

```
In [19]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [20]: df = pd.read_csv(r"C:\Users\HP\Desktop\train project.csv")
```

```
In [21]: df.head()
```

Out[21]:

	Video_id	category_id	channel_title	subscriber	title	tags
0	HDR9SQc79	22	CaseyNeistat	9086142.0	WE WANT TO TALK ABOUT OUR MARRIAGE	SHANtell martin
1	KNH52UF?48	24	LastWeekTonight	5937292.0	The Trump Presidency: Last Week Tonight with J...	last week tonight trump presidency last week t...
2	QTW28IRG36	23	Rudy Mancuso	4191209.0	Racist Superman Rudy Mancuso, King Bach & Le...	superman rudy mancuso king bach racist ... racist
3	MGL76WlJ26	24	Good Mythical Morning	13186408.0	Nickelback Lyrics: Real or Fake?	rhett and link gmm good mythical morning rhett...
4	TWP93KXT70	24	nigahiga	20563106.0	I Dare You: GOING BALD!?	ryan higa higatv nigahiga i dare you idy rhpc ...



```
In [22]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3198 entries, 0 to 3197
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Video_id                             3198 non-null   object
1   category_id                           3192 non-null   object
2   channel_title                         3195 non-null   object
3   subscriber                            3175 non-null   float64
4   title                                3195 non-null   object
5   tags                                  3046 non-null   object
6   description                           3133 non-null   object
7   Trend_day_count                       3197 non-null   float64
8   Tag_count                             3197 non-null   object
9   Trend_tag_count                       3197 non-null   object
10  comment_count                         3198 non-null   object
11  comment_disabled                      3198 non-null   object
12  like dislike disabled                 3198 non-null   object
13  likes                                 3198 non-null   object
14  dislike                               3198 non-null   object
15  tag appered in title                 3197 non-null   object
16  views                                3198 non-null   object
17  Unnamed: 17                          1 non-null      float64
18  Unnamed: 18                          1 non-null      object
dtypes: float64(3), object(16)
memory usage: 474.8+ KB

```

```
In [23]: df.describe()
```

```

Out[23]:
```

	subscriber	Trend_day_count	Unnamed: 17
count	3.175000e+03	3197.000000	1.0
mean	3.823981e+06	7.964342	2544.0
std	2.865771e+07	78.556055	NaN
min	0.000000e+00	0.000000	2544.0
25%	2.428800e+05	4.000000	2544.0
50%	1.241220e+06	7.000000	2544.0
75%	3.812622e+06	10.000000	2544.0
max	1.576229e+09	4444.000000	2544.0

```
In [24]: df.columns
```

```

Out[24]: Index(['Video_id', 'category_id', 'channel_title', 'subscriber', 'title',
               'tags', 'description', 'Trend_day_count', 'Tag_count',
               'Trend_tag_count', 'comment_count', 'comment_disabled',
               'like dislike disabled', 'likes', 'dislike', 'tag appered in title',
               'views', 'Unnamed: 17', 'Unnamed: 18'],
              dtype='object')

```

```
In [25]: df.shape
```

```
Out[25]: (3198, 19)
```

```
In [26]: df = pd.DataFrame(df)
```

```
In [27]: df
```

Out[27]:

	Video_id	category_id	channel_title	subscriber	title			
0	HDR9SQc79	22	CaseyNeistat	9086142.0	WE WANT TO TALK ABOUT OUR MARRIAGE	SHANT		
1	KNH52UF748	24	LastWeekTonight	5937292.0	The Trump Presidency: Last Week Tonight with J...	last week tonight trump presidency las		
2	QTW28IRG36	23	Rudy Mancuso	4191209.0	Racist Superman Rudy Mancuso, King Bach & Le...	racist superman rudy mancuso king bac		
3	MGL76WIJ26	24	Good Mythical Morning	13186408.0	Nickelback Lyrics: Real or Fake?	rhett and link gmm gooc morni		
4	TWP93KXT70	24	nigahiga	20563106.0	I Dare You: GOING BALD!?	ryan higa higatv nigahiga i dare you i		
...		
3193	OKR48DOE67	20	EA SPORTS FIFA	3150213.0	FIFA 18 - FUT Champions Cup Barcelona - Day 1	fifa fifa ultimate team fut fifa pack c		
3194	QJK69DS791	10	JackWhiteVEVO	261596.0	Jack White - Corporation (Audio)	Alternative Corporation Jack White Thir		
3195	VHF51NVr11	10	JamesBlakeVEVO	28321.0	James Blake - If The Car Beside You Moves Ahead...	James Blake If The Car Beside You Move:		
3196	XHU22OAJ39	26	Refinery29	890739.0	Lucie Fink Trains Like A Professional Gymnast ...	refinery29 refinery 29 r29 r29 video r		
3197	IFD79NSG47	22	MN khan	0.0	Man drops magnum of Champagne on the floor in ...			

3198 rows × 19 columns



In [28]: `df.isnull().value_counts().sum()`

Out[28]: 3198

In [29]: `df.isnull().sum()`

```
Out[29]: Video_id          0
category_id        6
channel_title      3
subscriber        23
title             3
tags             152
description        65
Trend_day_count    1
Tag_count          1
Trend_tag_count    1
comment_count      0
comment_disabled   0
like dislike disabled 0
likes             0
dislike           0
tag appered in title 1
views             0
Unnamed: 17        3197
Unnamed: 18        3197
dtype: int64
```

```
In [30]: df.nunique()
```

```
Out[30]: Video_id          3198
category_id          17
channel_title       1501
subscriber         1452
title             3191
tags             2946
description        3097
Trend_day_count     14
Tag_count           22
Trend_tag_count     10
comment_count       2174
comment_disabled     3
like dislike disabled 3
likes             2210
dislike           2155
tag appered in title 2
views             3194
Unnamed: 17          1
Unnamed: 18          1
dtype: int64
```

```
In [31]: df.duplicated().sum()
```

```
Out[31]: 0
```

```
In [32]: df.drop(columns=['Unnamed: 17', 'Unnamed: 18'], inplace=True)
```

```
In [33]: df.dropna(subset=['views', 'likes', 'dislike', 'comment_count'], inplace=True)
```

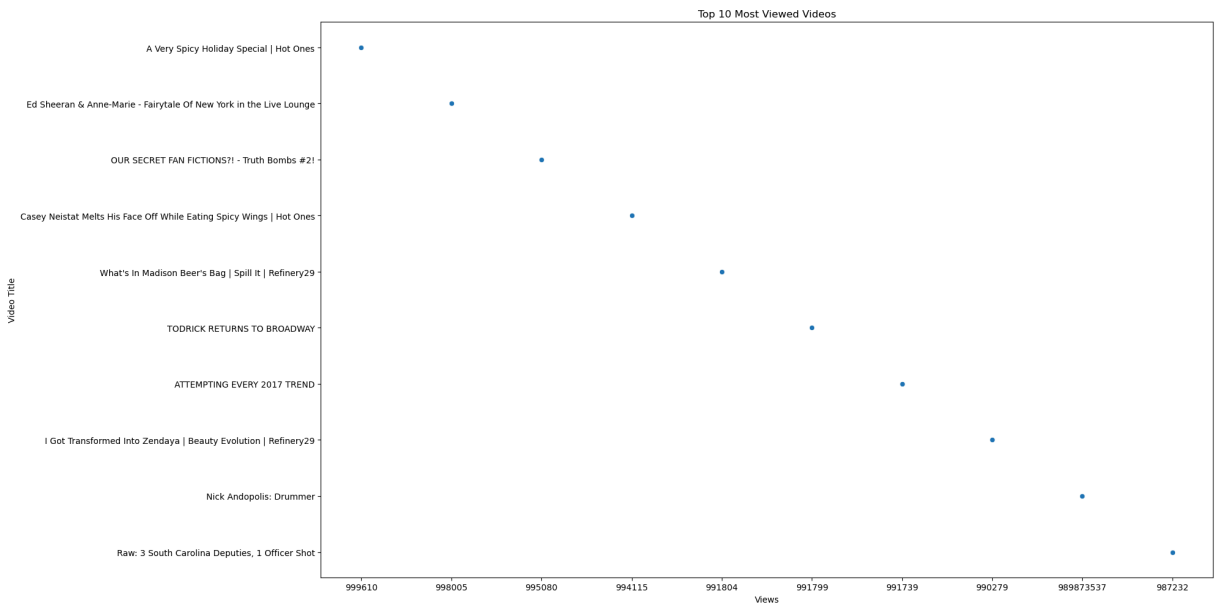
```
In [34]: df.head()
```

Out[34]:

	Video_id	category_id	channel_title	subscriber	title	tags
0	HDR9SQc79	22	CaseyNeistat	9086142.0	WE WANT TO TALK ABOUT OUR MARRIAGE	SHANtell martin
1	KNH52UF?48	24	LastWeekTonight	5937292.0	The Trump Presidency: Last Week Tonight with J...	last week tonight trump presidency last week t...
2	QTW28IRG36	23	Rudy Mancuso	4191209.0	Racist Superman Rudy Mancuso, King Bach & Le...	racist superman rudy mancuso king bach racist ...
3	MGL76WlJ26	24	Good Mythical Morning	13186408.0	Nickelback Lyrics: Real or Fake?	rhett and link gmm good mythical morning rhett...
4	TWP93KXT70	24	nigahiga	20563106.0	I Dare You: GOING BALD!?	ryan higa higatv nigahiga i dare you idy rhpc ...

```
In [40]: # plot top 10 most viewed videos
top_videos = df.sort_values('views', ascending=False).head(10)

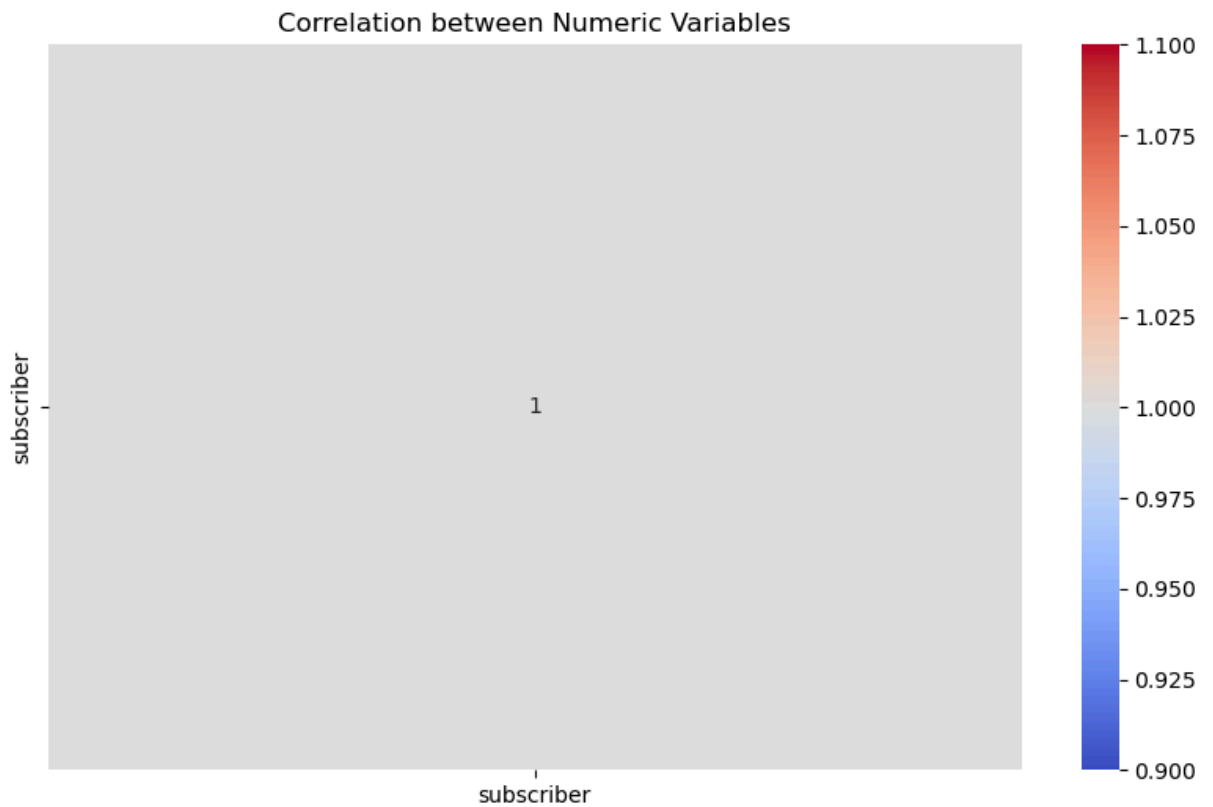
plt.figure(figsize=(20,10))
sns.scatterplot(x='views', y='title', data=top_videos)
plt.title('Top 10 Most Viewed Videos')
plt.xlabel('Views')
plt.ylabel('Video Title')
plt.tight_layout()
plt.show()
```



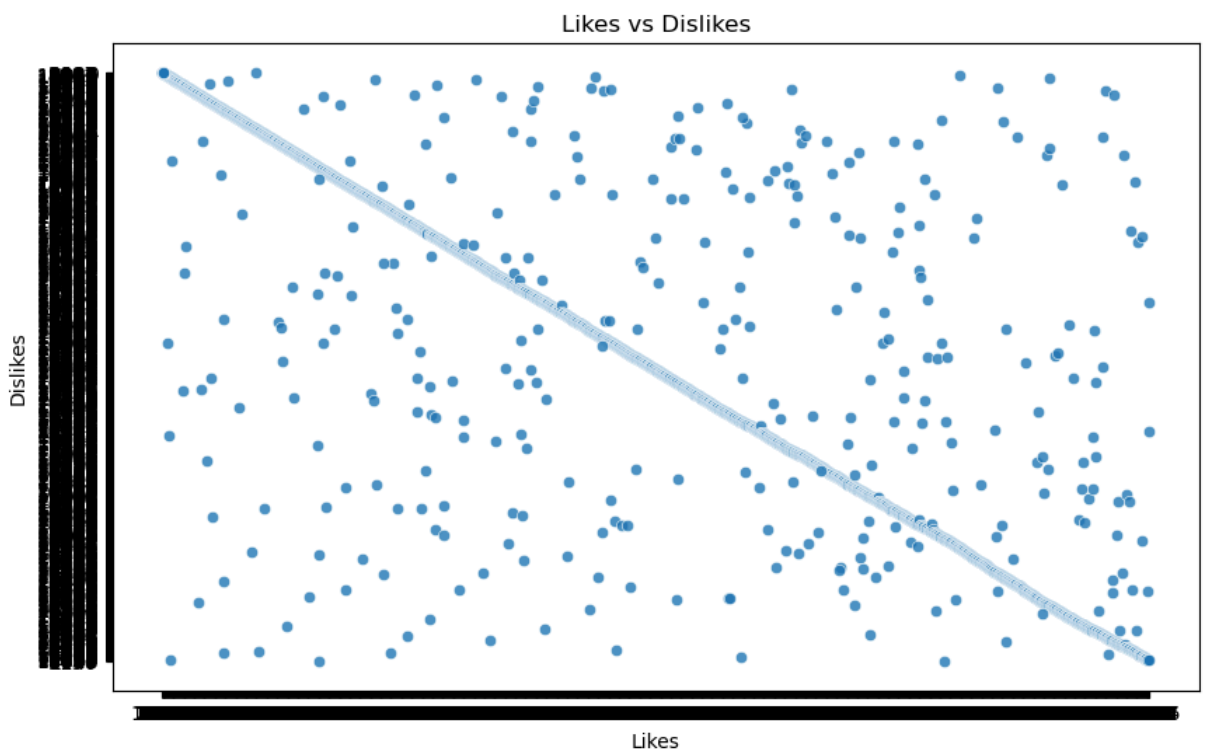
```
In [41]: # correlation of heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(df[['views', 'likes', 'dislike', 'comment_count', 'subscriber']].corr(), annot=True)
plt.title('Correlation between Numeric Variables')
plt.show()
```

C:\Users\HP\AppData\Local\Temp\ipykernel_27788\201440524.py:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
sns.heatmap(df[['views', 'likes', 'dislike', 'comment_count', 'subscriber']].corr(), annot=True, cmap='coolwarm')
```

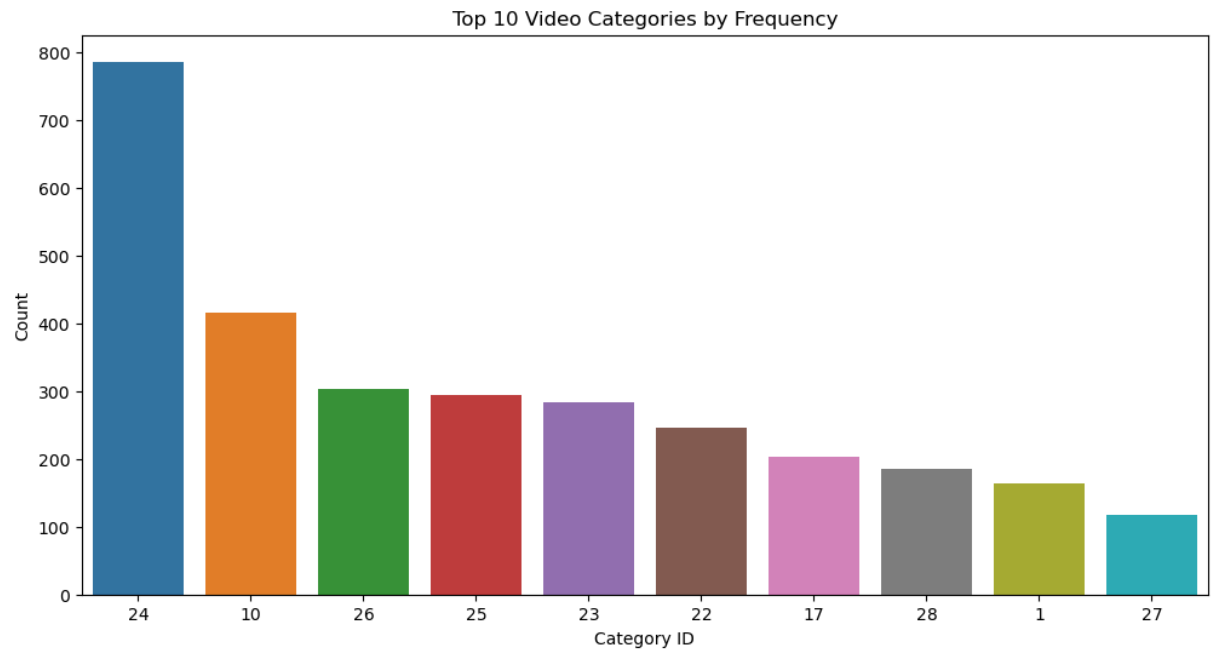


```
In [44]: # Like vs Dislike
plt.figure(figsize=(10,6))
sns.scatterplot(data=df, x='likes', y='dislike', alpha=0.8)
plt.title('Likes vs Dislikes')
plt.xlabel('Likes')
plt.ylabel('Dislikes')
plt.show()
```



In [45]: *# trend by Category*

```
plt.figure(figsize=(12,6))
top_categories = df['category_id'].value_counts().head(10)
sns.barplot(x=top_categories.index, y=top_categories.values)
plt.title('Top 10 Video Categories by Frequency')
plt.xlabel('Category ID')
plt.ylabel('Count')
plt.show()
```



Thank You !

In []: