

```
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\91863\AppData\Roaming\nltk_data...
[nltk_data] Unzipping corpora\stopwords.zip.
```

Out[2]: True

```
In [3]: data = pd.read_csv('twitter_training.csv')
v_data = pd.read_csv('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...
5	2402	Borderlands	Positive	So I spent a few hours making something for fu...
6	2402	Borderlands	Positive	So I spent a couple of hours doing something f...
7	2402	Borderlands	Positive	So I spent a few hours doing something for fun...
8	2402	Borderlands	Positive	So I spent a few hours making something for fu...
9	2402	Borderlands	Positive	2010 So I spent a few hours making something f...
10	2402	Borderlands	Positive	was
11	2403	Borderlands	Neutral	Rock-Hard La Varlope, RARE & POWERFUL, HANDSOM...
12	2403	Borderlands	Neutral	Rock-Hard La Varlope, RARE & POWERFUL, HANDSOM...
13	2403	Borderlands	Neutral	Rock-Hard La Varlope, RARE & POWERFUL, HANDSOM...
14	2403	Borderlands	Neutral	Rock-Hard La Vita, RARE BUT POWERFUL, HANDSOME...
15	2403	Borderlands	Neutral	Live Rock - Hard music La la Varlope, RARE & t...
16	2403	Borderlands	Neutral	I-Hard like me, RARE LONDON DE, HANDSOME 2011,...
17	2404	Borderlands	Positive	that was the first borderlands session in a lo...
18	2404	Borderlands	Positive	this was the first Borderlands session in a lo...
19	2404	Borderlands	Positive	that was the first borderlands session in a lo...
20	2404	Borderlands	Positive	that was the first borderlands session in a lo...
21	2404	Borderlands	Positive	that I was the first real borderlands session ...
22	2404	Borderlands	Positive	that was the first borderlands session in a ho...
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...
28	2405	Borderlands	Negative	the biggest dissappoinment in my life came bac...
29	2406	Borderlands	Positive	WE FINISHED BORDERLANDS 3 FINALLY YAS! Thank y...
...	...	...	...	...
74651	9196	Nvidia	Negative	Cheap doesn't mean better btw! . . techsall.co...
74652	9196	Nvidia	Negative	Cheap does not mean better by the way!.. techs...
74653	9196	Nvidia	Negative	??...
74654	9196	Nvidia	Negative	Cheap will mean better<unk>!.. techsall.com/go...
74655	9196	Nvidia	Negative	... Cheap doesn't t mean anyone better btw!.. ...
74656	9196	Nvidia	Negative	really doesn't look bad btw!.. net.com/google-...
74657	9197	Nvidia	Neutral	Nvidia doesn't want to give up its 2017 'crypt...

	2401	Borderlands	Positive	im getting on borderlands and i will murder you all ,
74658	9197	Nvidia	Neutral	Nvidia plans to release its 2017 "Crypto Craze...
74659	9197	Nvidia	Neutral	Nvidia does not want to give up its "cryptoins...
74660	9197	Nvidia	Neutral	Nvidia doesn't intend to give away its 2017 ad...
74661	9197	Nvidia	Neutral	Nvidia therefore doesn ' t want to give up its...
74662	9197	Nvidia	Neutral	is doesn't should I give up its password 'cryp...
74663	9198	Nvidia	Negative	Nvidia really delayed the 3070 2 weeks .
74664	9198	Nvidia	Negative	Nvidia really delayed the 3070 by 2 weeks.
74665	9198	Nvidia	Negative	Nvidia did delay by 3070 2 weeks.
74666	9198	Nvidia	Negative	Nvidia really delayed the 3070 several weeks.
74667	9198	Nvidia	Negative	Nvidia really only delayed the 3070 2 flight w...
74668	9198	Nvidia	Negative	Nvidia really delayed the next 2 weeks.
74669	9199	Nvidia	Positive	Let no elim go unnoticed. . . . NVIDIA Highlig...
74670	9199	Nvidia	Positive	t let Elim go unnoticed.... NVIDIA Highlights ...
74671	9199	Nvidia	Positive	Let no elite go unnoticed... NVIDIA Highlights...
74672	9199	Nvidia	Positive	Let no elim go unnoticed.... NVIDIA Highlights...
74673	9199	Nvidia	Positive	Let a no information elim that go unnoticed.....
74674	9199	Nvidia	Positive	<unk> my elim be no.... NVIDIA Highlights Pict...
74675	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
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 I...

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 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 🤔
3364		Facebook	Irrelevant	
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...
5	7925	MaddenNFL	Positive	Thank you @EAMaddenNFL!! \n\nNew TE Austin Hoo...
6	11332	TomClancysRainbowSix	Positive	Rocket League, Sea of Thieves or Rainbow Six: ...
7	1107	AssassinsCreed	Positive	my ass still knee-deep in Assassins Creed Odys...
8	2069	CallOfDuty	Negative	FIX IT JESUS ! Please FIX IT ! What In the wor...
9	3185	Dota2	Positive	The professional dota 2 scene is fucking explo...
10	1172	AssassinsCreed	Positive	Itching to assassinate \n\n#TCCGif #AssassinsC...
11	11783	Verizon	Negative	@FredTJoseph hey fred, Comcast cut the cable a...
12	4286	CS-GO	Neutral	CSGO Wingman (Im Silver dont bully) twitch.tv/...
13	8431	NBA2K	Negative	@NBA2K game sucks... down by 2 with 38 seconds...
14	9135	Nvidia	Positive	Congrats to the NVIDIA NeMo team for the 1.0.0...
15	4822	GrandTheftAuto(GTA)	Positive	yeah and it's fun
16	3068	Dota2	Negative	fuck my life 😡
17	10537	RedDeadRedemption(RDR)	Positive	happy birthday red dead redemption that shit c...
18	8056	Microsoft	Negative	What does that say about Microsoft hardware & ...
19	2131	CallOfDuty	Negative	The new @CallofDuty for ps5 is 🔥🔥🔥\nOh God 🤔🤔
20	5450	Hearthstone	Neutral	Anyone that plays a bad luck albatross deck in...
21	2286	CallOfDuty	Irrelevant	Call of duty warzone (livestream) w/ subs #War...
22	4038	CS-GO	Negative	Finally played Rainbow Six Siege for the first...

I mentioned on Facebook that I was struggling for motivation to go for a run the other day, which has been 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 🤔

3364		Facebook	Irrelevant	
23	526	ApexLegends	Neutral	Umm @PlayApex when I died it said Bug This pi...
24	8977	Nvidia	Neutral	#gtc20 - nice, motivational, and very accessi...
25	11995	Verizon	Negative	Yo! @Verizon just added a \$120 'fee' to my acc...
26	9449	Overwatch	Irrelevant	They might not be the last team that has to ma...
27	10193	PlayerUnknownsBattlegrounds(PUBG)	Irrelevant	Best squad yet#pubg #pubgmobile #pubgkenya ins...
28	2419	Borderlands	Negative	@Borderlands how do I submit a complaint? Your...
29	8857	Nvidia	Positive	Watching NVIDIA position itself as not just a ...
...	...	...	...	...
969	11957	Verizon	Negative	Old lady threatened me and said she was going ...
970	8318	Microsoft	Negative	Why didn't anyone think of this acronym for BL...
971	3080	Dota2	Positive	FUCKING FINALLY
972	2734	Borderlands	Positive	surprise i still love #borderlands ' jack pic....
973	397	Amazon	Positive	#Amazon Best Seller!\n\nScreen Cleaner Kit - B...
974	12949	Xbox(Xseries)	Irrelevant	Microsoft Apparently Has 16 First-Party Xbox S...
975	2219	CallOfDuty	Positive	this quarantine got me bored af playing call o...
976	4444	Google	Neutral	This is.....odd. @Google suggests the most dead...
977	10128	PlayerUnknownsBattlegrounds(PUBG)	Irrelevant	How Abhijeet Bhayya ur efforts are incredible ...
978	11431	TomClancysRainbowSix	Negative	@Rainbow6Game when will you ban mnk/cronos pla...
979	5495	Hearthstone	Positive	Team 5 what a great media/announcement day. Yo...
980	12689	WorldOfCraft	Positive	@WarcraftDevs this would be incredible!
981	9446	Overwatch	Negative	Removing mcree's noose because ur pussies enit...
982	3108	Dota2	Negative	wtf
983	1373	Battlefield	Negative	I have noticed streamers I watch who are now p...
984	10588	RedDeadRedemption(RDR)	Neutral	@6th__man playing red dead redemption-\n\n"Oh ...

3364		Facebook	Irrelevant	I mentioned on Facebook that I was struggling for motivation to go for a run the other day, which has been 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 🤔
985	3980	Cyberpunk2077	Positive	This looks like a poster for a 90s action movi...
986	4905	GrandTheftAuto(GTA)	Irrelevant	The ultimate pruning machine! @stihlusa GTA 26...
987	7516	LeagueOfLegends	Neutral	♥ Suikoden 2\n 1 Alex Kidd in Miracle World\...
988	5708	HomeDepot	Positive	Thank you to Matching funds Home Depot RW paym...
989	11504	TomClancysRainbowSix	Positive	Tom Clancy's Rainbow Six Siege is a great game...
990	2165	CallOfDuty	Neutral	Late night stream with the boys! Come watch so...
991	2962	Dota2	Positive	guess i'll broke.
992	314	Amazon	Negative	Please explain how this is possible! How can t...
993	9701	PlayStation5(PS5)	Positive	Good on Sony. As much as I want to see the new...
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...
5	2402	Borderlands	Positive	So I spent a few hours making something for fu...
6	2402	Borderlands	Positive	So I spent a couple of hours doing something f...
7	2402	Borderlands	Positive	So I spent a few hours doing something for fun...
8	2402	Borderlands	Positive	So I spent a few hours making something for fu...
9	2402	Borderlands	Positive	2010 So I spent a few hours making something f...
10	2402	Borderlands	Positive	was
11	2403	Borderlands	Neutral	Rock-Hard La Varlope, RARE & POWERFUL, HANDSOM...
12	2403	Borderlands	Neutral	Rock-Hard La Varlope, RARE & POWERFUL, HANDSOM...
13	2403	Borderlands	Neutral	Rock-Hard La Varlope, RARE & POWERFUL, HANDSOM...
14	2403	Borderlands	Neutral	Rock-Hard La Vita, RARE BUT POWERFUL, HANDSOME...
15	2403	Borderlands	Neutral	Live Rock - Hard music La la Varlope, RARE & t...
16	2403	Borderlands	Neutral	I-Hard like me, RARE LONDON DE, HANDSOME 2011,...
17	2404	Borderlands	Positive	that was the first borderlands session in a lo...
18	2404	Borderlands	Positive	this was the first Borderlands session in a lo...
19	2404	Borderlands	Positive	that was the first borderlands session in a lo...
20	2404	Borderlands	Positive	that was the first borderlands session in a lo...
21	2404	Borderlands	Positive	that I was the first real borderlands session ...
22	2404	Borderlands	Positive	that was the first borderlands session in a ho...
23	2405	Borderlands	Negative	the biggest dissappointment 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 dissappointment in my life coming o...
27	2405	Borderlands	Negative	For the biggest male dissappointment in my life...
28	2405	Borderlands	Negative	the biggest dissappointment in my life came bac...
29	2406	Borderlands	Positive	WE FINISHED BORDERLANDS 3 FINALLY YAS! Thank y...
...	...	...	...	...
74651	9196	Nvidia	Negative	Cheap doesn't mean better btw! . . techsall.co...
74652	9196	Nvidia	Negative	Cheap does not mean better by the way!.. techs...
74653	9196	Nvidia	Negative	??...
74654	9196	Nvidia	Negative	Cheap will mean better<unk>!.. techsall.com/go...
74655	9196	Nvidia	Negative	... Cheap doesn't t mean anyone better btw!.. ...
74656	9196	Nvidia	Negative	really doesn't look bad btw!.. net.com/google-...
74657	9197	Nvidia	Neutral	Nvidia doesn't want to give up its 2017 'crypt...

	id	game	sentiment	text
74658	9197	Nvidia	Neutral	Nvidia plans to release its 2017 "Crypto Craze...
74659	9197	Nvidia	Neutral	Nvidia does not want to give up its "cryptoins...
74660	9197	Nvidia	Neutral	Nvidia doesn't intend to give away its 2017 ad...
74661	9197	Nvidia	Neutral	Nvidia therefore doesn ' t want to give up its...
74662	9197	Nvidia	Neutral	is doesn't should I give up its password 'cryp...
74663	9198	Nvidia	Negative	Nvidia really delayed the 3070 2 weeks .
74664	9198	Nvidia	Negative	Nvidia really delayed the 3070 by 2 weeks.
74665	9198	Nvidia	Negative	Nvidia did delay by 3070 2 weeks.
74666	9198	Nvidia	Negative	Nvidia really delayed the 3070 several weeks.
74667	9198	Nvidia	Negative	Nvidia really only delayed the 3070 2 flight w...
74668	9198	Nvidia	Negative	Nvidia really delayed the next 2 weeks.
74669	9199	Nvidia	Positive	Let no elim go unnoticed. . . . NVIDIA Highlig...
74670	9199	Nvidia	Positive	t let Elim go unnoticed.... NVIDIA Highlights ...
74671	9199	Nvidia	Positive	Let no elite go unnoticed... NVIDIA Highlights...
74672	9199	Nvidia	Positive	Let no elim go unnoticed.... NVIDIA Highlights...
74673	9199	Nvidia	Positive	Let a no information elim that go unnoticed.....
74674	9199	Nvidia	Positive	<unk> my elim be no.... NVIDIA Highlights Pict...
74675	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
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 I...

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...
5	7925	MaddenNFL	Positive	Thank you @EAMaddenNFL! \n\nNew TE Austin Hoo...
6	11332	TomClancysRainbowSix	Positive	Rocket League, Sea of Thieves or Rainbow Six: ...
7	1107	AssassinsCreed	Positive	my ass still knee-deep in Assassins Creed Odys...
8	2069	CallOfDuty	Negative	FIX IT JESUS ! Please FIX IT ! What In the wor...
9	3185	Dota2	Positive	The professional dota 2 scene is fucking explo...
10	1172	AssassinsCreed	Positive	Itching to assassinate \n\n#TCCGif #AssassinsC...
11	11783	Verizon	Negative	@FredTJoseph hey fred, Comcast cut the cable a...
12	4286	CS-GO	Neutral	CSGO Wingman (Im Silver dont bully) twitch.tv/...
13	8431	NBA2K	Negative	@NBA2K game sucks... down by 2 with 38 seconds...
14	9135	Nvidia	Positive	Congrats to the NVIDIA NeMo team for the 1.0.0...
15	4822	GrandTheftAuto(GTA)	Positive	yeah and it's fun
16	3068	Dota2	Negative	fuck my life 😞
17	10537	RedDeadRedemption(RDR)	Positive	happy birthday red dead redemption that shit c...
18	8056	Microsoft	Negative	What does that say about Microsoft hardware & ...
19	2131	CallOfDuty	Negative	The new @CallofDuty for ps5 is 🔥 🔥🔥🔥\nOh God 🤔🤔
20	5450	Hearthstone	Neutral	Anyone that plays a bad luck albatross deck in...
21	2286	CallOfDuty	Irrelevant	Call of duty warzone (livestream) w/ subs #War...
22	4038	CS-GO	Negative	Finally played Rainbow Six Siege for the first...
23	526	ApexLegends	Neutral	Umm @PlayApex when I died it said Bug This pi...
24	8977	Nvidia	Neutral	#gtc20 - nice, motivational, and very accessi...

	id	game	sentiment	text
25	11995	Verizon	Negative	Yo! @Verizon just added a \$120 'fee' to my acc...
26	9449	Overwatch	Irrelevant	They might not be the last team that has to ma...
27	10193	PlayerUnknownsBattlegrounds(PUBG)	Irrelevant	Best squad yet#pubg #pubgmobile #pubgkenya ins...
28	2419	Borderlands	Negative	@Borderlands how do I submit a complaint? Your...
29	8857	Nvidia	Positive	Watching NVIDIA position itself as not just a ...
...	...	...	...	...
969	11957	Verizon	Negative	Old lady threatened me and said she was going ...
970	8318	Microsoft	Negative	Why didn't anyone think of this acronym for BL...
971	3080	Dota2	Positive	FUCKING FINALLY
972	2734	Borderlands	Positive	surprise i still love #borderlands ' jack pic....
973	397	Amazon	Positive	#Amazon Best Seller!\n\nScreen Cleaner Kit - B...
974	12949	Xbox(Xseries)	Irrelevant	Microsoft Apparently Has 16 First-Party Xbox S...
975	2219	CallOfDuty	Positive	this quarantine got me bored af playing call o...
976	4444	Google	Neutral	This is....odd. @Google suggests the most dead...
977	10128	PlayerUnknownsBattlegrounds(PUBG)	Irrelevant	How Abhijeet Bhayya ur efforts are incredible ...
978	11431	TomClancysRainbowSix	Negative	@Rainbow6Game when will you ban mnk/cronos pla...
979	5495	Hearthstone	Positive	Team 5 what a great media/announcement day. Yo...
980	12689	WorldOfCraft	Positive	@WarcraftDevs this would be incredible!
981	9446	Overwatch	Negative	Removing mcree's noose because ur pussies enit...
982	3108	Dota2	Negative	wtf
983	1373	Battlefield	Negative	I have noticed streamers I watch who are now p...
984	10588	RedDeadRedemption(RDR)	Neutral	@6th__man playing red dead redemption-\n\n"Oh ...
985	3980	Cyberpunk2077	Positive	This looks like a poster for a 90s action movi...
986	4905	GrandTheftAuto(GTA)	Irrelevant	The ultimate pruning machine! @stihlusa GTA 26...
987	7516	LeagueOfLegends	Neutral	♥ Suikoden 2\n 1 Alex Kidd in Miracle World\...
988	5708	HomeDepot	Positive	Thank you to Matching funds Home Depot RW paym...

	id	game	sentiment	text
989	11504	TomClancysRainbowSix	Positive	Tom Clancy's Rainbow Six Siege is a great game...
990	2165	CallOfDuty	Neutral	Late night stream with the boys! Come watch so...
991	2962	Dota2	Positive	guess i'll broke.
992	314	Amazon	Negative	Please explain how this is possible! How can t...
993	9701	PlayStation5(PS5)	Positive	Good on Sony. As much as I want to see the new...
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	Microsoft	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

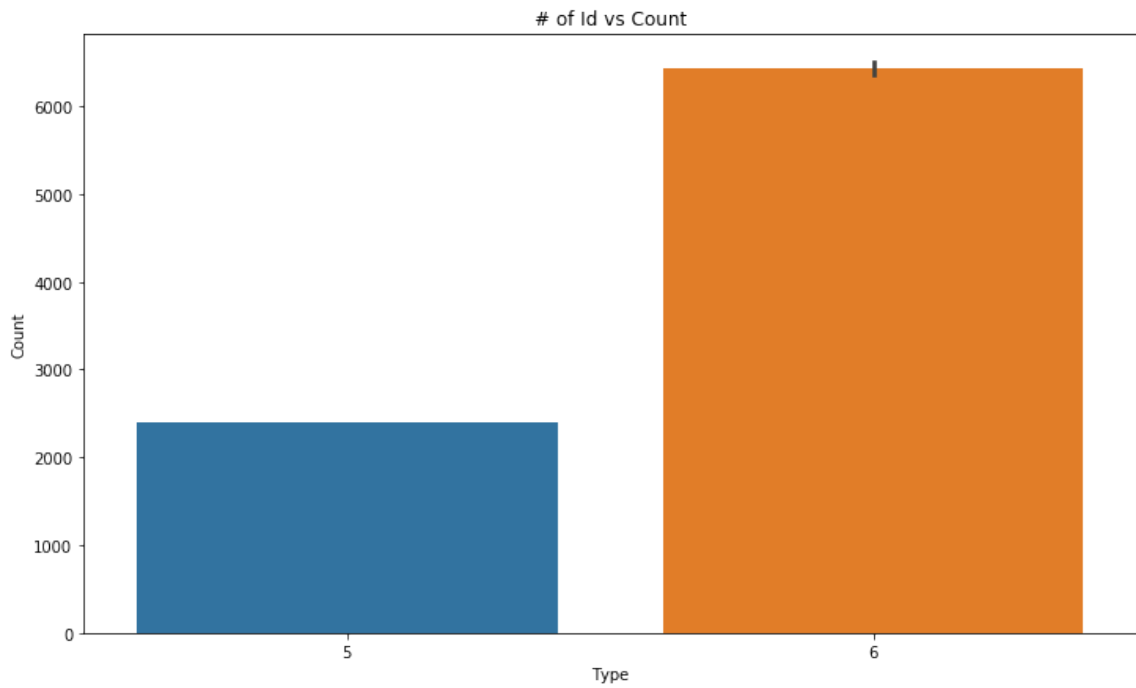
```
In [12]: id_types = data['id'].value_counts()
id_types
```



```
Out[12]: 2047      6
          2452      6
          4695      6
          6742      6
          597       6
          2644      6
          12883     6
          8785      6
          10832     6
          4631      6
          6678      6
          533       6
          2580      6
          8721      6
          10768     6
          4567      6
          6614      6
          469       6
          2516      6
          12755     6
          8657      6
          10704     6
          4503      6
          10896     6
          8849      6
          12947     6
          8977      6
          2900      6
          13139     6
          9041      6
          ..
          7081      6
          5032      6
          11119     6
          9070      6
          13164     6
          5224      6
          7273      6
          1130      6
          7401      6
          9518      6
          3371      6
          1322      6
          7465      6
          5416      6
          11503     6
          9454      6
          3307      6
          1258      6
          5352      6
          3179      6
          11439     6
          9390      6
          3243      6
          1194      6
          7337      6
          5288      6
          11375     6
          9326      6
          2049      6
```

```
2401      5  
Name: id, 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()
print(game_types)
```

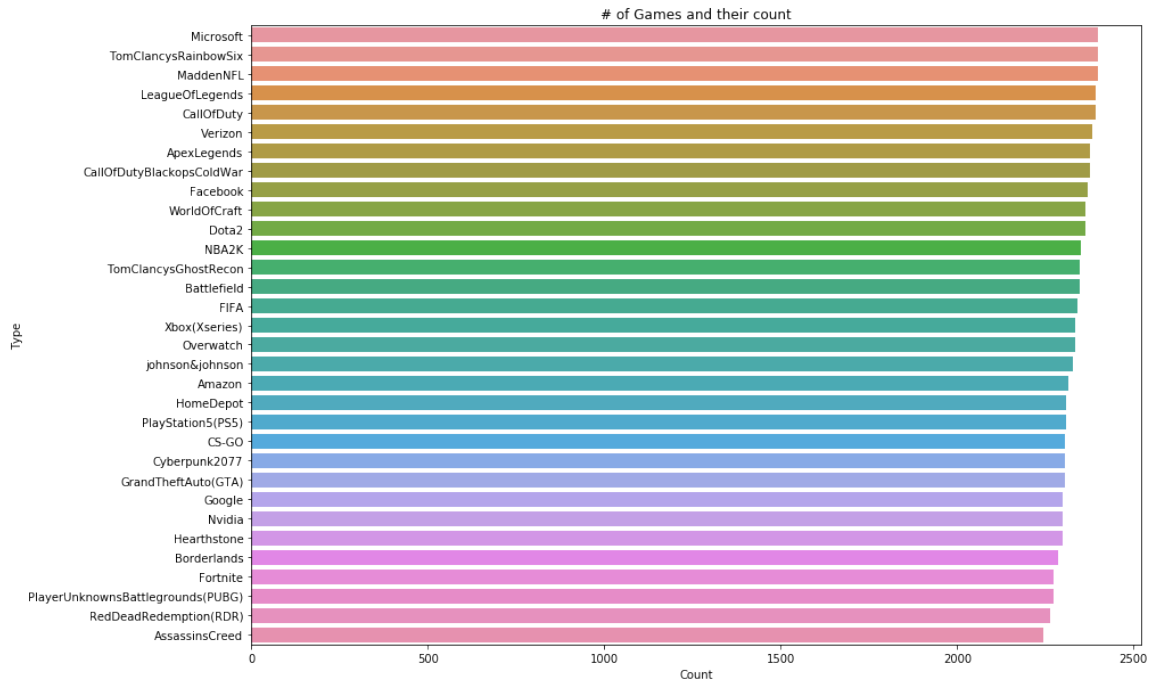
Microsoft	2400
TomClancysRainbowSix	2400
MaddenNFL	2400
LeagueOfLegends	2394
CallOfDuty	2394
Verizon	2382
ApexLegends	2376
CallOfDutyBlackopsColdWar	2376
Facebook	2370
WorldOfCraft	2364
Dota2	2364
NBA2K	2352
TomClancysGhostRecon	2346
Battlefield	2346
FIFA	2340
Xbox(Xseries)	2334
Overwatch	2334
johnson&johnson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

Name: game, 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()
```

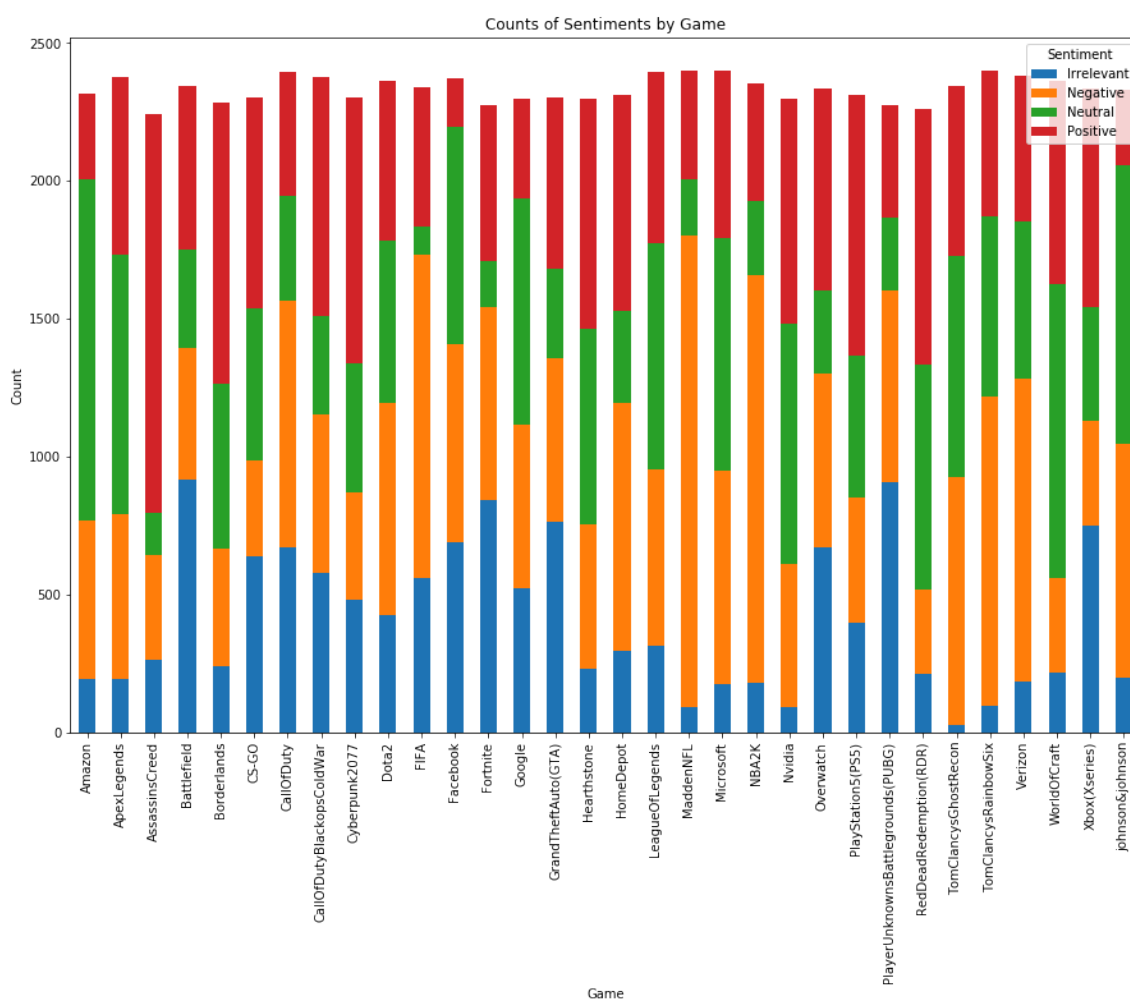


```
In [24]: #sns.catplot(x="game", hue="sentiment", kind="count", height=10, aspect=3, d
import matplotlib.pyplot as plt

# Group the data by 'game' and 'sentiment', and count the occurrences
grouped_data = data.groupby(['game', 'sentiment']).size().unstack()

# Plot the data using matplotlib
fig, ax = plt.subplots(figsize=(15, 10))
grouped_data.plot(kind='bar', stacked=True, ax=ax)

# Customize the plot
plt.xlabel('Game')
plt.ylabel('Count')
plt.title('Counts of Sentiments by Game')
plt.xticks(rotation=90)
plt.legend(title='Sentiment')
plt.show()
```



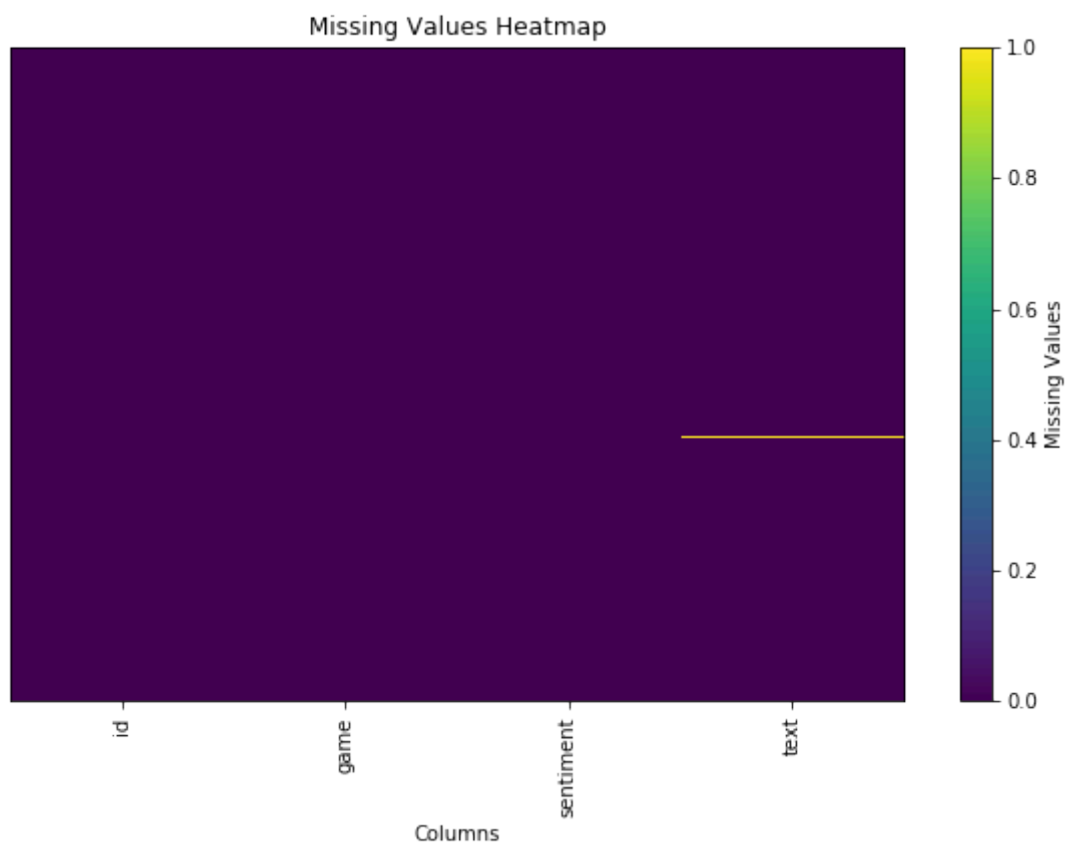
```
In [26]: #sns.heatmap(data.isnull(),yticklabels=False,cbar=False,cmap='viridis')
import matplotlib.pyplot as plt
import numpy as np

# Create a matrix indicating missing values in the data
missing_values_matrix = data.isnull().values.astype(int)

# Plot the heatmap using matplotlib
plt.figure(figsize=(10, 6))
plt.imshow(missing_values_matrix, cmap='viridis', aspect='auto')

# Customize the plot
plt.xlabel('Columns')
plt.title('Missing Values Heatmap')
plt.xticks(range(len(data.columns)), data.columns, rotation=90)
plt.yticks([]) # Disable y-axis ticks
plt.colorbar(label='Missing Values')

plt.show()
```



```
In [27]: total_null=data.isnull().sum().sort_values(ascending=False)
percent = ((data.isnull().sum()/data.isnull().count()*100).sort_values(ascending=False))
print("Total records = ", data.shape[0])
missing_data = pd.concat([total_null,percent.round(2)],axis=1,keys=['Total Missing','Percent'])
missing_data.head(10)
```

Total records = 74681

Out[27]:

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

```
In [28]: 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(ascending=False))
print("Total records = ", data.shape[0])
missing_data = pd.concat([total_null,percent.round(2)],axis=1,keys=['Total Missing','Percent'])
missing_data.head(10)
```

Total records = 73995

Out[28]:

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

```
In [29]: train0=data[data['sentiment']=="Negative"]
train1=data[data['sentiment']=="Positive"]
train2=data[data['sentiment']=="Irrelevant"]
train3=data[data['sentiment']=="Neutral"]
```

```
In [30]: train0.shape, train1.shape, train2.shape, train3.shape
```

Out[30]: ((22358, 4), (20654, 4), (12875, 4), (18108, 4))

```
In [31]: 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 [32]: train0.shape, train1.shape, train2.shape, train3.shape
```

Out[32]: ((1863, 4), (1721, 4), (1072, 4), (1509, 4))

```
In [33]: data=pd.concat([train0,train1,train2,train3],axis=0)
print(data)
```

	id	game	sentiment	\
23	2405	Borderlands	Negative	
24	2405	Borderlands	Negative	
25	2405	Borderlands	Negative	
26	2405	Borderlands	Negative	
27	2405	Borderlands	Negative	
28	2405	Borderlands	Negative	
35	2407	Borderlands	Negative	
36	2407	Borderlands	Negative	
37	2407	Borderlands	Negative	
38	2407	Borderlands	Negative	
39	2407	Borderlands	Negative	
40	2407	Borderlands	Negative	
53	2410	Borderlands	Negative	
54	2410	Borderlands	Negative	
55	2410	Borderlands	Negative	
56	2410	Borderlands	Negative	
57	2410	Borderlands	Negative	
58	2410	Borderlands	Negative	



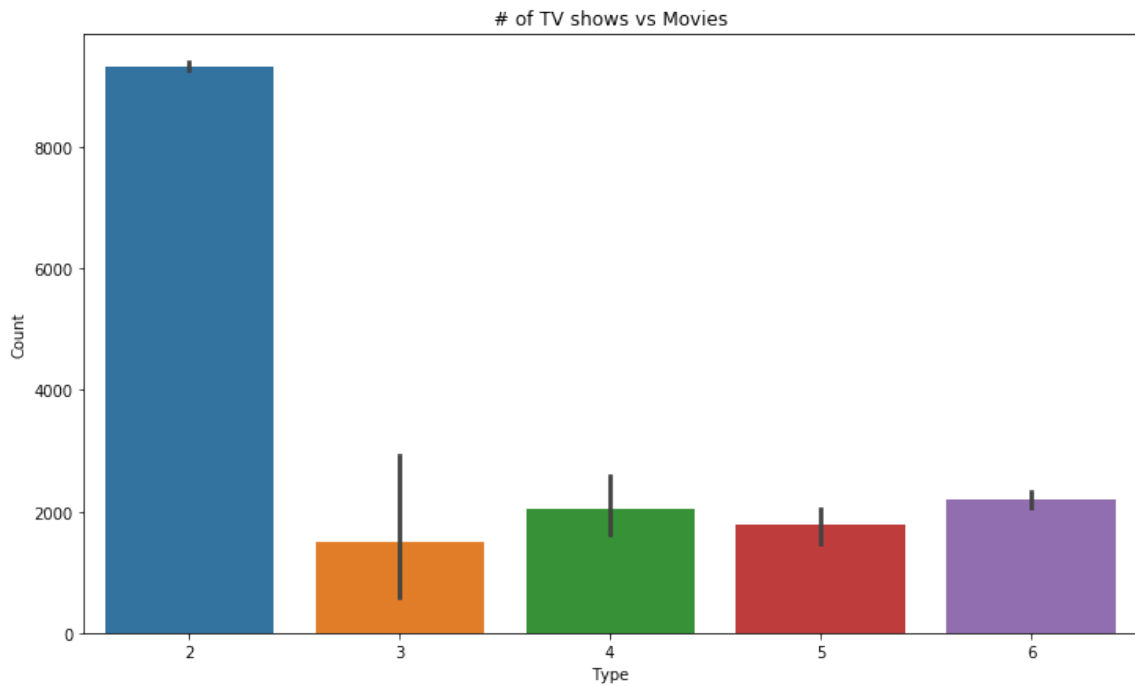
```
In [35]: id_types = data['id'].value_counts()
print(id_types)
```

1999	6
2584	6
2636	6
2632	6
2628	6
2624	6
2620	6
2616	6
2612	6
2608	6
2604	6
2600	6
2596	6
2592	6
2588	6
2580	6
2644	6
2576	6
2572	6
2564	6
2556	6
2552	6
2548	6
2544	6
2540	6
2536	6
2528	6
2524	6
2520	6
2516	6
	..
2532	5
1737	5
1717	5
2503	5
1708	5
1613	5
1779	5
2411	5
2401	5
1731	5
1622	5
2595	4
1632	4
1863	4
333	3
1880	3
1929	3
16	3
1602	3
42	3
1959	3
9256	3
235	3
1763	3
1719	3
134	3
37	3
157	3
9267	2

```
9388    2
Name: id, Length: 1040, dtype: int64
```

```
In [36]: 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()
```

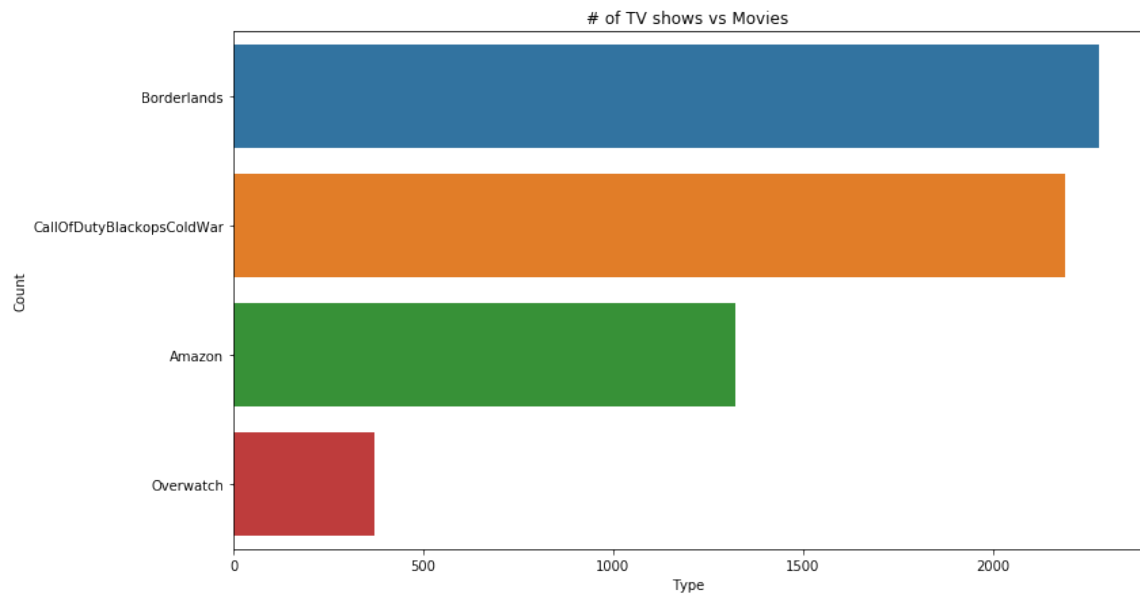


```
In [37]: game_types = data['game'].value_counts()
print(game_types)
```

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

```
In [38]: plt.figure(figsize=(12,7))
sns.barplot(x=game_types.values,y=game_types.index)

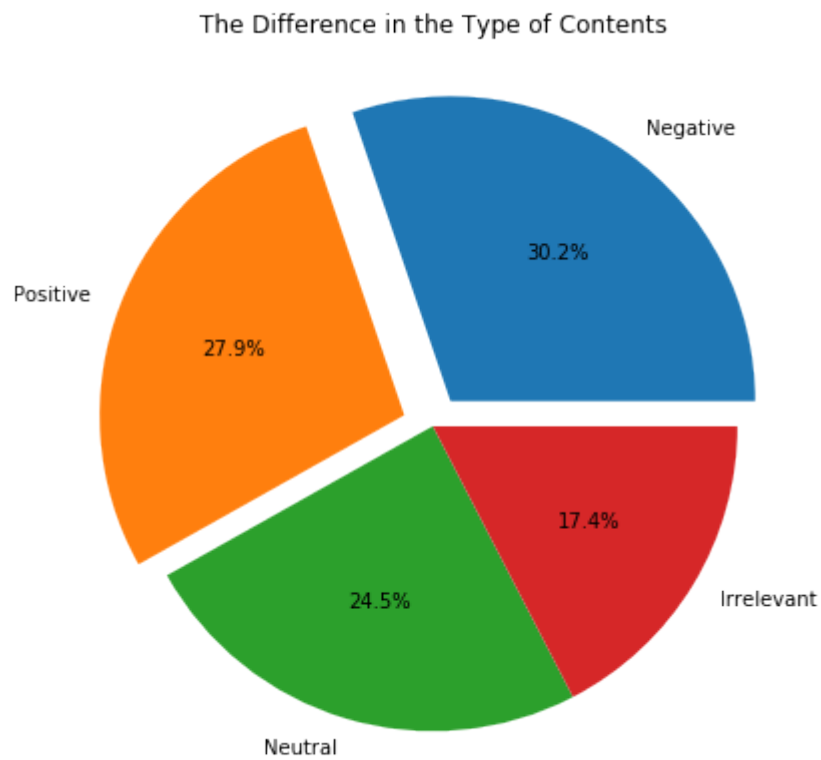
plt.xlabel('Type')
plt.ylabel('Count')
plt.title('# of TV shows vs Movies')
plt.show()
```



```
In [39]: sentiment_types = data['sentiment'].value_counts()
print(sentiment_types)
```

```
Negative      1863
Positive      1721
Neutral       1509
Irrelevant    1072
Name: sentiment, dtype: int64
```

```
In [41]: plt.figure(figsize=(12,7))  
plt.pie(x=sentiment_types.values, labels=sentiment_types.index, autopct='%1  
plt.title('The Difference in the Type of Contents')  
plt.gca().set_aspect('equal')  
plt.show()
```



```
In [44]: import matplotlib.pyplot as plt
import numpy as np

# Group the data by 'game' and 'sentiment', and count the occurrences
grouped_data = data.groupby(['game', 'sentiment']).size().unstack()

# Get the categories and sentiments
games = grouped_data.index
sentiments = grouped_data.columns

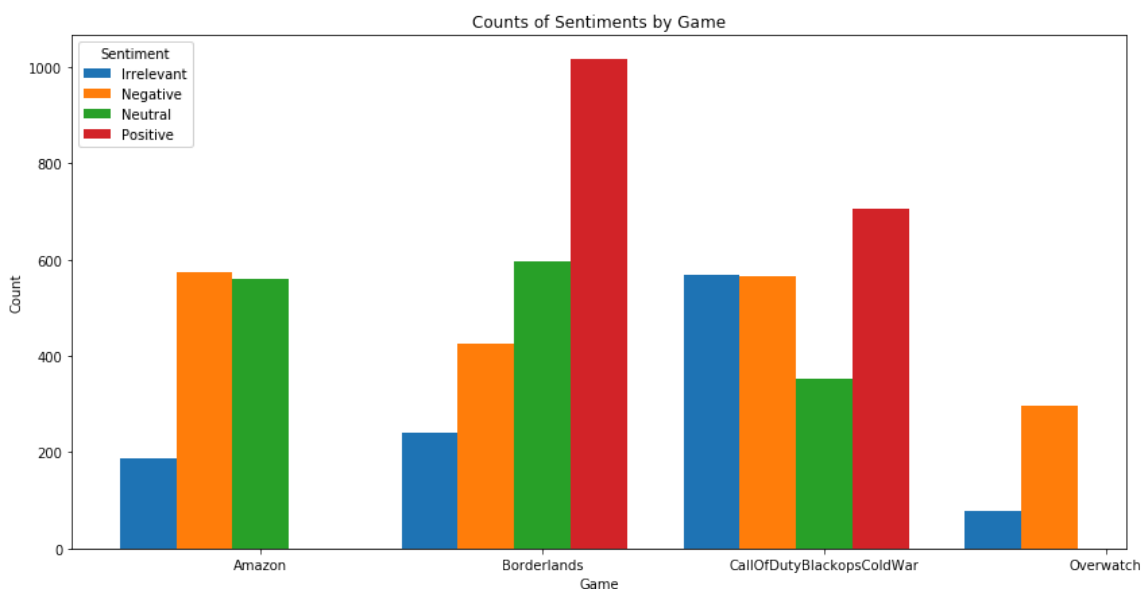
# Set up the plot
fig, ax = plt.subplots(figsize=(14, 7))

# Plot each sentiment category for each game
bar_width = 0.2
index = np.arange(len(games))

for i, sentiment in enumerate(sentiments):
    ax.bar(index + (i * bar_width), grouped_data[sentiment], width=bar_width)

# Add labels and title
ax.set_xlabel('Game')
ax.set_ylabel('Count')
ax.set_title('Counts of Sentiments by Game')
ax.set_xticks(index + (bar_width * len(sentiments)) / 2)
ax.set_xticklabels(games)
ax.legend(title='Sentiment')

plt.show()
```



```
In [45]: from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
```

```
In [46]: 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 [47]: data = data.drop(['id'],axis=1)
print(data)
```

	game	sentiment	text
23	1	1	the biggest dissappointment 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 dissappointment in my life coming o...
27	1	1	For the biggest male dissappointment in my life...
28	1	1	the biggest dissappointment in my life came bac...
35	1	1	Man Gearbox really needs to fix this dissapoi...
36	1	1	Man Gearbox really needs to fix these disappoi...
37	1	1	Man Gearbox really needs to fix this dissapoi...
38	1	1	Man Bethesda really needs to fix this dissapoi...
39	1	1	Man Gearbox really needs to fix this dissapoi...
40	1	1	<unk> Gearbox really time to fix this 10 drops...
53	1	1	why does like every man in borderlands have sl...
54	1	1	Why, like every man in border countries, have ...
55	1	1	Why, like everyone else in the border countrie...
56	1	1	why does like<unk> man in borderlands have sli...
57	1	1	why Beth does like every man in borderlands ha...
58	1	1	why does practically every man in France have ...
95	1	1	Grounded almost looked pretty cool even despit...
96	1	1	Grounded looked cool despite the borderline un...
97	1	1	Grosskreutz looked pretty cool, even despite t...
98	1	1	Grounded almost looked pretty cool here despit...
99	1	1	Deep Grounded almost looked pretty cool even d...
100	1	1	Grounded almost was pretty cool even despite t...
107	1	1	@Borderlands how do I submit a complaint? Your...
108	1	1	@ Borderlands, how can I file a complaint? You...
109	1	1	@ Borderlands how to file a complaint? Your CE...
110	1	1	@Borderlands how did I submit a complaint? You...
111	1	1	@Borderlands how do suppose I submit a complai...
112	1	1	Now how do I submit any complaint? Your CEO is...
...	...	...	...
5572	0	2	Plain Janie. Smart girl. And the good sister t...
5573	0	2	Starting season 4 of Bosch on Amazon Prime vid...
5574	0	2	Starts Season 4 of Bosch on Amazon Prime Video...
5575	0	2	The fourth season of the series starts on Amaz...
5576	0	2	Starting season 4 in Bosch on Amazon Prime vid...
5577	0	2	I Starting season 4 of Bosch on an Amazon Prim...
5578	0	2	Starting Part 4 of Bosch on Amazon Prime video...
5579	0	2	I played this interesting quiz on Amazon - Try...
5580	0	2	ve played this interesting quiz on Amazon - Tr...
5581	0	2	I played this interesting quiz on Amazon - Try...
5582	0	2	I played this interesting quiz on Amazon - Try...
5583	0	2	I played this interesting quiz on Amazon - Try...
5584	0	2	us take this interesting quiz on Amazon - Wish...
5585	0	2	Agents of Mayhem is now Just \$8.56. . Get it h...
5586	0	2	Agents of Mayhem now costs just \$8.56.. Get it...
5587	0	2	Mayhem's agent now costs only \$8.56. You can g...
5588	0	2	Agents of Mayhem Out now Just \$8.56.. Get it h...
5589	0	2	Agents of Infinite Mayhem there is and now cos...
5590	0	2	Agents Against Mayhem Are A Just a.56.. Found ...
5591	0	2	I played this interesting quiz on Amazon - Try...
5592	0	2	ve played this interesting quiz on Amazon - Tr...
5593	0	2	I played this interesting quiz on Amazon - Try...
5594	0	2	I played this interesting quiz on Amazon - Try...
5595	0	2	I played all this really interesting quiz on A...
5596	0	2	mind get this interesting chance on Amazon - T...
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 x 3 columns]
```

```
In [48]: data.nunique()
```

```
Out[48]: game          4  
sentiment         4  
text           5854  
dtype: int64
```

```
In [49]: v_data.nunique()
```

```
Out[49]: id           999  
game            32  
sentiment        4  
text           998  
dtype: int64
```