Youtube Paper

Abstract:

This study investigates the interplay between various factors influencing the performance of YouTube videos, focusing on video metadata, thumbnail characteristics, and title sentiment. Leveraging advanced feature engineering, data analytics, and predictive modeling, it aims to reveal patterns that guide content creators in optimizing their strategies to attract and retain audiences. By examining related literature, conducting extensive exploratory data analysis, and employing both classification and regression models using Random Forest algorithms, this research highlights how publication timing, video length, title sentiment, and thumbnail design collectively shape audience engagement and a video's likelihood of how fast it will trending. The findings provide a comprehensive, data-driven foundation for content creators to refine their content strategies and achieve higher visibility, stronger engagement, and sustained growth in a competitive digital landscape.

1. Introduction

The explosion of digital media platforms has dramatically reshaped the ways in which content is created, distributed, and consumed. YouTube, as one of the foremost platforms, has become a focal point of digital engagement with over 2 billion logged-in monthly users and billions of uploaded videos. This vast and constantly expanding content ecosystem challenges creators to stand out in an increasingly saturated environment. Key issues involve navigating changing platform algorithms, capturing audience attention amidst a deluge of competing content, and maintaining consistent levels of viewer interaction over time.

In this complex ecosystem, creators grapple with several questions: **How do sentiment analysis and semantic search in video titles influence audience engagement? What role do thumbnail characteristics—such as the presence of human faces or certain color schemes—play in attracting viewers? How do video metadata factors like duration, resolution, category (gaming, politics, education), and aspect ratio impact viewer retention? How does timing of video publication affect visibility and the speed at which videos begin to trend?** Addressing these inquiries is crucial for optimizing content strategies, ensuring sustainable audience growth, and enhancing the overall impact of a creator's presence on YouTube.

This paper explores these multifaceted dynamics by integrating natural language processing, thumbnail image processing, and temporal feature engineering. By applying advanced modeling techniques and analyzing large-scale data, we aim to produce actionable insights that can guide content creators in refining their production, publishing, and optimization strategies for better engagement and increased trending potential.

2. Review of Related Literature

The existing body of research related to social media engagement, particularly on YouTube, is extensive. This literature review provides a structured overview of key areas and influential studies that inform the present research. The following comprehensive list is organized into thematic categories, reflecting the complexity of factors—from sentiment to scheduling—that influence content performance.

2.1 Comprehensive List of Relevant Papers for Literature Review

1. Sentiment Analysis of Video Titles and Impact on Audience Engagement

1.1 Influence of Titles on YouTube Trending Videos

Description: This paper investigates how sentiment polarity in video titles, examined using the Valence Aware Dictionary Sentiment Reasoner (VADER), influences video visibility and engagement. It provides insights into the direct impact of title sentiment on audience responses and video discoverability.

Relevance: Directly relates to understanding how emotional or tonal cues in titles affect audience behavior, a central aspect of this research.

1.2 Sentiment Analysis of Social Media Usernames and Titles

Description: This Master's thesis focuses on sentiment embedded in usernames and video titles, examining its influence on video performance and overall viewership. It highlights the subtleties of language use in digital platforms.

Relevance: Offers valuable perspectives on how linguistic and affective qualities of titles shape viewer expectations and engagement patterns.

2. Effect of Content Publication Timing on Visibility and Engagement

2.1 Scheduling Content on Social Media

Description: A theoretical framework describes how timing attributes—such as time of day—affect the relationship between social media content and user engagement [3].

Relevance: Crucial for understanding how strategic timing of video publication can maximize audience reach and interaction.

2.2 Digital Content Marketing on Social Media

Description: This work underscores how content timeliness drives engagement, particularly in B2B contexts, but with implications extending to consumer audiences. Timely publication increases visibility and user responsiveness [4].

Relevance: Reinforces that timing, even outside typical consumer contexts, is vital for capturing audience attention at peak interest moments.

2.3 Toward Maximizing the Visibility of Content in Social Media

Description: Explores methods to identify optimal posting times, showing that strategic scheduling enhances content discoverability and ensures sustained visibility [5].

Relevance: Directly informs strategies for timing video releases on YouTube.

3. Integrated Studies on Multiple Factors for Predicting Content Performance

3.1 Clicks for Money: Predicting Video Views Through Title and Thumbnail Sentiment

Description: Analyzes both title and thumbnail sentiment to predict video views, demonstrating that positive or emotionally resonant visuals and language can increase audience engagement [6].

Relevance: Aligns closely with the research question, linking visual and textual elements to measurable outcomes.

3.2 How Marketers' Video Optimization Practices Influence Performance

Description: Examines the effect of optimization techniques, including sentiment-rich titles, on video views. Combines text and sentiment analysis to offer a holistic view of video performance drivers [7].

Relevance: Suggests best practices for applying sentiment strategies to optimize engagement.

3.3 The Effect of Emotion in Thumbnails and Titles of Video Clips

Description: Investigates emotional cues in thumbnails and titles, using emotion theories and dual coding theory to understand how these factors shape viewer decisions and engagement.

Relevance: Informs the analysis of emotional triggers that increase audience involvement.

3.4 Visual Representation in Thumbnails and Video Content

Description: Discusses how thumbnails represent a distilled visual summary of the video, influencing initial viewer impressions and click-through rates [8].

Relevance: Essential for understanding how the viewer's decision-making process is visually guided.

4. Influence of Thumbnail Images on Click-Through Rates

4.1 Human Faces in Thumbnails

Description: Studies highlight that thumbnails featuring human faces can increase engagement and CTR by fostering a personal connection and curiosity [9].

Relevance: Direct evidence that specific visual elements in thumbnails affect engagement outcomes.

4.2 Psychological Impact of Visuals in Thumbnails

Description: Explores how rapidly the human brain processes images compared to text, influencing immediate user decisions and thus affecting CTR [10].

Relevance: Reinforces the importance of visually appealing thumbnails as a key driver of engagement.

4.3 Case Studies on Thumbnail Effectiveness

Description: Various case studies demonstrate how thumbnails with human faces and expressive emotions can boost CTR by significant margins (e.g., a 38% increase), underscoring visual psychology's role in engagement [11].

Relevance: Quantitative evidence supporting strategic thumbnail design.

5. Color Psychology in Thumbnails and Its Effect on Viewer Behavior

5.1 Color Psychology and Viewer Behavior

Description: Investigates how different colors in thumbnails evoke emotional responses, guiding user decisions on whether to watch a video [12].

Relevance: Provides a framework for using color strategically to influence audience emotions and behaviors.

5.2 Impact of Color Complexity on User Attention

Description: Research from the University of Notre Dame shows that color complexity in images increases user attention and engagement [13].

Relevance: Suggests that rich color palettes may contribute to thumbnails' effectiveness.

5.3 Color and Emotional Triggers in Thumbnails

Description: Examines how color choices can tap into emotional triggers, enabling creators to design thumbnails that resonate more deeply with their target audiences [14].

Relevance: Offers creators practical guidelines to leverage color for eliciting desired viewer responses.

5.4 Color in Marketing and Digital Content

Description: Explores color's significance in marketing, influencing brand perception, user decision-making, and overall content appeal [15].

Relevance: Shows the broader marketing context of color usage and its relevance to YouTube thumbnails.

6. YouTube Engagement Dynamics and Impact of Video Metadata

6.1 Engagement Dynamics and Sensitivity Analysis of YouTube Videos

Description: Examines meta-level features' influence on video popularity, including views, likes, and subscriber dynamics [16].

Relevance: Underlines the complexity of engagement and how various metadata attributes, beyond thumbnails and titles, affect performance.

6.2 Impact of YouTube's Recommendation System on Video Engagement

Description: Investigates how algorithmic recommendations shape video views, demonstrating platform-driven dynamics in engagement outcomes [17].

Relevance: Highlights the importance of external factors like recommendation systems in video visibility.

6.3 To Engage or Not Engage? Features of Video Content on YouTube

Description: Presents a model identifying key drivers of YouTube video popularity and engagement (e.g., views, likes, comments), offering comprehensive insights into content performance [18].

Relevance: Provides a theoretical framework that supports the importance of multiple factors in shaping engagement.

This literature review establishes the theoretical and empirical foundations of our study, highlighting the multifaceted factors—sentiment, timing, visuals, color psychology, and platform algorithms—impacting video performance. Collectively, these works lay the groundwork for the integrated analysis undertaken in this research.

---

3. Methodology

This study employs a data-driven and feature engineering-centric approach to analyze YouTube video performance. We begin with a comprehensive dataset of trending YouTube videos, enrich it through feature extraction techniques, and then apply modeling to identify key factors influencing engagement and trending behavior.

Data Collection:

Our dataset encompasses information on trending YouTube videos and associated channel characteristics collected daily, spanning multiple countries and content categories. Key attributes include:

- **Video-Level Details:** Publication date, trending date, video duration, category ID, view count, like count, comment count, tags, and descriptions.
- **Channel-Level Details:** Channel title, channel creation date, country, total view count, subscriber count, total video count, and channel descriptions.

The global and daily nature of this dataset provides a rich, temporal snapshot of how videos rise to trend status, capturing variations in user preferences and platform dynamics over time.

Feature Engineering:

To extract the most predictive and interpretable variables, we performed extensive feature engineering, targeting aspects like time-based publishing, thumbnail analysis, and language sentiment.

1. **Temporal Features:**
   - **Published Time Segmentation:** Each video's publication timestamp was categorized into morning, afternoon, evening, or night.
   - **Is Weekend:** A binary feature indicating weekend publication to capture potential differences in audience availability and engagement patterns.
2. **Video Characteristics:**
   - **Video Duration in Seconds:** Converting ISO 8601 durations into seconds to precisely quantify video length and correlate it with viewer retention.
   - **Video Category Encoding:** Translating categorical information (e.g., entertainment, gaming, news) into numerical form for modeling.
3. **Channel Attributes:**
   - **Channel Video Count:** Reflects a channel's productivity and possibly audience familiarity.
   - **Channel Age Difference:** The time gap between channel creation and video upload date indicates channel maturity and possibly trust or loyalty among viewers.
4. **Thumbnail Analysis:**
   - **Object Detection (YOLO):** Employed the YOLO algorithm to detect objects, including identifying human faces in thumbnails. The presence of people or recognizable objects can strongly influence CTR.

- **Color Analysis:** Extracted dominant RGB color values, brightness levels, and calculated color diversity. High color complexity and brightness may attract user attention and encourage clicks.
- **Visual Complexity and Emotional Resonance:** Emphasizing features like the presence of human faces, warm or cool color dominance, and overall vibrancy informs how these visual cues affect engagement.

5. **Title Sentiment Analysis:**

- **Sentiment Extraction (BERT-based Models):** Employed 'joeddav/distilbert-base-uncased-go-emotions-student' to determine the emotional tone of titles—ranging from excitement and joy to anger or curiosity. Although constrained computationally, focusing on titles alone provides a strong signal of content positioning.
- **Dense Embeddings:** Utilized transformer-based models to generate dense vector representations of video titles, enabling semantic similarity analysis. This approach helps identify patterns in successful titles by clustering semantically similar content and analyzing their performance metrics, providing insights into which title structures and themes resonate most with viewers.

Analytical Modeling Overview:

Following feature engineering, next is Exploratory Data Analysis (EDA) phase provided insights into data distributions and relationships. Also, we applied predictive modeling to address two key tasks: a classification model to predict rapid trending and a regression model to predict engagement rates. In both cases, we leveraged Random Forest algorithms known for their robustness, ability to handle complex feature interactions, and straightforward feature importance computation.

4. Exploratory Data Analysis and Results

4.1 Preprocessing

Before EDA, we ensured data quality and consistency:

1. **Date Parsing:** Converted publication and trending dates into datetime objects, ensuring accurate computations of temporal differences (e.g., time-to-trend).
2. **Timezone Normalization:** Removed timezone inconsistencies to standardize timestamps.
3. **Handling Invalid or Zero Values:** Removed entries with impossible values (e.g., channels with zero views and zero videos), as these likely indicate incomplete or corrupted data.
4. **Missing Values:**
    - Filled missing textual fields (descriptions, tags) with placeholders (“None”) to preserve records.
    - Retained only rows with valid category IDs.
    - Replaced zero view counts in videos as missing data points, acknowledging that no-views metrics offer limited analytical value.

4.2 Feature Engineering Recap

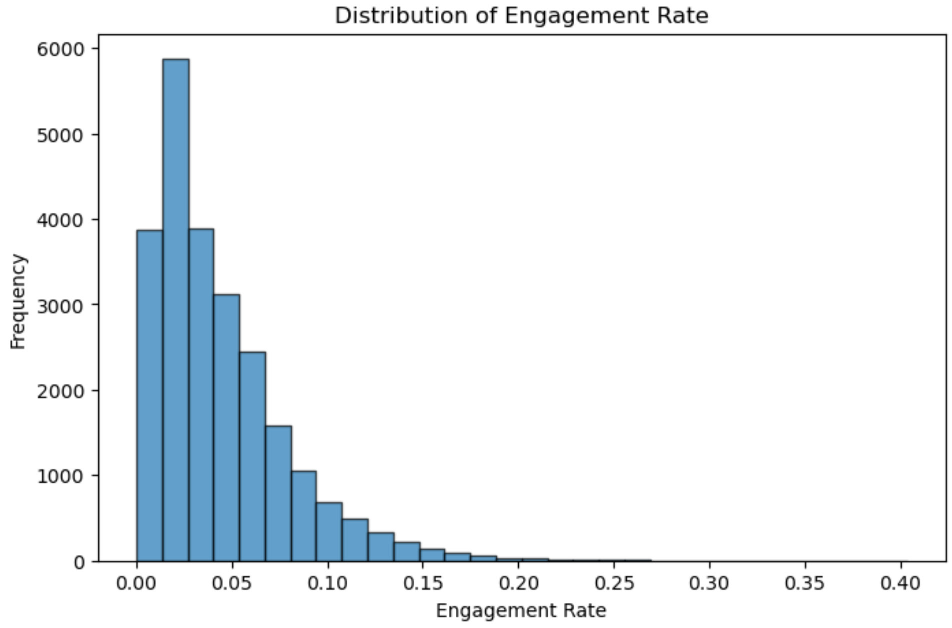
Having prepared our dataset, we introduced several new variables:

- **Engagement Rate:** (Likes + Comments) / Views. This normalized metric allows cross-comparisons of viewer interaction.
- **Time to Trend:** Computed as the time elapsed from video publication to the date it appeared on trending lists. This measures how quickly a video gains momentum.
- **Video Duration in Seconds:** Facilitates correlation analysis between content length and audience retention or engagement.
- **Thumbnail-Derived Features:** Indicators of human faces, dominant colors, brightness, and color diversity are now key explanatory variables.
- **Title Sentiment Detection:** Titles are analyzed using the **specifics fine-tuned BERT** models to identify specific emotions such as joy, sadness, anger, and surprise.
- **Title Embeddings:** Convert video titles into dense vector representations using a language model. Each title is transformed into a high-dimensional embedding that captures semantic nuances beyond simple keyword matching.

4.3 Exploratory Data Analysis (EDA)

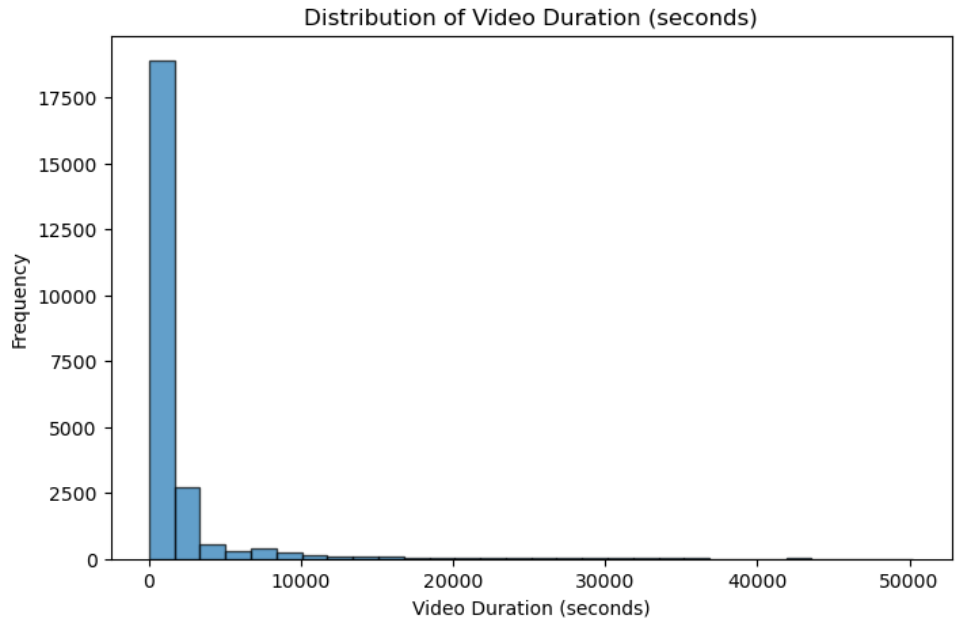
Engagement Rate Distribution

- Most videos cluster around low engagement rates (<0.05), but a subset shows significantly higher engagement.
- High engagement outliers may reflect content that resonates deeply with audiences, possibly due to emotional appeal or strong community ties.



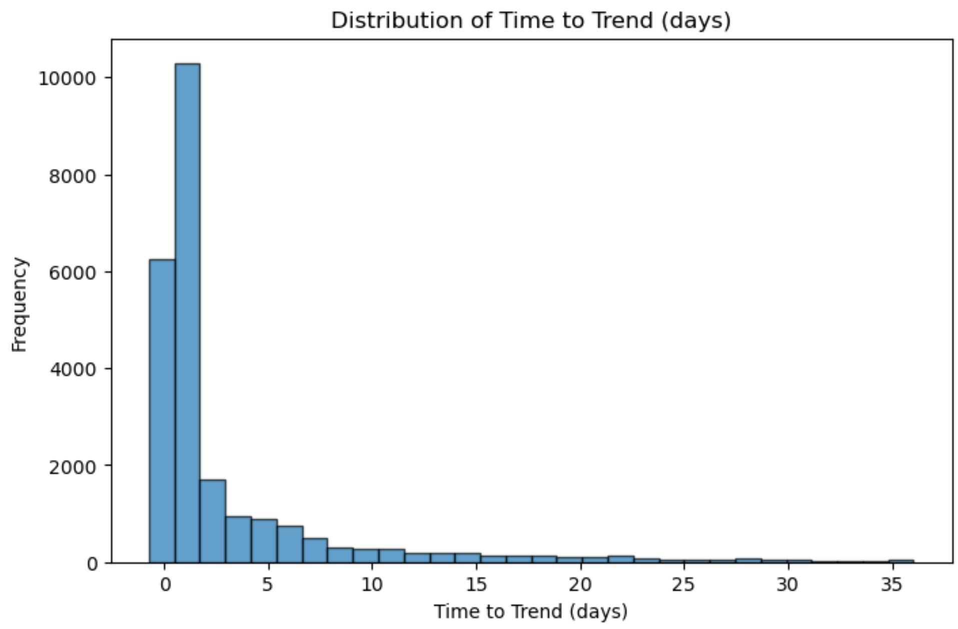
Video Duration Distribution

- The majority of videos are relatively short (under a few minutes), with fewer long-form pieces.
- Shorter videos often align with contemporary audience preferences for concise and digestible content, possibly leading to higher completion rates and engagement.



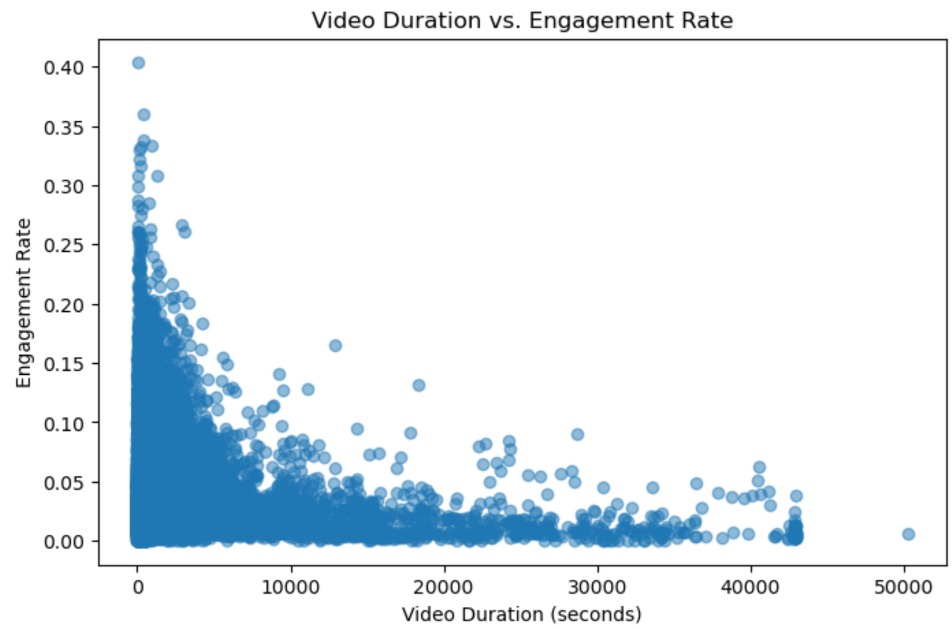
Time to Trend Distribution

- Most videos trend take less than 10 days of upload.
- Early-trending videos often address timely events, viral challenges, or benefit from algorithmic boosts. Later-trending content may be niche and rely on gradual community discovery.



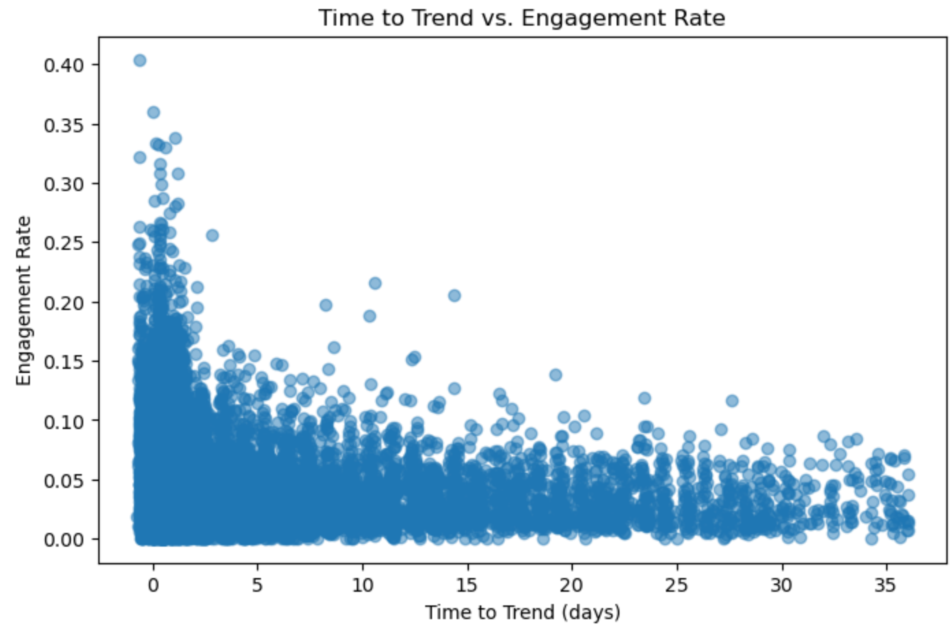
Correlation Between Duration and Engagement

- Shorter videos often show higher engagement rates.
- Viewers more easily complete shorter videos and feel more inclined to like, comment, or share, reinforcing engagement loops.



Correlation Between Time to Trend and Engagement

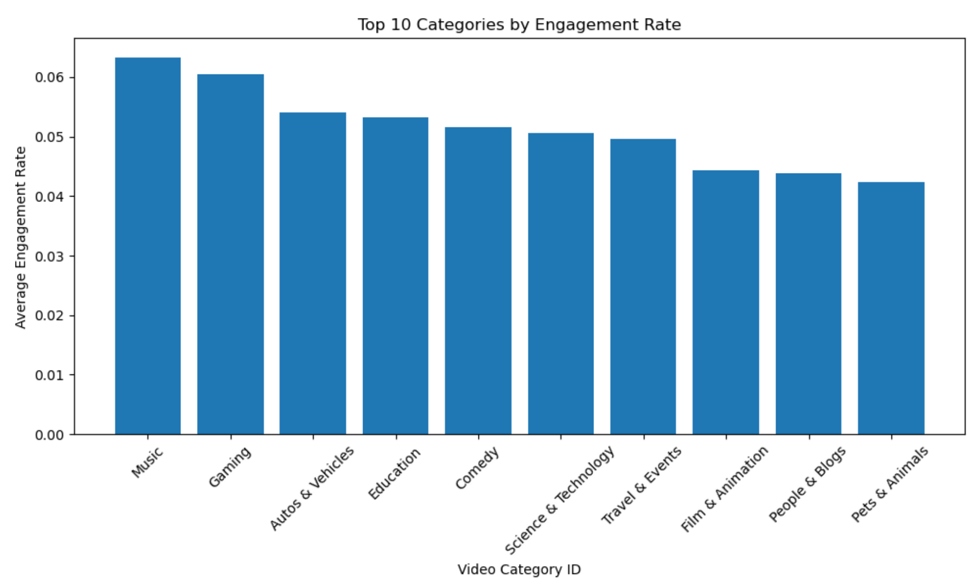
- Videos that trend sooner frequently exhibit higher engagement rates early on.
- Immediate and intense audience response can trigger platform signals that boost discoverability, creating a feedback loop.



Category and Country Analyses

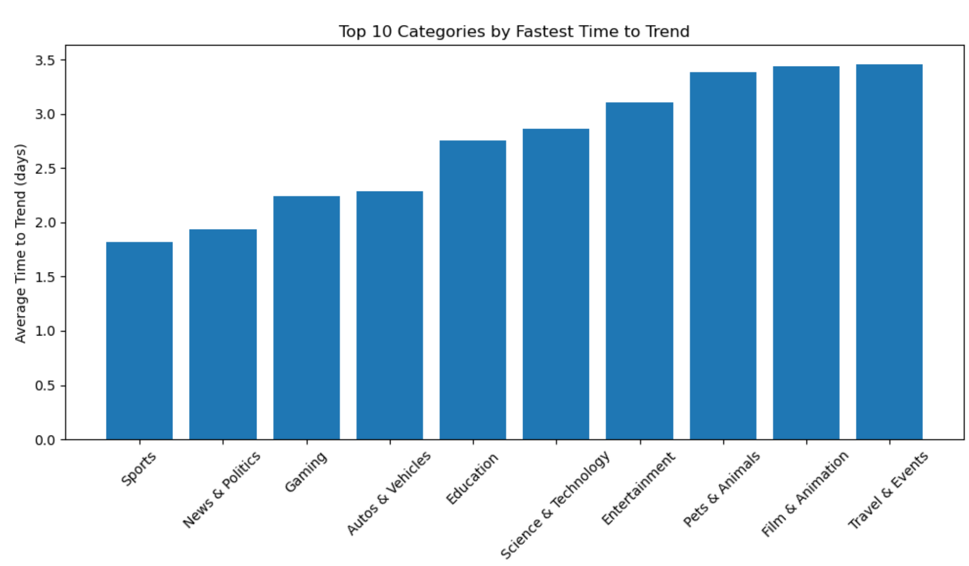
- **Categories:**
    - **Engagement:**

      Certain categories such as **Music** and **Gaming** consistently show higher engagement rates, suggesting these genres drive strong viewer interaction. Conversely, categories like **Pets & Animals** and **People & Blogs** see relatively lower engagement rates, possibly due to the nature of content being less interaction-oriented.



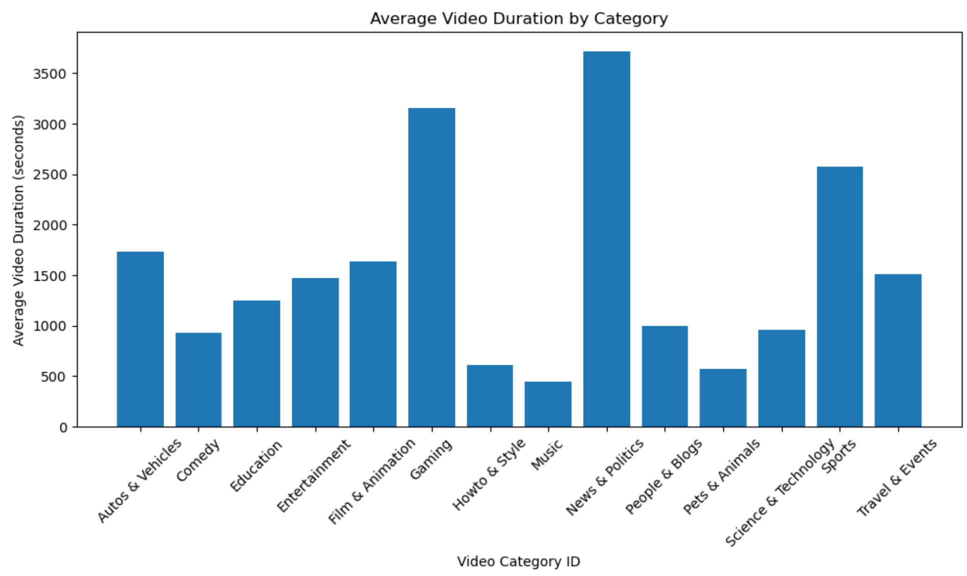
    - **Time to Trend:**

      Categories like **Sports** and **News & Politics** trend the fastest, likely due to their relevance to current events and the immediacy of viewer interest. In contrast, categories like **Film & Animation** and **Travel & Events** take longer to trend, as they might rely more on organic discovery or broader appeal over time.

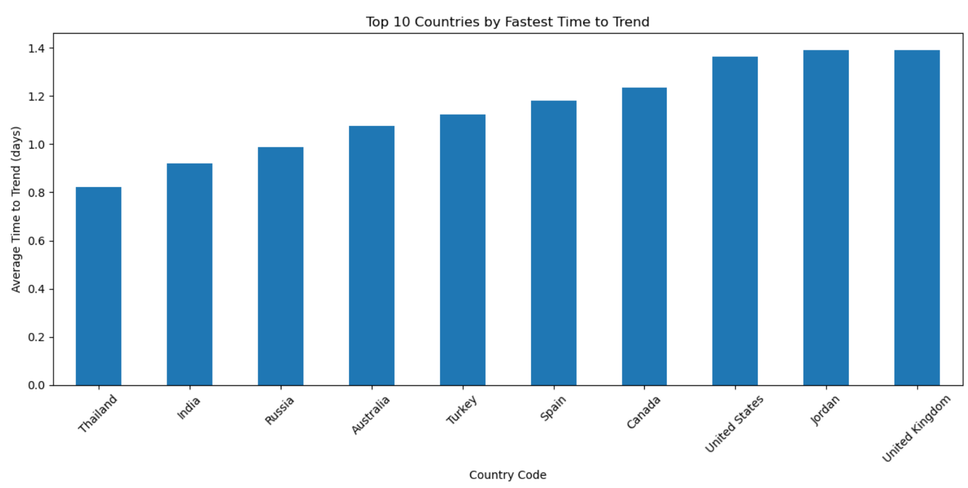


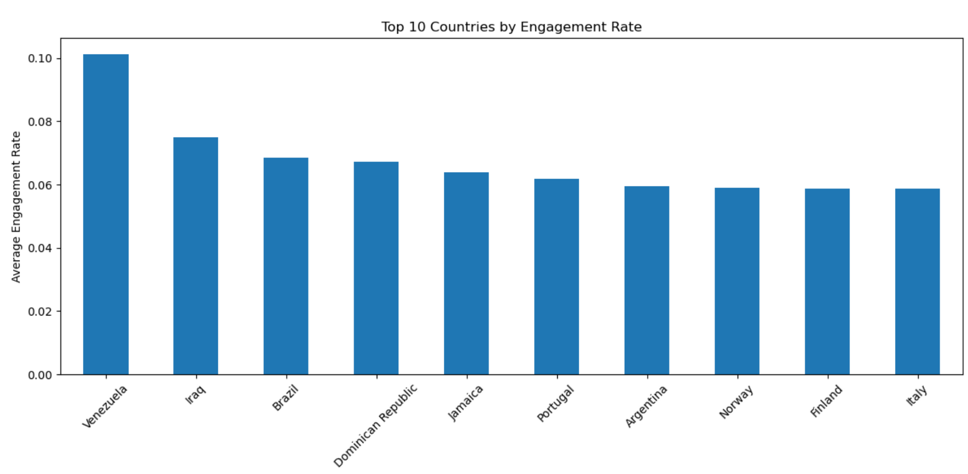
    - **Video Duration:**

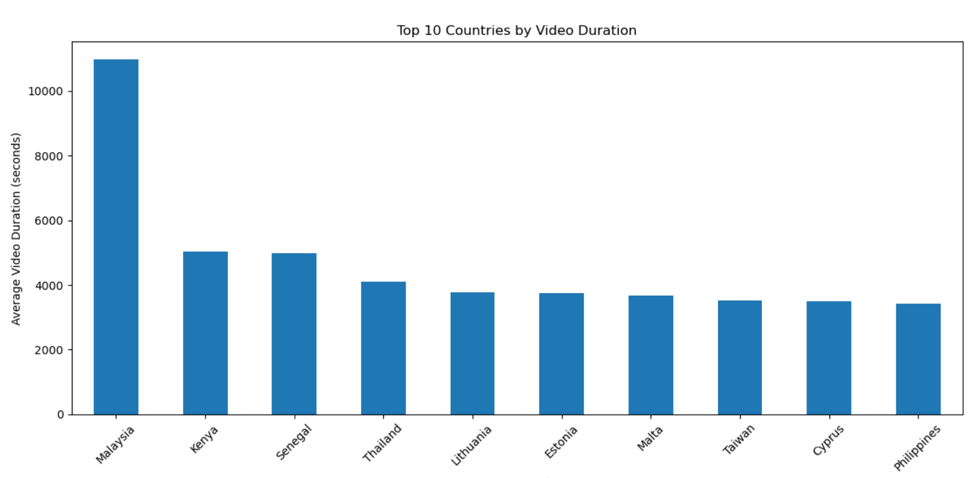
      Categories such as **News & Politics** and **Sports** tend to have longer average durations, possibly due to the depth or comprehensiveness of the content. Shorter durations are more common in **Comedy** and **Pets & Animals**, aligning with their focus on quick, engaging, and digestible content.



- **Countries:**
  - Regional differences in engagement rates and content preferences emerge.
  - Some countries show higher average engagement, reflecting cultural viewing habits or platform penetration.



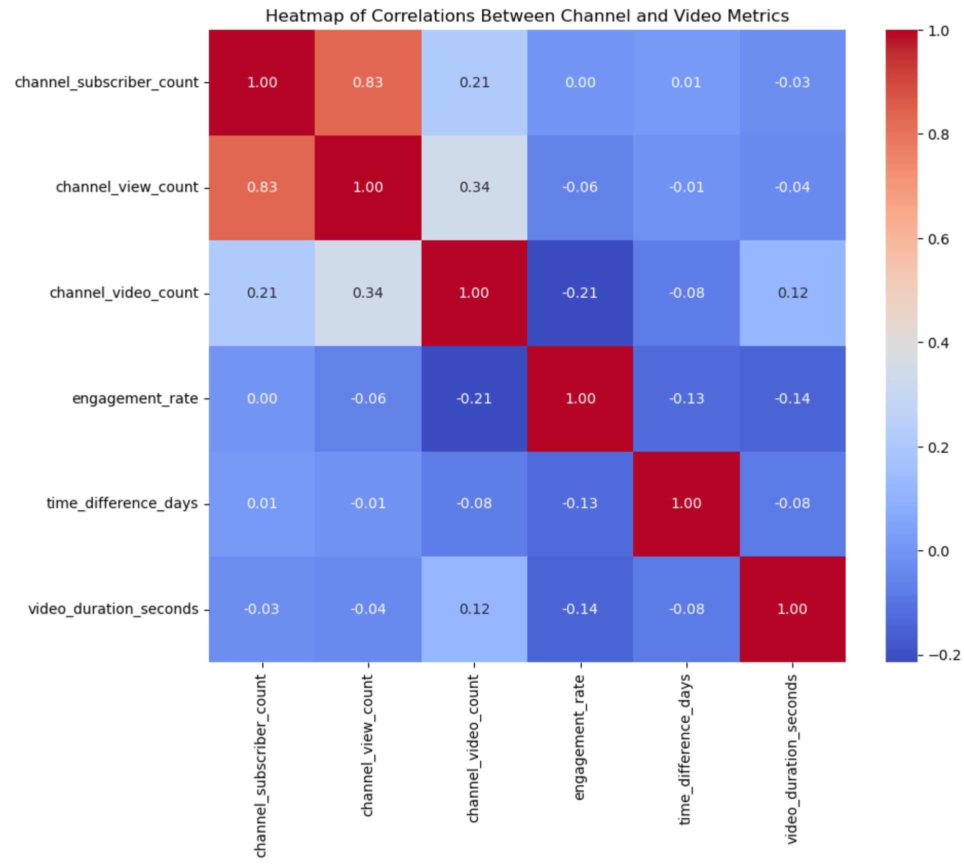




Correlation Analysis

- **Channel Subscriber Count:**
  - Strong positive correlation with **channel view count** (0.83), suggesting that larger subscriber bases drive more total views.
- **Channel View Count:**
  - Moderate positive correlation with **channel video count** (0.34), implying that more content leads to higher total views.
- **Channel Video Count:**
  - Weak negative correlation with **engagement rate** (-0.21), indicating that higher volume content might dilute individual video engagement.
- **Engagement Rate:**
  - Weakly negative correlations with **time difference (days)** (-0.13) and **video duration** (-0.14), suggesting that shorter videos trend faster and maintain better engagement.
- **Time Difference (Days):**
  - Minimal to no correlation with other metrics, highlighting its independence from channel size or video count.
- **Video Duration:**

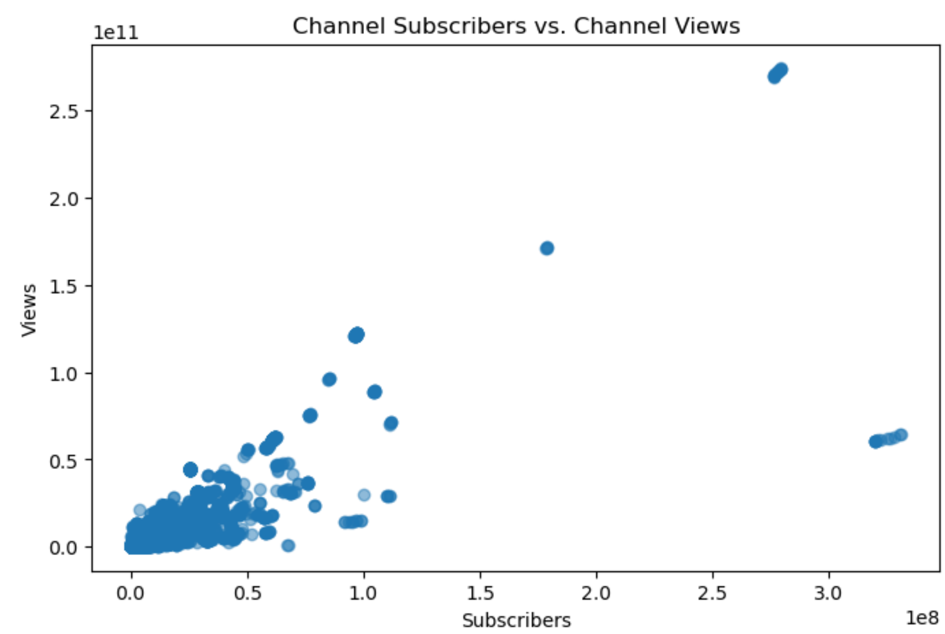
- Slight positive correlation with **channel video count** (0.12), implying longer videos may be more common on prolific channels.
- Weak negative correlation with **engagement rate**, showing that shorter videos tend to perform better in terms of audience interaction.



Channel Analysis

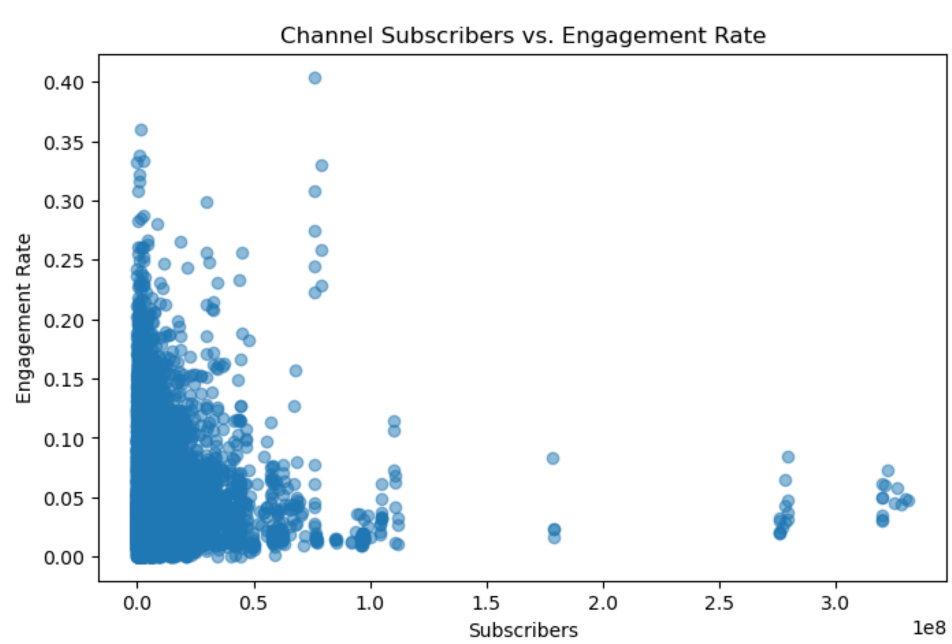
1. Channel Subscribers vs. Channel Views

- **Trend:** There is a clear positive correlation between subscribers and views, but with diminishing returns at higher subscriber counts.
- **Outliers:** Channels with extremely high views relative to their subscriber count may indicate viral content or widespread audience reach beyond subscribers.



2. Channel Subscribers vs. Engagement Rate

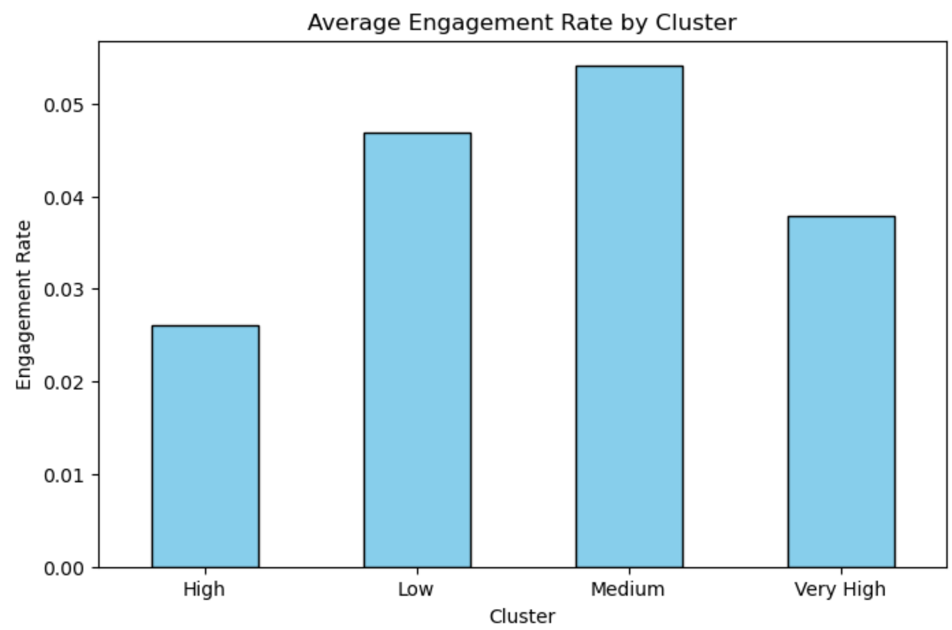
- **Trend:** Higher subscriber counts generally correlate with lower engagement rates, suggesting that smaller or niche channels may foster stronger viewer interaction.



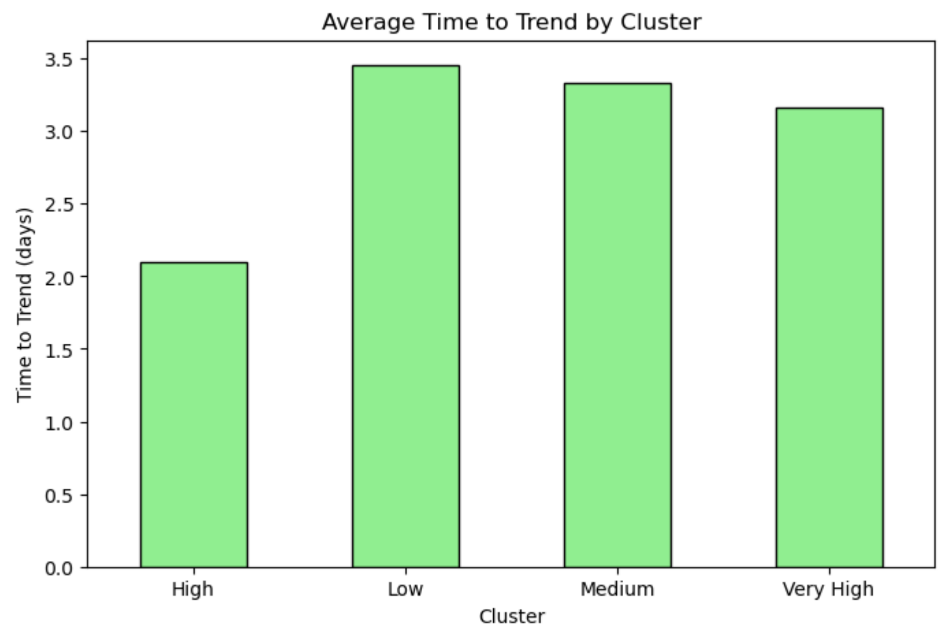
Clustering Analysis

- **Engagement Rate by Cluster:**
  - The **Medium** cluster has the highest engagement rate, followed by the **Low** cluster, suggesting that mid-range and smaller channels foster stronger audience interaction.

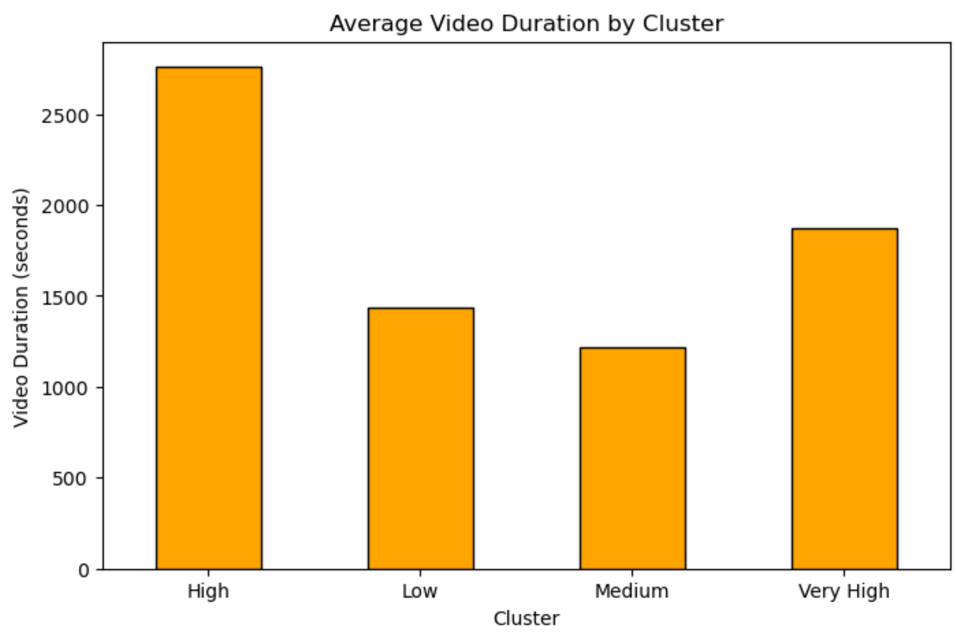
- The **High** and **Very High** clusters exhibit lower engagement rates, likely due to a broader and less niche audience base.



- **Time to Trend by Cluster:**
  - Channels in the **High** cluster trend the fastest, with an average time to trend of around 2 days, likely due to the immediate reach of larger audiences.
  - The **Low** and **Medium** clusters take longer to trend, averaging over 3 days, possibly due to slower organic discovery.

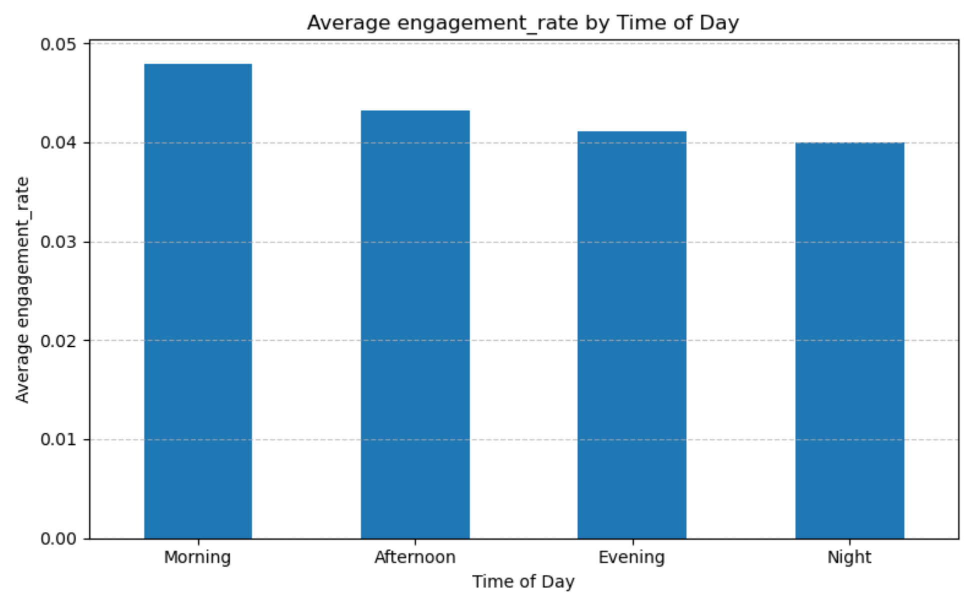


- **Video Duration by Cluster:**
  - Channels in the **High** cluster produce significantly longer videos, indicating a focus on in-depth or comprehensive content.
  - The **Very High** cluster also leans toward longer videos, while the **Medium** and **Low** clusters prefer shorter, more concise content.

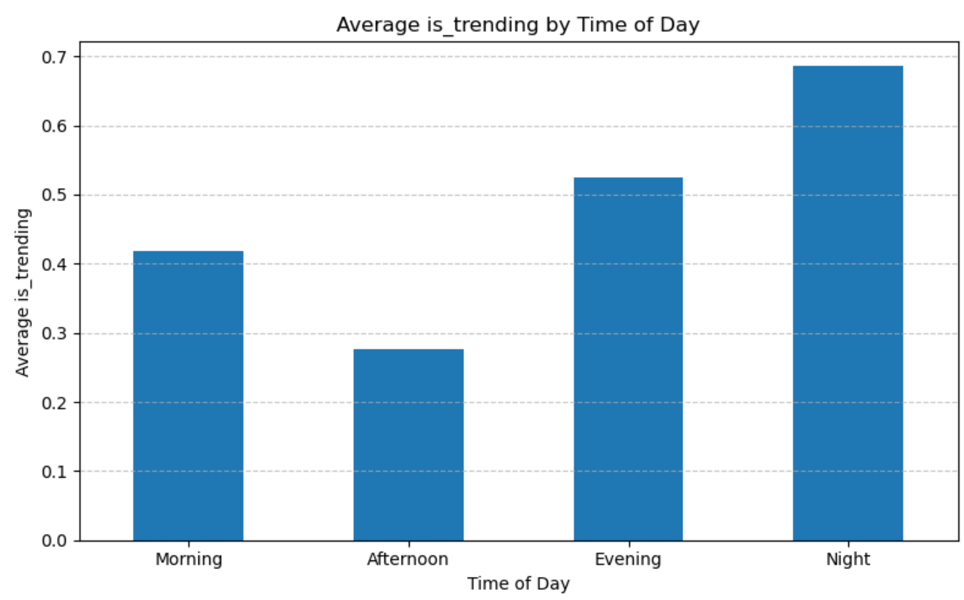


Time of Day Analysis

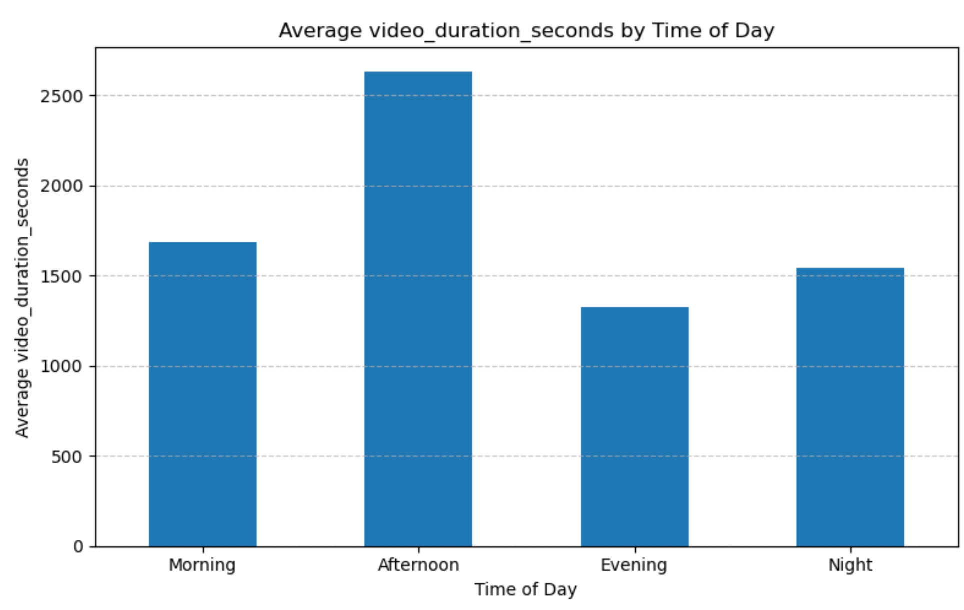
- **Average Engagement Rate by Time of Day:**
  - **Morning** videos achieve the highest engagement rates, suggesting that early content captures viewers' attention better.



- Average is_trending_1_day by Time of Day:
  - **Night** videos are the most likely to trend within 1 day, followed by **Evening** content. This could indicate that uploading later in the day aligns with audience activity spikes.
  - **Afternoon** videos are the least likely to trend quickly, possibly due to lower user activity during this time.



- Average Video Duration by Time of Day:
  - **Evening** videos are generally shorter, potentially catering to viewers looking for quick and digestible content.



---

5. Training Models

The modeling phase transforms the insights derived from EDA into predictive frameworks. Two modeling approaches were taken: (1) a classification model predicting whether a video would trend immediately (within one day) versus taking longer (2–10 days) and (2) a regression model predicting engagement rate based on multiple features.

5.1 Pretrained Image and Text Models

Leveraging pretrained models for both image and text modalities enriches the feature space beyond traditional metadata attributes (e.g., view counts, duration, category). By integrating computer vision and natural language processing (NLP) techniques, we gain nuanced insights into how visual and linguistic cues drive audience engagement. Thumbnails serve as the viewer's first impression and can significantly influence click-through rates. Simultaneously, video titles, as the initial textual hook, shape viewer expectations and emotional response. Therefore, applying robust image and text analysis allows us to capture these subtle yet critical signals that standard metadata often overlooks.

Image Processing:

The visual appeal and content of a video's thumbnail can profoundly influence a viewer's decision to engage with the content. To translate these visual signals into usable numeric features for modeling, we adopted a two-step approach: object detection using YOLO and the extraction of color-related metrics.

1. **Object Detection Using YOLO:**

	contain_1	contain_2	contain_3
0	person	car	NaN
1	person	person	NaN
2	person	person	NaN
3	person	person	person
4	donut	person	person
...	...	...	...
28662	person	person	NaN

We employed the YOLO (You Only Look Once) algorithm to detect and categorize objects present in each thumbnail. YOLO's efficiency and state-of-the-art performance make it well-suited for large-scale datasets. After running YOLO across all thumbnails, we captured the top three most salient objects for each video, storing them in the fields **contain_1**, **contain_2**, and **contain_3**. For example, a thumbnail might yield:

- **contain_1 = "person"**
- **contain_2 = "car"**
- **contain_3 = "donut"**

Identifying these objects provides a granular representation of the thumbnail's content. Prior research suggests that the presence of a person often correlates with higher engagement, possibly due to human relatability and the sense of personal connection it fosters. Other objects like food, animals, or vehicles may signal thematic interests or trending topics, serving as subtle cues to the viewer about the video's subject matter and appeal.

1. **Color-Based Features for Visual Aesthetics:**

   Beyond object presence, color and brightness play crucial roles in attracting viewer attention. To quantify these attributes, we extracted the following features:

	dominant_color_r	dominant_color_g	dominant_color_b	brightness	color_diversity
0	78.0037	65.1272	45.8112	12.786193	6950
1	90.7597	95.0496	94.4673	13.171799	4574
2	87.9835	71.3292	88.6419	13.388571	7131
3	108.4453	105.3869	105.3105	13.463079	7201
4	119.5454	95.2846	88.1445	13.349830	7472

- **dominant_color_r, dominant_color_g, dominant_color_b:**

  These three features represent the RGB components of the thumbnail's dominant color. By identifying the most prominent hue in each image, we gain insight into the overall visual theme and emotional resonance—e.g., warmer tones may evoke excitement or comfort, while cooler tones might suggest calmness or professionalism.

- **brightness:**

  This metric captures the overall lightness of the thumbnail. Brighter images tend to stand out more prominently on a viewer's screen, potentially increasing the likelihood of a click. Average brightness was computed after resizing images for computational efficiency, ensuring consistent comparisons across all thumbnails.

- **color_diversity:**

  A more varied color palette can signal complexity or richness, potentially intriguing viewers. By counting the number of distinct colors in the thumbnail, we quantify how diverse or uniform the image's visual presentation is. Thumbnails with greater color diversity may catch viewers' eyes more effectively and pique their curiosity.

These image-derived features, combined, provide a holistic view of a thumbnail's visual appeal and psychological impact on viewers.

Text Processing:

Video titles function as a linguistic framing device. They set expectations, convey tone, and can evoke emotional responses that influence user decisions. To systematically evaluate titles, we incorporated advanced NLP techniques:

1. **Sentiment Analysis:**

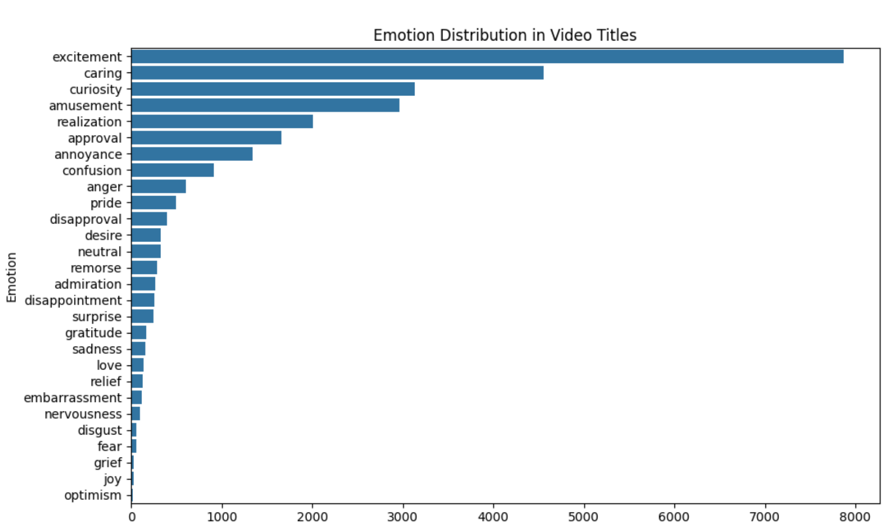
   We generated sentiment-based features that reflect the emotional spectrum conveyed by the video title. Emotions like excitement, anger, curiosity, or joy can shape viewer interest and engagement. By quantifying the emotional valence and category, we can correlate specific emotions with higher or lower CTRs and engagement rates.

2. **Models Used for Sentiment Analysis:**

   To accommodate linguistic variability and ensure robust sentiment detection, we employed two models:

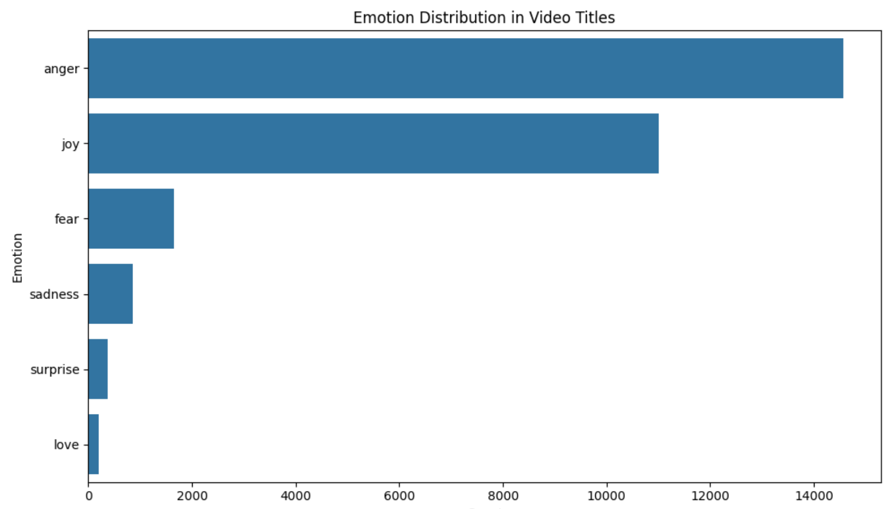
   - `joeddav/distilbert-base-uncased-go-emotions-student`:

     This is a compact, English-focused model fine-tuned on the GoEmotions dataset, making it adept at identifying a wide range of nuanced emotions in English titles. Its lightweight architecture allows efficient batch processing of large volumes of titles.



   - `google-bert/bert-base-multilingual-cased`:

     Considering that YouTube content is globally accessible, many video titles are non-English. Using a multilingual BERT model broadens our coverage, allowing basic sentiment detection across various languages. While originally trained to handle multiple languages, its direct applicability to nuanced emotional contexts in non-English titles may be limited due to data constraints.

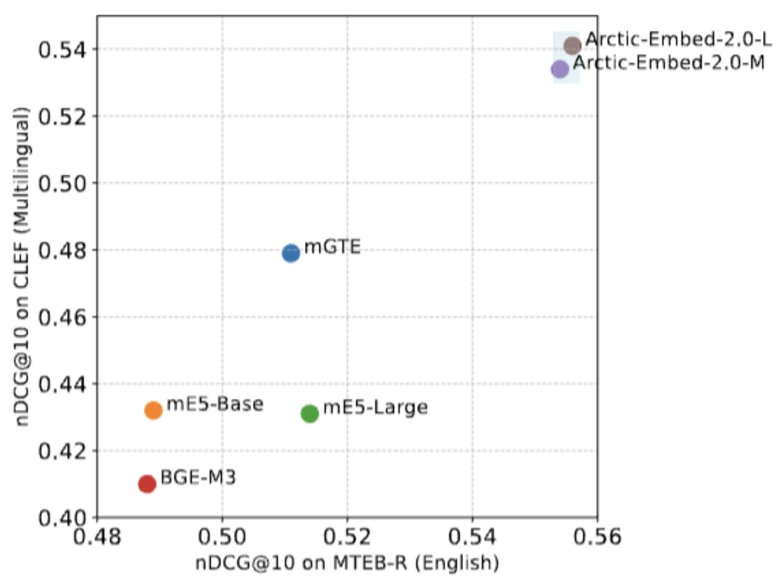


3. **Multilingual Considerations and Embedding-Based Approaches:**

Although we integrated a multilingual model, sentiment analysis remains inherently biased toward English due to the nature of available fine-tuning datasets. This limitation means that while the multilingual model can process non-English text, the accuracy and granularity of emotional detection may be lower for these languages.

To address these concerns and capture deeper semantic relationships between video titles, we utilized Snowflake's arctic-embed-l-v2.0 model for semantic search. This state-of-the-art model not only provides high-quality multilingual text retrieval without sacrificing performance in English, but also enables us to understand the contextual meaning and thematic similarities between titles. By embedding titles into a shared semantic vector space, we can identify related content, discover trending topics, and understand viewer preferences across different languages and cultural contexts. This approach provides several benefits:

- **Language-Agnostic Representations:** Converting titles into embeddings allows the model to focus on semantic meaning rather than specific linguistic tokens, improving cross-language comparisons.



Single-vector dense retrieval performance of open source multilingual embedding models with fewer than 1B parameters. Scores are average nDCG@10 on MTEB Retrieval and the subset of CLEF (ELRA, 2006) covering English, French, Spanish, Italian and German.

- **Dimensionality Reduction via PCA:** High-dimensional embeddings can be computationally expensive. After generating multilingual embeddings, we applied Principal Component Analysis (PCA) to reduce the feature space to 16 dimensions, striking a balance between preserving semantic richness and maintaining computational efficiency.
- **Enhancing Non-English Title Interpretation:** While direct sentiment detection might remain imprecise for some languages, embedding-based semantic representations help identify content topics, detect similarity patterns, and infer likely emotional tone indirectly. This provides a more holistic foundation for understanding how international audiences engage with content.

By integrating these image and text processing techniques, we expand our analytical capabilities beyond conventional metadata, capturing the richer tapestry of visual appeal, emotional nuance, and cultural diversity inherent in YouTube's global content ecosystem.

5.2 Tabular Classical Models

Beyond deep learning or embedding-based methods, we employed classical machine learning models (Random Forest) on the engineered tabular dataset. This choice provides several benefits:

- **Interpretability:** Feature importance metrics from Random Forest models help identify the most influential variables.
- **Robustness to Noise:** The ensemble nature of Random Forest helps manage complex, non-linear relationships without overfitting easily.
- **Scalability:** Random Forest models can handle large datasets efficiently, making them suitable for YouTube-scale data.

5.2.1 Classification Task

Objective:

Predict whether a video will trend within one day (rapid trending) versus taking longer (2–10 days) to reach the trending list. This rapid ascent can be referred to as its "Rate of Virality," reflecting how quickly it gains traction.

Target Variable:

Binary classification: 1 if a video trends within a day of publishing, 0 if it trends later.

Feature Set:

Incorporates metadata (duration, category), channel characteristics (age, number of videos), publication timing (time-of-day, weekend), thumbnail features (object detection, color), and title sentiment.

Why This Matters:

Identifying factors that lead to early trending can guide creators to optimize aspects known to generate immediate viewer interest. Rapidly trending videos may receive additional platform support, recommended placements, and virality boosts, ultimately maximizing reach.

Model Selection and Training:

- **Algorithm:** Random Forest Classifier with a grid search for hyperparameter tuning (e.g., number of trees, max depth) to balance model complexity and performance.
- **Training and Validation:** Utilized stratified cross-validation to ensure balanced class representation and stable performance estimates.

- **Evaluation Metrics:** Accuracy, precision, recall, and F1-score measure predictive quality. Additionally, a confusion matrix provide more nuanced insights.

Interpretation of Results:

A well-performing classification model that accurately predicts early-trending status can reveal which features—be it sentiment, thumbnail brightness, or posting time—most strongly influence a video's initial breakthrough.

5.2.2 Regression Task

Objective:

Predict the engagement rate (a continuous variable) for a given video.

Target Variable:

Engagement Rate = (Likes + Comments) / Views

Feature Set:

Similar to the classification task, but optimized to explain variance in a continuous measure of engagement. The inclusion of thumbnail metrics, sentiment, video length, and timing allows the model to discern subtle patterns that contribute to viewer interaction.

Model Selection and Training:

- **Algorithm:** Random Forest Regressor to capture non-linear relationships between features and engagement.
- **Evaluation Metrics:** Mean Squared Error (MSE) and R-squared scores gauge predictive accuracy and the proportion of variance in engagement explained by the model.

Value of the Regression Model:

By accurately predicting engagement rates, creators can calibrate their content strategies to target metrics known to boost interaction. For example, if results show that highly positive title sentiment and evening publishing significantly lift engagement, creators can adapt accordingly.

6. Results, Feature Importance and Model Interpretation

Both the classification and regression Random Forest models produce feature importance rankings, offering transparency into which factors exert the most influence. These insights help in refining content strategies and verifying theoretical assumptions drawn from the literature and EDA.

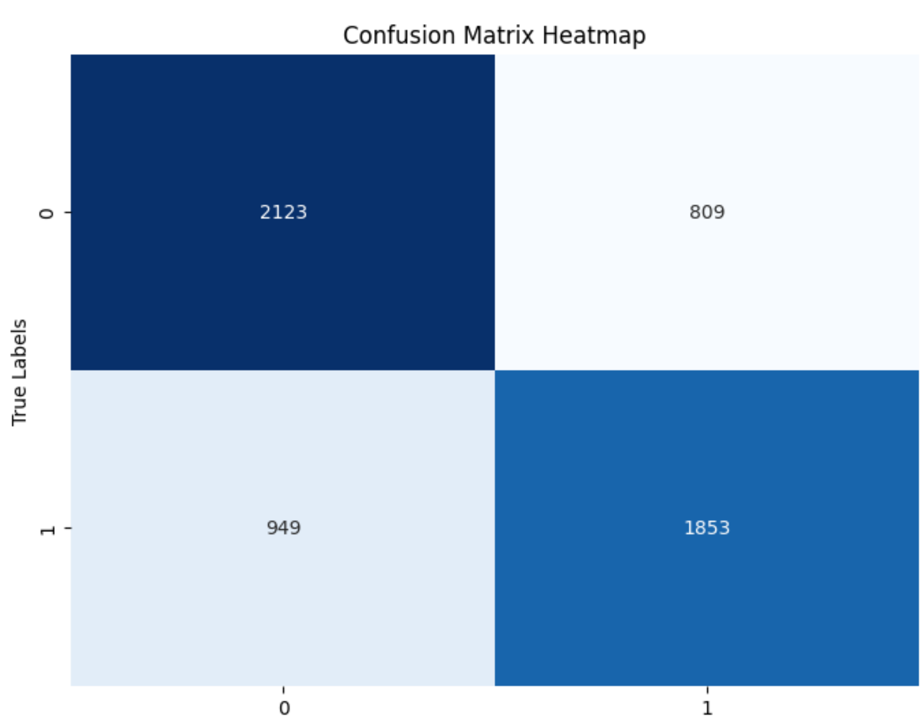
- **Common Influential Features:**
  - **Title Sentiment:** Emotional resonance in titles can significantly alter immediate click behavior and sustained engagement.
  - **Thumbnail Brightness and Human Faces:** Visually appealing and relatable thumbnails rank highly, corroborating the literature on human-centric imagery.
  - **Time of Publication:** Afternoon and evening uploads frequently correlate with better engagement and earlier trending.
  - **Video Duration:** Shorter videos often see higher engagement rates, reflecting viewer preferences for easily consumable content.

Visualizing these feature importances (e.g., via bar plots) provides an immediate, intuitive understanding of what drives model predictions. This transparency aids creators, marketers, and platform strategists in focusing their efforts on elements that matter most.

Classification Findings (Rate of Virality):

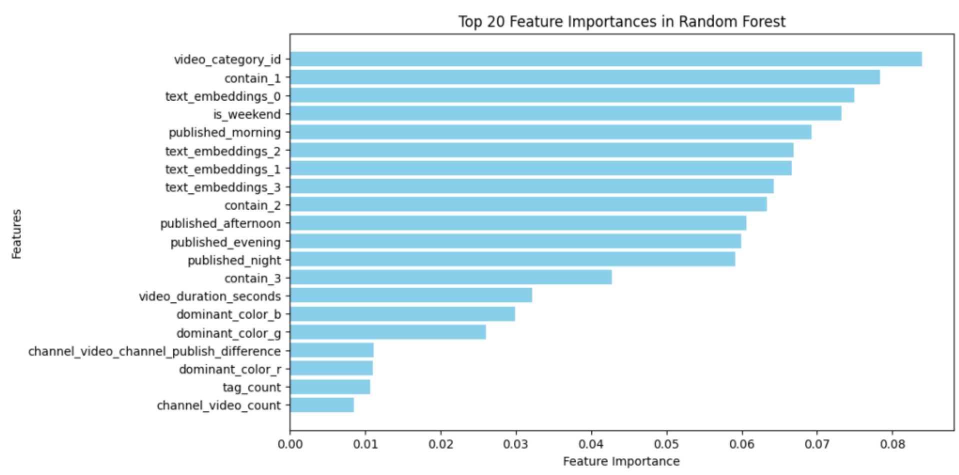
The classification model effectively distinguishes between videos that achieve rapid virality (trending within one day) and those requiring more time. Performance metrics and the confusion matrix demonstrate a balanced capacity to capture patterns associated with immediate audience engagement, while still leaving opportunities for refinement.

```
Model Accuracy: 0.6934077432856645
              precision    recall  f1-score   support

           0       0.69      0.72      0.71      2932
           1       0.70      0.66      0.68      2802

    accuracy                           0.69      5734
   macro avg       0.69      0.69      0.69      5734
weighted avg       0.69      0.69      0.69      5734
```



Key insights from the classification feature importance analysis reveal that certain factors consistently influence early trending behavior:

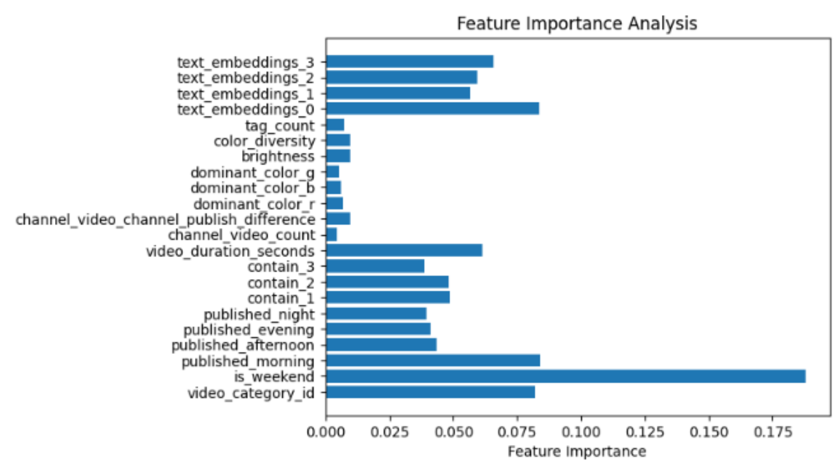
- **Video Content Category:** Certain categories inherently lend themselves to quicker uptake, reflecting audience preferences for specific genres or trending topics.
- **Thumbnail Objects:** The top detected object in the thumbnail can signal thematic relevance, enticing viewers to click almost immediately.
- **Textual Embeddings and Timing Signals:** Nuanced language in titles and strategic publishing times (e.g., weekend, evening) shape the likelihood of immediate virality.
- **Color and Aesthetic Attributes:** While less dominant, color composition and brightness subtly affect first impressions and prompt initial engagement.

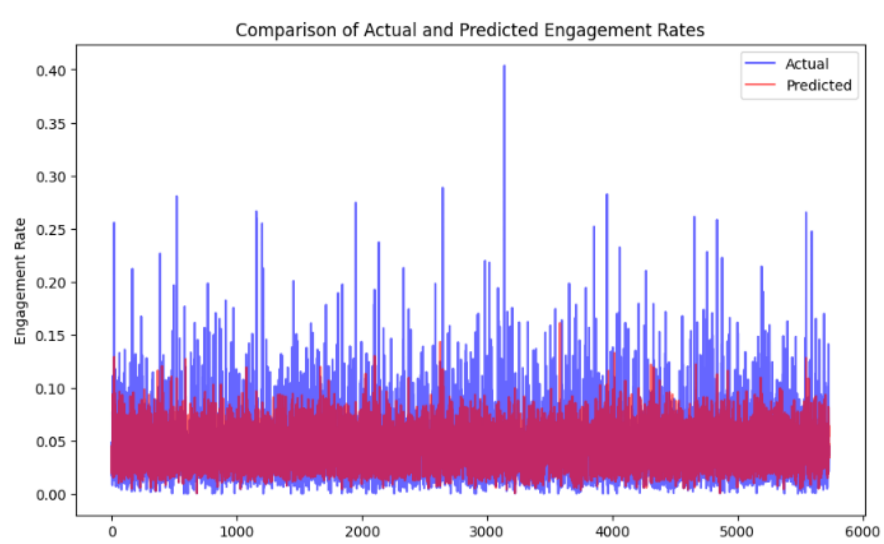


Regression Findings (Engagement Rate Prediction):

The regression model aims to quantify the engagement level a video achieves, reflecting the depth of viewer interaction relative to its reach. Although the accuracy for engagement rates is more moderate than the classification task, it still provides valuable guidance on continuous interaction drivers.

A comparison of actual and predicted engagement rates shows that while predictions track general trends, discrepancies exist that highlight the complexity of modeling nuanced viewer behaviors.





SHAP force plots further illuminate how certain features raise or lower the predicted engagement rate for individual videos:

- **Textual Embeddings:** Emotionally resonant titles, thematic cues, or narrative framing in the text remain critical. Even subtle linguistic nuances can shift engagement expectations.
- **Channel Dynamics and Timing:** Consistent publishing practices and strategic time slots correlate with higher predicted engagement.
- **Duration and Color Metrics:** Shorter videos and particular color profiles can nudge engagement upward, though their influence is more modest compared to text and timing factors.

Holistic Interpretation:

Taken together, the classification and regression results underscore a multifaceted ecosystem of factors driving YouTube video performance. Rapid virality hinges on content category, thumbnail intrigue, and timing strategies, while sustained engagement rates benefit from strong narrative cues, reliable publishing routines, and appealing visual aesthetics.

Creators and strategists can use these insights to refine both the "hook" that secures initial traction and the "substance" that fosters deeper audience involvement over time. By attending to these elements—supported by data-driven evidence—they can enhance their content's discoverability, resonance, and long-term success.

7. Conclusion

This paper presents a comprehensive, multi-faceted exploration of the factors influencing YouTube video engagement and trending behavior. By integrating research from sentiment analysis, color psychology, scheduling theory, and platform dynamics, we reveal how textual, visual, and temporal cues shape viewer engagement patterns.

Key Takeaways:

- **Timing and Visibility:** Videos posted during periods of high viewer activity (e.g., afternoon and evening) or on weekends often experience more rapid engagement, increasing their likelihood of early trending.
- **Title Sentiment and Emotional Appeal:** Positive, intriguing, or strongly emotional titles encourage initial clicks and sustained interaction, reflecting the importance of linguistic nuance.
- **Thumbnail Strategy:** Human faces, vibrant colors, and brightness in thumbnails create immediate visual connections with viewers, raising curiosity and willingness to engage.
- **Video Duration and Content Structure:** Shorter videos with high viewer completion rates foster more likes and comments, improving engagement metrics and potential trending performance.

Practical Implications:

Creators can use these insights to refine their content strategies. By aligning upload times with peak engagement windows, using emotionally resonant titles, and crafting visually compelling thumbnails, they can more effectively capture audience attention. Data-driven approaches help navigate platform algorithms and stand out in a competitive digital ecosystem.

Future Work:

Expanding the dataset's geographic and categorical scope, incorporating richer textual analysis of descriptions and tags, and applying advanced modeling techniques can deepen our understanding of engagement patterns. Additionally, analyzing emerging trends like short-form vertical video content and live streaming will ensure continued relevance as digital media landscapes evolve.

---

8. References

[1] The Influence of Titles on YouTube Trending Videos. Available at:

https://www.researchgate.net/publication/379971848_The_Influence_of_Titles_on_YouTube_Trending_Videos

[2] Sentiment Analysis of Social Media Usernames and Titles on YouTube and Twitch. Available at:

https://www.researchgate.net/publication/356377410_Sentiment_Analysis_of_Social_Media_Usernames_and_Titles_on_YouTube_and_Twitch

[3] Scheduling Content on Social Media: Theory, Evidence, and Application. Available at:

https://www.researchgate.net/publication/328179895_Scheduling_Content_on_Social_Media_Theory_Evidence_and_Application

[4] Digital Content Marketing on Social Media, 2024. Available at:

https://www.sciencedirect.com/science/article/pii/S0019850124000221

[5] Toward Maximizing the Visibility of Content in Social Media. Available at:

https://www.researchgate.net/publication/323208558_Toward_maximizing_the_visibility_of_content_in_social_media_brand_pages_a_temporal_analysis

[6] Clicks for Money: Predicting Video Views Through a Sentiment Analysis of Titles and Thumbnails. Available at:

https://ideas.repec.org/a/eee/jbrese/v183y2024ics0148296324003539.html

[7] YouTube Marketing: How Marketers' Video Optimization Practices Influence Video Views. Available at:

https://www.researchgate.net/publication/342178230_YouTube_marketing_How_marketers'_video_optimization_practices_influence_video_views

[8] An Exploration of the Relation Between the Visual Representation in Thumbnails and Video Content. Available at:

http://www.byungwan.com/papers/Thumbnails.pdf

[9] Human Faces in Thumbnails. Available at:

https://thumbnailtest.com/guides/face-in-youtube-thumbnail

[10] Psychological Impact of Visuals in Thumbnails. Available at:

https://www.linkedin.com/pulse/why-your-thumbnails-matter-look-data-behind-user-meer-shah--lpjjf

[11] Case Studies on Thumbnail Effectiveness. Available at:

https://www.linkedin.com/pulse/why-your-thumbnails-matter-look-data-behind-user-meer-shah--lpjjf

[12] Color Psychology and Viewer Behavior. Available at:

https://mention.com/en/blog/social-media-psychology-color/

[13] Impact of Color Complexity. Available at:

https://news.nd.edu/news/high-color-complexity-in-social-media-images-proves-more-eye-catching-increases-user-engagement/

[14] Color and Emotional Triggers in Thumbnails. Available at:

https://medium.com/@jyotijdash/the-art-and-science-behind-irresistible-youtube-thumbnails-db75ec5a627a

[15] Color in Marketing and Digital Content. Available at:

https://mention.com/en/blog/social-media-psychology-color/

[16] Engagement Dynamics and Sensitivity Analysis of YouTube Videos. Available at:

https://arxiv.org/abs/1611.00687

[17] The Impact of YouTube Recommendation System on Video Views. Available at:

https://www.researchgate.net/publication/220269659_The_impact_of_YouTube_recommendation_system_on_video_views

[18] To Engage or Not Engage? The Features of Video Content on YouTube Affecting Digital Consumer Engagement. Available at:

https://www.researchgate.net/publication/350465651_To_engage_or_not_engage_The_features_of_video_content_on_YouTube_affecting_digital_consumer_engagement

How to Calculate the Engagement Rate for All Social Media Platforms

https://www.socialinsider.io/blog/engagement-rate/#a