Youtube Data Analysis

June 26, 2024

[23]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

Dataset Import

[6]:
```
data=pd.read_csv(r"D:\Analytics Projects\Python\Jupyter\UScomments.
↪csv",on_bad_lines='skip')
```

[7]:
```
data.shape
```

[7]:
```
(10000, 4)
```

[8]:
```
data.head()
```

[8]:
```
      video_id                                      comment_text  likes  \
0  XpVt6Z1Gjjo                  Logan Paul it's yo big day         4
1  XpVt6Z1Gjjo  I've been following you from the start of your…      3
2  XpVt6Z1Gjjo                 Say hi to Kong and maverick for me      3
3  XpVt6Z1Gjjo                              MY FAN . attendance      3
4  XpVt6Z1Gjjo                                        trending       3

   replies
0        0
1        0
2        0
3        0
4        0
```

Data Cleansing

[9]:
```
data.isnull().sum()
```

[9]:
```
video_id        0
comment_text    1
likes           0
replies         0
dtype: int64
```

[10]:
```
data.dropna(inplace=True)
```

[11]:
```
data.isnull().sum()
```

[11]:
```
video_id        0
comment_text    0
likes           0
replies         0
dtype: int64
```

Sentiment Analysis using TextBlob

[12]:
```
from textblob import TextBlob
```

[13]:
```
data.shape
```

[13]:
```
(9999, 4)
```

[14]:
```
polarity=[]
for comment in data['comment_text']:
    try:
        polarity.append(TextBlob(comment).sentiment.polarity)
    except:
        polarity.append(0)
```

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

[15]:
```
9999
```

[16]:
```
data['polarity'] = polarity
```

[17]:
```
data.head()
```

[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

Wordcloud Analysis - Graphical Representation of Text Frequency

Complete Positive & Negative Polarity

[18]:
```
Filter1=data['polarity']==1
```

[19]:
```
Comments_Positive=data[Filter1]
```

[20]:
```
Filter2=data['polarity']==-1
```

[21]:
```
Comments_Negative=data[Filter2]
```

[22]:
```
Comments_Positive.head(5)
```

[22]:

	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

[23]:
```
from wordcloud import WordCloud, STOPWORDS
```

[24]:
```
set(STOPWORDS)
```

[24]:
```
{'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'}
```

[25]:
```
type(data['comment_text'])
```

[25]:
```
pandas.core.series.Series
```

[26]:
```
total_comments_positive=' '.join(Comments_Positive['comment_text'])
```

[27]:
```
#To WordCloud the data needs to be passed as String Data Structure but␣
 ↪currently it is in Series Data Structure
wordcloud_positive=WordCloud(stopwords=set(STOPWORDS)).
 ↪generate(total_comments_positive)
```

[28]:
```
plt.imshow(wordcloud_positive)
plt.axis('off')
```

[28]: (-0.5, 399.5, 199.5, -0.5)

[29]:
```
total_comments_negative=' '.join(Comments_Negative['comment_text'])
```

[30]:
```
wordcloud_negative=WordCloud(stopwords=set(STOPWORDS)).
↪generate(total_comments_negative)
```

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

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

Emoji Analysis

[32]:
```
import emoji
```

[33]:
```
emoji.__version__
```

[33]:
```
'2.12.1'
```

[34]:
```
emoji_list=[]
for comment in data['comment_text'].dropna():
    for char in comment:
        if char in emoji.EMOJI_DATA:
            emoji_list.append(char)
```

[35]:
```
#First 10 Emojis
emoji_list[0:10]
```

[35]:
```
[' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ']
```

[36]:
```
#Get the count of each emojis
from collections import Counter
```

[37]:
```
Counter(emoji_list).most_common(10)
```

[37]:
```
[(' ', 610),
 (' ', 323),
 (' ', 268),
 (' ', 68),
 (' ', 68),
 (' ', 56),
 (' ', 54),
 (' ', 52),
 (' ', 51),
 (' ', 47)]
```

[38]:
```
#Emojis List
emojis = [Counter(emoji_list).most_common(10)[i][0] for i in range(10)]
```

[39]:
```
#Emoji Frequency
Frequencies = [Counter(emoji_list).most_common(10)[i][1] for i in range(10)]
```

[50]:
```
import plotly.graph_objs as go
from plotly.offline import iplot
```

[51]:
```
trace = go.Bar(x=emojis , y=Frequencies)
```

[52]:
```
iplot([trace])
```

Entire Data Collection

[1]:
```
import os
```

[13]:
```
files=os.listdir(r"D:\Analytics Projects\Python\Jupyter\additional_data")
```

[14]:
```
files
```

[14]:
```
['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']
```

[16]:
```
files_csv=[file for file in files if '.csv' in file]
```

[17]:
```
files_csv
```

[17]:
```
['CAvideos.csv',
 'DEvideos.csv',
 'FRvideos.csv',
 'GBvideos.csv',
 'INvideos.csv',
 'JPvideos.csv',
 'KRvideos.csv',
 'MXvideos.csv',
 'RUvideos.csv',
 'USvideos.csv']
```

[19]:
```
import warnings
from warnings import filterwarnings
filterwarnings('ignore')
```

[44]:
```
full_df=pd.DataFrame()
path=r"D:\Analytics Projects\Python\Jupyter\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)
```

[45]:
```
full_df.shape
```

[45]: (375942, 16)

Export Data to csv,Json and Databases

[46]:
```
# Check number of duplicate rows in full Data Frame
full_df[full_df.duplicated()].shape
```

[46]: (36417, 16)

[47]:
```
# Remove duplicate rows in full Data Frame
full_df = full_df.drop_duplicates()
```

[48]:
```
# Check number of duplicate rows in full Data Frame after removing Duplicates
full_df.shape
```

[48]: (339525, 16)

[49]:
```
sample_full_df=full_df[0:1000]
```

[50]:
```
sample_full_df.shape
```

[50]: (1000, 16)

[51]:
```
sample_full_df.to_csv(r'D:\Analytics Projects\Python\Jupyter/youtube_sample.
 ↪csv', index=False)
```

[52]:
```
sample_full_df.to_json(r'D:\Analytics Projects\Python\Jupyter/youtube_sample.
 ↪json')
```

[54]:
```
from sqlalchemy import create_engine
```

[56]:
```
engine=create_engine(r'sqlite:///D:\Analytics Projects\Python\Jupyter/
 ↪youtube_sample.sqlite')
```

[57]:
```
full_df[0:1000].to_sql('Users',con=engine, if_exists='append')
```

[57]: 1000

Analysing most liked Category

[58]:
```
full_df.head()
```

[58]:

```
      video_id trending_date  \
0  n1WpP7iowLc      17.14.11
1  0dBIkQ4Mz1M      17.14.11
2  5qpjK5DgCt4      17.14.11
3  d380meD0W0M      17.14.11
4  2Vv-BfVoq4g      17.14.11

                                               title channel_title  \
0        Eminem - Walk On Water (Audio) ft. BeyoncÃ©    EminemVEVO
1                         PLUSH - Bad Unboxing Fan Mail     iDubbbzTV
2  Racist Superman | Rudy Mancuso, King Bach & Le…  Rudy Mancuso
3                            I Dare You: GOING BALD!?      nigahiga
4        Ed Sheeran - Perfect (Official Music Video)    Ed Sheeran

   category_id              publish_time  \
0           10  2017-11-10T17:00:03.000Z
1           23  2017-11-13T17:00:00.000Z
2           23  2017-11-12T19:05:24.000Z
3           24  2017-11-12T18:01:41.000Z
4           10  2017-11-09T11:04:14.000Z

                                                tags     views    likes  \
0  Eminem|"Walk"|"On"|"Water"|"Aftermath/Shady/In…  17158579   787425
1  plush|"bad unboxing"|"unboxing"|"fan mail"|"id…   1014651   127794
2  racist superman|"rudy"|"mancuso"|"king"|"bach"…   3191434   146035
3  ryan|"higa"|"higatv"|"nigahiga"|"i dare you"|"…   2095828   132239
4  edsheeran|"ed sheeran"|"acoustic"|"live"|"cove…  33523622  1634130

   dislikes  comment_count                                thumbnail_link  \
0     43420         125882  https://i.ytimg.com/vi/n1WpP7iowLc/default.jpg
1      1688          13030  https://i.ytimg.com/vi/0dBIkQ4Mz1M/default.jpg
2      5339           8181  https://i.ytimg.com/vi/5qpjK5DgCt4/default.jpg
3      1989          17518  https://i.ytimg.com/vi/d380meD0W0M/default.jpg
4     21082          85067  https://i.ytimg.com/vi/2Vv-BfVoq4g/default.jpg

   comments_disabled  ratings_disabled  video_error_or_removed  \
0              False             False                   False
1              False             False                   False
2              False             False                   False
3              False             False                   False
4              False             False                   False

                                        description
0  Eminem's new track Walk on Water ft. BeyoncÃ© …
1  STill got a lot of packages. Probably will las…
```

```
2  WATCH MY PREVIOUS VIDEO â ¶ \n\nSUBSCRIBE â º …
3  I know it's been a while since we did this sho…
4  ð §: https://ad.gt/yt-perfect\nð °: https://…
```

[59]:
```
#we only have Category Id but not Category Name
full_df['category_id'].unique()
```

[59]:
```
array([10, 23, 24, 25, 22, 26,  1, 28, 20, 17, 29, 15, 19,  2, 27, 43, 30,
       44], dtype=int64)
```

[60]:
```
json_df=pd.read_json(r'D:\Analytics␣
↪Projects\Python\Jupyter\additional_data\US_category_id.json')
```

[62]:
```
json_df['items'] #Its a dictionary
```

[62]:
```
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
```

[65]:
```
json_df['items'][0]
```

[65]:
```
{'kind': 'youtube#videoCategory',
 'etag': '"m2yskBQFythfE4irbTIeOgYYfBU/Xy1mB4_yLrHy_BmKmPBggty2mZQ"',
 'id': '1',
 'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
  'title': 'Film & Animation',
  'assignable': True}}
```

[67]:
```
Cat_dict={}
```

[71]:
```
cat_dict={}

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

[72]:
```
cat_dict
```

[72]:
```
{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'}
```

[74]:
```
full_df['category_name']=full_df['category_id'].map(cat_dict)
```

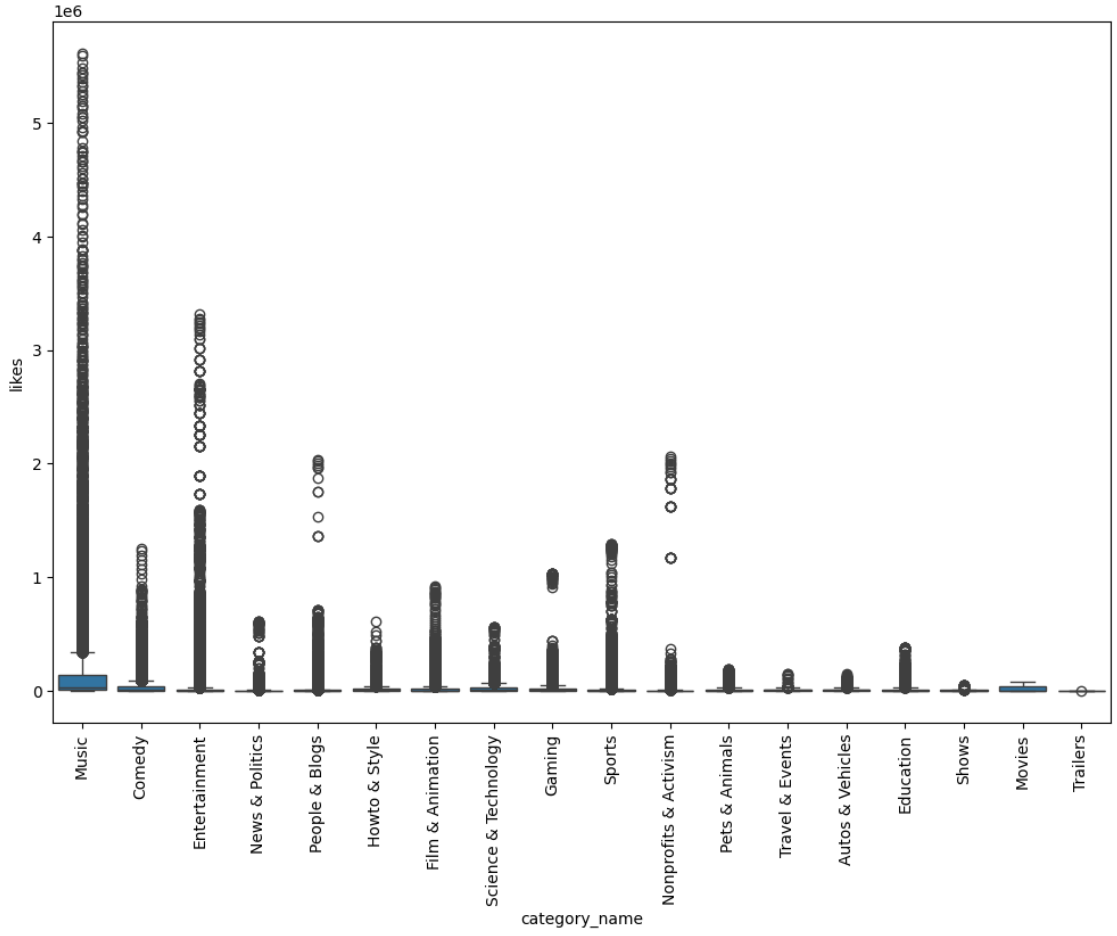
[75]:
```
full_df.head(4)
```

[75]:

```
      video_id trending_date  \
0  n1WpP7iowLc      17.14.11
1  0dBIkQ4Mz1M      17.14.11
2  5qpjK5DgCt4      17.14.11
3  d38OmeDOWOM      17.14.11

                                               title channel_title  \
0       Eminem - Walk On Water (Audio) ft. BeyoncÃ©    EminemVEVO
1                        PLUSH - Bad Unboxing Fan Mail     iDubbbzTV
2  Racist Superman | Rudy Mancuso, King Bach & Le…  Rudy Mancuso
3                           I Dare You: GOING BALD!?      nigahiga

   category_id              publish_time  \
0           10  2017-11-10T17:00:03.000Z
1           23  2017-11-13T17:00:00.000Z
2           23  2017-11-12T19:05:24.000Z
3           24  2017-11-12T18:01:41.000Z

                                              tags     views   likes  \
0  Eminem|"Walk"|"On"|"Water"|"Aftermath/Shady/In…  17158579  787425
1  plush|"bad unboxing"|"unboxing"|"fan mail"|"id…   1014651  127794
2  racist superman|"rudy"|"mancuso"|"king"|"bach"…   3191434  146035
3  ryan|"higa"|"higatv"|"nigahiga"|"i dare you"|"…   2095828  132239

   dislikes  comment_count                                thumbnail_link  \
0     43420         125882  https://i.ytimg.com/vi/n1WpP7iowLc/default.jpg
1      1688          13030  https://i.ytimg.com/vi/0dBIkQ4Mz1M/default.jpg
2      5339           8181  https://i.ytimg.com/vi/5qpjK5DgCt4/default.jpg
3      1989          17518  https://i.ytimg.com/vi/d38OmeDOWOM/default.jpg

   comments_disabled  ratings_disabled  video_error_or_removed  \
0              False             False                   False
1              False             False                   False
2              False             False                   False
3              False             False                   False

                                        description  category_name
```

```
0  Eminem's new track Walk on Water ft. BeyoncÃ© …          Music
1  STill got a lot of packages. Probably will las…         Comedy
2  WATCH MY PREVIOUS VIDEO â ¶ \n\nSUBSCRIBE â º …         Comedy
3  I know it's been a while since we did this sho…  Entertainment
```

[83]:
```
plt.figure(figsize=(12,8))
sns.boxplot(x='category_name',y='likes',data=full_df)
plt.xticks(rotation='vertical')
plt.show()
```



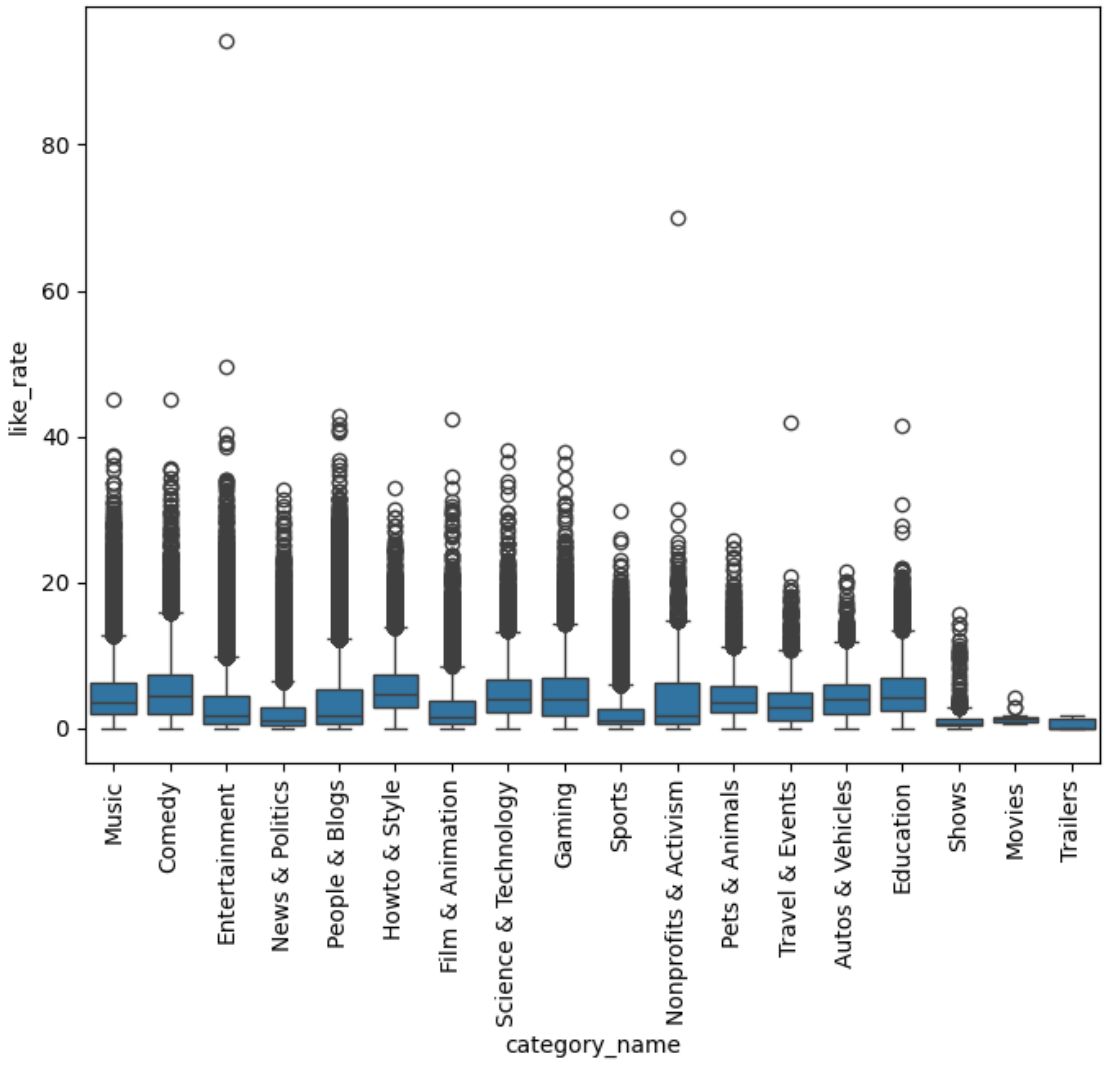
Analyse Whether audience are engaged or not

[80]:
```
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
```

[81]:
```
full_df.columns
```

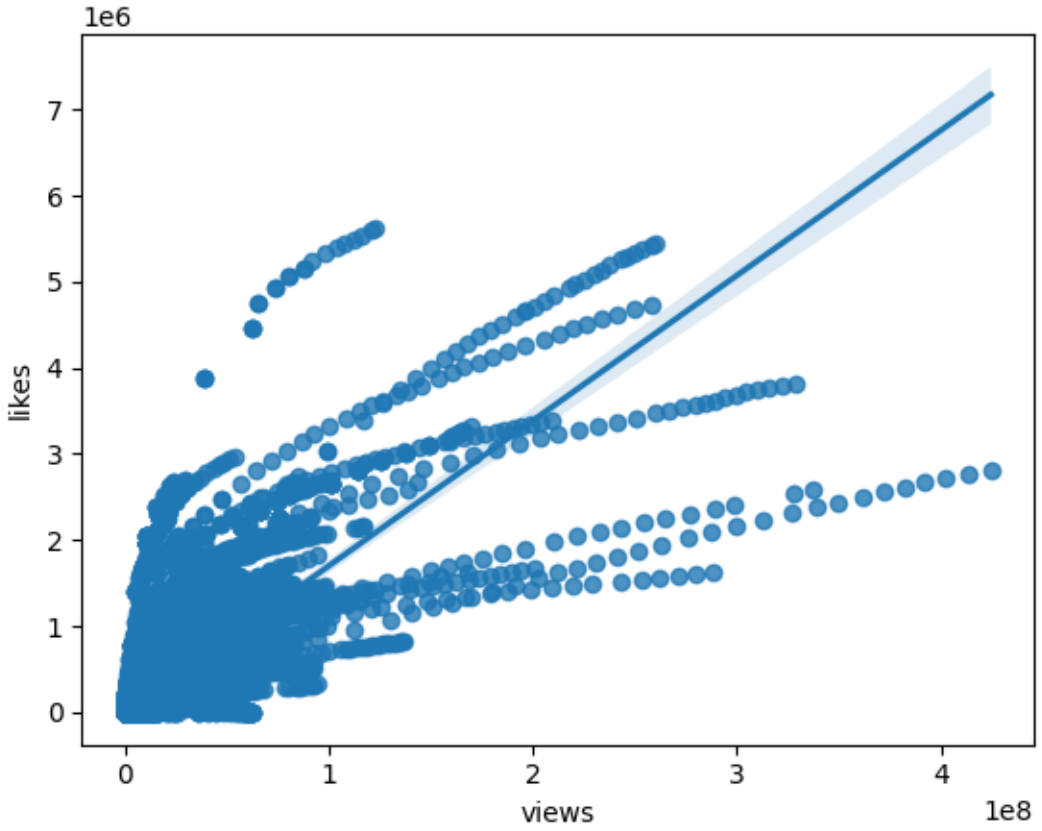
[81]:
```
Index(['video_id', 'trending_date', 'title', 'channel_title', 'category_id',
       'publish_time', 'tags', 'views', 'likes', 'dislikes', 'comment_count',
       'thumbnail_link', 'comments_disabled', 'ratings_disabled',
       'video_error_or_removed', 'description', 'category_name', 'like_rate',
       'dislike_rate', 'comment_count_rate'],
      dtype='object')
```

[82]:
```
plt.figure(figsize=(8,6))
sns.boxplot(x='category_name',y='like_rate',data=full_df)
plt.xticks(rotation='vertical')
plt.show()
```



[84]:
```
sns.regplot(x='views',y='likes',data=full_df)
```

[84]: <Axes: xlabel='views', ylabel='likes'>



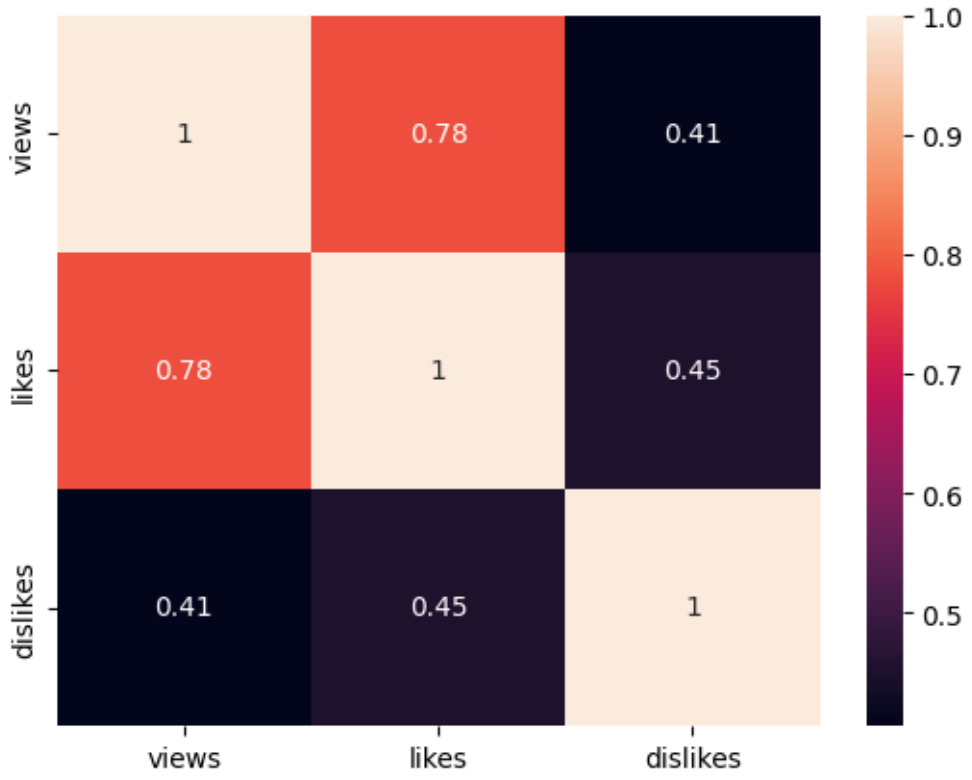
[88]:
```
#Get Correlation between views, likes and dislikes as table
full_df[['views','likes','dislikes']].corr()
```

[88]:

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

[89]:
```
#Get Correlation between views, likes and dislikes as heatmap
sns.heatmap(full_df[['views','likes','dislikes']].corr(), annot=True)
```

[89]: <Axes: >



Analyze Trending Videos by channels

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

[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')
```

[94]:
```
#Get channel count
full_df['channel_title'].value_counts()
```

[94]:
```
channel_title
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: count, Length: 37824, dtype: int64
```

[96]:
```
cdf=full_df.groupby(['channel_title']).size().sort_values(ascending=False).
↪reset_index()
```

[97]:
```
cdf
```

[97]:

	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 x 2 columns]

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

[100]:
```
cdf
```

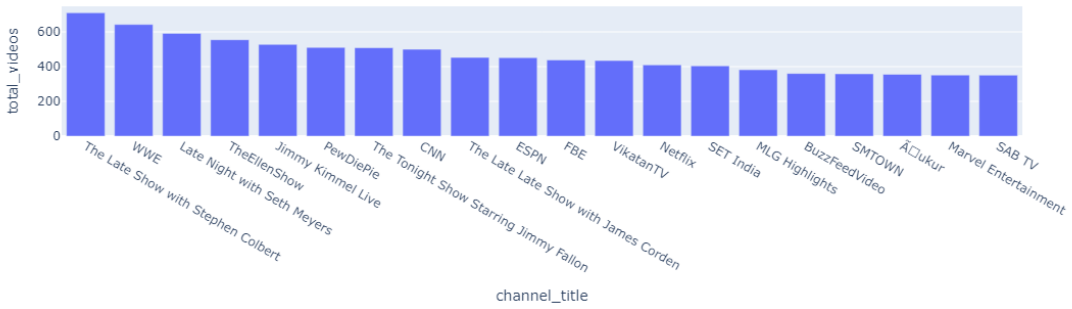
[100]:

	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 x 2 columns]

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

[109]:
```
#Top 20 channel with more videos
px.bar(data_frame=cdf[0:20],x='channel_title', y='total_videos')
```



Analyse if Punctuations in title and tags have any relation with views, dislikes and comments

[110]:
```
full_df['title'][0]
```

[110]:
```
'Eminem - Walk On Water (Audio) ft. BeyoncÃ©'
```

[111]:
```
import string
```

[112]:
```
string.punctuation
```

[112]:
```
'!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

[114]:
```
len([char for char in full_df['title'][0] if char in string.punctuation])
```

[114]:
```
4
```

[115]:
```
def punc_count(text):
    return len([char for char in text if char in string.punctuation])
```

[116]:
```
sample_df=full_df[0:10000]
```

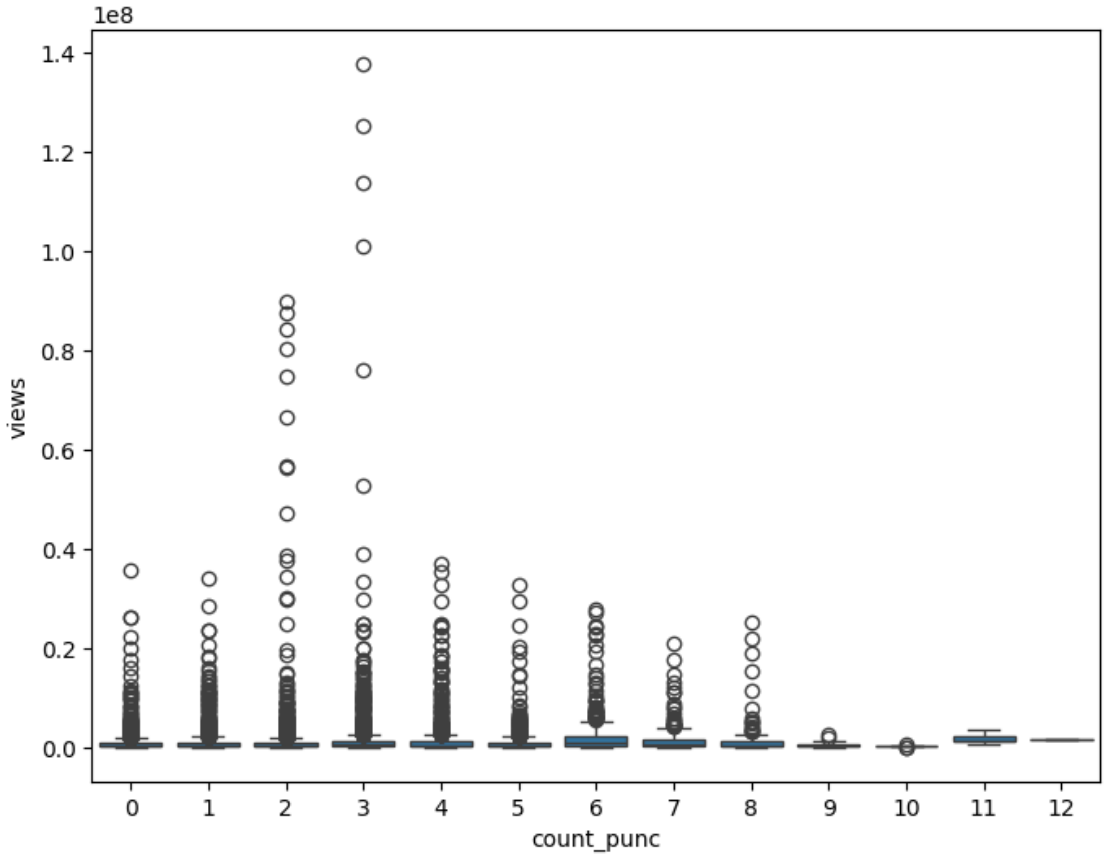
[118]:
```
sample_df['count_punc']= sample_df['title'].apply(punc_count)
```

[119]:
```
sample_df['count_punc']
```

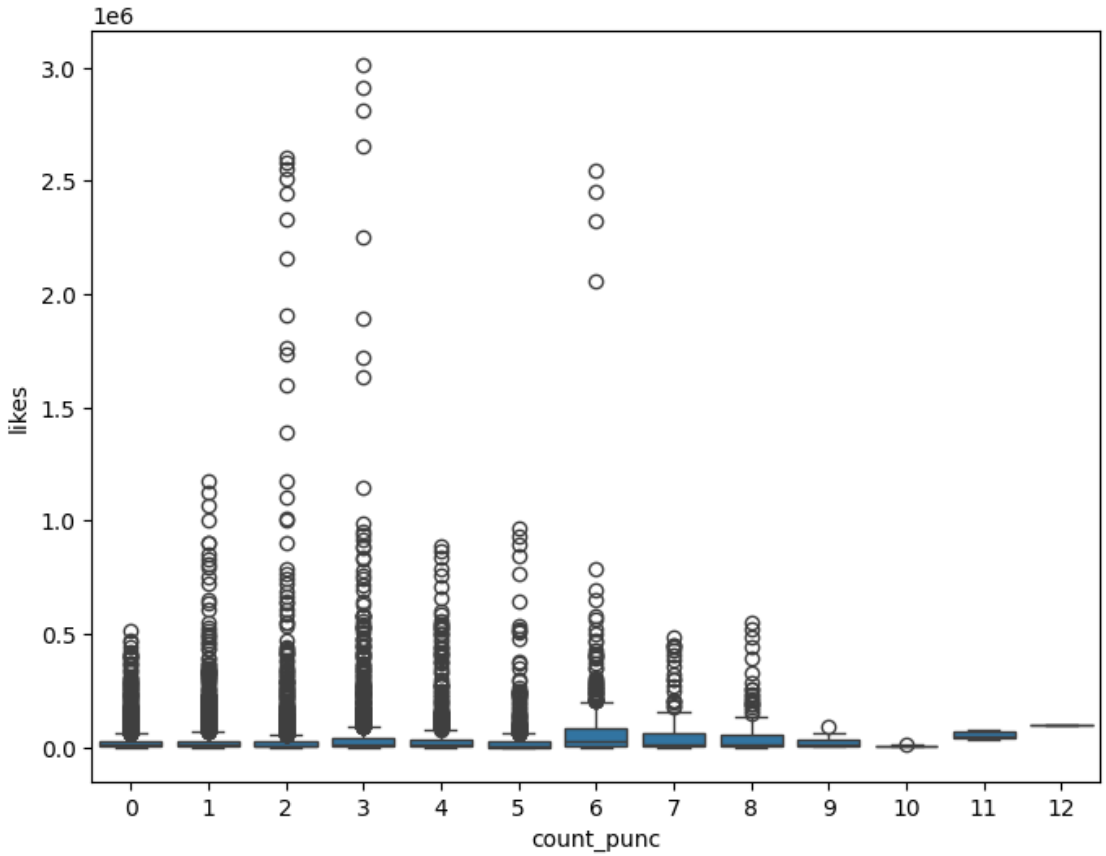
[119]:
```
0       4
1       1
2       3
3       3
4       3
       ..
9995    6
```

```
9996    0
9997    1
9998    0
9999    6
Name: count_punc, Length: 10000, dtype: int64
```

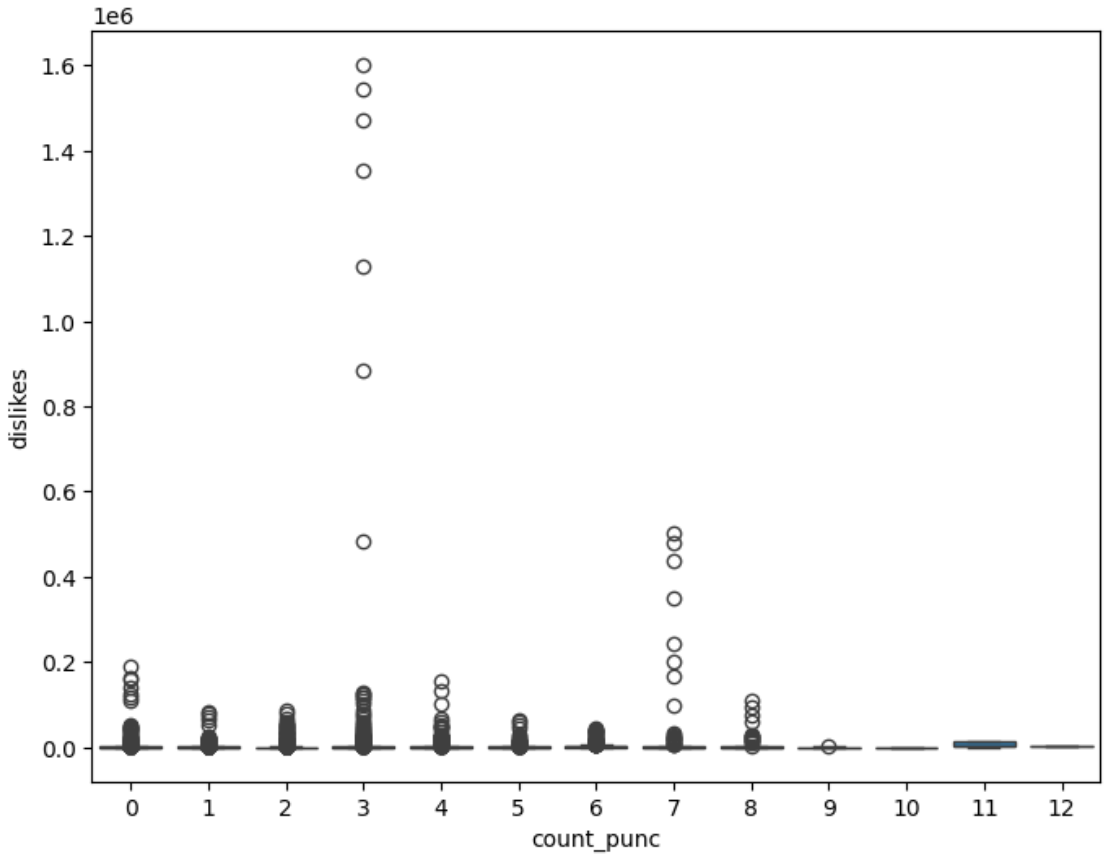
[120]:
```
plt.figure(figsize=(8,6))
sns.boxplot(x='count_punc',y='views',data=sample_df)
plt.show()
```



[121]:
```
plt.figure(figsize=(8,6))
sns.boxplot(x='count_punc',y='likes',data=sample_df)
plt.show()
```



[123]:
```
plt.figure(figsize=(8,6))
sns.boxplot(x='count_punc',y='dislikes',data=sample_df)
plt.show()
```



[122]:
```
sample_df[['count_punc','views','likes','dislikes']].corr()
```

[122]:

	count_punc	views	likes	dislikes
count_punc	1.000000	0.065100	0.066798	0.036223
views	0.065100	1.000000	0.881644	0.662583
likes	0.066798	0.881644	1.000000	0.530162
dislikes	0.036223	0.662583	0.530162	1.000000

[ ]: