Unifying Data Science - Causal Inference Analysis

Effects of Social Media Virality on Baseline Engagement Levels

May 2, 2024

Antara Bhide
Jiwon Shin
Matthew Holden
Titus Robin Arun

Overview & Contents

Executive Summary

The potency of social media platforms like Instagram is undeniable, serving as crucial conduits for business-to-consumer interactions. This study focuses on the Instagram platform within the fashion industry, seeking to understand how viral posts influence baseline engagement levels. Given the ephemeral nature of viral content, it is imperative to discern whether these spikes in engagement translate into sustained increases or if they merely represent temporary surges.

Utilizing data from SocialInsider, our analysis delved into the Instagram activities of 20 fashion brands over the past year. We employed a Difference-in-Differences (DiD) analysis on companies paired on the basis of their similarities. This was coupled with linear regression to isolate the impact of viral posts. Viral posts were defined as those exceeding three standard deviations above the mean engagement.

Our findings suggest that while viral posts drive significantly higher engagement levels for the isolated post, they do not establish a new norm in engagement levels; rather, engagement tends to revert to pre-viral levels. The immediate aftermath of a viral event often shows a brief decline in engagement, possibly due to the viral post causing the influx of non-regular visitors who do not sustain their interaction with the brand after the post reaches virality. We would recommend deeper analysis into why there may be an initial dip in the engagement levels after a viral post and then an increase in the engagement levels over time after a viral post. While increased brand visibility and the acquisition of new followers is a likely reason for this increase in engagement over time, an additional possibility that could be researched further is the effect of the viral post potentially impacting the engagement levels on posts that were uploaded in the past.

Over time (20-30 posts after virality) a slight increase in baseline engagement is observable, but the significance of this increase is negligible. The strategic implications of these findings are significant. For brands, especially in the highly visual and dynamic fashion sector, our study suggests that relying solely on virality is not an effective method for gaining consistent content engagement growth.

Introduction

In today's dynamic digital marketing landscape, social media platforms emerge as essential arenas for business-to-consumer (B2C) interactions. Companies don't merely establish social media profiles for visibility; they actively use these platforms to engage with and broaden their audience reach. This digital engagement acts as a direct conduit for broadcasting a brand's narrative, aiming to capture consumer attention in increasingly competitive markets.

The direct link between social media engagement and revenue generation remains complex and often elusive, but the primary intention behind businesses' social media activities is clear: no organization ventures into social media with the aim of limiting its audience. Consequently, understanding the nuances of social media engagement is crucial for businesses, particularly in evaluating the impact of viral content.

Viral posts can dramatically enhance a company's visibility and interactiveness with users on specific posts, yet the subsequent effects of these engagement surges are not well-documented. Does the spike in engagement that accompanies a viral post establish a new norm for audience interaction, or do engagement levels eventually return to their pre-viral standards? Answering these questions is vital for crafting effective social media and SEO strategies, especially when considering paid promotions and content amplification.

This study is designed to provide businesses with strategic insights on harnessing the power of viral content effectively. By exploring whether the fleeting nature of virality can be transformed into sustained engagement advances, we aim to enhance the strategic deployment of marketing resources across social platforms, potentially reshaping how businesses approach content creation and promotional activities.

Research Question

For our causal inference analysis, we specifically examined the characteristics of marketing campaigns (posts) within the Fashion industry on Instagram. This platform was chosen due to our earlier exploratory analysis, which identified TikTok and Instagram as the most popular platforms for B2C marketing across industries. Identifying potential areas for improvement in marketing efforts led us to focus on Instagram.

➢ Does a viral Instagram post in the fashion industry consistently boost baseline engagement levels, or is it just a temporary spike?

Data Source

Since companies consider their marketing data confidential assets, we relied on third-party data from SocialInsider, a Romanian company established in 2017, specializing in social media analytics. Their web-based solutions offer users comprehensive raw data on marketing performances across various social media platforms, including TikTok, Instagram, LinkedIn, Facebook, Twitter, and YouTube, either on different timeframes or per-post basis. While their API generally functions effectively, it occasionally lacks complete raw data, necessitating a validation check for data integrity. Our team diligently monitored data for our analysis of target brands.

For our causal analysis, we focused on the Instagram accounts of 20 brands, examining data such as posts, engagement, reach, impressions, and more over the past year, as this was the maximum data range available from SocialInsider. To ensure accuracy, we utilized predefined metrics provided by SocialInsider, such as follower count, number of posts, and engagement levels, avoiding estimations like estimated reach or impressions. Instead, we prioritized factual features such as post types and content without relying on estimated numerical values.

Key Definitions and Metrics

The definitions of the terms used in this report are provided by our data source 'Social Insider'. Here are the explanations:

1. **Engagement:** The total number of likes and comments received organically for the posts published during the specified period.
2. **Engagement Rate/Post:** The average organic engagement per post, calculated by dividing the total engagement of the posts by the number of followers the profile has, and then multiplied by 100.
3. **Follower Count:** The total number of individuals who are following the profile.
4. **Posts:** The total number of posts published during the specified period.
5. **Baseline:** The range within 2 standard deviations of the mean square root engagement.
6. **Virality:** Engagement that exceeds 3 standard deviations from the mean square root engagement.

Methodology & Design

Data Preparation, Operational Definitions and Analytical Techniques

To explore the impact of virality on baseline social media engagement, we employed a Difference-in-Differences (DiD) analysis combined with linear regression techniques. The DiD method is particularly effective for estimating causal effects where the intervention is a viral post. This technique facilitates the isolation of virality's causal impact by comparing engagement changes before and after a post goes viral against a control group without such virality spikes.

1. Defining Virality and Baseline Engagement

Virality was quantitatively defined as any post with engagement levels exceeding three standard deviations above the mean, based on normalized data. To effectively measure baseline engagement, we applied logarithmic and square root transformations to the engagement data. While logarithmic transformation was initially used to reduce skewness and stabilize variance, it somewhat obscured inherent viral trends. We thus opted for square root transformation which maintained a balance between normalization and visibility of trends, crucial for tracking subtle changes in engagement due to virality. After square root transformation we applied a Z-score transformation to ensure the response variable scale was consistent across all companies. In effect, the units of this analysis are standard deviations of square root engagement levels.

2. Data Segmentation and Matching

To accurately assess the impact of virality, independent of confounding variables of each company, we segmented the data by matching pairs of companies and dividing the matched pairs into control and treatment groups. This was done to ensure that companies compared were of similar scales in terms of engagement. To align companies for comparison, we focused on matching pairs based on key variables such as follower count and average post frequency, which are indicative of their visibility and activity on social media. This careful matching ensured that differences observed in our analysis could more confidently be attributed to virality effects rather than disparities in audience size or posting habits.

3. Data Preparation and Window Creation

Each viral post treatment and equivalent control was grouped into a “window”. This window consisted of ‘n’ posts before and ‘n’ posts after the viral post. The variable ‘n’ determined the number of consecutive posts required to establish a baseline. The baseline was defined as remaining within 2 standard deviations of the mean. This approach enabled a comprehensive analysis of virality by matching similar windows from similar companies where the treatment window was the concatenation of a baseline-viral-baseline, and the control was the concatenation of two consecutive baselines. The control and treatment windows were taken from the matched pairs of companies instead of from a single company to ensure there was no data leakage. The window sizes were varied, ranging from 5 to 40 posts, to determine the optimal span that provides a comprehensive view while maintaining the practicality of sequence availability within our data set. A 20-post window, we found, was effective in capturing the engagement trend context around viral content.

4. Balancing and Matching Treatment and Control Windows

A crucial final step in our data preparation was to ensure an equal number of treatment and control windows. This balancing act was necessary to minimize any baseline differences between the groups, thereby enhancing the robustness and reliability of our subsequent regression analysis. Between the control observations and treatment observations found across all companies, equal number of windows was taken from each of the matched pairs of companies to minimize baseline differences and maximize generalizability across companies of different scales.

In conclusion, our methodology, through rigorous data handling and innovative analytical techniques, was designed to capture and quantify the effects of virality on social media engagement comprehensively minimizing baseline differences.

Results

Causal Inference and Treatment Effects

Our causal inference analysis applied a difference-in-differences approach combined with linear regression modeling to assess the impact of viral posts on baseline Instagram engagement. The study particularly focused on the coefficients derived from viral posts to determine their average treatment effect on engagement metrics. The following predictors were included in the

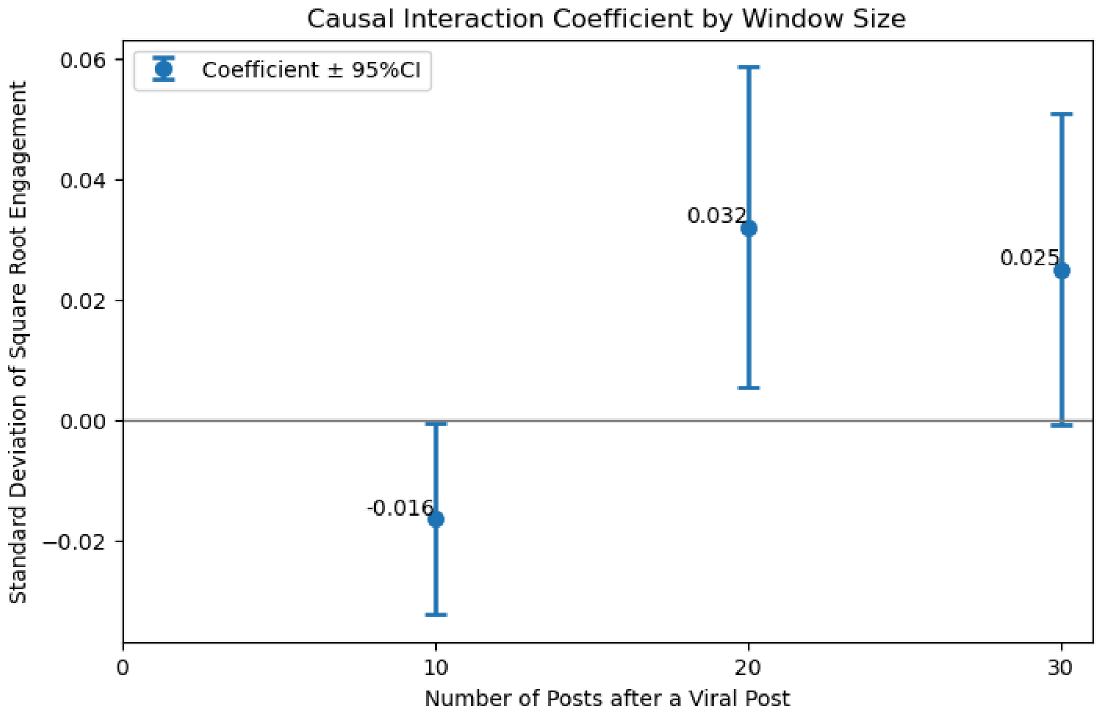


Figure 1: Resultant Effect of

Analysis of Coefficient Estimation by Window Size

- 10 Post Window: Shows a negative coefficient, suggesting that immediately following a viral post, there is a decrease in engagement.
- 20 Post Window: Displays a relatively positive coefficient with a significant confidence interval. This indicates that the effect of a viral post becomes positive as more time passes.
- 30 Post Window: Similar to the 20-post window, the effect remains positive showing consistent results after the short term window immediately after the viral post.

This pattern suggests a peculiar phenomenon where the immediate aftermath of a viral post may involve a drop in engagement. However, this drop may not be an indication of a reduction in engagement post-virality, but instead an artificially elevated engagement rate immediately prior to the viral post. When a post goes viral, there's a high influx of visitors directed to the profile's page. This unusual activity is naturally directed towards the posts at the top of the profile - the posts immediately before the "current" viral post. This behavior would artificially inflate the posts before the viral post and provide a misleading reduction in engagement immediately after the viral post, when in fact engagement levels return to normal. However, over time, as the window expands, the increased exposure is shown to lead to a stabilization and slight increase in baseline engagement, likely driven by a mix of new followers and heightened visibility.

Limitations

- Limited Data Availability: The analysis was constrained by the availability of data, as only a maximum of one year's worth of data was accessible. This posed a challenge when we needed enough data to observe trends. This problem was exacerbated when the minimum required posts to establish a baseline reached higher levels.

- Variability in Marketing Strategies: Each company employs unique marketing approaches, including the creation of multiple viral posts in a small timeframe. This variability complicates the determination of which viral posts have the most significant impact on baseline engagement levels. Our methodology grouped viral posts together and found the per-viral-post effect when multiple viral posts were found between baselines. However, this approach does not account for the content for each viral post.

Conclusion

Virality's Impact on Engagement

Our analysis suggests that virality, characterized by extreme spikes in engagement, has a slight positive impact on social media baseline engagement rates when averaging, at minimum, 20 posts after the viral post. However, the practicality of this increase is not as clear. When looking at the value of the increase (0.032 standard deviations of square root engagement) the data indicates that while viral posts do elevate visibility and interaction momentarily, they do not alter the baseline engagement levels. This finding aligns with our hypothesis that virality might not consistently benefit overall engagement strategies. Thus, it appears suboptimal to focus marketing strategies solely on achieving virality.

Observations on Post-Viral Engagement Trends

A notable observation from our study is the immediate post-viral engagement trend, which often shows a downward shift. This phenomenon could be attributed to an anticlimactic drop following the heightened activity during viral events. Often, viral posts might attract transient visitors whose interactions spike initially but do not translate into lasting engagement. This surge in activity could elevate the pre-viral engagement baseline, by directing new visitors to the posts immediately before the viral post. This would make immediate subsequent engagement levels appear diminished. Such patterns suggest that while viral content captures attention, it does not necessarily foster sustained interaction, possibly due to the transient nature of the audiences it attracts.

Strategic Recommendations

Given these insights, we recommend that brands, especially those in the dynamic landscape of fashion marketing on social media, should not prioritize virality as a core strategy. According to additional research in this field, focusing on consistent content delivery and maintaining a steady engagement rate is more beneficial. The algorithms governing visibility and engagement on platforms like Instagram tend to favor consistency, which can, over time, enhance engagement more reliably than sporadic viral hits. See references for more details.

Furthermore, brands should consider diversifying their content strategies to include a mix of engagement tactics that do not rely solely on virality. These might include:

- Engaging directly with followers through regular interactions.
- Utilizing data-driven insights to tailor content to audience preferences.
- Implementing scheduled postings during peak engagement times.

Closing Thoughts

In conclusion, while viral posts can serve as exciting highlights in a social media strategy, they do not necessarily establish a durable foundation for ongoing audience engagement. Brands should aim for a balanced approach that cultivates a dedicated following through consistent, high-quality content rather than chasing the fleeting spikes of viral fame. This strategy not only enhances engagement but also builds a loyal community that supports sustained digital marketing success.

Reference

(1) Conversion Rate Statistics Across Social Media Platforms

"Disparity between platforms such as Facebook, Instagram, and LinkedIn, with specific figures ranging from 0.54% to 9.21%. These statistics underscore the challenge marketers face in optimizing social media strategies to maximize conversion rates effectively."

Source: "Conversion rates across social platforms," Social Directions Agency Blog, Published 4 months ago. [LINK]

(2) Social Media Marketing Budget Allocation (CMO Survey - Duke University)

"Marketing expenses account for what percent of your company's overall budget?" Slide 18.

Source: The CMO Survey, Deloitte LLP, Duke University's Fuqua School of Business, and the American Marketing Association. [LINK]

(3) Why Content Consistency Is Key To Your Marketing Strategy(Forbes)
"While posting on social media or your blog sporadically might capture the attention of some customers, it will not help you develop meaningful relationships with your audience. To fully connect with your audience and hit all of the necessary touch points, you need to be a consistent content creator"
LINK

(4) Why Consistent Engagement on Social Media is Essential for Business Marketing (Medium.com)
"Posting consistently on social media can increase engagement and reach."
LINK

Appendix

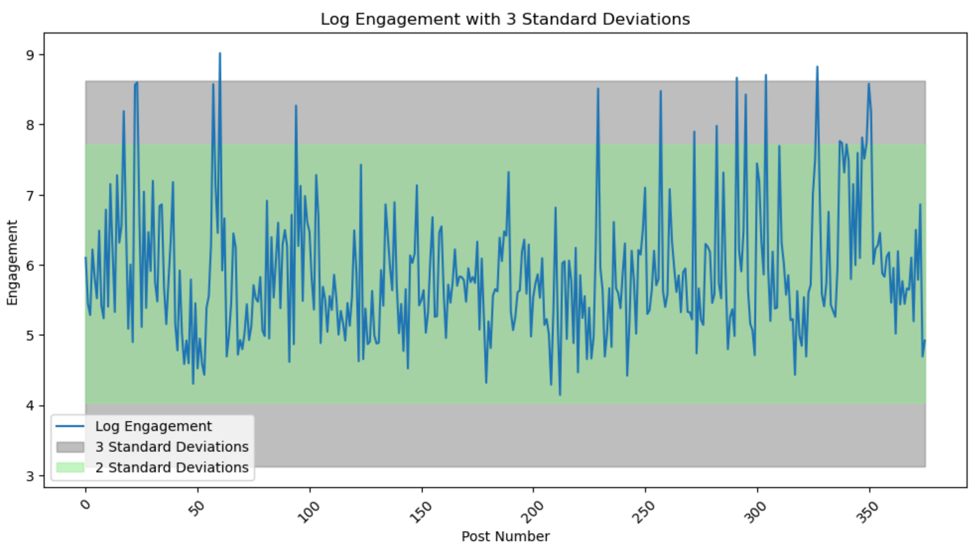
Appendix 1: Methodology Implementation

A. Data Handling and Transformation Techniques

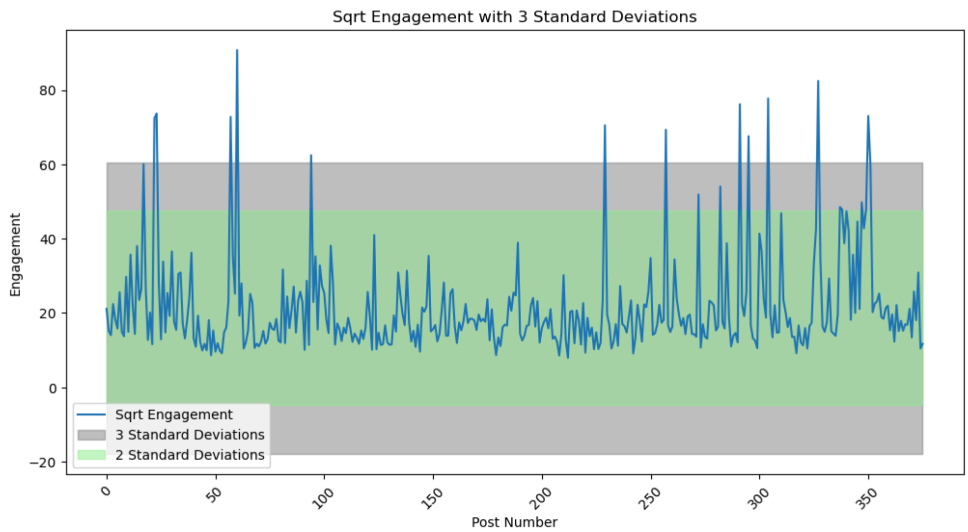
The detailed data handling steps involved reading and preprocessing data from Social Insider, followed by various transformations to stabilize variance and facilitate easier identification of trends and outliers. We explored multiple methods of determining baseline engagement and defining virality thresholds, which are crucial for our DiD setup. These included:

Data Transformations: We applied logarithmic and square root transformations to the raw engagement data, along with a final Z-Score normalization.

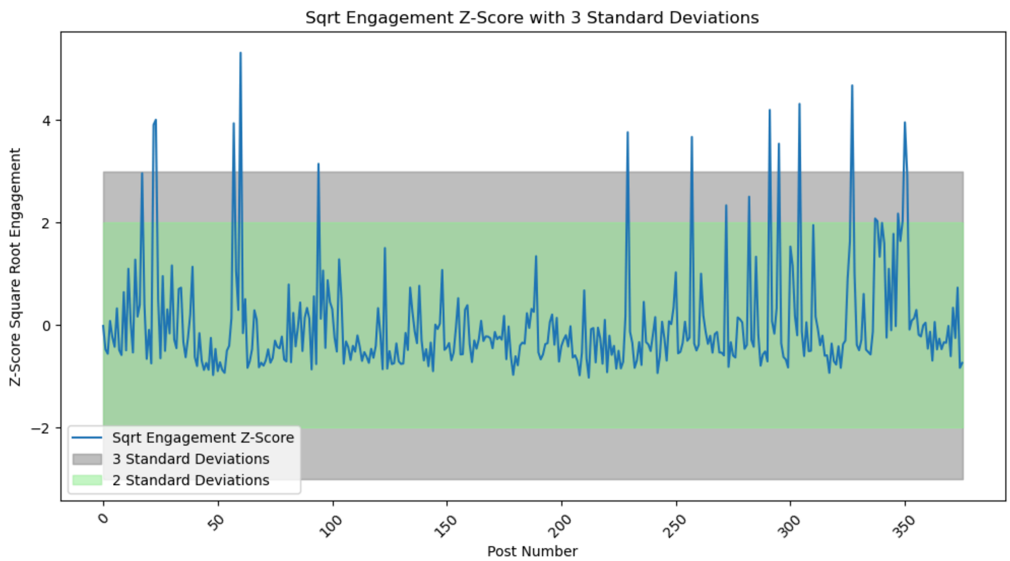
- **Logarithmic Transformation:** This transformation helps stabilize variance in data with a wide range. It also makes the data more amenable to statistical techniques that assume normality. However, this approach can obscure certain data characteristics, such as smaller fluctuations that are significant in engagement analysis.



- **Square Root Transformation:** This approach was found to be more suitable for our data as it adequately normalized the data while preserving more of the inherent variability than the logarithmic transformation, making it ideal for identifying subtle changes in engagement trends. However, the for square root engagement was still very different across companies.



- **Z-Score Square Root Transformation:** The Z-score square root transformation not only stabilized the variance of the data, but also normalized the scale across all companies. This transformation standardizes the engagement levels across all companies, making the results more interpretable in terms of standard deviations rather than raw engagement levels.



B. Baseline Engagement and Virality Detection Methods

A series of methodological tests to determine the most effective way of measuring baseline engagement and identifying virality.

1. **Baseline Engagement Levels:**

   Method 1: Defined baselines as periods where engagement levels remained within

±X% of a moving average calculated over N posts for D consecutive posts.

Method 2: Utilized rate of change in engagement, ensuring it remained lower than Y% of total engagement for M consecutive posts.

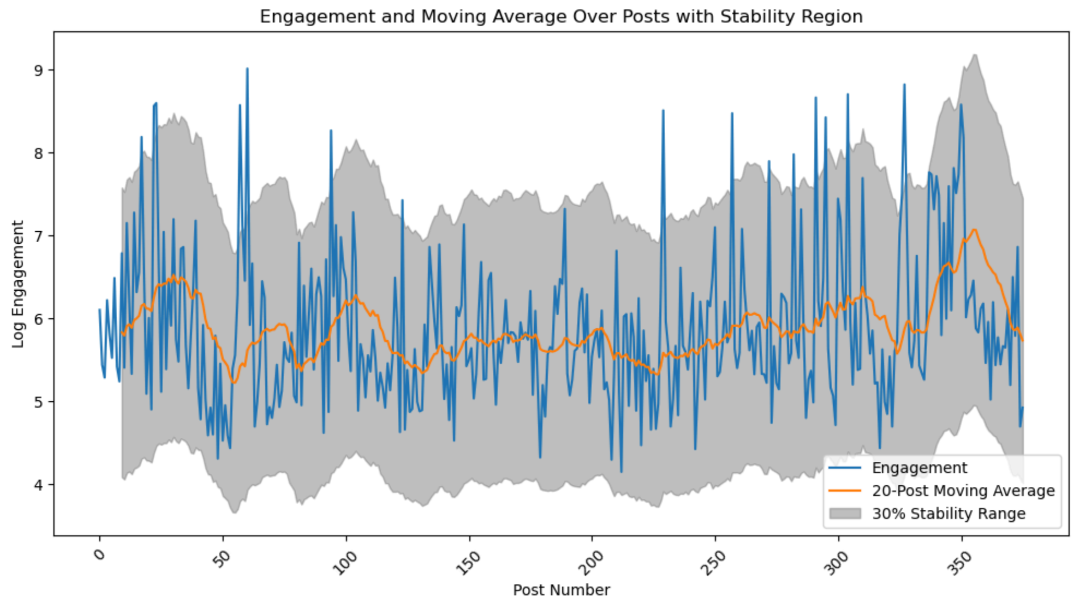
Method 3: Engagement levels must remain within S standard deviations from overall mean engagement level.

2. **Virality Detection:**

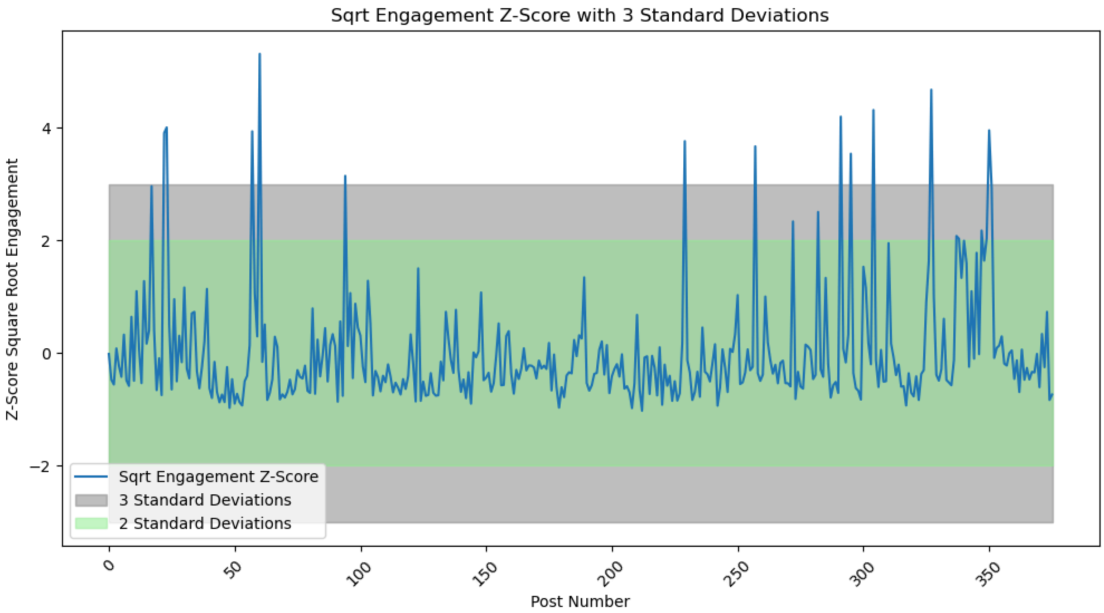
   Method 1 (Multiplicative Approach): Defined virality as instances where engagement reached W times the baseline average.

   Method 2 (Standard Deviation Approach): Considered posts as viral if their engagement exceeded S standard deviations beyond the baseline mean.

Each of these methods was robustly analyzed, including combinations of the data transformations as well. An example of Log-Transformed data using Method 1 baseline calculations with a 20 post moving average and 30% stability range is shown below.



Final Decision: The standard deviation approach using z-score square root-transformed data was selected. This method was found to strike the best balance between sensitivity to changes in engagement levels and robustness against random fluctuations, thus effectively highlighting unusually high engagement levels indicative of virality. This method involved calculating the mean and standard deviation of engagement over a specific baseline period. Posts then were flagged as viral if their engagement levels exceeded the mean by at least three standard deviations, a statistical threshold that empirically minimized false positives while effectively capturing significant spikes in engagement.



C. Data Segmentation and Matching Process

The segmentation and matching of data were crucial for ensuring the validity of our comparative analysis between different entities. Here's a detailed overview on the approach:

- **Categorization:** Companies were categorized based on engagement levels into high, medium, and low engagement tiers. This categorization helped in matching companies within similar operational scales.
- **Criteria for Matching:** The primary criteria included follower count ranges and average engagement rates. Companies within similar ranges were considered for pairing.
- **Algorithmic Pairing Using the Hungarian Method:** We applied the Hungarian algorithm to optimize the matching process. This algorithm calculated the minimal total distance between potential pairs based on their engagement and follower metrics, ensuring that each company was paired with the most similar counterpart.

For each matched pair, we conducted a parallel analysis to measure the impact of viral posts. This involved tracking engagement changes pre- and post-virality within each pair, thus enabling a controlled comparison across different segments.

D. Data Preparation and Window Creation

Data was sorted chronologically to maintain the sequence of posts.

1. Defining Windows: To effectively analyze the impact of virality on engagement, we defined 'windows' as sequences of 'n' consecutive posts. This method helps in understanding the engagement pattern before and after a viral event.

2. Selection of Window Sizes:
    - Experimental Ranges: We tested window sizes ranging from 3 to 20 posts. The objective was to balance engagement data for post continuity.
    - Optimal Window Size Determination: Through iterative testing, a 20-post window was found to be most effective. This size allowed for a robust analysis of engagement trends while ensuring that windows contained enough posts to be statistically meaningful.
    - Centering Method: Each viral post was used as the center of a window. This symmetric approach allowed us to analyze the engagement trends both before and after the viral event, ensuring a balanced view of its impact.

Trade-off between window size and the comprehensiveness of engagement context:

- Smaller Windows (e.g., 5-10 posts): While smaller windows increased the number of viral events that could be analyzed, they provided limited context for understanding the prolonged impact of virality on engagement.
- Larger Windows (e.g., 30-40 posts): Larger windows offered more extensive data around each viral post, allowing for a better understanding of the sustained effects of virality. However, the larger the window, the fewer the viral events that could be included without overlap or data interruption.

Window Size (Posts)	**Total Posts**	**Total Companies Contributed**	**Contribution Mean**	**Contribution Std. Dev.**
5	500	10	60.0	48.98
10	780	10	78.0	64.26
20	920	10	92.0	53.50
30	1080	10	108.0	47.33
40	960	8	96.0	73.52

Appendix 2: Exploratory Analysis Overview

Before delving into our causal analysis and research question, we conducted exploratory analysis, focusing on three key inquiries. From this analysis, we identified the sectors and platforms with the greatest potential for examination among the four most active social media industries: fashion, tech, travel, and FMCG, as well as across platforms including Instagram, TikTok, Facebook, Twitter, and YouTube.

1. Platform Preferences Across Industries

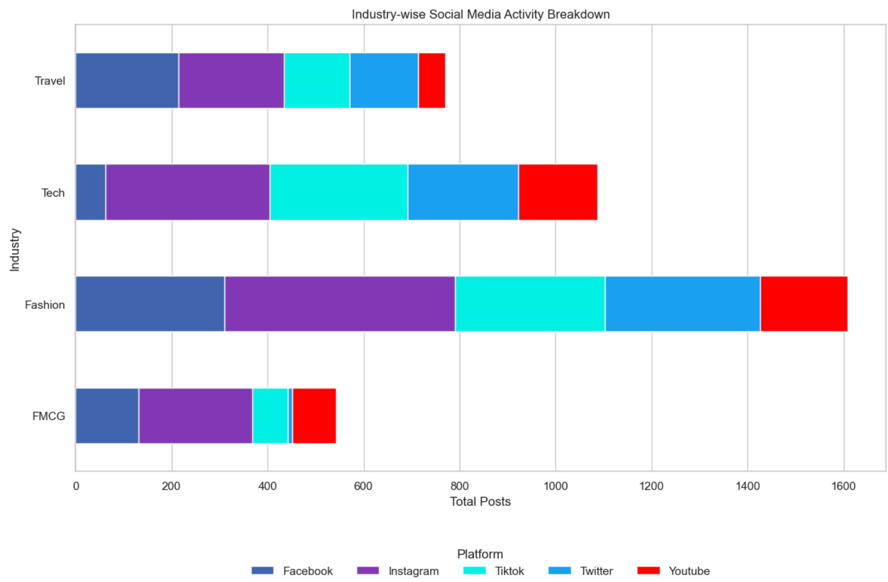


Fig 1. Social Media Usage Across Industries

- Instagram emerges as the dominant platform across industries, with FMCG (43.83%) and Tech (31.43%) showing the highest percentages, indicating a universal preference for its visual-centric approach.
- Industries invest heavily in visual content on Instagram to engage their audience effectively, leveraging its combination of images, short videos, and stories for B2C marketing strategies.
- TikTok demonstrates a significant presence in Fashion (19.40%) and Tech (26.38%) industries, surpassing established platforms like Facebook in the latter.
- Industries are eager to leverage TikTok's user base, known for preferring dynamic and interactive content, highlighting a shift towards innovative content strategies.

2. Effective Platforms for Customer Engagement

- We analyzed engagement rates across various social media platforms for each company, identifying the most effective platforms for engaging customers within different industries.

- Instagram traditionally leads in FMCG and Tech industries for maximizing customer engagement, while TikTok has emerged as dominant for the Travel and Fashion sectors recently.

3. Posting Frequency in the Fashion Industry

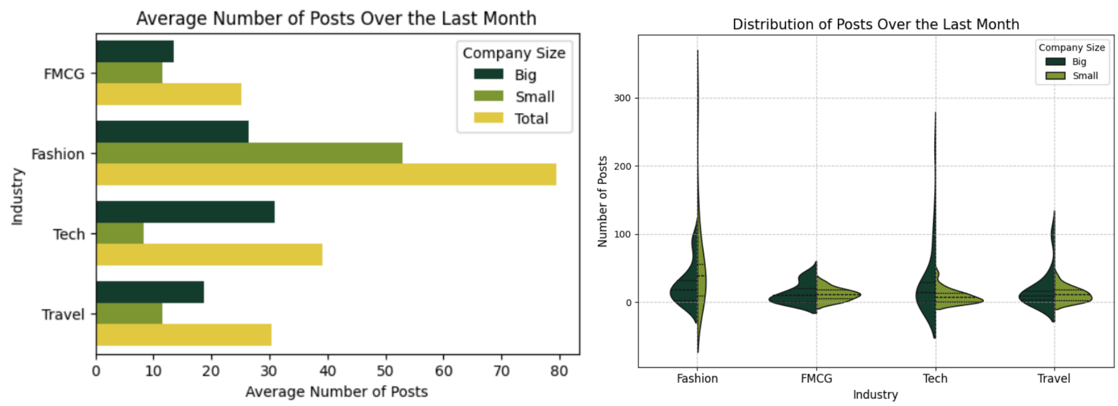


Fig 2, 3. Average Posts and Posting Distribution Over the Past Month

- The fashion industry averages 79.38 posts per company in the most recent 30-day period, regardless of their size.
- Smaller fashion companies post nearly twice as often as larger ones, suggesting different strategic approaches to social media marketing.
- Fluctuations in posting frequency likely correlate with factors such as product launches or sales seasons.
- Most fashion companies consistently post over 60 times a month on Instagram and more than 20 times a month on TikTok, reflecting their responsiveness to trends and robust social media marketing efforts, especially among luxury brands.
- The fashion industry leads in social media engagement, averaging nearly 230.4 posts per profile annually.

Key Takeaways

1. Adaptability and Platform-Specific Strategies

- TikTok's growing influence in the Fashion and Tech sectors highlights a preference for dynamic content experiences.
- The fashion industry's high posting rate, particularly among smaller companies, reflects an aggressive strategy for continuous consumer engagement.

2. Engagement and Follower Dynamics

- Metrics offer insights into how different industries and companies benefit from social media marketing.
- Identifying virality outliers underscores the impact of targeted campaigns and disparities in engagement rates across industries.

3. Platform Preferences

- Instagram and TikTok are favored platforms for engaging content, while Twitter and YouTube serve specific roles in communication and content depth.

4. Industry Growth on Social Media

- Certain industries, like Fashion and Tech, exhibit higher interaction rates.
- Despite potential, effectively adapting to the social media landscape remains challenging for most companies.

Appendix 3: Regression Outputs and Interpretations

Detailed exploration of the regression analyses conducted to understand the impact of virality on baseline engagement within various time windows following a viral event on Instagram. Each component of our regression output is elaborated here to provide a deeper insight into our findings and their implications.

Regression Equation Setup

The following baseline regression equation was used to perform the diff in diff analysis:

$$Y_{it} = \beta_0 + \beta_1 T_t + \beta_2 D_i + \beta_3 (T_t \times D_i)$$

Where:

Y_{it} : Square Root Engagement (SRE), expressed as Z-Score transformation

T_t : Binary indicator for pre-treatment vs post-treatment

D_i : Number of Viral Posts (Intensity of treatment, not binary indicator)

β_0 : Expected z-score SRE

β_1 : Expected change in z-score SRE with no viral post

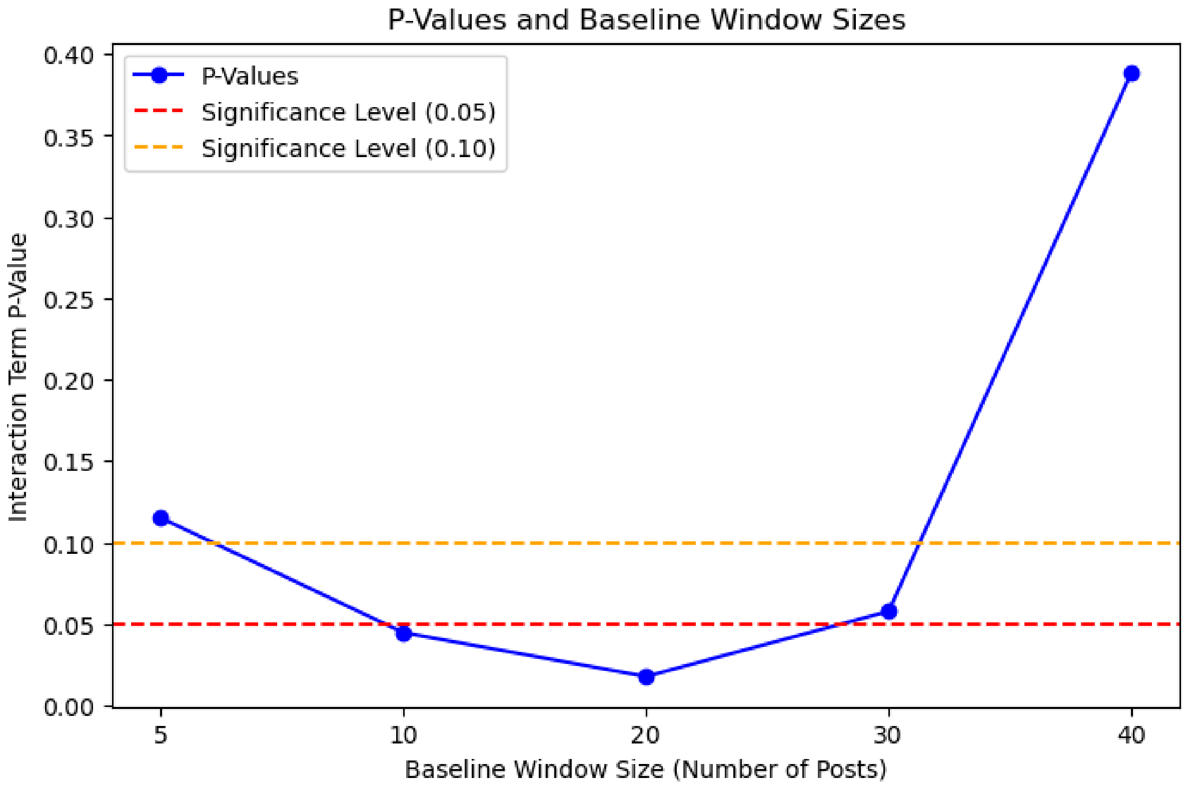
β_2 : Expected change in z-score SRE per viral post, independent of time

β_3 : Expected change in z-score SRE per viral post, controlling for effects of time

This equation setup makes an important distinction between a binary indicator for treatment and a treatment of varying intensities (multiple viral posts). This experimental design was chosen because of the difficulty of isolating single viral posts between two baselines, especially when the length of consecutive posts required to establish a baseline gets increasingly larger. By allowing for varying degrees of treatment, we can isolate the "per viral post" effect denoted as the interaction term β_3.

Regression Analysis Overview

The regression model developed aimed to quantify the effects of virality, operationalized through an interaction term between the presence of viral content and specific post-viral time windows. Our model focused on how engagement metrics change in response to viral posts, considering different temporal contexts in which these viral events occur.



P-Values by Window Size

The p-values associated with the interaction term across different baseline window sizes, ranging from 5 to 40 posts. The p-values measure the statistical significance of the interaction between virality and the number of posts following a viral post, indicating the robustness of the virality's impact on engagement across varying window sizes.

Graph Analysis

The plot displaying p-values across varying baseline window sizes illustrates a nuanced dynamic in determining the optimal context for analyzing the impact of virality on engagement. For smaller windows of 10 and 20 posts, we observe p-values below the 0.05 significance level, indicating strong statistical support for the virality's impact on engagement within these confines. These window sizes provide a balanced view, capturing enough post-viral interaction without losing contextual relevance or diluting the viral post's immediate effects. However, as window sizes increase, particularly at 30 and 40 posts, the p-values rise sharply, surpassing the 0.10 significance threshold in the largest window. This suggests that the impact of virality becomes statistically insignificant with larger windows, likely due to the dilution of the viral effect over time and a reduction in available data for these extended sequences, highlighting the challenge in maintaining analytical precision with expanding post-viral timelines.

The trends observed suggest that while smaller windows may capture the immediate effects of virality, they risk lacking sufficient data to fully contextualize these effects. Conversely, very large windows might incorporate more comprehensive engagement patterns but dilute the specific impact of virality and suffer from lower data density.

Detailed Regression Table

Here we provide the full regression output for the model focusing on the 20-post window, where the interaction of viral presence and time frames showed significant effects:

Term	Coefficient	Std Error	t-value	P-value	95% Conf. Interval
const	-0.1663	0.026	-6.357	0.000	(-0.218, -0.115)
Time_Frame	-0.0290	0.037	-0.783	0.434	(-0.102, 0.044)
Treatment	-0.0097	0.010	-1.014	0.311	(-0.028, 0.009)
interaction	0.0321	0.014	2.372	0.018	(0.006, 0.059)

Interpretation:

Constant: A negative intercept suggests that the baseline engagement level, absent viral influences, is below the normalized mean of our dataset.

Time_Frame: This variable's non-significant effect suggests that time alone, without the interaction with viral posts, does not significantly influence engagement levels.

Treatment (Viral Post Presence): Similarly, the presence of viral posts alone does not significantly alter engagement, indicating that virality's impact is context-dependent.

Interaction Term: The statistically significant positive coefficient for the interaction term reinforces the hypothesis that viral posts, when combined with specific time frames, have a statistically significant positive effect on engagement. However that effect is is extremely small raising concerns over the practical significance of striving for virality as a primary social media strategy

The regression analysis provides substantial evidence that while virality alone does not guarantee sustained increases in engagement, its strategic use within defined post-viral windows can enhance audience interaction. This nuanced understanding emphasizes the importance of timing and contextual engagement strategies in maximizing the benefits of viral marketing campaigns on social media platforms like Instagram.

Appendix 4: Control Variables

Our analysis used the bare minimum base variables to conduct the difference in difference analysis - binary indicators for Treatment/ Control and Time Frame along with the interaction between the two. It is important to consider the effects of adding control variables to the analysis and determine if the effect is consistent or deviates from the patterns previously described. In the table below we describe the coefficient and P-Value for the interaction term using a window size of 20 posts while various combinations of control variables are added to the regression analysis. A check indicates the control variable was included in the regression. The first column represents the base model with no control variables, as described earlier in this report. From the results we see that adding control variables did not drastically change the interaction term coefficient, justifying the usage of the baseline model in our analysis.

- **Month**: Categorical Date (Month) that the post was made to account for potential seasonality effects.
- **Post Type:** Categorical type of post, including Instagram TV, Reels, or Images
- **Total Posts:** Total number of posts associated with the company for the entire 1 year dataset. This variable accounts for variable post frequencies between companies.

	Base Model	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Month	✖	✔	✖	✖	✔
Post Type	✖	✖	✔	✖	✔
Total Posts	✖	✖	✖	✔	✔
20-Post Interaction Coefficient	0.0321	0.0330	0.0328	0.0321	0.0314
P-Value	0.018	0.026	0.016	0.018	0.037