Netflix India YouTube Data Analysis

Presentated by **Masihuzzaman Sayyed**

Introduction

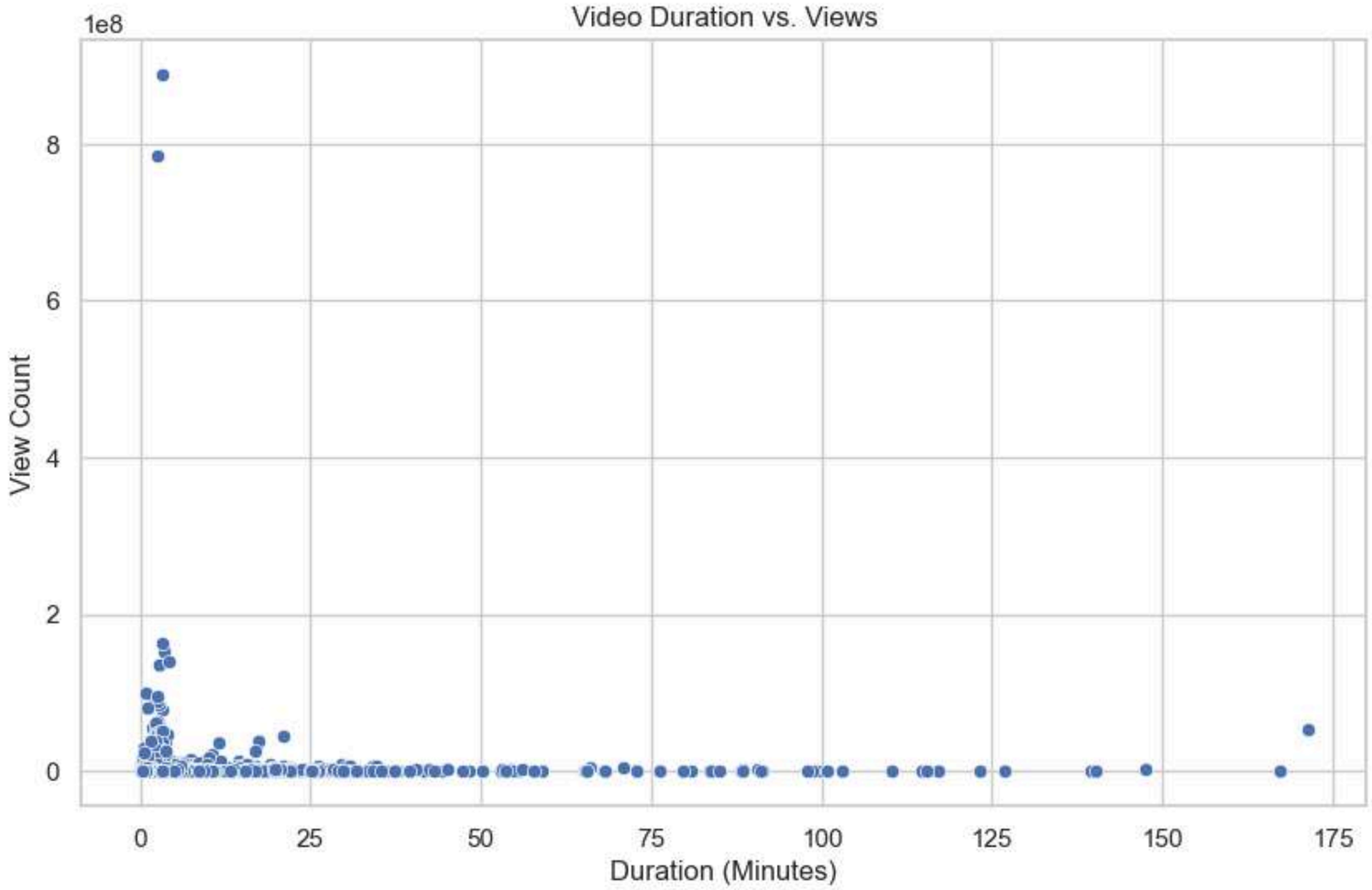
Our Social Media Analytics Team at AtliQ has been analysing successful YouTube channels to improve strategies for one of our clients. Netflix India’s YouTube channel is a great example. To help us learn from them, we’ve collected their channel data with metadata and prepared a set of questions that need answers.

📊 Question 1
Does the duration of the video influence views and comments?

code

```
import matplotlib.pyplot as plt
import seaborn as sns

# Set style
sns.set(style="whitegrid")

# Duration vs. Views
plt.figure(figsize=(10,6))
sns.scatterplot(data=df, x="DurationMinutes", y="viewCount")
plt.title("Video Duration vs. Views")
plt.xlabel("Duration (Minutes)")
plt.ylabel("View Count")
plt.show()

# Duration vs. Comments
plt.figure(figsize=(10,6))
sns.scatterplot(data=df, x="DurationMinutes", y="commentCount")
plt.title("Video Duration vs. Comments")
plt.xlabel("Duration (Minutes)")
plt.ylabel("Comment Count")
plt.show()
```

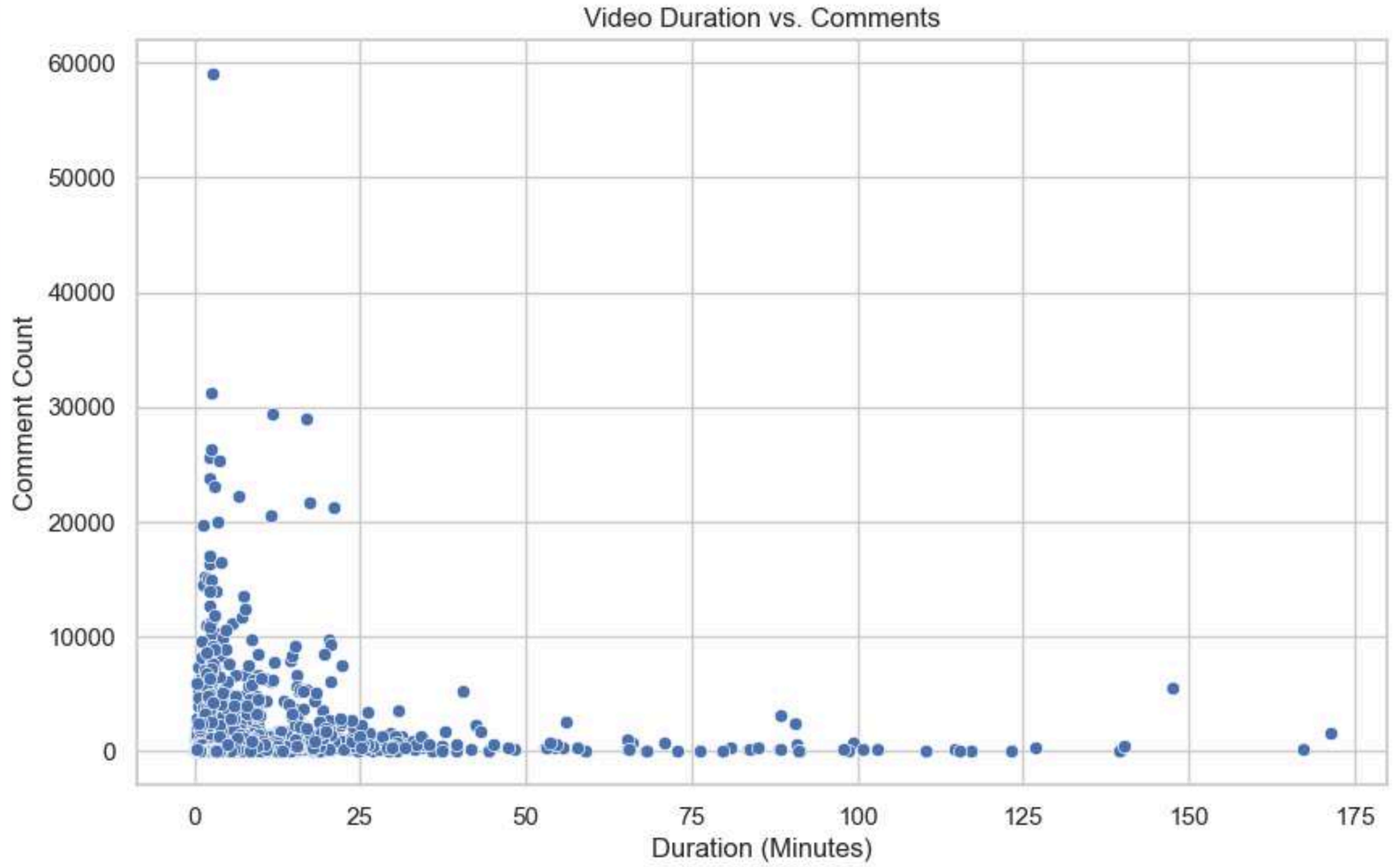
scatter plot vs views



The correlation between duration and views was **0.001**, and between duration and comments was **0.056**.

These results show that video length has no measurable impact on views or comments.

scatter plot vs comments



The correlation between duration and views was **0.001**, and between duration and comments was **0.056**.

These results show that video length has no measurable impact on views or comments.

📊 Question 2
Is there a relation between views and comments?

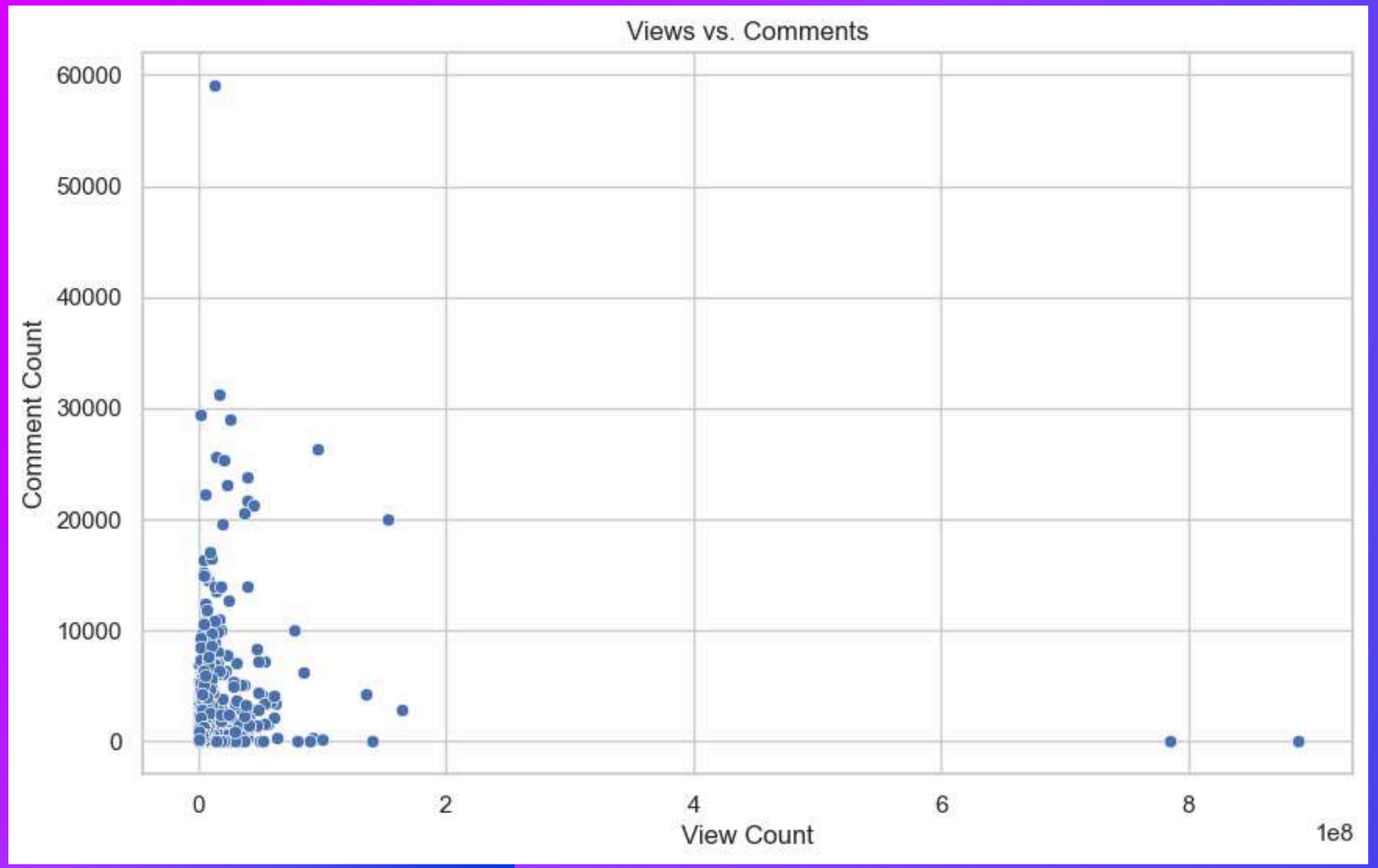
code

```
# Scatter plot
plt.figure(figsize=(10,6))
sns.scatterplot(data=df, x="viewCount", y="commentCount")
plt.title("Views vs. Comments")
plt.xlabel("View Count")
plt.ylabel("Comment Count")
plt.show()

# Correlation
corr_views_comments = df["viewCount"].corr(df["commentCount"])
print("Correlation between Views and Comments:", round(corr_views_comments, 3))
```

✓ 0.1s

The correlation was **0.143**, indicating a weak positive relationship.
So, videos with more views do get slightly more comments, but the relationship isn't strong.

scatter plot - view vs comments



The correlation was **0.143**, indicating a weak positive relationship.
So, videos with more views do get slightly more comments, but the relationship isn't strong.

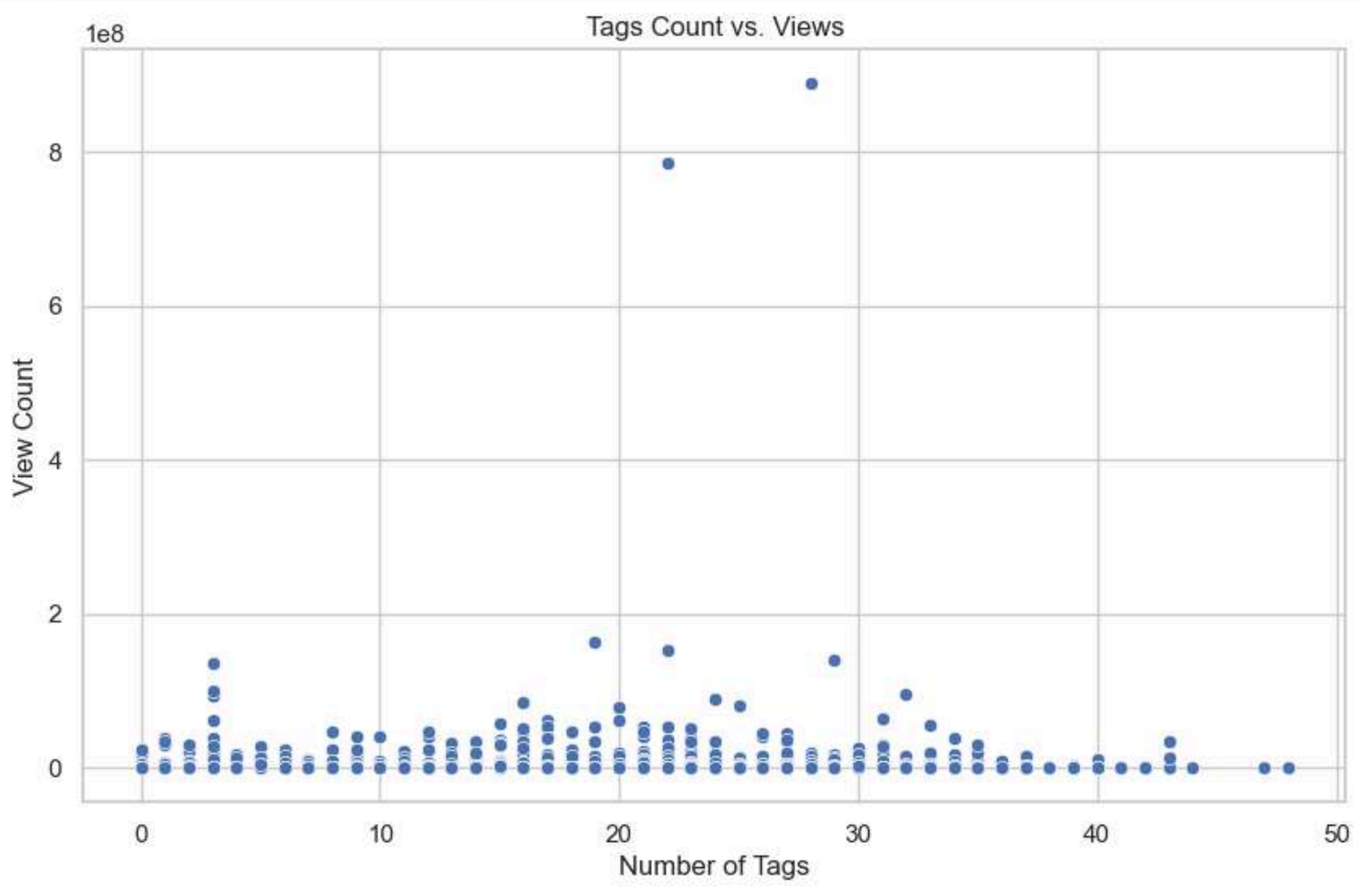
📊 Question 3
Does tag count matter to get more views?

code

```
# Scatter plot: TagsCount vs. Views
plt.figure(figsize=(10,6))
sns.scatterplot(data=df, x="TagsCount", y="viewCount")
plt.title("Tags Count vs. Views")
plt.xlabel("Number of Tags")
plt.ylabel("View Count")
plt.show()

# Correlation
corr_tags_views = df["TagsCount"].corr(df["viewCount"])
print("Correlation between Tags Count and Views:", round(corr_tags_views, 3))
```

The correlation was **0.007**, indicating that the tag count has no significant impact on views.

scatter plot - tag count vs views



The correlation was **0.007**, indicating that the tag count has no significant impact on views.

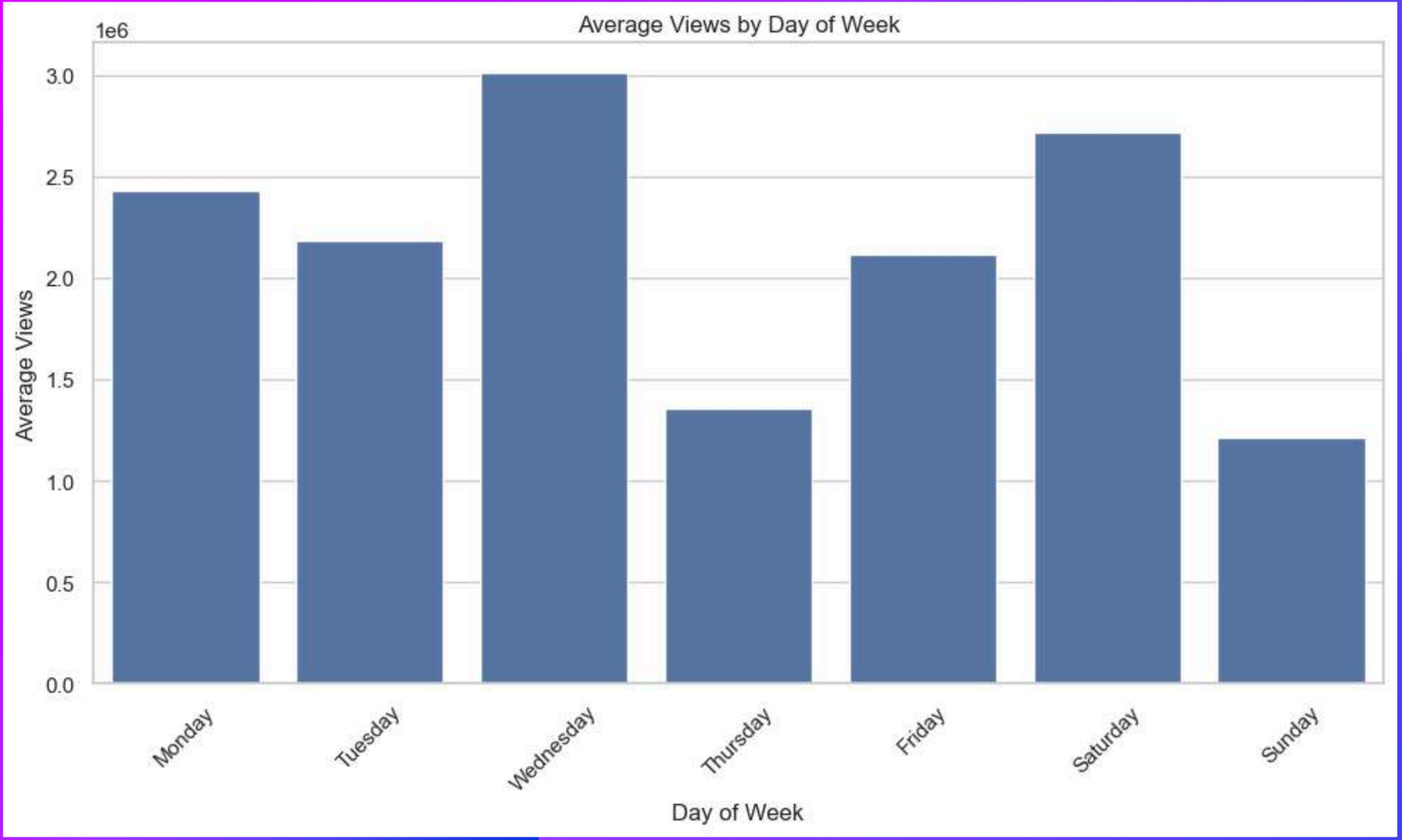
📊 Question 4
Does video publishing day or
time influence engagement?

code

```
# Make sure to parse the publishedAt column to datetime
df["publishedAt_dt"] = pd.to_datetime(df["publishedAt"])

# Day of week (Monday=0, Sunday=6)
df["PublishDay"] = df["publishedAt_dt"].dt.day_name()

# Hour of day (0-23)
df["PublishHour"] = df["publishedAt_dt"].dt.hour

# Quick check
print(df[["publishedAt", "PublishDay", "PublishHour"]].head())
```

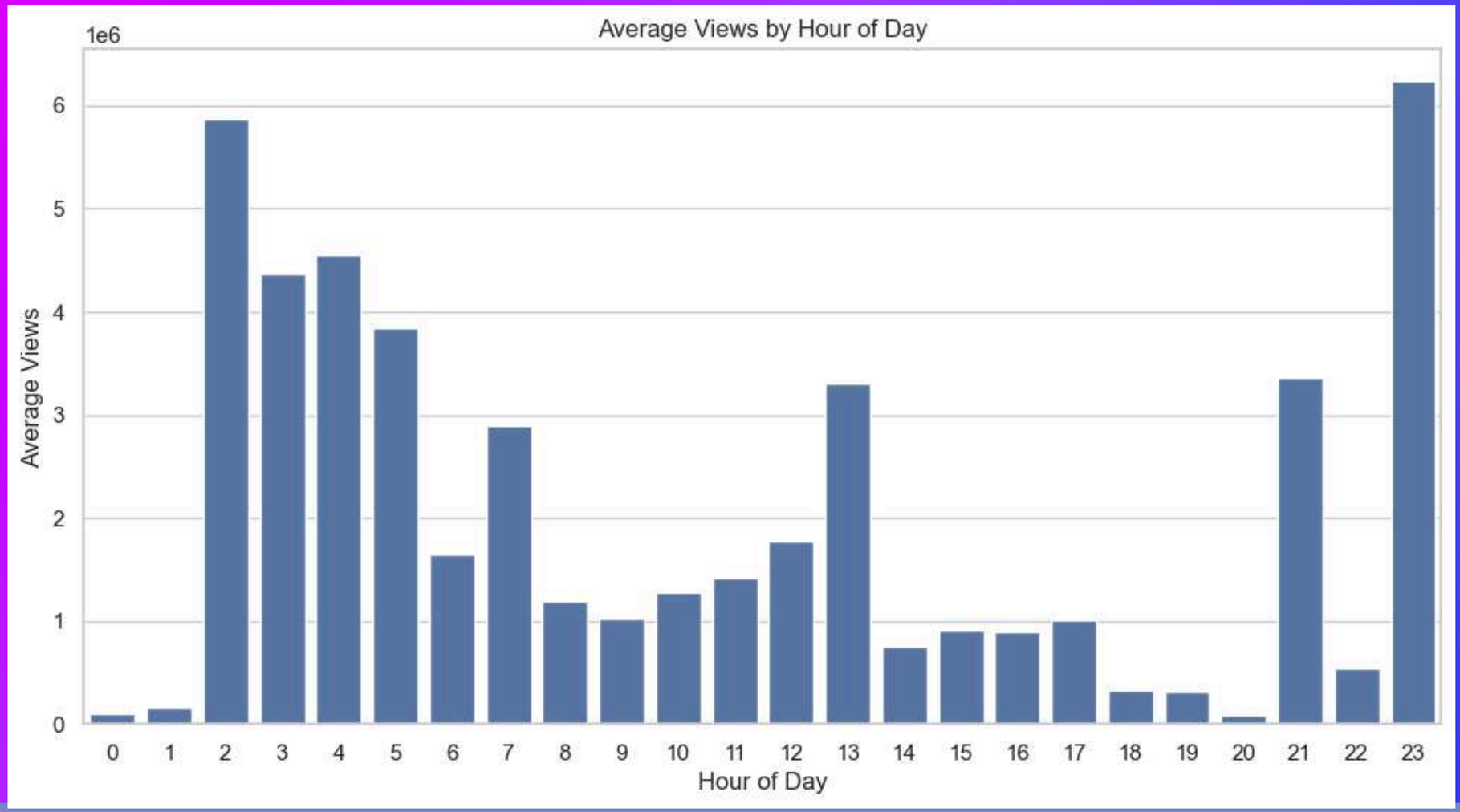
I found that **Wednesday** had the highest average views, and videos published late at night performed better.

Bar Chart - Avg Views By Day

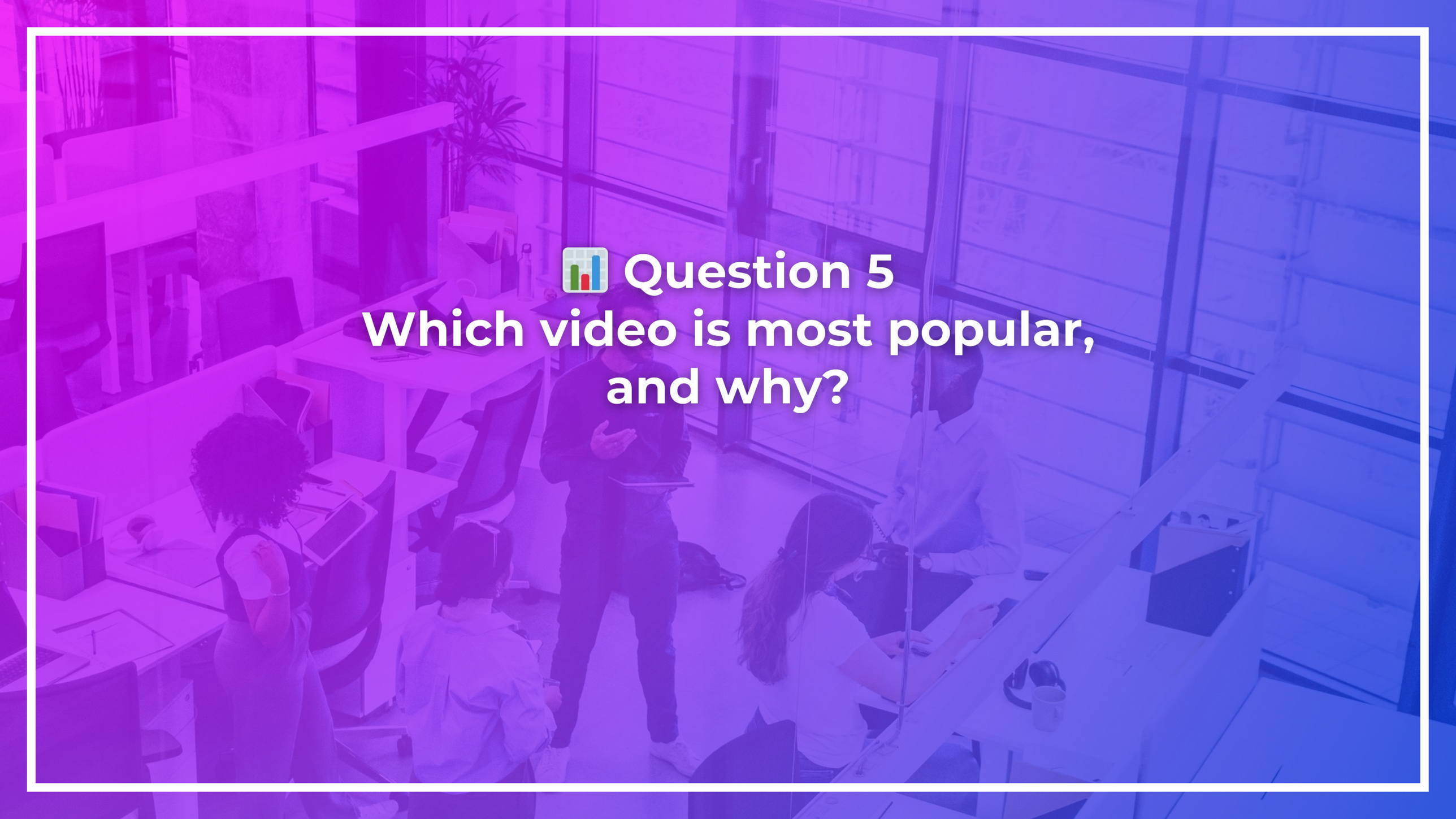


I found that **Wednesday** had the highest average views, and videos published late at night performed better.

Bar Chart - Avg Views By Hour



I found that **Wednesday** had the highest average views, and videos published late at night performed better.


📊 Question 5
Which video is most popular,
and why?

code

```
most_viewed = df.loc[df["viewCount"].idxmax()]
print(most_viewed)
```

The most viewed video was **Mighty Little Bheem | Bheem with a Beat**, with **888** million views.

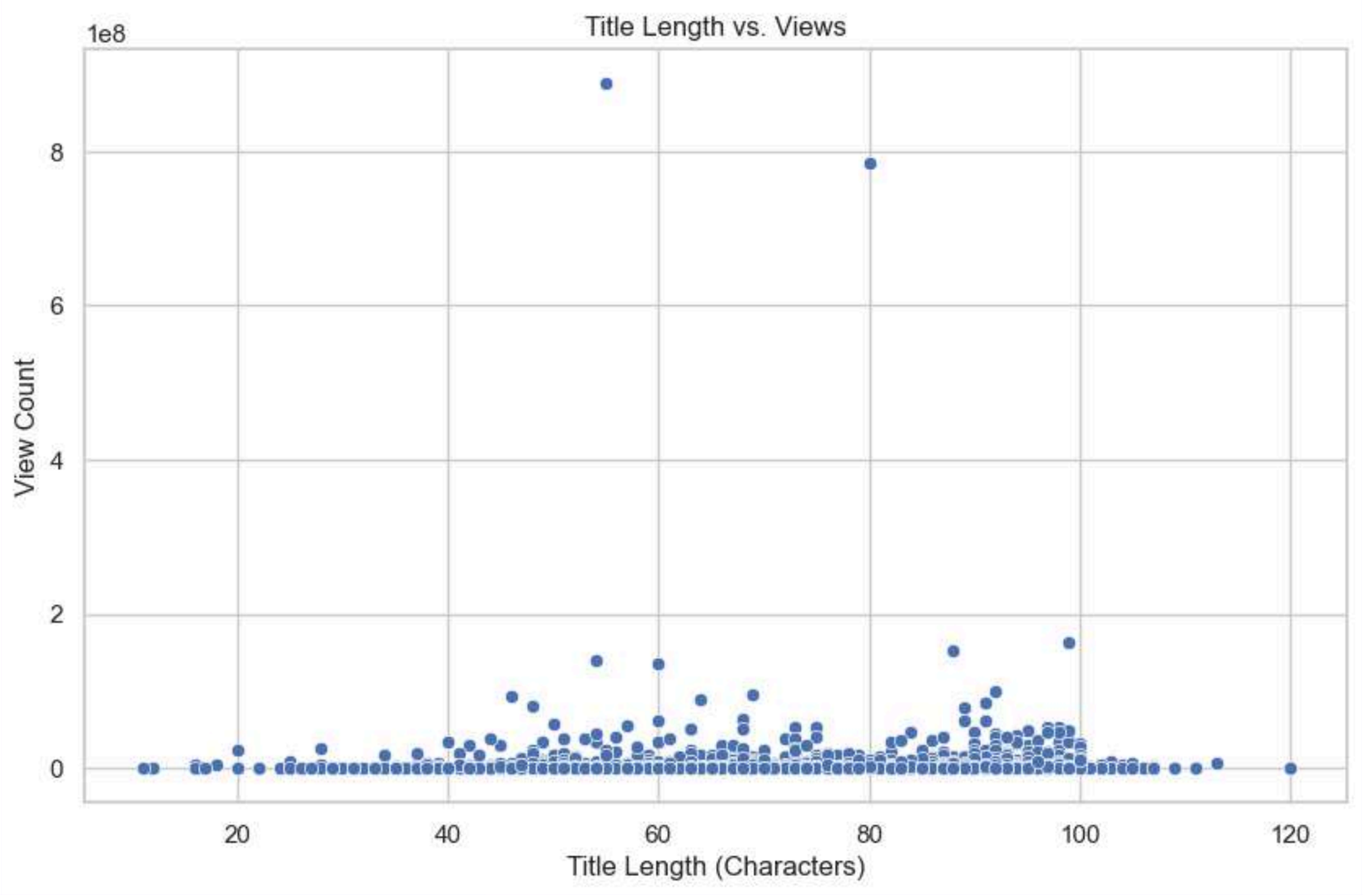
📊 Question 6
Does title length influence
views?

code

```
# Scatter plot: Title Length vs. Views
plt.figure(figsize=(10,6))
sns.scatterplot(data=df, x="TitleLength", y="viewCount")
plt.title("Title Length vs. Views")
plt.xlabel("Title Length (Characters)")
plt.ylabel("View Count")
plt.show()

# Correlation
corr_title_views = df["TitleLength"].corr(df["viewCount"])
print("Correlation between Title Length and Views:", round(corr_title_views, 3))
```

The correlation was **0.004**, proving no relationship between title length and views.

scatter plot - Video len vs views



The correlation was **0.004**, proving no relationship between title length and views.

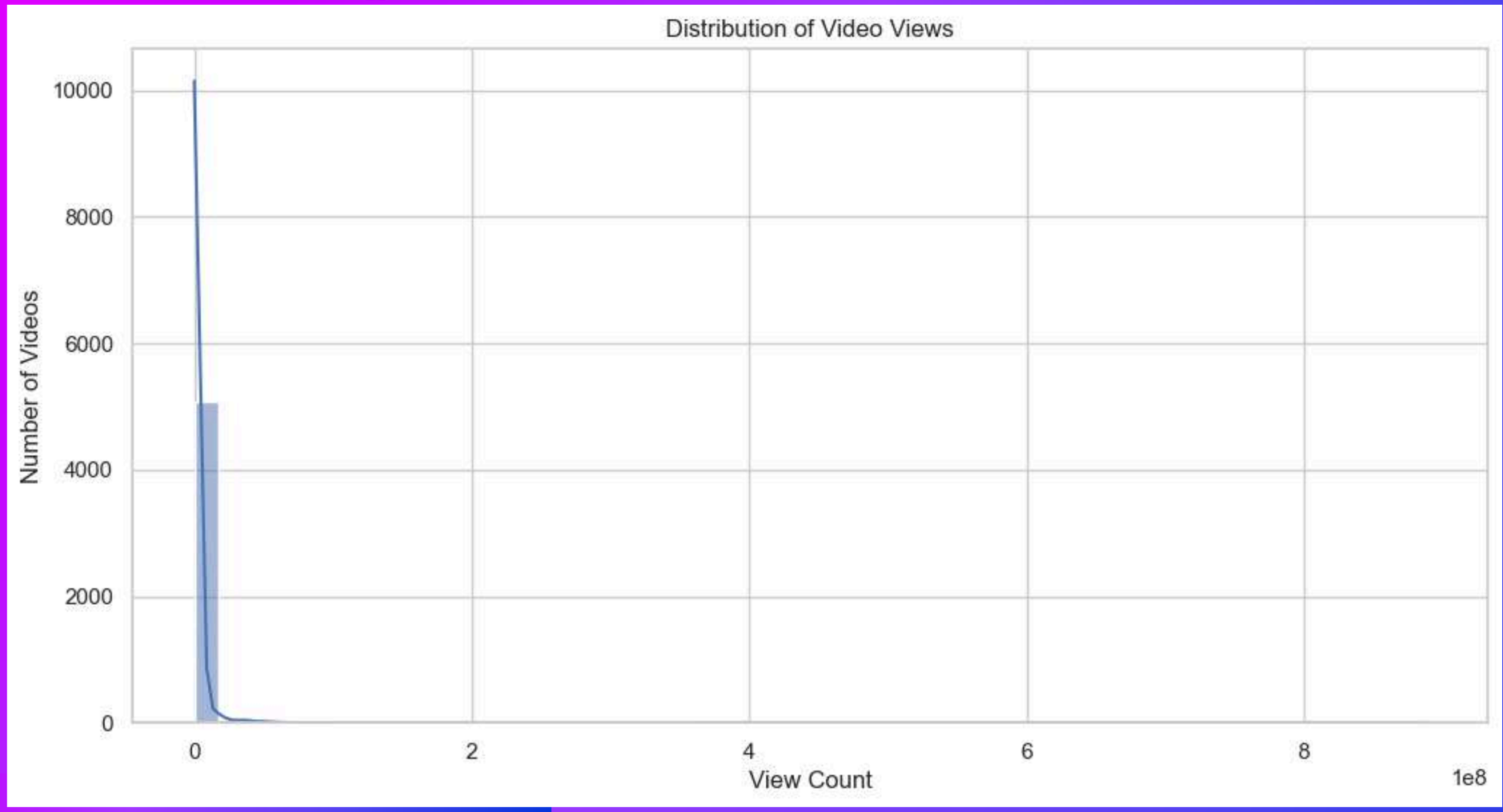
📊 Question 7
What is the distribution of views?

code

```
plt.figure(figsize=(12,6))
sns.histplot(df["viewCount"], bins=50, kde=True)
plt.title("Distribution of Video Views")
plt.xlabel("View Count")
plt.ylabel("Number of Videos")
plt.show()

print(df["viewCount"].describe())
```

The dataset is heavily skewed, with most videos having **fewer than 1 million views**, and a few blockbusters driving the average up.

Histogram - Video views



The correlation was **0.004**, proving no relationship between title length and views.

📊 Question 8
Which topics or tags perform best?

code

```
tag_rows = []
for idx, row in df.iterrows():
    tags = row["tags"]
    if pd.isna(tags) or tags.strip() == "":
        continue
    tag_list = [tag.strip() for tag in tags.split(",")]
    for tag in tag_list:
        tag_rows.append({"tag": tag, "viewCount": row["viewCount"]})

tags_df = pd.DataFrame(tag_rows)
tag_views = tags_df.groupby("tag")["viewCount"].mean().reset_index()
tag_views = tag_views.sort_values("viewCount", ascending=False)
```

Top tags included '**chhota bheem cartoon**' and '**pongal video**', showing kids content dominates.

Table Comparison

Rank	Tag	Average Views
1	'chhota bheem cartoon'	~784 million
2	'pongal video'	~392 million
3	'Shows for kids'	~310 million
4	'Drums'	~224 million
5	'kite festival'	~196 million
6	'ankur warikoo'	~164 million
7	'fabulous lives of bollywood lives'	~164 million
8	'tejasswi prakash'	~164 million
9	'tony kakkar'	~164 million
10	'playback 2022'	~164 million

Top tags included '**chhota bheem cartoon**' and '**pongal video**', showing kids content dominates.

📊 Question 9
Did any videos without tags
perform well?

code

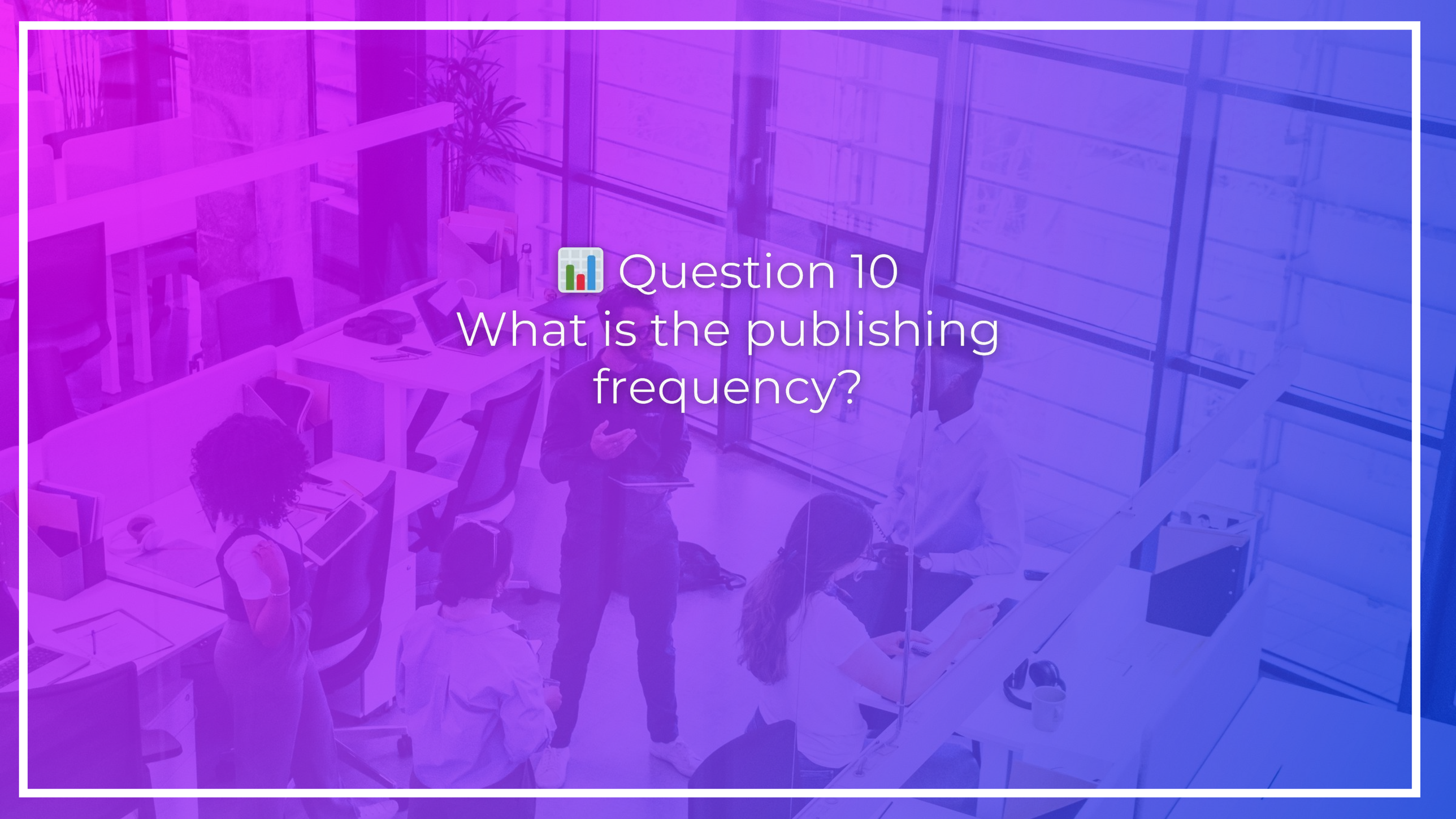
```
no_tags_df = df[df["tags"].isna() | (df["tags"].str.strip() == "")]
no_tags_sorted = no_tags_df.sort_values("viewCount", ascending=False)
```

Some untagged videos like **Heating Up** and **Ghoul** trailers still got millions of views, likely due to brand power.

Table Comparison

Rank	Title	Views	Likes	Comments	Published At
1	Heating Up | Netflix	24,051,981	26,609	2,410	2018-04-02 07:10:31
2	Ghoul - Silent Trailer | Netflix | HD	18,523,444	40,118	2,379	2018-08-03 06:58:10
3	Ghoul: The Room | Official [VR] | Netflix	18,407,319	41,185	1,823	2018-09-24 09:52:48
4	To-Dos | Netflix	4,419,563	36	8	2018-04-02 07:09:58
5	Ad Break | Netflix	3,582,965	146	10	2018-04-02 07:10:07
6	Dangal | Main Trailer | Netflix	1,253,173	5,052	128	2017-07-04 13:47:08
7	Marvel's Iron Fist | 7 Days of Iron Fist | Netflix	295,551	108	5	2017-03-10 06:10:04
8	The Ivory Game – Official Trailer – Netflix Documentary	292,647	239	88	2016-10-21 16:30:00
9	Little Things S3 | Mithila Palkar Helps...	231,591	4,711	75	2019-11-16 10:40:21
10	New on Netflix October 2018 | India	172,218	2,483	270	2018-10-02 03:46:05

Some untagged videos like **Heating Up** and **Ghoul** trailers still got millions of views, likely due to brand power.

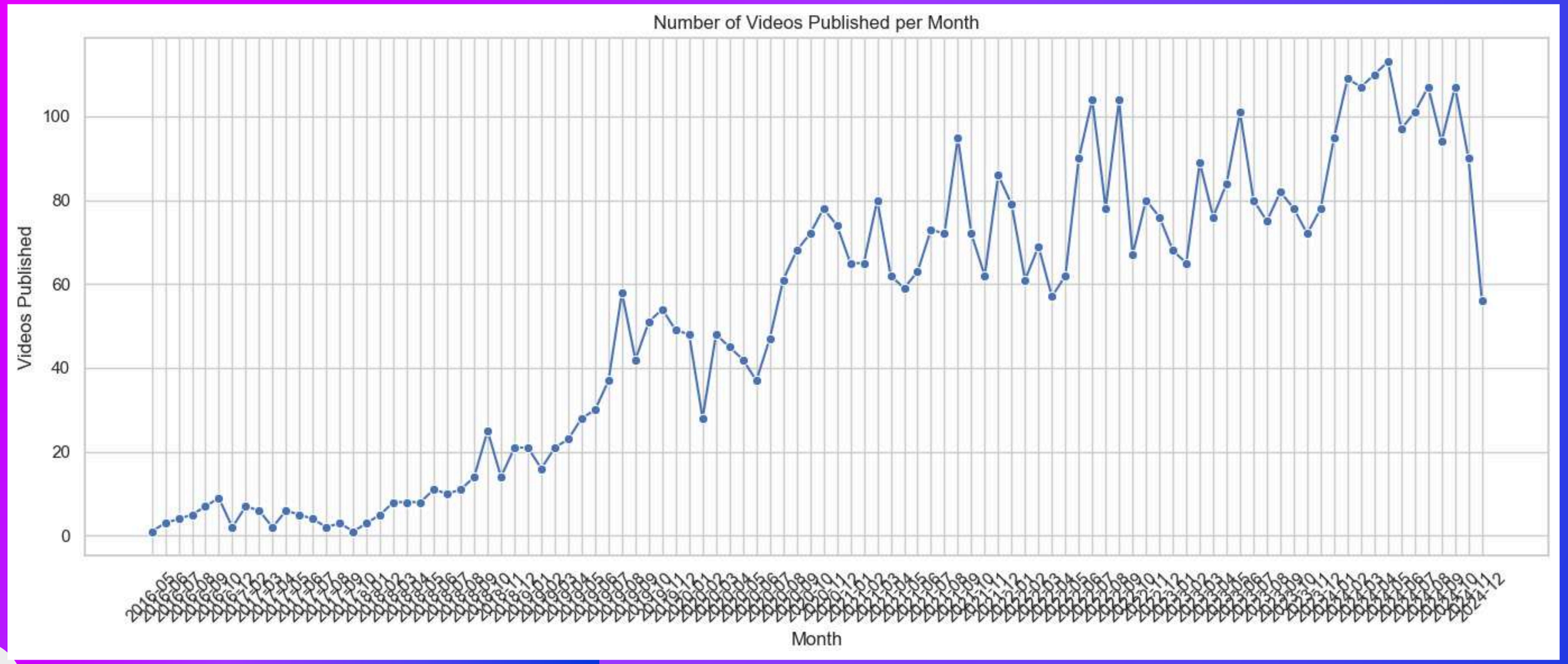

📊 Question 10
What is the publishing frequency?

code

```
df["PublishMonth"] = df["publishedAt_dt"].dt.to_period("M")
monthly_counts = df.groupby("PublishMonth").size().reset_index(name="VideoCount")
```

Publishing started slow in 2016 and scaled to over **100 videos per month** by 2020.

line chart per month



Publishing started slow in 2016 and scaled to over **100 videos per month** by 2020.

📊 Question 11
What are the most used tags overall?

code

```
all_tags = []
for idx, row in df.iterrows():
    tags = row["tags"]
    if pd.isna(tags) or tags.strip() == "":
        continue
    tag_list = [tag.strip() for tag in tags.split(",")]
    all_tags.extend(tag_list)

tags_series = pd.Series(all_tags)
tag_counts = tags_series.value_counts().reset_index()
```

Top tags were **'netflix', 'netflix india', and 'netflix shows',** reflecting strong brand consistency.

Top 20 Most Frequently Used Tags

Rank	Tag	Frequency
1	'netflix'	2,861
2	['netflix india' *(trailing bracket)*	2,458
3	'netflix shows'	2,297
4	'Netflix'	656
5	'netflix india'	545
6	'Netflix India'	436
7	'netflix shows'] *(trailing bracket)*	389
8	'netflix movies'	190
9	'kapil sharma'	164
10	'ranbir kapoor'	160
11	['Netflix' *(trailing bracket)*	160
12	'kapil sharma comedy'	154
13	'sunil grover'	152
14	'comedy'	149
15	'sunil grover and kapil sharma'	146
16	'bollywood'	143
17	'sunil grover comedy'	142
18	'kapil sharma show funny episode'	129
19	'new on netflix'	129
20	'trailer'	118

Publishing started slow in 2016 and scaled to over **100 videos per month** by 2020.

📊 Question 12
Which videos had the highest engagement rate?

code

```
df["EngagementRate"] = (df["likeCount"] + df["commentCount"]) / df["viewCount"].replace(0, 1)
engaged_videos = df.sort_values("EngagementRate", ascending=False)
```

The most engaging videos were influencer collaborations with **Yo Yo Honey Singh, Triggered Insaan, and Mythpat**.

Top 10 Videos by Engagement Rate (Likes + Comments per View)

Rank	Title	Views	Likes	Comments	Engagement Rate
1	Geeked Week 2024 LIVE Showcase | Ft. Squid Game...	0	246	0	246
2	Meet THE REAL Honey Singh | Yo Yo Honey Singh: Netflix Documentary	52,864	7,111	394	0.142
3	Game Hai? Ya Sachhi Ka Extraction? | @Thugesh	269,422	28,804	1,429	0.112
4	Love And Life Lessons By Kamal Haasan | Anbe Sivam	24,545	2,435	261	0.11
5	Survival Game Show With @FukraInsaan | Kaala Paani	1,231,836	122,532	8,441	0.106
6	@triggeredinsaan's Engineering Life Was Like Kaala Paani	1,365,480	133,907	4,142	0.101
7	What is @triggeredinsaan's Lakshya? | Triggered Insaan	1,140,810	110,630	3,389	0.1
8	Which Anime Universe Would You Like to Visit?	51,368	3,828	1,002	0.094
9	@Mythpat Takes The Ultimate Mimicry Challenge	2,082,222	186,586	8,938	0.094
10	Why Bhuvan Bam has a Problem with Netflix | @Bhuvan Bam	1,509,471	135,659	1,904	0.091

The most engaging videos were influencer collaborations with **Yo Yo Honey Singh, Triggered Insaan, and Mythpat.**

Thank you

masihuzzaman sayyed

data analyst

masisayyedofficial@gmail.com