```
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
```

+ Code + Text

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

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
True
```

```
data = pd.read_csv('/content/twitter_training.csv')
v_data = pd.read_csv('/content/twitter_validation.csv')
```

```
data
```

	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...
...	...	...	...	...
60949	4847	GrandTheftAuto(GTA)	Irrelevant	Whoever vandalizing the speed cameras around t...
60950	4847	GrandTheftAuto(GTA)	Irrelevant	Whoever is vandalizing the speed cameras aroun...
60951	4847	GrandTheftAuto(GTA)	Irrelevant	OH Whoever is vandalizing the video speed came...
60952	4847	GrandTheftAuto(GTA)	Irrelevant	Whoever is mounting the speed cameras around f...
60953	4848	GrandTheftAuto(GTA)	Irrelevant	The first one looks like the shit you see in h...

60954 rows × 4 columns

```
v_data
```

	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 🤣
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...

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

```
data
```

	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...
...	...	...	...	...
60949	4847	GrandTheftAuto(GTA)	Irrelevant	Whoever vandalizing the speed cameras around t...
60950	4847	GrandTheftAuto(GTA)	Irrelevant	Whoever is vandalizing the speed cameras aroun...
60951	4847	GrandTheftAuto(GTA)	Irrelevant	OH Whoever is vandalizing the video speed came...
60952	4847	GrandTheftAuto(GTA)	Irrelevant	Whoever is mounting the speed cameras around f...

```
v_data
```

	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

```
data.shape
```

```
(60954, 4)
```

```
data.columns
```

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

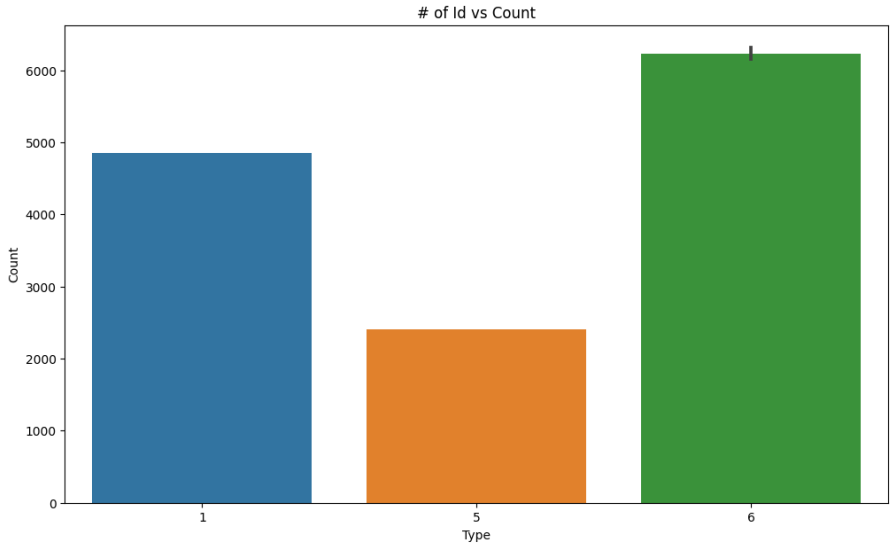
```
data.describe(include='all')
```

	id	game	sentiment	text
count	60954.000000	60954	60954	60400
unique	NaN	27	4	56655
top	NaN	TomClancysRainbowSix	Negative	At the same time, despite the fact that there ...
freq	NaN	2400	17646	137
mean	6231.609886	NaN	NaN	NaN
std	3953.822916	NaN	NaN	NaN
min	1.000000	NaN	NaN	NaN
25%	2608.250000	NaN	NaN	NaN
50%	5990.000000	NaN	NaN	NaN
75%	9793.000000	NaN	NaN	NaN
max	13200.000000	NaN	NaN	NaN

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

```
7238     6
1375     6
1368     6
1369     6
1370     6
        ..
12682    6
12683    6
4846     6
2401     5
4848     1
Name: id, Length: 10160, dtype: int64
```

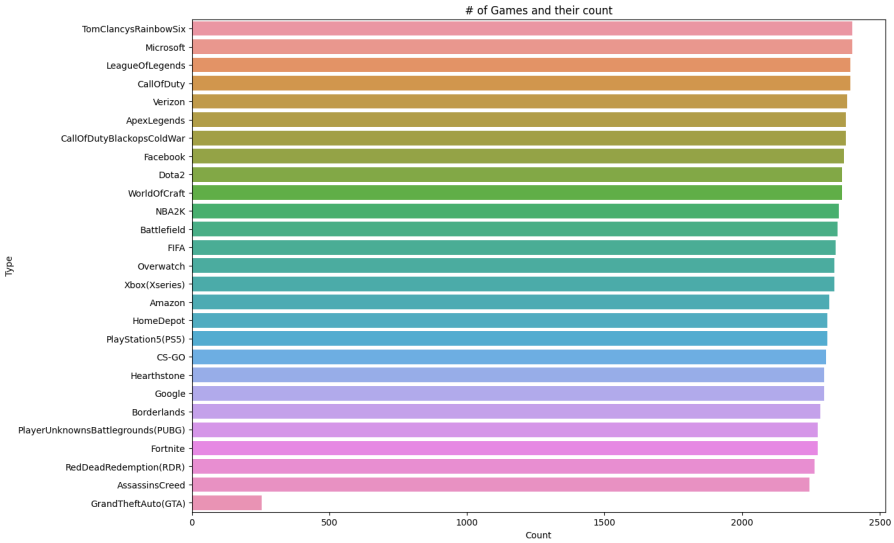
```
plt.figure(figsize=(12,7))
sns.barplot(y=id_types.index,x=id_types.values)
plt.xlabel('Type')
plt.ylabel('Count')
plt.title('# of Id vs Count')
plt.show()
```




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

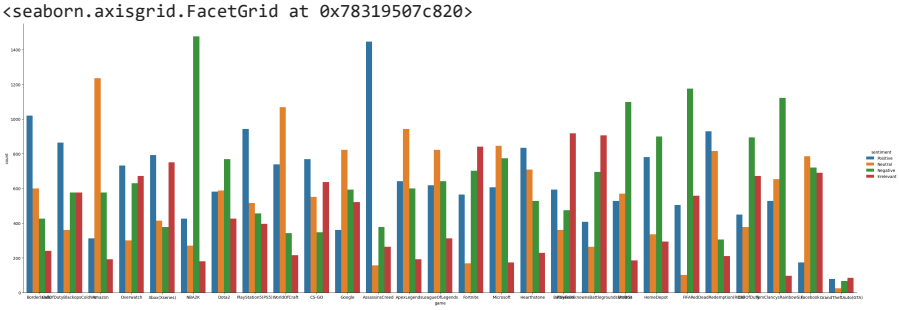
```
TomClancysRainbowSix                2400
Microsoft                           2400
LeagueOfLegends                     2394
CallOfDuty                          2394
Verizon                             2382
ApexLegends                         2376
CallOfDutyBlackopsColdWar           2376
Facebook                            2370
Dota2                               2364
WorldOfCraft                        2364
NBA2K                               2352
Battlefield                         2346
FIFA                                2340
Overwatch                           2334
Xbox(Xseries)                       2334
Amazon                              2316
HomeDepot                           2310
PlayStation5(PS5)                   2310
CS-GO                               2304
Hearthstone                         2298
Google                              2298
Borderlands                         2285
PlayerUnknownsBattlegrounds(PUBG)   2274
Fortnite                            2274
RedDeadRedemption(RDR)              2262
AssassinsCreed                      2244
GrandTheftAuto(GTA)                  253
Name: game, dtype: int64
```

```
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()
```



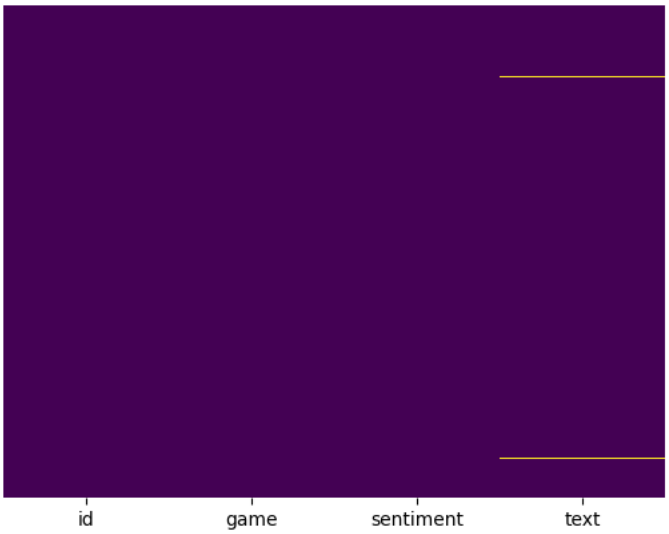

```
sns.catplot(x="game",hue="sentiment", kind="count",height=10,aspect=3, data=data)
```

```
<seaborn.axisgrid.FacetGrid at 0x78319507c820>
```



```
sns.heatmap(data.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

```
<Axes: >
```




```
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','In Percent'])
missing_data.head(10)
```

```
Total records =  60954
```

	Total Missing	In Percent
text	554	0.91
id	0	0.00
game	0	0.00
sentiment	0	0.00

```
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','In Percent'])
missing_data.head(10)
```

```
Total records =  60400
```

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

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

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

```
((17497, 4), (17080, 4), (11332, 4), (14491, 4))
```

```
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)]
```

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

```
((1458, 4), (1423, 4), (944, 4), (1207, 4))
```

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

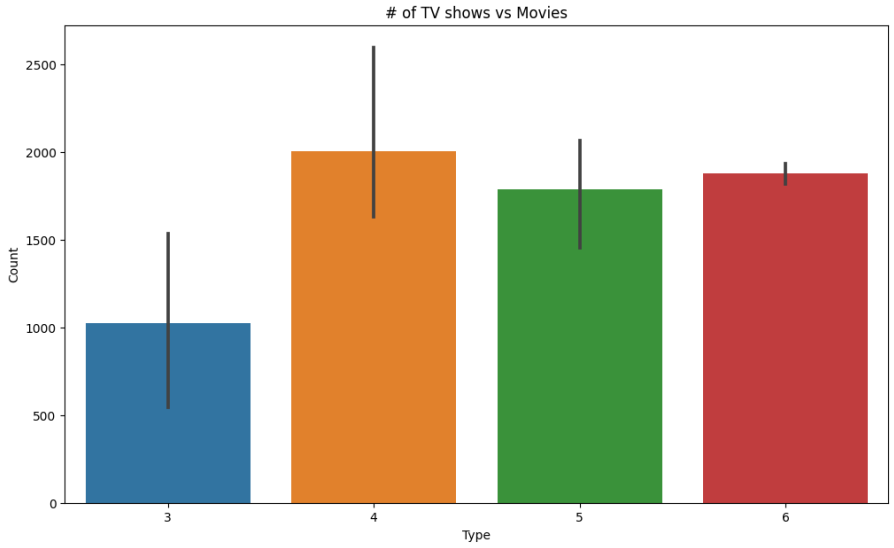
	id	game	sentiment	text
23	2405	Borderlands	Negative	the biggest dissappoinment in my life came out...
24	2405	Borderlands	Negative	The biggest disappointment of my life came a y...
25	2405	Borderlands	Negative	The biggest disappointment of my life came a y...
26	2405	Borderlands	Negative	the biggest dissappoinment in my life coming o...
27	2405	Borderlands	Negative	For the biggest male dissappoinment in my life...
...	...	...	...	...
5085	74	Amazon	Neutral	Are. 5 buff. 5 ly / → 3jdEiGn → https://t.co/C...
5086	74	Amazon	Neutral	a.<unk>.ly/3jdEiGn https://t.co/CdCiSpUNOy]
5087	75	Amazon	Neutral	RT @richardturrin: Amazon and Goldman partner....
5088	75	Amazon	Neutral	RT @ richardturrin: Amazon and Goldman partner...
5089	75	Amazon	Neutral	RT @ richardturrin: Amazon and Goldman Sachs p...

5032 rows × 4 columns

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

```
2405    6
1777    6
1698    6
1700    6
1709    6
       ..
1959    3
16      3
1719    3
42      3
75      3
Name: id, Length: 848, dtype: int64
```

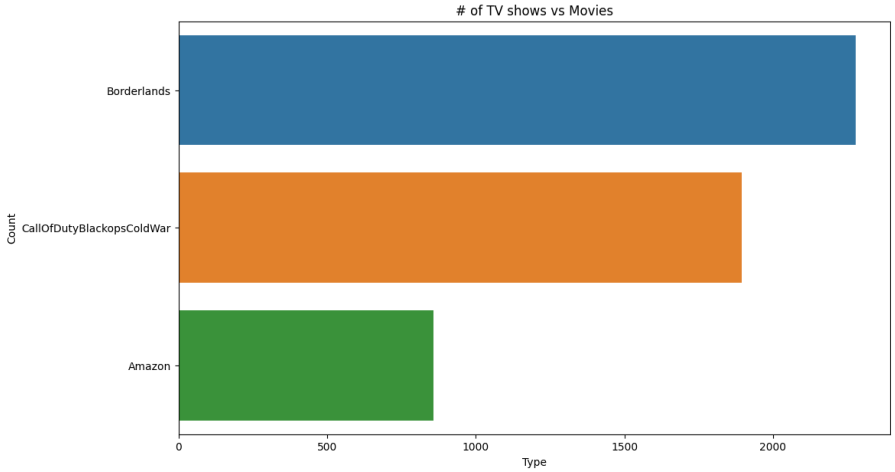
```
plt.figure(figsize=(12,7))
sns.barplot(x=id_types.values,y=id_types.index)
plt.xlabel('Type')
plt.ylabel('Count')
plt.title('# of TV shows vs Movies')
plt.show()
```




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

```
Borderlands                  2279
CallOfDutyBlackopsColdWar    1894
Amazon                        859
Name: game, dtype: int64
```

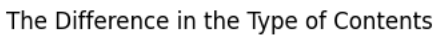
```
plt.figure(figsize=(12,7))
sns.barplot(x=game_types.values,y=game_types.index)
plt.xlabel('Type')
plt.ylabel('Count')
plt.title('# of TV shows vs Movies')
plt.show()
```




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

```
Negative      1458
Positive      1423
Neutral       1207
Irrelevant     944
Name: sentiment, dtype: int64
```

```
plt.figure(figsize=(12,7))
plt.pie(x=sentiment_types.values, labels=sentiment_types.index, autopct='%.1f%%', explode=[0.1, 0.1,0,0])
plt.title('The Difference in the Type of Contents')
plt.show()
```




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

```
<seaborn.axisgrid.FacetGrid at 0x783195221150>
```