

```
In [3]: import pandas as pd
import numpy as np
from nltk.tokenize import sent_tokenize, word_tokenize
from sklearn.feature_extraction.text import CountVecorizer
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 [4]: nltk.download('stopwords')
```

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

```
Out[4]: True
```

```
In [5]: v_data = pd.read_csv('twitter_training.csv')
v_data = pd.read_csv('twitter_validation.csv')
```

```
In [6]: 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...	
...	...	...	...	...	...
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	x 4 columns			

```
In [7]: 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 dai...
1	6312	Microsoft	Negative		@Microsoft Why do I pay for WORD when it fund...
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		H @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		IHS 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 tal baby po...
...	...	...	...	...	...
999	rows	x 4 columns			

```
In [8]: data.columns = ['id', 'game', 'sentiment', 'text']
v_data.columns = ['id', 'game', 'sentiment', 'text']
```

```
In [9]: 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...	
...	...	...	...	...	...
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	x 4 columns			

```
In [10]: v_data
```

		id	game	sentiment	text
0	352	Amazon	Neutral		BBC News - Amazon boss Jeff Bezos rejects dai...
1	6312	Microsoft	Negative		@Microsoft Why do I pay for WORD when it fund...
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		H @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		IHS 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 tal baby po...
...	...	...	...	...	...
999	rows	x 4 columns			

```
In [11]: data.shape
```

```
Out[11]: (74681, 4)
```

```
In [12]: data.columns
```

```
Out[12]: Index(['id', 'game', 'sentiment', 'text'], dtype='object')
```

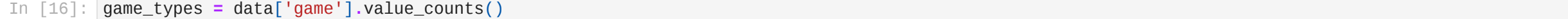
```
In [13]: data.describe(include='all')
```

		id	game	sentiment	text
count	74681	0.000000	74681	74681	73995
unique	NaN	NaN	32	4	69490
top	NaN	TomClancysRainbowSix	Negative		
freq	NaN	2400	22542	172	
mean	6432.640149	NaN	NaN	NaN	NaN
std	3740.423819	NaN	NaN	NaN	NaN
min	1.000000	NaN	NaN	NaN	NaN
25%	3195.000000	NaN	NaN	NaN	NaN
50%	6422.000000	NaN	NaN	NaN	NaN
75%	9601.000000	NaN	NaN	NaN	NaN
max	13200.000000	NaN	NaN	NaN	NaN

```
In [14]: id_types = data['id'].value_counts()
id_types
```

id	count
5283	6
6164	6
6141	6
6142	6
6143	6
...	...
4678	6
4679	6
4680	6
4681	6
2481	5
Name: count, Length: 12447, dtype: int64	

```
In [15]: plt.figure(figsize=(12,7))
sns.barplot(y=id_types.index, x=id_types.values)
```



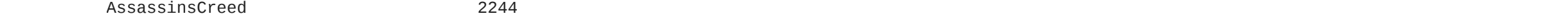
```
In [16]: game_types = data['game'].value_counts()
game_types
```

game	count
TomClancysRainbowSix	2400
MaddenNFL	2400
Microsoft	2400
LeagueOfLegends	2394
CallOfDuty	2394
Verizon	2382
CallOfDutyBlackopsColdwar	2376
ApexLegends	2376
Facebook	2379
WorldOfCraft	2364
Data2	2364
NBA2K	2352
TomClancysGhostRecon	2346
Battlefield	2346
FIFA	2340
Xbox(Xseries)	2334
Overwatch	2334
Johnson&Johnson	2328
Amazon	2326
PlayStation5(PSS)	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	

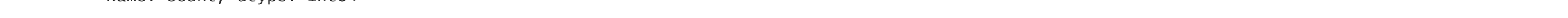
```
In [17]: plt.figure(figsize=(14,10))
sns.barplot(x=game_types.values, y=game_types.index)
```



```
In [18]: sns.catplot(x='game', hue='sentiment', kind='count', height=10, aspect=3, data=data)
```



```
In [19]: sns.heatmap(data.isnull(), yticklabels=False, cbar=False, cmap='viridis')
```



```
In [20]: 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 Missing	In Percent
text	688	0.92	
id	0	0.00	
game	0	0.00	
sentiment	0	0.00	

```
In [21]: 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 Missing	In Percent
id	0	0.0	
game	0	0.0	
sentiment	0	0.0	
text	0	0.0	

```
In [22]: train0=data[data['sentiment']=='Negative']
train1=data[data['sentiment']=='Positive']
train2=data[data['sentiment']=='Irrelevant']
train3=data[data['sentiment']=='Neutral']
```

```
In [23]: train0.shape, train1.shape, train2.shape, train3.shape
```

```
Out[23]: ((22358, 4), (20654, 4), (12875, 4), (18108, 4))
```

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

```
data
```

		id	game	sentiment	text
23	2405	Borderlands	Negative	the biggest disappointment 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 disappointment in my life coming o...	
27	2405	Borderlands	Negative	For the biggest male disappointment in my life...	
...	...	...	...	...	...
74658	9197	Nvidia	Neutral	Nvidia plans to release its 2017 "Crypto Craze...	
74659	9197	Nvidia	Neutral	Nvidia does not want to give up its "cryptoin...	
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...	
...	...	...	...	...	...
73995	rows	x 4 columns			

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

id	count
16	6
6649	6
6619	6
6631	6
6632	6
...	...
3288	3
1384	3
10250	3
12519	3
Name: count, Length: 12447, dtype: int64	

```
In [26]: plt.figure(figsize=(12,7))
sns.barplot(x=id_types.values, y=id_types.index)
```



```
In [27]: game_types = data['game'].value_counts()
game_types
```

game	count
MaddenNFL	2377
LeagueOfLegends	2377
CallOfDuty	2376
Verizon	2365
TomClancysRainbowSix	2364
Microsoft	2364
Facebook	2362
WorldOfCraft	2357
NBA2K	2353
CallOfDutyBlackopsColdwar	2343
FIFA	2324
Johnson&Johnson	2324
TomClancysGhostRecon	2321
Battlefield	2316
Overwatch	2316
GrandTheftAuto(GTA)	2293
HomeDepot	2292
PlayStation5(PSS)	2291
Hearthstone	2286
CS-GO	2284
Xbox(Xseries)	2283
Borderlands	2279
Amazon	2276
Google	2274
Nvidia	2271
Cyberpunk2077	2262
RedDeadRedemption(RDR)	2249
Fortnite	2234
PlayerUnknownsBattlegrounds(PUBG)	2234
AssassinsCreed	2234
Name: count, dtype: int64	

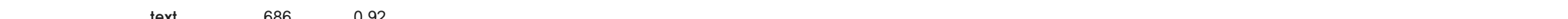
```
In [28]: plt.figure(figsize=(12,7))
sns.barplot(x=game_types.values, y=game_types.index)
```



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

sentiment	count
Negative	22358
Positive	20654
Neutral	18108
Irrelevant	12875
Name: count, dtype: int64	

```
In [30]: plt.figure(figsize=(12,7))
plt.pie(sentiment_types.values, labels=sentiment_types.index, autopct='%1.0f%%', explode=[0.1, 0.1, 0.8])
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)
```



```
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
```

		game	sentiment	text
23	4	1	the biggest disappointment in my life came out...	
24	4	1	The biggest disappointment of my life came a y...	
25	4	1	The biggest disappointment of my life came a y...	
26	4	1	the biggest disappointment in my life coming o...	
27	4	1	For the biggest male disappointment in my life...	
...	...	...	...	...
74658	21	2	Nvidia plans to release its 2017 "Crypto Craze...	
74659	21	2	Nvidia does not want to give up its "cryptoin...	
74660	21	2	Nvidia doesn't intend to give away its 2017 ad...	
74661	21	2	Nvidia therefore doesn't want to give up its...	
74662	21	2	is doesn't should I give up its password 'cryp...	
...	...	...	...	...
73995	rows	x 3 columns		

```
In [35]: data.nunique()
```

game	32
sentiment	4
text	69490
dtype: int64	

```
In [36]: v_data.nunique()
```

id	999
game	32
sentiment	4
text	998
dtype: int64	

```
In [ ]:
```