

Avinash Bunga

Information Systems and Business Analytics, Park University

CIS607HOF1P2024 Applied Business Forecasting

Professor: Dr. Abdelmonaem Jornaz

Aug 29, 2024

Unit 3: Discussion

How Video Length and Quality Influence Viewer Engagement Metrics

Introduction

In the digital age, understanding the relationships between various data points is crucial for making strategic decisions, particularly in content creation and marketing. On platforms like YouTube, where video content competes for viewer attention, knowing how factors like video length, quality, and engagement metrics interact can provide valuable insights for optimizing content strategies. This discussion explores two such relationships using data from the YouTube Trending Videos dataset obtained from Kaggle. This dataset includes various attributes such as video length, view count, like count, and video quality, offering a rich source of data for analyzing viewer behavior and engagement trends.

The first example examines the impact of video length on viewer retention, measured by view count, while the second investigates how video length and quality together influence viewer engagement, measured by like count. Both examples highlight whether these relationships are linear or non-linear and suggest appropriate regression models to capture these dynamics. The knowledge gathered from these analyses can help content creators, and marketers understand viewer behavior more deeply and tailor their strategies for better engagement (Ali, n.d.).

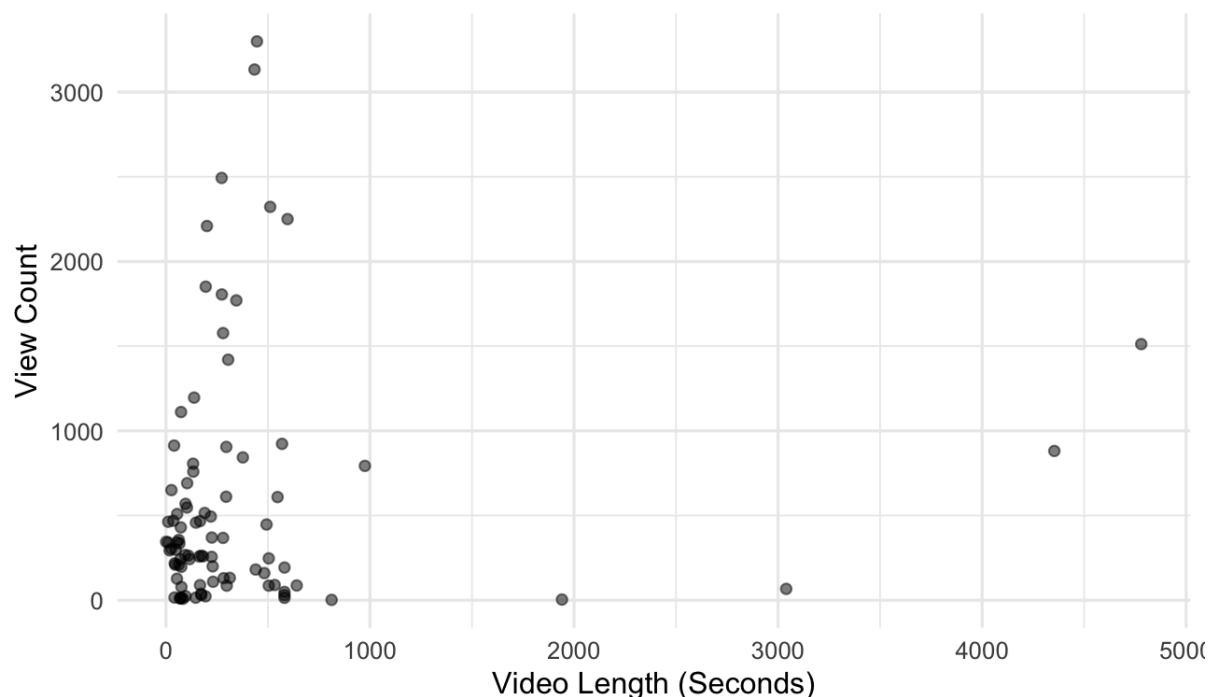
Example 1: Relationship Between Video Length and View Count

The first example focuses on the relationship between **video length** (durationSec) and **view count** (viewCount). This relationship is significant for content creators and marketers looking to maximize engagement.

Analysis:

A scatter plot was created using R to visualize the connecting between **video length** on the x-axis and **view count** on the y-axis (see Figure 1). The plot shows that shorter videos (around 0-200 seconds) generally have much higher view counts compared to longer ones. As the video length increases, there is a noticeable drop in views, suggesting that viewers are more likely to engage with shorter videos and tend to lose interest in longer content (Dean, 2017; Datacamp, 2024).

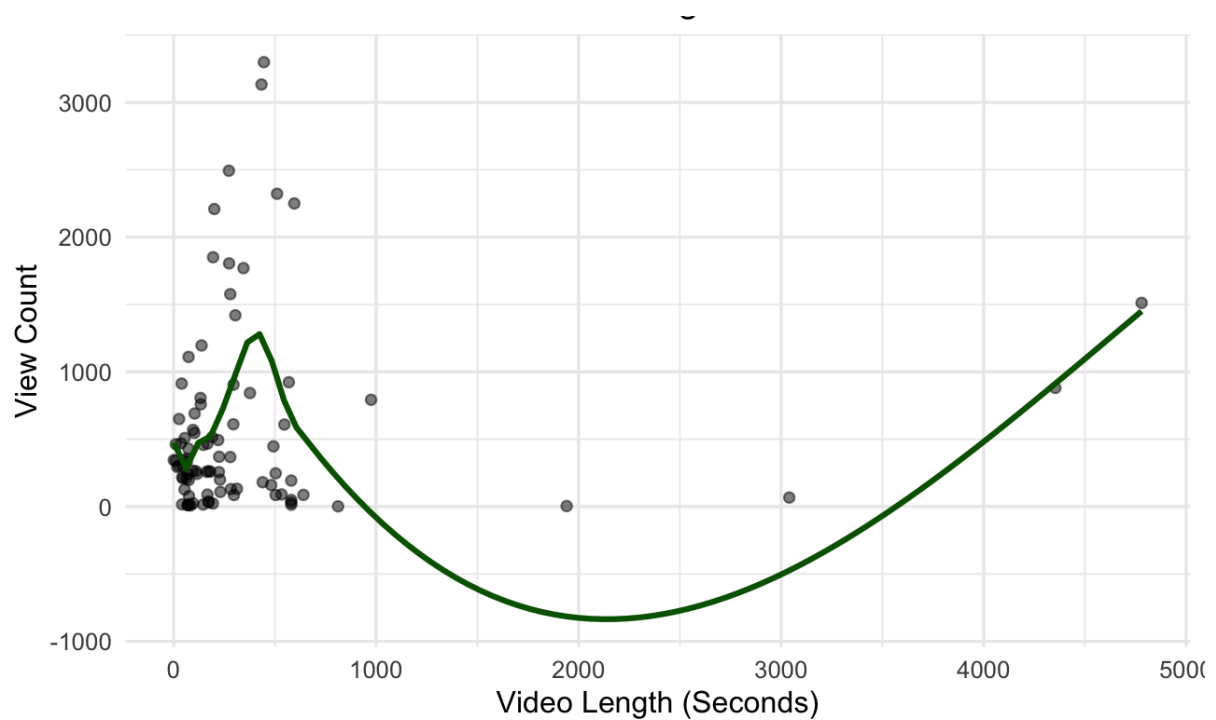
Figure 1: Scatter Plot of Video Length vs View Count



The scatter plot highlights how the connection between video length and view count is **non-linear**. A linear regression line would not effectively capture the rapid decline in view counts as video length increases since it assumes a steady rate of change across all video lengths. Instead, a **LOESS smoothing curve** was applied to better represent the data (see

Figure 2). This curve is more flexible and closely follows the actual pattern of the points, displaying a sharp rise in views for short videos and a steep decline as videos get longer (Bobbitt, 2022; Datacamp, 2023).

Figure 2: Scatter Plot with LOESS Smoothing Curve for Video Length and View Count



Nature of the Relationship:

This non-linear trend can be explained by several factors related to viewer behavior. Generally, Shorter videos are easier to watch through completely and more engaging, leading to higher view counts. Shorter content is also more likely to be shared, recommended, and promoted by platform algorithms, which further boosts visibility and views. Conversely, as videos get longer, fewer viewers are willing to commit to watching the entire content, causing a rapid drop in view counts (Datacamp, 2023).

Suggested Regression Curve:

For this data, a **LOESS smoothing curve** is the most effective regression model. The curve begins with a high point at shorter video lengths, indicating high view counts, and then

sharply declines as the video length increases. After this steep drop, the curve flattens out, suggesting that view counts remain consistently low for longer videos. This model adapts well to the nuances in the data, making it the best fit to represent the relationship (Bobbitt, 2022; Datacamp, 2023).

Conclusion for Example 1:

The relationship between video length and view count on YouTube is non-linear and is best captured by a LOESS smoothing curve. This model effectively illustrates the initial surge in views for shorter videos and the subsequent decline as videos get longer. These insights are valuable for optimizing content strategies to enhance viewer engagement.

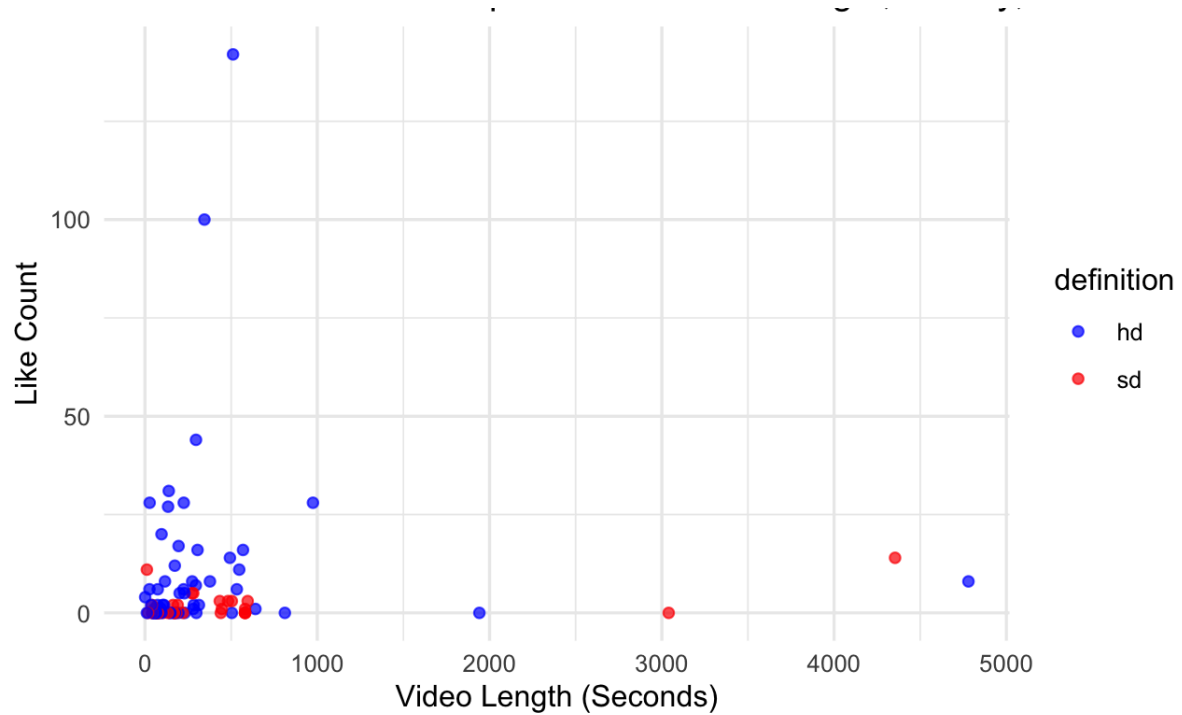
Example 2: Relationship Between Video Length, Video Quality, and Like Count

For the second example, the combined impact of **video length** (durationSec) and **video quality** (definition) on **like count** (likeCount) is examined. Understanding how these elements interact is essential for optimizing video production and strategy on platforms like YouTube (Ali, n.d.; Dean, 2017).

Analysis:

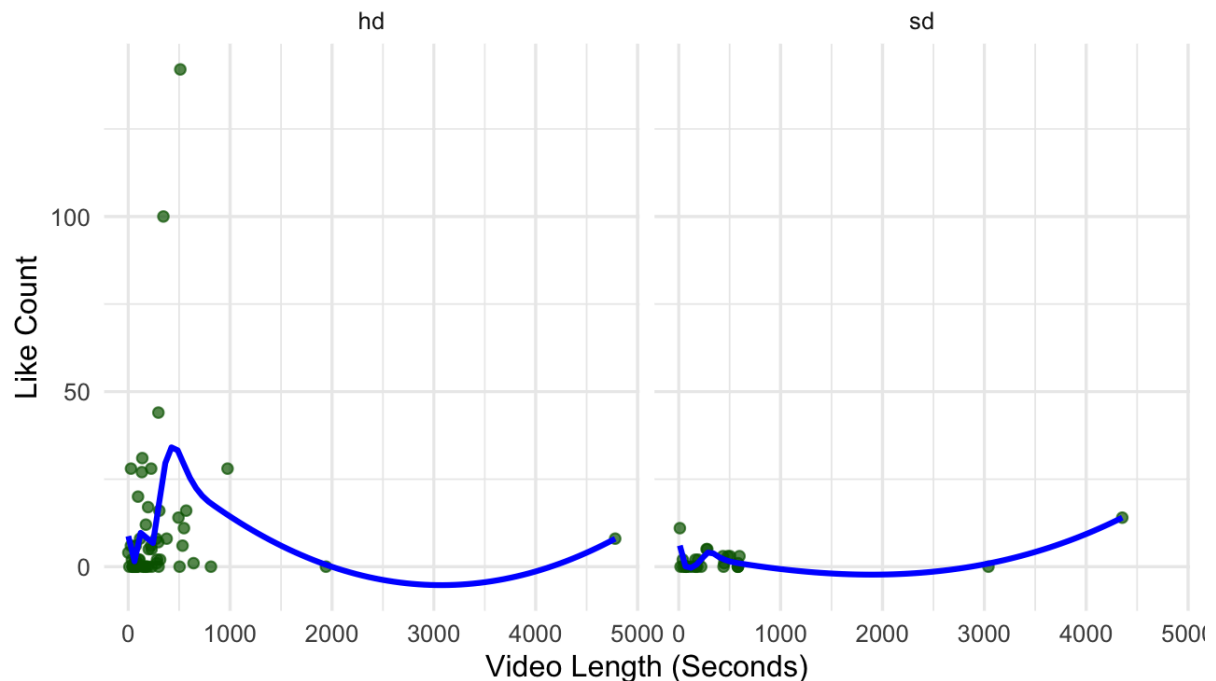
To explore this relationship, a scatter plot was created with **video length** on the x-axis and **like count** on the y-axis, where points are color coded by **video quality, high definition (hd)**, and **standard definition (sd)** (see Figure 3). The plot reveals that HD videos (blue points) tend to receive more likes than SD videos (red points), especially when the videos are shorter. HD and SD videos cluster like counts at shorter lengths, with very few videos extending to longer durations (Dean, 2017; Datacamp, 2024).

Figure 3: Scatter Plot: Relationship Between Video Length, Quality, and Like Count



For a clearer view of these patterns, a **faceted scatter plot** was used to separate the analysis for HD and SD videos (see Figure 4). By incorporating a **LOESS smoothing curve** in each facet, the non-linear relationships within each video quality category become evident (Bobbitt, 2022; Dean, 2017; Datacamp, 2024).

Figure 4: Faceted Scatter Plot with Smoothing Curve: Video Length and Like Count by Video Quality



The curve for **HD videos** (left panel) shows a noticeable **peak** in like counts for shorter videos (under 200 seconds), followed by a **sharp decline** as the video length increases. This suggests that viewers are more engaged with shorter HD videos, likely due to the combination of high-quality visuals and concise content. After 1000 seconds, the curve flattens out and shows a minor increase for some longer videos, though the overall trend indicates higher engagement for shorter HD videos.

For **SD videos** (right panel), the smoothing curve is relatively flat, showing consistently low like counts regardless of video length. This pattern suggests that lower video quality significantly reduces viewer engagement, regardless of video length.

Nature of the Relationship:

The relationship between **video length**, **video quality**, and **like count** is clearly **non-linear**. HD videos are more likely to receive higher likes, especially when they are shorter, reflecting an optimal balance between quality and length. In contrast, SD videos consistently show lower engagement, implying that video quality is a more decisive factor for attracting likes than the video's length (Datacamp, 2023).

Suggested Regression Model:

To capture these interactions more precisely, a **multivariate regression model** that includes interaction terms between **video length** and **video quality** would be suitable. Such a model would effectively quantify the independent effects of these variables and their combined impact on like counts. While the **LOESS smoothing curves** provide a good visual representation, a formal multivariate model would offer more detailed insights (Bobbitt, 2022; Dean, 2017; Datacamp, 2024; Statsdirect, n.d.; UCLA, n.d.).

Conclusion for Example 2:

Overall, video length and quality significantly affect like counts on YouTube, with their impact being non-linear and dependent on their interaction. HD videos generally perform better in terms of likes, especially when shorter, while SD videos experience lower engagement across the board. Using a multivariate regression model could provide a deeper understanding of how these factors influence viewer engagement.

Conclusion

The analysis of YouTube Trending Videos data reveals important insights into how video length and quality affect viewer engagement. Both examples illustrate the importance of understanding non-linear relationships in content strategy. Shorter videos tend to capture more views and likes, especially when presented in high definition, reflecting the need for concise, high-quality content to maximize engagement. The findings also emphasize that video quality plays a crucial role in viewer interaction, often outweighing the impact of video length alone. Leveraging these insights, content creators can optimize their strategies to align better with viewer preferences, focusing on the balance between content length and quality to drive higher engagement. Advanced regression models that consider these interactions can provide even more precise guidance for future content strategies (Bobbitt, 2022; Dean, 2017; Datacamp, 2024; Statsdirect, n.d.; UCLA, n.d.).

References

- Ali, M. F. (n.d.). *YOUTUBE TRENDING VIDEOS DATA*. Kaggle. Retrieved August 29, 2024, from <https://www.kaggle.com/datasets/faisaljanjua0555/trending-videos-on-youtube-dataset>
- Bobbitt, Z. (2022, May 17). *How to Perform LOESS Regression in R (With Example)*. Statology. <https://www.statology.org/loess-regression-in-r/>
- Dean, B. (2017, February 18). *Here's What We Learned About YouTube SEO*. Backlinko. <https://backlinko.com/youtube-ranking-factors>
- Datacamp (2023, September 1). *Introduction to Non-Linear Models and Insights Using R*. https://www.datacamp.com/tutorial/introduction-to-non-linear-model-and-insights-using-r?utm_source=google&utm_medium=paid_search&utm_campaignid=19589720830&utm_adgroupid=157156377351&utm_device=c&utm_keyword=&utm_matchtype=&utm_network=g&utm_adposition=&utm_creative=710011990745&utm_targetid=aud-517318242147:dsa-2218886984060&utm_loc_interest_ms=&utm_loc_physical_ms=9052025&utm_content=&utm_campaign=230119_1-sea~dsa~tofu_2-b2c_3-us_4-prc_5-na_6-na_7-le_8-pdsh-go_9-nb-e_10-na_11-na-fawaug24&gad_source=1&gclid=CjwKCAjwuMC2BhA7EiwAmJKRrK972LgKeDmqyp80Y1euzwG8YC6RurQLhlX2sUiu0z9zKgVCNply9BoCMpIQAvD_BwE
- Datacamp (2024, May). *Scatterplot in R*. <https://www.datacamp.com/tutorial/scatterplot-in-r>
- Statsdirect (n.d.). *LOESS Curve Fitting (Local Polynomial Regression)*. Retrieved August 29, 2024, from https://www.statsdirect.com/help/nonparametric_methods/loess.htm
- UCLA (n.d.). *Multivariate Regression Analysis | Stata Data Analysis Examples*. Retrieved August 29, 2024, from <https://stats.oarc.ucla.edu/stata/dae/multivariate-regression-analysis/>

