

YouTube

Analysis Python Project

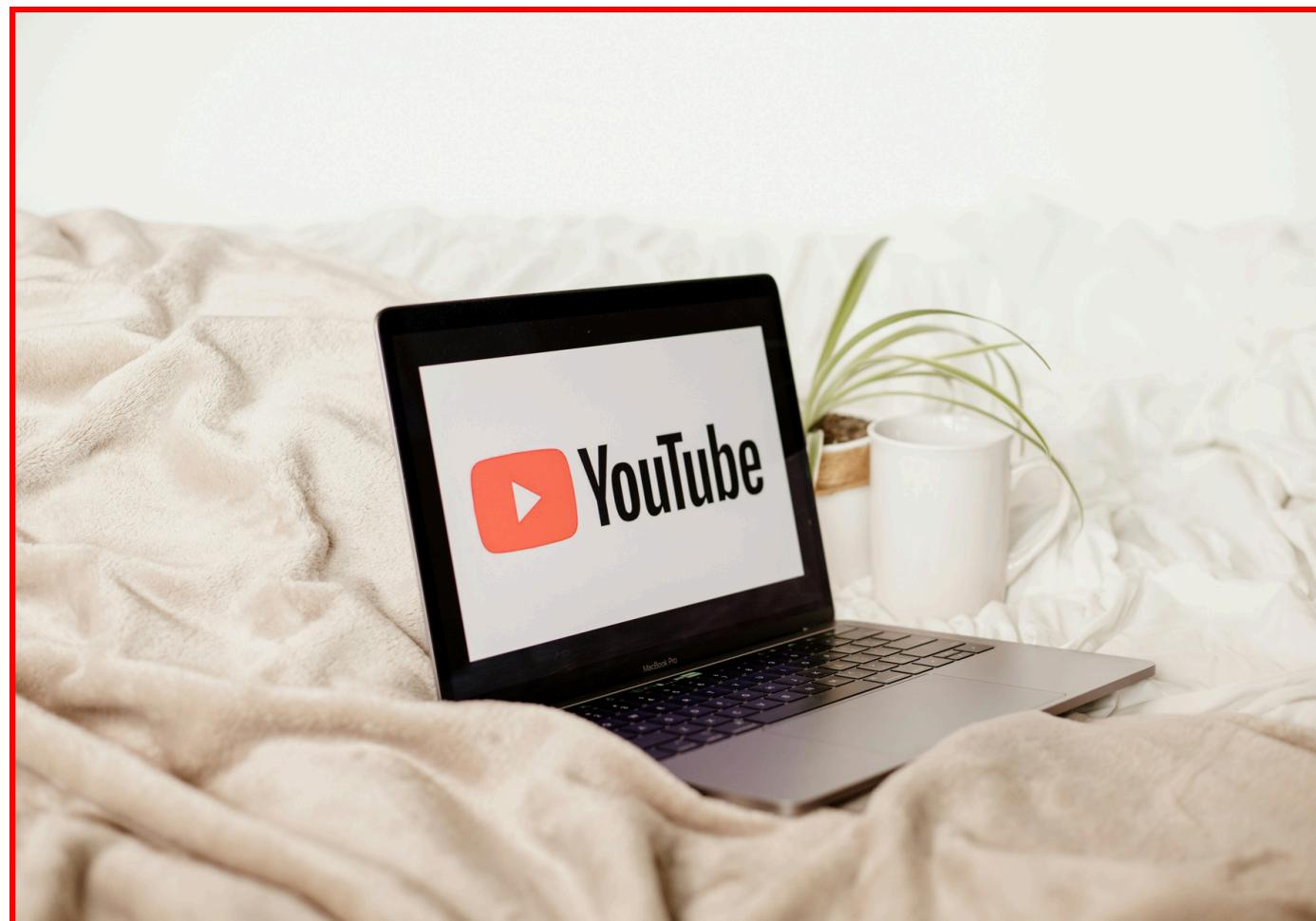
Scraping, Analyzing & Visualizing YouTube Data

Leveraging the YouTube Data API
for Insights into Top Channels



Project Overview

You Tube



- The project focuses on extracting and analyzing YouTube channel statistics using the YouTube Data API.
- Built an end-to-end pipeline for data extraction, cleaning, exploratory data analysis (EDA), and visualization.
- Special case study: 'T-Series' channel analysis.

Project Goal

You Tube

Understand trends and performance of popular YouTube channels.



Identify metrics driving subscriber growth and engagement.



Extract actionable insights from real-world YouTube data.



Showcase how Python can automate, analyze, and visualize API-based data.



Project Workflow

You Tube



1. Read and understand YouTube API documentation.
2. Generate YouTube API Key.
3. Create virtual environment & install dependencies.
4. From top 50 channels, select 15 channels based on subscriber count.
5. Scrape, analyze, and visualize channel statistics.
6. Deep-dive into 'T-Series' channel for detailed insights.

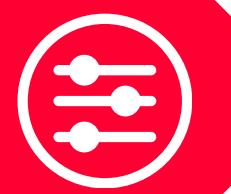
Performance Highlights Tools and Technologies Used

You Tube



Environment: Jupyter Notebook

A Jupyter Notebook virtual environment is an isolated Python setup where notebooks run with their own interpreter and packages, preventing conflicts and ensuring reproducibility.



Programming Language: Python

Libraries:

Data Handling: Numpy, Pandas

Visualization: Matplotlib, Seaborn

API Integration: Googleapiclient

Formatting/Text: Textwrap

Scrape, Analyze and Visualization Channels Statistics

You Tube

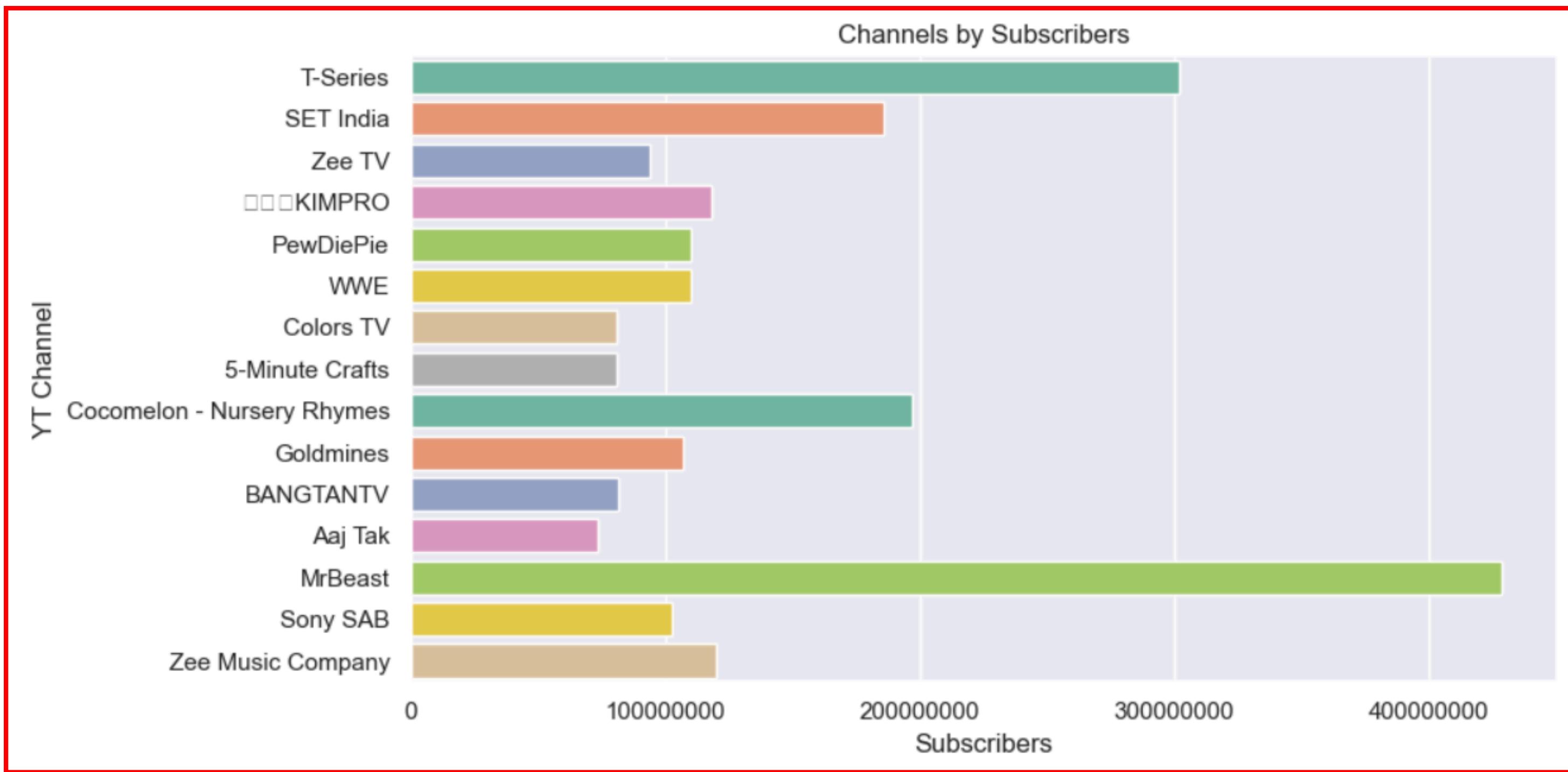
From top 50 channels, select 15 channels based on subscriber count:

	Channel_name	Country	Language	Joined_YT	Subscribers	Views	Total_videos	playlist_id
0	T-Series	IN	Not Available	2006-03-13T14:27:05Z	302000000	310365998401	24243	UUq-Fj5jknLsUf-MWSy4_brA
1	SET India	IN	Not Available	2006-09-20T22:24:59Z	186000000	181193920254	159489	UUpEhnqL0y41EpW2TvWAHD7Q
2	Zee TV	IN	en	2005-12-11T06:23:26Z	94100000	108787457676	219359	UUppHT7SZKKvar4Oc9J4oljQ
3	김프로KIMPRO	KR	Not Available	2017-11-11T17:13:52Z	118000000	103490129843	3463	UUiVs2pnGW5mLlc1jS2nxhjg
4	PewDiePie	JP	Not Available	2010-04-29T10:54:00Z	110000000	29399862973	4648	UU-IHJZR3Gqxm24_Vd_AJ5Yw
5	WWE	Not Available	Not Available	2007-05-11T01:20:02Z	110000000	97784798271	87486	UUJ5v_MCY6GNUBTO8-D3XoAg
6	Colors TV	IN	Not Available	2008-06-13T11:09:32Z	81000000	81343692575	281295	UU55IWqFLDH1Xp7iu1_xknRA
7	5-Minute Crafts	US	Not Available	2016-11-15T09:25:37Z	81000000	28258563563	7806	UU295-Dw_tDNTZXFeAPA6Aw
8	Cocomelon - Nursery Rhymes	US	en	2006-09-01T22:18:49Z	197000000	208186897881	1688	UUbCmjCuTUZos6Inko4u57UQ
9	Goldmines	IN	Not Available	2012-01-21T08:42:21Z	107000000	30753488490	11511	UUyoXW-Dse7fURq30EWI_CUA
10	BANGTANTV	Not Available	Not Available	2012-12-17T03:20:38Z	81300000	25531146524	2916	UULkAepWjdylmXSItofFvsYQ
11	Aaj Tak	IN	Not Available	2009-08-27T11:54:24Z	73400000	40941536568	499743	UUt4t-jeY85JegMIZ-E5UWtA
12	MrBeast	US	en	2012-02-20T00:43:50Z	429000000	93874727227	898	UX6OQ3DkcsbYNNE6H8uQQuVA
13	Sony SAB	IN	Not Available	2007-08-04T16:05:01Z	103000000	133854318153	102473	UU6-F5tO8uklgE9Zy8lvbdFw
14	Zee Music Company	IN	Not Available	2014-03-12T11:59:25Z	120000000	81079162560	14728	UUFFbnve3yF62-tVXkTyHqg

Scrape, Analyze and Visualization Channels Statistics

You Tube

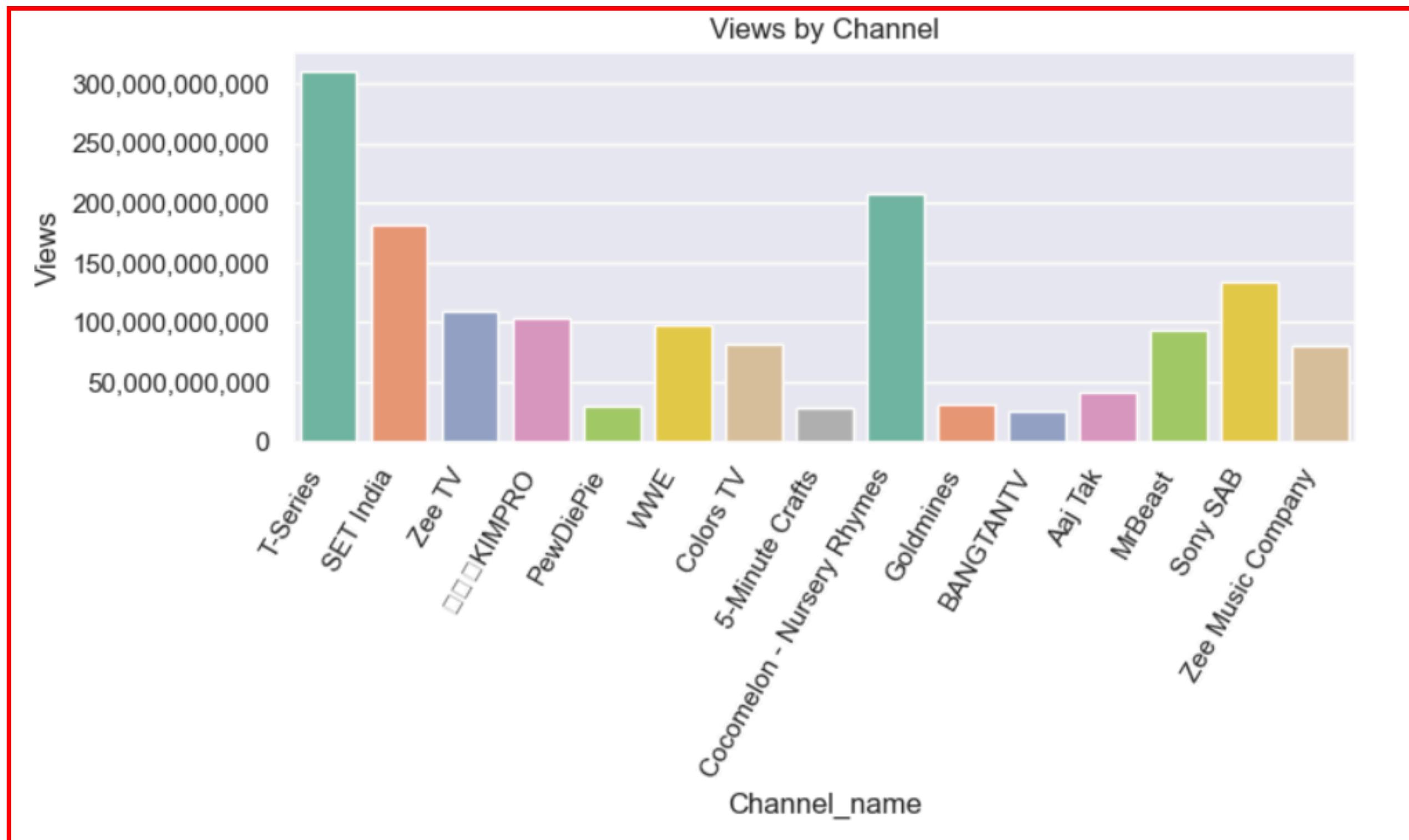
Exploratory Data Analysis (EDA):



Scrape, Analyze and Visualization Channels Statistics

You Tube

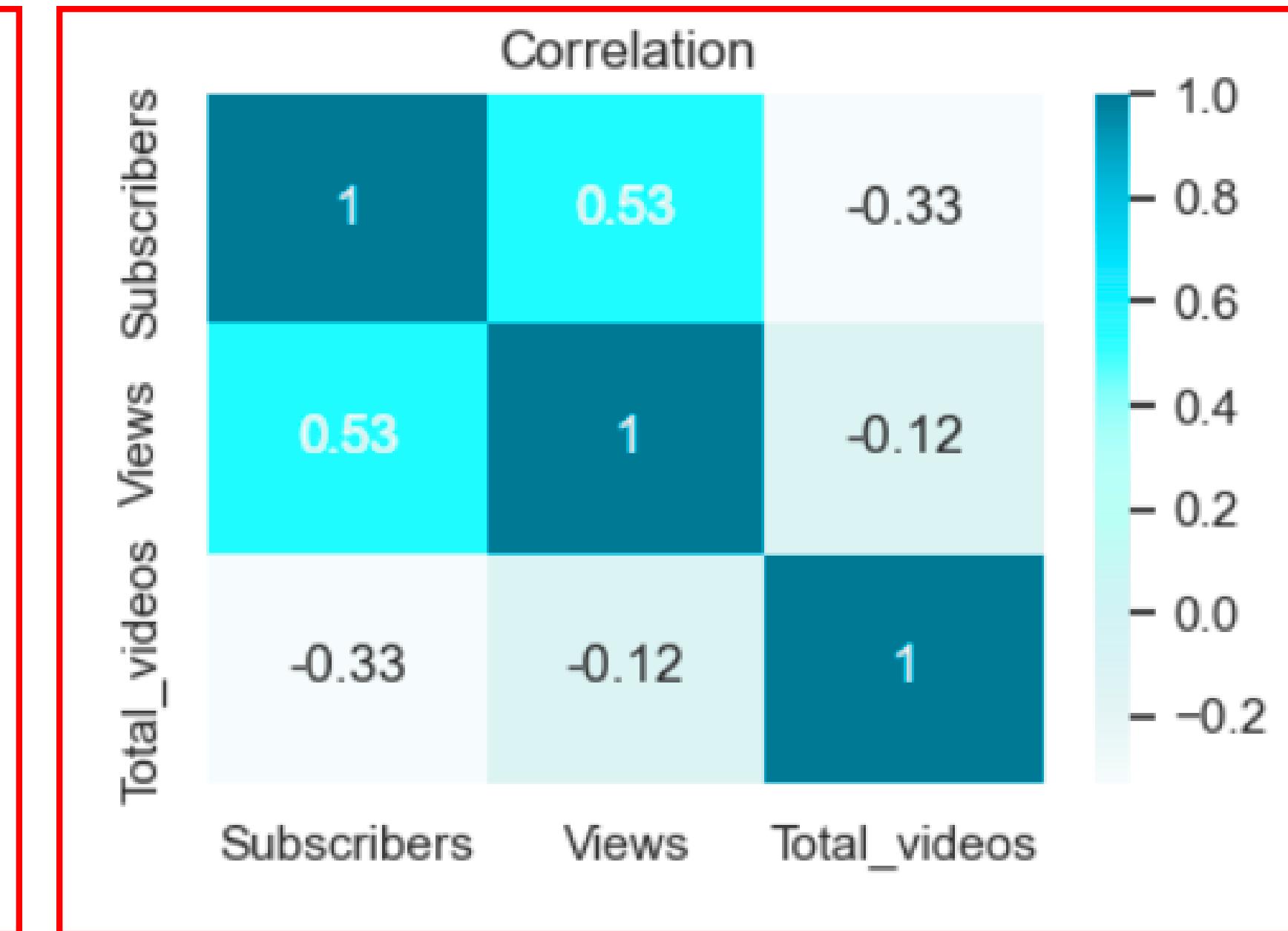
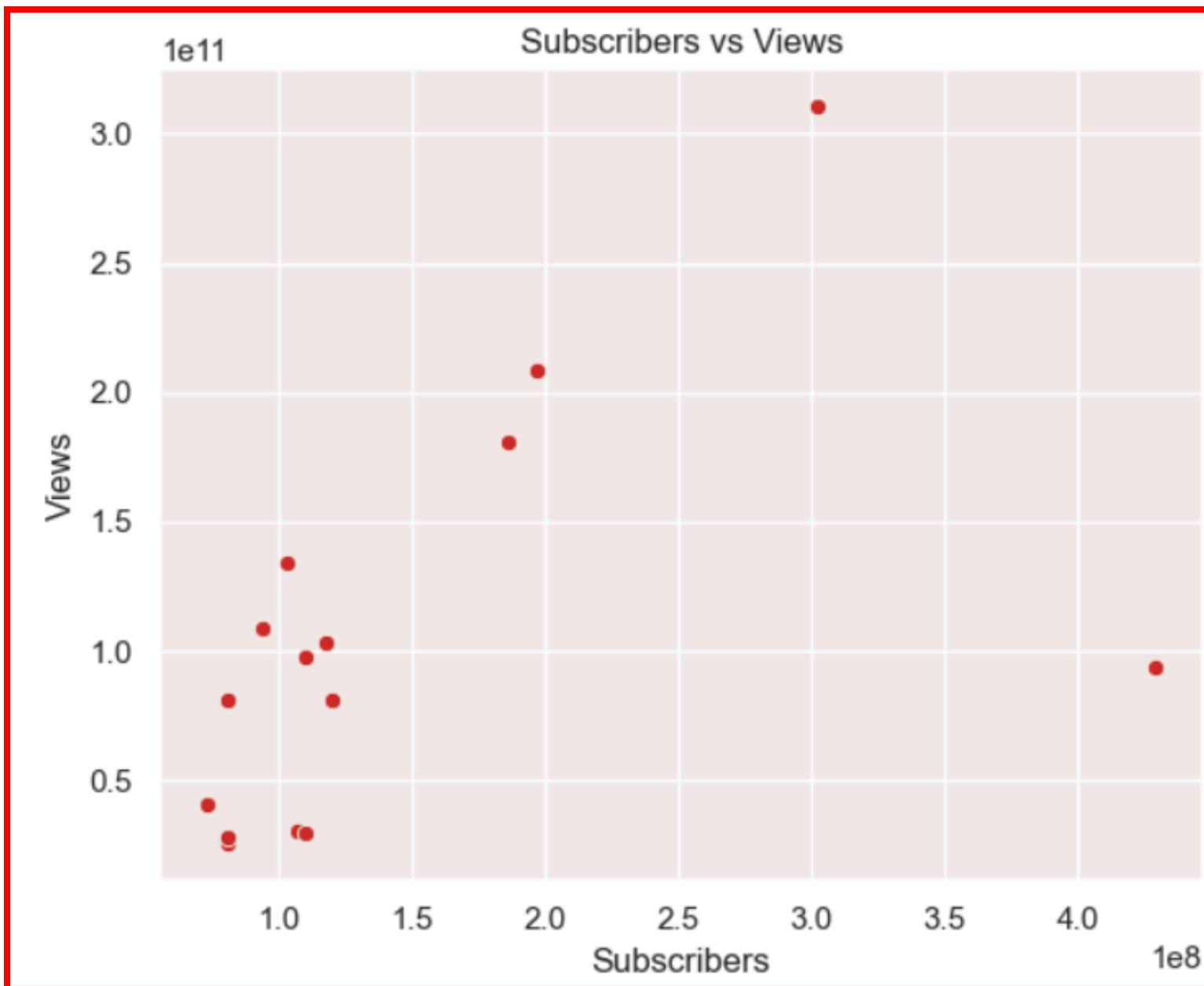
Exploratory Data Analysis (EDA):



Scrape, Analyze and Visualization Channels Statistics

You Tube

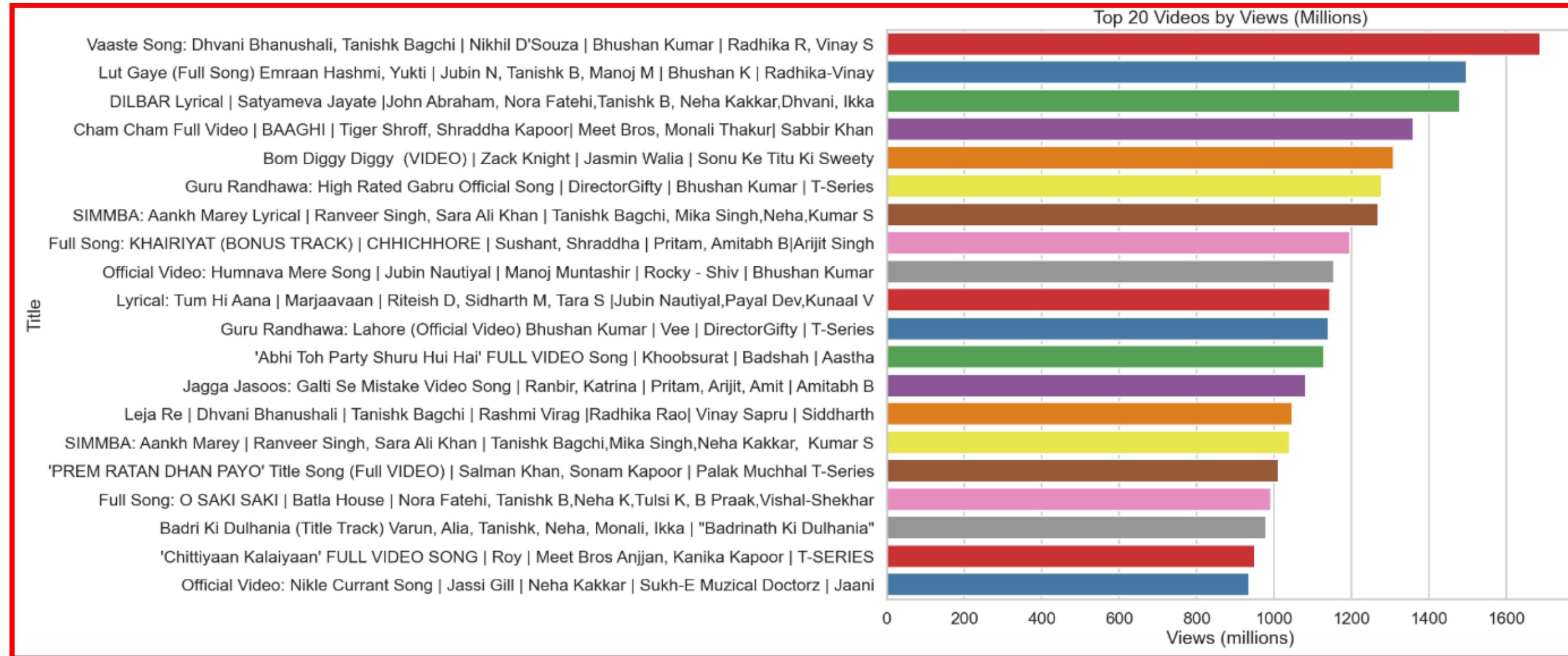
Exploratory Data Analysis (EDA):



Deep-dive into 'T-Series' channel

You Tube

Exploratory Data Analysis (EDA):



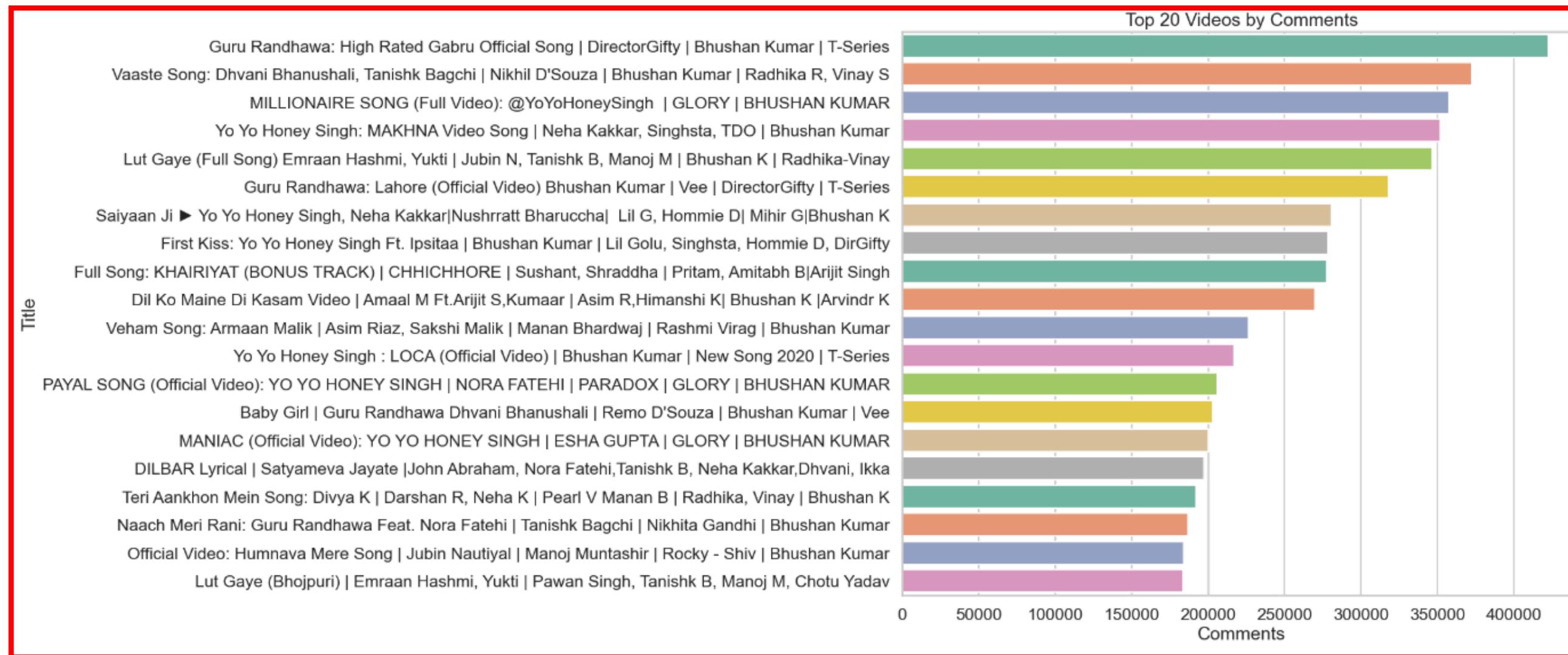
Insights from Analysis:

- A single title leads views by a noticeable margin, crossing roughly 1.6B, ahead of the pack.
- Rest of the top 20 cluster between ~0.9B and ~1.4B views, indicating a strong but tighter second tier.
- Several repeat marquee artists appear frequently among top-viewed videos, signaling artist/brand pull on reach.
- After the top few, view counts decline gradually rather than sharply, showing a long high-performing tail.

Deep-dive into 'T-Series' channel

You Tube

Exploratory Data Analysis (EDA):



Insights from Analysis:

- A few titles dominate comment volume, with the top video substantially ahead of the rest.
- The top 5–8 videos cluster in a high band, then totals drop gradually, indicating a long but decaying tail.
- Multiple high-comment videos feature recurring marquee artists, hinting that star power and franchise songs drive discussion.
- Beyond the leaders, mid-tier videos still generate significant conversations, suggesting broad audience engagement across releases.

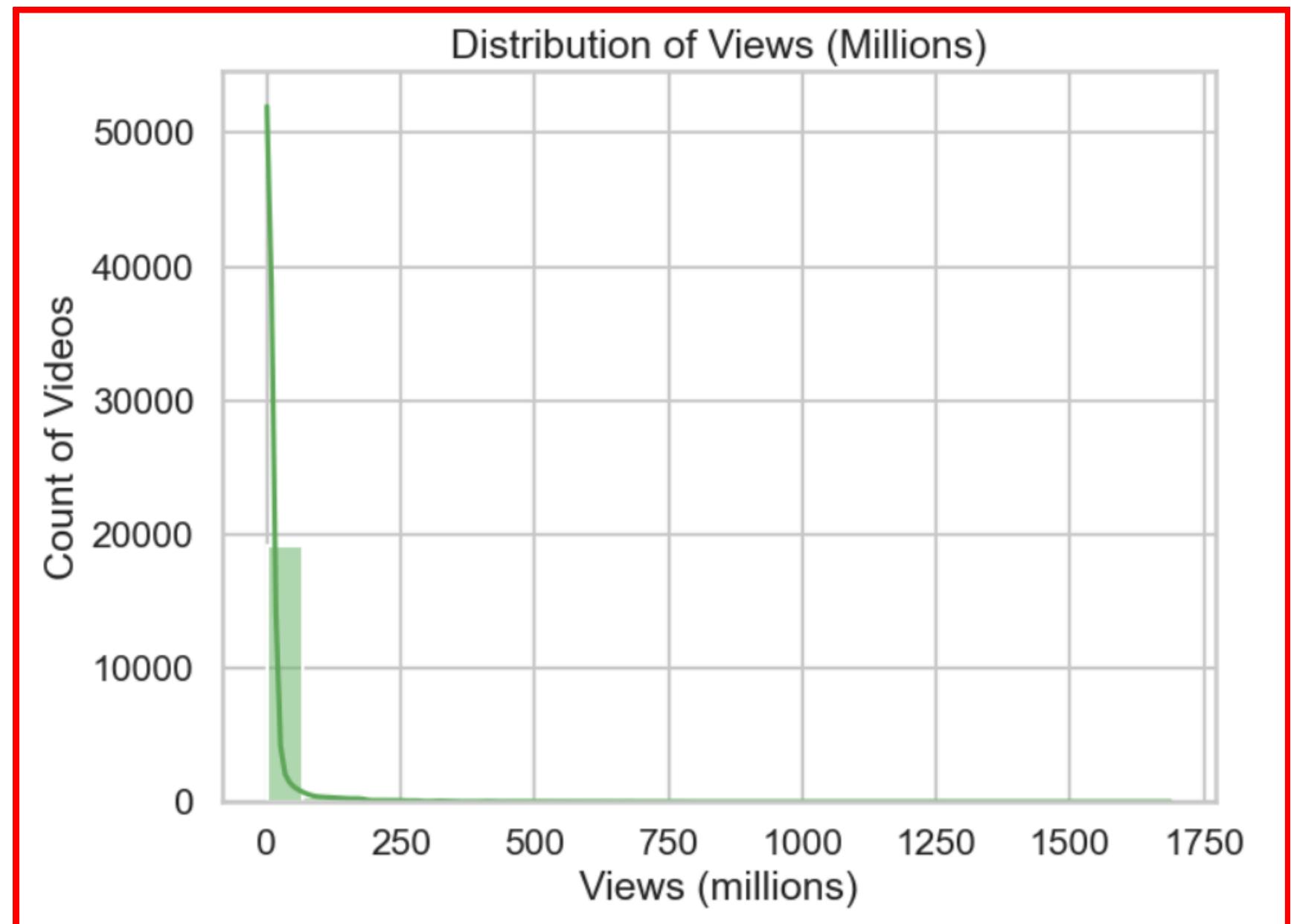
Deep-dive into 'T-Series' channel

You Tube

Exploratory Data Analysis (EDA):

Insights from Analysis:

- Views are extremely right-skewed: the vast majority of videos cluster at low view counts, with a long tail of rare high-view hits.
- A small number of blockbuster videos account for a disproportionate share of total views, while most uploads get modest traction.
- For analysis and modeling, log-scaling or percentile cuts are appropriate to handle the heavy-tailed distribution.



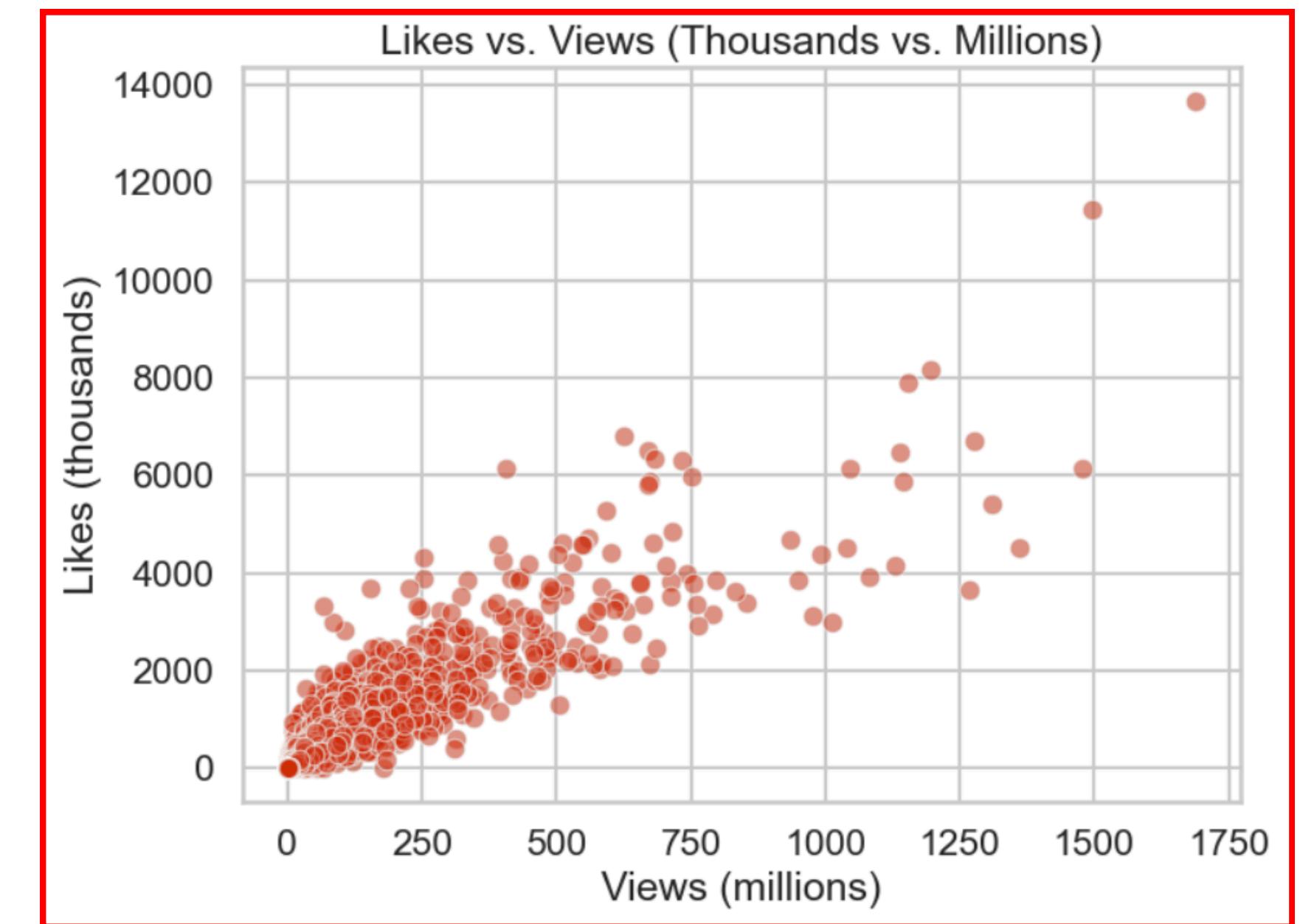
Deep-dive into 'T-Series' channel

You Tube

Exploratory Data Analysis (EDA):

Insights from Analysis:

- Clear positive relationship: more views generally correspond to more likes.
- Wide spread at similar view levels means like-through-rate varies by video (content, audience, timing).
- Several standout points show exceptionally high likes for their view counts, indicating unusually engaging videos.
- Dense cluster at low-to-mid views with modest likes reflects the typical performance band for the channel.

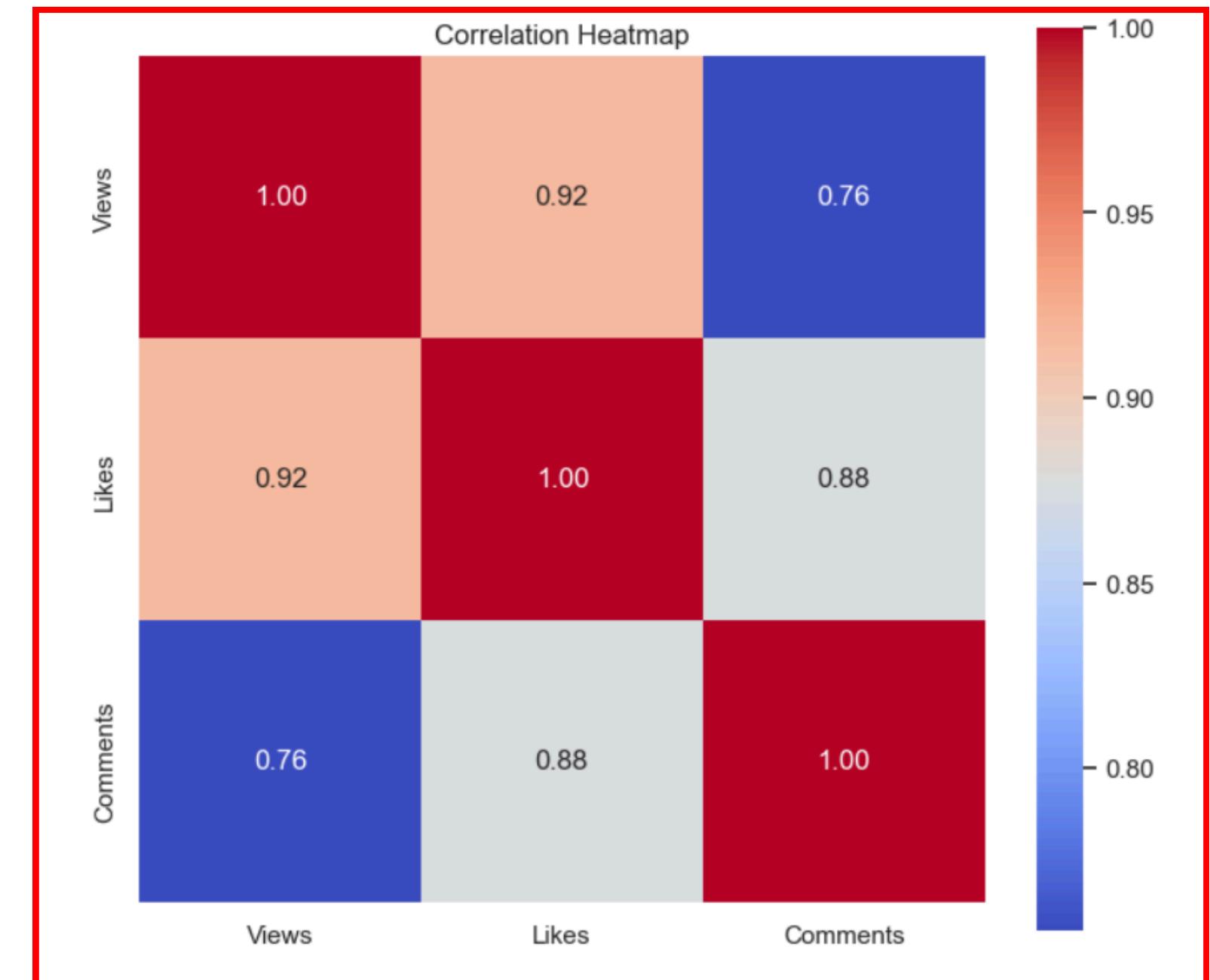


Deep-dive into 'T-Series' channel

Exploratory Data Analysis (EDA):

Insights from Analysis:

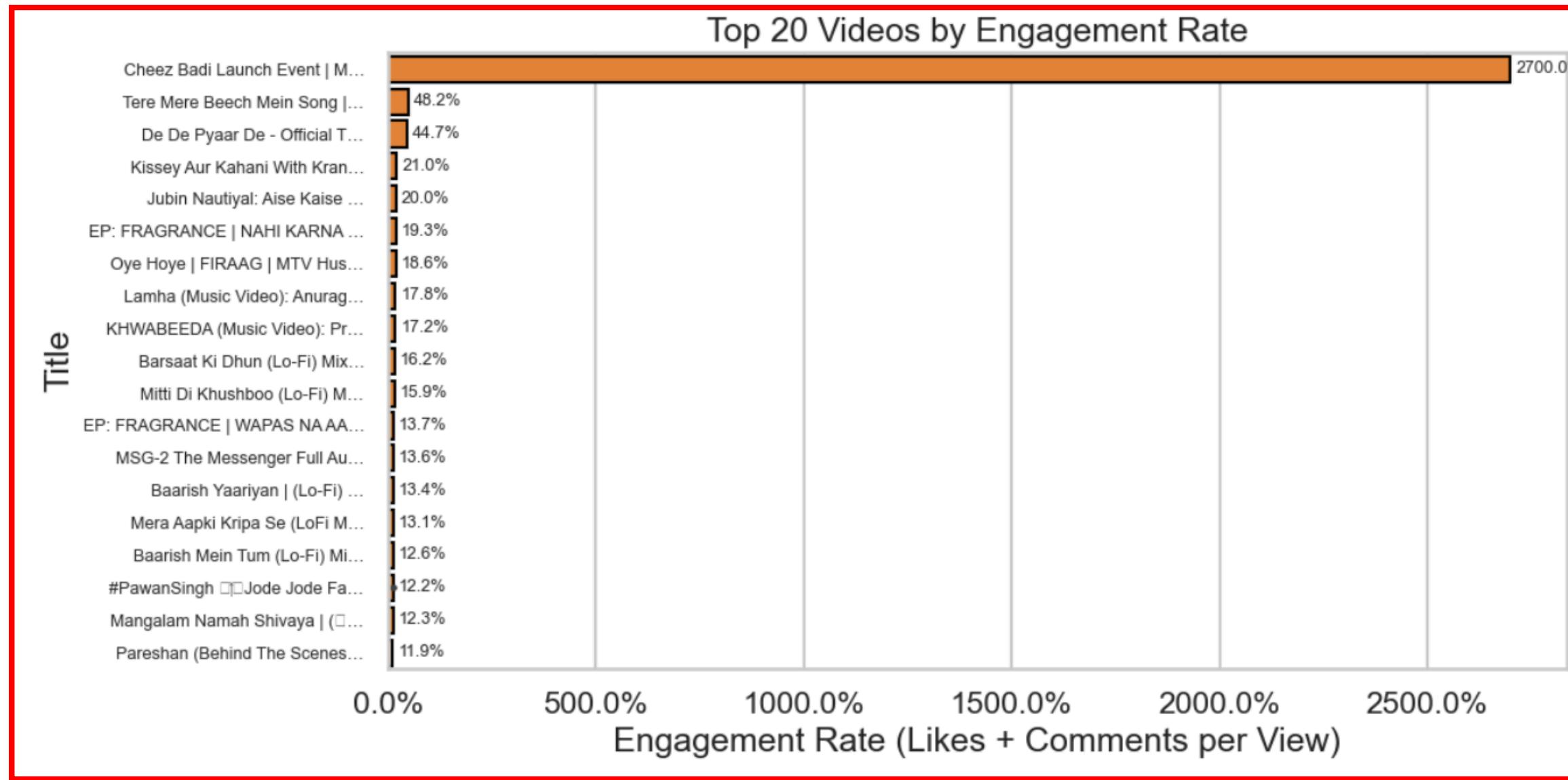
- Views and Likes move almost in lockstep (corr ≈ 0.92), so higher reach strongly aligns with more likes.
- Likes and Comments are very closely related (corr ≈ 0.88), indicating videos that get likes also tend to spark discussion.timing).
- Views and Comments are strongly but comparatively less correlated (corr ≈ 0.76), suggesting comments depend on more than just reach (e.g., content type, prompts).



Deep-dive into 'T-Series' channel

You Tube

Exploratory Data Analysis (EDA):



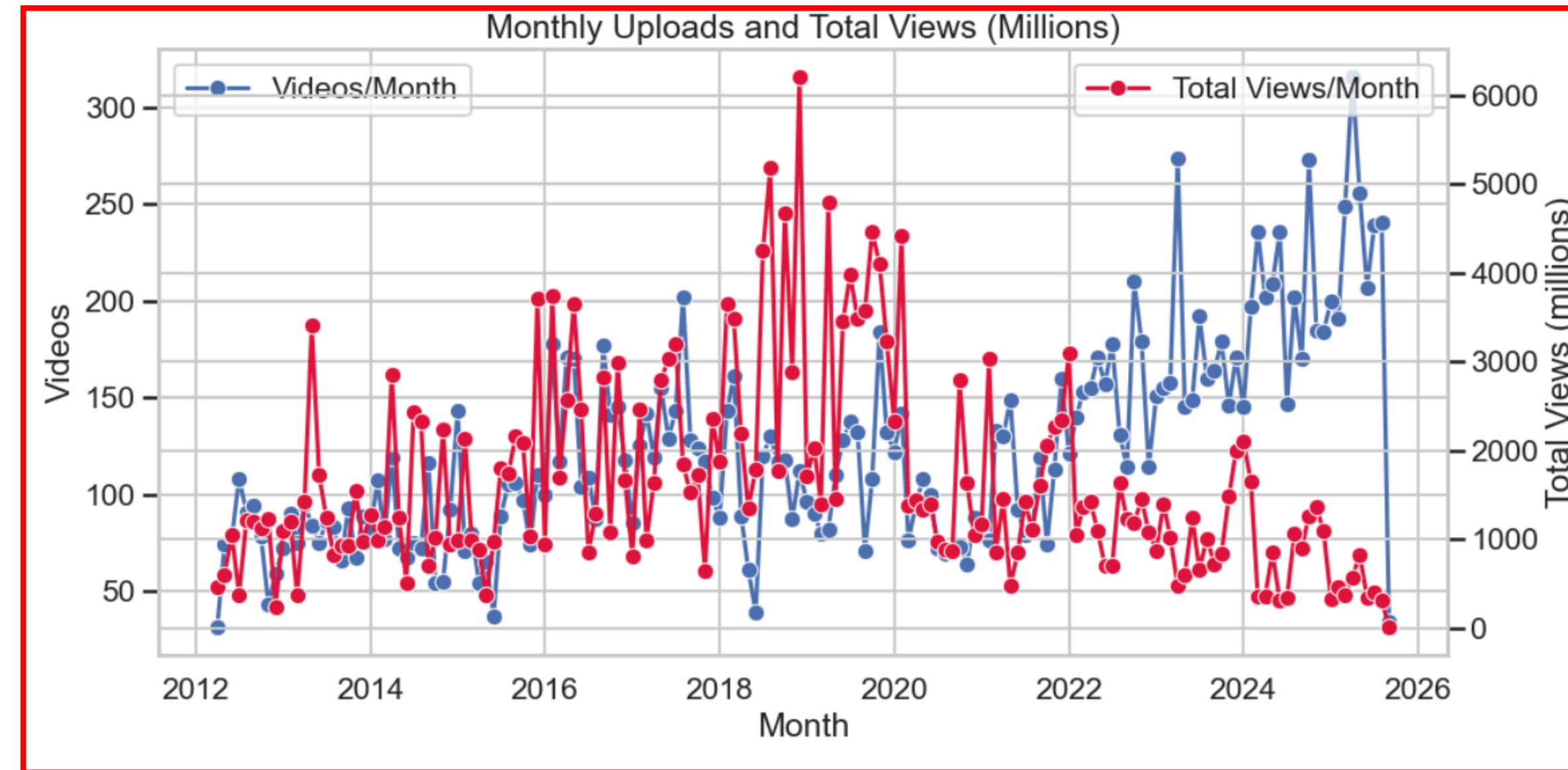
Insights from Analysis:

- One extreme outlier dominates: the top video's engagement rate dwarfs all others, suggesting anomalous counts or uniquely viral behavior.
- The next best group clusters around roughly 12–21% engagement, indicating a more typical high-engagement band for the channel.
- Large gap between rank 1 and the rest implies the leaderboard is top-heavy; small changes won't displace the leader. Validate data for that video (disabled views/like spikes) to ensure the metric isn't distorted.

Deep-dive into 'T-Series' channel

You Tube

Exploratory Data Analysis (EDA):

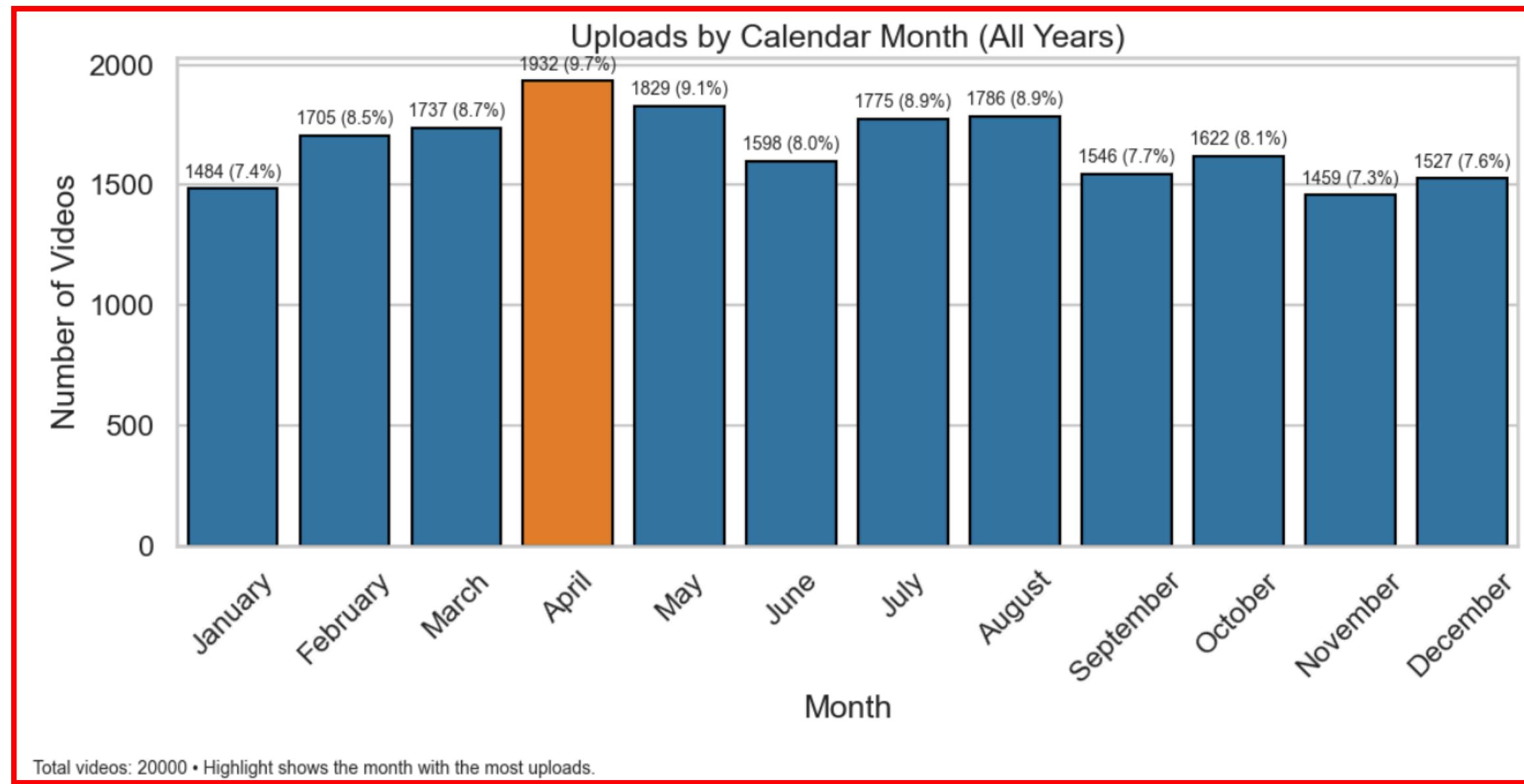


Insights from Analysis:

- Upload volume has climbed steadily over the years, reaching the highest sustained levels in recent periods.
- Total monthly views peaked in the 2018–2020 window, then trended down despite rising upload counts.
- Post-2021, uploads rose while views generally fell, suggesting diminishing returns per video.
- Short spikes in views around 2022 and 2024 didn't persist, indicating brief campaigns or breakout releases rather than a structural recovery.

Deep-dive into 'T-Series' channel

Exploratory Data Analysis (EDA):



Insights from Analysis:

- April has the most uploads; about 9–10% of all videos were published in April.
- Uploads rise from January to April, then stay relatively high from May through August.
- September and November are comparatively lower-upload months, with November among the lowest.
- Despite month-to-month swings, uploads are fairly spread across the year; seasonality is moderate.

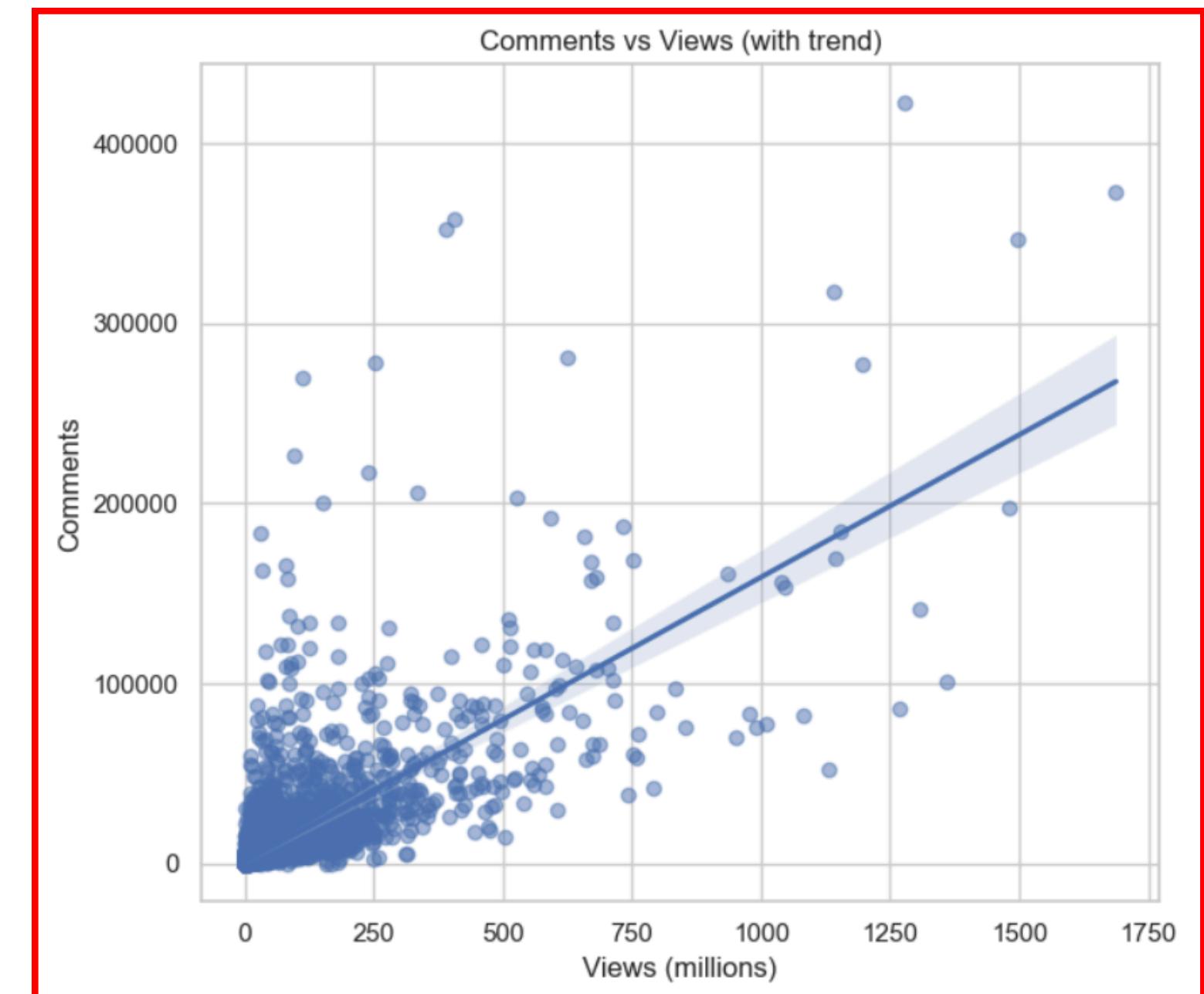
Deep-dive into 'T-Series' channel

You Tube

Exploratory Data Analysis (EDA):

Insights from Analysis:

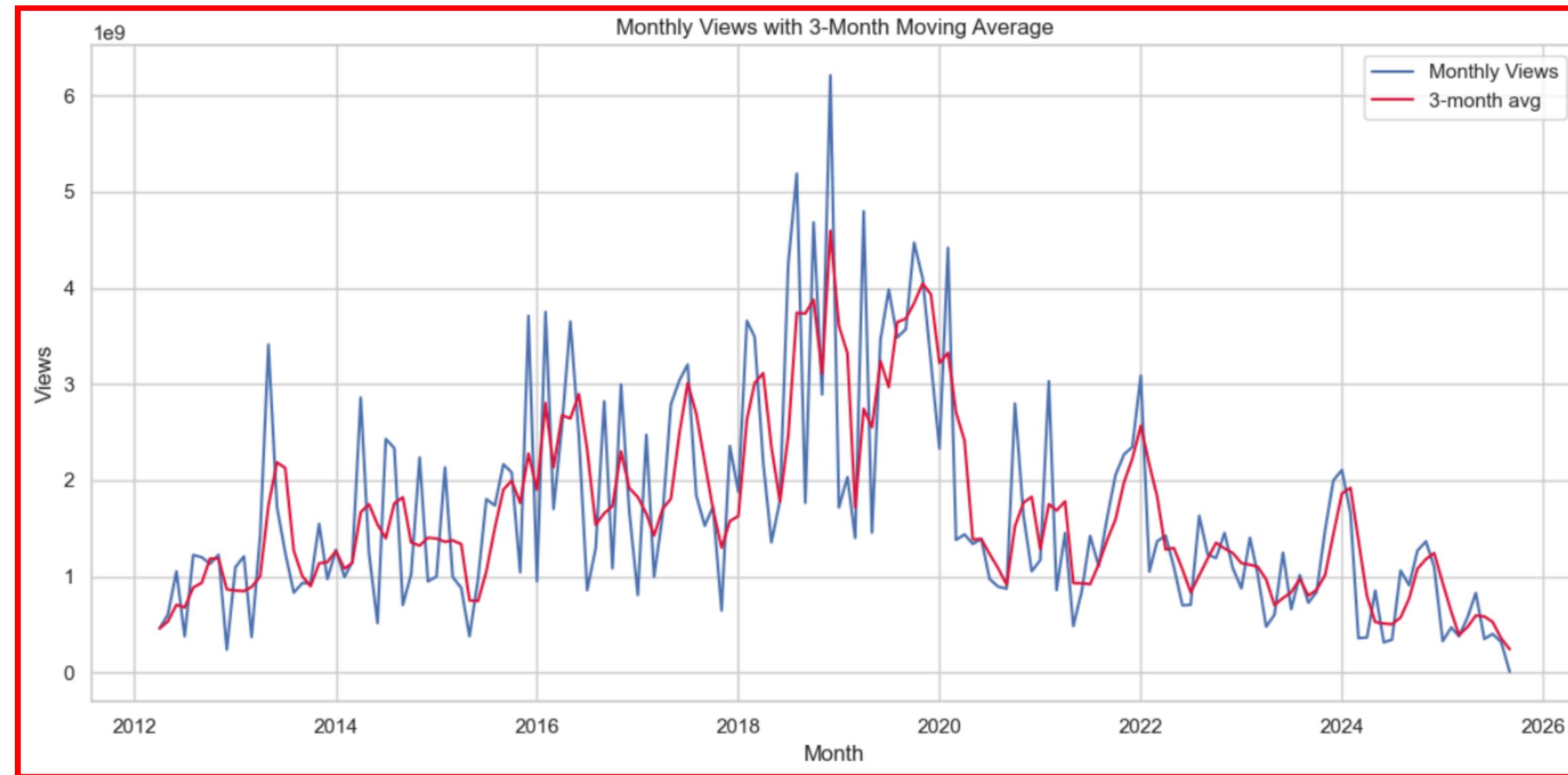
- Comments rise with views; the positive regression slope shows higher-view videos generally attract more comments.
- Wide vertical spread at similar view levels implies large variability in commenting behavior across videos with comparable views.
- Several high-view outliers still have moderate comments, suggesting views don't always convert to discussion equally (content type matters).
- Low-view region is dense with low comments, consistent with typical engagement scaling up only after substantial reach.



Deep-dive into 'T-Series' channel

You Tube

Exploratory Data Analysis (EDA):



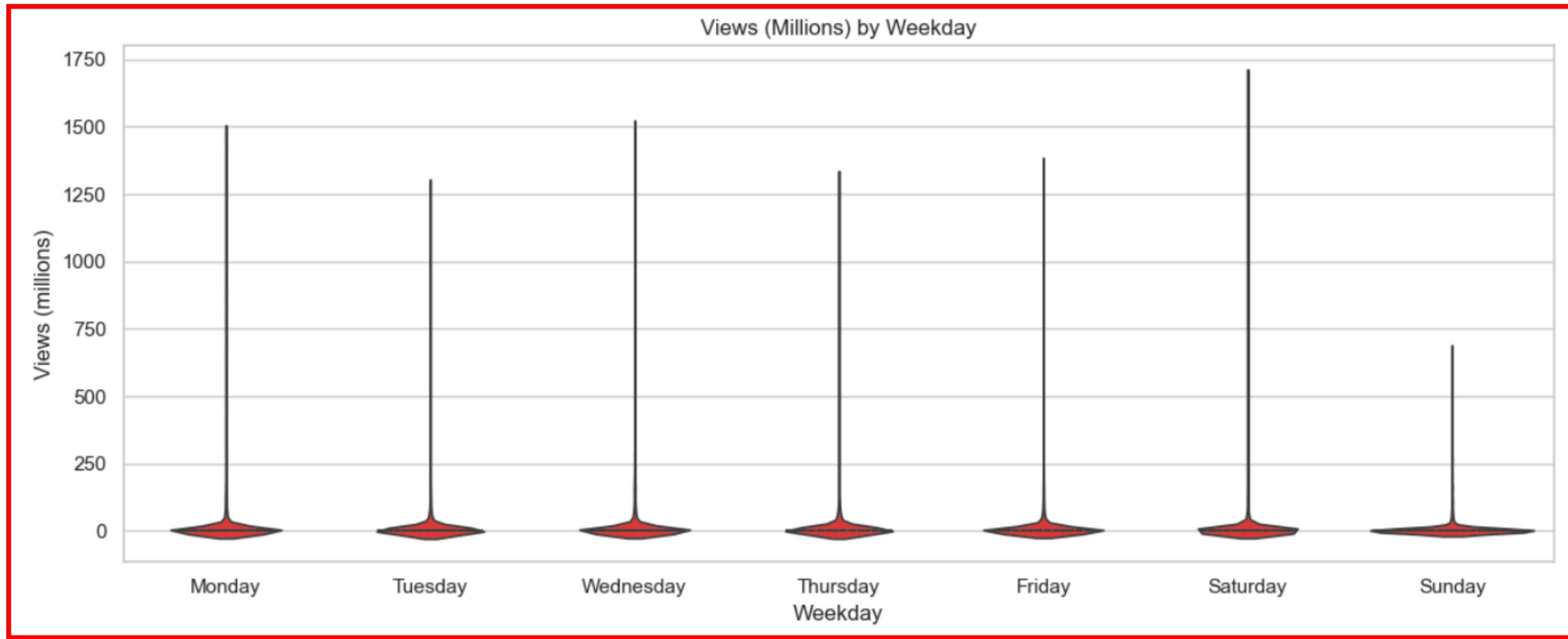
Insights from Analysis:

- 2012–2016: Steady growth in monthly views with periodic spikes; overall upward trend begins.
- 2018–2020: Peak era; multiple months exceed prior highs, with the 3-month average also topping out.
- 2022 and 2024 show short-lived recoveries, but each fades, keeping the longer-term trajectory lower.
- Current period: Views and the moving average are at multi-year lows versus peak years, indicating reduced sustained reach.

Deep-dive into 'T-Series' channel

You Tube

Exploratory Data Analysis (EDA):



Insights from Analysis:

- Views are heavily skewed across all weekdays: most videos get modest views, with a few extreme outliers driving the upper tails.
- Typical performance looks similar Monday–Sunday; the central bands and quartiles cluster low for every day.
- Saturday has the highest single outlier, indicating at least one exceptionally viewed upload on that day.
- Sunday's upper range is comparatively lower, suggesting fewer extreme hits vs. other days in this dataset.

Final Recommendation

You Tube

For Creators

- Focus on consistent uploads to boost engagement.
- Optimize for viewer retention rather than sheer video volume.

For Researchers

- Extend analysis to include comments & sentiment analysis for deeper audience insights.



For Brands/Marketers

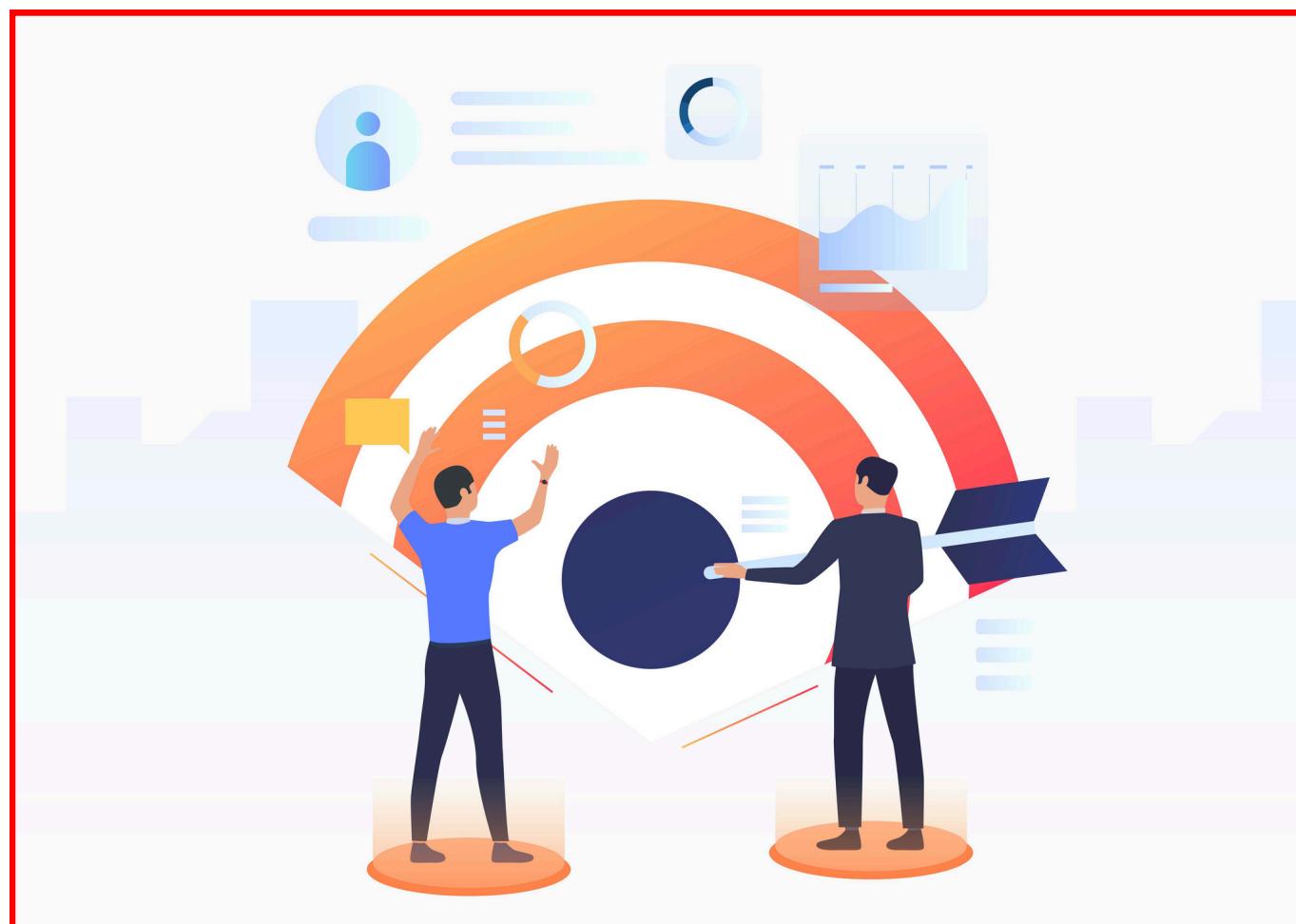
- Partner with channels that show steady subscriber growth + high view count per video.

For Platform Strategy

- Monitor regional content trends to predict emerging markets and audience shifts.

Key Outcomes

You Tube



- Built an end-to-end pipeline for API-based YouTube data extraction.
- Successfully analyzed 15 top channels.
- Derived key growth insights from comparative analysis.
- Conducted a deep dive into 'T-Series', confirming its dominance.

Conclusion

You Tube



Demonstrated the power of Python & APIs for real-world data analysis.



Highlighted trends that explain why certain channels dominate YouTube.



The framework can be extended to automate monitoring and predict channel performance.

Thank You

