What Makes a YouTube Video Trending?

A Data-Driven Analysis on the Factors Behind Trending Videos

Saksham Ghimire
Kathmandu Model Secondary School, Bagbazar, Kathmandu
sakshamghimire05@gmail.com

Abstract

In an age where digital content is consumed on an unprecedented scale, YouTube remains one of the most influential platforms for user-generated media. This study investigates the multifaceted drivers behind the trending videos on YouTube. We use a comprehensive data set of trending videos to evaluate both quantitative metrics (such as views, likes, and comments) and qualitative features (such as video category and title content). Our methodology combines data cleaning, descriptive statistics, correlation analysis, categorical aggregation, and feature engineering, culminating in a comparative analysis of clickbait effectiveness. The results indicate a strong interdependence among numerical engagement metrics, a substantial influence of content categories on visibility, and a counterintuitive relationship between clickbait-laden titles and popularity. These findings suggest that while engagement metrics are key indicators of a video's trending status, authentic content and thoughtful categorization play equally pivotal roles.

Keywords: YouTube, trending videos, data analysis, clickbait, engagement metrics, content category, video popularity

1. Introduction

YouTube has become a dominant platform for video sharing and consumption in the digital era, with more than 2 billion monthly logged-in users. A diverse audience is attracted to the huge range of content available on YouTube, which includes news, personal vlogs, and education alongside entertainment. Content creators, marketers, and researchers have become increasingly interested in understanding the factors that contribute to the popularity of YouTube videos, as the number of videos uploaded daily continues to increase.

But what makes a video trend? Is it raw numerical performance—views, likes, comments—or qualitative elements such as the title or category of the video? As creators attempt to reverse engineer virality, understanding the underlying dynamics becomes crucial. This study addresses this question by analyzing a dataset of trending YouTube videos with the aim of identifying the most significant features that contribute to their trending status.

2. Methodology

2.1. Data Source

This study uses the Kaggle dataset "Trending YouTube Video Statistics". Although the data set includes trending video data from multiple countries, only the United States data was used for our analysis. Covers daily trending videos from 2017 to 2018 and includes 16 variables in CSV format, offering insights into the factors that drive video popularity.

2.2. Variable Selection

The following variables were selected as the basis of analysis:

Variable	**Description**
`clickbait_score`	Count of clickbait keywords found in video title
`title`	Video title
`publish_time`	Time the video was published
`trending_date`	Date video appeared on trending page
`views`	Number of views
`likes`	Number of likes
`dislikes`	Number of dislikes
`comment_count`	Number of comments
`category_id`	Category identifier

2.3. Method Introduction

To investigate the factors that influence the popularity of YouTube videos, this study conducted a structured exploratory data analysis (EDA) using the Trending YouTube Video Statistics dataset available on Kaggle. For consistency and specificity, only the dataset corresponding to the United States was used, containing metadata for videos that appeared on YouTube's trending page between 2017 and 2018.

The dataset was first loaded into a structured format using Python's `pandas` library. Initial data cleaning steps involved handling missing values (filling numerics with 0 and text fields with empty strings), removing duplicate rows, and converting temporal variables (`trending_date` and `publish_time`) into datetime objects. The `category_id` was mapped to human-readable names using a JSON file. Video titles were also preprocessed to remove punctuation and normalize text for further analysis.

Following preprocessing, descriptive statistics were computed to understand the distribution and central tendencies of key numerical features such as views, likes, and comment counts. Category-based analysis was performed to explore the distribution and average engagement metrics across different video categories. Correlation analysis using Pearson's coefficient helped determine the strength of relationships among views, likes, and comments.

To assess the qualitative influence of video titles, a custom `clickbait_score` was introduced. This score was derived by scanning titles for the presence of predefined keywords commonly associated with clickbait, such as "shocking," "amazing," and "you won't believe."

3. Results and Discussion

3.1. The Numbers Are Interconnected

Correlation analysis revealed strong positive relationships between views, likes, and comments. Videos that excelled in one metric often performed well in the others. This confirms that numerical popularity metrics are tightly linked, and gaining traction in one domain (e.g., likes) likely propels others due to algorithmic reinforcement.

	views	likes	comment_count
views	1.000000	0.849179	0.617657
likes	0.849179	1.000000	0.803088
comment_count	0.617657	0.803088	1.000000

Figure 1: Correlation matrix between views, likes, and comments

3.2. Content Category Influences Performance

Our analysis indicates that the category of a video plays a significant role in its likelihood of trending. Although complete category names were unavailable for all entries (requiring analysis based on `category_id`), we observed a clear imbalance in the distribution of trending videos across different categories. Certain `category_id`s accounted for a disproportionately high number of trending videos and also demonstrated higher average view counts compared to others in the dataset.

This finding suggests that the inherent nature of the content type heavily influences a video's potential to trend on the platform, potentially reflecting prevalent audience interests or the way the YouTube algorithm favors certain content types.

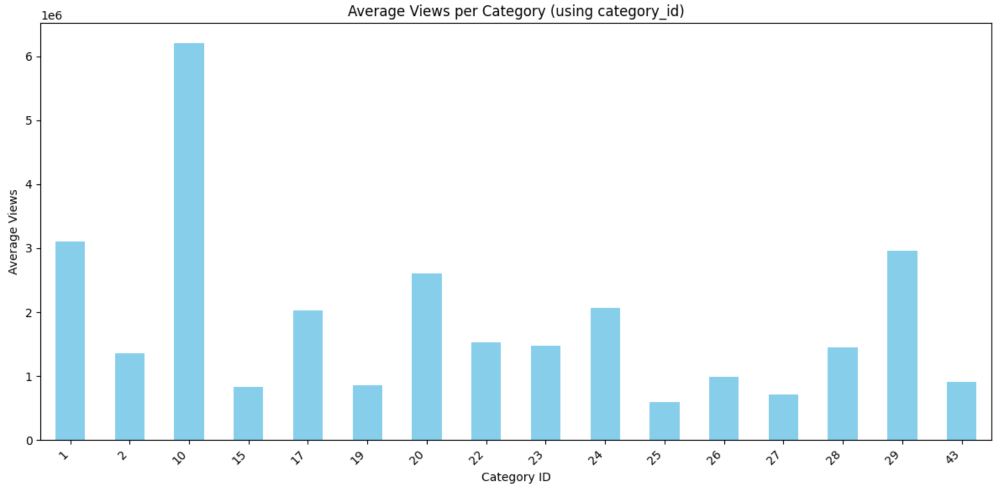

Figure 2: Average Views Per Category

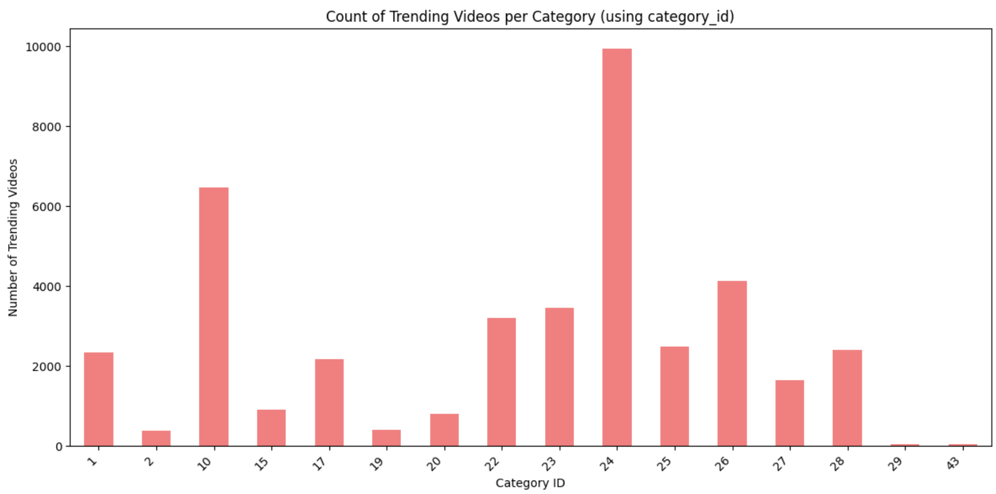

Figure 3: Distribution of trending videos across different categories

3.3. Clickbait Is Not Always Effective

Contrary to expectations, videos with lower clickbait scores had slightly higher average views compared to those with higher scores. While sensational titles are often assumed to drive clicks, our data suggest that overuse of clickbait may negatively impact viewer trust or lead to diminished engagement.

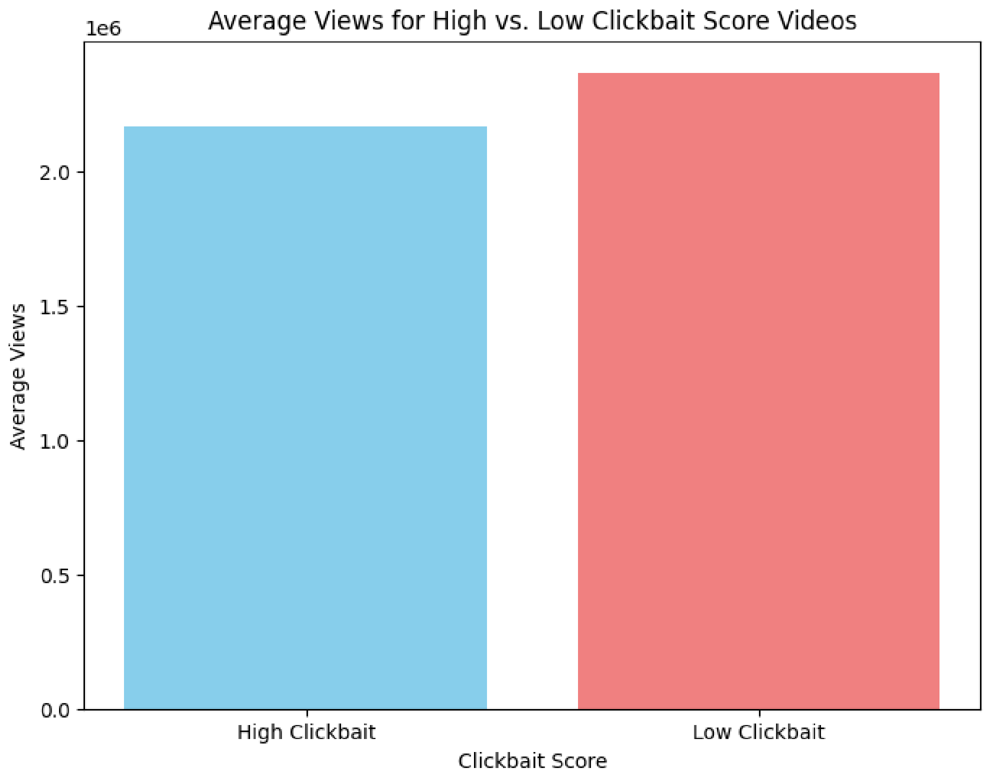

Figure 4: Comparison of average views for different clickbait scores

4. Conclusion

This analysis investigated factors influencing YouTube video trending status, revealing that both numerical engagement and content characteristics play roles. Our findings show that key engagement metrics like views, likes, and comments are strongly interconnected, highlighting that high numerical popularity is central to trending. Furthermore, video category significantly impacts trending likelihood; certain content types are considerably more represented among trending videos, indicating the strong influence of content alignment with platform trends or audience interest. Interestingly, our analysis of simple "clickbait" titles did not find a strong correlation with higher average views, suggesting viewers might respond better to authentic or relevant titles than overt sensationalism.

In essence, trending on YouTube appears to be a balance: achieving strong early engagement, creating content in high-potential categories, and employing thoughtful titles rather than relying on simple clickbait.

References

[1] YouTube Trending Dataset: `https://www.kaggle.com/datasets/datasnaek/youtube-new?resource=download`

[2] pandas documentation: `https://pandas.pydata.org/`

[3] matplotlib & seaborn for data visualization