



SOC471A

YouTube Algorithm Analysis

Group 10

Pulkit Dhamija (200738)

Rakesh Sankhla (200767)

Ritick Gupta (200801)

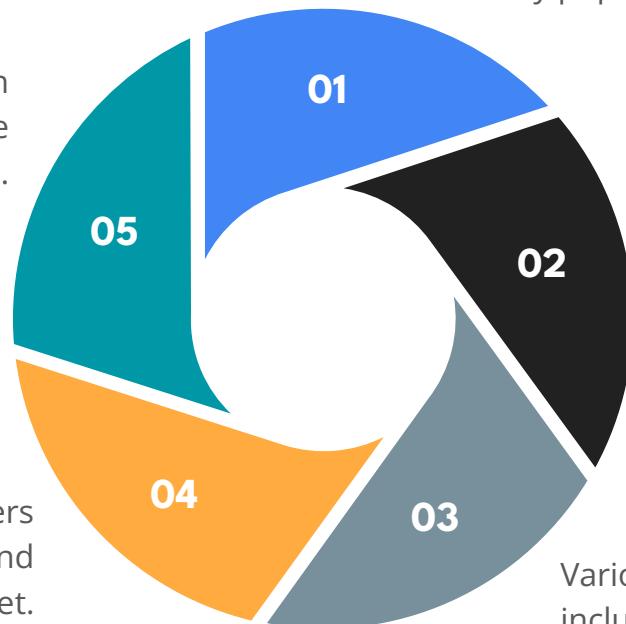
Ritik Raj (200803)

Instructor: Dr. Pradip Swarnakar

Introduction

This project focuses on trending videos on YouTube in India.

The analysis uncovers patterns, preferences, and insights within the dataset.



In the digital era, online video content has become immensely popular.

YouTube is a key player in delivering diverse content to global audiences.

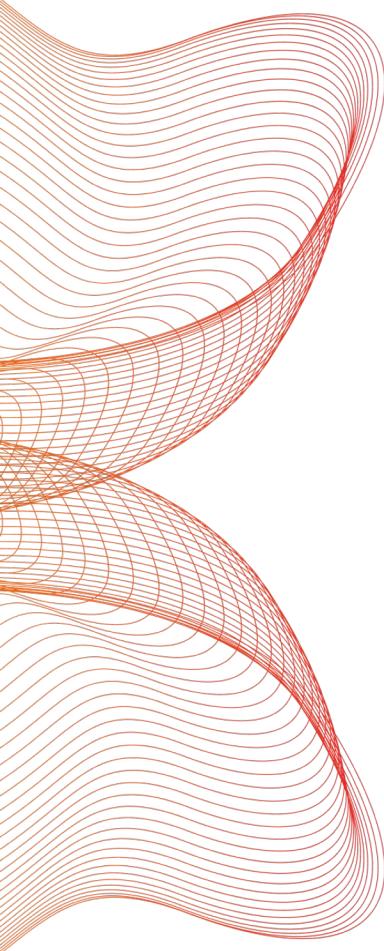
Various aspects are explored, including channel popularity, content categories, viewer engagement metrics, and linguistic patterns in video titles.



Objectives of the Project

- Exploratory Data Analysis (EDA)
- Content Category Analysis
- Channel Metrics
- Engagement Analysis
- Title Analysis
- Time Analysis
- Word Cloud Generation





Benefits

- Provides valuable insights for content creators, marketers, and YouTube enthusiasts
 - Understands trends and patterns of trending videos in the Indian YouTube space
 - Contribute to the broader discourse on digital content consumption
 - Sheds light on the preferences of the YouTube audience in India
- 



Current Practices

- **Consultant based Strategies:**
 - Hiring consultants or taking courses from YouTubers.
 - Non-algorithmic and reliant on personal experience.
 - Limited accessibility due to high costs.
- **Trending Topics and Short Term Gains:**
 - Relying on trending social media topics or news for quick reach.
 - Genre-specific and effective only in the short run.
- **Challenges:**
 - Lack of a concrete methods, genre independent model.
 - Limited use of big data and analytics.
 - Accessibility issues for creators with budget constraints.





Innovative Approach

- **Big Data Analytics for YouTube Growth:**
 - Utilizing the latest YouTube analytical data.
 - Statistical analysis for relevant conclusions.
 - Leveraging major characteristics of big data.
- **Efficiency and Accessibility:**
 - Efficient results through an algorithmic approach.
 - Affordable and accessible to content creators of all sizes.
 - No need for expensive consultants or courses.
- **Long Term Success:**
 - Easily implementable for any content creator.
 - Provides consistent, long-term growth strategies.
 - Accelerates channel reach without depending on short-term trends.





Data Preprocessing for YouTube Trending Dataset

- Dataset contains information on trending videos like title, channel, views, likes, comments, and category.
- Convert date columns to datetime objects.
- Map category IDs to corresponding labels for better interpretability.
- Convert tags into lists and Handle special cases like "[none]."
- Clean and process video titles. Remove unnecessary words, special characters.



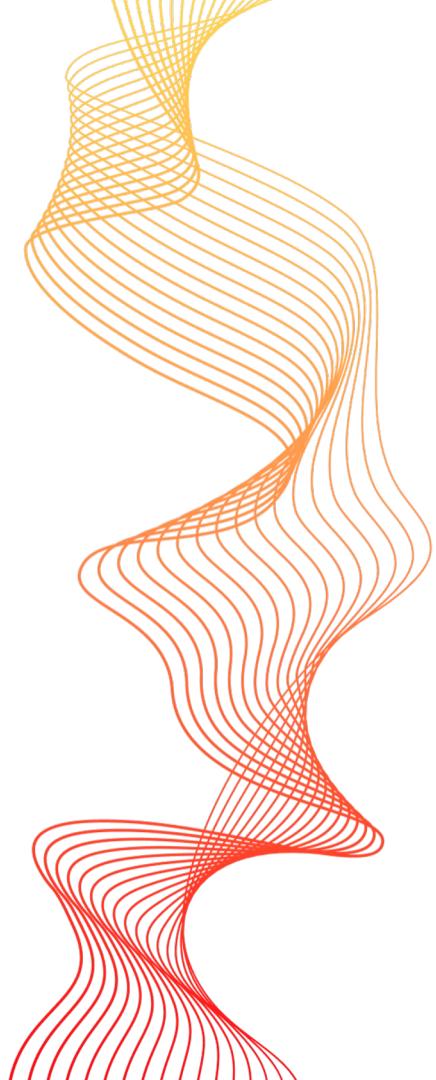
Exploratory Data Analysis (EDA) for YouTube Trending Dataset

- **Top Viewed Channels:** Visualize the top 20 most viewed channels.
- **Category ID Mapping:** Explore distribution across different content categories.
- **Video Titles Analysis:** Indepth exploration of word choice and frequently used words.
- **Views, Likes, and Comments:** Visualize distributions and identify patterns in viewer engagement.
- **Time Based Analysis:** Examine publishing days for temporal trends.
- **Word Cloud Generation:** Visually represent common words in video titles.

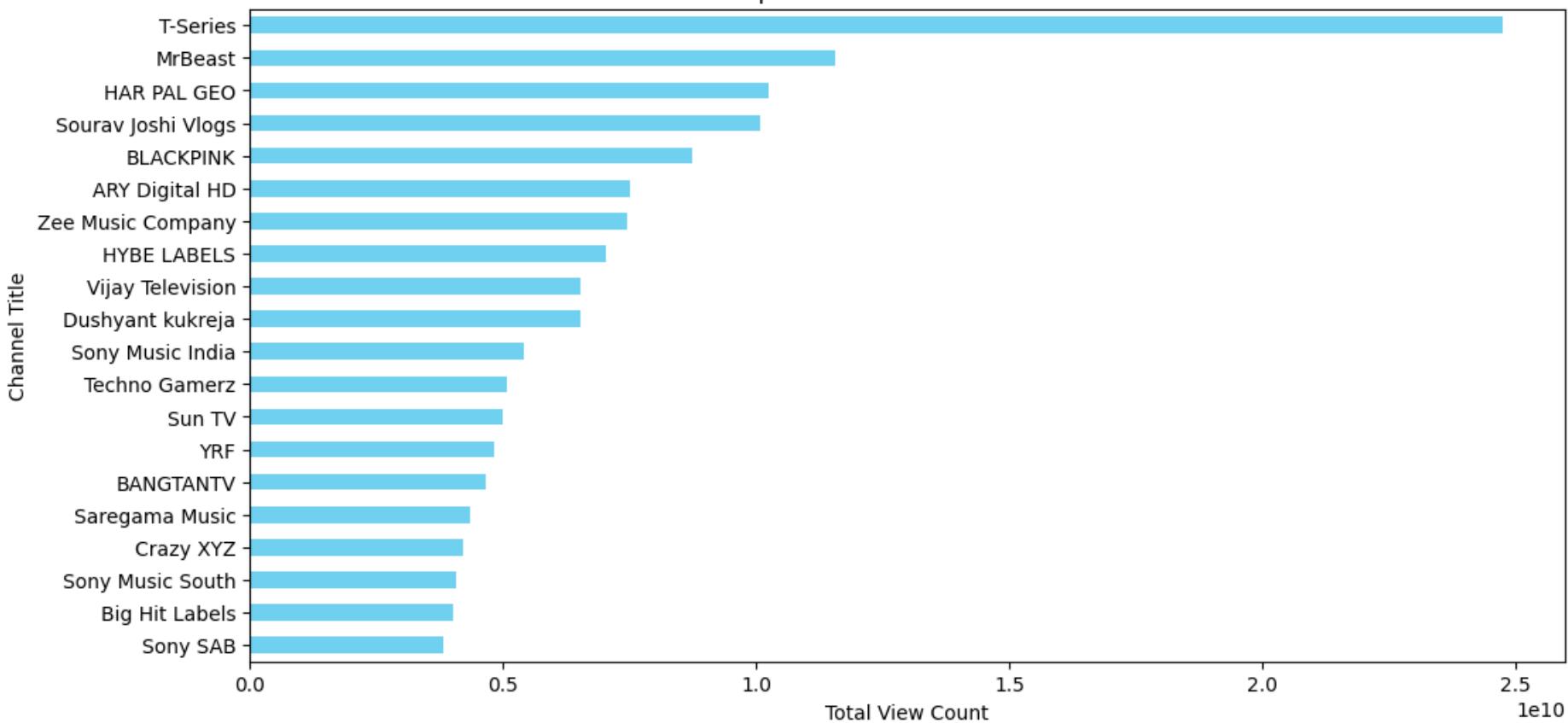




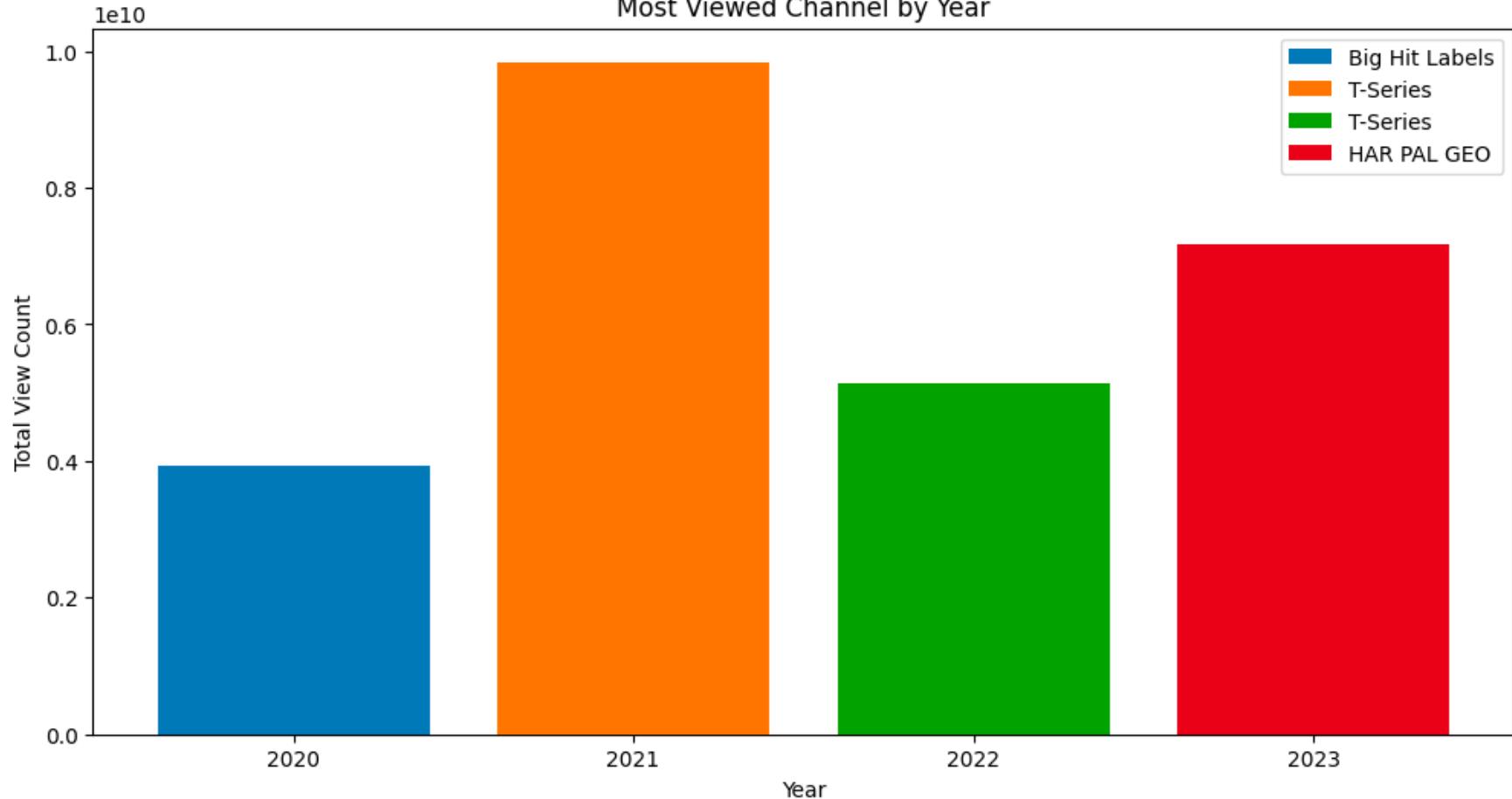
Most Viewed Videos



Top 20 Most Viewed Channels



Most Viewed Channel by Year





Occurrence of Words in Titles

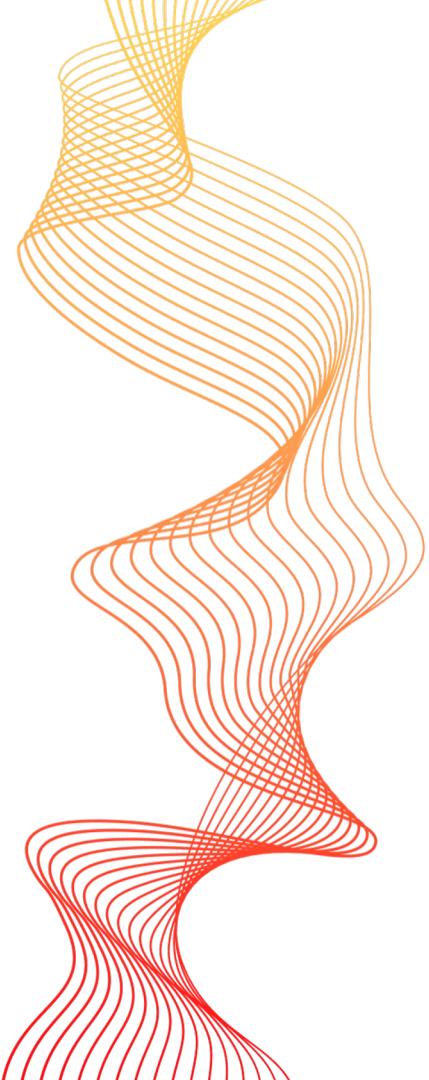
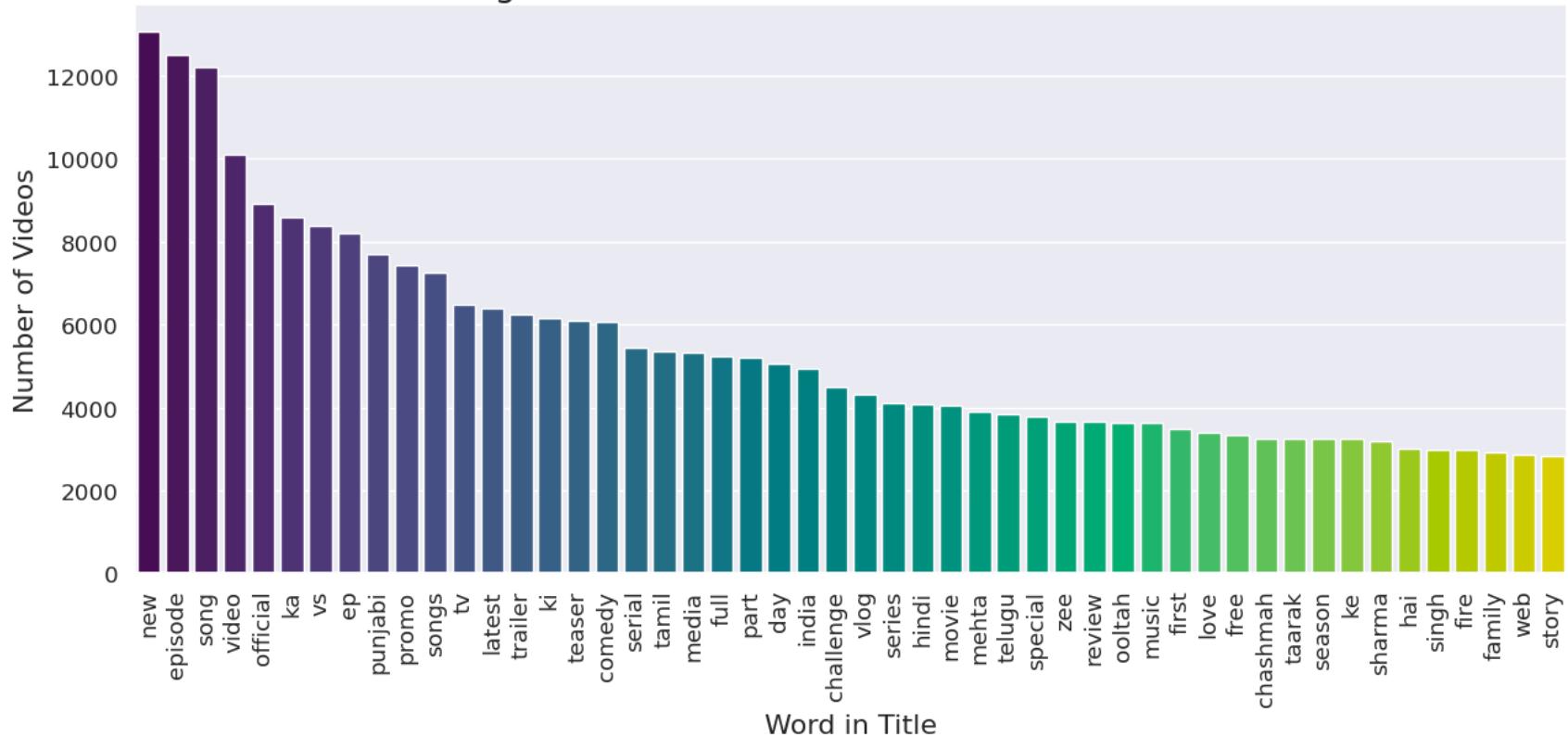


Figure 1: Occurrence of Words in Video Titles



Word Cloud

zee live infinitum kundali friend
video interview comedy ki tv diwali haryanvi krishna
kyah ho etv mr 1st hour tamada exam love
malayalam film free dance rahul singh moment food time
ravi promo april prank studio gta car may last web
akhtar event ipl premiere bhai sister real show
hindu home full girl july story pyaar
year sir fire team vlog finally watch raja khan
unboxing music scene making chashmah ali sad
teaser first ne ke sun indian star bigg game drama
part one update babu original test gift house emotional
unboxing family wedding marriage sun indian star bigg game drama
tamil middle celebration end garena back happy
boss digital sharma oct preview aur bhagya
teaser official sidhu september tere recipe pro haryanvi
trailer serie mehta exclusive
episode shankar surprise funny boy holi
presented 2nd dil talk big yadav
digitally稀有
final tour match



Most frequently used Tags

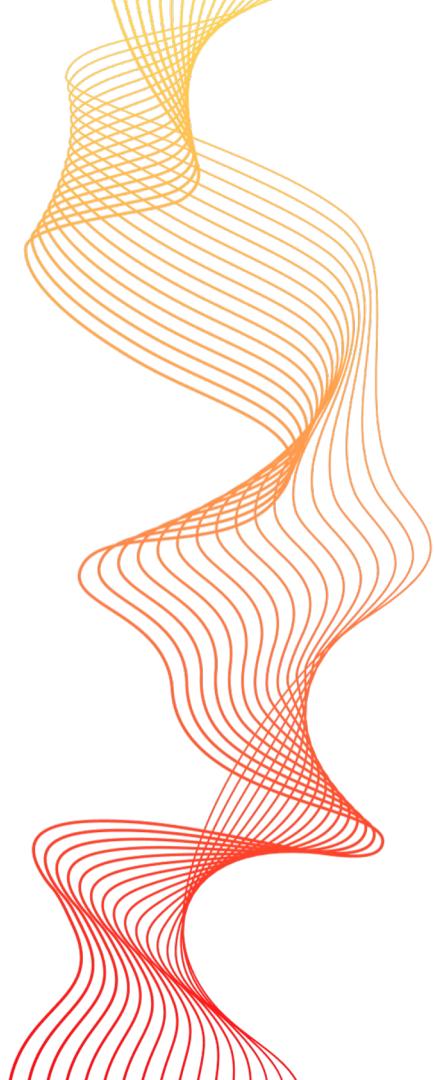
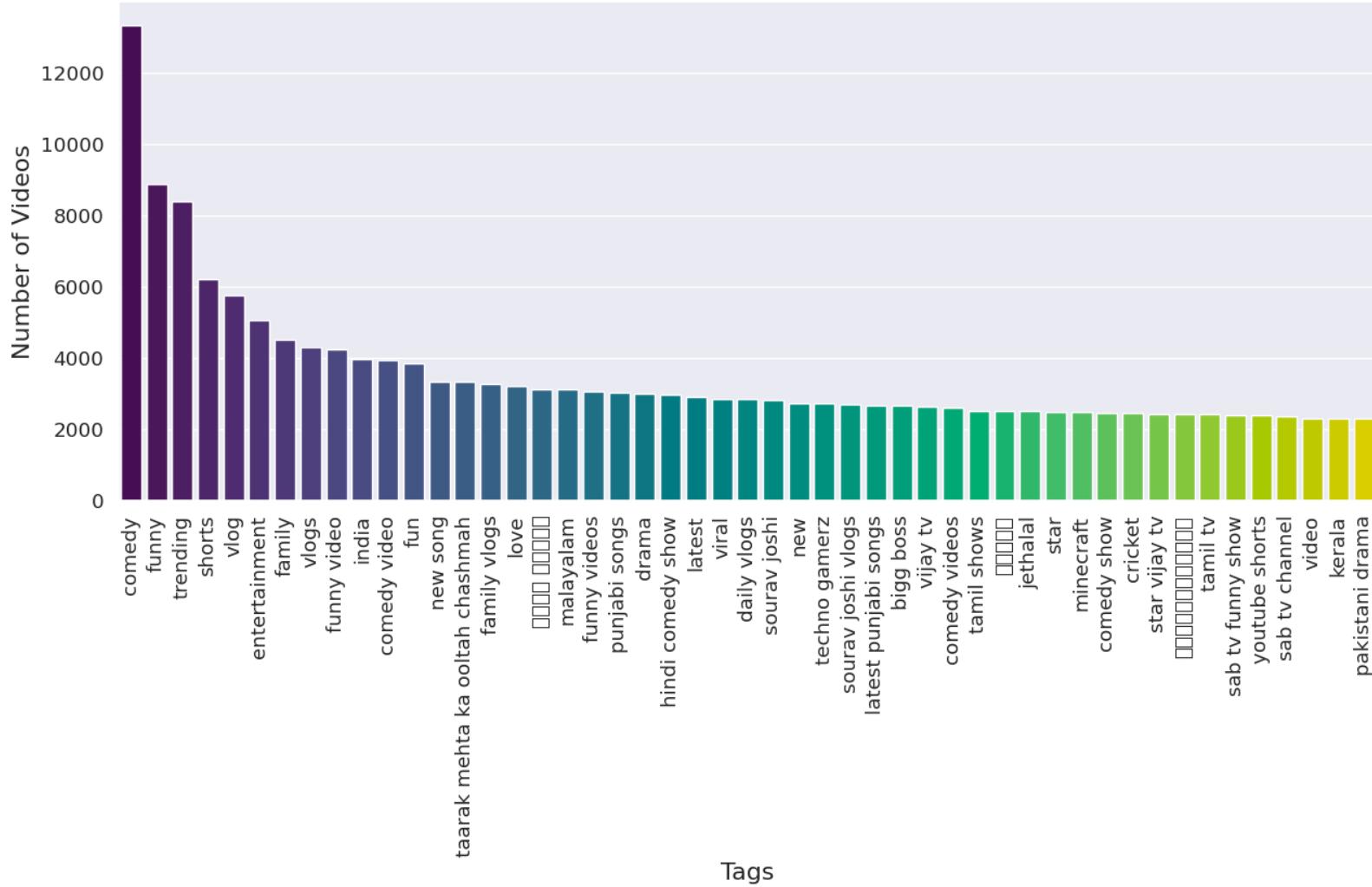


Figure 1: Most frequent tags used

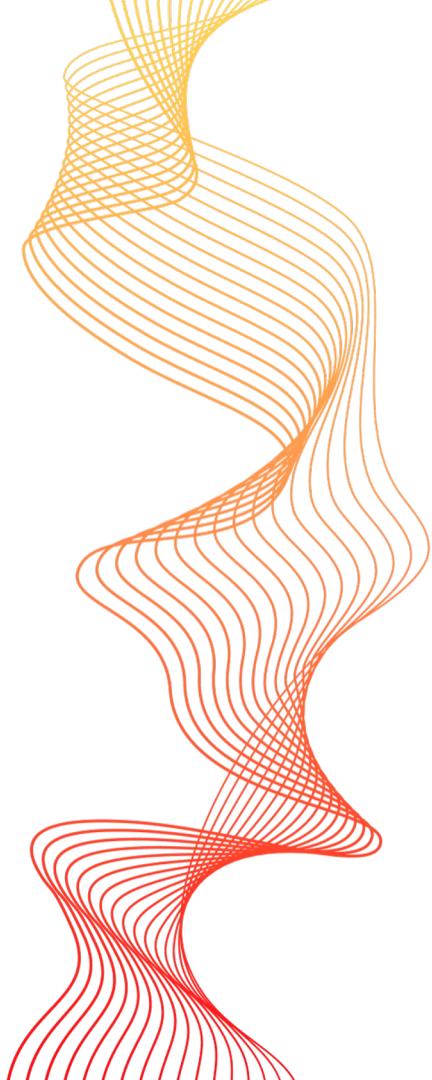


Word Cloud



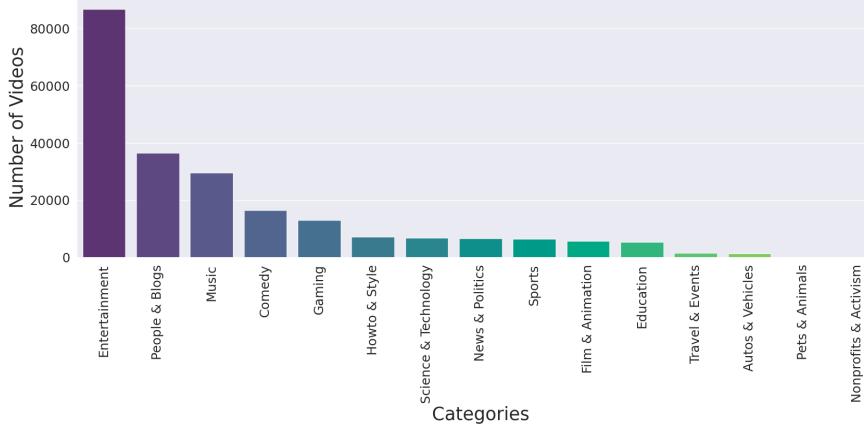


Analysis of Trending Videos



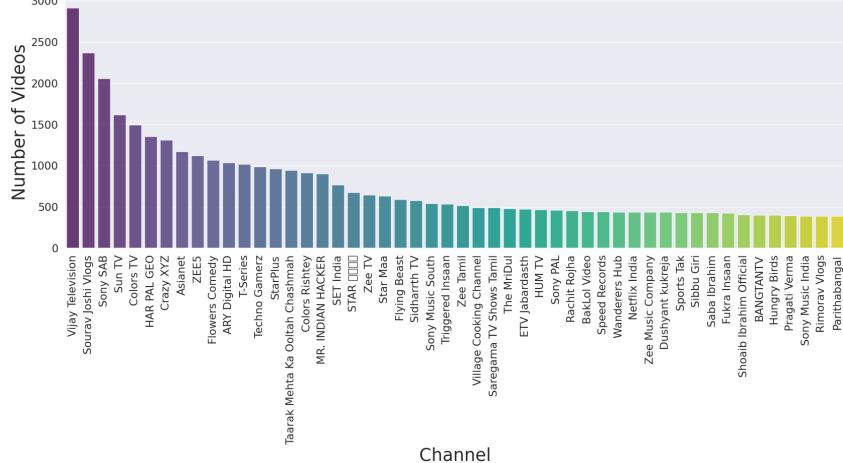
Trending Videos by Category

Figure 2: Trending Videos by Category



Trending Videos by Channel

Figure 3: Trending Videos by Channel



- Videos of Entertainment category dominates the top trending videos

- T-series does have a large number of videos (~1000) on the trending page

Figure 4: Categories of Top 10 Channels with Most Trending Videos

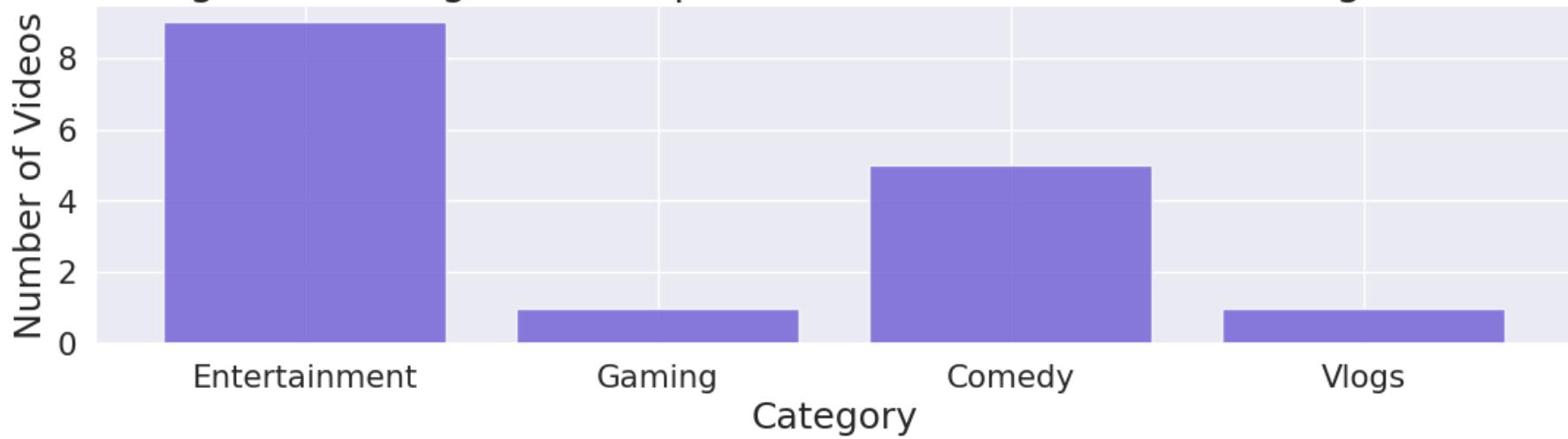
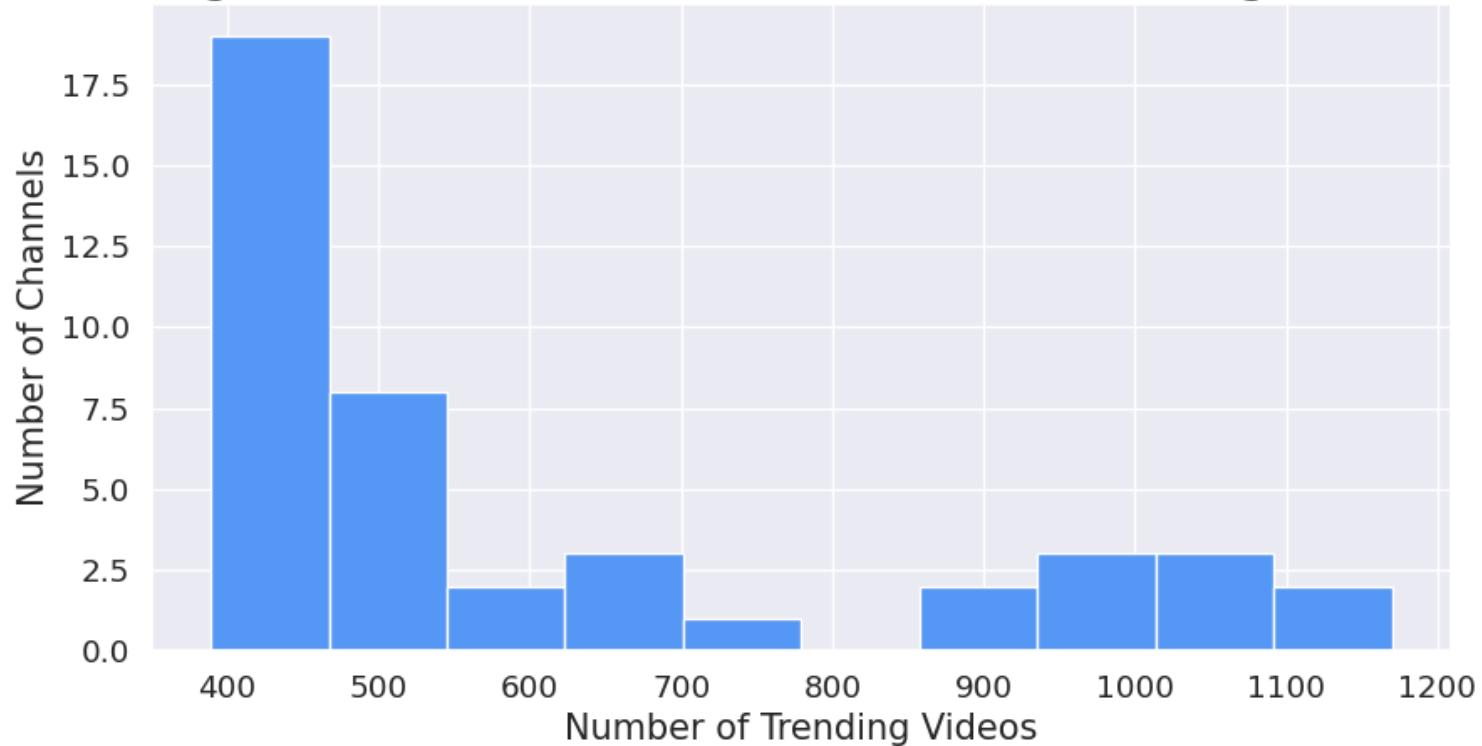


Figure 6: Channels with Less than 700 Trending Videos





Exploration of Views

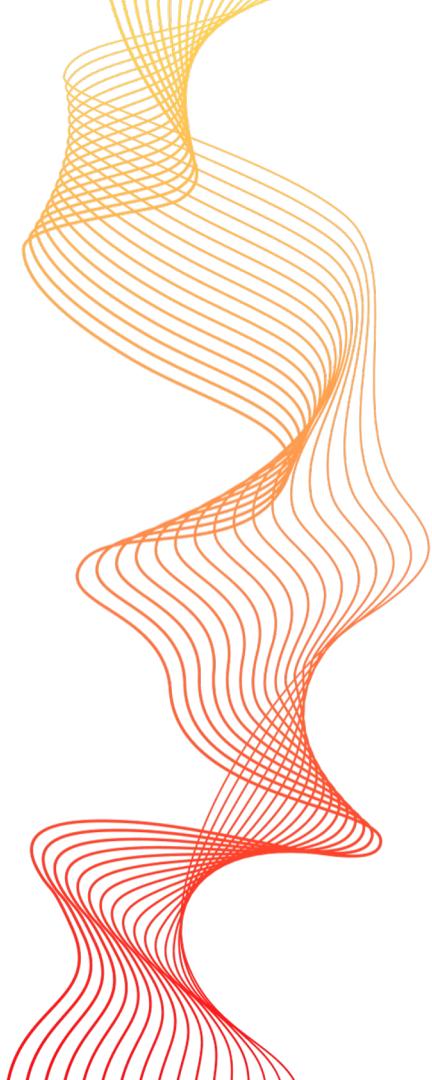


Figure 7: Views of Trending Videos

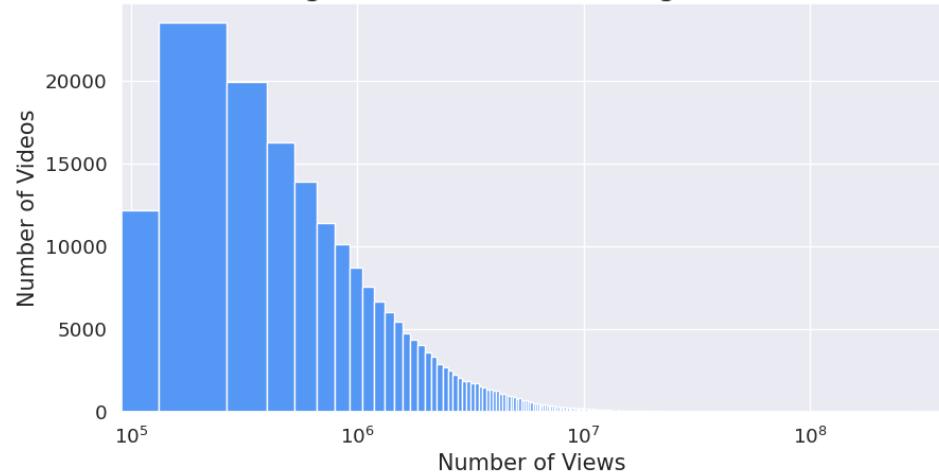
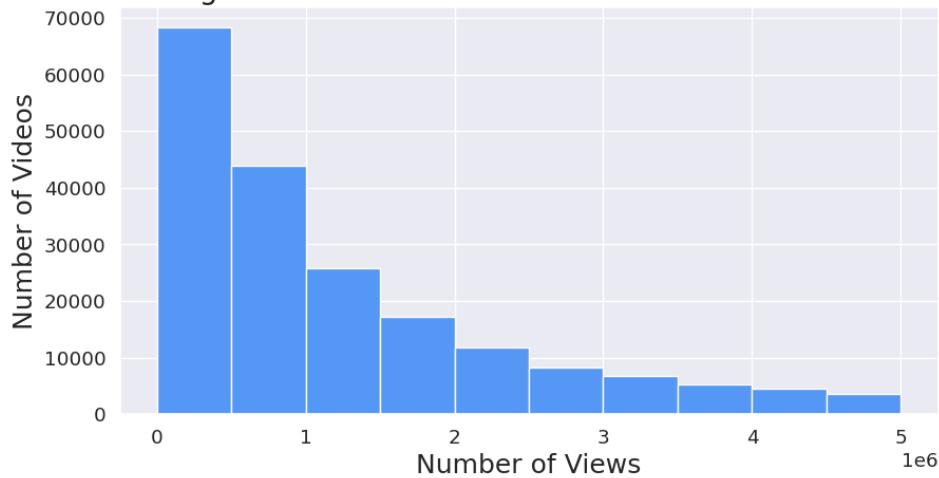


Figure 8: Videos with Fewer than 5 Million Views



- Average number of views: **2885734**
- Median number of views: **991707**
- Percentage of videos under 5 million views: **87.53%**
- Percentage of videos under 1 million views: **50.24%**



Exploration of Likes

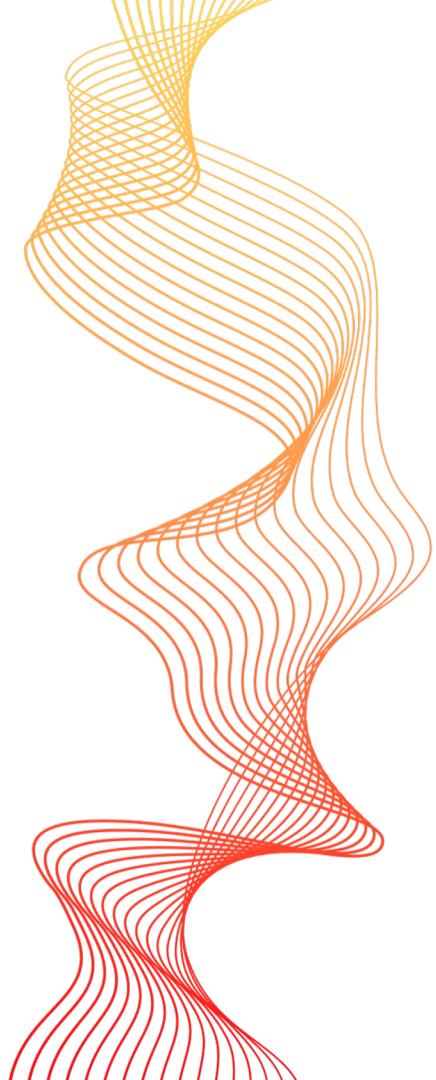


Figure 9: Total Likes for Trending Videos

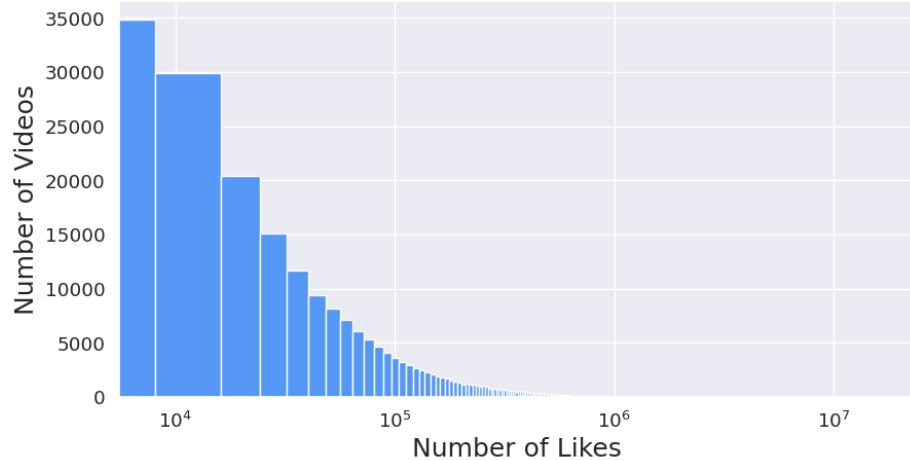
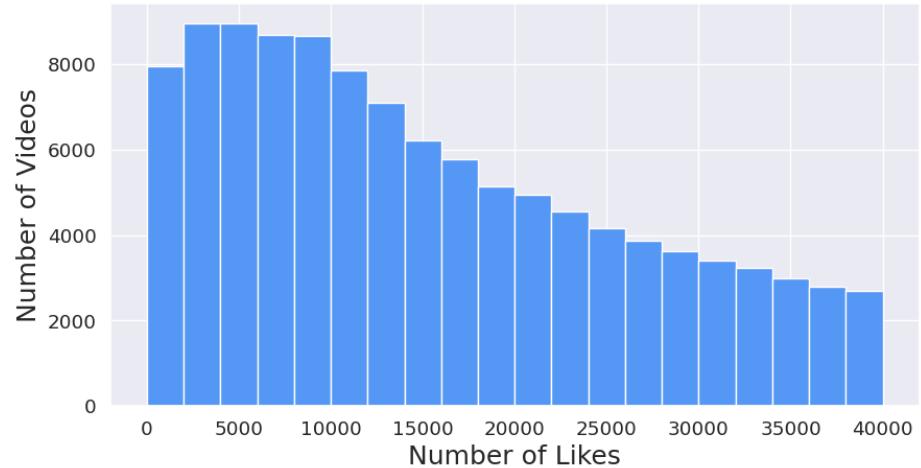


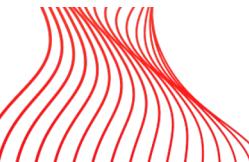
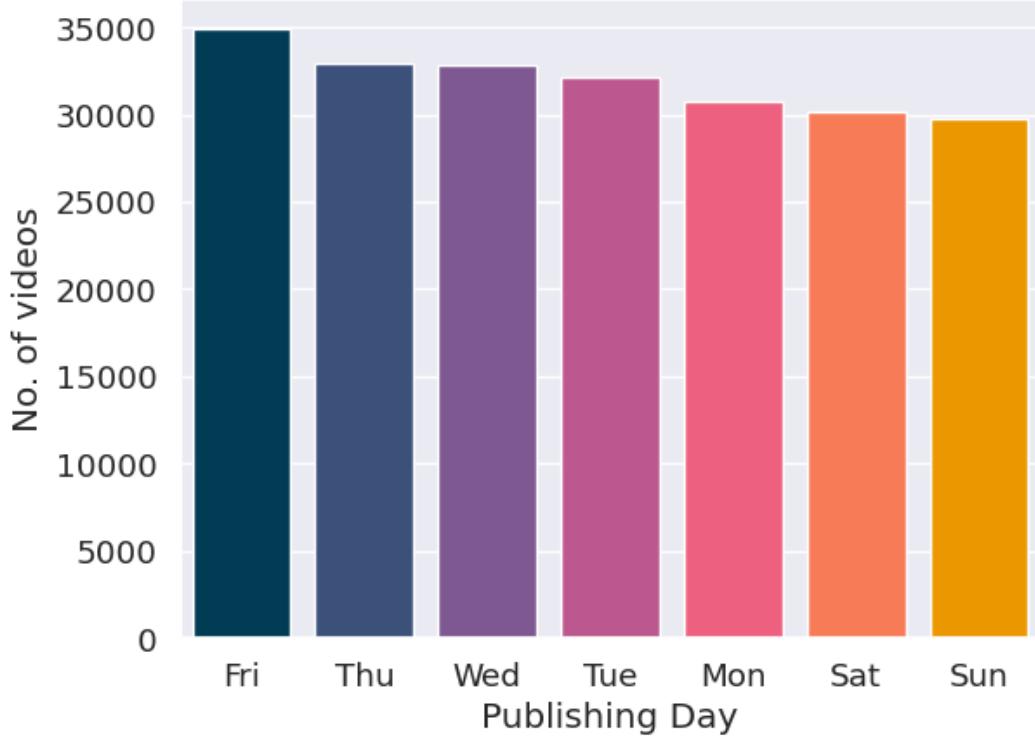
Figure 10: Total Likes for Trending Videos Below 40,000 Likes



- Average number of likes: **146219**
- Median number of likes: **40163**
- This tells us that the distribution of views is strongly skewed to the right, with the majority of videos under **40,613** likes



Publishing Day Analysis





Channel-wise Comparison

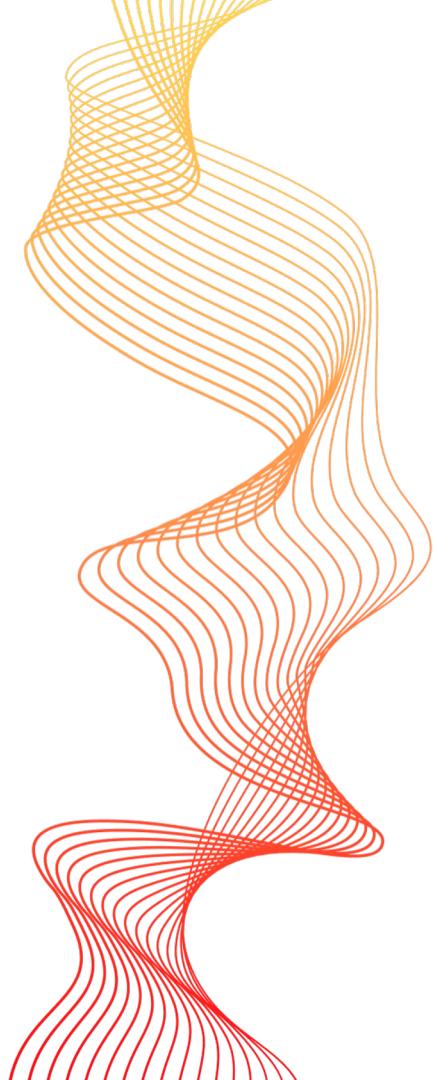
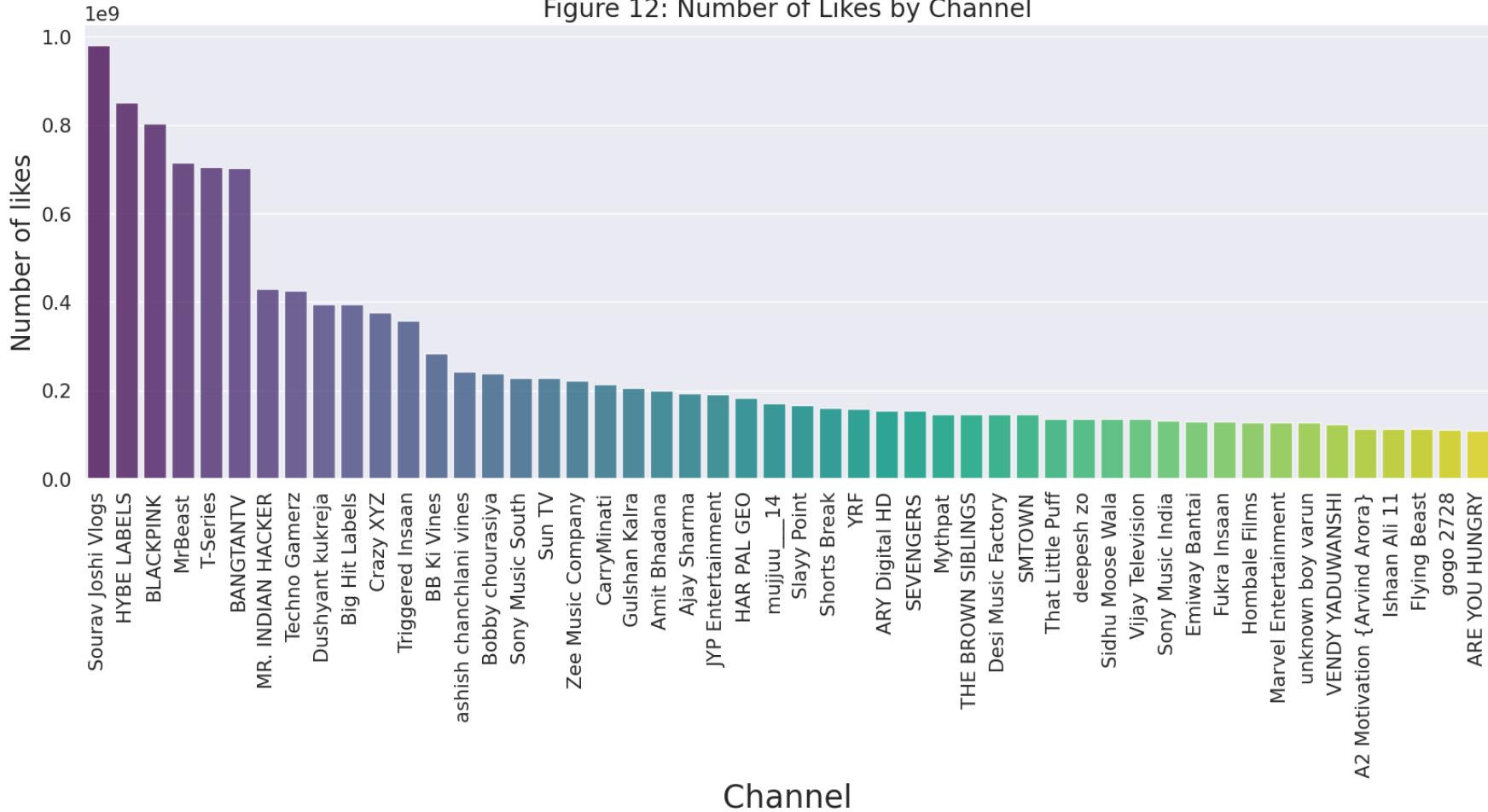


Figure 12: Number of Likes by Channel



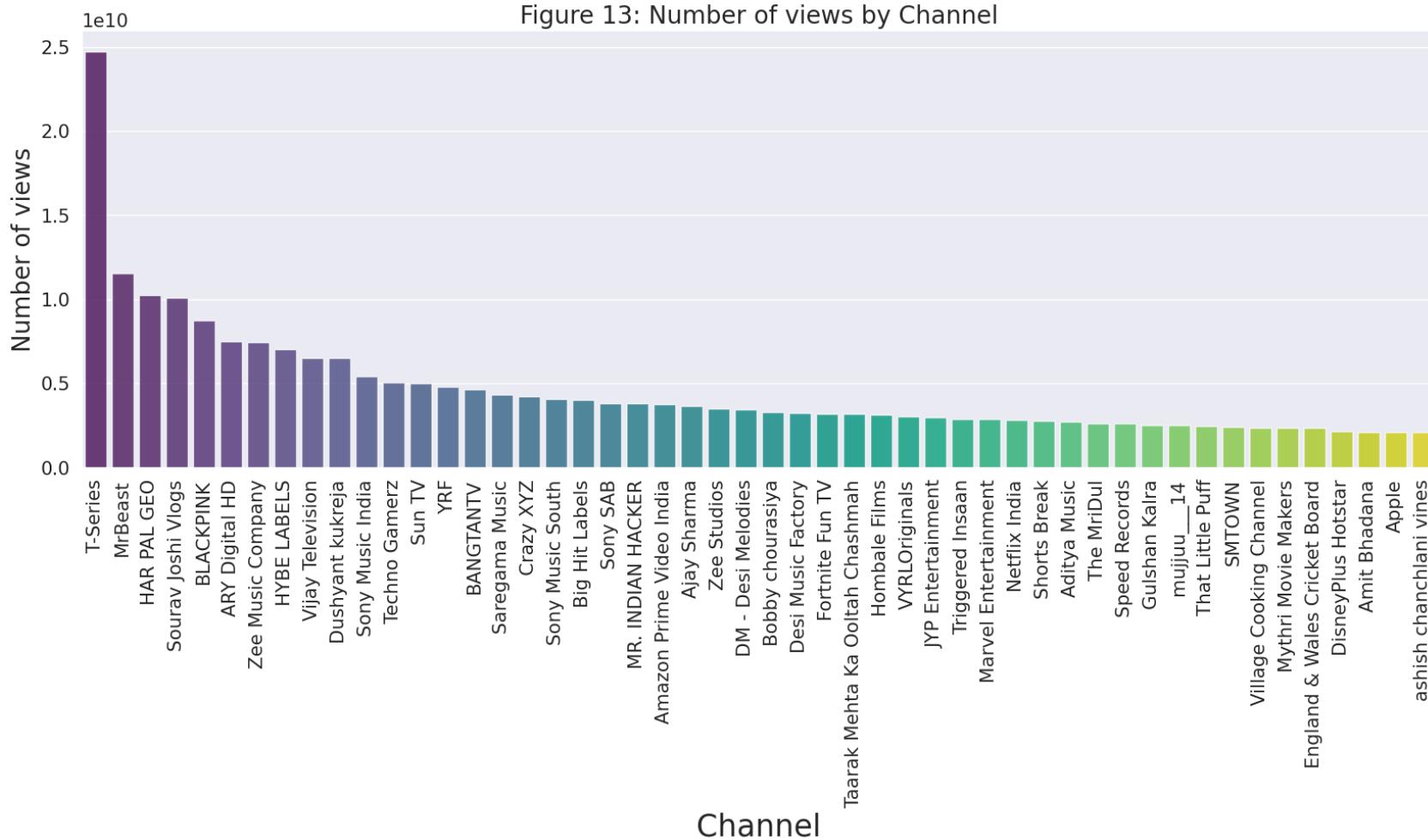
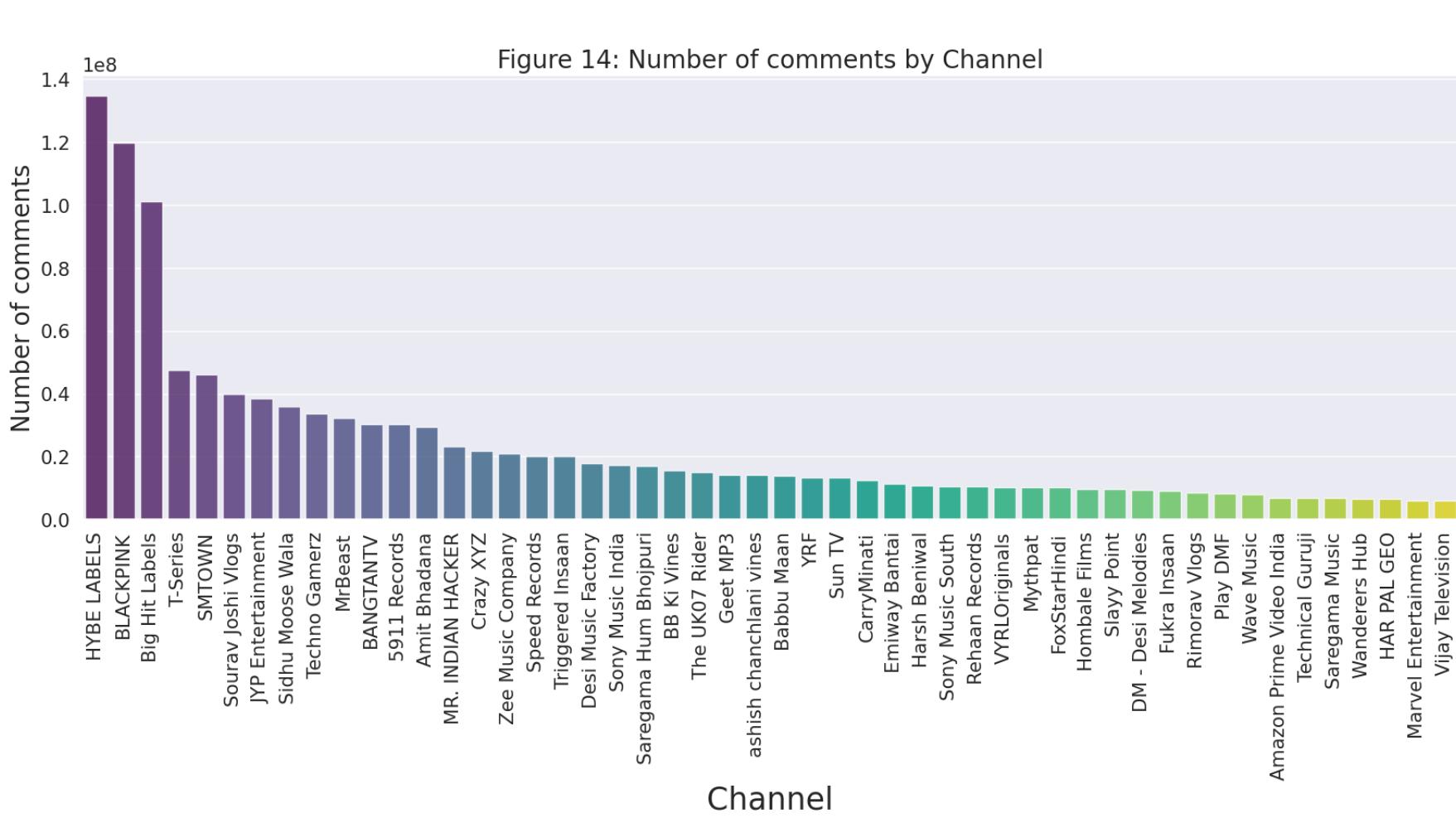
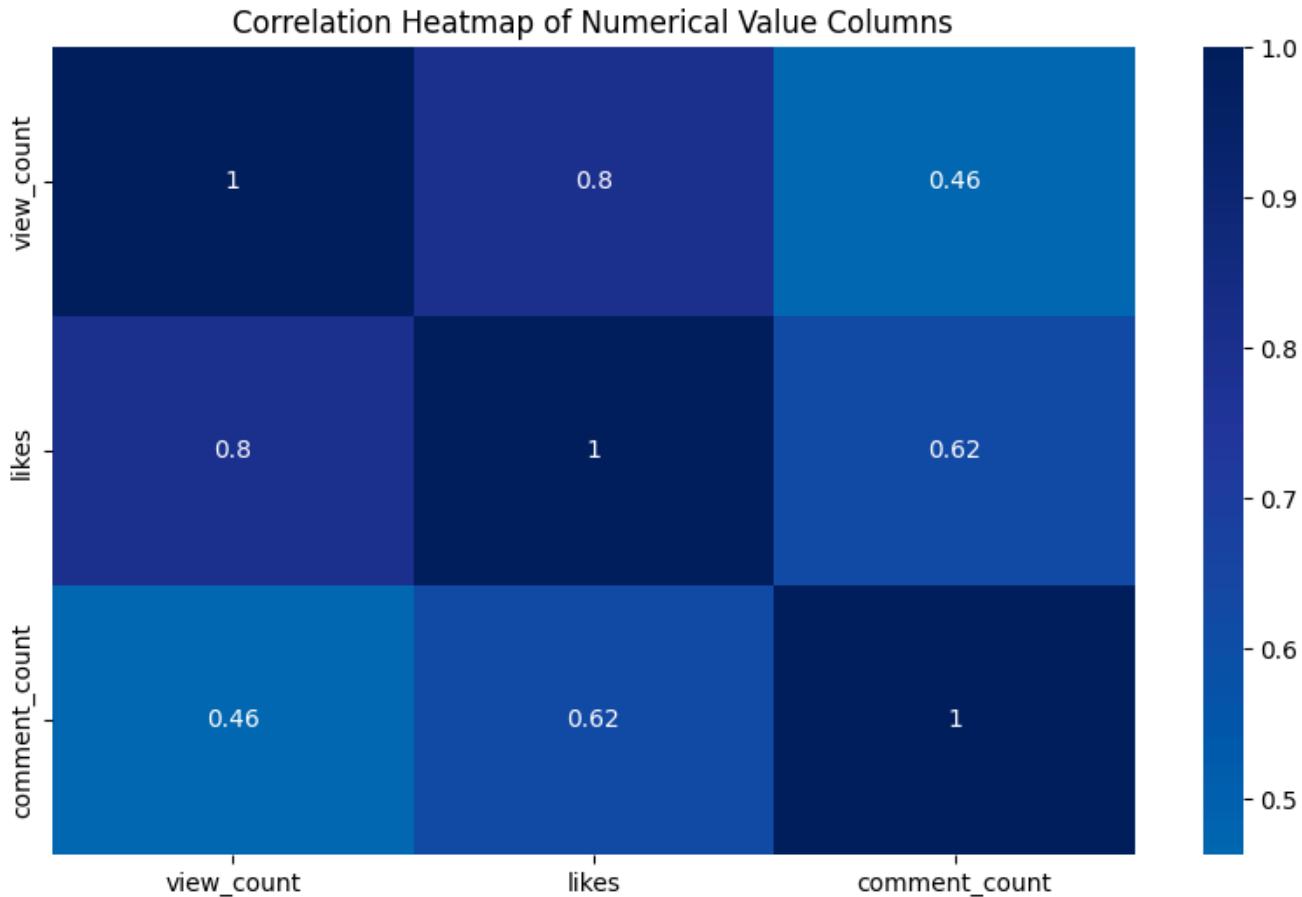


Figure 14: Number of comments by Channel



Analysis



Inference:

- Views Count and Likes have the **max correlation of 0.8**
- Views Count and Comments Count have **least correlation of 0.46**
- Likes and Comments Count have correlation of **0.62**



Conclusion

The analysis of the YouTube algorithm for trending videos in India provided valuable insights into various aspects of content creation and viewer engagement.

Here are some key findings and conclusions from the project:

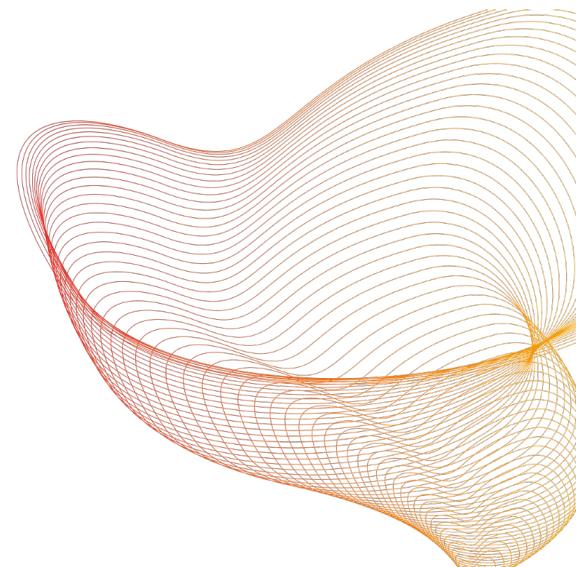
- Top Viewed Channels
- Word Usage in Titles
- Tag Analysis
- Category Dominance
- Channel Performance vs. Views
- Distribution of Views, Likes, and Comments
- Day of Week for Video Release
- Correlation Analysis





Remarks

- This YouTube algorithm analysis provides a comprehensive understanding of the factors influencing video trends in India.
- Content creators, marketers, researchers, educators, and viewers can benefit from these insights to optimise their strategies, create engaging content, and navigate the dynamic landscape of YouTube.
- It is crucial to note that ongoing monitoring and adaptation to algorithm changes are necessary for sustained success in the ever-evolving digital space.





Submitted By:



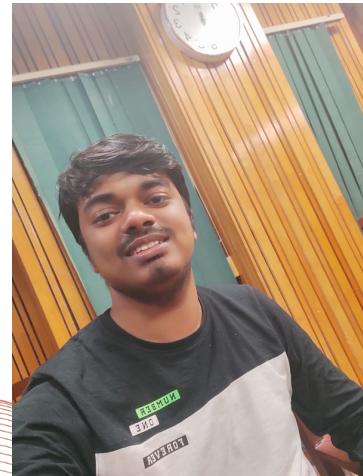
Pulkit Dhamija
(200738)



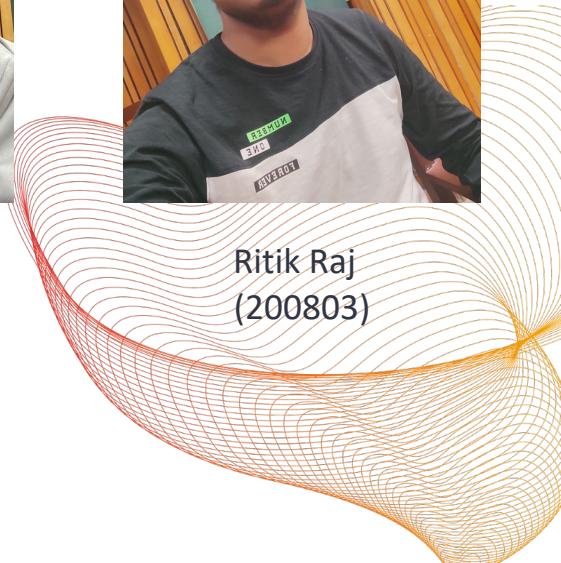
Rakesh Sankhla
(200767)



Ritick Gupta
(200801)



Ritik Raj
(200803)





Thank You

