Evaluating the Effect of the Public Dislike Feature on Viewership

Qichen Liu, Eric Le, Jason Rudianto

Introduction

Over the past decade, watching online videos has become widely popular due to the expansion of online video platforms. Many users have the experience of clicking a “dislike” button when finding a video offensive. YouTube introduced the dislike button in 2010. However, having such a feature has both pros and cons. On the one hand, it helps protect people from fake information and perilous videos, as a large number of social “dislikes” on a video might be interpreted as bad social proof and aid in its removal. On the other hand, the feature exposes creators to potential abuse and harms the platform ecosystem through “dislike attacks”, in which assailants actively work to drive up the number of dislikes on a creator’s videos, according to Susan Wojcicki, CEO of YouTube.

In November 2021, YouTube removed the public dislike number from all of its videos, quickly followed by widespread disagreement concerning the adverse effects of the shift. Nowadays, there are more than 37 million YouTube channels worldwide vying for viewers’ attention. A decrease in user engagement with videos will definitely cause a significant loss of a creator’s revenue, which implies that YouTube’s decision is somewhat unjustifiable. Thus, data-based approaches are required to resolve the dispute over the impact of the feature on user engagement.

This study evaluates the effectiveness of having the public dislike number on the platform empirically, utilizing observations of the top trending YouTube videos each day as of August 2020. The data contains each video’s category information as well as various metrics for user engagement, such as the quantity of views, comments, likes, and dislikes. Applying a set of regression models, we estimate the change of viewership on the second day for videos in each category when the dislike number is public.

Data and Methodology

The data in this study comes from the YouTube Trending Video Dataset. It was collected and made publicly available by the YouTube API sourced from Kaggle. The dataset is a daily record of the top trending YouTube videos, so each unit of observation represents a single trending video recorded from August 2020 to July 2022 with its associated metadata and metrics.

According to the data dictionary, the publication date variable indicates when a video is originally published, while the trending date represents the time when a video is recorded as an observation. The difference between these two dates can be regarded as the age of a video and we split the data into several groups by looking at the age. Because we planned to investigate the viewership on the second day, we filtered out 127195 observations the age of which is larger than one and used the remaining 11.294372%, totaling 16195 rows, to generate the statistics in this report. All exploration and model building were performed on the subsample that was filtered out.

Given the practice in the industry, viewership is the most widely used metric of user engagement. To operationalize viewership, we use the view count as our outcome variable. Meanwhile, we analyze the tag information to count the number of tags and use it as an explanatory variable, but the main variable that we focus on is the indicator for public dislike feature. Because we don’t know the exact date when YouTube removed the dislike number, a video that was active in November 2021 is likely to be published with the public dislike number but recorded without, and thus is not a valid observation following a definite pattern (with or without the feature). We therefore filter out 717 videos of which the trending date is later than November 1 and the publication date is earlier than November 30. There are also 312 duplicated videos in our data, which take up 1.9265205%, as some videos are repeatedly displayed as trending videos on the same date. Finally, we remove 27 outliers whose view count is larger than $3 \cdot 10^7$, leaving 15141 observations. Additionally, we include a fixed effect for each category that is interacted with

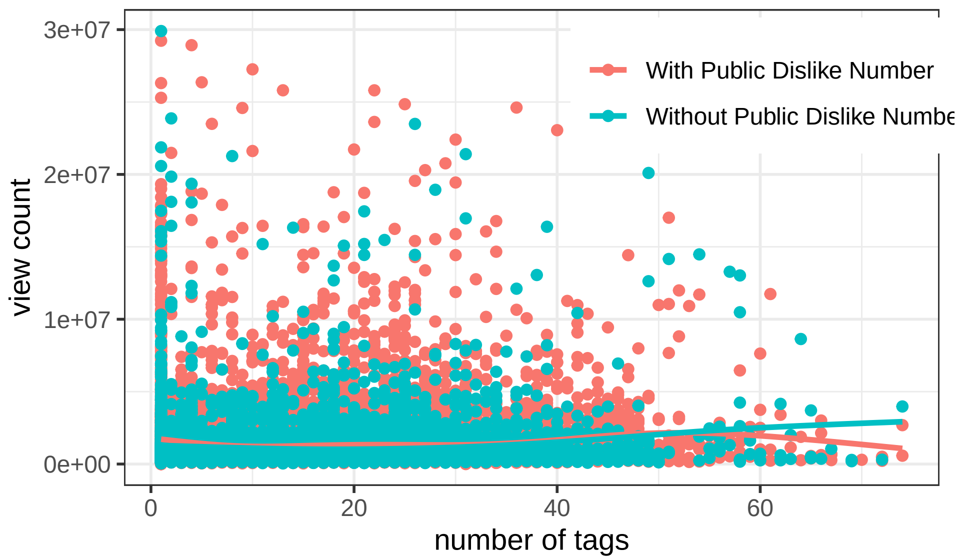

Figure 1: Video Viewership as a Function of Age

I am interested in the difference in value between two counterfactuals: one in which a video has a public dislike button, and another in which it does not. Importantly, we will have to consider the categories these video fall under as they cater to different populations

We are also interested in a number of additional features. These include the number of tags a video has has and the ratio of dislikes a video has. We suspect that more tags should lead to more views as that improves the searchability and relevance of a video to users. Further, we suspect that videos with a high dislike ratio could possibly lead to less views due to it being poor content, but it is also possible that a number would have more views due to it being "controversial".

We represent this overall as the below equation:

$$\widehat{view\ count} = \beta_0 + \beta_1 \cdot D + \mathbf{I}\beta_I + D \cdot \mathbf{I}\beta_D + \mathbf{Z}\gamma$$

Results

As shown above, we see that that the ratio of dislike button and the presence of a dislike button can play a critical role in viewership, however the statistical significance is dependent on the model. We see that the ratio of dislikes by itself are not indicative of whether the viewership for a video would increase. A similar story can be said of whether the dislike button plays a role in increased viewership. However, this regression output importantly leaves out the relationship between categories on each of the above feature. However, we do see that an increase in tags would actually lead to a decrease in total viewership and this is statistically significant.

When including the interaction between categories, we see that each coefficient is statistically significant and that the dislike button's effect varies between which category of video we are analyzing.

Table 1: Estimated Regressions

	Output Variable: total views		
	(1)	(2)	(3)
as.numeric(if_dislike_public)	241,214.70***	−806,943.30	−797,648.40
	(35,743.99)	(481,643.00)	(483,581.60)
ratio.of.dislike		42,698,850.00	42,935,020.00
		(24,381,425.00)	(24,459,659.00)
log(num.tags)			−90,444.31***
			(17,595.46)
Constant	1,172,425.00***	589,290.80***	796,659.30***
	(27,149.36)	(101,739.60)	(110,490.00)
Observations	15,141	15,032	15,032
R^2	0.003	0.04	0.05
Residual Std. Error	2,203,508.00 (df = 15139)	2,159,581.00 (df = 14987)	2,156,959.00 (df = 14986)
Note:	*p<0.05; **p<0.01; ***p<0.001 HC_1 robust standard errors in parentheses. Additional features are number of film/animation videos, people/blogs videos, travel/events videos, howto/style videos, autos/vehicles videos, and pets/animals videos.		

Limitations

Although not explicitly defined in the research paper behind this dataset, we believe that this dataset is not IID. 200 Trending video data was collected during the 2020-2022 year from the same date. Since no other information was given about the data selection process, and even though the chances that videos are related is low, we cannot definitively say that the videos are independent from each other. To error on the side of caution, we evaluate this assumption to be violated.

A BLP exists when the covariances between the variables are finite (finite mean and finite variance). To determine this we can examine the correlation matrix between the input variables and the output variable. We conclude that there was no infinite covariance between the variables. We can examine if the BLP is unique by evaluating if there is perfect collinearity among the variables. Since none of our linear models dropped any input variables, we can conclude that there is no perfect collinearity and that there exists a unique BLP. Therefore the assumption is met.

As far as structural limitations, several omitted variables may bias estimates. In a classic omitted variables framework, the omitted variable is assumed not to interact with the key variable in the true model. There were several variables that were omitted from the model that could have had an effect on the model, whether positively or negatively. We intentionally left out these variables because our research question asks whether the dislike feature can impact viewership. One variable not related to the dislike feature that could have caused significant omitted variable bias is state of the world which indicates whether a video had more viewership due to a significant event. If an event is highly controversial then viewership should reflect that, in most cases. So we can reason that our coefficient for the state of the world variable should be positive. Since both coefficients are positive, we can conclude that the direction of the omitted variable bias is away from zero. Another variable not related to the dislike feature that could cause significant omitted variable bias is fake accounts driving up the amount of viewership for a given video. Therefore we can also reason that our coefficient for the fake accounts variable should be positive. An unintentionally omitted variable could be celebrity appearance which can also increase the likelihood of views on a particular video. We can reason that individuals are more likely to view a video if their favorite celebrities are included, indicating that the coefficient between celebrity appearance and viewership is positive. Since both coefficients are positive, we can conclude that the direction of the omitted variable bias is away from zero. In the future, we can collect data that asks the student how much they enjoy studying that subject.

Conclusion

The goal of our research was to evaluate the relationship between the dislike feature and view count on a particular genre of video. We conclude that the dislike feature variable was not statistically significant in explaining the outcome variable in any of our three models. Based on our findings, we did not find that including the dislike feature increased viewership count.

We found our results to be practically significant and can serve as a foundation for future studies. There were omitted variables that could have impacted viewership. In the future research, new datasets may generate data that incorporate additional input variables and output.