YouTube Trending Analytics: Exploring the Path to Video and Channel Success

Project Overview:

This project focuses on analysing YouTube trending video data collected via the YouTube API. The analysis is performed on both video-specific and channel-specific data, with an emphasis on identifying key factors that contribute to video virality and channel growth. The project aims to derive insights that can help content creators optimize their strategies for better engagement and visibility on the platform.

Data Collection:

Trending videos and their associated channels were collected using the YouTube API. The data includes various attributes such as video views, likes, comments, duration, tags, and channel statistics like subscriber count, upload rate, and channel age.

Data Analysis: The analysis was conducted in two stages:

1. **Video Data Analysis:**

    - **Outliers Analysis:** The video data was analysed twice once with outliers included and once with outliers removed. This approach helped in understanding the impact of extreme values on the overall trends and correlations.
    - **Correlation Analysis:** Spearman correlation was used to assess the relationship between key variables like view count, like count, comment count, video duration, and the number of tags. This method was chosen for its robustness to outliers.

2. **Channel Statistics Analysis:**

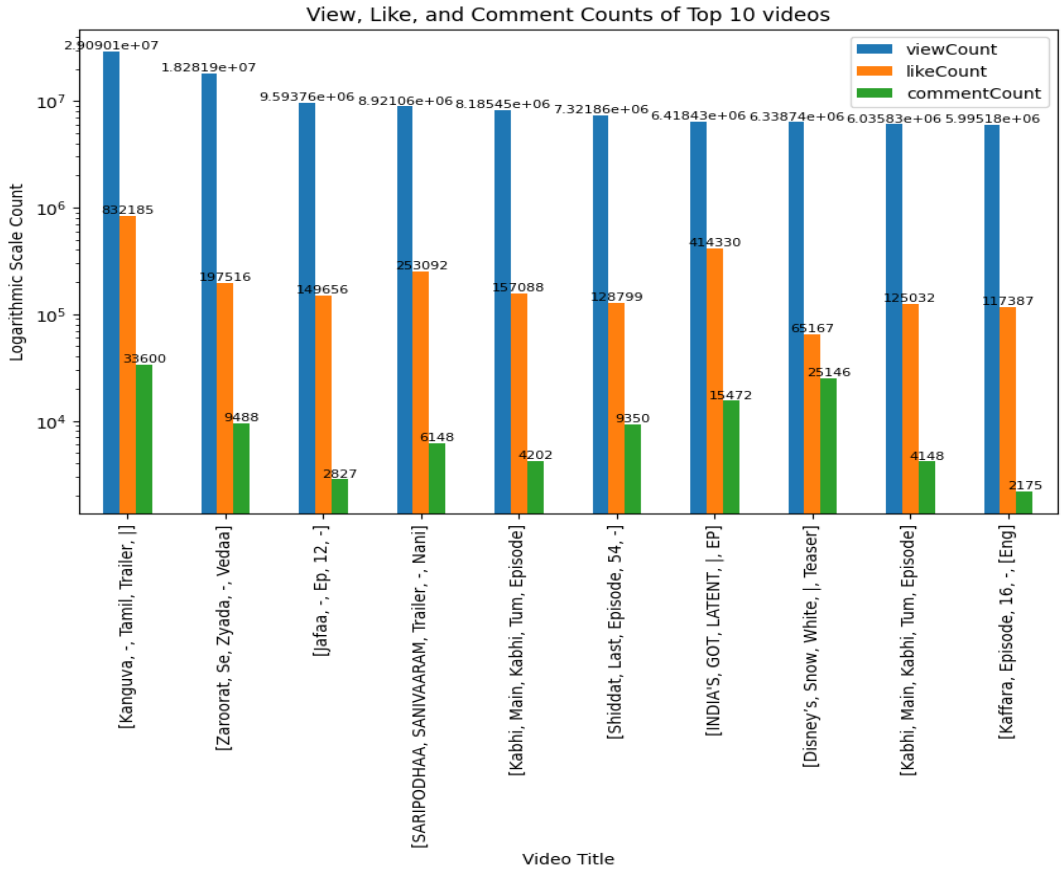
    - **Correlation Analysis:** The relationship between subscriber count, upload rate, and the number of trending videos was analysed using Spearman correlation. This helped in understanding how frequently uploading content and the existing subscriber base influence a channel's success.
    - **Regression Analysis:** Linear regression models were built to predict the number of trending videos based on subscriber count and upload rate.

The key questions that guided your YouTube trending video data analysis project:

- What factors contribute to a video becoming trending on YouTube?
- How do outliers affect the overall trends and correlations in video data?
- What is the relationship between video duration and viewer engagement?
- How does the number of tags used in a video influence its visibility and engagement?
- What are the most common tags that has been used in these trending videos?
- Does more upload rate attract more subscriber count?
- Can subscriber count and upload rate predict the number of trending videos a channel produces?
- How does the age of a YouTube channel correlate with its performance in terms of trending videos?
- Does caption availability influence a video's likelihood of becoming trending?
- How does the default audio language of a video impact its success?
- What is the ideal length of a video for reaching success?

Analysis and findings:

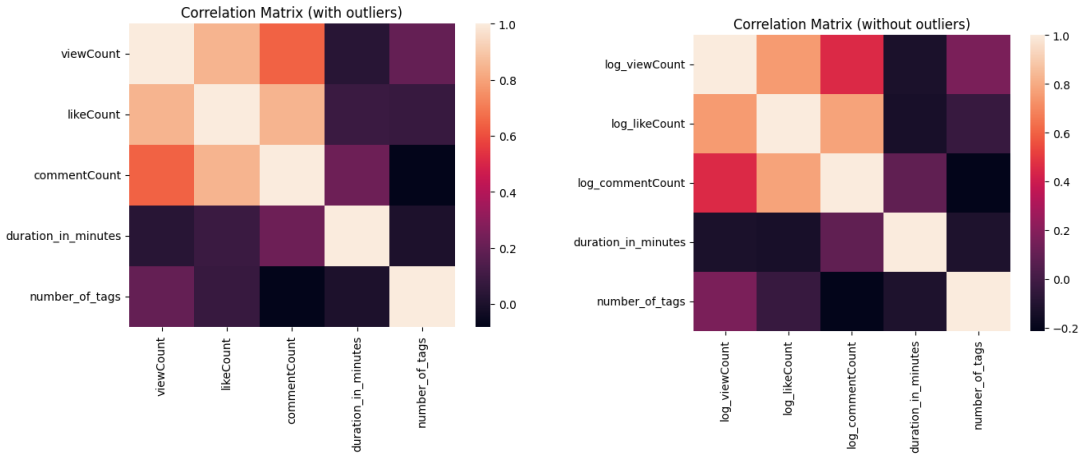
- **Top 10 most trending videos based on their views counts, likes counts and comment counts**



- **Relationship between viewers engagement and video duration or number of tags used in the video**

WITH OUTLIERS	viewCount	likeCount	commentCount	duration_in_minutes	number_of_tags
viewCount	1.000000	0.844647	0.641216	0.034719	0.197185
likeCount	0.844647	1.000000	0.844856	0.084613	0.078016
commentCount	0.641216	0.844856	1.000000	0.218718	-0.082123
duration_in_minutes	0.034719	0.084613	0.218718	1.000000	-0.001226
number_of_tags	0.197185	0.078016	-0.082123	-0.001226	1.000000

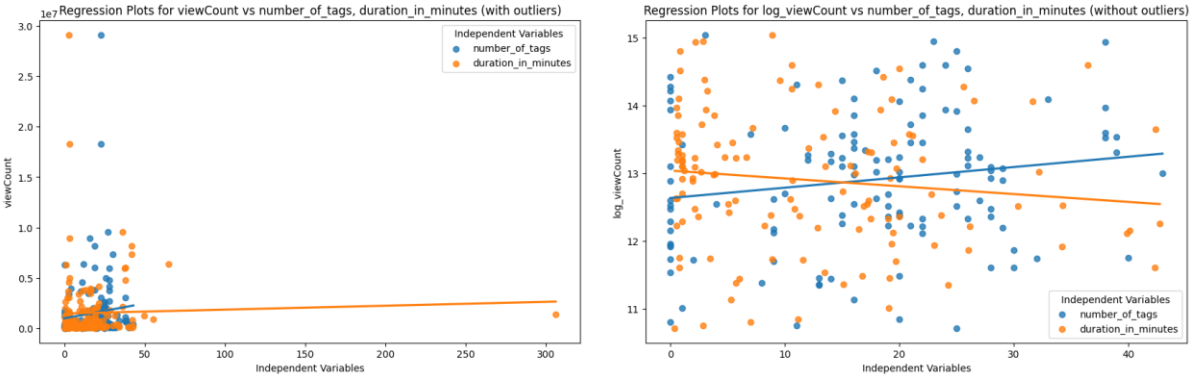
WITHOUT OUTLIERS	log_viewCount	log_likeCount	log_commentCount	duration_in_minutes	number_of_tags
log_viewCount	1.000000	0.748790	0.454281	-0.127848	0.164823
log_likeCount	0.748790	1.000000	0.776837	-0.135009	-0.035957
log_commentCount	0.454281	0.776837	1.000000	0.086860	-0.214538
duration_in_minutes	-0.127848	-0.135009	0.086860	1.000000	-0.115747
number_of_tags	0.164823	-0.035957	-0.214538	-0.115747	1.000000



It does not show a strong correlation between the variables in both the cases with or without outliers.

There is a correlation showing higher view count tend to attract higher like counts.

- **Regression analysis:**

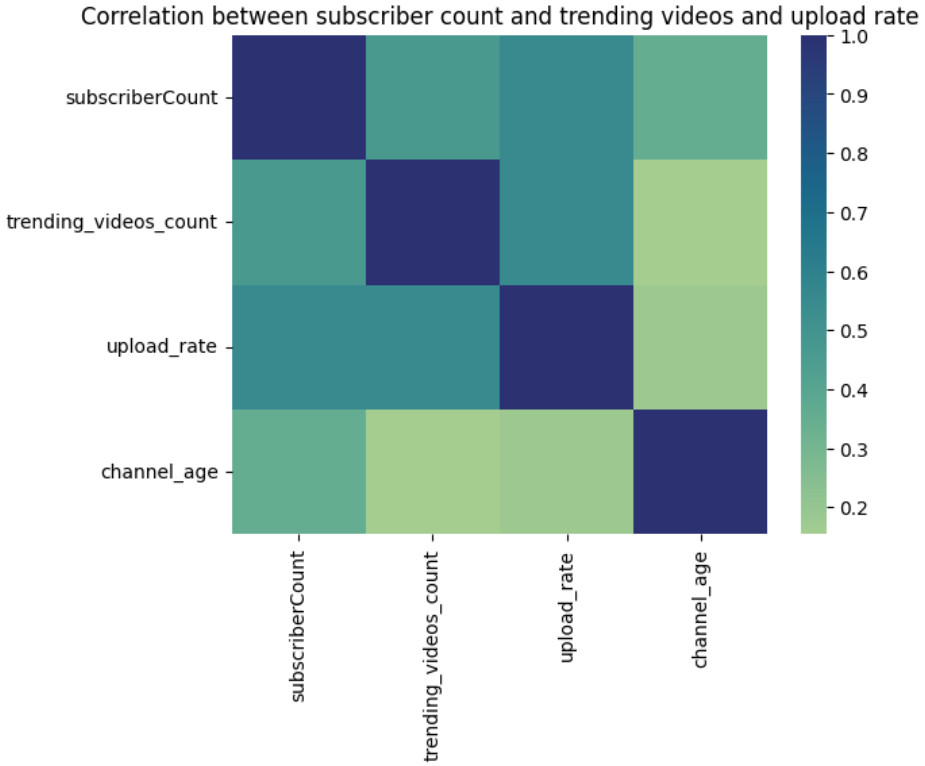


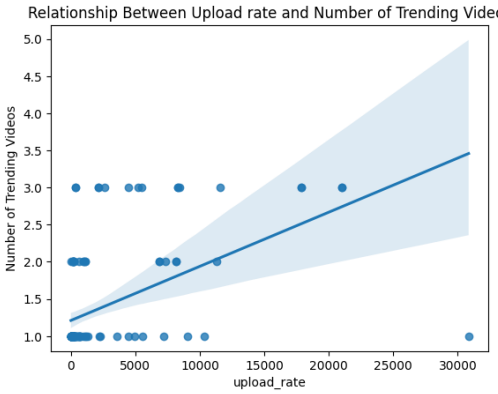
Though without outliers we have a clearer picture to see how the trend is but still not strong enough to say that there is a strong positive or negative correlation

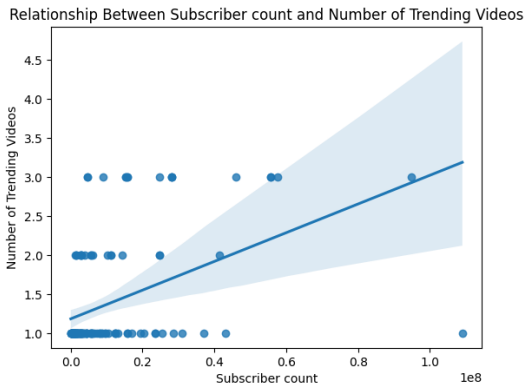
➢ **The most common tags that has been used in these trending videos**

The bigger the size of the word, higher the number of videos where this word has been used

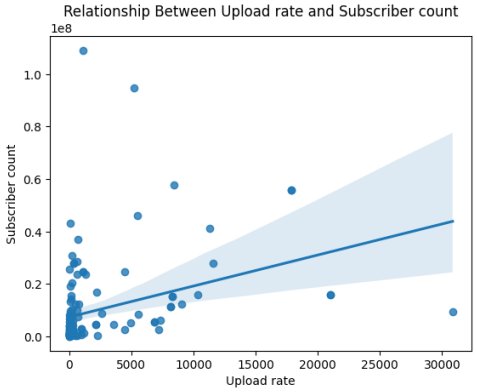
- **Does more upload rate attract more subscriber count?**
- **Can subscriber count and upload rate predict the number of trending videos a channel produces?**





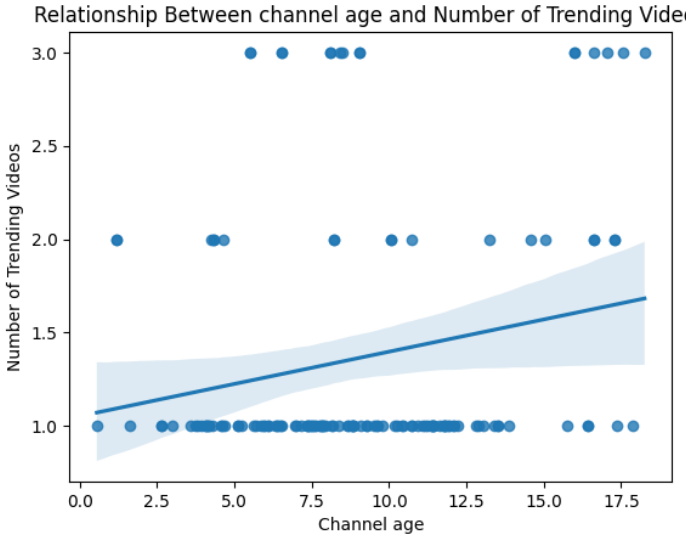


Upload rate and subscriber count tend to have a linear relation with the number of trending videos (has a moderate correlation)



More upload rate tend to attract more subscribers (moderate correlation)

- **How does the age of a YouTube channel correlate with its performance in terms of trending videos?**



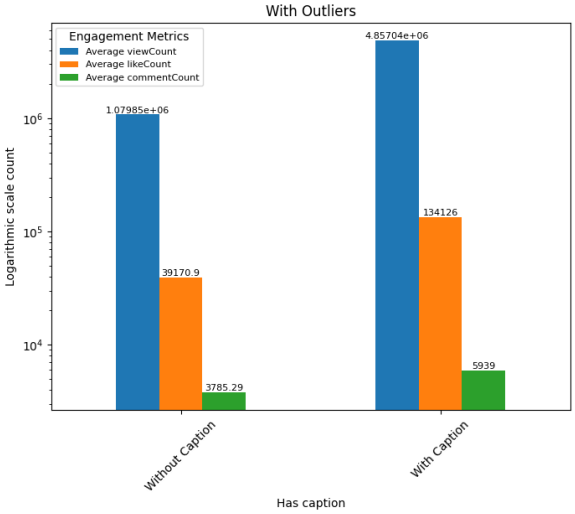
Very less correlation

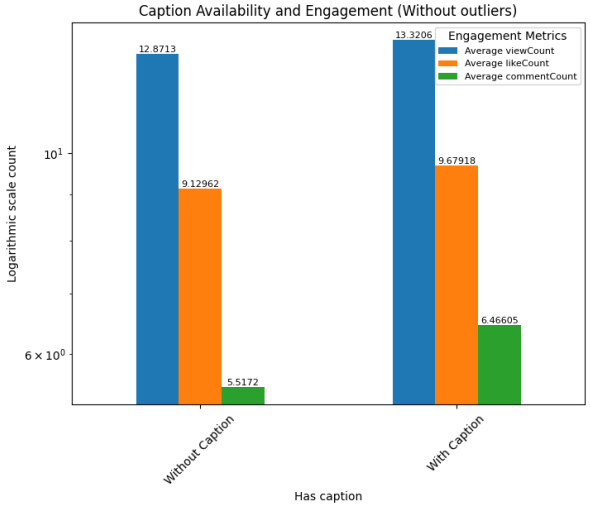
- **We have used 'Channel's upload rate', 'Subscriber count' and 'Channel's age' to predict the number of trending videos of a particular using "linear regression model".**

Regression summary:

- R-squared Value: 0.3414
    - An R-squared value of 0.3414 means that approximately 34.14% of the variance in trending_videos_count (our dependent variable) can be explained by the independent variables subscriberCount and upload_rate.
    - **Implication:** This is a moderate level of explanatory power. It suggests that while our model does capture some of the variability in the number of trending videos, a significant portion of the variability is explained by other factors not included in your model.
- Coefficients: [1.19925416e-08, 5.72320143e-05, 1.11637335e-02]
    - The first coefficient 1.19925416e-08 corresponds to subscriberCount. This value indicates that for each additional subscriber, the number of trending videos is expected to increase by approximately 1.19925416e-08, assuming upload_rate and channel_age remains constant.
    - The second coefficient 5.72320143e-05 corresponds to upload_rate. This value indicates that for each additional video uploaded per year (increase in upload_rate), the number of trending videos is expected to increase by approximately 5.72320143e-05, assuming subscriberCount and channel_age remains constant.
    - The third coefficient 1.11637335e-02 corresponds to channel_age. This value indicates that for each additional year (increase in channel_age), the number of trending videos is expected to increase by approximately 1.11637335e-02, assuming subscriberCount and upload rate remains constant.
- Intercept: 1.0244333175741462

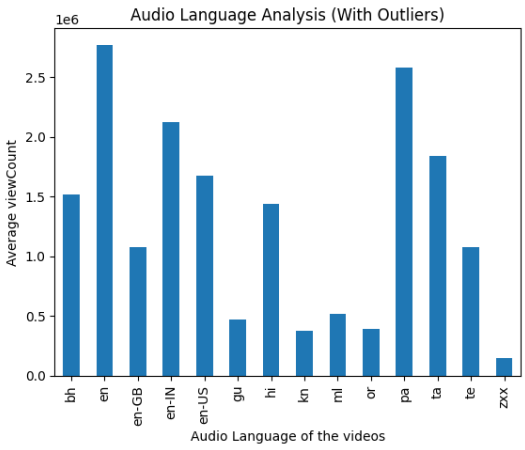
➢ Does caption availability influence a video's likelihood of becoming trending?

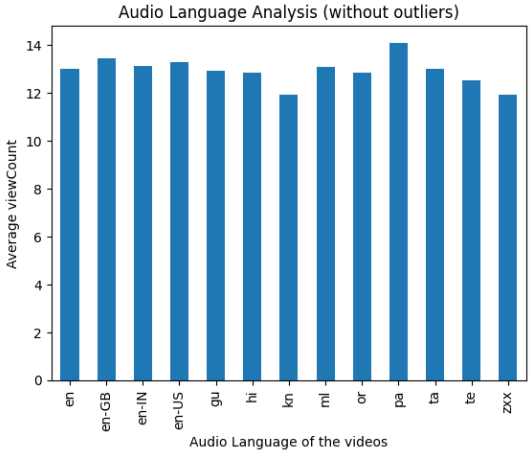




When we tried to observe the general trend by removing the outliers we could observe that caption availability does has a very less influence the engagement metrics

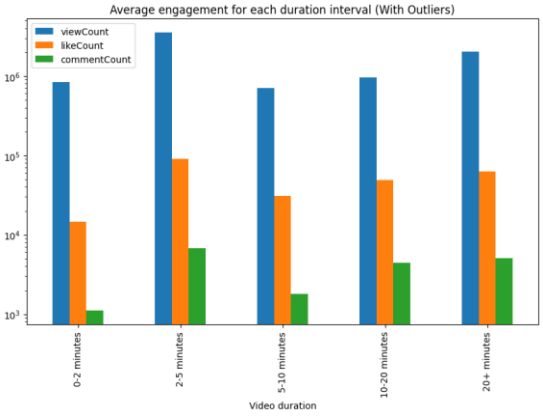
➢ How does the default audio language of a video impact its success?

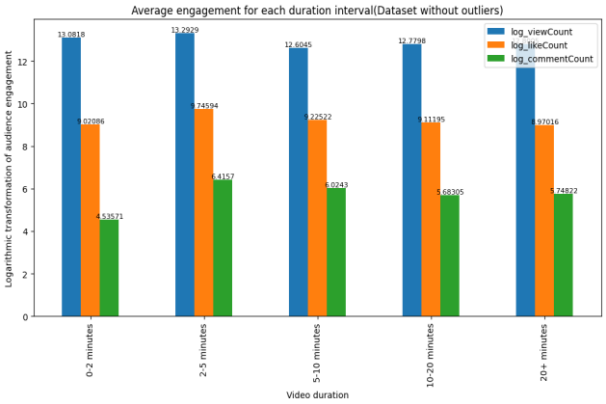




We can observe a huge difference after removing the outliers, for exceptional cases (with outliers) we observed that the videos with the audio language 'English' has the highest Viewcount, whereas for general trend (without outliers) we can observe the audio language has very less influence on the Viewcount

- **What is the ideal length of a video for reaching success?**





We can say that videos with duration less than 5 minutes has a slightly more viewer engagement than others

Conclusion:

This study explored various factors that influence the success of YouTube videos and channels, using trending video data collected via the YouTube API. Our analysis covered both video-specific and channel-specific metrics, focusing on the relationship between engagement metrics (such as views, likes, and comments) and variables like video duration, the number of tags, and caption availability.

Key findings from the analysis include:

1. **Viewer Engagement**: A moderate positive correlation was found between view count and other engagement metrics like likes and comments, suggesting that higher view counts tend to be associated with increased interaction.
2. **Video Duration**: While the duration of a video did not show a strong correlation with viewer engagement, shorter videos (less than 5 minutes) tend to have slightly better engagement.
3. **Tags**: The number of tags used in a video did not exhibit a significant impact on its engagement metrics. However, certain common tags like “Entertainment”, “Drama”, “Trending”, “Latest” etc. were identified as being more frequently used in trending videos.
4. **Outliers Impact**: Outlier removal provided a clearer picture of general trends, though it did not drastically change the overall conclusions. The presence of outliers mostly affected the perceived strength of correlations.
5. **Channel Growth**: A moderate correlation was observed between a channel's upload rate and its subscriber count, indicating that more frequent uploads might help in attracting more subscribers. Additionally, a channel’s age showed little to no correlation with the number of trending videos, suggesting that newer channels can compete with older ones in achieving viral content.
6. **Language and Captions**: English audio language was associated with higher view counts, although this effect diminished after outlier removal. Caption availability was found to have a minimal impact on video engagement.
7. **Predictive Modeling**: A linear regression model predicting the number of trending videos based on subscriber count and upload rate explained about 34% of the variance, highlighting that while these factors are important, other unexamined variables also play a significant role in video success.

In conclusion, while some factors like upload rate and subscriber count contribute to a channel's success, the complex and multifaceted nature of YouTube's platform means that content creators must consider a variety of strategies to optimize their visibility and engagement. This analysis provides a foundational understanding, but further research incorporating additional variables could offer deeper insights into the dynamics of YouTube virality and channel growth.