

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

```
In [378... comments = pd.read_csv(r'C:\Users\me\Desktop\utube\UScomments.csv', on_bad_lines='s
```

```
In [45]: comments.head()
```

```
Out[45]:
```

	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 [47]: comments.isnull()
```

```
Out[47]:
```

	video_id	comment_text	likes	replies
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...
691395	False	False	False	False
691396	False	False	False	False
691397	False	False	False	False
691398	False	False	False	False
691399	False	False	False	False

691400 rows × 4 columns

```
In [ ]: comments.ISNULL()
```

```
In [51]: comments.isnull()
```

```
Out[51]:
```

	video_id	comment_text	likes	replies
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...
691395	False	False	False	False
691396	False	False	False	False
691397	False	False	False	False
691398	False	False	False	False
691399	False	False	False	False

691400 rows × 4 columns

```
In [53]: comments.isnull().sum()
```

```
Out[53]: video_id      0
comment_text  26
likes        0
replies      0
dtype: int64
```

```
In [55]: comments.dropna(inplace=True)
```

```
In [57]: comments.isnull().sum()
```

```
Out[57]: video_id      0
comment_text  0
likes        0
replies      0
dtype: int64
```

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

```
In [61]: from textblob import TextBlob
```

```
In [63]: comments.head(6)
```

Out[63]:

	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 AYYYYEEEE	3	0

In [65]: `comments.head(1000)`

Out[65]:

	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
...
995	Ayb_2qbZHm4	Generic attention comment	0	0
996	Ayb_2qbZHm4	Zero amount of funny	0	0
997	Ayb_2qbZHm4	Honest? I see no random color hairgender studi...	0	0
998	Ayb_2qbZHm4	What, no joke about how you can't apply if you...	0	0
999	CsdzflTXBVQ	Floyd Mayweather, the swimmer.	3331	261

1000 rows × 4 columns

In [69]: `TextBlob("Logan Paul it's yo big day !!!!!").sentiment.polarity`

Out[69]: 0.0

```
In [ ]: polarity=[]

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

In [75]: `sample_df = comments[0:1000]`

In [103... `sample_df.shape`

Out[103... `(1000, 4)`

In [109... `comments['polarity'] = polarity`

In [111... `comments.head(6)`

Out[111...

	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
5	XpVt6Z1Gjjo	#1 on trending AYYYYEEEE	3	0	0.0

In [117... `polarity[0]`

Out[117... `0.0`

In [119... `filter1=comments['polarity']==1`

In [123... `comments_positive=comments[filter1]`

In [127... `comments_positive`

Out [127...

	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
...
691373	EoejGgUNmVU	Perfect. Every version	0	0	1.0
691383	EoejGgUNmVU	Me encantaaaaa <3	0	0	1.0
691388	EoejGgUNmVU	Best song ever 🙌	0	0	1.0
691389	EoejGgUNmVU	excellent performance	0	0	1.0
691393	EoejGgUNmVU	<3	0	0	1.0

19390 rows × 5 columns

In [129...

```
filter2=comments['polarity']==-1
```

In [131...

```
comments_negative=comments[filter2]
```

In [133...

```
comments_negative
```

Out[133...

	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
...
690788	Hlt3rA-oDao	Like Kelly she evil	0	0	-1.0
690865	Ni5Uc-DdMqk	R U FUCKING KIDDING ME?!?!?!?!?	11	1	-1.0
691073	4kmSATD25iY	This is horribly offensive please report	0	0	-1.0
691180	xlu6i6lT_vk	Sink holes looks terrifying sinkholes sink you...	0	0	-1.0
691224	qRoVIH1Ocl4	Trump talked to the president of US Virgin Isl...	0	0	-1.0

3508 rows × 5 columns

In []: !pip install wordcloud

In [137... from wordcloud import WordCloud, STOPWORDS

In [139... set(STOPWORDS)

```
Out[139... {'a',
            'about',
            'above',
            'after',
            'again',
            'against',
            'all',
            'also',
            'am',
            'an',
            'and',
            'any',
            'are',
            "aren't",
            'as',
            'at',
            'be',
            'because',
            'been',
            'before',
            'being',
            'below',
            'between',
            'both',
            'but',
            'by',
            'can',
            "can't",
            'cannot',
            'com',
            'could',
            "couldn't",
            'did',
            "didn't",
            'do',
            'does',
            "doesn't",
            'doing',
            "don't",
            'down',
            'during',
            'each',
            'else',
            'ever',
            'few',
            'for',
            'from',
            'further',
            'get',
            'had',
            "hadn't",
            'has',
            "hasn't",
            'have',
            "haven't",
            'having',
```

'he',
"he'd",
"he'll",
"he's",
'hence',
'her',
'here',
"here's",
'hers',
'herself',
'him',
'himself',
'his',
'how',
"how's",
'however',
'http',
'i',
"i'd",
"i'll",
"i'm",
"i've",
'if',
'in',
'into',
'is',
"isn't",
'it',
"it's",
'its',
'itself',
'just',
'k',
"let's",
'like',
'me',
'more',
'most',
"mustn't",
'my',
'myself',
'no',
'nor',
'not',
'of',
'off',
'on',
'once',
'only',
'or',
'other',
'otherwise',
'ought',
'our',
'ours',
'ourselves',

'out',
'over',
'own',
'r',
'same',
'shall',
"shan't",
'she',
"she'd",
"she'll",
"she's",
'should',
"shouldn't",
'since',
'so',
'some',
'such',
'than',
'that',
"that's",
'the',
'their',
'theirs',
'them',
'themselves',
'then',
'there',
"there's",
'therefore',
'these',
'they',
"they'd",
"they'll",
"they're",
"they've",
'this',
'those',
'through',
'to',
'too',
'under',
'until',
'up',
'very',
'was',
"wasn't",
'we',
"we'd",
"we'll",
"we're",
"we've",
'were',
"weren't",
'what',
"what's",
'when',

```
"when's",
'where',
"where's",
'which',
'while',
'who',
"who's",
'whom',
'why',
"why's",
'with',
"won't",
'would',
"wouldn't",
'www',
'you',
"you'd",
"you'll",
"you're",
"you've",
'your',
'yours',
'yourself',
'yourselves'}
```

In []:

In [141... type(comments['comment_text'])

Out[141... pandas.core.series.Series

In [143... total_comments_positive=' '.join(comments_positive['comment_text'])

In [145... wordcloud=WordCloud(stopwords=set(STOPWORDS)).generate(total_comments_positive)

In [147... plt.imshow(wordcloud)
plt.axis('off')

Out[147... (-0.5, 399.5, 199.5, -0.5)



```
In [149...] total_comments_negative=' '.join(comments_negative['comment_text'])
```

```
In [151...] wordcloud1=WordCloud(stopwords=set(STOPWORDS)).generate(total_comments_negative)
```

```
In [153...] plt.imshow(wordcloud1)
plt.axis('off')
```

```
Out[153...] (-0.5, 399.5, 199.5, -0.5)
```



```
In [ ]:
```

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

```
In [158...] import emoji
```

```
In [160...] comments['comment_text']
```

```

Out[160... 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          Isliği yeter...
691399  Kocham tą piosenkę 🥰❤❤❤byłam zakochana po uszy ...
Name: comment_text, Length: 691374, dtype: object

```

In []:

```

In [165... all_emojis_list=[]

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

```

```

In [167... all_emojis_list[0:10]

```

```

Out[167... ['!!', '!!', '!!', '😊', '😭', '👍', '🇲🇽', '🇲🇽', '🇲🇽', '🇲🇽']

```

In []:

```

In [170... from collections import Counter

```

```

In [180... Counter(all_emojis_list).most_common(10)

```

```

Out[180... [('😭', 36987),
('😭', 33453),
('❤', 31119),
('🔥', 8694),
('😭', 8398),
('👍', 5719),
('😊', 5545),
('👍', 5476),
('❤', 5359),
('❤', 5147)]

```

In []:

```

In [182... emojis=[Counter(all_emojis_list).most_common(10)[i][0] for i in range(10)]

```

```

In [184... freqs=[Counter(all_emojis_list).most_common(10)[i][1] for i in range(10)]

```

```

In [186... freqs

```

```

Out[186... [36987, 33453, 31119, 8694, 8398, 5719, 5545, 5476, 5359, 5147]

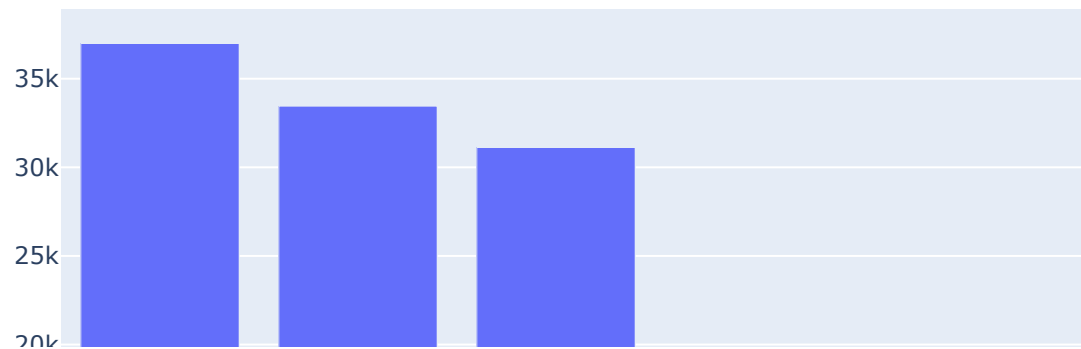
```

In []:

```
In [190... import plotly.graph_objs as go
from plotly.offline import iplot
```

```
In [192... trace=go.Bar(x=emojis,y=freqs)
```

```
In [194... iplot([trace])
```



In []:

```
In [196... import os
```

```
In [204... files=os.listdir(r'C:\Users\me\Downloads\dataset\additional_data')
```

```
In [206... files
```

```
Out[206... ['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 [210... files_csv=[file for file in files if '.csv' in file]
```

```
In [212... files_csv
```

```
Out[212... ['CAvideos.csv',
            'DEvideos.csv',
            'FRvideos.csv',
            'GBvideos.csv',
            'INvideos.csv',
            'JPvideos.csv',
            'KRvideos.csv',
            'MXvideos.csv',
            'RUvideos.csv',
            'USvideos.csv']
```

```
In [ ]:
```

```
In [214... import warnings
from warnings import filterwarnings
filterwarnings('ignore')
```

```
In [216... # defining blank dataframe
```

```
In [220... #data collection
full_df=pd.DataFrame()
path=r'C:\Users\me\Downloads\dataset\additional_data'

for file in files_csv:
    current_df=pd.read_csv(path+'/'+file , encoding='iso-8859-1' , on_bad_lines='skip')
    full_df=pd.concat([full_df,current_df],ignore_index=True)
```

```
In [222... full_df.shape
```

Out[222... (375942, 16)

In []:

```
In [227... #remove duplicates  
full_df[full_df.duplicated()].shape
```

Out[227... (36417, 16)

```
In [229... full_df=full_df.drop_duplicates()
```

```
In [231... full_df.shape
```

Out[231... (339525, 16)

```
In [233... #exporting  
full_df[0:1000].to_csv(r'E:\youtube_analysis/youtube_sample.csv',index=False)
```

```
In [237... #storing in db  
#create a engine to connect db  
from sqlalchemy import create_engine
```

```
In [241... engine=create_engine(r'sqlite:///E:\youtube_analysis/youtube_sample.sqlite')
```

```
In [243... full_df[0:1000].to_sql('Users',con=engine,if_exists='append')
```

Out[243... 1000

```
In [ ]: # which category has the maximum likes ?
```

```
In [245... full_df.head(5)
```

Out[245...

	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
1	0dBlkQ4Mz1M	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
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 [247...

```
full_df['category_id'].unique()
```

Out[247...

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

In [253...

```
json_df=pd.read_json(r'C:\Users\me\Downloads\dataset\additional_data\US_category_id
```

In [255...

```
json_df
```


		kind	etag
0	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
1	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
2	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
3	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
4	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
5	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
6	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
7	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
8	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
9	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
10	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
11	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
12	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et
13	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	'youtube#video' 'et

	kind	etag
14	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
15	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
16	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
17	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
18	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
19	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
20	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
21	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
22	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
23	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
24	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
25	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
26	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'
27	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..." 'youtube#videoCategoryListResponse'

	kind	etag
28	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..."
29	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..."
30	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..."
31	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv..."

In [257... json_df['items']

```
Out[257... 0    {'kind': 'youtube#videoCategory', 'etag': '"m2...
1    {'kind': 'youtube#videoCategory', 'etag': '"m2...
2    {'kind': 'youtube#videoCategory', 'etag': '"m2...
3    {'kind': 'youtube#videoCategory', 'etag': '"m2...
4    {'kind': 'youtube#videoCategory', 'etag': '"m2...
5    {'kind': 'youtube#videoCategory', 'etag': '"m2...
6    {'kind': 'youtube#videoCategory', 'etag': '"m2...
7    {'kind': 'youtube#videoCategory', 'etag': '"m2...
8    {'kind': 'youtube#videoCategory', 'etag': '"m2...
9    {'kind': 'youtube#videoCategory', 'etag': '"m2...
10   {'kind': 'youtube#videoCategory', 'etag': '"m2...
11   {'kind': 'youtube#videoCategory', 'etag': '"m2...
12   {'kind': 'youtube#videoCategory', 'etag': '"m2...
13   {'kind': 'youtube#videoCategory', 'etag': '"m2...
14   {'kind': 'youtube#videoCategory', 'etag': '"m2...
15   {'kind': 'youtube#videoCategory', 'etag': '"m2...
16   {'kind': 'youtube#videoCategory', 'etag': '"m2...
17   {'kind': 'youtube#videoCategory', 'etag': '"m2...
18   {'kind': 'youtube#videoCategory', 'etag': '"m2...
19   {'kind': 'youtube#videoCategory', 'etag': '"m2...
20   {'kind': 'youtube#videoCategory', 'etag': '"m2...
21   {'kind': 'youtube#videoCategory', 'etag': '"m2...
22   {'kind': 'youtube#videoCategory', 'etag': '"m2...
23   {'kind': 'youtube#videoCategory', 'etag': '"m2...
24   {'kind': 'youtube#videoCategory', 'etag': '"m2...
25   {'kind': 'youtube#videoCategory', 'etag': '"m2...
26   {'kind': 'youtube#videoCategory', 'etag': '"m2...
27   {'kind': 'youtube#videoCategory', 'etag': '"m2...
28   {'kind': 'youtube#videoCategory', 'etag': '"m2...
29   {'kind': 'youtube#videoCategory', 'etag': '"m2...
30   {'kind': 'youtube#videoCategory', 'etag': '"m2...
31   {'kind': 'youtube#videoCategory', 'etag': '"m2...
Name: items, dtype: object
```

In [259... json_df['items'][0]

```
Out[259... {'kind': 'youtube#videoCategory',
            'etag': '"m2yskBQFythfE4irbTieOgYYfBU/Xy1mB4_yLrHy_BmKmpBggtY2mZQ"',
            'id': '1',
            'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
                        'title': 'Film & Animation',
                        'assignable': True}}
```

```
In [261... json_df['items'][1]
```

```
Out[261... {'kind': 'youtube#videoCategory',
            'etag': '"m2yskBQFythfE4irbTieOgYYfBU/UZ1oLIiz2dxIhO45ZTFR3a3NyTA"',
            'id': '2',
            'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
                        'title': 'Autos & Vehicles',
                        'assignable': True}}
```

```
In [263... #items - its a dictionary itself
cat_dict={}
for item in json_df['items'].values:
    cat_dict[int(item['id'])]=item['snippet']['title']
```

```
In [265... full_df['category_id'].map(cat_dict)
```

```
Out[265... 0           Music
1           Comedy
2           Comedy
3      Entertainment
4           Music
...
375936  People & Blogs
375938  People & Blogs
375939      Entertainment
375940  Film & Animation
375941           Gaming
Name: category_id, Length: 339525, dtype: object
```

```
In [273... cat_dict
```

```
Out[273... {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 [275... full_df['category_name']=full_df['category_id'].map(cat_dict)
```

```
In [277... full_df.head(5)
```

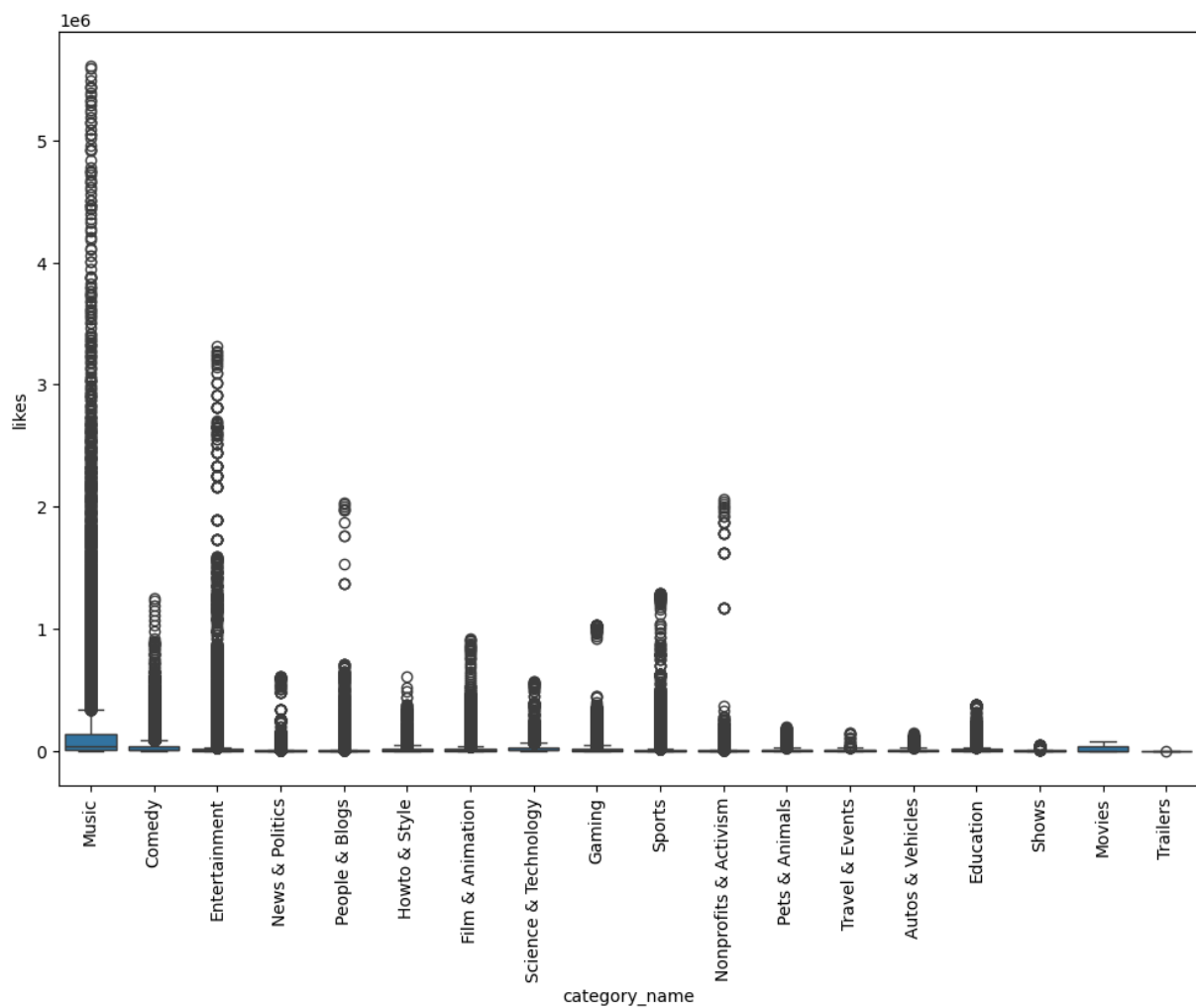
Out[277...

	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
1	0dBlkQ4Mz1M	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
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 []: *#plotting percentile values*

In [287...
plt.figure(figsize=(12,8))
sns.boxplot(x='category_name',y='likes',data=full_df)
plt.xticks(rotation='vertical')

```
Out[287...] ([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')])
```



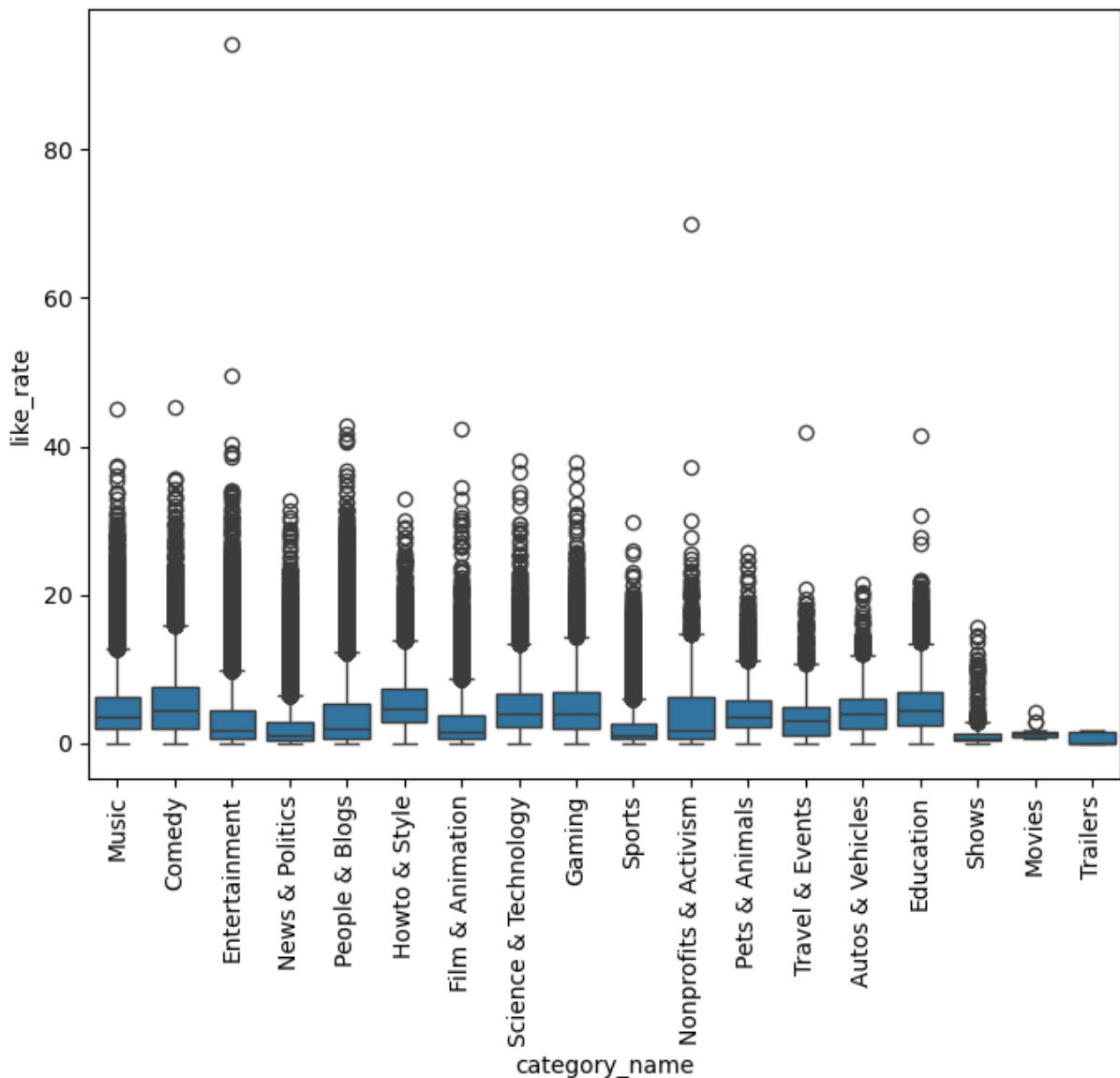
```
In [ ]: #find out whether audience is engaged or not
#like_rate
#dislikes
#comment_count
```

```
In [291...] full_df['like_rate']=(full_df['likes']/full_df['views'])*100
full_df['dislike_rate']=(full_df['dislikes']/full_df['views'])*100
full_df['comment_count_rate']=(full_df['comment_count']/full_df['views'])*100
```

```
In [293...] full_df.columns
```

```
Out[293...] Index(['video_id', 'trending_date', 'title', 'channel_title', 'category_id',
      'publish_time', 'tags', 'views', 'likes', 'dislikes', 'comment_count',
      'thumbnail_link', 'comments_disabled', 'ratings_disabled',
      'video_error_or_removed', 'description', 'category_name', 'like_rate',
      'dislike_rate', 'comment_count_rate'],
      dtype='object')
```

```
In [295...] plt.figure(figsize=(8,6))
sns.boxplot(x='category_name',y='like_rate',data=full_df)
plt.xticks(rotation='vertical')
plt.show()
```

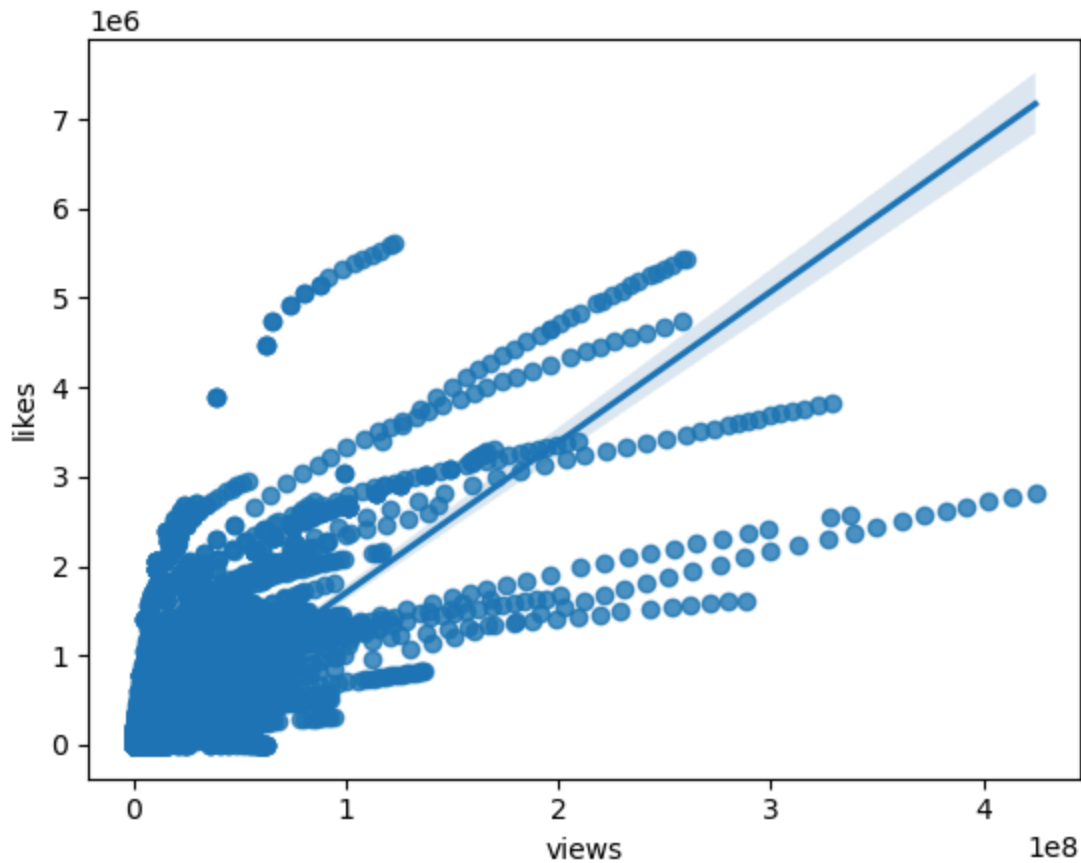


```
In [297...] #regression_plot
sns.regplot(x='views',y='likes',data=full_df)
```



```
# views incre likes also incre
```

```
Out[297... <Axes: xlabel='views', ylabel='likes'>
```



```
In [299... full_df.columns
```

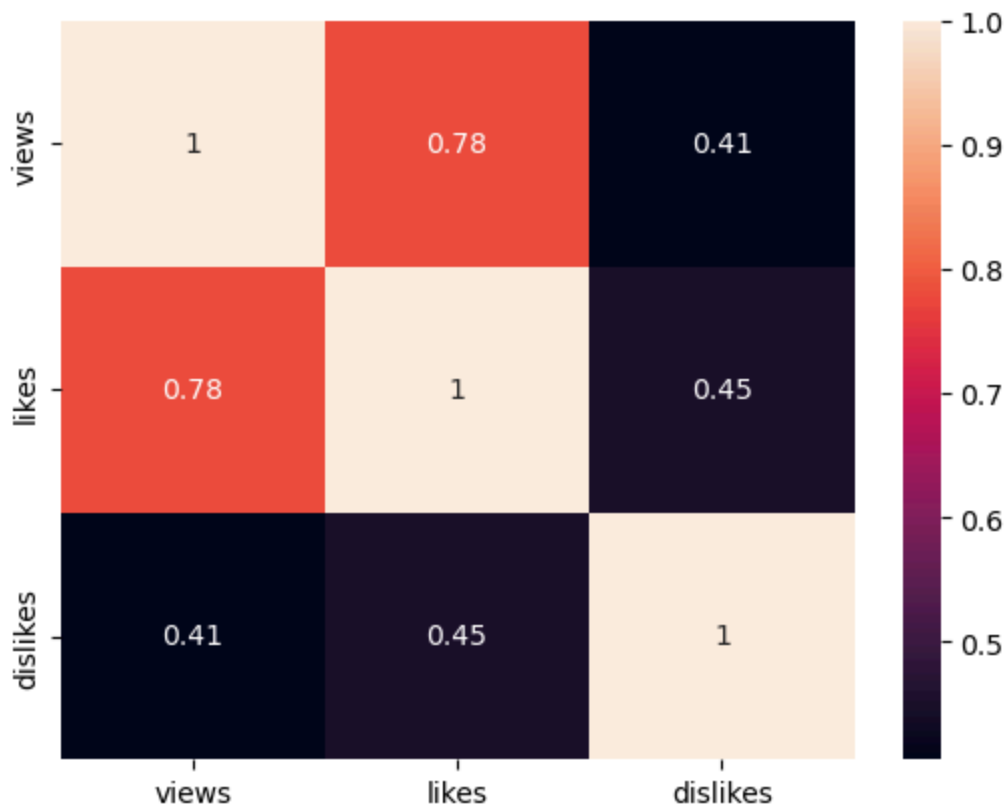
```
Out[299... Index(['video_id', 'trending_date', 'title', 'channel_title', 'category_id',  
      'publish_time', 'tags', 'views', 'likes', 'dislikes', 'comment_count',  
      'thumbnail_link', 'comments_disabled', 'ratings_disabled',  
      'video_error_or_removed', 'description', 'category_name', 'like_rate',  
      'dislike_rate', 'comment_count_rate'],  
      dtype='object')
```

```
In [303... full_df[['views', 'likes', 'dislikes']].corr() #correlation
```

```
Out[303...      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 [307... sns.heatmap(full_df[['views', 'likes', 'dislikes']].corr(),annot=True)
```

```
Out[307... <Axes: >
```



In [309... `full_df.head(6)`

Out[309...

	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
1	0dBlkQ4Mz1M	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
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

In [337...

```
cdf=full_df.groupby(['channel_title']).size().sort_values(ascending=False).reset_in
```

In [339...

```
cdf
```

Out[339...

	channel_title	0
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 [341...

```
cdf=cdf.rename(columns={0:'total_videos'})
```

In [343...

```
cdf
```

Out[343...

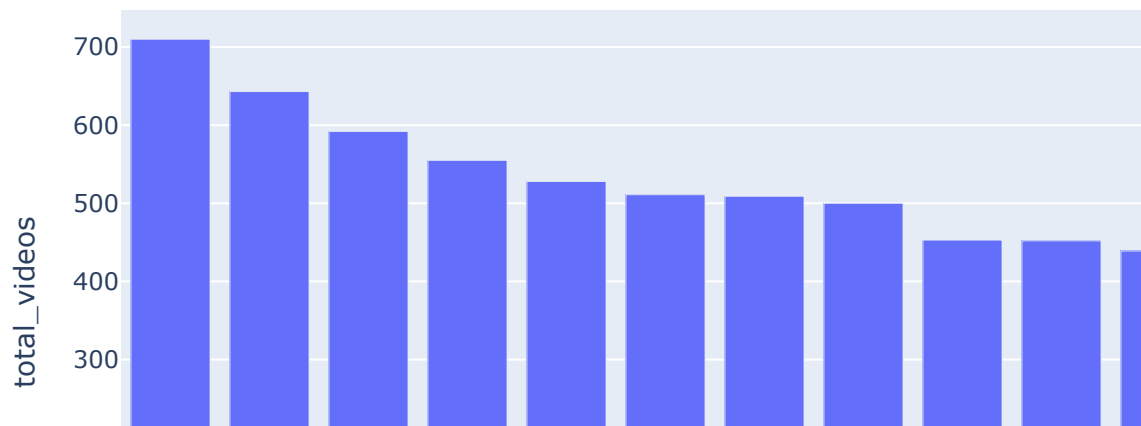
	channel_title	total_videos
0	The Late Show with Stephen Colbert	710
1	WWE	643
2	Late Night with Seth Meyers	592
3	TheEllenShow	555
4	Jimmy Kimmel Live	528
...
37819	Kd Malts	1
37820	Zedan TV	1
37821	Kc Kelly - Rocketprenuer	1
37822	Kbaby	1
37823	Pavel Sidorik TV	1

37824 rows × 2 columns

In [345...

```
import plotly.express as px
```

```
In [347... px.bar(data_frame=cdf[0:20],x='channel_title',y='total_videos')
```



```
In [ ]: #Does punctiations in title and tags have any relation with views,likes,dislikes co
```

```
In [349... full_df['title']
```

```
Out[349... 0          Eminem - Walk On Water (Audio) ft. BeyoncÃ©
1          PLUSH - Bad Unboxing Fan Mail
2    Racist Superman | Rudy Mancuso, King Bach & Le...
3          I Dare You: GOING BALD!?!
4    Ed Sheeran - Perfect (Official Music Video)
...
375936    BTS Plays With Puppies While Answering Fan Que...
375938          True Facts : Ant Mutualism
375939    I GAVE SAFIYA NYGAARD A PERFECT HAIR MAKEOVER ...
375940          How Black Panther Should Have Ended
375941    Official Call of Duty®: Black Ops 4 â Mult...
Name: title, Length: 339525, dtype: object
```

```
In [351... full_df['title'][0]
```

```
Out[351... 'Eminem - Walk On Water (Audio) ft. BeyoncÃ©'
```

```
In [ ]: #logic - punct inc views inc or punct dec and views dec
```

```
In [353... import string #help to calculate punctutation in any dataset
```

```
In [355... string.punctuation
```

```
Out[355... '!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

```
In [357... len([char for char in full_df['title'][0] if char in string.punctuation])
```

```
Out[357... 4
```

```
In [359... def punc_count(text):  
    return len([char for char in text if char in string.punctuation])
```

```
In [361... sample=full_df[0:10000]
```

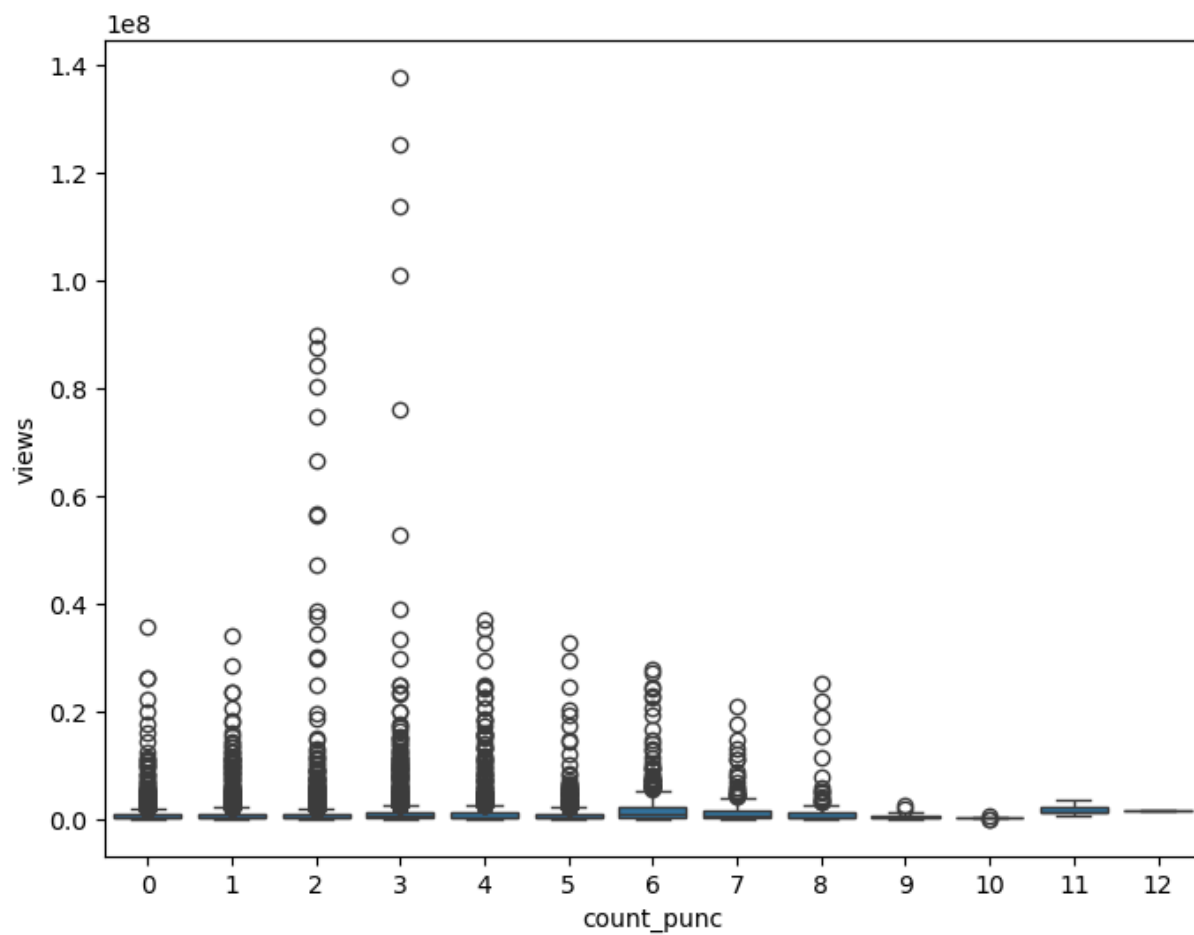
```
In [363... sample['count_punc']=sample['title'].apply(punc_count)
```

```
In [370... sample['count_punc']
```

```
Out[370... 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
```

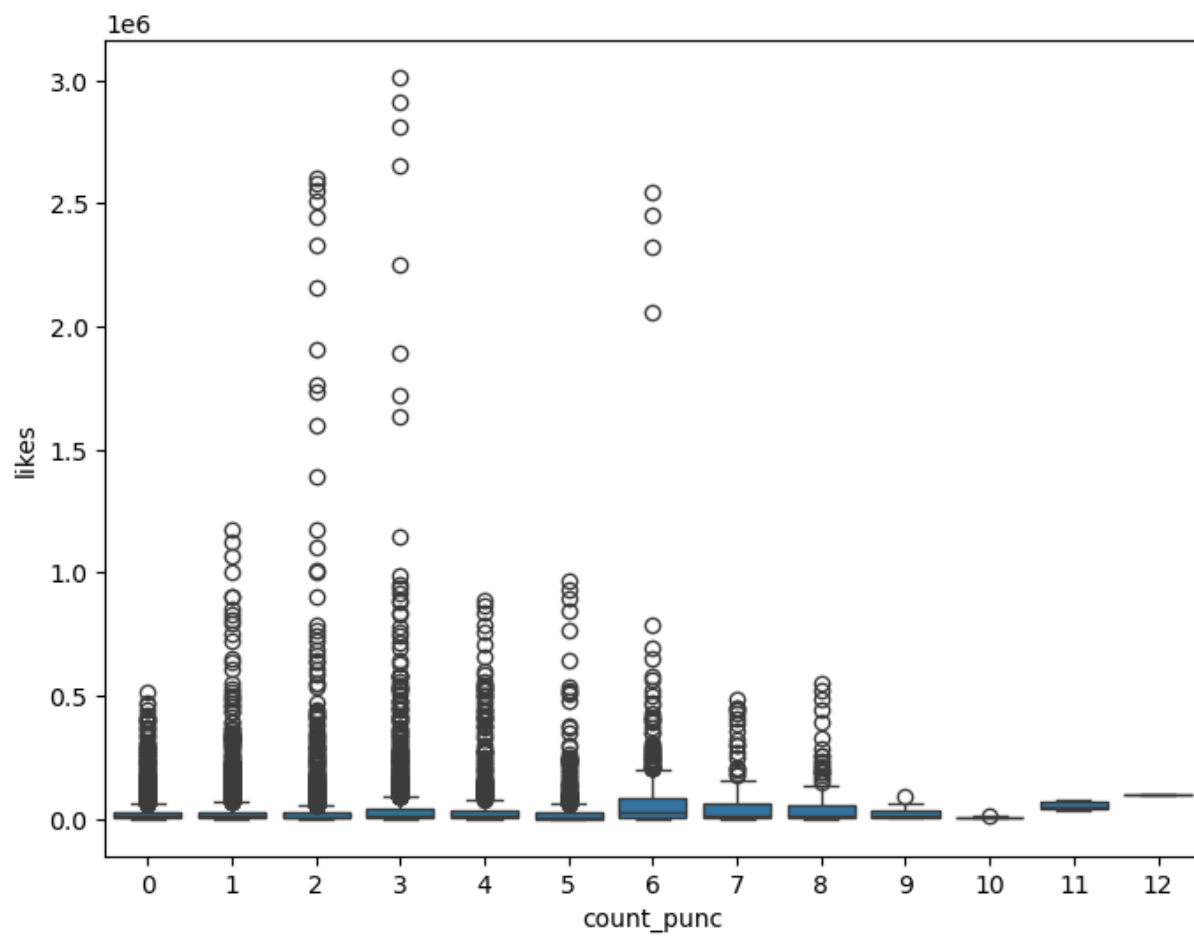
```
In [ ]: #box plot
```

```
In [374... plt.figure(figsize=(8,6))  
sns.boxplot(x='count_punc',y='views',data=sample)  
plt.show()
```



```
In [ ]: # when punct is 2 or 3 the views are generally high
```

```
In [376... plt.figure(figsize=(8,6))
sns.boxplot(x='count_punc',y='likes',data=sample)
plt.show()
```



In []: *# when punct is 2 or 3 or 5 the Likes are generally high*

In []: