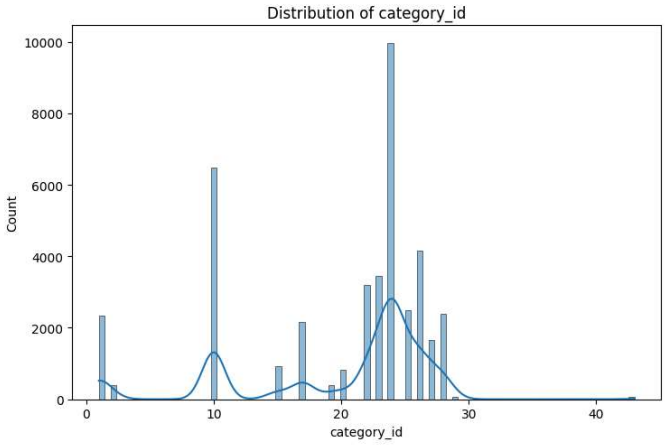
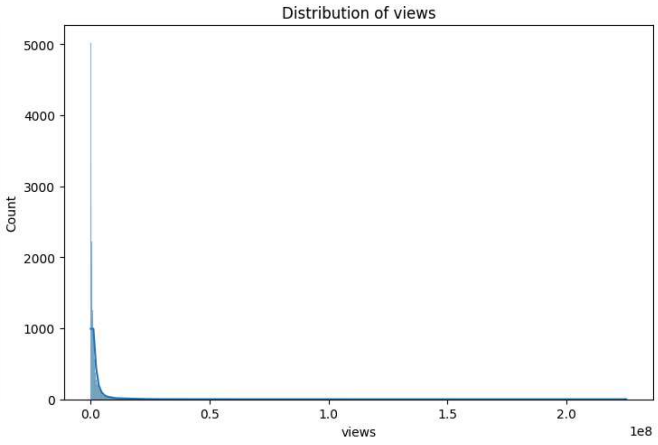
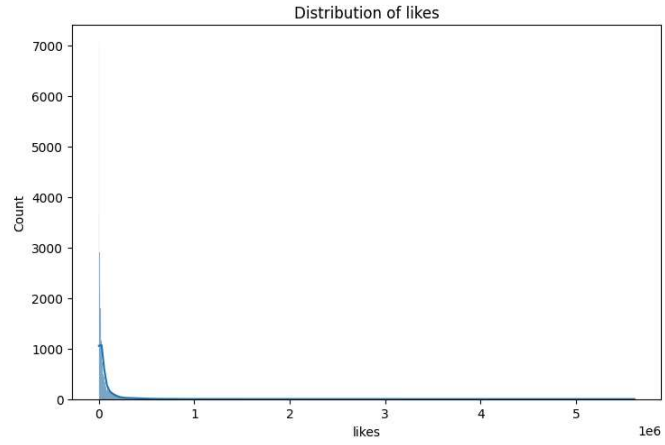
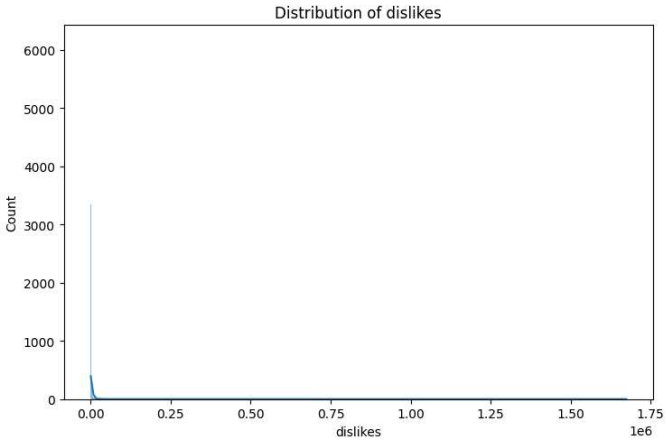
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from scipy import stats

# Load the dataset (replace 'your_dataset.csv' with the actual file name)
df = pd.read_csv('USvideos.csv')

# Display basic information about the dataset
print(df.info())
print("\nSummary statistics:")
print(df.describe())

# Create histograms for numerical variables
numerical_cols = df.select_dtypes(include=[np.number]).columns
fig, axes = plt.subplots(nrows=(len(numerical_cols) + 1) // 2, ncols=2, figsize=(15, 5 * ((len(numerical_cols) + 1) // 2)))
axes = axes.flatten()

for i, col in enumerate(numerical_cols):
    sns.histplot(data=df, x=col, kde=True, ax=axes[i])
    axes[i].set_title(f'Distribution of {col}')

plt.tight_layout()
plt.show()

# Create box plots to identify outliers
fig, axes = plt.subplots(nrows=(len(numerical_cols) + 1) // 2, ncols=2, figsize=(15, 5 * ((len(numerical_cols) + 1) // 2)))
axes = axes.flatten()

for i, col in enumerate(numerical_cols):
    sns.boxplot(data=df, y=col, ax=axes[i])
    axes[i].set_title(f'Box plot of {col}')

plt.tight_layout()
plt.show()

# Create a correlation heatmap
correlation_matrix = df[numerical_cols].corr()
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1, center=0)
plt.title('Correlation Heatmap')
plt.show()

# Identify and print highly correlated variable pairs
high_corr_pairs = []
for i in range(len(numerical_cols)):
    for j in range(i + 1, len(numerical_cols)):
        corr = correlation_matrix.iloc[i, j]
        if abs(corr) > 0.7:  # You can adjust this threshold
            high_corr_pairs.append((numerical_cols[i], numerical_cols[j], corr))

print("\nHighly correlated variable pairs:")
for pair in high_corr_pairs:
    print(f"{pair[0]} and {pair[1]}: {pair[2]:.2f}")

# Create scatter plots for highly correlated pairs
for pair in high_corr_pairs:
    plt.figure(figsize=(10, 6))
    sns.scatterplot(data=df, x=pair[0], y=pair[1])
    plt.title(f'Scatter plot: {pair[0]} vs {pair[1]}')
    plt.show()

# Identify potential outliers using Z-score
z_scores = np.abs(stats.zscore(df[numerical_cols]))
potential_outliers = (z_scores > 3).any(axis=1)
print(f"\nNumber of potential outliers: {potential_outliers.sum()}")
print("Rows with potential outliers:")
print(df[potential_outliers])
```

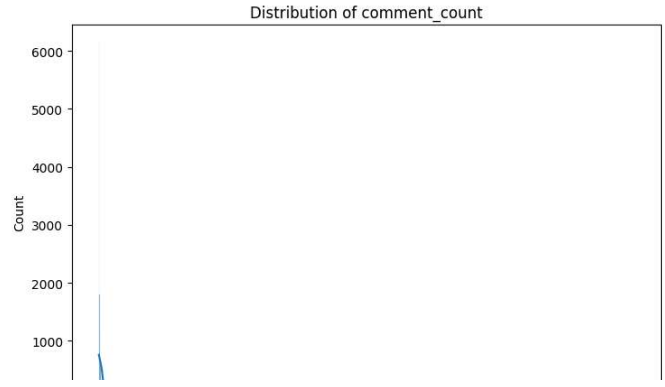
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40949 entries, 0 to 40948
Data columns (total 16 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   video_id                40949 non-null  object
 1   trending_date           40949 non-null  object
 2   title                   40949 non-null  object
 3   channel_title           40949 non-null  object
 4   category_id             40949 non-null  int64
 5   publish_time            40949 non-null  object
 6   tags                    40949 non-null  object
 7   views                   40949 non-null  int64
 8   likes                   40949 non-null  int64
 9   dislikes                40949 non-null  int64
 10  comment_count           40949 non-null  int64
 11  thumbnail_link          40949 non-null  object
 12  comments_disabled       40949 non-null  bool
 13  ratings_disabled        40949 non-null  bool
 14  video_error_or_removed  40949 non-null  bool
 15  description             40379 non-null  object
dtypes: bool(3), int64(5), object(8)
memory usage: 4.2+ MB
None

Summary statistics:
        category_id         views         likes      dislikes  comment_count
count  40949.000000  4.094900e+04  4.094900e+04  4.094900e+04   4.094900e+04
mean      19.972429  2.360785e+06  7.426670e+04  3.711401e+03   8.446804e+03
std        7.568327  7.394114e+06  2.288853e+05  2.902971e+04   3.743049e+04
min        1.000000  5.490000e+02  0.000000e+00  0.000000e+00   0.000000e+00
25%       17.000000  2.423290e+05  5.424000e+03  2.020000e+02   6.140000e+02
50%       24.000000  6.818610e+05  1.809100e+04  6.310000e+02   1.856000e+03
75%       25.000000  1.823157e+06  5.541700e+04  1.938000e+03   5.755000e+03
max       43.000000  2.252119e+08  5.613827e+06  1.674420e+06   1.361580e+06
```

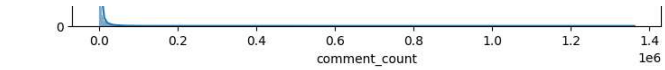


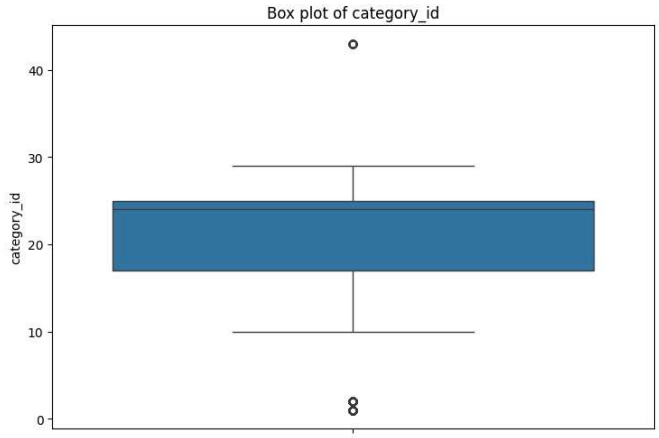


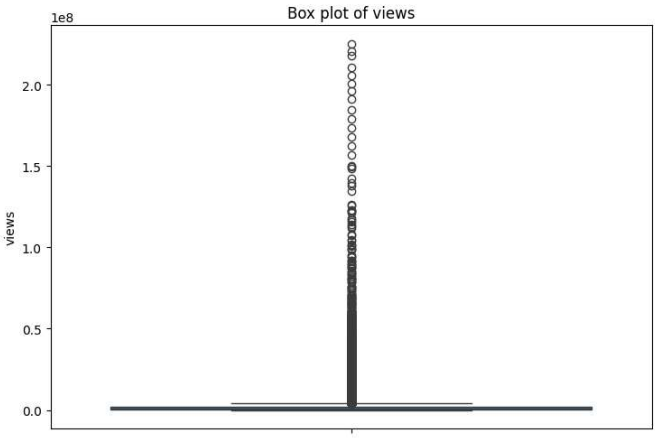


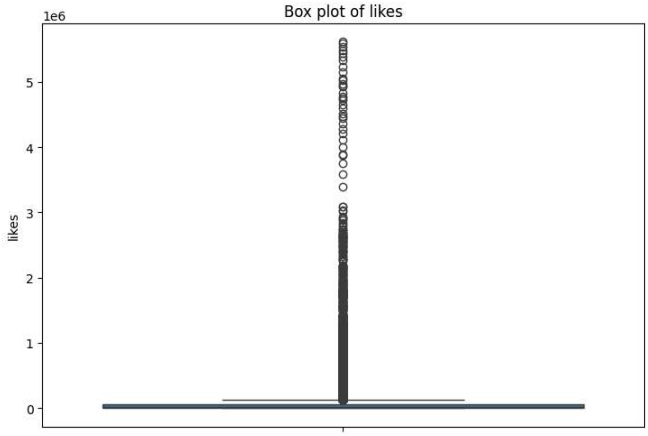


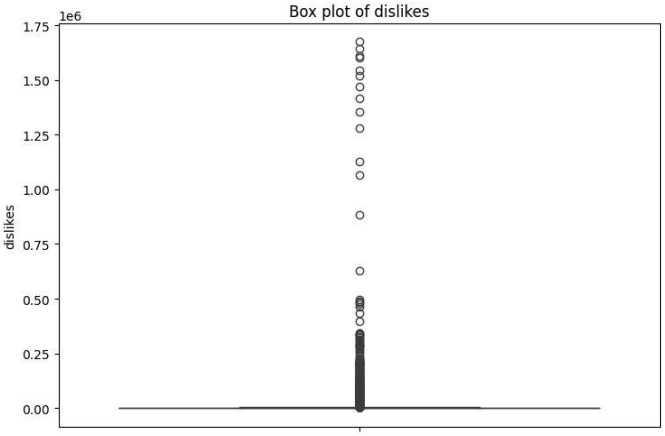


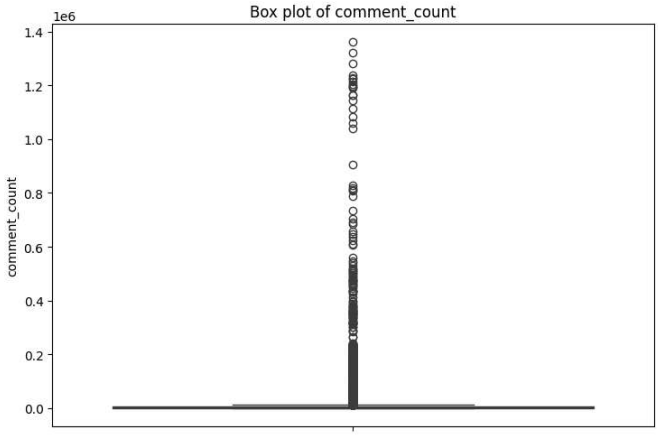

comment_count
1e6

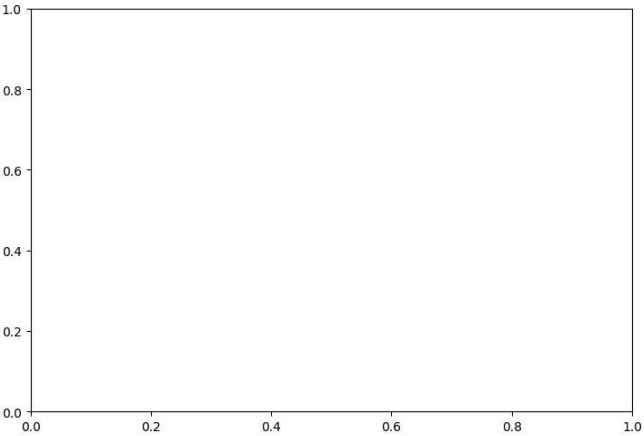

Box plot of category_id
category_id

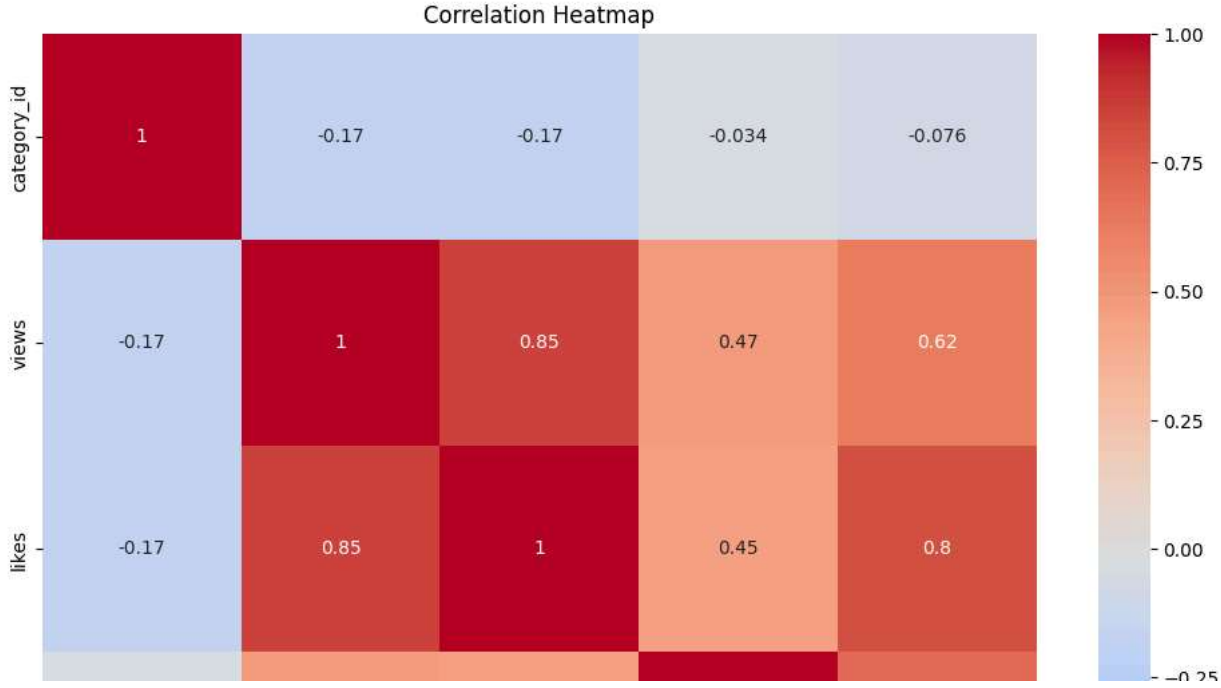

Box plot of views
1e8
views

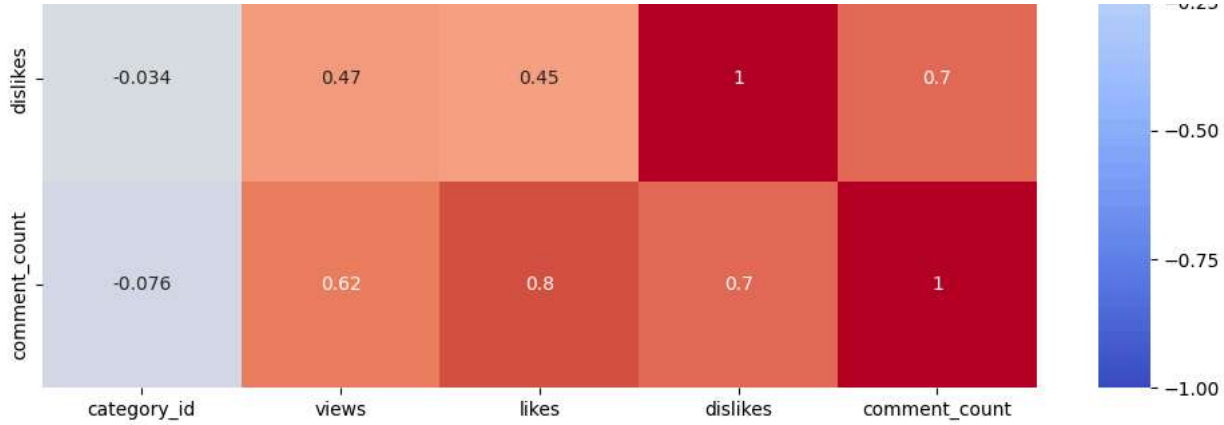

Box plot of likes
1e6
likes


Box plot of dislikes
1e6
dislikes

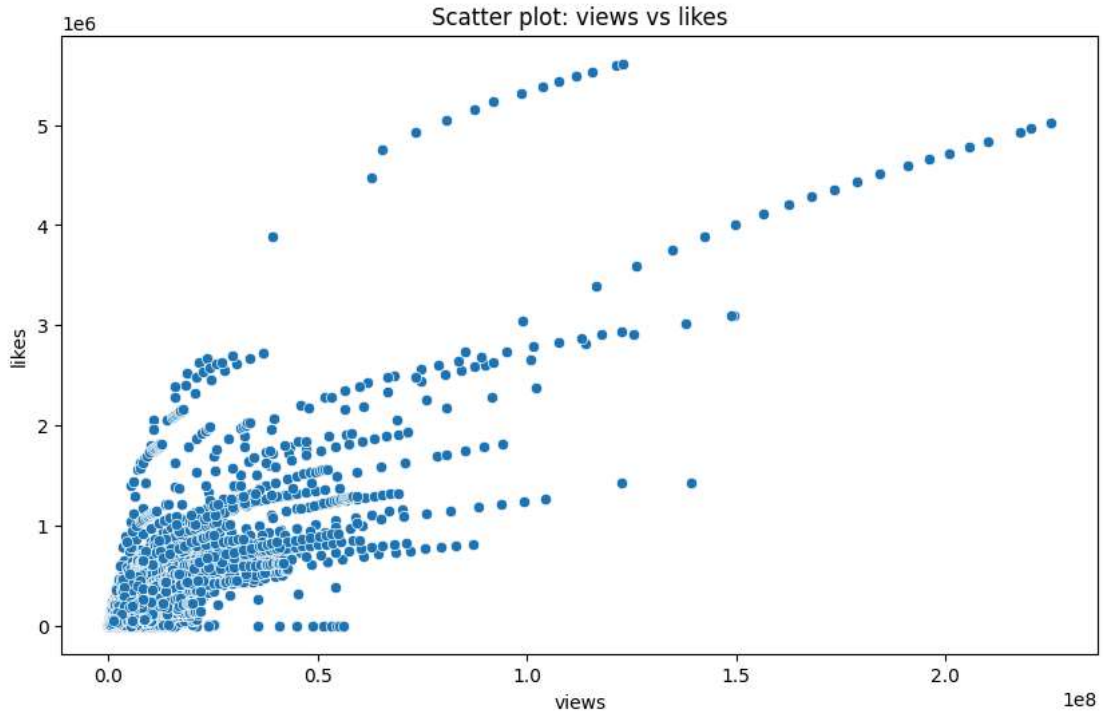

Box plot of comment_count
1e6
comment_count

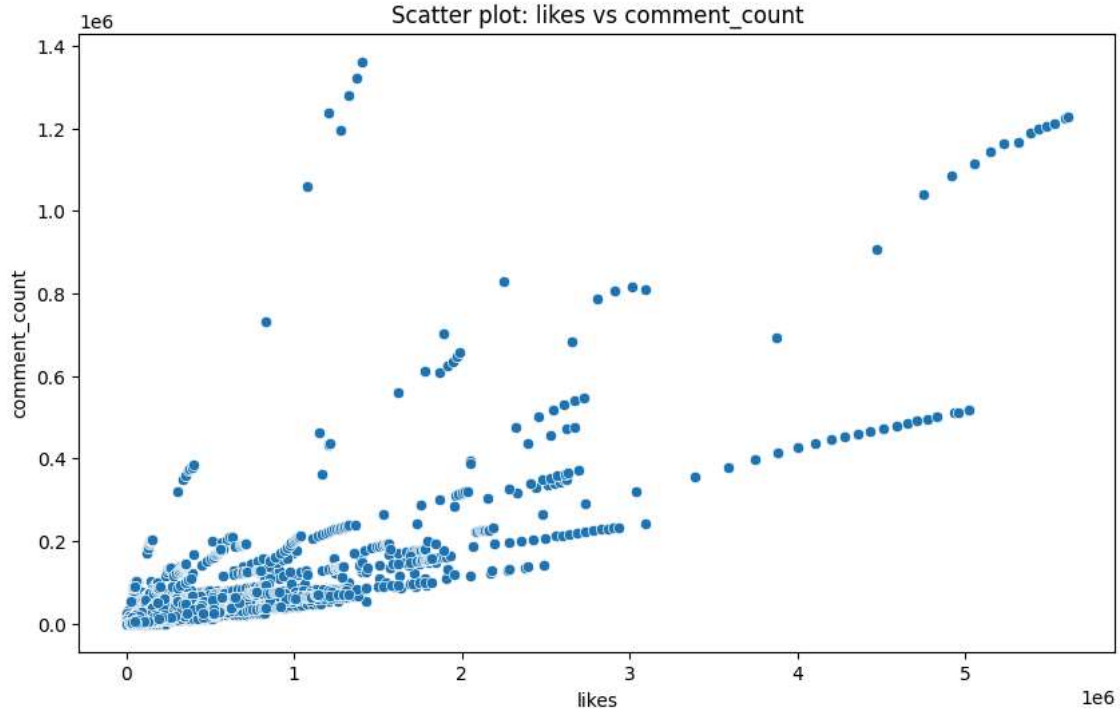


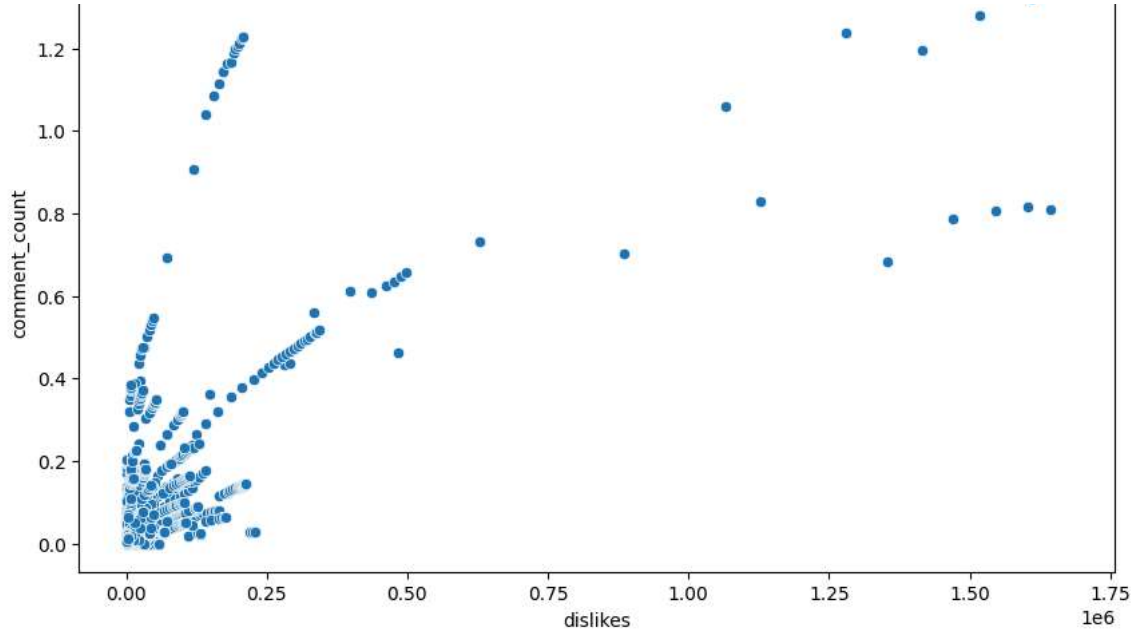

Correlation Heatmap
category_id
views
likes
1
-0.17
-0.17
-0.034
-0.076
-0.17
1
0.85
0.47
0.62
-0.17
0.85
1
0.45
0.8
1.00
0.75
0.50
0.25
0.00




```
Highly correlated variable pairs:
views and likes: 0.85
likes and comment_count: 0.80
dislikes and comment_count: 0.70
```








```
Number of potential outliers: 920
Rows with potential outliers:
         video_id trending_date  \
32     n1WpP7iowLc      17.14.11
70     2Vv-BfVoq4g      17.14.11
149    9wg3v-01yKQ      17.14.11
298    n1WpP7iowLc      17.15.11
336    2Vv-BfVoq4g      17.15.11
...            ...           ...
40896  nQySbNGu4g0      18.14.06
40907  aEM2kOrrNJI      18.14.06
40925  QgOXIEhHU1Y      18.14.06
40938  n_W54baizX8      18.14.06
40948  ooyjaVdt-jA      18.14.06

                                            title      channel_title  \
32            Eminem - Walk On Water (Audio) ft. Beyoncé         EminemVEVO
70           Ed Sheeran - Perfect (Official Music Video)         Ed Sheeran
149                                Harry Styles - Kiwi    HarryStylesVEVO
298           Eminem - Walk On Water (Audio) ft. Beyoncé         EminemVEVO
336          Ed Sheeran - Perfect (Official Music Video)         Ed Sheeran
...                                                ...                ...
40896  [CHOREOGRAPHY] BTS (방탄소년단) 'FAKE LOVE' Dance P...           BANGTANTV
40907     Jennifer Lopez - Dinero ft. DJ Khaled, Cardi B  JenniferLopezVEVO
40925  Diplo, French Montana & Lil Pump ft. Zhavia - ...              Diplo
40938               Daddy Yankee - Hielo (Video Oficial)       Daddy Yankee
40948  Official Call of Duty®: Black Ops 4 — Multipla...       Call of Duty

       category_id              publish_time  \
32              10  2017-11-10T17:00:03.000Z
70              10  2017-11-09T11:04:14.000Z
149             10  2017-11-08T13:00:01.000Z
298             10  2017-11-10T17:00:03.000Z
336             10  2017-11-09T11:04:14.000Z
...            ...                       ...
40896           10  2018-05-27T11:00:03.000Z
40907           10  2018-05-24T12:00:02.000Z
40925           10  2018-05-21T14:12:22.000Z
40938           10  2018-05-18T14:00:04.000Z
40948           20  2018-05-17T17:09:38.000Z

                                                  tags     views    likes  \
32     Eminem|"Walk"|"On"|"Water"|"Aftermath/Shady/In...  17158531   787419
70     edsheeran|"ed sheeran"|"acoustic"|"live"|"cove...  33523622  1634124
149                Columbia|"Harry Styles"|"Kiwi"|"Pop"   9632678   810895
298    Eminem|"Walk"|"On"|"Water"|"Aftermath/Shady/In...  20539417   840642
336    edsheeran|"ed sheeran"|"acoustic"|"live"|"cove...  39082222  1721383
...                                                  ...       ...      ...
40896  방탄소년단|"BTS"|"BANGTAN"|"HIPHOP"|"랩몬스터"|"RapMons...  11381059  1141726
40907  jennifer lopez|"jlo"|"jennifer lopez live"|"jl...  30599645   455949
40925  Welcome to the party|"Diplo"|"Lil Pump"|"Frenc...  40087764   835657
40938  daddy yankee reggaeton|"daddy yankee youtube"|...  41803845   628861
40948      call of duty|"cod"|"activision"|"Black Ops 4"  10306119   357079

       dislikes  comment_count  \
32        43420         125882
70        21082          85067
149       16139          59473
298       47715         124236
336       23137          90352
...         ...            ...
40896      4696          69934
```

```
40907     42374          25679
40925     25283          38305
40938     42833          39363
40948    212976         144795

                                      thumbnail_link  comments_disabled  \
32     https://i.ytimg.com/vi/n1WpP7iowLc/default.jpg              False
70     https://i.ytimg.com/vi/2Vv-BfVoq4g/default.jpg              False
149    https://i.ytimg.com/vi/9wg3v-01yKQ/default.jpg              False
298    https://i.ytimg.com/vi/n1WpP7iowLc/default.jpg              False
336    https://i.ytimg.com/vi/2Vv-BfVoq4g/default.jpg              False
...                                              ...                ...
40896  https://i.ytimg.com/vi/nQySbNGu4g0/default.jpg              False
40907  https://i.ytimg.com/vi/aEM2kOrrNJI/default.jpg              False
40925  https://i.ytimg.com/vi/QgOXIEhHU1Y/default.jpg              False
40938  https://i.ytimg.com/vi/n_W54baizX8/default.jpg              False
40948  https://i.ytimg.com/vi/ooyjaVdt-jA/default.jpg              False

       ratings_disabled  video_error_or_removed  \
32                False                   False
70                False                   False
149               False                   False
298               False                   False
336               False                   False
...                 ...                     ...
40896             False                   False
40907             False                   False
40925             False                   False
40938             False                   False
40948             False                   False

                                            description
32     Eminem's new track Walk on Water ft. Beyoncé i...
70     🎧: https://ad.gt/yt-perfect\n💰: https://atlant...
149    Harry Styles' self-titled debut album is avail...
298    Eminem's new track Walk on Water ft. Beyoncé i...
336    🎧: https://ad.gt/yt-perfect\n💰: https://atlant...
...                                                  ...
40896  BTS Official Homepage http://bts.ibighit.comBT...
40907  Dinero Available at: Spotify: http://smarturl....
40925  Official Video | Diplo, French Montana & Lil P...
40938  Daddy Yankee - Hielo (Video Oficial)Spotify: h...
40948  Call of Duty: Black Ops 4 Multiplayer raises t...

[920 rows x 16 columns]
```