Department of Mathematics

Howard University

Washington, DC 20059

Math 014(01) Intro to Data Science

Spring 2025

Final: Global Social Media Virality: Unpacking What Makes Content Go Viral

By

Godfred Asamoah

Submitted To:

Dr. Meenakshi Nerolu

Objective:

The objective of this project is to analyze a comprehensive dataset of viral social media posts to

Godfred Asamoah

Global Social Media Virality: Unpacking What Makes Content Go Viral

Dr. Meenakshi Nerolu

04/29/2025

uncover patterns in content virality across different platforms. We aim to examine how various factors – such as the platform, content type, topic (hashtags), and engagement metrics – contribute to a post's viral success. Through data wrangling, visualization, and statistical analysis, we seek to identify key insights (e.g. which platforms foster the most engagement, what content types trend more often, how engagement is distributed) that can inform social media strategy and our understanding of viral phenomena.

Introduction

Social media platforms have become incubators for viral content, where a single post can rapidly gain millions of views, likes, shares, and comments. Understanding why certain posts go viral while others languish is a topic of both academic interest and practical importance for marketers, influencers, and platform designers. Prior research suggests that content popularity often follows a **power-law (Pareto)** distribution – a small fraction of posts garner the majority of attentiontutorchase.comtutorchase.com – making virality hard to predict. Nevertheless, by examining data on past viral posts, we can observe common features of highly engaging content.

In this report, we analyze a **Viral Social Media Trends** dataset containing **5,000 posts** that went viral on four major platforms: **YouTube, TikTok, Instagram, and Twitter**. Each record includes the platform, a prominent **hashtag** categorizing the content (e.g. #challenge, #education), the **content type** (format of the post, such as video, image post, live stream, etc.), the **region** of

origin, and engagement metrics (**views, likes, shares, comments**). We also derive additional metrics like total engagement (sum of likes, shares, comments) and engagement rate (the percentage of viewers who interacted). The data spans multiple regions (USA, India, Brazil, UK, etc.), reflecting the global nature of social media virality. By comparing platforms and content characteristics, we attempt to tell the story of what drives a post to become "viral."

The following sections detail our methodology for cleaning and analyzing the data, present key findings with visualizations, and discuss the narrative those findings reveal about viral social media content. The analysis provides a cross-platform perspective – for example, contrasting the high interaction rates of TikTok videos with the more modest engagement on Twitter – to illustrate how platform dynamics and content choices can influence viral success.

Method

Data Wrangling: The dataset was first inspected for structure and consistency. It contains 5,000 entries with 10 attributes (Post ID, Platform, Hashtag, Content_Type, Region, Views, Likes, Shares, