Do viral sounds boost TikTok video performance? Analyzing the influence of TikTok sounds on engagement metrics

Bennett Blanco, Camila Daza, Sofia Lopez-Somohano, Jose Salerno
Group 1

03/07/2024

Introduction

TikTok's popularity has skyrocketed since its release in 2016. With its 1.5 billion users spending over 4.43 billion minutes watching videos, understanding the key factors that make TikTok videos viral has become crucial for businesses and content creators (West, 2024). This experiment explores how adding sound to posts from the platform's Top 10 songs affects audience engagement metrics, by analyzing the views metric. Based on our analysis, we find that within our experiment, song inclusion actually leads to a decrease in viewership per TikTok posting. However, the scope of our experiment is rather limited, and our estimates are noisy. Ultimately, these findings should aim to encourage further and more in depth analyses to find more precise estimates.

Research Question

How does the inclusion of trending sounds affect viewership for a TikTok post?

Hypothesis

The incorporation of a trending TikTok sound into a TikTok post boosts viewership within an hour of posting.

H_0: There is no effect of including viral songs on TikTok post-viewership

H_a: There is an effect of including viral songs on TikTok post-viewership

Literature Review

While comprehensive literature on the workings of TikTok's algorithm is scarce due to its proprietary nature, the platform employs a distinctive and highly personalized recommendation system founded on users preferences and past engagement. This implies that each user's For You Page on TikTok is tailored to their individual interests, ensuring a unique browsing experience for every user [?]. Therefore, it is reasonable to hypothesize that incorporating viral songs into TikTok posts factors into TikTok's algorithm and will influence post-viewership.

Methodology

Participants

In this study, participants were not individuals because of our limited access to TikTok data. Instead, the participant was the unit of randomization, which was a TikTok post on a completely new account that was created for the study. The account's content was focused on a dog, Indiana (Indy), starting with 0 followers and 0 posts.

Experimental Design and Randomization

The experiment consisted of 30 videos of Indy, all with a treatment and control instance, resulting in 60 observations. Data was collected in a six-day window, with postings occurring from approximately 8 am to 10 pm. The order of postings was randomly shuffled at the date and hour level to ensure heterogeneity. Furthermore, the outcome variable is collected at the following hour and before the next post. This is to ensure non-interference and prevent spillover. In addition to randomizing the order, the trending rank of the sound utilized in the treatment was a randomized number from one to ten, to further ensure heterogeneity.

Pre-Experiment Randomization Check

After randomizing both videos and songs, we conducted a randomization check test to ensure proper randomization of posts. We utilized a proportion z-test on our posting schedule data. As anticipated, the proportions test yielded a result of 1.0, indicating that we failed to reject the null hypothesis of proper randomization. In simpler terms, this implies that there is no significant difference between the treatment group and the population proportion.

Sample	% in Treatment	P-Value
Treatment	50%	1.0

Table 1: Pre-Experiment Randomization Check

Measurements

The outcome of interest is the viewership of a TikTok post within the first hour of going live. As previously stated, this window of observation had to be employed to prevent spillover and ensure non-interference. Besides monitoring viewership, we also gathered information on comments, likes, favorites, and shares to track a wider spectrum of audience engagement. This collection spanned from February 27, 2024, to March 4, 2024. While it may not entirely prevent this issue, by posting and collecting data every hour, we could mitigate potential interference with the treatment for users who potentially would navigate to another video through clicking on the profile page. This procedure eliminates the risk of new videos contaminating the treatment for previous videos where viewership was already collected but cannot ensure the inverse.

We ensured that users who viewed one of our posts wouldn't subsequently view another post on the same profile, thus preserving the integrity of our analysis regarding the treatment effect.

Data Analysis

Statistical Methods Used

We performed three main analyses with the data collected: average treatment effect (ATE), regressions, and power analysis.

While the true average treatment effect can never be fully accounted for since we can't observe the effect of the same TikTok post being both part of the treatment and control groups, it's still valuable to estimate and calculate the average treatment effect ($\hat{\text{A}}$TE). This was calculated by taking the difference between the views in the treatment and control group. Understanding the ATE in this experiment is crucial for assessing causality and facilitating comparative analysis.

Following this, we ran two regressions. The first was a simple ordinary least squares regression to find the effect of treatment on views. Then, we conducted a regression with multiple fixed effects. These included date, time, and video to account for the variation these three variables had on views. This was also done to obtain more precise estimates of the treatment effect.

Finally, we measured statistical power to understand the probability of rejecting the null hypothesis when there is a true treatment effect of some size. We also calculated Cohen's D to measure effect size, providing further insights into the practical significance of our findings.

Shadow Banning Assumption

Shadow banning on TikTok is a process where a user's content is quietly restricted or made less visible on the platform, often without any notification, leading to a significant decrease in engagement and reach since the content doesn't show up in other users' For You pages [?]. Our experiment suggested that the application imposed a shadow ban on our posts, as evidenced by the exponential decrease in views. However, given that shadow bans cannot be measured or verified with the information that is available to users, our analysis did not control for its impact on our dependent variable. Therefore, we operated under the assumption that the shadow ban did not affect our outcome. Including shadow banning as a control would reduce noise and provide clearer insights, given its strong correlation with the outcome.

Covariates

It was difficult to include covariates to control for in this experiment, as the observable ones were captured in the fixed effects for the most part. Covariates that would be ideal to control for would be related to how the TikTok algorithm was operating, or what was trending at the time of the post and engagement observation period for the post. Additionally, if we could've randomized on the user level, we would have been able to introduce user-related covariates, but at the post level mostly everything gets absorbed by the video fixed effect.

Results

Difference in Means

From our difference in means, we found that there was no statistically significant difference, as our independent t-test returned a p-value < 0.05, meaning that we could not reject the null hypothesis that there is no effect on post views as a result of our treatment.

Estimate of ATE	-9.1333
t-score (t)	-0.3359
P-value (p)	7.3841e-01

Table 2: Summary of ATE estimation

Regression Analysis

As the difference in means cannot provide an interpretable unit effect or control for other externalities, we performed a regression analysis. From a simple OLS regression, we found a non-statistically significant effect of a 9 view decrease for a given post as a result of imposing the treatment. As this doesn't control for much of the observable variation in our experiment, we employed a multi-way fixed effect model to more accurately observe the treatment effect.

In the multi-way fixed effects model, we fixed for video, as within each video of the dog, there may be certain kinds of content, a variation in video length, or other un-observable effects that could cause variation in the outcome. Hour and date were also fixed for, as there may be unobservable variations within those time windows that may result in different view outcomes. When controlling for these factors, we found that the effect of including a trending sound resulted in a statistically significant decrease of approximately 65 views for a given post, at the 99% confidence interval. However, this is a rather noisy coefficient, with a standard error of approximately 24. Randomization ensures that this standard error is robust to heteroskedasticity.

	OLS	**Multi-Way Fixed Effects**
Outcome	views	views
Intercept	32.933 (22.905)	-
treatment	-9.133 (27.194)	-65.375** (24.189)
video	-	x
day_posted	-	x
time_posted	-	x
R2	0.002	0.912
S.E. type	hetero	hetero
Observations	60	60

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Regression Results

Statistical Power

Using our treatment effect from the multi-way fixed effects model, we calculate a Cohen's D of -0.625. Utilizing this information with the size of our treatment and control, we found that the statistical power for this experiment was extremely low, at 0.6635. Though it is quite low, it makes a lot of sense given the noise in our treatment effect estimate and its high standard error.

Limitations

Throughout our experiment, we encountered six limitations that could possibly be improved upon if we were to extend this experiment for a longer period of time.

Shadow Banning

The first limitation of this study is the potential impact of shadow banning on the observed engagement metrics. Due to the time constraints of the experiment, we had to post more frequently than advised. Specifically, content creators recommend 1 - 4 times a day, and as we greatly surpassed this, we believe this led to shadow banning (Glover, 2024). This potentially influenced the visibility of posts and subsequent engagement metrics, particularly after our first initial posts. Consequently, the observed engagement levels may not accurately represent the true impact of music on audience engagement.

If we were to re-run this experiment, it would be advisable to stick to the recommended guidelines and post less frequently. This could potentially ensure more accurate assessments.

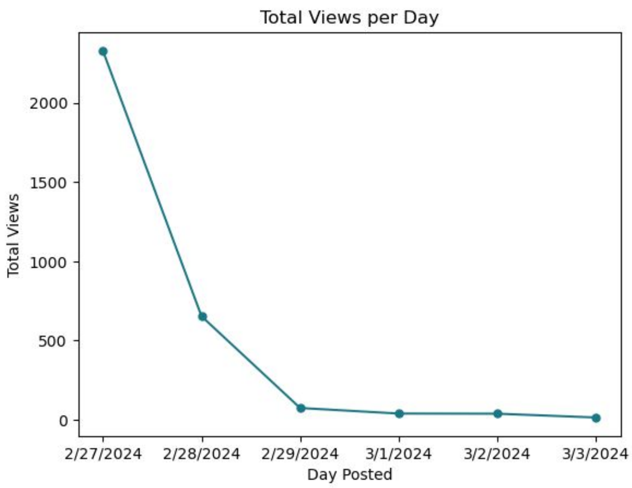

Figure 1: Decrease viewership following the first five posts

Sparse Results For Other Engagement Metrics

Due to the possibility of shadowban, our results for all outcome variables, besides views, were small. Across the duration of our experiment, we observed moderately high views, minimal changes in likes, and no activity in terms of comments, favorites, or shares. This restricted our ability to conduct in-depth analysis regarding the impact of the treatment on these metrics. However, extending the scope and duration of the experiment would provide a more comprehensive understanding of how music influences these engagement metrics.

Limited Posting Hours

Another limitation is the decision to post content during specific hours (8 am to 10 pm) to maintain a normal sleep schedule. By limiting posting to these hours, the study may not capture variations in engagement levels that occur during different times of the day, potentially overlooking valuable insights.

Duration of Experiment and Data Collection

The experiment, posting of TikTok videos, spanned six days (Tuesday to Sunday), which may not fully account for variations in engagement levels across different days of the week. Factors such as weekday versus weekend engagement patterns or specific events occurring on certain days could influence audience behavior and engagement metrics, thus limiting the generalizability of our findings.

Hourly Metrics Collection

Due to the necessity of avoiding metric overlap, hourly metrics were collected after posting. However, this approach may overlook fluctuations in engagement that occur over a longer period of time. Consequently a comprehensive analysis of engagement patterns should include both short-term and long-term metrics to provide a more accurate understanding of audience behavior and content performance.

Spillover

Given that we had no control over TikTok's algorithm, users might encounter an older post on their For You Page and interact with the most recently posted one (the one currently under evaluation). In essence, a user could access our account and encounter both the treatment and control videos we had posted. Therefore, it's crucial to consider users who not only encountered the videos on their For You Pages but also actively sought out more of Indy's content could be included in our metrics.

Limited Content

As a team, our pool of Indy videos was limited, thereby restricting our options for posting. Due to this constraint, we selected thirty videos, which inherently introduced bias into our experiment. Moreover, this constrained content pool implies that the scope of our experiment is confined to posts related to dogs, or more specifically, to posts featuring Indy. Therefore, our findings should not be generalized to all TikTok posts.

During the interpretation of the experiment's findings, the limitations were taken into account. In order to obtain a more comprehensive understanding of the correlation between music integration and audience engagement on TikTok, it is recommended that future research attempts to address these constraints, and the experiment is repeated. By doing so, the results can be further refined and provide a more detailed insight into the relationship between music incorporation and audience engagement on TikTok.

Conclusion

In this experiment, we aimed to investigate the impact of trending sounds on engagement with video posts on TikTok. We believed that including sounds from the app's Top 10 list would significantly improve a video's performance. However, our analysis revealed a more nuanced reality.

To conduct the experiment, we created a TikTok account for a team member's dog named Indiana. We observed the effects of trending sounds on various engagement metrics, such as views, likes, shares, and saves. Despite our best efforts, our findings showed no substantial evidence that our hypothesis was true.

Our statistical analysis indicated that there was no effect on posts due to our treatment. These results can, therefore, not be generalized since there is a lot of noise in the treatment effect estimate. This suggests that the influence of trending sounds on TikTok video posts may not be as straightforward as we had originally hypothesized. However, these findings could still be useful for content creators and businesses by reminding them to consider these nuances when strategizing for engagement, and acknowledging that success on TikTok might require more than using trending sounds. Further a more in-depth research is needed to explore the black box that is TikTok's algorithm, in order to offer clearer guidelines to be able to maximize video engagement more effectively.

References

[1] Geyser, W. (2024, January 30). Everything you need to know about TikTok shadow ban. *Influencer Marketing Hub*. Retrieved from https://influencermarketinghub.com/tiktok-shadow-ban/#:~:text=A%20TikTok%20shadow%20ban%20is,in%20the%20app%27s%20hashtags%20section

[2] Glover, R. (2024, February 20). How often to post on TikTok for the most views & followers (+7 Tips to Post Faster). *WordStream*. Retrieved from https://www.wordstream.com/blog/ws/2023/09/20/how-often-to-post-on-tiktok

[3] Savolainen, L. (2022, March 12). The Shadow Banning Controversy: Perceived governance and algorithmic... *Sage Journals*. Retrieved from https://journals.sagepub.com/doi/full/10.1177/01634437221077174

[4] West, C. (2024, February 20). 27 TikTok statistics marketers need to know in 2024. *Sprout Social*. Retrieved from https://sproutsocial.com/insights/tiktok-stats/

[5] Zeng, J. (2022, February 21). From content moderation to visibility moderation: A case study... *Wiley Online Library*. Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1002/poi3.287

[6] Zhao, Z. (2021). Analysis on the "Douyin (Tiktok) mania" phenomenon... *Analysis on the "Douyin (Tiktok) Mania" Phenomenon Based on Recommendation Algorithms*. Retrieved from https://www.e3s-conferences.org/articles/e3sconf/pdf/2021/11/e3sconf_netid2021_03029.pdf