

```
In [272... import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import re #regular expression
import string
import nltk #natural language toolkit for text processing
```

```
In [273... twitter= pd.read_csv('C:/Users/DELL/Downloads/twitter_training.csv', names=['ID', 'Topic', 'Sentiment', 'Tweet'])
```

```
In [274... twitter['Topic'] = twitter['Topic'].astype(str)
twitter['Sentiment'] = twitter['Sentiment'].astype(str)
twitter['Tweet'] = twitter['Tweet'].astype(str)
twitter.head()
```

```
Out[274... 
```

	ID	Topic	Sentiment	Tweet
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...

```
In [275... twitter.tail()
```

```
Out[275... 
```

	ID	Topic	Sentiment	Tweet
74677	9200	Nvidia	Positive	Just realized that the Windows partition of my...
74678	9200	Nvidia	Positive	Just realized that my Mac window partition is ...
74679	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
74680	9200	Nvidia	Positive	Just realized between the windows partition of...
74681	9200	Nvidia	Positive	Just like the windows partition of my Mac is I...

```
In [276... twitter.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 74682 entries, 0 to 74681
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0    ID          74682 non-null  int64
1    Topic       74682 non-null  object
2    Sentiment   74682 non-null  object
3    Tweet       74682 non-null  object
dtypes: int64(1), object(3)
memory usage: 2.3+ MB
```

```
In [277... #removing special characters,numbers and punctuations
# Function to remove special characters, numbers, and punctuation
def remove_special_characters(text):
    import re
    return re.sub(r'^[a-zA-Z#]', ' ', text)

# Apply the function to the 'Tweet' column
twitter['Tweet'] = twitter['Tweet'].apply(remove_special_characters)
```

```
In [278... twitter.head(30)
```

Out [278..

	ID	Topic	Sentiment	Tweet
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands and i will murder ...
5	2401	Borderlands	Positive	im getting into borderlands and i can murder y...
6	2402	Borderlands	Positive	So I spent a few hours making something for fu...
7	2402	Borderlands	Positive	So I spent a couple of hours doing something f...
8	2402	Borderlands	Positive	So I spent a few hours doing something for fun...
9	2402	Borderlands	Positive	So I spent a few hours making something for fu...
10	2402	Borderlands	Positive	So I spent a few hours making something f...
11	2402	Borderlands	Positive	was
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 Varlope RARE POWERFUL HANDSOM...
15	2403	Borderlands	Neutral	Rock Hard La Vita RARE BUT POWERFUL HANDSOME...
16	2403	Borderlands	Neutral	Live Rock Hard music La la Varlope RARE t...
17	2403	Borderlands	Neutral	I Hard like me RARE LONDON DE HANDSOME ...
18	2404	Borderlands	Positive	that was the first borderlands session in a lo...
19	2404	Borderlands	Positive	this 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 was the first borderlands session in a lo...
22	2404	Borderlands	Positive	that I was the first real borderlands session ...
23	2404	Borderlands	Positive	that was the first borderlands session in a ho...
24	2405	Borderlands	Negative	the biggest dissappointment in my life came out...
25	2405	Borderlands	Negative	The biggest disappointment of my life came a y...
26	2405	Borderlands	Negative	The biggest disappointment of my life came a y...
27	2405	Borderlands	Negative	the biggest dissappointment in my life coming o...
28	2405	Borderlands	Negative	For the biggest male dissappointment in my life...
29	2405	Borderlands	Negative	the biggest dissappointment in my life came bac...

In [279..

```
#removing short words
twitter['Tweet']=twitter['Tweet'].apply(lambda x: " ".join([w for w in x.split() if len(w) > 3]))
```

In [280..

```
twitter.head(20)
```

Out [280...	ID	Topic	Sentiment	Tweet
0	2401	Borderlands	Positive	getting borderlands will murder
1	2401	Borderlands	Positive	coming borders will kill
2	2401	Borderlands	Positive	getting borderlands will kill
3	2401	Borderlands	Positive	coming borderlands will murder
4	2401	Borderlands	Positive	getting borderlands will murder
5	2401	Borderlands	Positive	getting into borderlands murder
6	2402	Borderlands	Positive	spent hours making something know HUGE Borderl...
7	2402	Borderlands	Positive	spent couple hours doing something know that h...
8	2402	Borderlands	Positive	spent hours doing something know HUGE Borderla...
9	2402	Borderlands	Positive	spent hours making something know HUGE Rhandle...
10	2402	Borderlands	Positive	spent hours making something know HUGE Rhandle...
11	2402	Borderlands	Positive	
12	2403	Borderlands	Neutral	Rock Hard Varlope RARE POWERFUL HANDSOME JACKP...
13	2403	Borderlands	Neutral	Rock Hard Varlope RARE POWERFUL HANDSOME JACKP...
14	2403	Borderlands	Neutral	Rock Hard Varlope RARE POWERFUL HANDSOME JACKP...
15	2403	Borderlands	Neutral	Rock Hard Vita RARE POWERFUL HANDSOME JACKPOT ...
16	2403	Borderlands	Neutral	Live Rock Hard music Varlope RARE POWERFUL Liv...
17	2403	Borderlands	Neutral	Hard like RARE LONDON HANDSOME Borderlands Xbo...
18	2404	Borderlands	Positive	that first borderlands session long time where...
19	2404	Borderlands	Positive	this first Borderlands session long time where...

```
In [281... #tokenizing words in each tweet as tokens
tokens=twitter['Tweet'].apply(lambda x: x.split())
tokens.head()
```

```
Out[281... 0    [getting, borderlands, will, murder]
1          [coming, borders, will, kill]
2    [getting, borderlands, will, kill]
3    [coming, borderlands, will, murder]
4    [getting, borderlands, will, murder]
Name: Tweet, dtype: object
```

```
In [282... #stem the words
from nltk.stem.porter import PorterStemmer
stemmer=PorterStemmer()
```

```
In [283... tokens=tokens.apply(lambda sentence:[stemmer.stem(word) for word in sentence])
tokens.head(15)
```

```
Out[283... 0          [get, borderland, will, murder]
1          [come, border, will, kill]
2          [get, borderland, will, kill]
3          [come, borderland, will, murder]
4          [get, borderland, will, murder]
5          [get, into, borderland, murder]
6    [spent, hour, make, someth, know, huge, border...
7    [spent, coupl, hour, do, someth, know, that, h...
8    [spent, hour, do, someth, know, huge, borderla...
9    [spent, hour, make, someth, know, huge, rhandl...
10   [spent, hour, make, someth, know, huge, rhandl...
11          []
12   [rock, hard, varlop, rare, power, handsom, jac...
13   [rock, hard, varlop, rare, power, handsom, jac...
14   [rock, hard, varlop, rare, power, handsom, jac...
Name: Tweet, dtype: object
```

```
In [284... #combine the new tokenized words into full sentences
for i in range(len(tokens)):
    tokens[i]=" ".join(tokens[i])
twitter['Tweet']=tokens
twitter.head()
```

Out[284]:

	ID	Topic	Sentiment	Tweet
0	2401	Borderlands	Positive	get borderland will murder
1	2401	Borderlands	Positive	come border will kill
2	2401	Borderlands	Positive	get borderland will kill
3	2401	Borderlands	Positive	come borderland will murder
4	2401	Borderlands	Positive	get borderland will murder

In [285...

Exploratory Data Analysis

Cell In[285], line 1
Exploratory Data Analysis

SyntaxError: invalid syntax

In [286...

```
# visualise the frequent words
all_words = " ".join([sentence for sentence in twitter['Tweet']])

from wordcloud import WordCloud
wordcloud = WordCloud(width=800, height=500, random_state=42, max_font_size=100).generate(all_words)

# plot the graph
plt.figure(figsize=(15, 8))
plt.imshow(wordcloud, interpolation='bilinear') # Correct the interpolation parameter
plt.axis('off')
plt.show()
```



In [287...

```
#Displaying Postive words

# Frequent words visualizaatoion for +ve
all_words = " ".join([sentence for sentence in twitter['Tweet'][twitter['Sentiment']=='Positive']])

from wordcloud import WordCloud
wordcloud = WordCloud(width=800, height=500, random_state=42, max_font_size=100).generate(all_words)

# plot the graph
plt.figure(figsize=(15, 8))
plt.imshow(wordcloud, interpolation='bilinear') # Correct the interpolation parameter
plt.axis('off')
plt.show()
```


