Data Collection

In [ ]:

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
comments=pd.read_csv(r'C:\Users\lenovo\Downloads\Youtube_project_shan_singh_
```

```
C:\Users\lenovo\AppData\Local\Temp\ipykernel_7292\3572458290.py:1: FutureW
arning: The error_bad_lines argument has been deprecated and will be remov
ed in a future version. Use on_bad_lines in the future.


  comments=pd.read_csv(r'C:\Users\lenovo\Downloads\Youtube_project_shan_si
ngh_Udemy (1)/UScomments.csv', error_bad_lines=False)
b'Skipping line 41589: expected 4 fields, saw 11\nSkipping line 51628: exp
ected 4 fields, saw 7\nSkipping line 114465: expected 4 fields, saw 5\n'
b'Skipping line 142496: expected 4 fields, saw 8\nSkipping line 189732: ex
pected 4 fields, saw 6\nSkipping line 245218: expected 4 fields, saw 7\n'
b'Skipping line 388430: expected 4 fields, saw 5\n'
C:\Users\lenovo\AppData\Local\Temp\ipykernel_7292\3572458290.py:1: DtypeWa
rning: Columns (2,3) have mixed types. Specify dtype option on import or s
et low_memory=False.
  comments=pd.read_csv(r'C:\Users\lenovo\Downloads\Youtube_project_shan_si
ngh_Udemy (1)/UScomments.csv', error_bad_lines=False)
```

In [ ]:

In [ ]:

In [ ]:

In [3]:
```
comments.head()
```

Out[3]:

	video_id	comment_text	likes	replies
0	XpVt6Z1Gjjo	Logan Paul it's yo big day !!!!!	4	0
1	XpVt6Z1Gjjo	I've been following you from the start of your...	3	0
2	XpVt6Z1Gjjo	Say hi to Kong and maverick for me	3	0
3	XpVt6Z1Gjjo	MY FAN . attendance	3	0
4	XpVt6Z1Gjjo	trending 😉	3	0

In [ ]:

In [4]:
```
comments.isnull().sum()
```

Out[4]:
```
video_id         0
comment_text    25
likes            0
replies          0
dtype: int64
```

In [5]:
```
comments.dropna(inplace=True)
```

In [6]:
```
comments.isnull().sum()
```

Out[6]:
```
video_id        0
comment_text    0
likes           0
replies         0
dtype: int64
```

In [ ]:

Perform sentiment Analysis

In [ ]:

In [ ]:

In [7]:
```
!pip install textblob
```

```
Requirement already satisfied: textblob in c:\users\lenovo\anaconda3\lib\s
ite-packages (0.18.0.post0)

WARNING: There was an error checking the latest version of pip.


Requirement already satisfied: nltk>=3.8 in c:\users\lenovo\anaconda3\lib
\site-packages (from textblob) (3.8.1)
Requirement already satisfied: click in c:\users\lenovo\anaconda3\lib\site
-packages (from nltk>=3.8->textblob) (8.0.4)
Requirement already satisfied: joblib in c:\users\lenovo\anaconda3\lib\sit
e-packages (from nltk>=3.8->textblob) (1.1.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\lenovo\anaconda
3\lib\site-packages (from nltk>=3.8->textblob) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\lenovo\anaconda3\lib\site-
packages (from nltk>=3.8->textblob) (4.64.1)
Requirement already satisfied: colorama in c:\users\lenovo\anaconda3\lib\s
ite-packages (from click->nltk>=3.8->textblob) (0.4.5)
```

In [8]:
```
from textblob import TextBlob
```

In [9]:
```
comments.head(6)
```

Out[9]:

	video_id	comment_text	likes	replies
0	XpVt6Z1Gjjo	Logan Paul it's yo big day !!!!!	4	0
1	XpVt6Z1Gjjo	I've been following you from the start of your...	3	0
2	XpVt6Z1Gjjo	Say hi to Kong and maverick for me	3	0
3	XpVt6Z1Gjjo	MY FAN . attendance	3	0
4	XpVt6Z1Gjjo	trending 😉	3	0
5	XpVt6Z1Gjjo	#1 on trending AYYEEEEE	3	0

In [10]:
```
TextBlob("Logan Paul it's yo big day !!!!!").sentiment.polarity
```

Out[10]:
```
0.0
```

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

Out[11]:
```
(691375, 4)
```

In [12]:
```
sample_df =comments[0:1000]
```

In [13]:
```
sample_df.shape
```

Out[13]:
```
(1000, 4)
```

In [ ]:
```

```

In [14]:
```
polarity=[]

for comment in comments['comment_text']:
    try:
        polarity.append(TextBlob(comment).sentiment.polarity)
    except:
        polarity.append(0)
```

In [15]:
```
len(polarity)
```

Out[15]:
```
691375
```

In [ ]:
```

```

In [16]:
```
comments['polarity'] = polarity
```

In [ ]:
```

```

In [17]:
```
comments.head(5)
```

Out[17]:

	video_id	comment_text	likes	replies	polarity
0	XpVt6Z1Gjjo	Logan Paul it's yo big day !!!!!	4	0	0.0
1	XpVt6Z1Gjjo	I've been following you from the start of your...	3	0	0.0
2	XpVt6Z1Gjjo	Say hi to Kong and maverick for me	3	0	0.0
3	XpVt6Z1Gjjo	MY FAN . attendance	3	0	0.0
4	XpVt6Z1Gjjo	trending 😉	3	0	0.0

In [ ]:
```
```

Wordcloud analysis of data

In [ ]:
```
```

In [ ]:
```
```

In [18]:
```
filter1 = comments['polarity']==1
```

In [19]:
```
comments_positive = comments[filter1]
```

In [ ]:
```
```

In [ ]:
```
```

In [20]:
```
filter2 = comments['polarity']==-1
```

In [21]:
```
comments_negative = comments[filter2]
```

In [22]:
```
comments_negative.head(5)
```

Out[22]:

	video_id	comment_text	likes	replies	polarity
512	8wNr-NQImFg	BEN CARSON IS THE MAN!!!!! THEY HATE HIM CAUSE...	0	0	-1.0
562	8wNr-NQImFg	Well… The brain surgeon Ben Carson just proved...	0	0	-1.0
952	Ayb_2qbZHm4	WHY DID YOU MAKE FURRY FORCE?! SO NASTY!!!	0	0	-1.0
1371	vu_9muoxT50	WTF BRUH!!!!!!	0	0	-1.0
1391	vu_9muoxT50	cheeseus christ thats insane!!!	0	0	-1.0

In [ ]:
```
```

In [23]:
```
comments_positive.head(5)
```

Out[23]:

	video_id	comment_text	likes	replies	polarity
64	XpVt6Z1Gjjo	yu are the best	1	0	1.0
156	cLdxuaxaQwc	Power is the disease. Care is the cure. Keep...	0	0	1.0
227	WYYvHb03Eog	YAS Can't wait to get it! I just need to sell ...	0	0	1.0
307	sjlHnJvXdQs	This is priceless	0	0	1.0
319	sjlHnJvXdQs	Summed up perfectly	0	0	1.0

In [ ]:
```
```

In [ ]:
```
```

In [24]:
```
!pip install wordcloud
```

```
Requirement already satisfied: wordcloud in c:\users\lenovo\anaconda3\lib
\site-packages (1.9.3)
Requirement already satisfied: numpy>=1.6.1 in c:\users\lenovo\anaconda3\l
ib\site-packages (from wordcloud) (1.21.5)
Requirement already satisfied: pillow in c:\users\lenovo\anaconda3\lib\sit
e-packages (from wordcloud) (9.2.0)
Requirement already satisfied: matplotlib in c:\users\lenovo\anaconda3\lib
\site-packages (from wordcloud) (3.5.2)
Requirement already satisfied: packaging>=20.0 in c:\users\lenovo\anaconda
3\lib\site-packages (from matplotlib->wordcloud) (21.3)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\lenovo\ana
conda3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\lenovo\anacond
a3\lib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: cycler>=0.10 in c:\users\lenovo\anaconda3\l
ib\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\lenovo\anacon
da3\lib\site-packages (from matplotlib->wordcloud) (1.4.2)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\lenovo\anacon
da3\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: six>=1.5 in c:\users\lenovo\anaconda3\lib\s
ite-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)

WARNING: There was an error checking the latest version of pip.
```

In [ ]:
```
```

In [25]:
```
from wordcloud import WordCloud , STOPWORDS
```

In [26]:
```
set(STOPWORDS)
```

Out[26]:
```
{'a',
 'about',
 'above',
 'after',
 'again',
 'against',
 'all',
 'also',
 'am',
 'an',
 'and',
 'any',
 'are',
 "aren't",
 'as',
 'at',
 'be',
 'because',
 'been',
```

In [ ]:
```

```

In [27]:
```
comments['comment_text']
```

Out[27]:
```
0                          Logan Paul it's yo big day ‼️‼️‼️
1         I've been following you from the start of your...
2                       Say hi to Kong and maverick for me
3                                  MY FAN . attendance
4                                           trending 😉
                              ...
691395                                              Лучшая
691396    qu'est ce que j'aimerais que tu viennes à Roan...
691397                            Ven a mexico! 😍 te amo LP
691398                                    Islığı yeter...
691399    Kocham tą piosenkę😍❤❤❤byłam zakochana po uszy ...
Name: comment_text, Length: 691375, dtype: object
```

In [28]:
```
type(comments['comment_text'])
```

Out[28]:
```
pandas.core.series.Series
```

In [29]:
```
Total_Comments_Positive=' '.join(comments_positive['comment_text'])
```

In [30]:
```
wordcloud=WordCloud(stopwords=set(STOPWORDS)).generate(Total_Comments_Positi
```

In [31]:
```
plt.imshow(wordcloud)
plt.axis('off')
```

Out[31]: (-0.5, 399.5, 199.5, -0.5)

In [ ]:

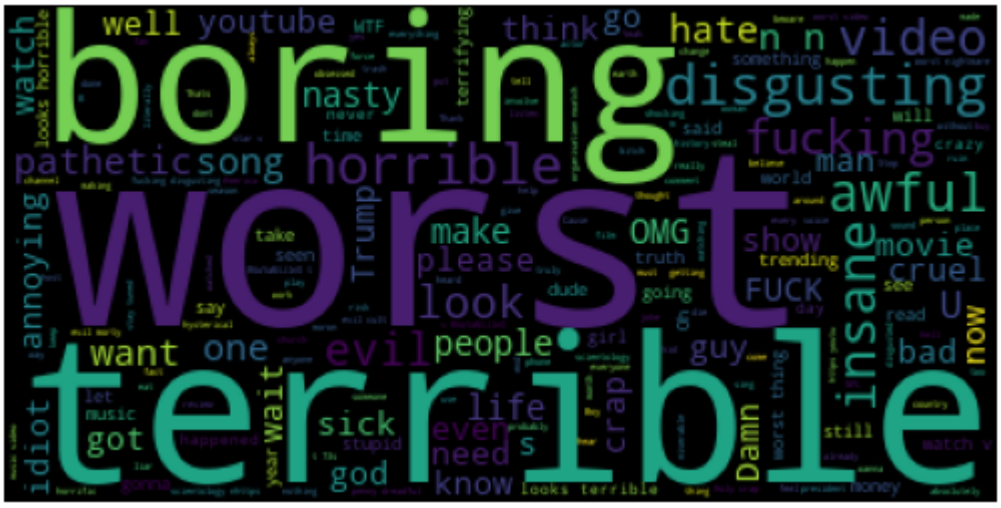
In [ ]:

In [32]:
```
Total_Comments_Negative=' '.join(comments_negative['comment_text'])
```

In [33]:
```
wordcloud2 = WordCloud(stopwords=set(STOPWORDS)).generate(Total_Comments_Neg
```

In [34]:
```
plt.imshow(wordcloud2 )
plt.axis('off')
```

Out[34]: (-0.5, 399.5, 199.5, -0.5)



In [ ]:

In [35]:
```
!pip install emoji==2.2.0
```

