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


data = pd.read_csv("/content/USvideos.csv")
data.shape
```

```
<ipython-input-29-ae7b1189f947>:1: DtypeWarning: Columns (12,13,14) have mixed types. Sp
  data = pd.read_csv("/content/USvideos.csv")
(34047, 16)
```

```
data = data.drop_duplicates()
data.describe()
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34047 entries, 0 to 34046
Data columns (total 16 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   video_id                34047 non-null  object
 1   trending_date           34047 non-null  object
 2   title                   34046 non-null  object
 3   channel_title           34046 non-null  object
 4   category_id             34046 non-null  float64
 5   publish_time            34046 non-null  object
 6   tags                    34046 non-null  object
 7   views                   34046 non-null  float64
 8   likes                   34046 non-null  float64
 9   dislikes                34046 non-null  float64
 10  comment_count           34046 non-null  float64
 11  thumbnail_link          34046 non-null  object
 12  comments_disabled       34046 non-null  object
 13  ratings_disabled        34046 non-null  object
 14  video_error_or_removed  34046 non-null  object
 15  description             33558 non-null  object
dtypes: float64(5), object(11)
memory usage: 4.2+ MB
```

```
columns = ['thumbnail_link','description']
data = data.drop(columns = columns)
```

```
nding_date'].apply(lambda x : datetime.datetime.strptime(x,'%y.%d.%m'))
```

```
data.head(5)
```

	video_id	trending_date	title	channel_title	category_id	publish_
0	2kyS6SvSYSE	2017-11-14	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	201 13T17:13:01.
1	1ZAPwfrtAFY	2017-11-14	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	201 13T07:30:00.
2	5qpjK5DgCt4	2017-11-14	Racist Superman | Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	201 12T19:05:24.
3	puqaWrEC7tY	2017-11-14	Nickelback Lyrics: Real or Fake?	Good Mythical Morning	24	201 13T11:00:04.
4	d380meD0W0M	2017-11-14	I Dare You: GOING BALD!?	nigahiga	24	201 12T18:01:41.

Next steps: Generate code with `data` View recommended plots

```
print(sorted(data['category_id'].unique()))
[1, 2, 10, 15, 17, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 43]
data['category_name'] = np.nan
data.loc[(data['category_id'] == 1), 'category_name'] = 'Film and Animat
data.loc[(data['category_id'] == 2), 'category_name'] = 'Autos and Vehic
data.loc[(data['category_id'] == 10), 'category_name'] = 'Music'
data.loc[(data["category_id"] == 15), "category_name"] = 'Pets and Anima
data.loc[(data ["category_id"] == 17 ), "category_name"] = 'Sports'
data.loc[(data["category_id"] == 19), "category_name"] =  'Travel and Ev
data.loc[(data["category_id"] == 20 ), "category_name"] = 'Gaming'
data.loc[(data["category_id"] == 22 ), "category_name"] = 'People and Bl
data.loc[(data["category_id"]== 23), "category_name"] = 'Comedy'
data.loc[(data["category_id"]== 24), "category_name"] = 'Entertainment'
data.loc[(data["category_id"] == 25), "category_name"] = 'News and Polit
data.loc[(data["category_id"] == 26), "category_name"] = 'How to and Sty
data.loc[(data["category_id"]== 27), "category_name"] =  'Education'
data.loc[(data["category_id"] == 28), "category_name"] = 'Science and Te
data.loc[(data["category_id"] == 29), "category_name"] = 'Non Profits ar
data.loc[(data["category_id"] == 30), "category_name"] = 'Movies'
data.loc[(data["category_id"] == 43), "category_name"] = 'Shows'
```

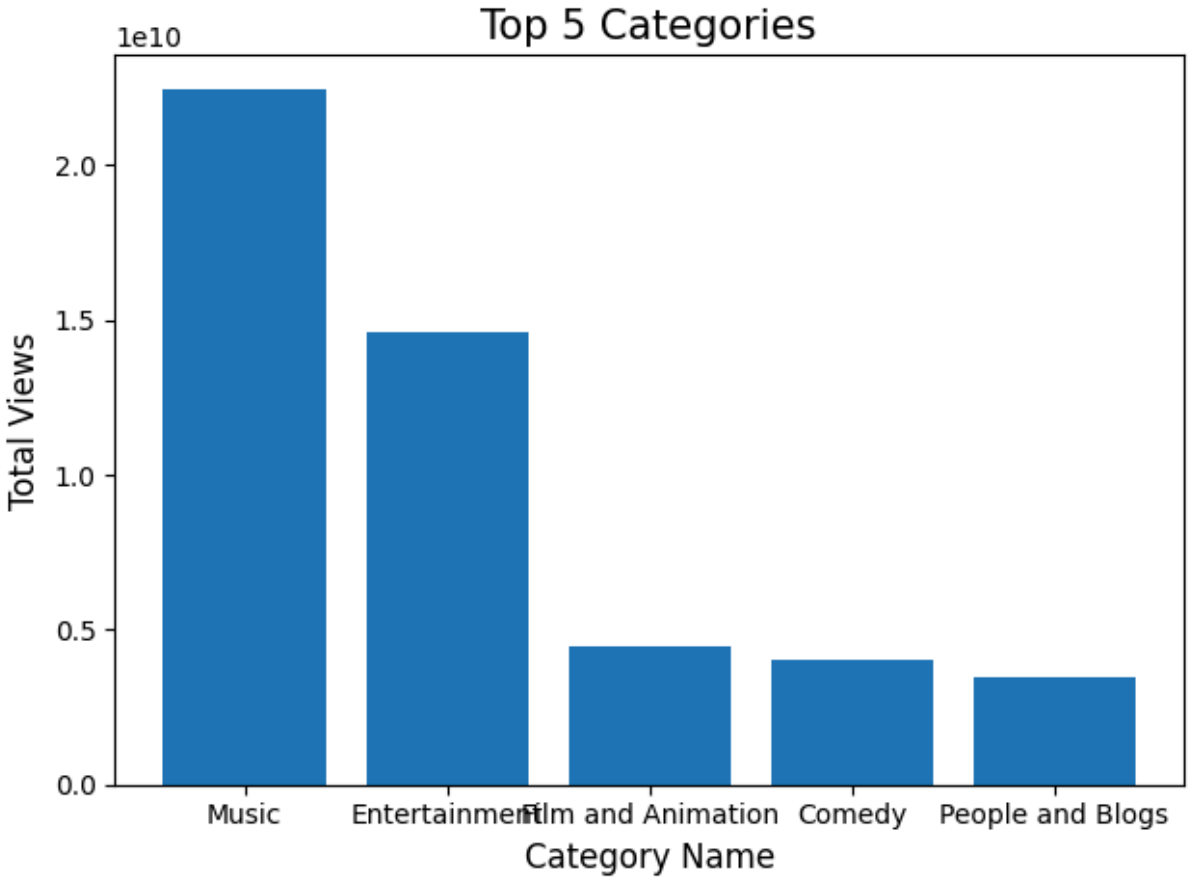
```
[1.0, 2.0, 10.0, 15.0, 17.0, 19.0, 20.0, 22.0, 23.0, 24.0, 25.0, 26.0, 27.0, 28.0, 29.0,
```

```
data.head()
```

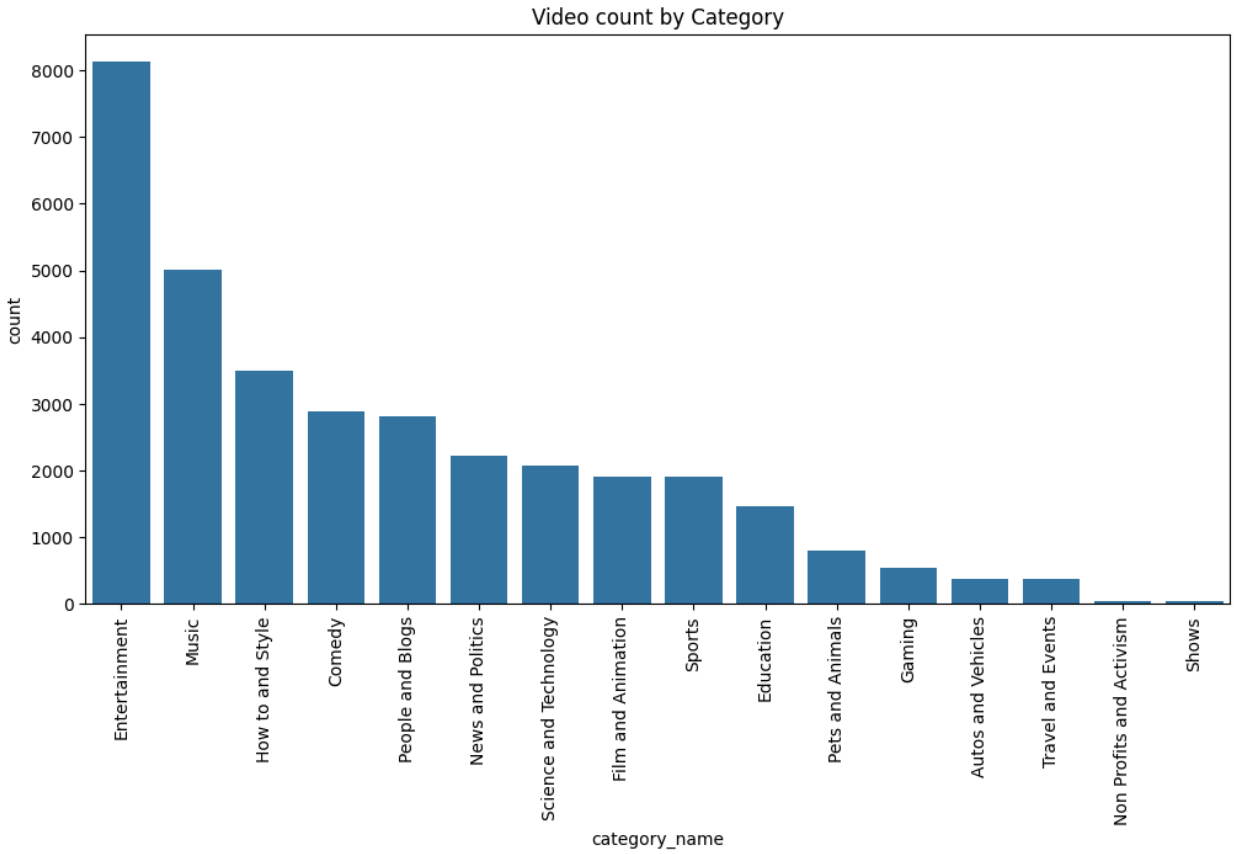
	video_id	trending_date	title	channel_title	category_id	publish_
0	2kyS6SvSYSE	2017-11-14	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22.0	201 13T17:13:01.
1	1ZAPwfrtAFY	2017-11-14	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24.0	201 13T07:30:00.
2	5qpjK5DgCt4	2017-11-14	Racist Superman | Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23.0	201 12T19:05:24.
3	puqaWrEC7tY	2017-11-14	Nickelback Lyrics: Real or Fake?	Good Mythical Morning	24.0	201 13T11:00:04.
4	d380meD0W0M	2017-11-14	I Dare You: GOING BALD!?	nigahiga	24.0	201 12T18:01:41.

Next steps: Generate code with `data` View recommended plots

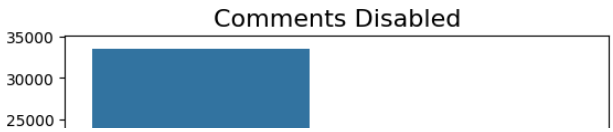
```
category_views = data.groupby('category_name')['views'].sum().reset_index
top_categories = category_views.sort_values(by='views', ascending = False
plt.bar(top_categories['category_name'], top_categories['views'])
plt.xlabel('Category Name', fontsize = 12)
plt.ylabel('Total Views', fontsize = 12)
plt.title('Top 5 Categories', fontsize = 15)
plt.tight_layout()
plt.show()
```

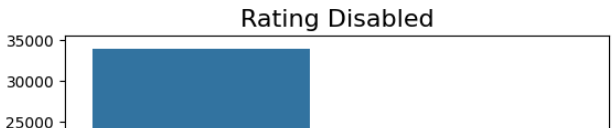



```
plt.figure(figsize = (12,6))
sns.countplot(x = 'category_name', data=data, order=data['category_name']
plt.xticks(rotation=90)
plt.title('Video count by Category')
plt.show()
```




```
plt.figure(figsize = (14,8))
plt.subplots_adjust(wspace = 0.2,hspace = 0.4, top = 0.9)
plt.subplot(2,2,1)
g = sns.countplot(x ='comments_disabled', data = data)
g.set_title("Comments Disabled",fontsize= 16)
plt.subplot(2,2,2)
g1 = sns.countplot(x = 'ratings_disabled', data = data)
g1.set_title("Rating Disabled",fontsize = 16)
plt.subplot(2,2,3)
g2 = sns.countplot(x = 'video_error_or_removed',data = data)
g2.set_title("Video Error or Removed",fontsize = 16)
plt.show()
```


Comments Disabled
35000
30000
25000


Rating Disabled
35000
30000
25000