#Displaying Neutral words

[illegible]

```
from collections import Counter
```

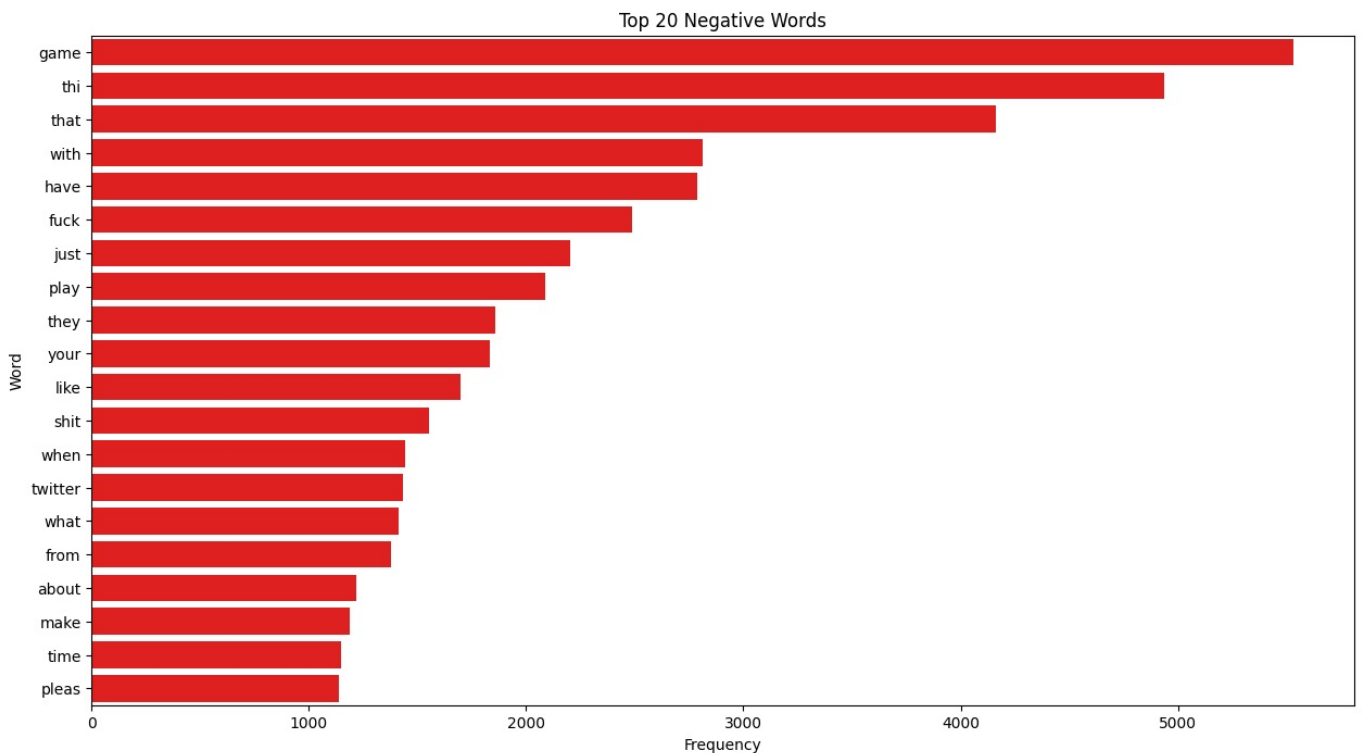
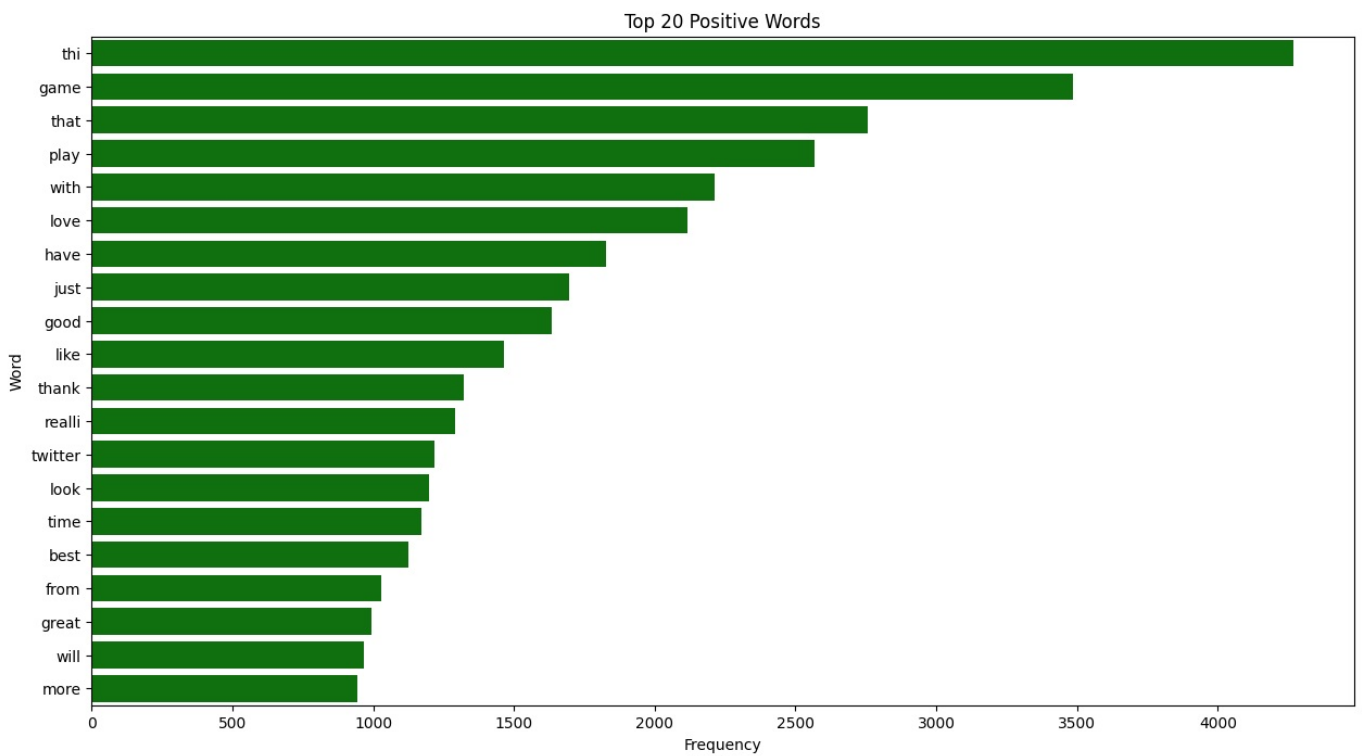
```
# Calculate word frequencies
positive_words = " ".join([sentence for sentence in twitter['Tweet'][twitter['Sentiment'] == 'Positive']])
negative_words = " ".join([sentence for sentence in twitter['Tweet'][twitter['Sentiment'] == 'Negative']])

positive_freq_dist = Counter(positive_words.split())
negative_freq_dist = Counter(negative_words.split())

# Convert to DataFrame for easier plotting
positive_df = pd.DataFrame(positive_freq_dist.items(), columns=['Word', 'Frequency']).sort_values(by='Frequency')
negative_df = pd.DataFrame(negative_freq_dist.items(), columns=['Word', 'Frequency']).sort_values(by='Frequency')

# Plotting
plt.figure(figsize=(15, 8))
sns.barplot(data=positive_df, x='Frequency', y='Word', color='green')
plt.title('Top 20 Positive Words')
plt.show()

plt.figure(figsize=(15, 8))
sns.barplot(data=negative_df, x='Frequency', y='Word', color='red')
plt.title('Top 20 Negative Words')
plt.show()
```

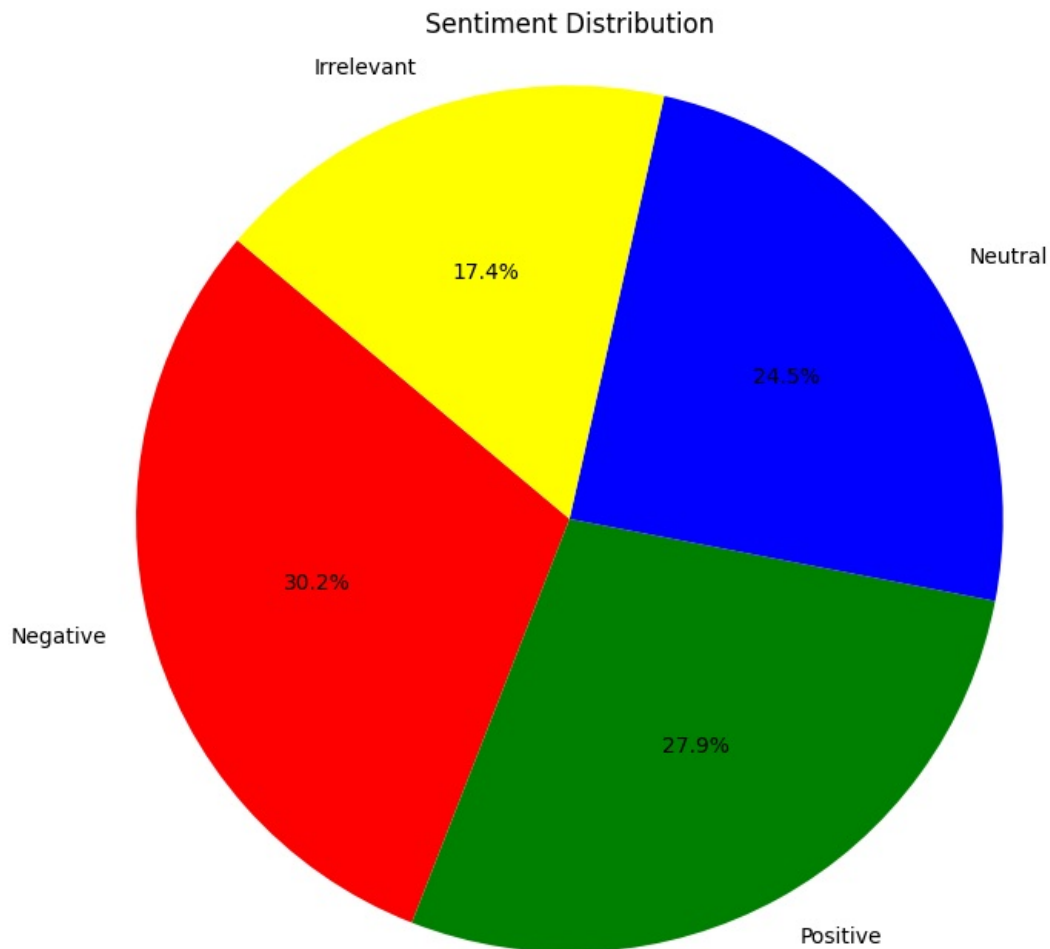



```
In [291]: # Check the data to confirm the sentiment values
print(twitter['Sentiment'].value_counts())

# Corrected Pie Chart for sentiment distribution
label_counts = twitter['Sentiment'].value_counts()
labels = label_counts.index.tolist() # Automatically get the labels from the data

plt.figure(figsize=(8, 8))
plt.pie(label_counts, labels=labels, autopct='%1.1f%%', startangle=140, colors=['red', 'green', 'blue', 'yellow'])
plt.title('Sentiment Distribution')
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
```

```
Sentiment
Negative    22542
Positive    20832
Neutral     18318
Irrelevant  12990
Name: count, dtype: int64
```



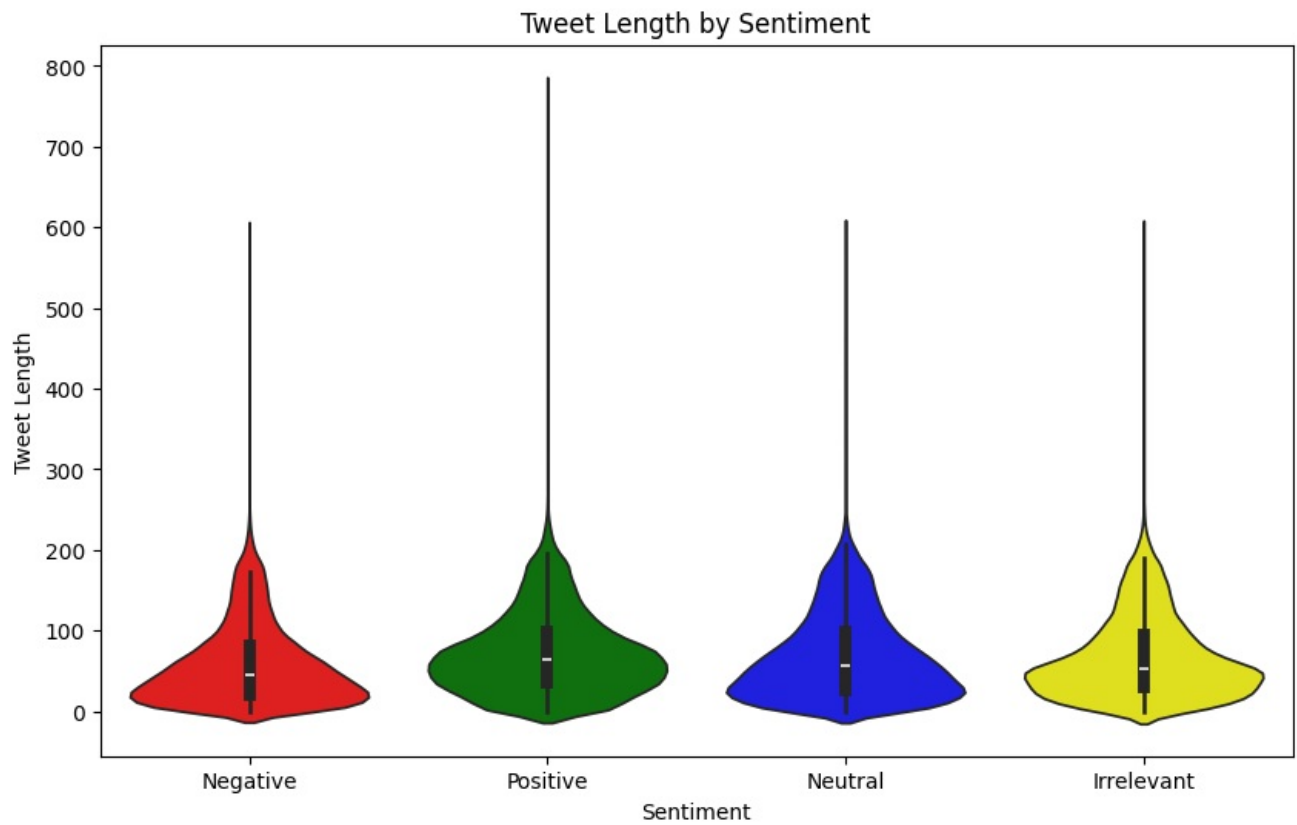
```
In [292... # Calculate tweet length (number of characters)
twitter['tweet_length'] = twitter['Tweet'].apply(len)

# Violin plot for tweet length by sentiment
plt.figure(figsize=(10, 6))
sns.violinplot(x='Sentiment', y='tweet_length', data=twitter, palette=['red', 'green', 'blue', 'yellow'])
plt.title('Tweet Length by Sentiment')
plt.xlabel('Sentiment')
plt.ylabel('Tweet Length')
plt.xticks([0, 1, 2, 3], ['Negative', 'Positive', 'Neutral', 'Irrelevant'])
plt.show()
```

C:\Users\DELL\AppData\Local\Temp\ipykernel_35236\3402404594.py:6: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

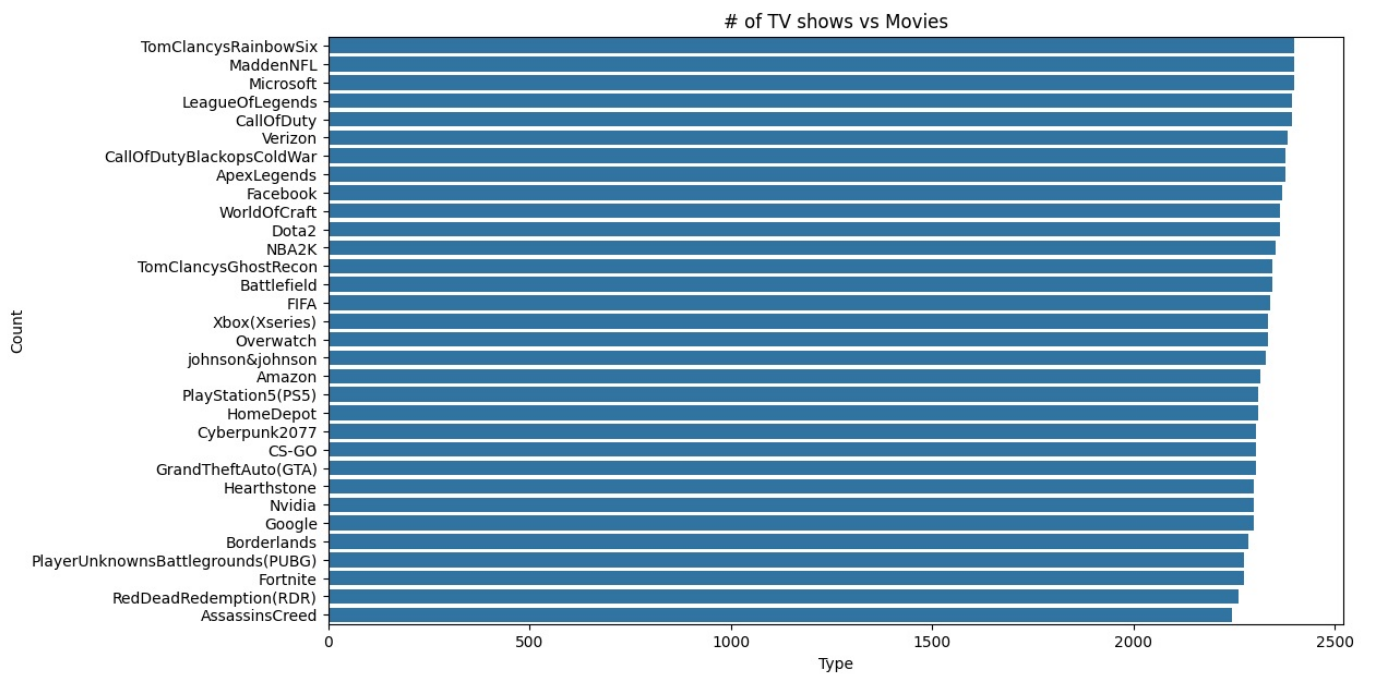
```
sns.violinplot(x='Sentiment', y='tweet_length', data=twitter, palette=['red', 'green', 'blue', 'yellow'])
```

```
In [293...] topic_types = twitter['Topic'].value_counts()
topic_types

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

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



```
In [294...] #splitting data into train and test part
from sklearn.model_selection import train_test_split

# Split your data into training and validation sets
x_train, x_test, y_train, y_test = train_test_split(twitter['Tweet'], twitter['Sentiment'], train_size=0.85, random_state=42)
len(x_train), len(x_test)
```

Out[294...] (63479, 11203)

```
In [295...] ##extracting features
from sklearn.feature_extraction.text import TfidfVectorizer

vector = TfidfVectorizer()
vector.fit(x_train)

# Get the feature names
Feature_Names = vector.get_feature_names_out()

print("number of feature words: ", len(Feature_Names))
```

number of feature words: 19621

```
In [296...] from sklearn.feature_extraction.text import TfidfVectorizer

# Converting text data to lowercase before vectorization
x_train = [text.lower() for text in x_train]

# Creating vectorsizer
vector = TfidfVectorizer()

# Transforming lowercase text data
x_train = vector.fit_transform(x_train)

# Transforming testing data in same manner
x_test = [text.lower() for text in x_test]
x_test = vector.transform(x_test)
```

```
In [297...] #Encoding data

from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression
# Create a LabelEncoder object
lb = LabelEncoder()
y_train = lb.fit_transform(y_train)
y_test = lb.fit_transform(y_test)
#Performing Logistic Regression
from sklearn.linear_model import LogisticRegression

LR = LogisticRegression(random_state=42)

# Fit the model to your training data
LR.fit(x_train, y_train)
```

```
C:\Users\DELL\anaconda3\Lib\site-packages\sklearn\linear_model\_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

Increase the number of iterations (max_iter) or scale the data as shown in:
<https://scikit-learn.org/stable/modules/preprocessing.html>
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
n_iter_i = check_optimize_result()

Out[297..

```
LogisticRegression
LogisticRegression(random_state=42)
```

In [298..

```
Train_Accuracy = LR.score(x_train , y_train)

LR_Predicted = LR.predict(x_test)

Test_Accuracy = accuracy_score(y_test , LR_Predicted)

Test_Precision = precision_score(y_test , LR_Predicted , average='weighted')
Test_Recall = recall_score(y_test , LR_Predicted , average='weighted')
Test_F1score = f1_score(y_test , LR_Predicted , average='weighted')
print(f"The training accuracy for logistic regression : {(Train_Accuracy*100):0.2f}%\n")
print(f"The testing accuracy for logistic regression : {(Test_Accuracy*100):0.2f}%\n")
print(f"The precision for logistic regression : {Test_Precision:0.2f}\n")
print(f"The recall for logistic regression : {Test_Recall:0.2f}\n")
print(f"The F1 score for logistic regression : {Test_F1score:0.2f}\n")
```

The training accuracy for logistic regression : 80.80%

The testing accuracy for logistic regression : 75.12%

The precision for logistic regression : 0.75

The recall for logistic regression : 0.75

The F1 score for logistic regression : 0.75

In [299..

```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import confusion_matrix

LR_Confusion_Matrix = confusion_matrix(y_test , LR_Predicted)
sns.heatmap(LR_Confusion_Matrix, annot=True,fmt='4g')
plt.show()
```



In [332..

```
#making predictions
test_twitter = pd.read_csv('C:/Users/DELL/Downloads/twitter_validation.csv',names=['ID', 'Topic', 'Sentiment',
```

In [333..

```
test_twitter['Topic'] = test_twitter['Topic'].astype(str)
test_twitter['Sentiment'] = test_twitter['Sentiment'].astype(str)
test_twitter['Tweet'] = test_twitter['Tweet'].astype(str)
test_twitter.head()
```

Out[333...	ID	Topic	Sentiment	Tweet
0	3364	Facebook	Irrelevant	I mentioned on Facebook that I was struggling ...
1	352	Amazon	Neutral	BBC News - Amazon boss Jeff Bezos rejects clai...
2	8312	Microsoft	Negative	@Microsoft Why do I pay for WORD when it funct...
3	4371	CS-GO	Negative	CSGO matchmaking is so full of closet hacking,...
4	4433	Google	Neutral	Now the President is slapping Americans in the...

In [334... *# Define the twitter_Review dictionary*

```
twitter_Review = {
    'Positive': 0,
    'Neutral': 1,
    'Negative': 2,
    'Irrelevant':3,
}
```

In [335... **def** getlabel(n):
for label, code **in** twitter_Review.items():
 if code == n:
 return label

In [336... label = getlabel(1)

In [341... **import** pandas **as** pd
from sklearn.feature_extraction.text **import** TfidfVectorizer
from sklearn.linear_model **import** LogisticRegression
from sklearn.preprocessing **import** LabelEncoder
from sklearn.metrics **import** accuracy_score, precision_score, recall_score, f1_score
from tqdm **import** tqdm
import seaborn **as** sns
import matplotlib.pyplot **as** plt

```
def make_predictions(row):
    # Randomly sample 10 rows from the data
    random_data = row.sample(n=10)
    random_tweets = random_data['Tweet'].values

    # Check if random sampling worked correctly
    print("Randomly sampled tweets:")
    print(random_tweets)

    # Transform the cleaned tweets using the vectorizer
    x_test = vector.transform(random_tweets).toarray()
    y_test = random_data['Sentiment'].values

    # Check if transformation worked correctly
    print("Transformed test data:")
    print(x_test)

    # Encode y_test
    y_test_encoded = lb.transform(y_test)

    # Make predictions using the Logistic Regression model
    lr_pred = LR.predict(x_test)

    # Check if prediction worked correctly
    print("Predictions:")
    print(lr_pred)

    # Print the results for each tweet
    for i in tqdm(range(10)):
        print(f"The original tweet: {random_tweets[i]}\n")
        print(f"The original Sentiment: {getlabel(y_test_encoded[i])}\n")
        print(f"The lr prediction is: {getlabel(lr_pred[i])}\n")
        print('-' * 120)

# Call the function with your data
make_predictions(test_twitter)
```


Randomly sampled tweets:

['Watching NVIDIA position itself as not just a leading hardware manufacturer but also providing meaningful software to consumers is a remarkable thing of beauty. What an incredibly lead company with clear focus and goals. Well done @nvidia.'

'@roundedtiktak thank you #Fortnite #XboxShare pic.twitter.com/zWq0rFk0JN'

'Really enjoying the new 50bmg in #ModernWarfare , find it very satisfying hitting little clips like this (im not the best sniper but enjoy it) #CallofDuty\n\nThe Build I use is:\nDictator barrel\nTac laser\nTac-Wrap stock\nStippled grip\nSleight of hand https://t.co/xZbblCYUZI'

'Transvaginal mesh: Johnson & Johnson fined \$344m for deceptive marketing to women\n\ntheguardian.com/business/2020/...'

"Glad I got my #Cyberpunk2077 strategy guide pre-ordered, I was going to do it last week but it was sold out! So I'm glad it back in stock pic.twitter.com/CrmoICs3w2"

"The Nigeria national team has been ranked as the 29th best team in the World and 3rd in Africa in the latest FIFA World rankings. □□□\n\nIt's the first time the team has made the top 30 since May 2013. @Kano, Nigeria instagram.com/p/CFPZvT0n8Ab/..."

'I was inspired by @miserabletop to get assassins creed on switch and now I couldn't be more excited to leave the party early to go home and play Black Flag pic.twitter.com/yHrybJDYnn'

"GIVE @2K THE ABILITIES TO MAKE SIM FOOTBALL GAMES NOW! EFF THAT CONTRACT. @EAMaddenNFL Y'ALL HAVE SWINDLED US ENOUGH! @NFL TELL THEM TO GIVE IT UP, THEY DON'T DESERVE THE EXCLUSIVITY LICENSE ANYMORE!!!!!"

'bhopping in csgo is so cozy' 'Yo the new event on PUBG is weird!']

Transformed test data:

```
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]
```

Predictions:

```
[3 0 3 2 3 0 3 1 3 2]
```

100%|██████████| 10/10 [00:00<?, ?it/s]

The original tweet: Watching NVIDIA position itself as not just a leading hardware manufacturer but also providing meaningful software to consumers is a remarkable thing of beauty. What an incredibly lead company with clear focus and goals. Well done @nvidia.

The original Sentiment: Irrelevant

The lr prediction is: Irrelevant

The original tweet: @roundedtiktak thank you #Fortnite #XboxShare pic.twitter.com/zWq0rFk0JN

The original Sentiment: Positive

The lr prediction is: Positive

The original tweet: Really enjoying the new 50bmg in #ModernWarfare , find it very satisfying hitting little clips like this (im not the best sniper but enjoy it) #CallofDuty

The Build I use is:
Dictator barrel
Tac laser
Tac-Wrap stock
Stippled grip
Sleight of hand https://t.co/xZbblCYUZI

The original Sentiment: Irrelevant

The lr prediction is: Irrelevant

The original tweet: Transvaginal mesh: Johnson & Johnson fined \$344m for deceptive marketing to women
theguardian.com/business/2020/...

The original Sentiment: Negative

The lr prediction is: Negative

The original tweet: Glad I got my #Cyberpunk2077 strategy guide pre-ordered, I was going to do it last week but it was sold out! So I'm glad it back in stock pic.twitter.com/CrmoICs3w2

The original Sentiment: Irrelevant

The lr prediction is: Irrelevant

The original tweet: The Nigeria national team has been ranked as the 29th best team in the World and 3rd in Africa in the latest FIFA World rankings. ☐☐☐

It's the first time the team has made the top 30 since May 2013. @ Kano, Nigeria [instagram.com/p/CFPZvT0n8Ab/...](https://www.instagram.com/p/CFPZvT0n8Ab/)

The original Sentiment: Positive

The lr prediction is: Positive

The original tweet: I was inspired by @miserabletop to get assassins creed on switch and now I couldn't be more excited to leave the party early to go home and play Black Flag pic.twitter.com/yHrybJDYnn

The original Sentiment: Irrelevant

The lr prediction is: Irrelevant

The original tweet: GIVE @2K THE ABILITIES TO MAKE SIM FOOTBALL GAMES NOW! EFF THAT CONTRACT. @EAMaddenNFL Y'ALL HAVE SWINDLED US ENOUGH! @NFL TELL THEM TO GIVE IT UP, THEY DON'T DESERVE THE EXCLUSIVITY LICENSE ANYMORE!!!!

The original Sentiment: Neutral

The lr prediction is: Neutral

The original tweet: bhopping in csgo is so cozy

The original Sentiment: Irrelevant

The lr prediction is: Irrelevant

The original tweet: Yo the new event on PUBG is weird!

The original Sentiment: Positive

The lr prediction is: Negative

In []: