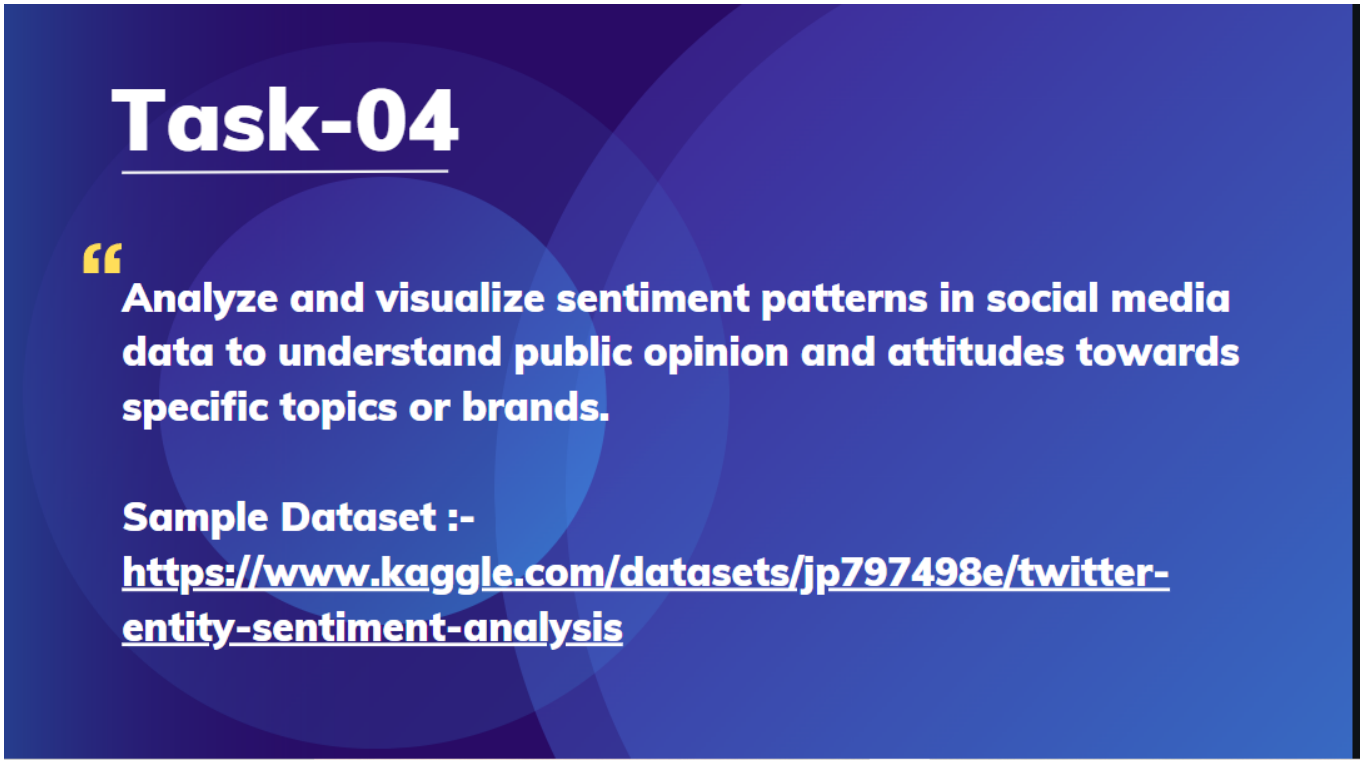



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

```
url= "/content/twitter_training.csv"
```

```
df=pd.read_csv(url)
```

```
df.head()
```

	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...

```
df.rename(columns={"2401":"id","Borderlands":"topic","Positive":"sentiment","im getting on borderlands and i will murder you all ,":"twe
```

```
df.tail()
```

	id	topic	sentiment	tweet
68426	3717	Cyberpunk2077	Positive	LETS N GOOOOOOOOOO
68427	3717	Cyberpunk2077	Positive	she LETS IN FUCKING OF GOOOOOOOOOO
68428	3717	Cyberpunk2077	Positive	LETS FUCKING LI
68429	3718	Cyberpunk2077	Positive	I can’t wait for this to come out
68430	3718	Cyberpunk2077	Positive	I can't wait for

```
df.shape
```

```
(68431, 4)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 68431 entries, 0 to 68430
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   id         68431 non-null  int64 
 1   topic      68431 non-null  object
 2   sentiment  68431 non-null  object
 3   tweet      67829 non-null  object
dtypes: int64(1), object(3)
memory usage: 2.1+ MB
```

```
df.isnull().sum()
```

```
id             0
topic          0
sentiment      0
tweet        602
dtype: int64
```

```
df["sentiment"].value_counts()
```

```
Negative      20862
Positive      18727
Neutral       16320
Irrelevant    12522
Name: sentiment, dtype: int64
```

```
df["topic"].value_counts()
```

```
Microsoft                           2400
TomClancysRainbowSix                2400
MaddenNFL                           2400
LeagueOfLegends                     2394
CallOfDuty                          2394
Verizon                             2382
ApexLegends                         2376
CallOfDutyBlackopsColdWar           2376
Facebook                            2370
Dota2                               2364
WorldOfCraft                        2364
NBA2K                               2352
Battlefield                         2346
FIFA                                2340
Xbox(Xseries)                       2334
Overwatch                           2334
johnson&johnson                     2328
Amazon                              2316
HomeDepot                           2310
PlayStation5(PS5)                   2310
GrandTheftAuto(GTA)                 2304
CS-GO                               2304
Google                              2298
Hearthstone                         2298
Borderlands                         2285
PlayerUnknownsBattlegrounds(PUBG)   2274
Fortnite                            2274
RedDeadRedemption(RDR)              2262
AssassinsCreed                      2244
Cyberpunk2077                        698
Name: topic, dtype: int64
```

```
df["sentiment"].nunique()
```

```
4
```

```
df["id"].nunique(),df.shape
```

```
(11406, (68431, 4))
```

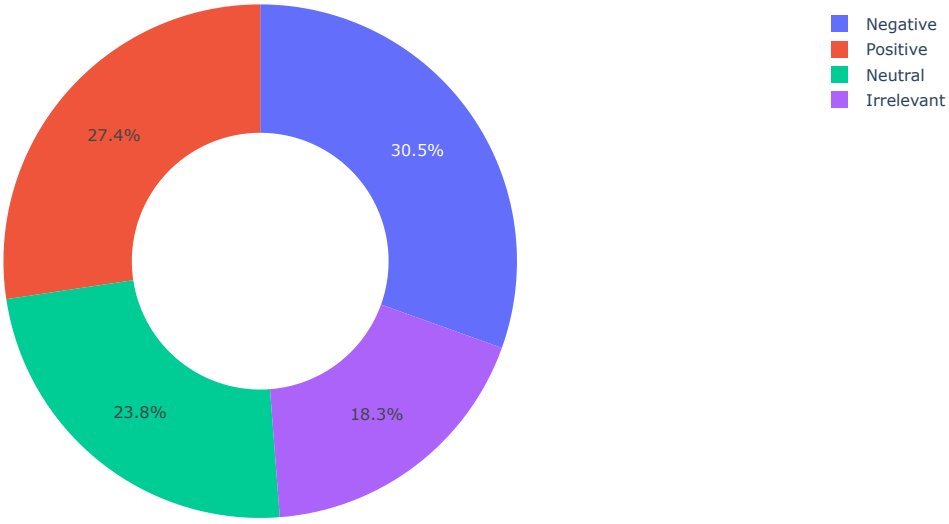
```
df.pivot_table(columns="sentiment",index="topic",aggfunc="count")
```

	id				tweet
sentiment	Irrelevant	Negative	Neutral	Positive	Irreleva
topic					
Amazon	192	576	1236	312	1
ApexLegends	192	600	942	642	1
AssassinsCreed	264	378	156	1446	2
Battlefield	918	474	360	594	9
Borderlands	240	426	600	1019	2
CS-GO	636	348	552	768	6
CallOfDuty	672	894	378	450	6
CallOfDutyBlackopsColdWar	576	576	360	864	5
Cyberpunk2077	126	132	144	296	1
Dota2	426	768	588	582	4
FIFA	558	1176	102	504	5
Facebook	690	720	786	174	6
Fortnite	840	702	168	564	8
Google	522	594	822	360	5
GrandTheftAuto(GTA)	762	594	324	624	7
Hearthstone	228	528	708	834	2
HomeDepot	294	900	336	780	2
LeagueOfLegends	312	642	822	618	3
MaddenNFL	90	1710	204	396	
Microsoft	174	774	846	606	1
NBA2K	180	1476	270	426	1
Overwatch	672	630	300	732	6
PlayStation5(PS5)	396	456	516	942	3
PlayerUnknownsBattlegrounds(PUBG)	906	696	264	408	8
RedDeadRedemption(RDR)	210	306	816	930	2
TomClancysRainbowSix	96	1122	654	528	
Verizon	186	1098	570	528	1
WorldOfCraft	216	342	1068	738	2
Xbox(Xseries)	750	378	414	792	7
johnson&johnson	198	846	1014	270	1

```
df.head()
```

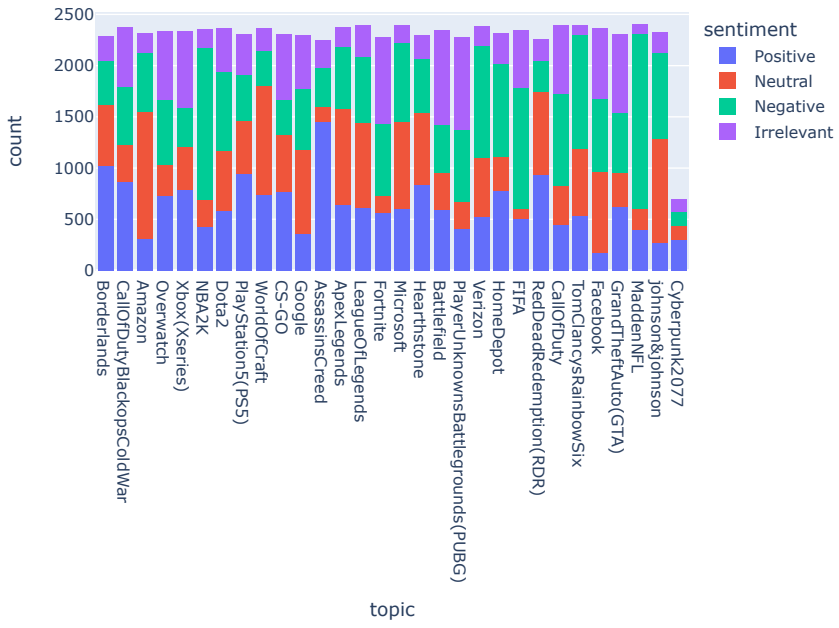
	id	topic	sentiment	tweet	polarity_score
0	2401	Borderlands	Positive	I am coming to the borders and I will kill you...	0.0
1	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...	0.0
2	2401	Borderlands	Positive	im coming on borderlands and i will murder you...	0.0
3	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...	0.0
4	2401	Borderlands	Positive	im getting into borderlands and i can murder y...	0.0

```
px.pie(df,"sentiment",hole=0.5,)
```




```
?px.histogram
```

```
px.histogram(df,x="topic",color="sentiment")
```



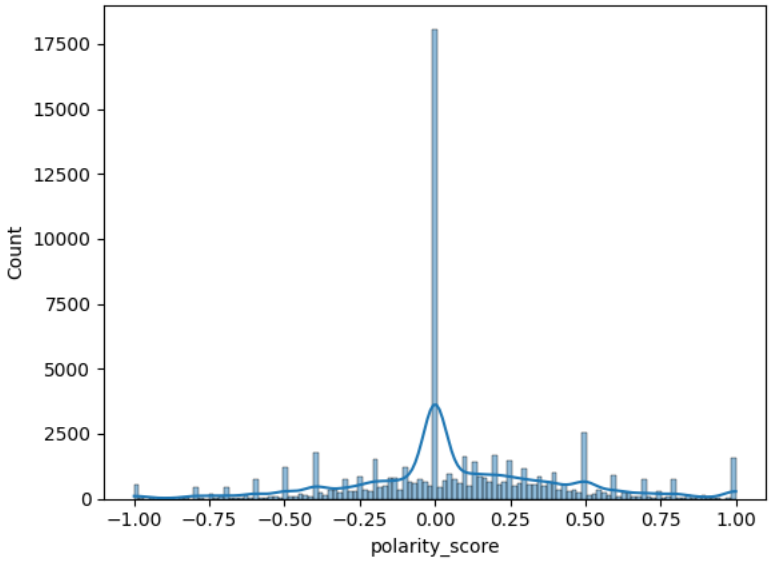

```
from textblob import TextBlob


def get_polarity(tweet):
    if isinstance(tweet, str):
        return TextBlob(tweet).sentiment.polarity
    else:
        return 0.0

df["polarity_score"] = df["tweet"].apply(get_polarity)
```

```
sns.histplot(data=df,x="polarity_score",kde=True)
```

```
<Axes: xlabel='polarity_score', ylabel='Count'>
```




```
def get_sentiment_label(polarity_score):
    if polarity_score > 0:
        return "Positive"
    elif polarity_score < 0:
        return "Negative"
    else:
        return "Neutral"

# Apply the get_sentiment_label function to create a new column "sentiment_label"
df['sentiment_by_polarity'] = df['polarity_score'].apply(get_sentiment_label)
```

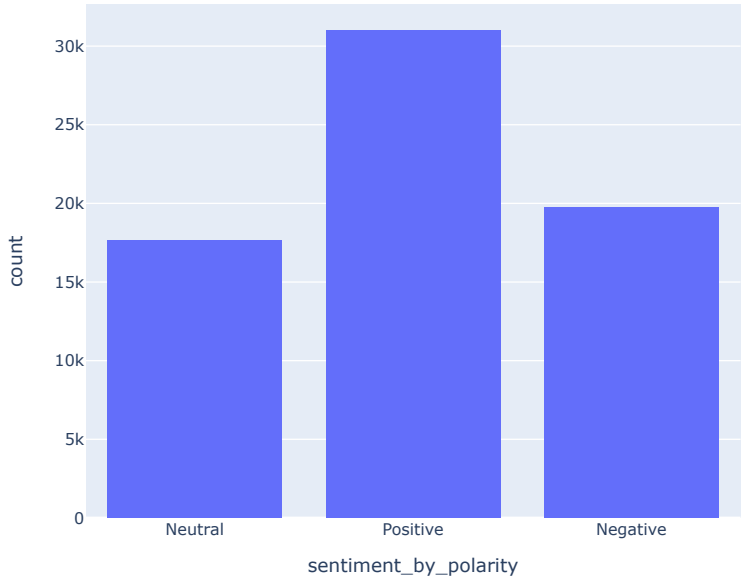
```
df.head()
```

	id	topic	sentiment	tweet	polarity_score	sentiment_label	sentiment
0	2401	Borderlands	Positive	I am coming to the borders and I will kill you...	0.0	Neutral	
1	2401	Borderlands	Positive	im getting on	0.0	Neutral	

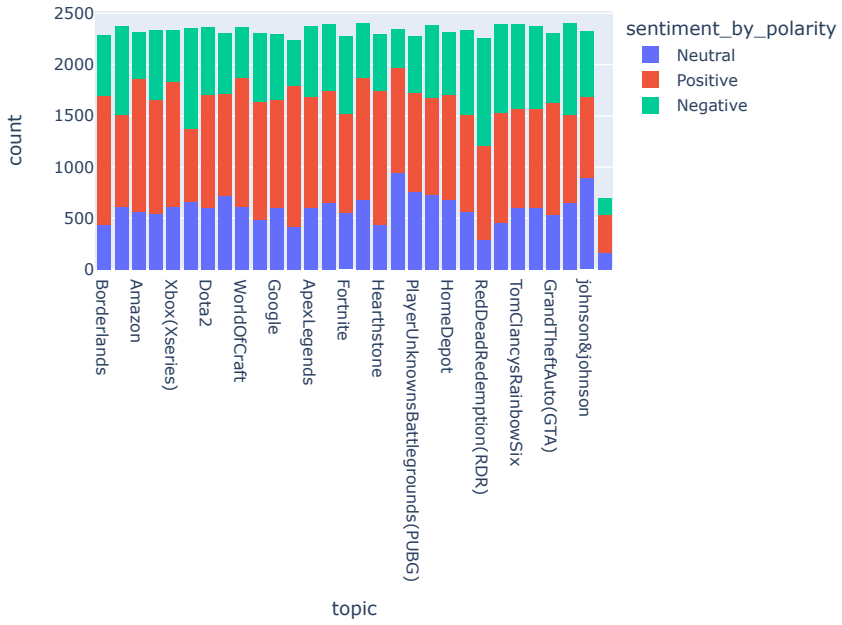
```
df["sentiment_by_polarity"].value_counts(),df["sentiment"].value_counts()
```

```
(Positive    31020
 Negative    19744
 Neutral     17667
 Name: sentiment_by_polarity, dtype: int64,
 Negative      20862
 Positive      18727
 Neutral       16320
 Irrelevant    12522
 Name: sentiment, dtype: int64)
```

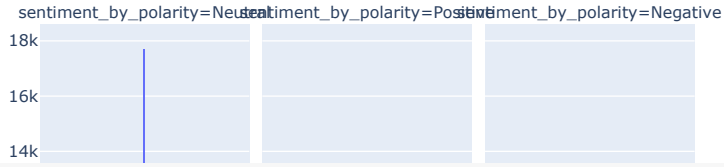
```
px.histogram(df,x="sentiment_by_polarity")
```




```
px.histogram(df,x="topic",color="sentiment_by_polarity")
```




```
px.histogram(df,x="polarity_score",facet_col="sentiment_by_polarity")
```

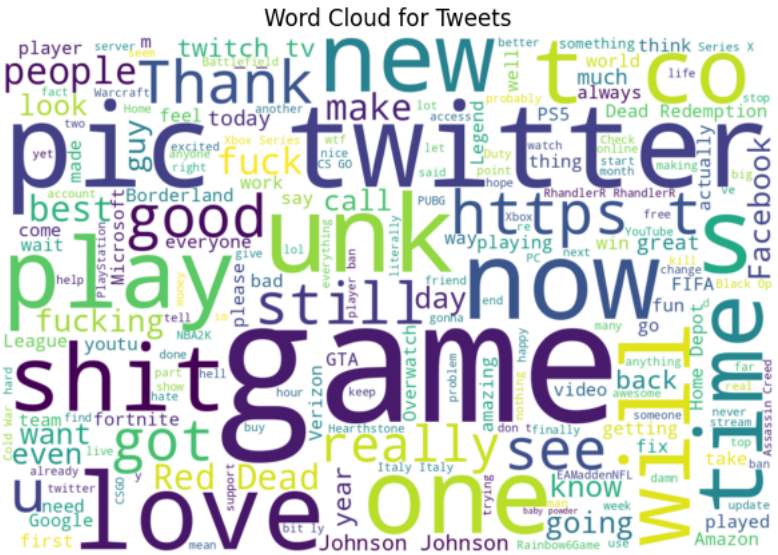



```
from wordcloud import WordCloud


# Join all the tweets into a single string
text = ' '.join(df['tweet'].dropna())

# Generate the word cloud
wordcloud = WordCloud(width=1200, height=800, background_color='white').generate(text)

# Plot the word cloud
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.title('Word Cloud for Tweets')
plt.axis('off')
plt.show()
```




```
fig=px.histogram(df,x="sentiment",y="polarity_score",animation_frame="topic",histfunc="avg",color="sentiment")#change the histfunc to coun
fig.update_traces(hovertemplate='Sentiment: %{x}<br>Polarity Score: %{y}<br>Count of Tweets: %{y}')
fig.update_traces(texttemplate='%{y}')

fig.show()
```