

1. Read & load data

```
In [1]: # Lets import all the necessary packages !
```

```
In [2]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [3]: comments= pd.read_csv(r'C:\Users\lenov\Desktop\Coding\Projects\1\UScomments.csv', error_bad_lines=False)
```

C:\Users\lenov\AppData\Local\Temp\ipykernel_5388\3340412043.py:1: FutureWarning: The error_bad_lines argument has been deprecated and will be removed in a future version. Use on_bad_lines in the future.

```
comments= pd.read_csv(r'C:\Users\lenov\Desktop\Coding\Projects\1\UScomments.csv', error_bad_lines=False)
b'Skipping line 41589: expected 4 fields, saw 11\nSkipping line 51628: expected 4 fields, saw 7\nSkipping line 114
465: expected 4 fields, saw 5\n'
b'Skipping line 142496: expected 4 fields, saw 8\nSkipping line 189732: expected 4 fields, saw 6\nSkipping line 24
5218: expected 4 fields, saw 7\n'
b'Skipping line 388430: expected 4 fields, saw 5\n'
C:\Users\lenov\AppData\Local\Temp\ipykernel_5388\3340412043.py:1: DtypeWarning: Columns (2,3) have mixed types. Sp
ecify dtype option on import or set low_memory=False.
comments= pd.read_csv(r'C:\Users\lenov\Desktop\Coding\Projects\1\UScomments.csv', error_bad_lines=False)
```

```
In [4]: # above is a warning (ignore)
```

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

```
Out[5]:
```

	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 [6]: ## Lets finding out missing values in data
comments.isnull().sum()
```

```
Out[6]: video_id      0
comment_text    25
likes           0
replies         0
dtype: int64
```

```
In [7]: ## drop missing values as we have very few & Lets update dataframe as well..
comments.dropna(inplace=True)
```

```
In [8]: comments.isnull().sum()
```

```
Out[8]: video_id      0
comment_text    0
likes           0
replies         0
dtype: int64
```

2.Performing Sentiment Analysis

```
In [9]: ### performing sentiment analysis using TextBlob which is a NLP Library built on top of NLTK ..  
!pip install textblob
```

```
Defaulting to user installation because normal site-packages is not writeable  
Requirement already satisfied: textblob in c:\users\lenov\appdata\roaming\python\python39\site-packages (0.17.1)  
Requirement already satisfied: nltk>=3.1 in c:\programdata\anaconda3\lib\site-packages (from textblob) (3.7)  
Requirement already satisfied: tqdm in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (4.64.1)  
Requirement already satisfied: click in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (8.0.4)  
Requirement already satisfied: regex>=2021.8.3 in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (2022.7.9)  
Requirement already satisfied: joblib in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (1.1.0)  
Requirement already satisfied: colorama in c:\programdata\anaconda3\lib\site-packages (from click->nltk>=3.1->textblob) (0.4.5)
```

```
In [10]: from textblob import TextBlob
```

```
In [11]: comments.head(6)
```

```
Out[11]:
```

	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 [12]: TextBlob("Logan Paul it's yo big day !!!!!").sentiment.polarity  
### its a neutral sentence !
```

```
Out[12]: 0.0
```

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

```
Out[13]: (691375, 4)
```

```
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 [16]: comments['polarity']=polarity  
### Inserting polarity values into comments dataframe while defining feature name as "polarity"
```

```
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

3.Wordcloud Analysis

```
In [18]: ### Lets perform EDA for the highly Positive sentences ie Polarity value will be 1
```

```
In [19]: filter1=comments['polarity']==1
```

```
In [20]: comments_positive = comments[filter1]
```

```
In [21]: filter2 = comments['polarity']==-1
```

```
In [22]: comments_negative = comments[filter2]
```

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

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: wordcloud in c:\users\lenov\appdata\roaming\python\python39\site-packages (1.9.3)
Requirement already satisfied: matplotlib in c:\programdata\anaconda3\lib\site-packages (from wordcloud) (3.5.2)
Requirement already satisfied: pillow in c:\programdata\anaconda3\lib\site-packages (from wordcloud) (9.2.0)
Requirement already satisfied: numpy>=1.6.1 in c:\programdata\anaconda3\lib\site-packages (from wordcloud) (1.21.5)
Requirement already satisfied: fonttools>=4.22.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\lib\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.4.2)
Requirement already satisfied: packaging>=20.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib->wordcloud) (21.3)
Requirement already satisfied: python-dateutil>=2.7 in c:\programdata\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-packages (from python-dateutil->matplotlib->wordcloud) (1.16.0)
```

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

```
set(STOPWORDS)
```

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

```
comments['comment_text']
```

```
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          Islgi yeter...
691399          Kocham tą piosenkę😍❤️❤️❤️byłam zakochana po uszy ...
Name: comment_text, Length: 691375, dtype: object
```

```
type(comments['comment_text'])
```

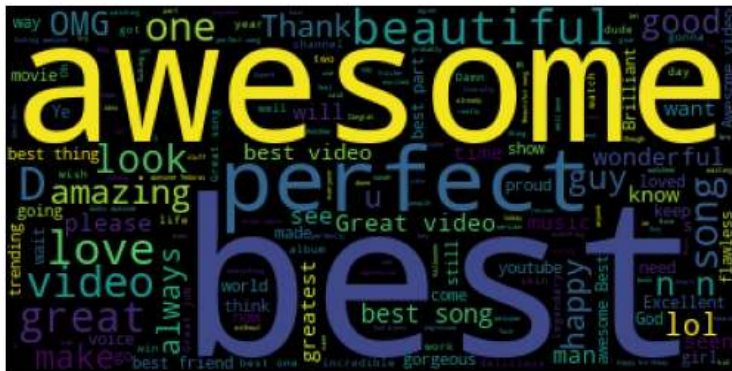
pandas.core.series.Series

```
### for wordcloud , we need to frame our 'comment_text' feature into string ..
total_comments_positive = ' '.join(comments_positive['comment_text'])
```

```
wordcloud = WordCloud(stopwords=set(STOPWORDS)).generate(total_comments_positive)
```

```
plt.imshow(wordcloud)
plt.axis('off')
```

$(-0.5, 399.5, 199.5, -0.5)$



Conclusion--> positive Users are emphasizing more on best , awesome , perfect , amazing , Look , happy etc..

```
In [33]: total_comments_negative = ' '.join(comments_negative['comment_text'])
```

```
In [34]: wordcloud2 = WordCloud(stopwords=set(STOPWORDS)).generate(total_comments_negative)
```

```
In [35]: plt.imshow(wordcloud2)
plt.axis('off')
```

```
Out[35]: (-0.5, 399.5, 199.5, -0.5)
```



In [36]: *### Conclusion--> Negative Users are emphasizing more on Terrible , worst ,horrible ,boring , disgusting etc..*

4. Performing Emoji's Analysis

```
In [37]: !pip install emoji==2.2.0
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: emoji==2.2.0 in c:\users\lenov\appdata\roaming\python\python39\site-packages (2.2.0)
```

```
In [38]: import emoji
```

```
In [39]: emoji.__version__
```

Out[39]: '2.2.0'

```
In [40]: comments['comment_text'].head(6)
```

```
Out[40]: 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 AYYEEEEEE
Name: comment_text, dtype: object
```

```
In [41]: comments['comment_text'].head(6)
```

```
Out[41]: 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 AYYEEEEEE
Name: comment_text, dtype: object
```

```
In [42]: ### Lets extract emoji from below comment
```

```
In [43]: comment = 'trending 😊'
```

```
In [44]: [char for char in comment if char in emoji.EMOJI_DATA]
```

```
Out[44]: ['😊']
```

```
In [45]: ## above code in a more simpler & readable way :  
emoji_list = []  
  
for char in comment:  
    if char in emoji.EMOJI_DATA:  
        emoji_list.append(char)
```

```
In [46]: emoji_list
```

```
Out[46]: ['😊']
```

```
In [47]: all_emojis_list = []  
  
for comment in comments['comment_text'].dropna(): ## in case have missing values , call dropna()  
    for char in comment:  
        if char in emoji.EMOJI_DATA:  
            all_emojis_list.append(char)
```

```
In [48]: all_emojis_list[0:10]
```

```
Out[48]: ['!', '!', '!', '😊', '😭', '👍', '👎', '👎', '😭', '🚫']
```

```
In [49]: ### Now we have to compute frequencies of each & every emoji in "all_emojis_list"..
```

```
In [50]: from collections import Counter
```

```
In [51]: Counter(all_emojis_list).most_common(10)
```

```
Out[51]: [('😭', 36987),  
          ('😊', 33453),  
          ('❤️', 31119),  
          ('👍', 8694),  
          ('😭', 8398),  
          ('👎', 5719),  
          ('😊', 5545),  
          ('👍', 5476),  
          ('❤️', 5359),  
          ('❤️', 5147)]
```

```
In [52]: Counter(all_emojis_list).most_common(10)[0]
```

```
Out[52]: ('😭', 36987)
```

```
In [53]:  
  
Counter(all_emojis_list).most_common(10)[1][0]
```

```
Out[53]: '😊'
```

```
In [54]:  
  
Counter(all_emojis_list).most_common(10)[2][0]
```

```
Out[54]: '❤️'
```

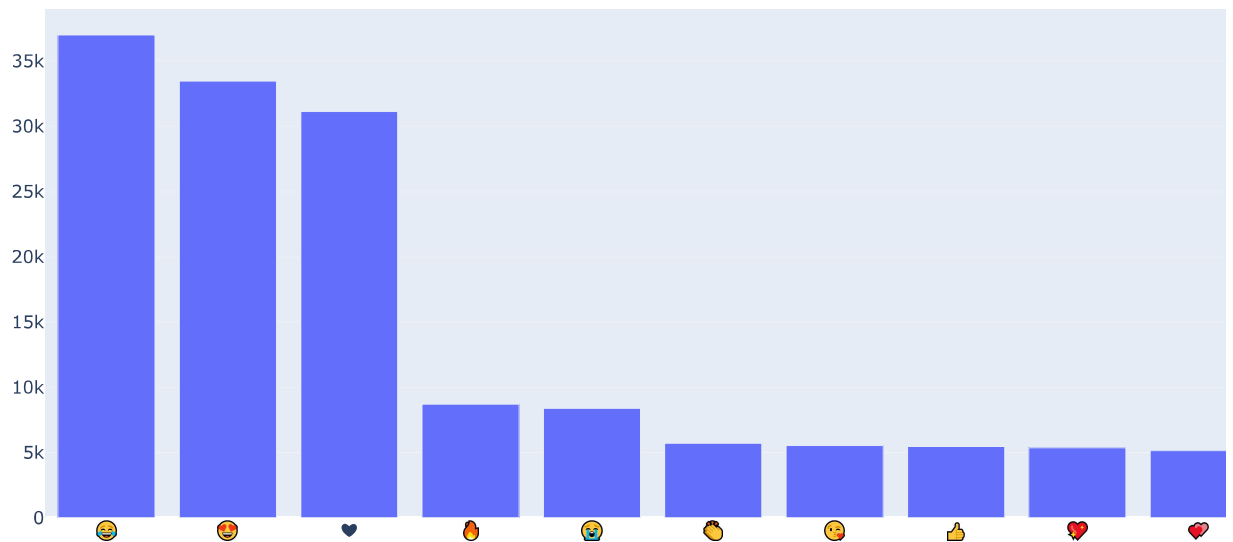
```
In [55]: emojis = [Counter(all_emojis_list).most_common(10)[i][0] for i in range(10)]
```

```
In [56]: freqs = [Counter(all_emojis_list).most_common(10)[i][1] for i in range(10)]
```

```
In [57]: import plotly.graph_objs as go
from plotly.offline import iplot
```

```
In [58]: trace = go.Bar(x=emojis , y=freqs)
```

```
In [59]: iplot([trace])
```



```
In [60]: ## Conclusions : Majority of the customers are happy as most of them are using emojis Like: funny , Love , heart ,et
```

5. Collect Entire data of Youtube !

```
In [61]: import os
```

```
In [62]: files= os.listdir(r'C:\Users\lenov\Desktop\Coding\Projects\1\additional_data-20231210T182859Z-001\additional_data')
```

```
In [63]: files
```

```
Out[63]: ['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 [64]: ## extracting csv files only from above list ..

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

```
In [65]: files_csv
```

```
Out[65]: ['CAvideos.csv',
          'DEvideos.csv',
          'FRvideos.csv',
          'GBvideos.csv',
          'INvideos.csv',
          'JPvideos.csv',
          'KRvideos.csv',
          'MXvideos.csv',
          'RUvideos.csv',
          'USvideos.csv']
```

```
In [66]: import warnings
from warnings import filterwarnings
filterwarnings('ignore')
```

```
In [67]: #different types of encoding-->
#Note : encoding may change depending upon data , country data , sometimes regional data as well.
#For more information on Encoding -- Follow below
# https://docs.python.org/3/library/codecs.html#standard-encodings%C2%B6
```

```
In [68]: full_df = pd.DataFrame()
path = r'C:\Users\lenov\Desktop\Coding\Projects\1\additional_data-20231210T182859Z-001\additional_data'

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

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

```
In [69]: full_df.shape
```

```
Out[69]: (375942, 16)
```

6. How to export data into (csv, json, db)

```
In [70]: #export your data into      a) csv      b) json      c) db
```



```
In [71]: full_df[full_df.duplicated()].shape
```

```
Out[71]: (36417, 16)
```

```
In [72]: #Storing data into csv ..  
full_df[0:1000].to_csv(r'C:\Users\lenov\Desktop\Coding\Projects\1\youtube_sample.csv' , index=False)
```

```
In [73]: #Storing data into json ..  
full_df[0:1000].to_json(r'C:\Users\lenov\Desktop\Coding\Projects\1\youtube_sample.json')
```

```
In [74]: # Storing data into database
```

```
In [75]: #create engine allows us to connect to database  
from sqlalchemy import create_engine
```

```
In [76]: # Lets create sqlalchemy engine by using create_engine method ie create engine allows us to connect to database  
engine = create_engine(r'sqlite:///C:\Users\lenov\Desktop\Coding\Projects\1\youtube_sample.sqlite')
```

```
In [77]: ### we will store first 1000 rows into Users table..  
full_df[0:1000].to_sql('Users' , con=engine , if_exists='append')
```

```
Out[77]: 1000
```

7.. Which Category has the maximum likes ?

```
In [78]: full_df.head(5)
```

```
Out[78]:
```

	video_id	trending_date	title	channel_title	category_id	publish_time	tags	views
0	n1WpP7iowLc	17.14.11	Eminem - Walk On Water (Audio) ft. Beyoncé	EminemVEVO	10	2017-11-10T17:00:03.000Z	Eminem "Walk " "On " "Water " "Aftermath/Shady/In...	171585
1	0dBlkQ4Mz1M	17.14.11	PLUSH - Bad Unboxing Fan Mail	iDubbbzTV	23	2017-11-13T17:00:00.000Z	plush "bad unboxing " "unboxing " "fan mail " "id...	10146
2	5qpjK5DgCt4	17.14.11	Racist Superman Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	2017-11-12T19:05:24.000Z	racist superman "rudy " "mancuso " "king " "bach"...	31914
3	d380meD0W0M	17.14.11	I Dare You: GOING BALD!?	nigahiga	24	2017-11-12T18:01:41.000Z	ryan "higa " "higatv " "nigahiga " "i dare you " "...	20958
4	2Vv-BfVoq4g	17.14.11	Ed Sheeran - Perfect (Official Music Video)	Ed Sheeran	10	2017-11-09T11:04:14.000Z	edsheeran "ed sheeran " "acoustic " "live " "cove...	335236

```
In [79]: full_df['category_id'].unique()
```

```
Out[79]: array([10, 23, 24, 25, 22, 26, 1, 28, 20, 17, 29, 15, 19, 2, 27, 43, 30,  
44], dtype=int64)
```

```
In [80]: ## Lets read json file ..  
json_df = pd.read_json(r'C:\Users\lenov\Desktop\Coding\Projects\1\additional_data-20231210T182859Z-001\additional_d...
```

```
In [81]: json_df
```

```
Out[81]:
```

	kind	etag	items
0	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
1	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
2	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
3	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
4	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
5	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
6	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
7	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
8	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
9	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
10	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
11	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
12	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
13	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
14	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
15	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
16	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
17	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
18	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
19	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
20	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
21	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
22	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
23	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
24	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
25	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
26	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
27	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
28	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
29	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
30	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...
31	youtube#videoCategoryListResponse	"m2yskBQFythfE4irbTleOgYYfBU/S730Ilt-Fi-emsQJv...	{'kind': 'youtube#videoCategory', 'etag': '"m2...

```
In [82]: json_df['items'][0]
```

```
### each row of 'Items' feature is dictionary ..
```

```
Out[82]: {'kind': 'youtube#videoCategory',
          'etag': '"m2yskBQFythfE4irbTleOgYYfBU/Xy1mB4_yLrHy_BmKmpBgty2mZQ"',
          'id': '1',
          'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
                      'title': 'Film & Animation',
                      'assignable': True}}
```

```
In [83]: json_df['items'][1]
```

```
Out[83]: {'kind': 'youtube#videoCategory',
          'etag': '"m2yskBQFythfE4irbTleOgYYfBU/UZ1oLIIZ2dxIhO45ZTFR3a3NyTA"',
          'id': '2',
          'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
                      'title': 'Autos & Vehicles',
                      'assignable': True}}
```

```
In [84]: cat_dict = {}

for item in json_df['items'].values:
    ## cat_dict[key] = value (Syntax to insert key:value in dictionary)
    cat_dict[int(item['id'])] = item['snippet']['title']
```

```
In [85]: cat_dict
```

Out[85]: {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 [86]: full_df['category_name']=full_df['category_id'].map(cat_dict)
```

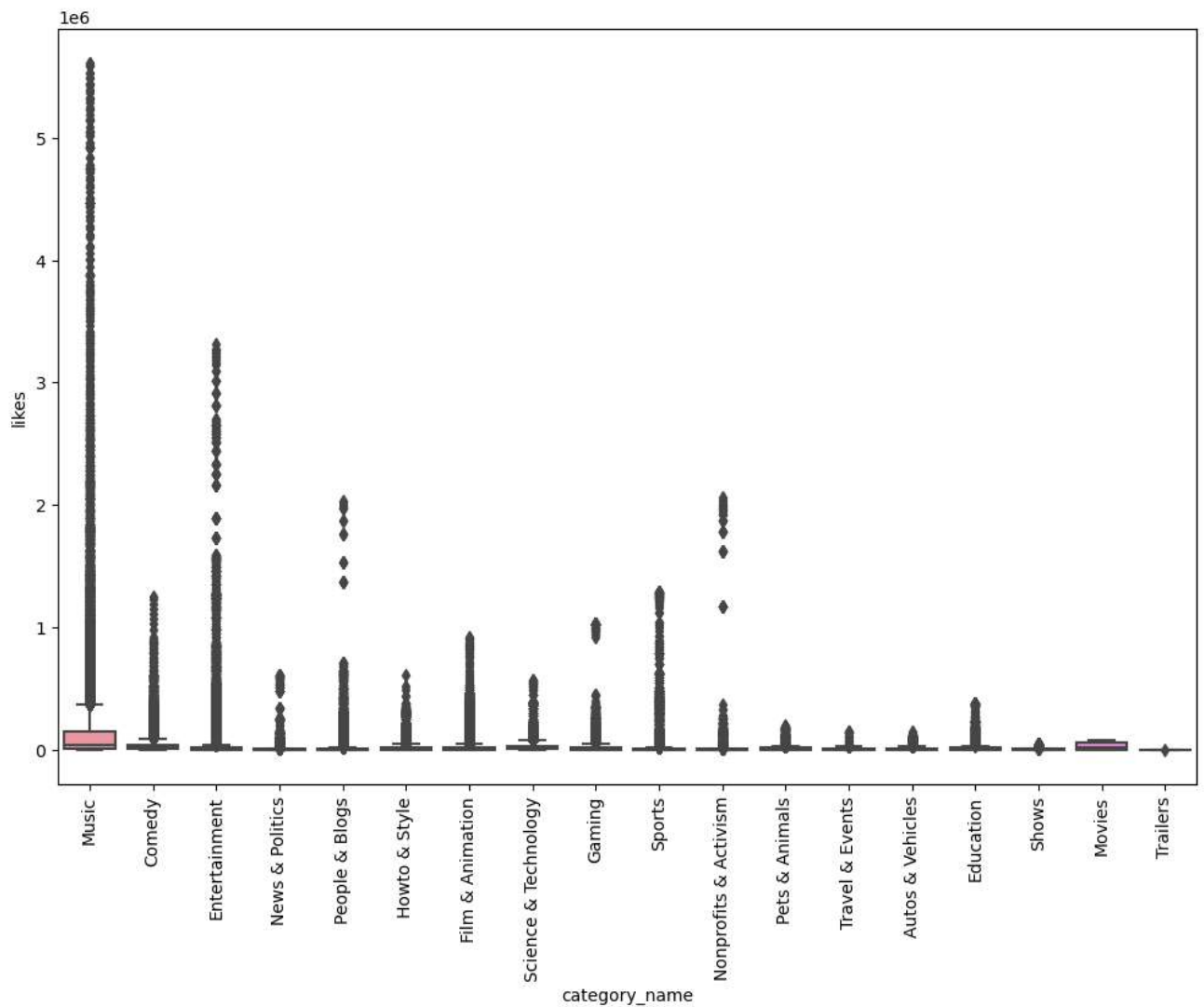
```
In [87]: full_df.head(4)
```

Out[87]:

likes	comment_count	thumbnail_link	comments_disabled	ratings_disabled	video_error_or_removed	description
3420	125882	https://i.ytimg.com/vi/n1WpP7iowLc/default.jpg	False	False	False	Eminem's new track Walk on Water ft. BeyoncÃ© ...
1688	13030	https://i.ytimg.com/vi/0dBlkQ4Mz1M/default.jpg	False	False	False	STill got a lot of packages. Probably will las...
5339	8181	https://i.ytimg.com/vi/5qpjK5DgCt4/default.jpg	False	False	False	WATCH MY PREVIOUS VIDEO â \\n\\nSUBSCRIBE â ...
1989	17518	https://i.ytimg.com/vi/d380meD0W0M/default.jpg	False	False	False	I know it's been a while since we did this sho...

```
In [88]: plt.figure(figsize=(12,8))
sns.boxplot(x='category_name' , y='likes' , data=full_df)
plt.xticks(rotation='vertical')
```

```
Out[88]: (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')])
```



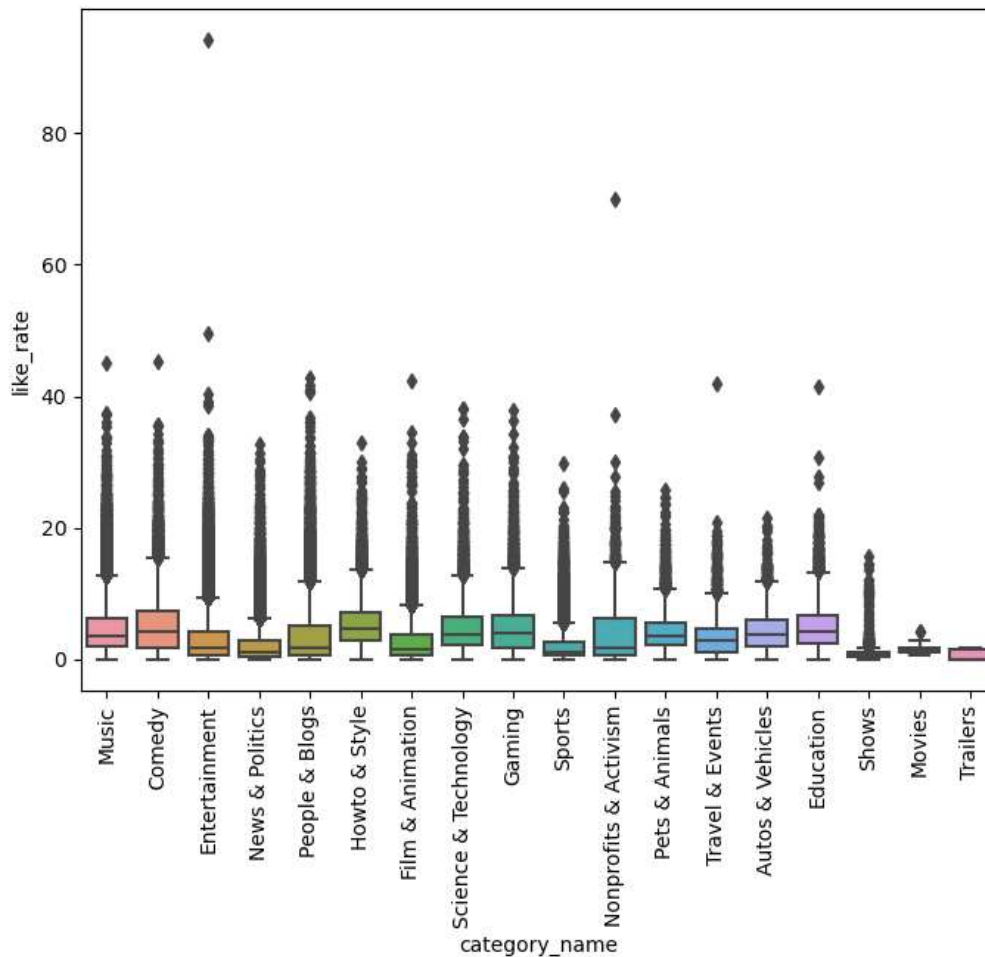
8.. Find out whether audience is engaged or not

```
In [89]: 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 [90]: full_df.columns
```

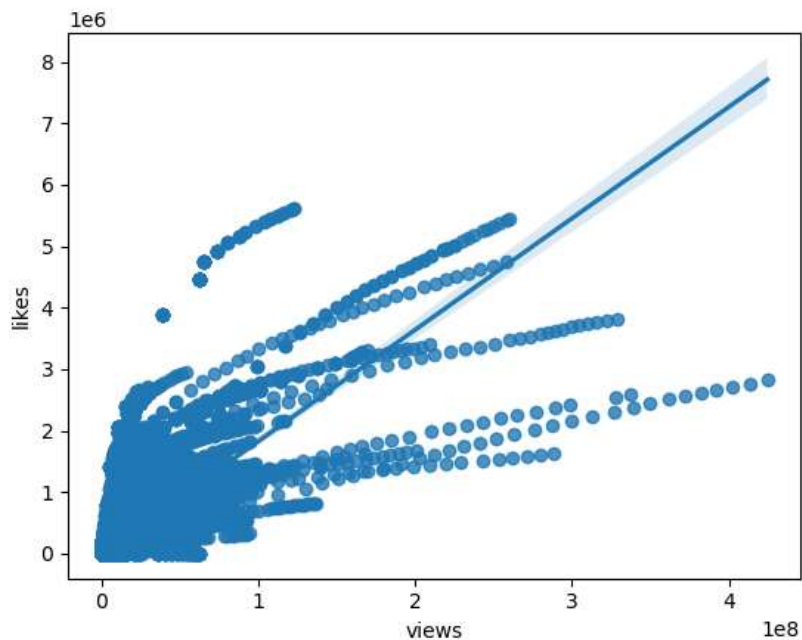
```
Out[90]: Index(['video_id', 'trending_date', 'title', 'channel_title', 'category_id',
'publish_time', 'tags', 'views', 'likes', 'dislikes', 'comment_count',
'thumbnail_link', 'comments_disabled', 'ratings_disabled',
'video_error_or_removed', 'description', 'category_name', 'like_rate',
'dislike_rate', 'comment_count_rate'],
dtype='object')
```

```
In [91]: plt.figure(figsize=(8,6))
sns.boxplot(x='category_name', y='like_rate', data=full_df)
plt.xticks(rotation='vertical')
plt.show()
```



```
In [92]: ## analysing relationship between views & likes  
sns.regplot(x='views' , y='likes' , data = full_df)
```

```
Out[92]: <AxesSubplot:xlabel='views', ylabel='likes'>
```



```
In [93]: full_df.columns
```

```
Out[93]: 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')
```

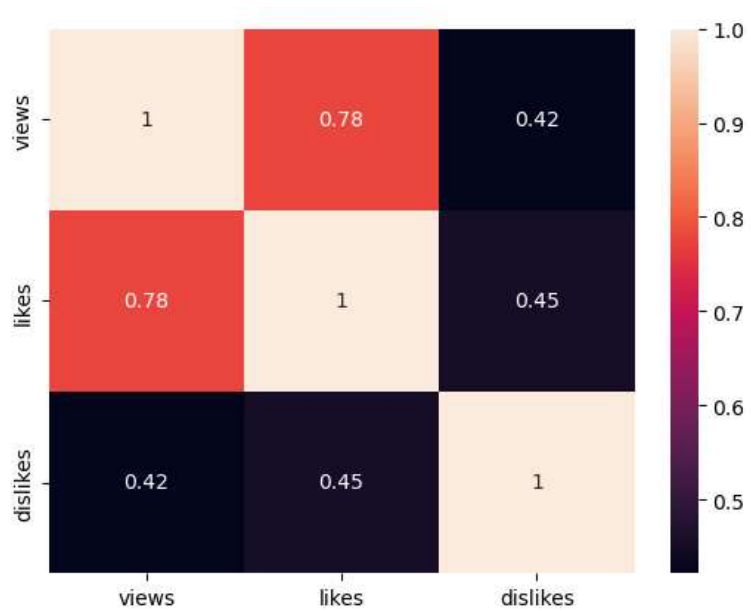
```
In [94]: full_df[['views', 'likes', 'dislikes']].corr() ### finding co-relation values between ['views', 'likes', 'dislikes']
```

```
Out[94]:
```

	views	likes	dislikes
views	1.000000	0.777796	0.421653
likes	0.777796	1.000000	0.453710
dislikes	0.421653	0.453710	1.000000

```
In [95]: sns.heatmap(full_df[['views', 'likes', 'dislikes']].corr() , annot=True)
```

Out[95]: <AxesSubplot:>



9. Which channels have the largest number of trending videos?

```
In [96]: full_df.head(6)
```

Out[96]:

	video_id	trending_date	title	channel_title	category_id	publish_time	tags	views
0	n1WpP7iowLc	17.14.11	Eminem - Walk On Water (Audio) ft. Beyonc�	EminemVEVO	10	2017-11-10T17:00:03.000Z	Eminem "Walk "On "Water "Aftermath/Shady/In...	171585
1	0dBikQ4Mz1M	17.14.11	PLUSH - Bad Unboxing Fan Mail	iDubbbzTV	23	2017-11-13T17:00:00.000Z	plush "bad unboxing "unboxing "fan mail "id...	10146
2	5qpjK5DgCt4	17.14.11	Racist Superman Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	2017-11-12T19:05:24.000Z	racist superman "rudy "mancuso "king "bach"...	31914
3	d380meD0W0M	17.14.11	I Dare You: GOING BALD!?	nigahiga	24	2017-11-12T18:01:41.000Z	ryan "higa "higatv "nigahiga "i dare you "...	20958
4	2Vv-BfVoq4g	17.14.11	Ed Sheeran - Perfect (Official Music Video)	Ed Sheeran	10	2017-11-09T11:04:14.000Z	edsheeran "ed sheeran "acoustic "live "cove...	335236
5	0yIWz1XEeyc	17.14.11	Jake Paul Says Alissa Violet CHEATED with LOGA...	DramaAlert	25	2017-11-13T07:37:51.000Z	#DramaAlert "Drama "Alert "DramaAlert "keem...	13096

```
In [97]: full_df['channel_title'].value_counts()
```

```
Out[97]: The Late Show with Stephen Colbert    984
WWE                                             804
Late Night with Seth Meyers                   773
VikatanTV                                     763
TheEllenShow                                 743
...
DFC Orrivals                                 1
haiblubbbblubb                               1
SOYER                                         1
GOLD CLAN                                    1
Herr Zymny                                   1
Name: channel_title, Length: 37824, dtype: int64
```

```
In [98]: ### Lets obtain above frequency table using groupby approach :
cdf = full_df.groupby(['channel_title']).size().sort_values(ascending=False).reset_index()
```

```
In [99]: cdf = cdf.rename(columns={0: 'total_videos'})
```

```
In [100]: cdf
```

```
Out[100]:
```

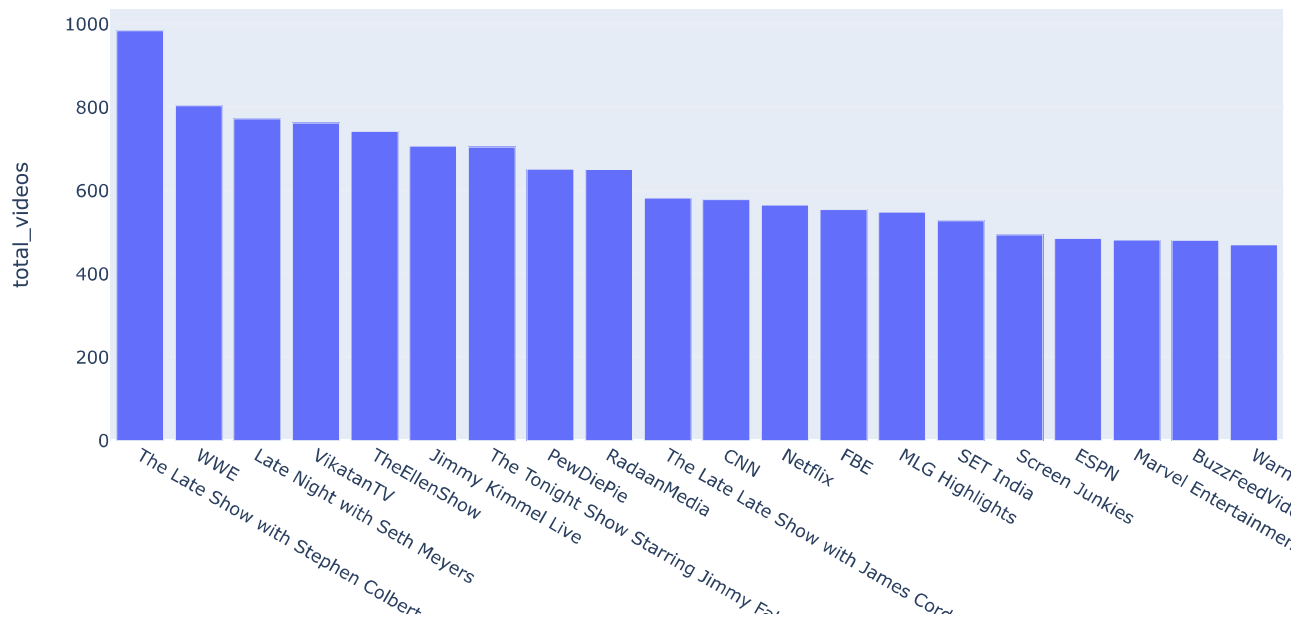
	channel_title	total_videos
0	The Late Show with Stephen Colbert	984
1	WWE	804
2	Late Night with Seth Meyers	773
3	VikatanTV	763
4	TheEllenShow	743
...
37819	LIGHTS - 001 jrnj	1
37820	bangtanist	1
37821	LIGAMX Femenil	1
37822	LIGA COLOMBIANA OFICIAL	1
37823	Pavel Sidorik TV	1

37824 rows × 2 columns

```
In [101]: import plotly.express as px
```



```
In [102]: px.bar(data_frame=cdf[0:20] , x='channel_title' , y='total_videos')
```



10. Does Punctuations in title and tags have any relation with views, likes, dislikes comments?

```
In [103]: full_df['title'][0]
```

```
Out[103]: 'Eminem - Walk On Water (Audio) ft. BeyoncÃ'
```

```
In [104]: import string
```

```
In [105]: string.punctuation
```

```
Out[105]: '!"#$%&'()*+,-./:;<=>?@[\\]^_`{|}~'
```

```
In [106]: len([char for char in full_df['title'][0] if char in string.punctuation])
```

```
Out[106]: 4
```

```
In [107]: def punc_count(text):
            return len([char for char in text if char in string.punctuation])
```

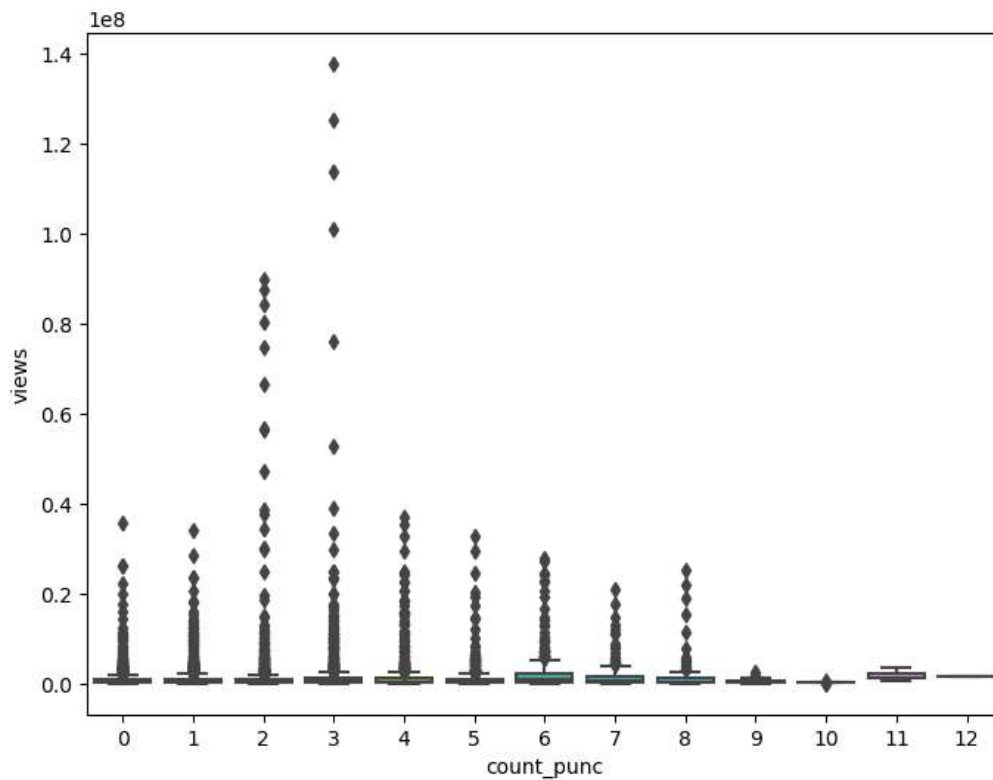
```
In [108]: sample = full_df[0:10000]
```

```
In [109]: sample['count_punc'] = sample['title'].apply(punc_count)
```

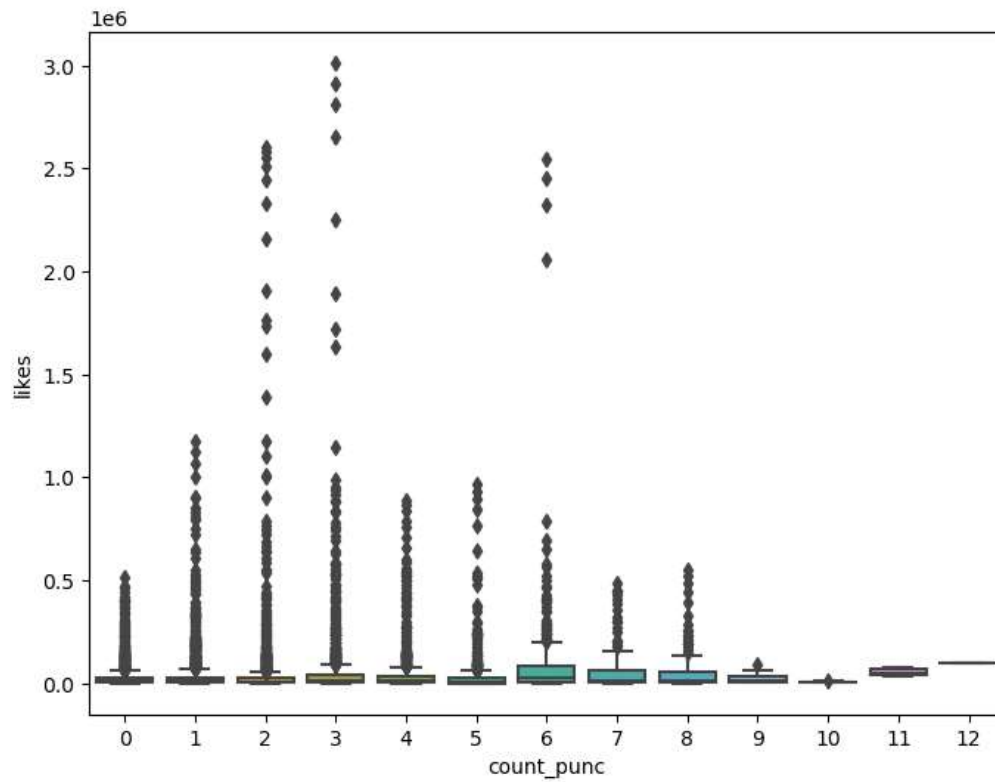
```
In [110]: sample['count_punc']
```

```
Out[110]: 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 [111]: plt.figure(figsize=(8,6))
          sns.boxplot(x='count_punc' , y='views' , data=sample)
          plt.show()
```



```
In [112]: plt.figure(figsize=(8,6))
sns.boxplot(x='count_punc' , y='likes' , data=sample)
plt.show()
```



```
In [ ]:
```