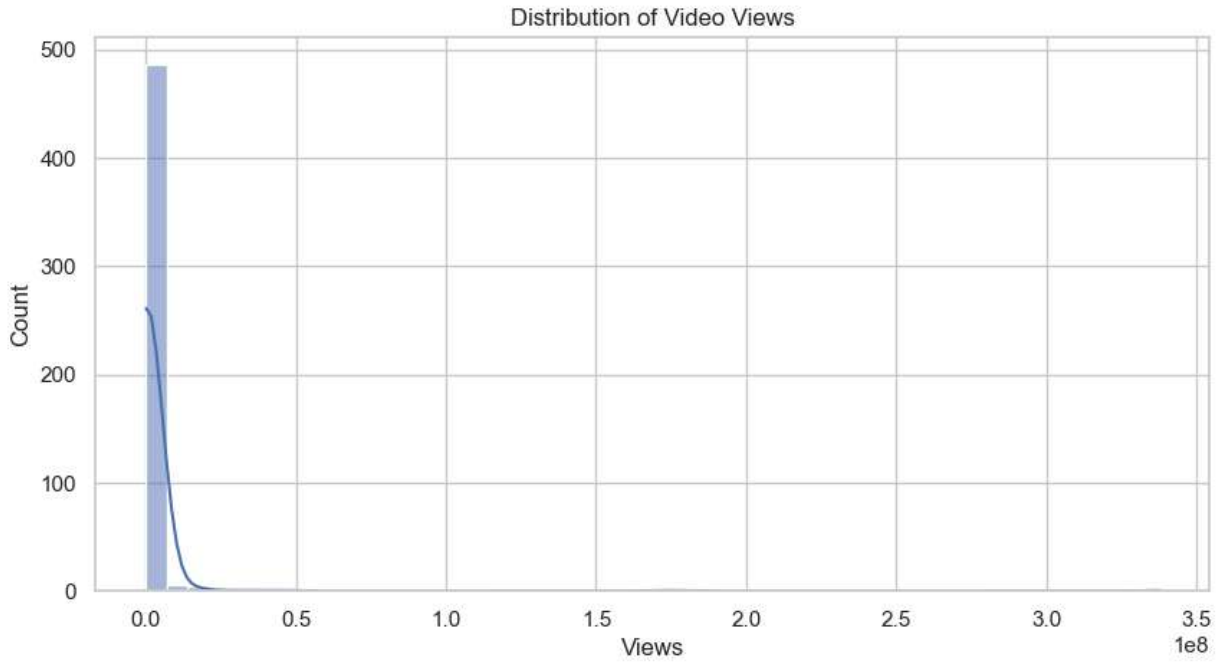
In [4]:
```
# 📌 Step 1: Import libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Set styles
sns.set(style="whitegrid")
plt.rcParams['figure.figsize'] = (10, 6)

# 📌 Step 2: Load dataset
df = pd.read_csv("C:/Users/User/Desktop/ELEVATE/small_cleaned_youtube_trending_pc

# Preview
df.head()
```

Out[4]:

	video_id	trending_date	title	channel_title	category_id	publish_time	
0	gS1DbvHHVH0	2018-06-06	Going in to brain surgery	Simone Giertz	28.0	2018-05-30 14:22:13+00:00	
1	n1MkZA7yG2U	2017-12-30	La Chute d'une IcÃ´ne 10 FIN Appo Firenze Esob...	esepelisa	24.0	2017-12-29 08:00:04+00:00	esepelisa'
2	m7IFyINZIBs	2018-04-19	Cardi B Confirms Pregnancy	Wendy Williams	24.0	2018-04-10 01:00:02+00:00	
3	8uP-fpQJT3s	2018-05-04	4 May 2018 - The Hindu Editorial News Paper An...	Study IQ education	27.0	2018-05-04 03:02:22+00:00	
4	os5mMTI3IBs	2017-12-03	VLOGMAS DAY 1: We're in Rehearsals!	Giovannasworld	24.0	2017-12-01 09:00:03+00:00	vlogmas|"2

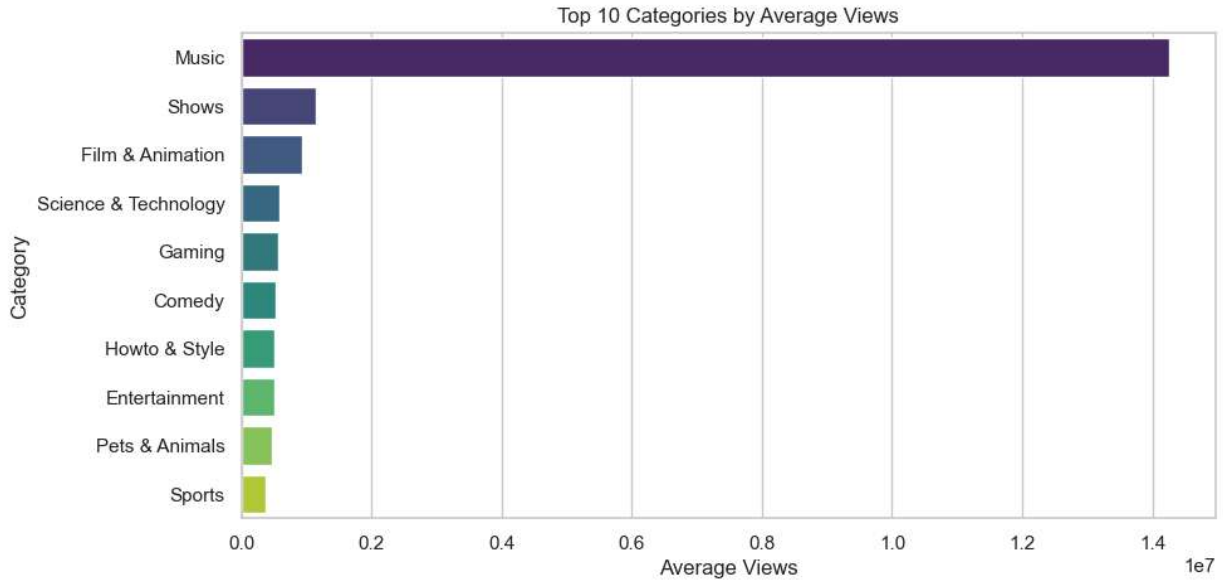
In [ ]:
```
# 1. Distribution of Views
```

In [5]:
```
plt.figure(figsize=(10,5))
sns.histplot(df['views'], bins=50, kde=True)
plt.title("Distribution of Video Views")
plt.xlabel("Views")
plt.ylabel("Count")
plt.show()
```



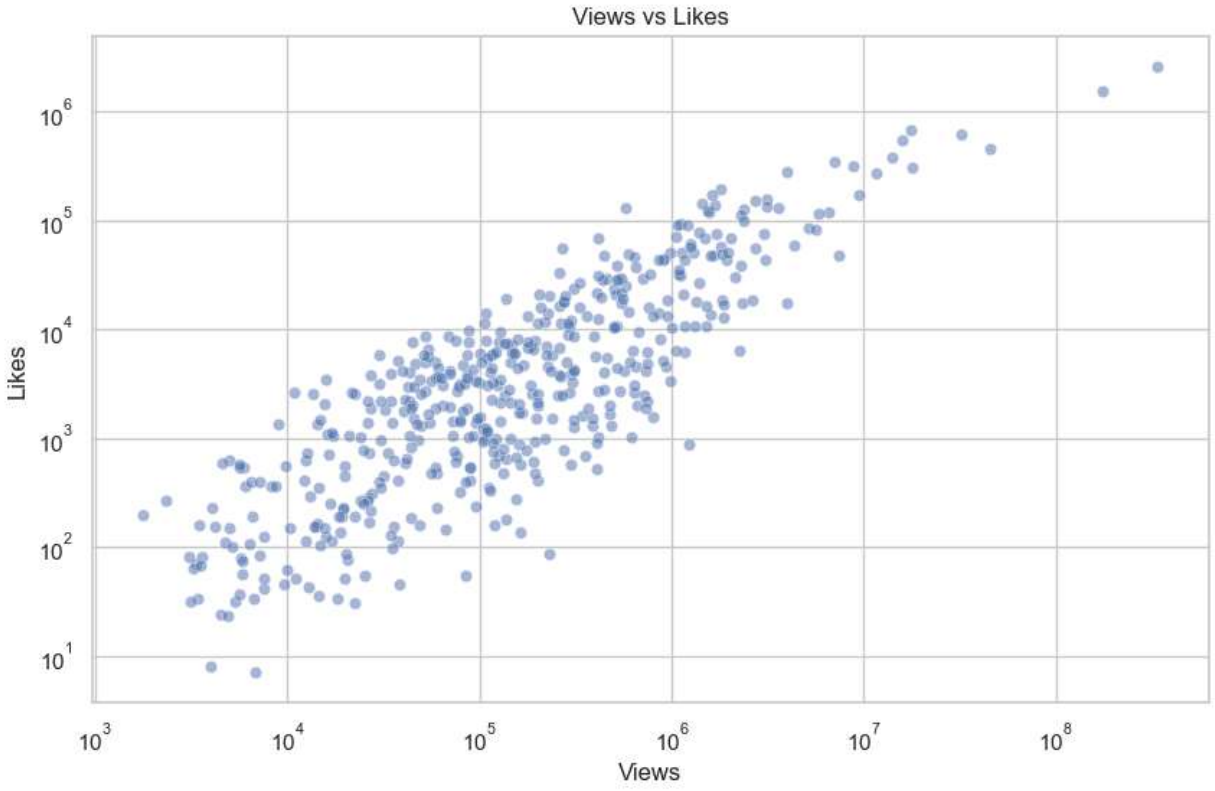
In [6]:
```
# 2. Top 10 Categories by Average Views
```

In [7]:
```
top_categories = df.groupby("category_name")['views'].mean().sort_values(ascendir

plt.figure(figsize=(10,5))
sns.barplot(x=top_categories.values, y=top_categories.index, palette="viridis")
plt.title("Top 10 Categories by Average Views")
plt.xlabel("Average Views")
plt.ylabel("Category")
plt.show()
```



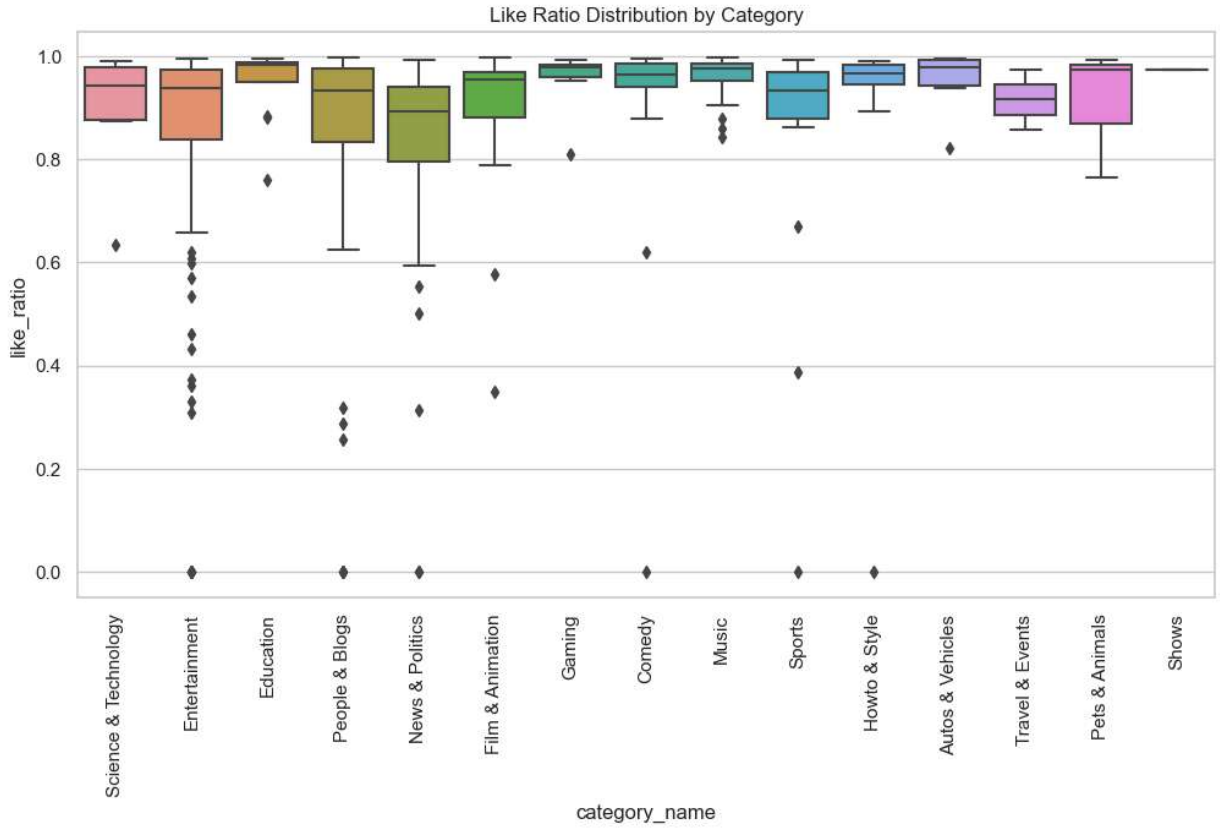
In [8]:
```
# 3. Engagement (Likes vs Views)
```

In [9]:
```
plt.figure(figsize=(10,6))
sns.scatterplot(data=df, x="views", y="likes", alpha=0.5)
plt.title("Views vs Likes")
plt.xlabel("Views")
plt.ylabel("Likes")
plt.xscale("log")  # log scale for better visualization
plt.yscale("log")
plt.show()
```



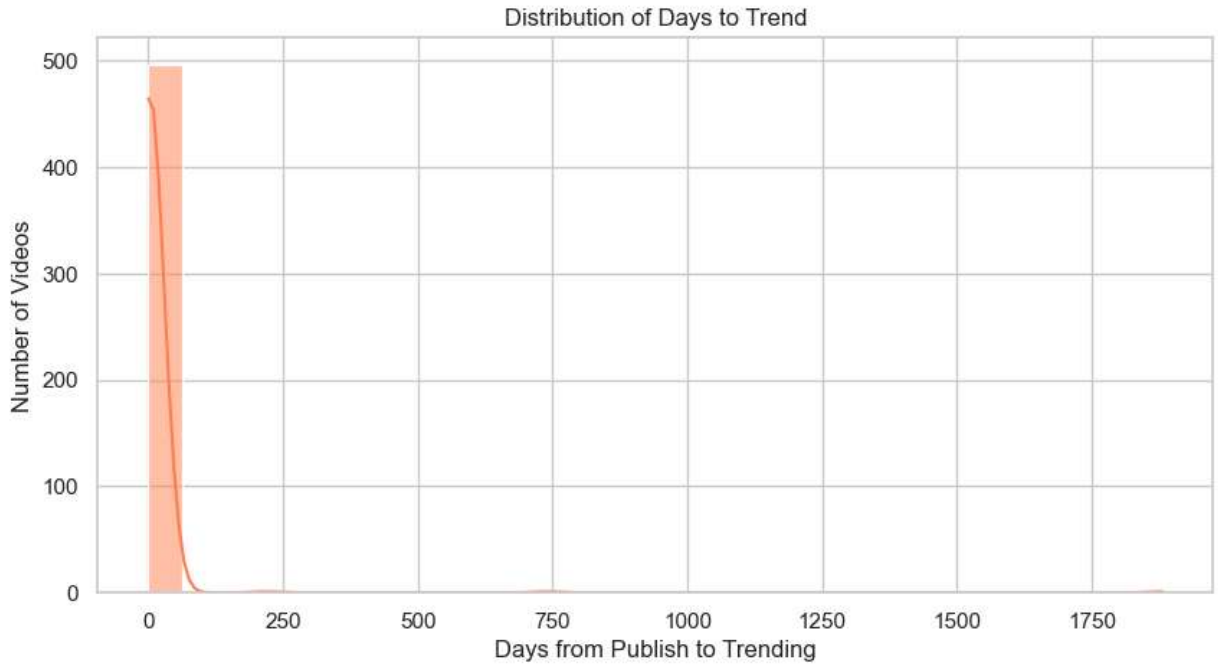
In [10]:
```
# 4. Like-Dislike Ratio by Category
```

In [11]:
```
df['like_ratio'] = df['likes'] / (df['likes'] + df['dislikes'] + 1)

plt.figure(figsize=(12,6))
sns.boxplot(x="category_name", y="like_ratio", data=df)
plt.title("Like Ratio Distribution by Category")
plt.xticks(rotation=90)
plt.show()
```



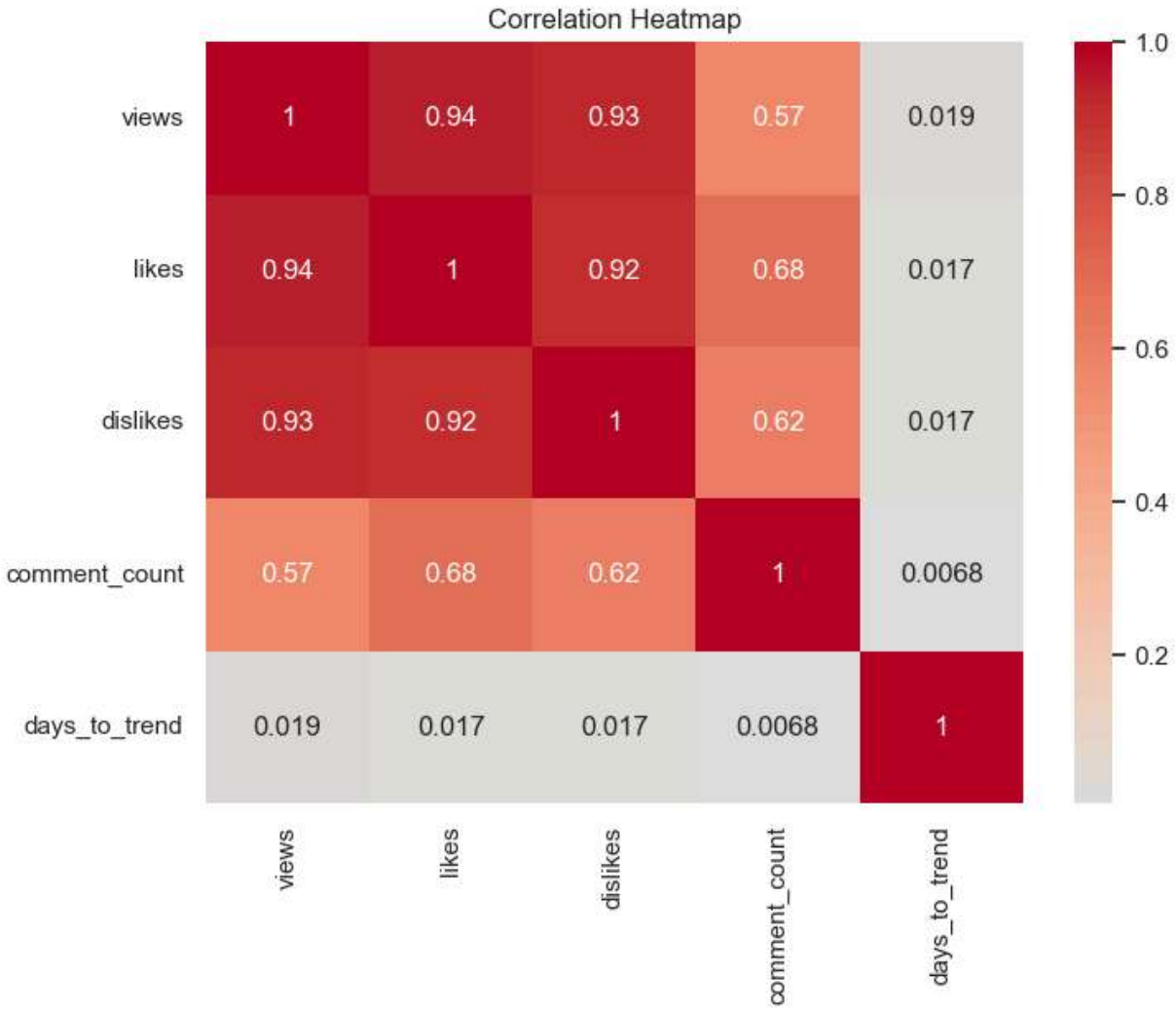
In [12]:
```
# 5. Days to Trend Analysis
```

In [13]:
```
plt.figure(figsize=(10,5))
sns.histplot(df['days_to_trend'], bins=30, kde=True, color="coral")
plt.title("Distribution of Days to Trend")
plt.xlabel("Days from Publish to Trending")
plt.ylabel("Number of Videos")
plt.show()
```



In [14]:
```
# 6. Heatmap of Correlations
```

In [15]:
```
corr = df[['views','likes','dislikes','comment_count','days_to_trend']].corr()

plt.figure(figsize=(8,6))
sns.heatmap(corr, annot=True, cmap="coolwarm", center=0)
plt.title("Correlation Heatmap")
plt.show()
```



In [ ]: