In [1]:

```
import pandas as pd
import numpy as np
from nltk.tokenize import sent_tokenize, word_tokenize
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.datasets import fetch_20newsgroups
from nltk.corpus import stopwords
import string
from nltk import pos_tag
from nltk.stem import WordNetLemmatizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import preprocessing
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:

```
import nltk
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\ADMIN\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\stopwords.zip.
```

Out[2]:

```
True
```

In [3]:

```
data = pd.read_csv(r"C:\Users\ADMIN\Downloads\twitter_training.csv.zip")
v_data = pd.read_csv(r"C:\Users\ADMIN\Downloads\twitter_validation.csv")
```

In [4]:
```
data
```

Out[4]:
```
       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...
  ...   ...          ...       ...                                                     ...
74676  9200       Nvidia  Positive        Just realized that the Windows partition of my...
74677  9200       Nvidia  Positive        Just realized that my Mac window partition is ...
74678  9200       Nvidia  Positive       Just realized the windows partition of my Mac ...
74679  9200       Nvidia  Positive       Just realized between the windows partition of...
74680  9200       Nvidia  Positive          Just like the windows partition of my Mac is l...

74681 rows × 4 columns
```

In [5]:
```
v_data
```

Out[5]:
```
                                          I mentioned on Facebook that I was struggling for
                                        motivation to go for a run the other day, which has been
     3364           Facebook  Irrelevant   translated by Tom’s great auntie as ‘Hayley can’t get
                                          out of bed’ and told to his grandma, who now thinks I’m
                                                                 a lazy, terrible person 🤣
  0   352             Amazon     Neutral        BBC News - Amazon boss Jeff Bezos rejects clai...
  1  8312          Microsoft    Negative      @Microsoft Why do I pay for WORD when it funct...
  2  4371             CS-GO    Negative        CSGO matchmaking is so full of closet hacking,...
  3  4433             Google     Neutral        Now the President is slapping Americans in the...
  4  6273               FIFA    Negative      Hi @EAHelp I’ve had Madeleine McCann in my cel...
...   ...                ...         ...                                                  ...
994  4891  GrandTheftAuto(GTA) Irrelevant       ⭐ Toronto is the arts and culture capital of ...
995  4359             CS-GO   Irrelevant     tHIS IS ACTUALLY A GOOD MOVE TOT BRING MORE
                                                                                        VI...
996  2652        Borderlands    Positive        Today sucked so it’s time to drink wine n play...
997  8069          Microsoft    Positive      Bought a fraction of Microsoft today. Small wins.
998  6960   johnson&johnson     Neutral      Johnson & Johnson to stop selling talc baby po...

999 rows × 4 columns
```

In [6]:
```
data.columns = ['id', 'game', 'sentiment', 'text']
v_data.columns = ['id', 'game', 'sentiment', 'text']
```

In [7]:
```
data
```

Out[7]:

	id	game	sentiment	text
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...
...	...	...	...	...
74676	9200	Nvidia	Positive	Just realized that the Windows partition of my...
74677	9200	Nvidia	Positive	Just realized that my Mac window partition is ...
74678	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
74679	9200	Nvidia	Positive	Just realized between the windows partition of...
74680	9200	Nvidia	Positive	Just like the windows partition of my Mac is l...

74681 rows × 4 columns

In [8]:
```
v_data
```

Out[8]:

	id	game	sentiment	text
0	352	Amazon	Neutral	BBC News - Amazon boss Jeff Bezos rejects clai...
1	8312	Microsoft	Negative	@Microsoft Why do I pay for WORD when it funct...
2	4371	CS-GO	Negative	CSGO matchmaking is so full of closet hacking,...
3	4433	Google	Neutral	Now the President is slapping Americans in the...
4	6273	FIFA	Negative	Hi @EAHelp I’ve had Madeleine McCann in my cel...
...	...	...	...	...
994	4891	GrandTheftAuto(GTA)	Irrelevant	⭐️ Toronto is the arts and culture capital of ...
995	4359	CS-GO	Irrelevant	tHIS IS ACTUALLY A GOOD MOVE TOT BRING MORE VI...
996	2652	Borderlands	Positive	Today sucked so it’s time to drink wine n play...
997	8069	Microsoft	Positive	Bought a fraction of Microsoft today. Small wins.
998	6960	johnson&johnson	Neutral	Johnson & Johnson to stop selling talc baby po...

999 rows × 4 columns

In [9]:
```
data.shape
```

Out[9]:
```
(74681, 4)
```

In [10]:
```
data.columns
```

Out[10]:
```
Index(['id', 'game', 'sentiment', 'text'], dtype='object')
```

In [11]:

```
data.describe(include='all')
```

Out[11]:

	id	game	sentiment	text
count	74681.000000	74681	74681	73995
unique	NaN	32	4	69490
top	NaN	TomClancysRainbowSix	Negative	
freq	NaN	2400	22542	172
mean	6432.640149	NaN	NaN	NaN
std	3740.423819	NaN	NaN	NaN
min	1.000000	NaN	NaN	NaN
25%	3195.000000	NaN	NaN	NaN
50%	6422.000000	NaN	NaN	NaN
75%	9601.000000	NaN	NaN	NaN
max	13200.000000	NaN	NaN	NaN

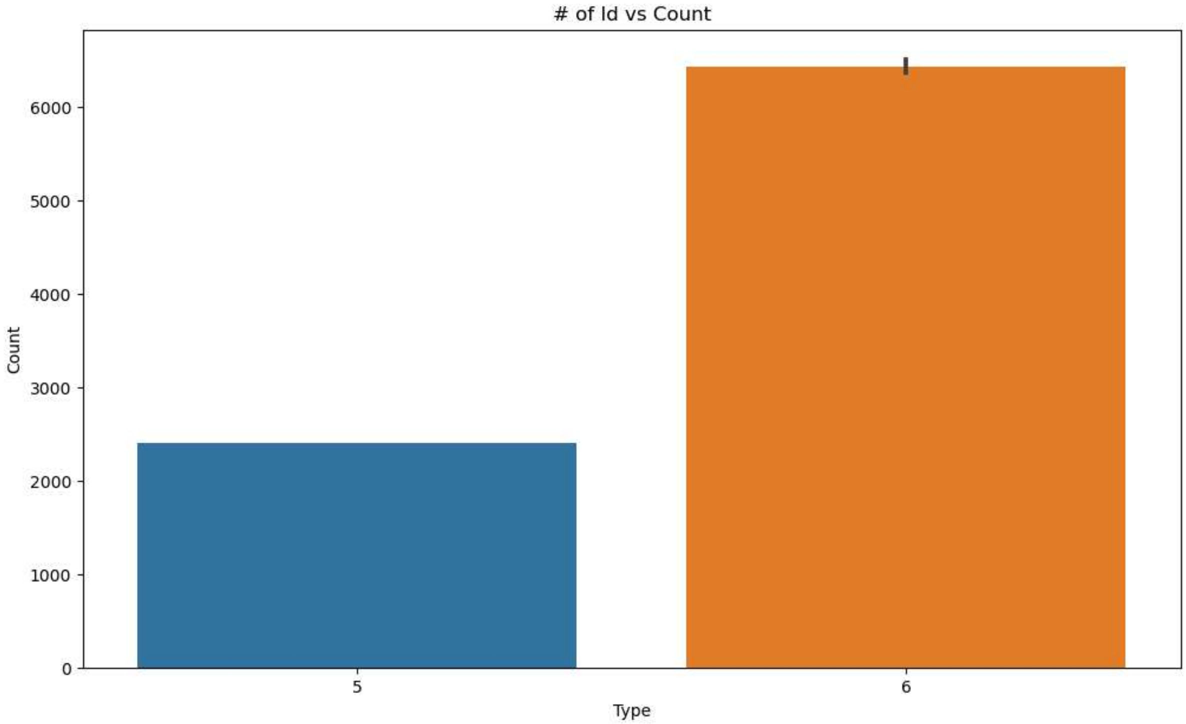
In [12]:

```
id_types = data['id'].value_counts()
id_types
```

Out[12]:

```
id
5203    6
6164    6
6141    6
6142    6
6143    6
       ..
4678    6
4679    6
4680    6
4681    6
2401    5
Name: count, Length: 12447, dtype: int64
```

```
In [13]: plt.figure(figsize=(12,7))
         sns.barplot(y=id_types.index, x=id_types.values)
         plt.xlabel('Type')
         plt.ylabel('Count')
         plt.title('# of Id vs Count')
         plt.show()
```



In [14]:

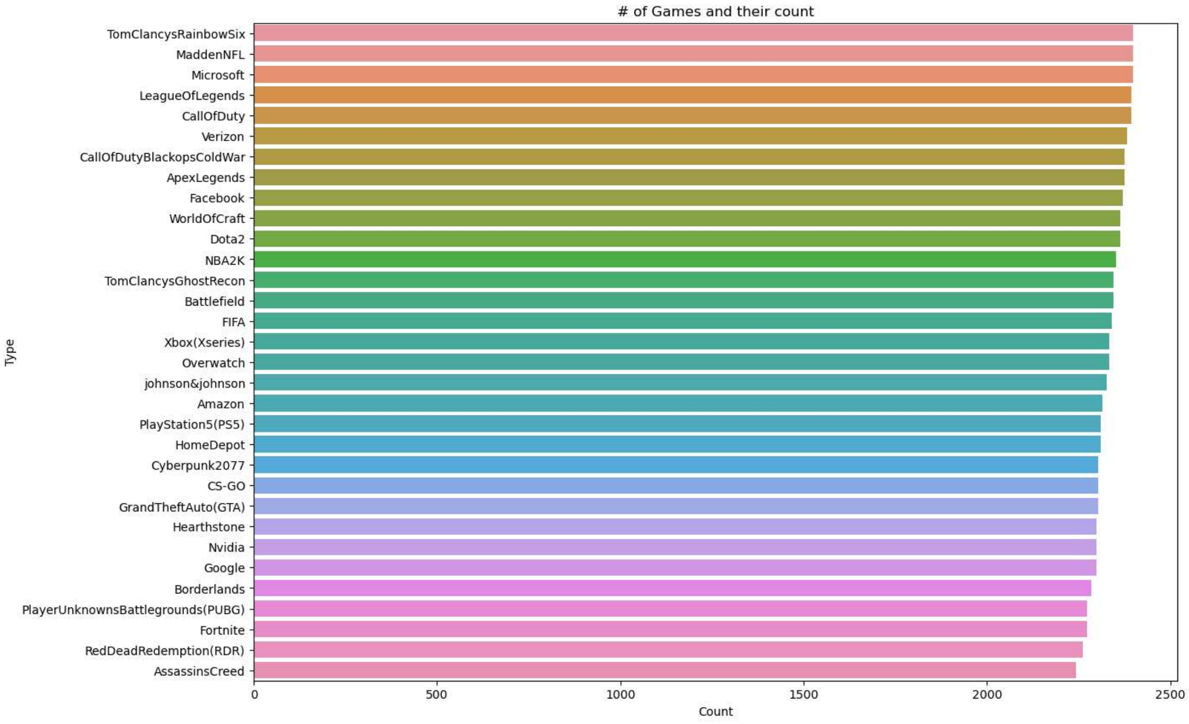
```
game_types = data['game'].value_counts()
game_types
```

Out[14]:

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

In [15]:

```
plt.figure(figsize=(14,10))

sns.barplot(x=game_types.values,y=game_types.index)
plt.title('# of Games and their count')
plt.ylabel('Type')
plt.xlabel('Count')

plt.show()
```



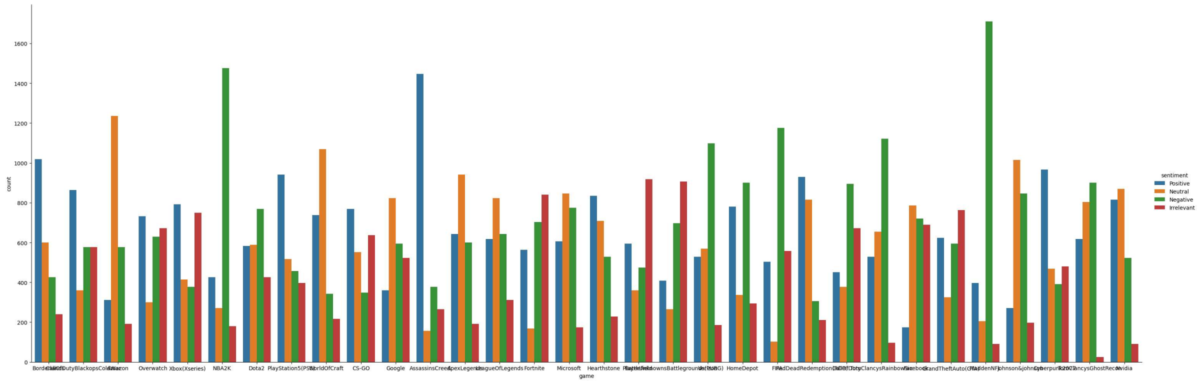
In [16]:

```
sns.catplot(x="game",hue="sentiment", kind="count",height=10,aspect=3, data=dat
```

```
C:\Users\ADMIN\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarni
ng: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

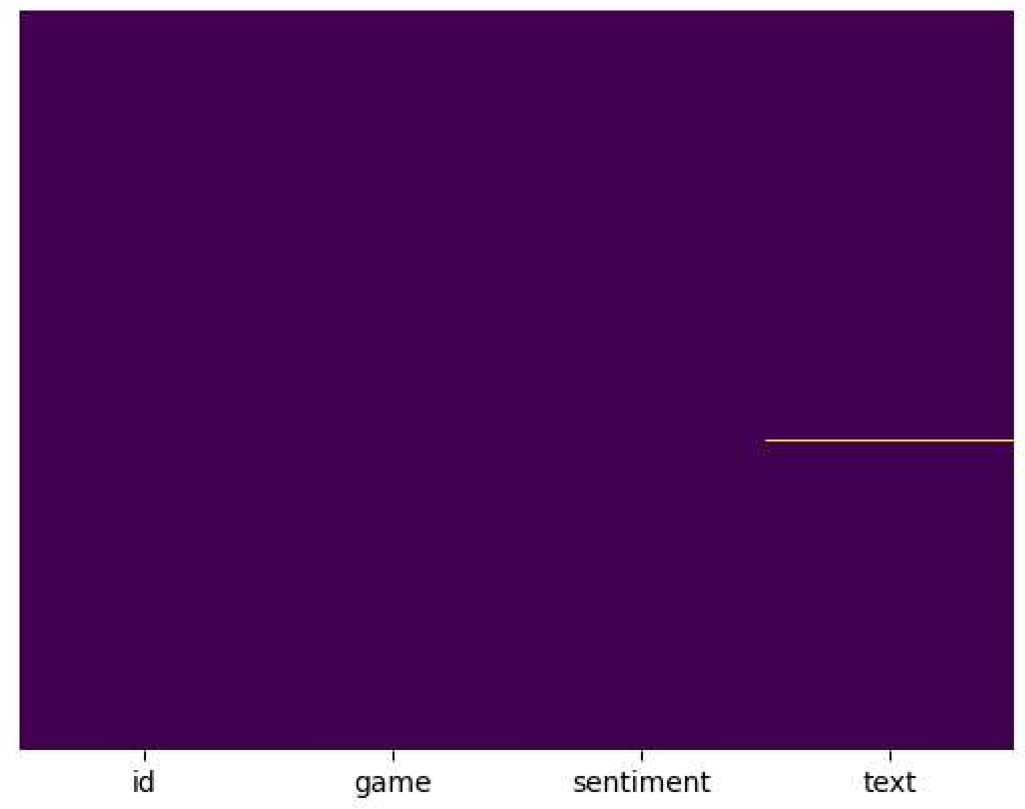
Out[16]:

```
<seaborn.axisgrid.FacetGrid at 0x1ba5dae63d0>
```



In [17]:
```
sns.heatmap(data.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

Out[17]:
```
<Axes: >
```



In [18]:
```
total_null=data.isnull().sum().sort_values(ascending=False)
percent = ((data.isnull().sum()/data.isnull().count())*100).sort_values(ascendi
print("Total records = ", data.shape[0])
missing_data = pd.concat([total_null,percent.round(2)],axis=1,keys=['Total Miss
missing_data.head(10)
```

```
Total records =  74681
```

Out[18]:

	Total Missing	In Percent
text	686	0.92
id	0	0.00
game	0	0.00
sentiment	0	0.00

In [19]:

```
data.dropna(subset=['text'],inplace=True)

total_null=data.isnull().sum().sort_values(ascending=False)
percent = ((data.isnull().sum()/data.isnull().count())*100).sort_values(ascendi
print("Total records = ", data.shape[0])
missing_data = pd.concat([total_null,percent.round(2)],axis=1,keys=['Total Miss
missing_data.head(10)
```

```
Total records =  73995
```

Out[19]:

	Total Missing	In Percent
id	0	0.0
game	0	0.0
sentiment	0	0.0
text	0	0.0

In [20]:

```
train0=data[data['sentiment']=="Negative"]
train1=data[data['sentiment']=="Positive"]
train2=data[data['sentiment']=="Irrelevant"]
train3=data[data['sentiment']=="Neutral"]
```

In [21]:

```
train0.shape, train1.shape, train2.shape, train3.shape
```

Out[21]:

```
((22358, 4), (20654, 4), (12875, 4), (18108, 4))
```

In [22]:

```
train0=train0[:int(train0.shape[0]/12)]
train1=train1[:int(train1.shape[0]/12)]
train2=train2[:int(train2.shape[0]/12)]
train3=train3[:int(train3.shape[0]/12)]
```

In [23]:

```
train0.shape, train1.shape, train2.shape, train3.shape
```

Out[23]:

```
((1863, 4), (1721, 4), (1072, 4), (1509, 4))
```

In [24]:
```
data=pd.concat([train0,train1,train2,train3],axis=0)
data
```

Out[24]:

	id	game	sentiment	text
23	2405	Borderlands	Negative	the biggest dissappoinment in my life came out...
24	2405	Borderlands	Negative	The biggest disappointment of my life came a y...
25	2405	Borderlands	Negative	The biggest disappointment of my life came a y...
26	2405	Borderlands	Negative	the biggest dissappoinment in my life coming o...
27	2405	Borderlands	Negative	For the biggest male dissappoinment in my life...
...	...	...	...	...
5603	165	Amazon	Neutral	An amazing read aloud book for you and your ch...
5604	165	Amazon	Neutral	An amazing reading book for you and your child...
5605	165	Amazon	Neutral	An amazing book to read aloud for you and your...
5606	165	Amazon	Neutral	An amazing read aloud book for you and your ch...
5607	165	Amazon	Neutral	and An amazing read aloud book for you and you...

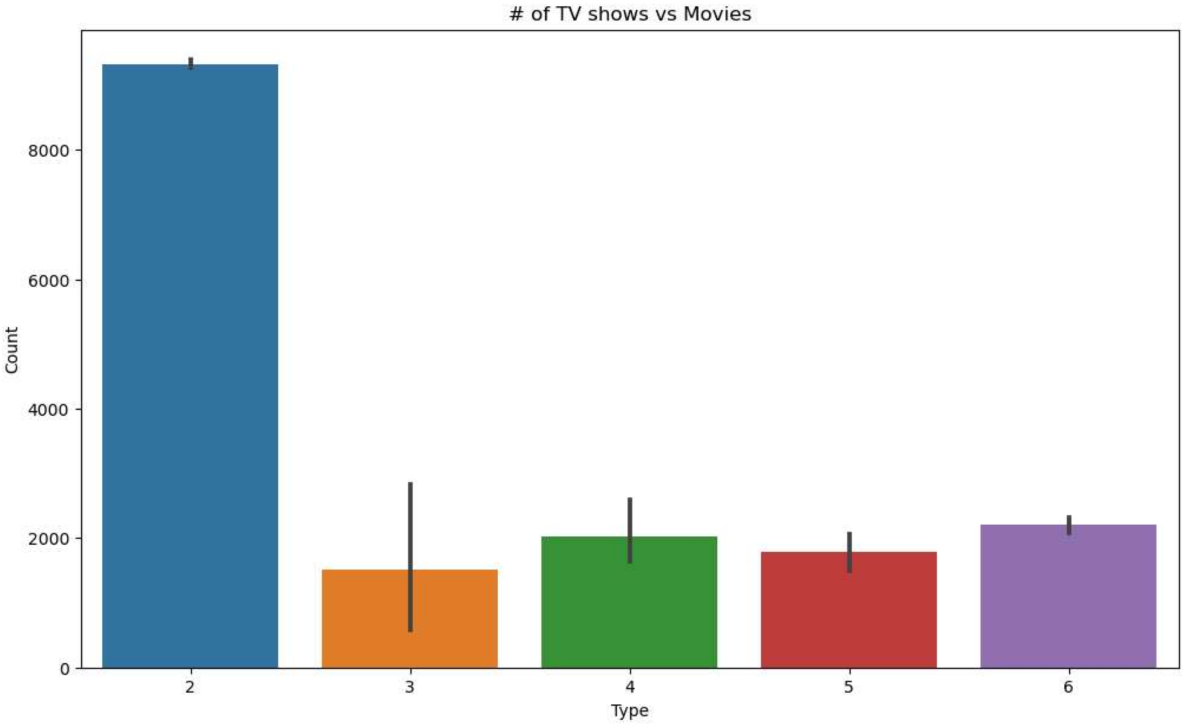
6165 rows × 4 columns

In [25]:
```
id_types = data['id'].value_counts()
id_types
```

Out[25]:
```
id
2405    6
1810    6
1748    6
1754    6
1760    6
       ..
1602    3
1880    3
333     3
9388    2
9267    2
Name: count, Length: 1040, dtype: int64
```

In [26]:

```
plt.figure(figsize=(12,7))
sns.barplot(x=id_types.values,y=id_types.index)

plt.xlabel('Type')
plt.ylabel('Count')
plt.title('# of TV shows vs Movies')
plt.show()
```



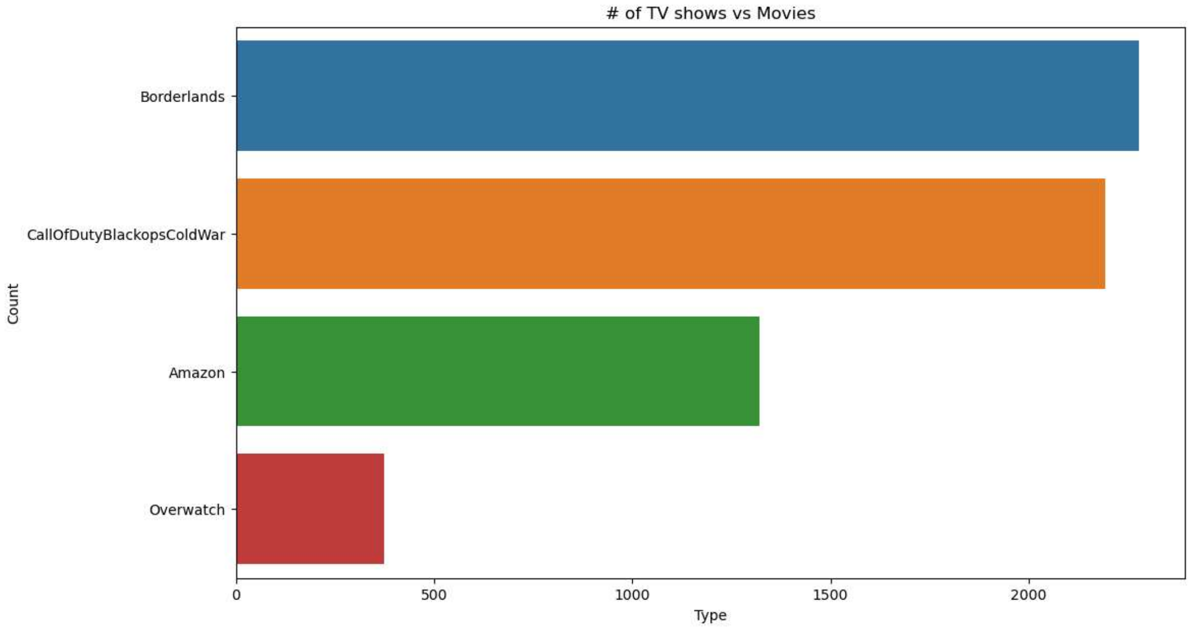
In [27]:

```
game_types = data['game'].value_counts()
game_types
```

Out[27]:

```
game
Borderlands                 2279
CallOfDutyBlackopsColdWar   2192
Amazon                      1321
Overwatch                    373
Name: count, dtype: int64
```

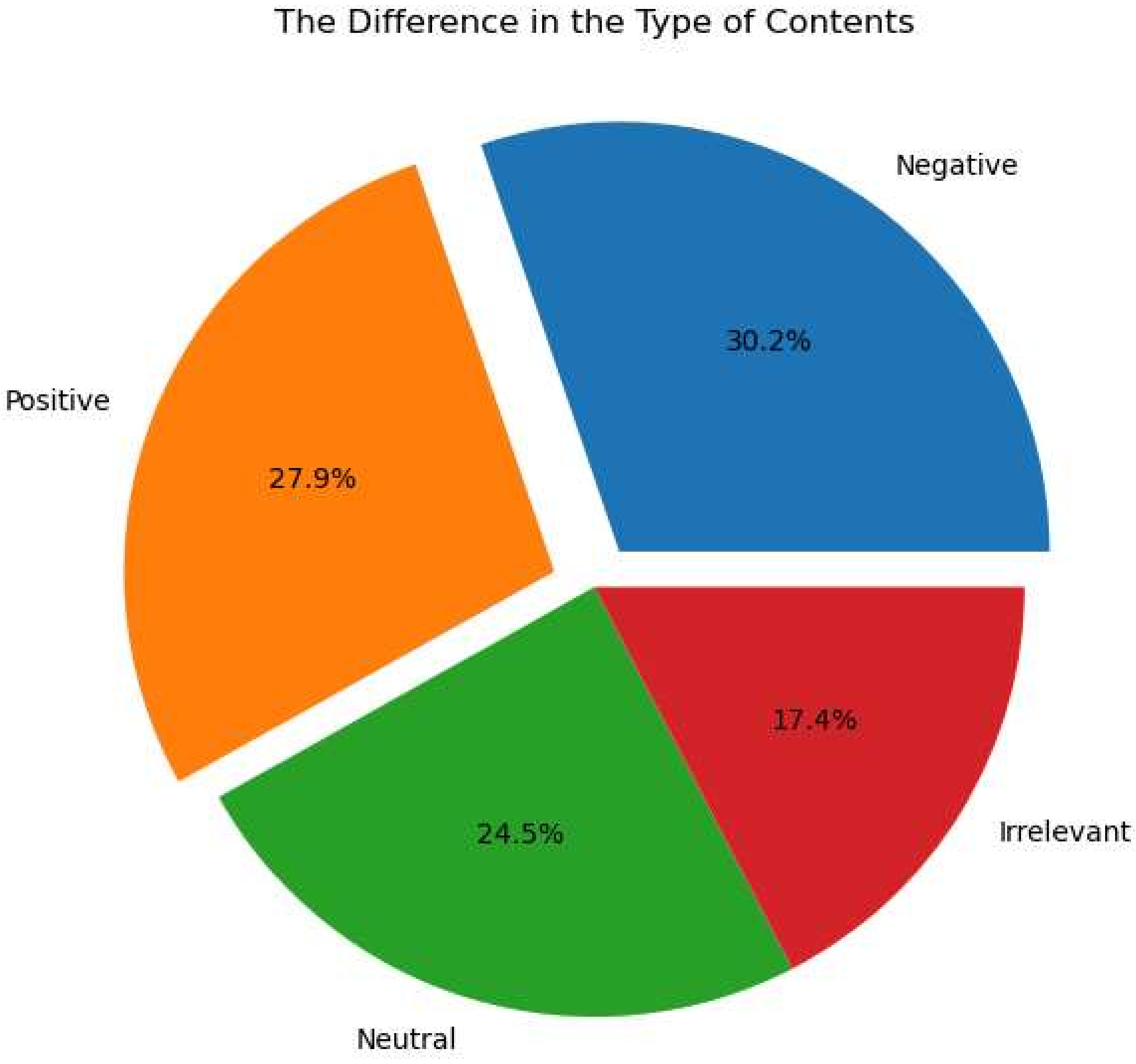
In [28]:
```
plt.figure(figsize=(12,7))
sns.barplot(x=game_types.values,y=game_types.index)

plt.xlabel('Type')
plt.ylabel('Count')
plt.title('# of TV shows vs Movies')
plt.show()
```



In [29]:
```
sentiment_types = data['sentiment'].value_counts()
sentiment_types
```

Out[29]:
```
sentiment
Negative      1863
Positive      1721
Neutral       1509
Irrelevant    1072
Name: count, dtype: int64
```

```
In [30]: plt.figure(figsize=(12,7))
         plt.pie(x=sentiment_types.values, labels=sentiment_types.index, autopct='%.1f%
         plt.title('The Difference in the Type of Contents')
         plt.show()
```



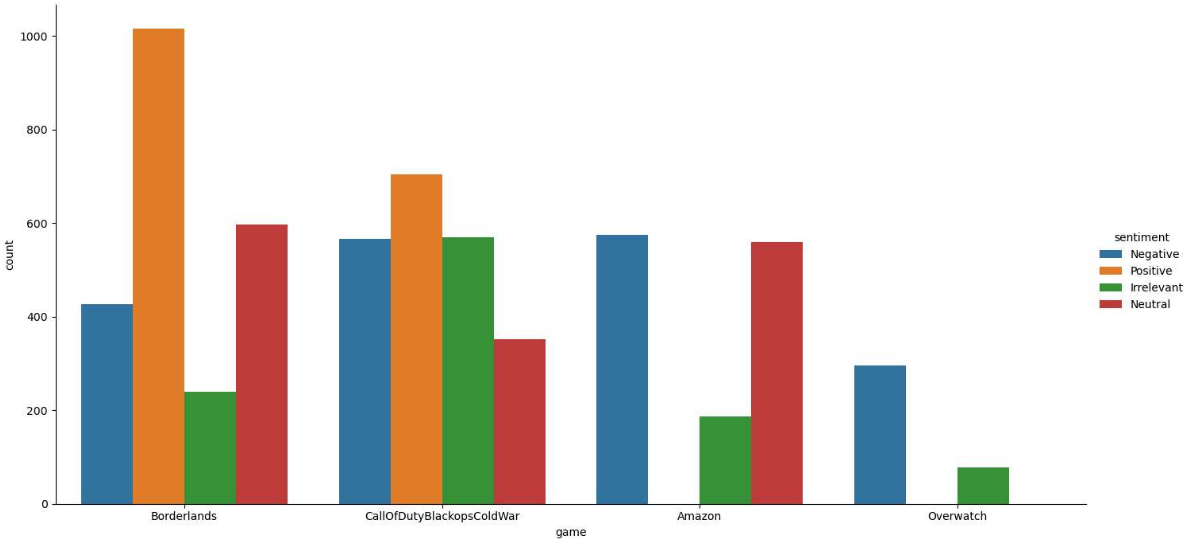
In [31]:

```
sns.catplot(x='game',hue='sentiment',kind='count',height=7,aspect=2,data=data)
```

```
C:\Users\ADMIN\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarni
ng: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

Out[31]:

```
<seaborn.axisgrid.FacetGrid at 0x1ba5dbe3990>
```



In [32]:

```
from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
```

In [33]:

```
data['sentiment']=label_encoder.fit_transform(data['sentiment'])
data['game']=label_encoder.fit_transform(data['game'])
v_data['sentiment']=label_encoder.fit_transform(v_data['sentiment'])
v_data['game']=label_encoder.fit_transform(v_data['game'])
```

In [34]:

```
data = data.drop(['id'],axis=1)

data
```

Out[34]:

	game	sentiment	text
23	1	1	the biggest dissappoinment in my life came out...
24	1	1	The biggest disappointment of my life came a y...
25	1	1	The biggest disappointment of my life came a y...
26	1	1	the biggest dissappoinment in my life coming o...
27	1	1	For the biggest male dissappoinment in my life...
...	...	...	...
5603	0	2	An amazing read aloud book for you and your ch...
5604	0	2	An amazing reading book for you and your child...
5605	0	2	An amazing book to read aloud for you and your...
5606	0	2	An amazing read aloud book for you and your ch...
5607	0	2	and An amazing read aloud book for you and you...

6165 rows × 3 columns

In [35]:

```
data.nunique()
```

Out[35]:

```
game            4
sentiment       4
text         5854
dtype: int64
```

In [36]:

```
v_data.nunique()
```

Out[36]:

```
id           999
game          32
sentiment      4
text         998
dtype: int64
```

In [ ]:

```
```