```
Requirement already satisfied: emoji==2.2.0 in c:\users\lenovo\anaconda3\l
ib\site-packages (2.2.0)

WARNING: There was an error checking the latest version of pip.
```

In [ ]:
```

```

Emoji Analysis

In [ ]:
```

```

In [36]:
```
import emoji
```

In [37]:
```
emoji.__version__
```

Out[37]:
```
'2.2.0'
```

In [38]:
```
comments['comment_text'].head(6)
```

Out[38]:
```
0                    Logan Paul it's yo big day !!!!!
1    I've been following you from the start of your...
2                   Say hi to Kong and maverick for me
3                                MY FAN . attendance
4                                          trending 😉
5                             #1 on trending AYYEEEEE
Name: comment_text, dtype: object
```

In [ ]:
```

```

In [39]:
```
comment = 'trending 😉'
```

In [40]:
```
[char for char in comment if char in emoji.EMOJI_DATA]
```

Out[40]:
```
['😉']
```

In [ ]:
```

```

In [41]:
```
emoji_list =[]
for char in comment:
    if char in emoji.EMOJI_DATA:
        emoji_list.append(char)
```

In [42]:
```
emoji_list
```

Out[42]:
```
['😉']
```

In [ ]:
```

```

In [43]:
```
all_emoji_list =[]

for comment in comments['comment_text'].dropna():
    for char in comment:
        if char in emoji.EMOJI_DATA:
            all_emoji_list.append(char)
```

In [44]:
```
all_emoji_list[0:10]
```

Out[44]: ['‼', '‼', '‼', '😉', '😭', '👍', '', '', '😍', '💋']

In [ ]:

In [45]:
```
from collections import Counter
```

In [46]:
```
Counter(all_emoji_list).most_common(10)
```

Out[46]:
```
[('😂', 36987),
 ('😍', 33453),
 ('♥', 31119),
 ('🔥', 8694),
 ('😭', 8398),
 ('👏', 5719),
 ('😘', 5545),
 ('👍', 5476),
 ('💔', 5359),
 ('💞', 5147)]
```

In [ ]:

In [47]:
```
Counter(all_emoji_list).most_common(10)[0]
```

Out[47]: ('😂', 36987)

In [48]:
```
Counter(all_emoji_list).most_common(10)[0][0]
```

Out[48]: '😂'

In [49]:
```
Counter(all_emoji_list).most_common(10)[2][0]
```

Out[49]: '♥'

In [50]:
```
emojis= [Counter(all_emoji_list).most_common(10)[i][0]for i in range(10)]
```

In [ ]:

In [51]:
```
frequency= [Counter(all_emoji_list).most_common(10)[i][1]for i in range(10)]
```

In [52]:
```
frequency
```

Out[52]: [36987, 33453, 31119, 8694, 8398, 5719, 5545, 5476, 5359, 5147]

In [ ]:

In [53]:
```
import plotly.graph_objs as go
from plotly.offline import iplot
```

In [54]:
```
trace = go.Bar(x=emojis, y=frequency)
```

In [105]:
```
iplot([trace])
```

In [ ]:
```

```

Collect Entire data of Youtube

In [ ]:
```

```

In [56]:
```
import os
```

In [57]:
```
files=os.listdir(r'C:\Users\lenovo\Downloads\Youtube_project_shan_singh_Uder
```

In [58]:
```
files
```

Out[58]:
```
['CAvideos.csv',
 'CA_category_id.json',
 'DEvideos.csv',
 'DE_category_id.json',
 'FRvideos.csv',
 'FR_category_id.json',
 'GBvideos.csv',
 'GB_category_id.json',
 'INvideos.csv',
 'IN_category_id.json',
 'JPvideos.csv',
 'JP_category_id.json',
 'KRvideos.csv',
 'KR_category_id.json',
 'MXvideos.csv',
 'MX_category_id.json',
 'RUvideos.csv',
 'RU_category_id.json',
 'USvideos.csv',
 'US_category_id.json']
```

In [ ]:
```

```

In [59]:
```
files_csv =[file for file in files if '.csv' in file]
```

In [60]:
```
files_csv
```

Out[60]:
```
['CAvideos.csv',
 'DEvideos.csv',
 'FRvideos.csv',
 'GBvideos.csv',
 'INvideos.csv',
 'JPvideos.csv',
 'KRvideos.csv',
 'MXvideos.csv',
 'RUvideos.csv',
 'USvideos.csv']
```

In [ ]:
```

```

In [61]:
```
import warnings
from warnings import filterwarnings
filterwarnings('ignore')
```

In [ ]:
```

```

In [62]:
```
full_df = pd.DataFrame()
path = r'C:\Users\lenovo\Downloads\Youtube_project_shan_singh_Udemy (1)\addi


for file in files_csv:
    current_df = pd.read_csv(path+'/'+file , encoding='iso-8859-1' , error_b

    full_df = pd.concat([full_df, current_df] , ignore_index=True)
```

In [63]:
```
full_df.shape
```

Out[63]:
```
(375942, 16)
```

In [ ]:
```

```

How to export your data into (csv, json, db)

In [ ]:
```

```

In [64]:
```
full_df[full_df.duplicated()].shape
```

Out[64]:
```
(36417, 16)
```

In [65]:
```
full_df = full_df.drop_duplicates()
```

In [66]:
```
full_df.shape
```

Out[66]:
```
(339525, 16)
```

In [67]:
```
full_df[0:1000].to_csv(r'D:\Software\manas/youtube_sample.csv' ,index=False)
```

In [68]:
```
full_df[0:1000].to_json(r'D:\Software\manas/youtube_sample.json')
```

In [ ]:
```

```

In [ ]:
```

```

In [69]:
```
from sqlalchemy import create_engine
```

In [70]:
```
engine=create_engine(r'sqlite:///D:\Software\manas/youtube_sample.sqlite')
```

In [ ]:
```

```

In [71]:
```
full_df[0:1000].to_sql('Users', con=engine, if_exists='append')
```

Out[71]:
```
1000
```

In [ ]:
```

```

Which category has the maximum likes

In [ ]:

In [72]:
```
full_df.head(5)
```

Out[72]:

	video_id	trending_date	title	channel_title	category_id	publish_time	
0	n1WpP7iowLc	17.14.11	Eminem - Walk On Water (Audio) ft. BeyoncÃ©	EminemVEVO	10	2017-11-10T17:00:03.000Z	Em
1	0dBIkQ4Mz1M	17.14.11	PLUSH - Bad Unboxing Fan Mail	iDubbbzTV	23	2017-11-13T17:00:00.000Z	
2	5qpjK5DgCt4	17.14.11	Racist Superman | Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	2017-11-12T19:05:24.000Z	ra
3	d380meD0W0M	17.14.11	I Dare You: GOING BALD!?	nigahiga	24	2017-11-12T18:01:41.000Z	
4	2Vv-BfVoq4g	17.14.11	Ed Sheeran - Perfect (Official Music Video)	Ed Sheeran	10	2017-11-09T11:04:14.000Z	

In [ ]:

In [73]:
```
full_df['category_id'].unique()
```

Out[73]:
```
array([10, 23, 24, 25, 22, 26,  1, 28, 20, 17, 29, 15, 19,  2, 27, 43, 30,
       44], dtype=int64)
```

In [ ]:

In [74]:
```
json_df = pd.read_json(r'C:\Users\lenovo\Downloads\Youtube_project_shan_sing
```

In [75]:
```
json_df['items'][0]
```

Out[75]:
```
{'kind': 'youtube#videoCategory',
 'etag': '"m2yskBQFythfE4irbTIeOgYYfBU/Xy1mB4_yLrHy_BmKmPBggty2mZQ"',
 'id': '1',
 'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
  'title': 'Film & Animation',
  'assignable': True}}
```

In [76]:
```
json_df['items'][1]
```

Out[76]:
```
{'kind': 'youtube#videoCategory',
 'etag': '"m2yskBQFythfE4irbTIeOgYYfBU/UZ1oLIIz2dxIhO45ZTFR3a3NyTA"',
 'id': '2',
 'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
  'title': 'Autos & Vehicles',
  'assignable': True}}
```

In [ ]:
```

```

In [77]:
```
cat_dict={}
for item in json_df['items'].values:
    cat_dict[int(item['id'])]= item['snippet']['title']
```

In [78]:
```
cat_dict
```

Out[78]:
```
{1: 'Film & Animation',
 2: 'Autos & Vehicles',
 10: 'Music',
 15: 'Pets & Animals',
 17: 'Sports',
 18: 'Short Movies',
 19: 'Travel & Events',
 20: 'Gaming',
 21: 'Videoblogging',
 22: 'People & Blogs',
 23: 'Comedy',
 24: 'Entertainment',
 25: 'News & Politics',
 26: 'Howto & Style',
 27: 'Education',
 28: 'Science & Technology',
 29: 'Nonprofits & Activism',
 30: 'Movies',
 31: 'Anime/Animation',
 32: 'Action/Adventure',
 33: 'Classics',
 34: 'Comedy',
 35: 'Documentary',
 36: 'Drama',
 37: 'Family',
 38: 'Foreign',
 39: 'Horror',
 40: 'Sci-Fi/Fantasy',
 41: 'Thriller',
 42: 'Shorts',
 43: 'Shows',
 44: 'Trailers'}
```

In [79]:
```
full_df['category_name']= full_df['category_id'].map(cat_dict)
```

In [80]:
```
full_df.head(5)
```

Out[80]:

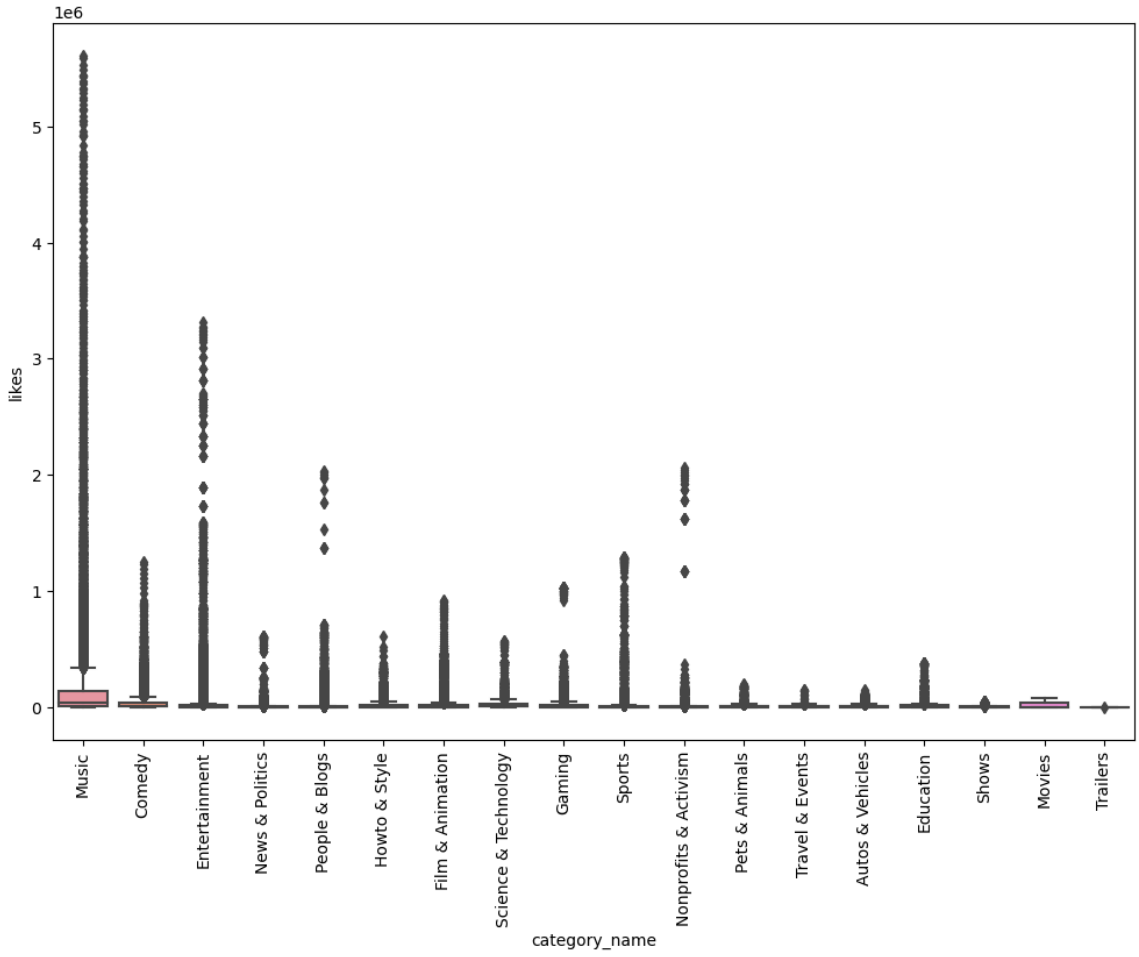
	video_id	trending_date	title	channel_title	category_id	publish_time	
0	n1WpP7iowLc	17.14.11	Eminem - Walk On Water (Audio) ft. BeyoncÃ©	EminemVEVO	10	2017-11-10T17:00:03.000Z	Em
1	0dBIkQ4Mz1M	17.14.11	PLUSH - Bad Unboxing Fan Mail	iDubbbzTV	23	2017-11-13T17:00:00.000Z	
2	5qpjK5DgCt4	17.14.11	Racist Superman | Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	2017-11-12T19:05:24.000Z	ra
3	d380meD0W0M	17.14.11	I Dare You: GOING BALD!?	nigahiga	24	2017-11-12T18:01:41.000Z	
4	2Vv-BfVoq4g	17.14.11	Ed Sheeran - Perfect (Official Music Video)	Ed Sheeran	10	2017-11-09T11:04:14.000Z	

In [ ]:

In [ ]:

In [81]:
```
plt.figure(figsize=(12,8))
sns.boxplot(x='category_name', y='likes' , data=full_df)
plt.xticks(rotation='vertical')
```

Out[81]:
```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
        17]),
 [Text(0, 0, 'Music'),
  Text(1, 0, 'Comedy'),
  Text(2, 0, 'Entertainment'),
  Text(3, 0, 'News & Politics'),
  Text(4, 0, 'People & Blogs'),
  Text(5, 0, 'Howto & Style'),
  Text(6, 0, 'Film & Animation'),
  Text(7, 0, 'Science & Technology'),
  Text(8, 0, 'Gaming'),
  Text(9, 0, 'Sports'),
  Text(10, 0, 'Nonprofits & Activism'),
  Text(11, 0, 'Pets & Animals'),
  Text(12, 0, 'Travel & Events'),
  Text(13, 0, 'Autos & Vehicles'),
  Text(14, 0, 'Education'),
  Text(15, 0, 'Shows'),
  Text(16, 0, 'Movies'),
  Text(17, 0, 'Trailers')])
```



In [ ]:

find out whether audience is engaged or not

In [ ]:

In [ ]:

In [82]:
```
full_df['like_rate'] = (full_df['likes']/full_df['views'])
full_df['dislike_rate'] = (full_df['dislikes']/full_df['views'])
full_df['Comment_count_rate'] = (full_df['comment_count']/full_df['views'])
```

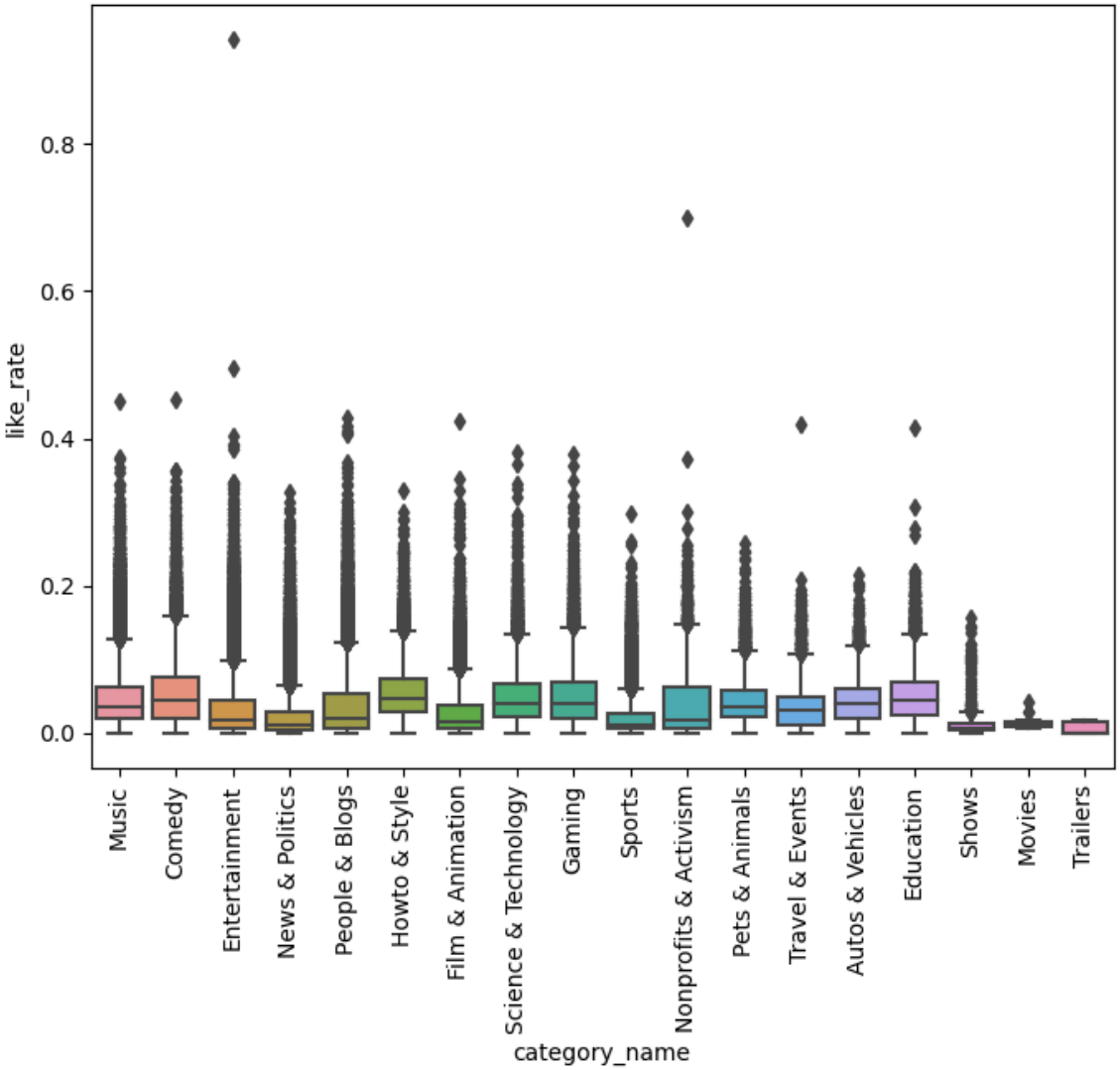
In [83]:
```
full_df.columns
```

Out[83]:
```
Index(['video_id', 'trending_date', 'title', 'channel_title', 'category_i
d',
       'publish_time', 'tags', 'views', 'likes', 'dislikes', 'comment_coun
t',
       'thumbnail_link', 'comments_disabled', 'ratings_disabled',
       'video_error_or_removed', 'description', 'category_name', 'like_rat
e',
       'dislike_rate', 'Comment_count_rate'],
      dtype='object')
```

In [ ]:

In [84]:
```
plt.figure(figsize=(8,6))
sns.boxplot(x='category_name', y='like_rate' , data=full_df)
plt.xticks(rotation='vertical')
plt.show
```

Out[84]: <function matplotlib.pyplot.show(close=None, block=None)>

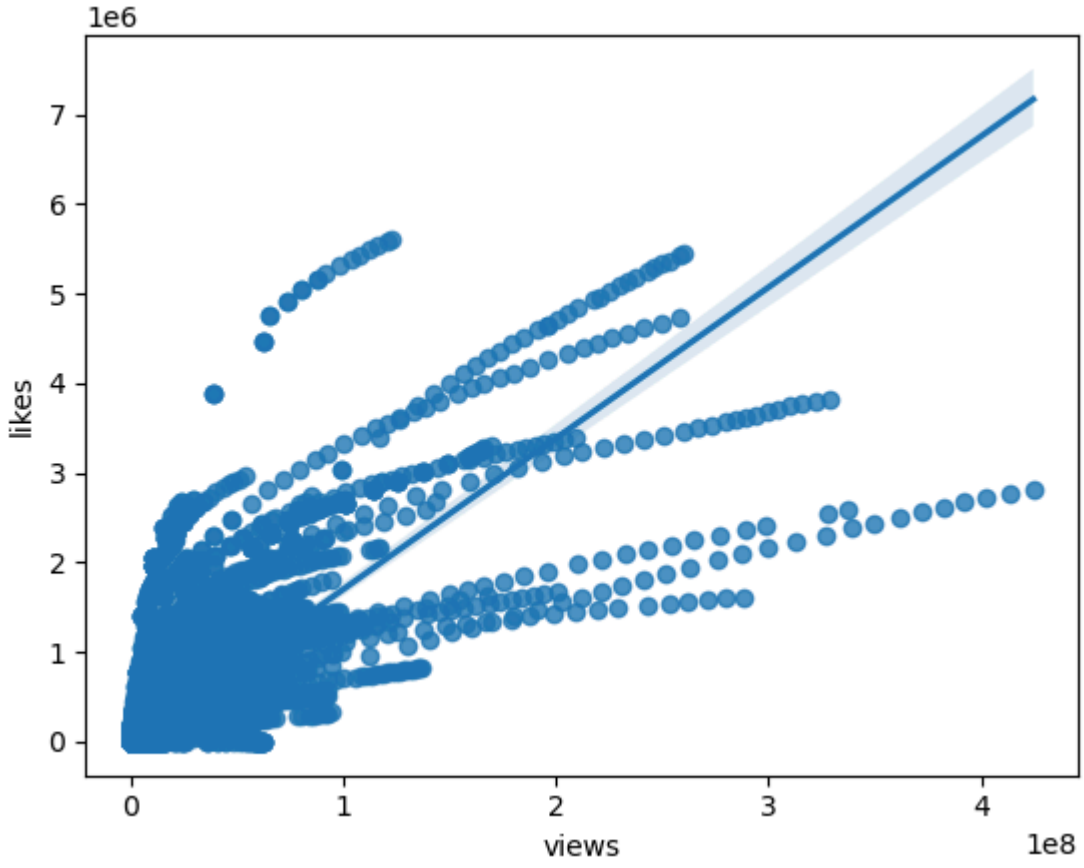


In [ ]:

In [ ]:

In [85]:
```
sns.regplot(x='views', y='likes', data=full_df)
```

Out[85]:
```
<AxesSubplot:xlabel='views', ylabel='likes'>
```



In [ ]:
```

```

In [86]:
```
full_df.columns
```

Out[86]:
```
Index(['video_id', 'trending_date', 'title', 'channel_title', 'category_i
d',
       'publish_time', 'tags', 'views', 'likes', 'dislikes', 'comment_coun
t',
       'thumbnail_link', 'comments_disabled', 'ratings_disabled',
       'video_error_or_removed', 'description', 'category_name', 'like_rat
e',
       'dislike_rate', 'Comment_count_rate'],
      dtype='object')
```

In [87]:
```
full_df[['views', 'likes', 'dislikes']].corr()
```

Out[87]:

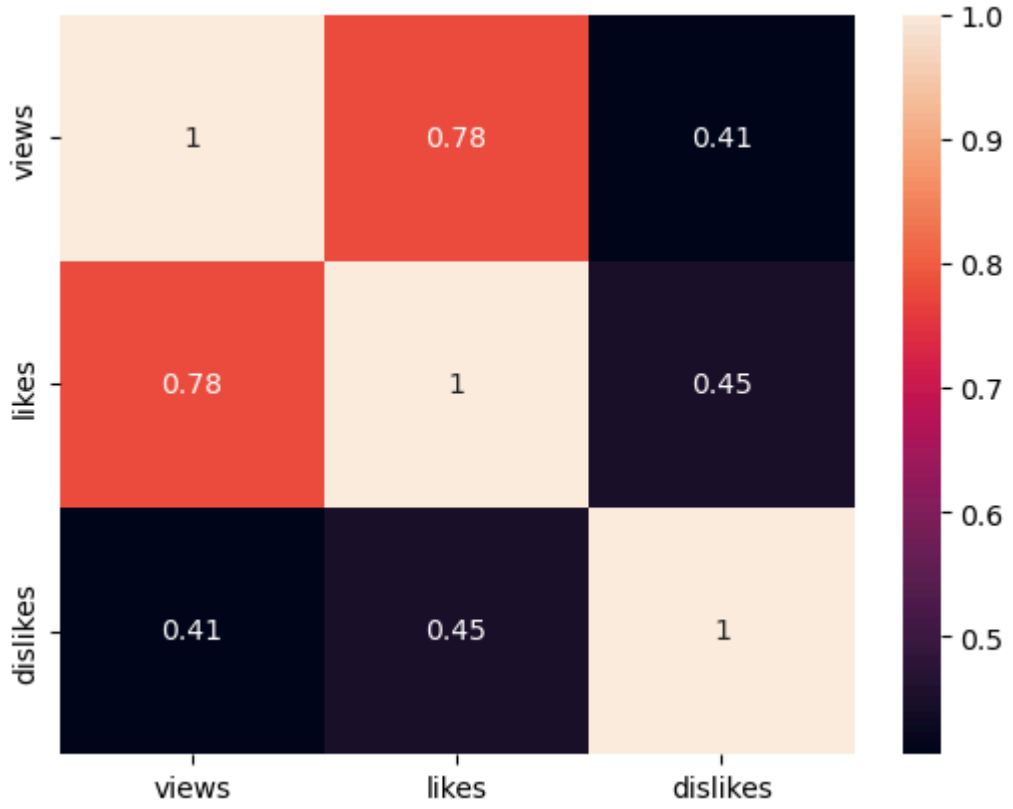
	views	likes	dislikes
views	1.000000	0.779531	0.405428
likes	0.779531	1.000000	0.451809
dislikes	0.405428	0.451809	1.000000

In [ ]:
```

```

In [88]:
```
sns.heatmap(full_df[['views', 'likes', 'dislikes']].corr() ,annot=True)
```

Out[88]: <AxesSubplot:>



In [ ]:

Which channels have the largest number of trending videos

In [ ]:

In [89]:
```
full_df.head(6)
```

Out[89]:

	video_id	trending_date	title	channel_title	category_id	publish_time	
0	n1WpP7iowLc	17.14.11	Eminem - Walk On Water (Audio) ft. BeyoncÃ©	EminemVEVO	10	2017-11-10T17:00:03.000Z	Em
1	0dBIkQ4Mz1M	17.14.11	PLUSH - Bad Unboxing Fan Mail	iDubbbzTV	23	2017-11-13T17:00:00.000Z	
2	5qpjK5DgCt4	17.14.11	Racist Superman | Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	2017-11-12T19:05:24.000Z	ra
3	d380meD0W0M	17.14.11	I Dare You: GOING BALD!?	nigahiga	24	2017-11-12T18:01:41.000Z	
4	2Vv-BfVoq4g	17.14.11	Ed Sheeran - Perfect (Official Music Video)	Ed Sheeran	10	2017-11-09T11:04:14.000Z	
5	0yIWz1XEeyc	17.14.11	Jake Paul Says Alissa Violet CHEATED with LOGA...	DramaAlert	25	2017-11-13T07:37:51.000Z	#D

In [ ]:
```
```

In [90]:
```
full_df['channel_title'].value_counts()
```

Out[90]:
```
The Late Show with Stephen Colbert    710
WWE                                   643
Late Night with Seth Meyers           592
TheEllenShow                          555
Jimmy Kimmel Live                     528
                                     ...
Daas                                    1
YT Industries                           1
BTLV Le mÃ©dia complÃ©mentaire           1
Quem Sabia ?                            1
Jessi Osorno                            1
Name: channel_title, Length: 37824, dtype: int64
```

In [ ]:
```
```

In [91]:
```
cdf=full_df.groupby(['channel_title']).size().sort_values(ascending=False).r
```

In [92]:
```
cdf= cdf.rename(columns={0:'total_videos'})
```

In [93]:
```
cdf
```

Out[93]:

	channel_title	total_videos
0	The Late Show with Stephen Colbert	710
1	WWE	643
2	Late Night with Seth Meyers	592
3	TheEllenShow	555
4	Jimmy Kimmel Live	528
...	...	...
37819	Kd Malts	1
37820	Zedan TV	1
37821	Kc Kelly - Rocketprenuer	1
37822	Kbaby	1
37823	Pavel Sidorik TV	1

37824 rows × 2 columns

In [94]:
```
import plotly.express as px
```

In [95]:
```
px.bar(data_frame=cdf[0:20], x='channel_title', y='total_videos')
```

Does Panctuations in title and tags have any relations with views, Dislikes, Comments

In [96]:
```
full_df['title'][0]
```

Out[96]:
```
'Eminem - Walk On Water (Audio) ft. BeyoncÃ©'
```

In [ ]:
```

```

In [97]:
```
import string
```

In [98]:
```
string.punctuation
```

Out[98]:
```
'!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

In [99]:
```
len([char for char in full_df['title'][0] if char in string.punctuation])
```

Out[99]:
```
4
```

In [106]:
```
def punc_count(text):
    return len([char for char in text if char in string.punctuation])
```

In [ ]:
```
```

In [107]:
```
sample = full_df[0:10000]
```

In [108]:
```
sample['count_punc'] = sample['title'].apply(punc_count)
```

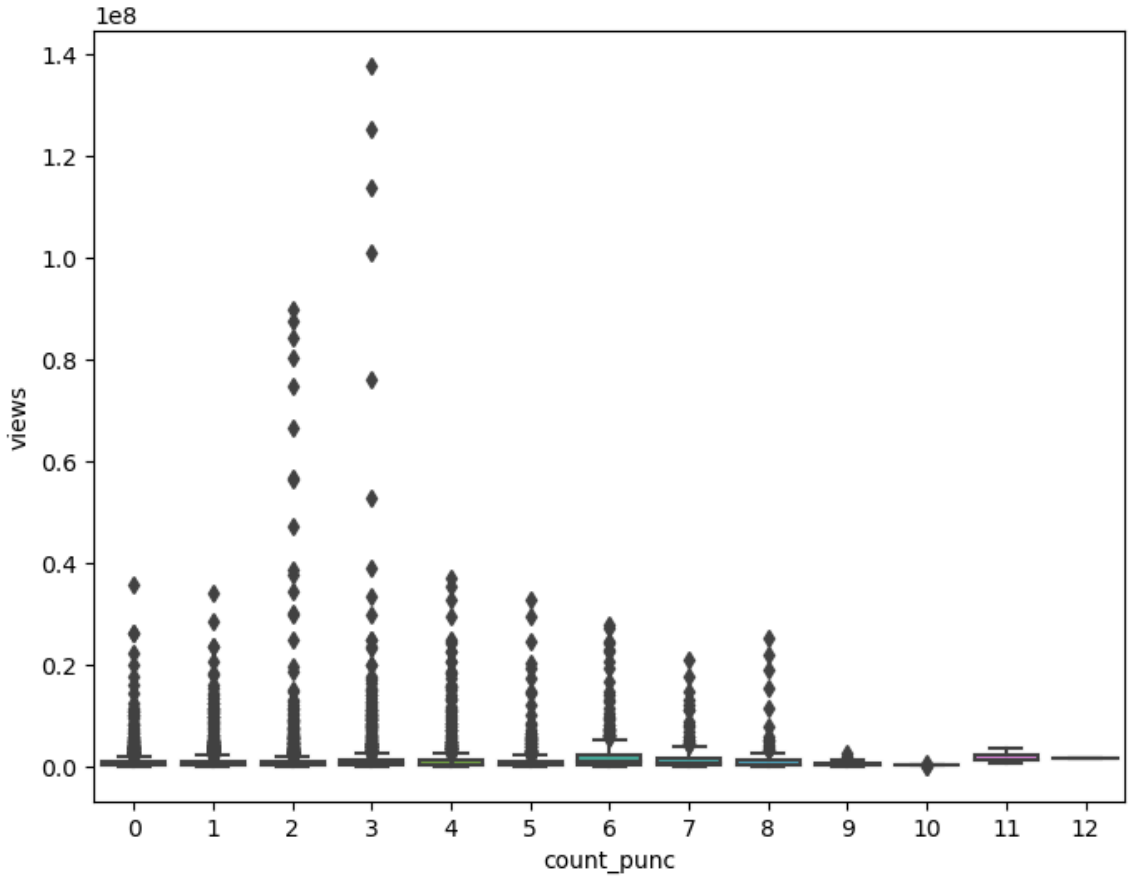
In [109]:
```
sample['count_punc']
```

Out[109]:
```
0       4
1       1
2       3
3       3
4       3
       ..
9995    6
9996    0
9997    1
9998    0
9999    6
Name: count_punc, Length: 10000, dtype: int64
```

In [ ]:
```
```

In [110]:
```
plt.figure(figsize=(8,6))
sns.boxplot(x='count_punc', y='views' , data=sample)
plt.show
```

Out[110]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: