In [7]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [8]:
```
pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\aktha\appdata\local\programs\py
thon\python312\lib\site-packages (2.2.3)
Requirement already satisfied: numpy>=1.26.0 in c:\users\aktha\appdata\local\prog
rams\python\python312\lib\site-packages (from pandas) (2.2.6)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\aktha\appdata\l
ocal\programs\python\python312\lib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\aktha\appdata\local\progr
ams\python\python312\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\aktha\appdata\local\pro
grams\python\python312\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\aktha\appdata\local\programs
\python\python312\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.
0)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 24.2 -> 25.1.1
[notice] To update, run: C:\Users\aktha\AppData\Local\Programs\Python\Python312\p
ython.exe -m pip install --upgrade pip
```

In [9]:
```
pip install matplotlib
```

```
Requirement already satisfied: matplotlib in c:\users\aktha\appdata\local\program
s\python\python312\lib\site-packages (3.10.3)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\aktha\appdata\local\p
rograms\python\python312\lib\site-packages (from matplotlib) (1.3.2)
Requirement already satisfied: cycler>=0.10 in c:\users\aktha\appdata\local\progr
ams\python\python312\lib\site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\aktha\appdata\local
\programs\python\python312\lib\site-packages (from matplotlib) (4.58.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\aktha\appdata\local
\programs\python\python312\lib\site-packages (from matplotlib) (1.4.8)
Requirement already satisfied: numpy>=1.23 in c:\users\aktha\appdata\local\progra
ms\python\python312\lib\site-packages (from matplotlib) (2.2.6)
Requirement already satisfied: packaging>=20.0 in c:\users\aktha\appdata\local\pr
ograms\python\python312\lib\site-packages (from matplotlib) (25.0)
Requirement already satisfied: pillow>=8 in c:\users\aktha\appdata\local\programs
\python\python312\lib\site-packages (from matplotlib) (11.2.1)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\aktha\appdata\local\p
rograms\python\python312\lib\site-packages (from matplotlib) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\aktha\appdata\loc
al\programs\python\python312\lib\site-packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in c:\users\aktha\appdata\local\programs
\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.1
7.0)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 24.2 -> 25.1.1
[notice] To update, run: C:\Users\aktha\AppData\Local\Programs\Python\Python312\p
ython.exe -m pip install --upgrade pip
```

In [10]:
```
pip install seaborn
```

```
Requirement already satisfied: seaborn in c:\users\aktha\appdata\local\programs\p
ython\python312\lib\site-packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\aktha\appdata\loc
al\programs\python\python312\lib\site-packages (from seaborn) (2.2.6)
Requirement already satisfied: pandas>=1.2 in c:\users\aktha\appdata\local\progra
ms\python\python312\lib\site-packages (from seaborn) (2.2.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\aktha\appdata
\local\programs\python\python312\lib\site-packages (from seaborn) (3.10.3)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\aktha\appdata\local\p
rograms\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seabor
n) (1.3.2)
Requirement already satisfied: cycler>=0.10 in c:\users\aktha\appdata\local\progr
ams\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn)
(0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\aktha\appdata\local
\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seabo
rn) (4.58.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\aktha\appdata\local
\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seabo
rn) (1.4.8)
Requirement already satisfied: packaging>=20.0 in c:\users\aktha\appdata\local\pr
ograms\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn)
(25.0)
Requirement already satisfied: pillow>=8 in c:\users\aktha\appdata\local\programs
\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (11.
2.1)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\aktha\appdata\local\p
rograms\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seabor
n) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\aktha\appdata\loc
al\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->sea
born) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\aktha\appdata\local\progr
ams\python\python312\lib\site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\aktha\appdata\local\pro
grams\python\python312\lib\site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\aktha\appdata\local\programs
\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.
1,>=3.4->seaborn) (1.17.0)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 24.2 -> 25.1.1
[notice] To update, run: C:\Users\aktha\AppData\Local\Programs\Python\Python312\p
ython.exe -m pip install --upgrade pip
```

In [13]:
```python
import pandas as pd

# Load US and India data
us = pd.read_csv("Downloads/archive (3)/INvideos.csv")
india = pd.read_csv("Downloads/archive (3)/USvideos.csv")

# Show first few rows
print("US Data:")
display(us.head())

print("India Data:")
display(india.head())
```

```
US Data:
```

	video_id	trending_date	title	channel_title	category_id	publish_time
0	kzwfHumJyYc	17.14.11	Sharry Mann: Cute Munda (Song Teaser) | Parmi...	Lokdhun Punjabi	1	2017-11-12T12:20:39.000Z
1	zUZ1z7FwLc8	17.14.11	पीरियड्स के समय, पेट पर पति करता ऐसा, देखकर दं...	HJ NEWS	25	2017-11-13T05:43:56.000Z
2	10L1hZ9qa58	17.14.11	Stylish Star Allu Arjun @ ChaySam Wedding Rece...	TFPC	24	2017-11-12T15:48:08.000Z
3	N1vE8iiEg64	17.14.11	Eruma Saani | Tamil vs English	Eruma Saani	23	2017-11-12T07:08:48.000Z
4	kJzGH0PVQHQ	17.14.11	why Samantha became EMOTIONAL @ Samantha naga ...	Filmylooks	24	2017-11-13T01:14:16.000Z

```
India Data:
```

	video_id	trending_date	title	channel_title	category_id	publish_tir
0	2kyS6SvSYSE	17.14.11	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	2017-1 13T17:13:01.00
1	1ZAPwfrtAFY	17.14.11	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	2017-1 13T07:30:00.00
2	5qpjK5DgCt4	17.14.11	Racist Superman | Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	2017-1 12T19:05:24.00
3	puqaWrEC7tY	17.14.11	Nickelback Lyrics: Real or Fake?	Good Mythical Morning	24	2017-1 13T11:00:04.00
4	d380meD0W0M	17.14.11	I Dare You: GOING BALD!?	nigahiga	24	2017-1 12T18:01:41.00

In [14]:
```
us["country"] = "US"
india["country"] = "India"
```

In [15]:
```
combined_df = pd.concat([us, india], ignore_index=True)
```

In [16]:
```
combined_df.shape  # To see rows and columns
combined_df.head()  # View first few rows
```

Out[16]:

video_id	trending_date	title	channel_title	category_id	publish_time	
HumJyYc	17.14.11	Sharry Mann: Cute Munda (Song Teaser) | Parmi...	Lokdhun Punjabi	1	2017-11-12T12:20:39.000Z	sharry song
z7FwLc8	17.14.11	पीरियड्स के समय, पेट पर पति करता ऐसा, देखकर दं...	HJ NEWS	25	2017-11-13T05:43:56.000Z	पीरिय ऐ
hZ9qa58	17.14.11	Stylish Star Allu Arjun @ ChaySam Wedding Rece...	TFPC	24	2017-11-12T15:48:08.000Z	Stylish @ Chay
E8iiEg64	17.14.11	Eruma Saani | Tamil vs English	Eruma Saani	23	2017-11-12T07:08:48.000Z	Erun Videos"
)PVQHQ	17.14.11	why Samantha became EMOTIONAL @ Samantha naga ...	Filmylooks	24	2017-11-13T01:14:16.000Z	Filr

In [17]:
```
combined_df.isnull().sum()
```

Out[17]:
```
video_id                   0
trending_date              0
title                      0
channel_title              0
category_id                0
publish_time               0
tags                       0
views                      0
likes                      0
dislikes                   0
comment_count              0
thumbnail_link             0
comments_disabled          0
ratings_disabled           0
video_error_or_removed     0
description             1131
country                    0
dtype: int64
```

In [18]:
```
# Drop rows where description is missing (optional)
combined_df = combined_df.dropna(subset=['description'])
```

```
# Remove duplicates
combined_df = combined_df.drop_duplicates()

# Confirm shape again
print("Shape after cleaning:", combined_df.shape)
```

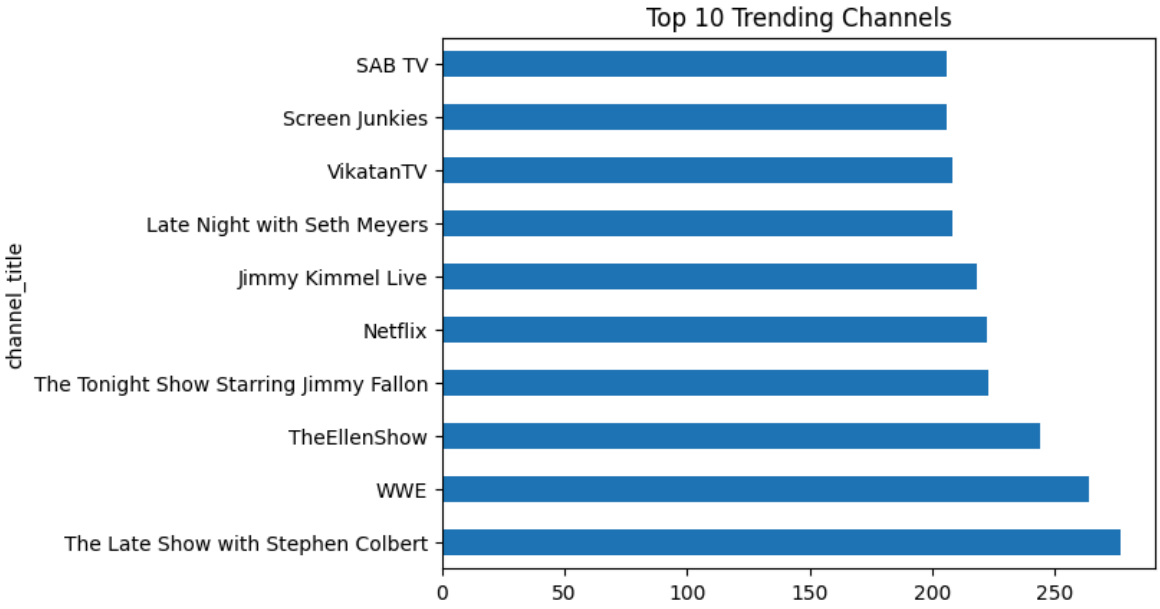
```
Shape after cleaning: (72894, 17)
```

EDA

1. TOP TRENDING CHANNELS

In [19]:
```
top_channels = combined_df['channel_title'].value_counts().head(10)
top_channels.plot(kind='barh', title='Top 10 Trending Channels')
```

Out[19]:
```
<Axes: title={'center': 'Top 10 Trending Channels'}, ylabel='channel_title'>
```



2. MOST VIEWED VIDEOS

In [20]:
```
top_videos = combined_df.sort_values(by='views', ascending=False)[['title', 'cha
top_videos
```

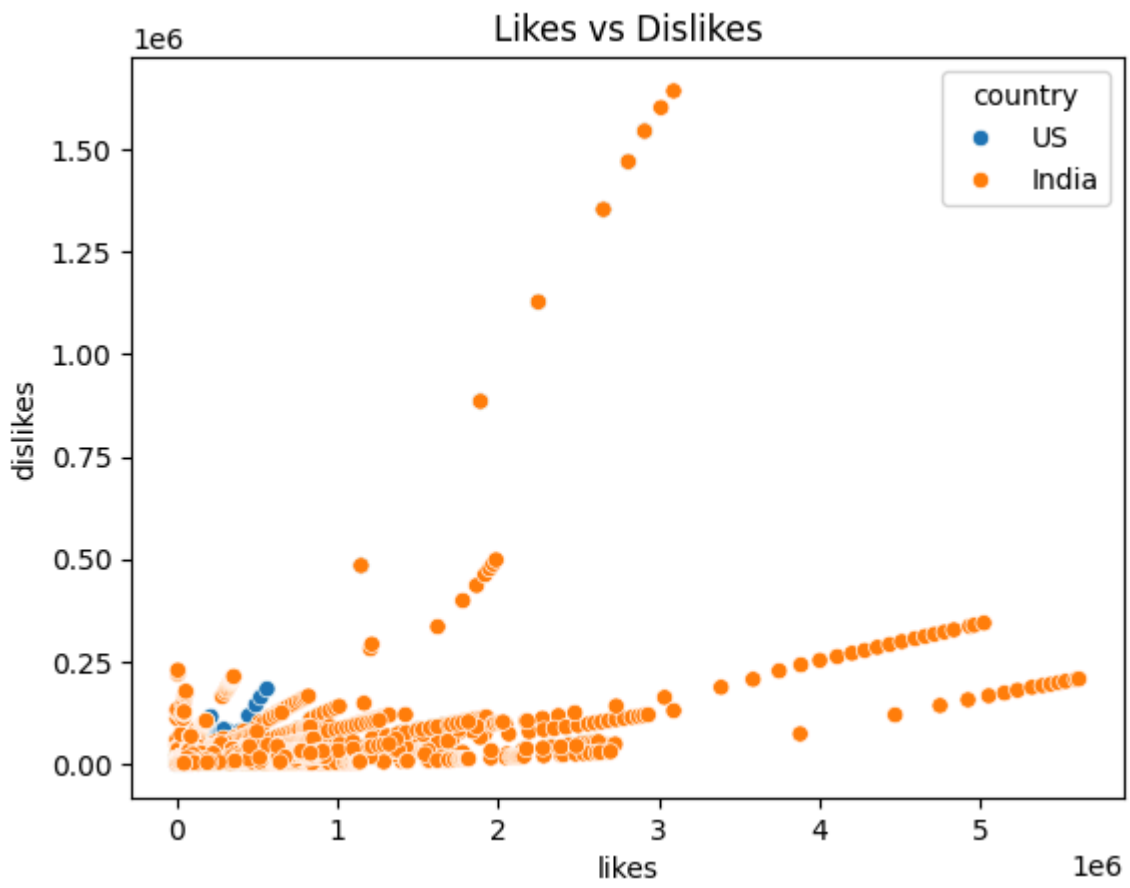
Out[20]:

	title	channel_title	views
75899	Childish Gambino - This Is America (Official V...	ChildishGambinoVEVO	225211923
75697	Childish Gambino - This Is America (Official V...	ChildishGambinoVEVO	220490543
75498	Childish Gambino - This Is America (Official V...	ChildishGambinoVEVO	217750076
75287	Childish Gambino - This Is America (Official V...	ChildishGambinoVEVO	210338856
75082	Childish Gambino - This Is America (Official V...	ChildishGambinoVEVO	205643016
74883	Childish Gambino - This Is America (Official V...	ChildishGambinoVEVO	200820941
74685	Childish Gambino - This Is America (Official V...	ChildishGambinoVEVO	196222618
74475	Childish Gambino - This Is America (Official V...	ChildishGambinoVEVO	190950401
74265	Childish Gambino - This Is America (Official V...	ChildishGambinoVEVO	184446490
74062	Childish Gambino - This Is America (Official V...	ChildishGambinoVEVO	179045286

3.LIKES VS DISLIKES SCATTER PLOT

In [21]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.scatterplot(data=combined_df, x='likes', y='dislikes', hue='country')
plt.title("Likes vs Dislikes")
plt.show()
```



4. CATEGORY-WISE TRENDING

In [23]:
```
import json

# Load category_id.json
with open("Downloads/archive (3)/IN_category_id.json", "r") as file:
    category_data = json.load(file)

# Extract category_id and title
category_dict = {}
for item in category_data['items']:
    category_dict[int(item['id'])] = item['snippet']['title']

# Display the dictionary (optional)
print(category_dict)
```

```
{1: 'Film & Animation', 2: 'Autos & Vehicles', 10: 'Music', 15: 'Pets & Animals',
17: 'Sports', 18: 'Short Movies', 19: 'Travel & Events', 20: 'Gaming', 21: 'Video
blogging', 22: 'People & Blogs', 23: 'Comedy', 24: 'Entertainment', 25: 'News & P
olitics', 26: 'Howto & Style', 27: 'Education', 28: 'Science & Technology', 30:
'Movies', 31: 'Anime/Animation', 32: 'Action/Adventure', 33: 'Classics', 34: 'Com
edy', 35: 'Documentary', 36: 'Drama', 37: 'Family', 38: 'Foreign', 39: 'Horror',
40: 'Sci-Fi/Fantasy', 41: 'Thriller', 42: 'Shorts', 43: 'Shows', 44: 'Trailers'}
```

In [24]:
```
# Map category_id to category name
combined_df['category_name'] = combined_df['category_id'].map(category_dict)

# See updated data
combined_df[['category_id', 'category_name']].head()
```

Out[24]:

	category_id	category_name
0	1	Film & Animation
1	25	News & Politics
2	24	Entertainment
3	23	Comedy
4	24	Entertainment

A.MOST TRENDING CATEGORIES

In [26]:
```
combined_df['category_name'].value_counts().head(10)
```

Out[26]:
```
category_name
Entertainment           24329
Music                    9723
News & Politics          7053
Comedy                   6380
People & Blogs           5298
Howto & Style            4930
Film & Animation         3793
Science & Technology     2871
Education                2761
Sports                   2754
Name: count, dtype: int64
```

B. AVERAGE VIEWS BY CATEGORY

In [27]:
```
combined_df.groupby('category_name')['views'].mean().sort_values(ascending=False
```

Out[27]:
```
category_name
Music                   4.968176e+06
Movies                  3.191953e+06
Film & Animation        2.777772e+06
Gaming                  2.696406e+06
Sports                  2.025643e+06
Entertainment           1.378427e+06
Science & Technology    1.282977e+06
Autos & Vehicles        1.228690e+06
Comedy                  1.177229e+06
People & Blogs          1.093853e+06
Name: views, dtype: float64
```

C. COUNTRY-WISE CATEGORY

In [28]:
```
combined_df.groupby(['country', 'category_name'])['views'].mean().sort_values(as
```

Out[28]:
```
country  category_name
India    Music                6.214107e+06
US       Gaming               3.606369e+06
         Movies               3.191953e+06
India    Film & Animation     3.107903e+06
         Gaming               2.634002e+06
US       Music                2.533100e+06
         Film & Animation     2.247295e+06
India    Entertainment        2.072942e+06
         Sports               2.070765e+06
US       Sports               1.873829e+06
Name: views, dtype: float64
```

5. TOP CHANNELS (PER COUNTRY)

In [29]:
```
# Top 10 channels by total views per country
top_channels = combined_df.groupby(['country', 'channel_title'])['views'].sum().

# Sort by views
top_channels = top_channels.sort_values(['country', 'views'], ascending=[True, F
```

```
# Get top 10 for each country
top_10_channels = top_channels.groupby('country').head(10)

# Display
top_10_channels
```

Out[29]:

	country	channel_title	views
345	India	ChildishGambinoVEVO	3758488765
2034	India	ibighit	2235906679
526	India	Dude Perfect	1870085178
1096	India	Marvel Entertainment	1808998971
107	India	ArianaGrandeVevo	1576959172
1076	India	MalumaVEVO	1551515831
2052	India	jypentertainment	1486972132
1568	India	Sony Pictures Entertainment	1432374398
633	India	FoxStarHindi	1238609854
173	India	BeckyGVEVO	1182971286
3166	US	T-Series	1748057724
2756	US	Marvel Entertainment	1058174340
2488	US	FoxStarHindi	982953616
2217	US	Amit Bhadana	855533181
3133	US	Speed Records	648890913
2524	US	Goldmines Telefilms	592596907
3128	US	Sony Pictures Entertainment	587519233
3461	US	Zee Music Company	556309816
3441	US	YRF	549492884
2422	US	Dude Perfect	527111130

6. TRENDING DURATION

In [30]:

```
# Convert trending_date to datetime (if not already)
combined_df['trending_date'] = pd.to_datetime(combined_df['trending_date'], form

# Count how many times (days) each video_id appears
trending_days = combined_df.groupby(['video_id', 'title', 'country']).size().res

# Sort by trending days
top_trending_videos = trending_days.sort_values(['country', 'trending_days'], as

# Top 10 most persistent videos per country
```

```
top_persistent_videos = top_trending_videos.groupby('country').head(10)

# Display
top_persistent_videos
```

Out[30]:

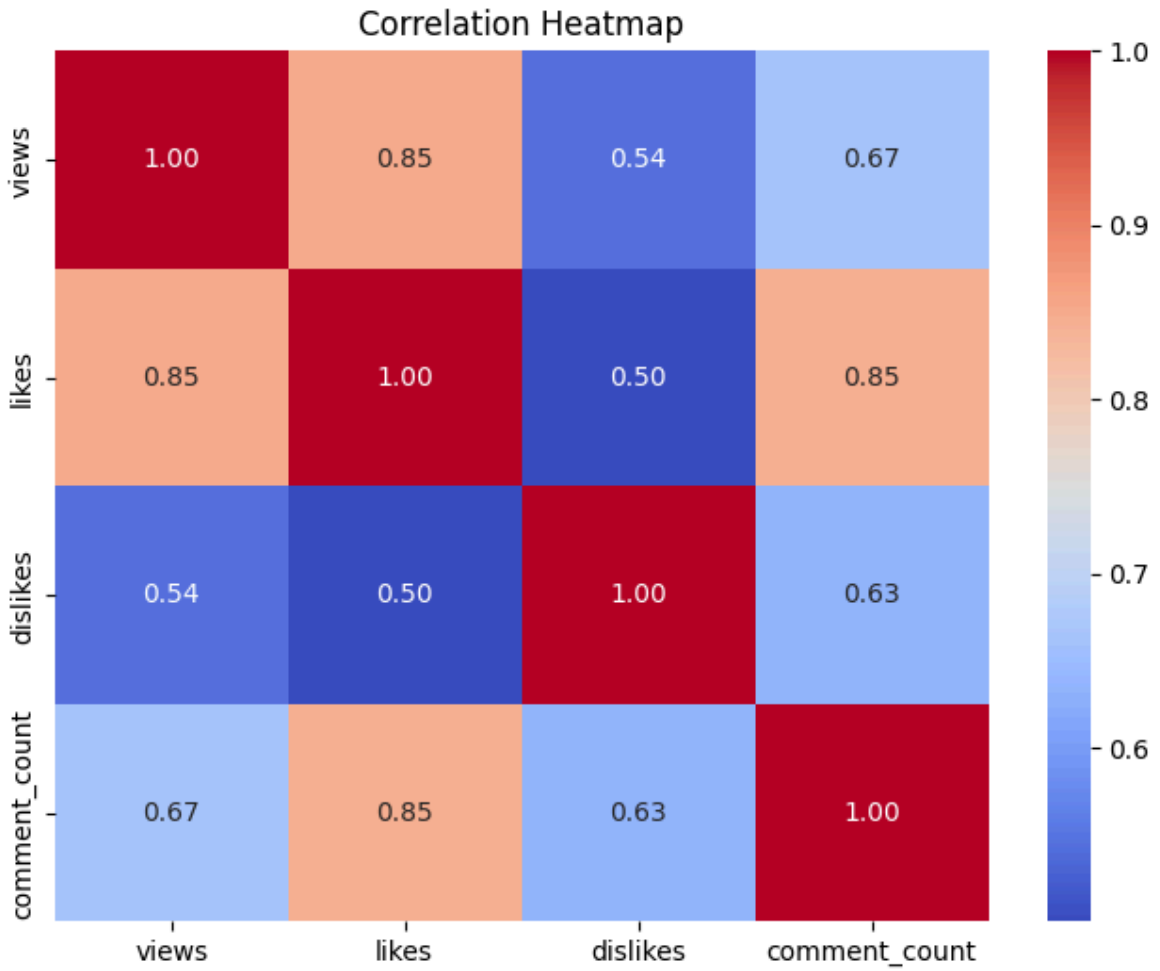
	video_id	title	country	trending_days
3529	8h--kFui1JA	Sam Smith - Pray (Official Video) ft. Logic	India	29
16879	j4KvrAUjn6c	WE MADE OUR MOM CRY...HER DREAM CAME TRUE!	India	29
2701	6S9c5nnDd_s	Bohemian Rhapsody | Teaser Trailer [HD] | 20th...	India	28
8687	NBSAQenU2Bk	Rooster Teeth Animated Adventures - Millie So ...	India	28
9801	QBL8IRJ5yHU	Why I'm So Scared (being myself and crying too...	India	28
11983	WIV3xNz8NoM	Cobra Kai Season 2	India	28
12657	YI3tsmFsrOg	The Deadliest Being on Planet Earth – The Bact...	India	28
16581	iILJvqrAQ_w	Charlie Puth - BOY [Official Audio]	India	28
19731	r-3iathMo7o	The ULTIMATE $30,000 Gaming PC Setup	India	28
20501	t4pRQ0jn23Q	YoungBoy Never Broke Again Goes Sneaker Shoppi...	India	28
19915	rRr1qiJRsXk	Sanju | Official Teaser | Ranbir Kapoor | Rajk...	US	13
5728	EyPXz6hKa_s	School Ke Wo Din - Amit Bhadana	US	11
825	1J76wN0TPI4	Sanju | Official Trailer | Ranbir Kapoor | Raj...	US	10
12155	Wm_vSSlVsV4	Kaala (Tamil) - Official Teaser | Rajinikanth ...	US	10
13727	aNwWdF8qq-M	Official Video: Raat Kamaal Hai | Guru Randhaw...	US	10
2093	4juJXyLX510	Avengers Infinity War with Ashish Chanchlani	US	9
5802	FCUPcNBpq4E	Kinjal Dave - Moj Ma (Ghate To Zindagi Ghate ...	US	9
5866	FPm7xM849-E	BB Ki Vines- | Maun Vrat |	US	9
7166	J-dv_DcDD_A	ZAYN - Let Me (Official Video)	US	9
14168	bYSRPuDEnTg	Garmi Ke Side-Effects | Ashish Chanchlani	US	9

CORRELATION HEATMAP

In [31]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Select only numerical columns
numeric_cols = combined_df[['views', 'likes', 'dislikes', 'comment_count']]

# Plot heatmap
plt.figure(figsize=(8, 6))
sns.heatmap(numeric_cols.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```



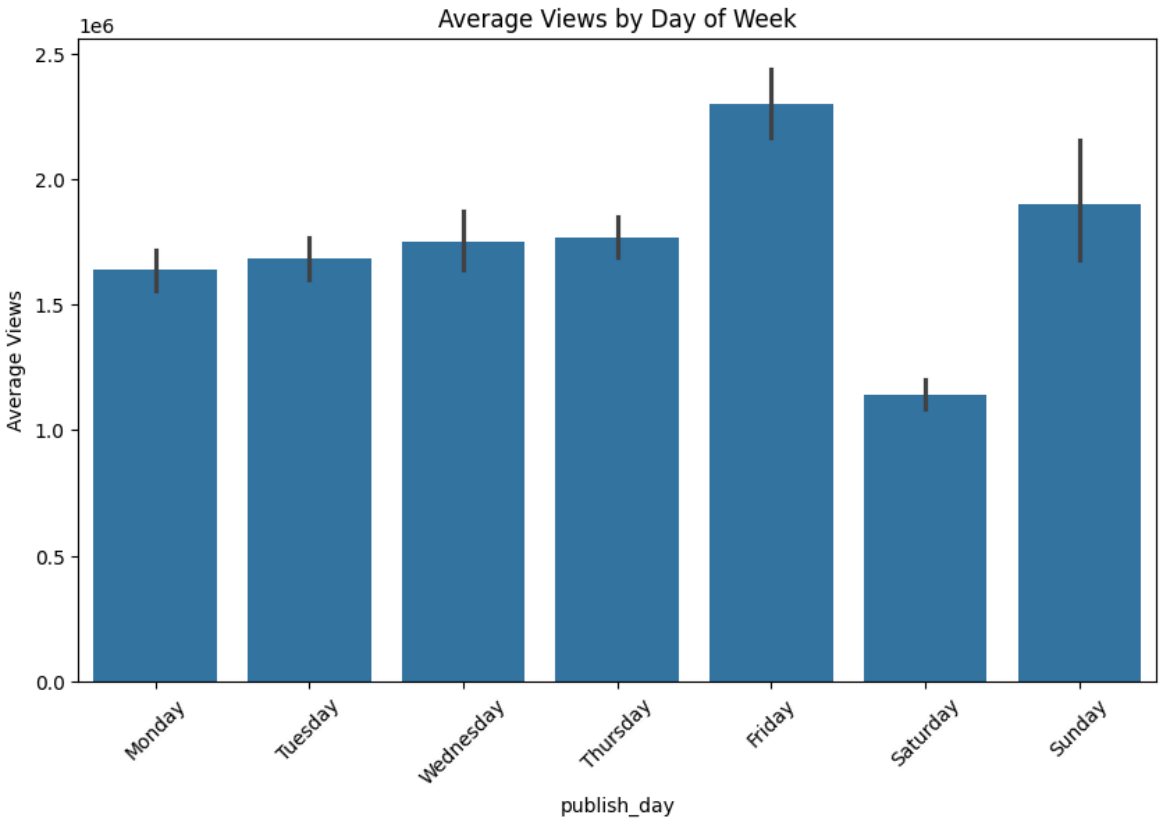
TIME-BASED ANALYSIS

In [33]:
```
combined_df['publish_time'] = pd.to_datetime(combined_df['publish_time'])
```

In [34]:
```
combined_df['publish_day'] = combined_df['publish_time'].dt.day_name()
combined_df['publish_hour'] = combined_df['publish_time'].dt.hour
```

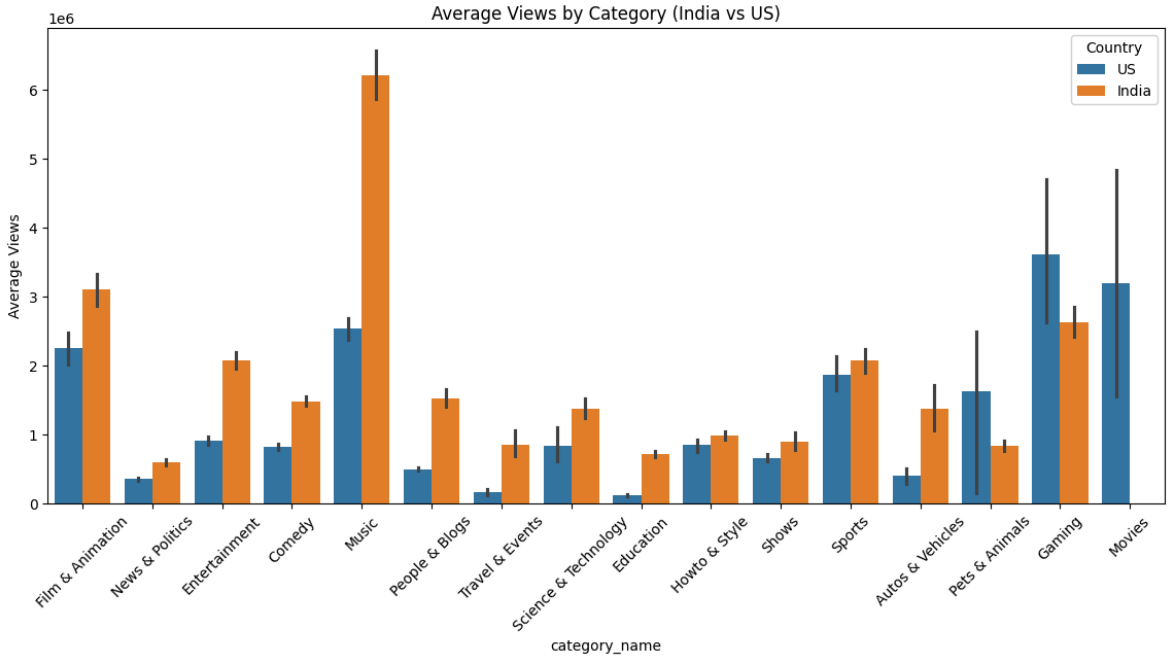
In [35]:
```
plt.figure(figsize=(10,6))
sns.barplot(x='publish_day', y='views', data=combined_df, estimator='mean', orde
    'Monday','Tuesday','Wednesday','Thursday','Friday','Saturday','Sunday'])
plt.title('Average Views by Day of Week')
```

```
plt.ylabel('Average Views')
plt.xticks(rotation=45)
plt.show()
```



BAR CHART: VIEWS BY CATEGORY (INDIA VS US)

In [36]:
```
plt.figure(figsize=(14, 6))
sns.barplot(data=combined_df, x='category_name', y='views', hue='country', estim
plt.title('Average Views by Category (India vs US)')
plt.ylabel('Average Views')
plt.xticks(rotation=45)
plt.legend(title='Country')
plt.show()
```



In [44]:
```
# Save your cleaned dataframe to a CSV file
combined_df.to_csv("youtube_cleaned_data.csv", index=False)
```

In [ ]:
```

```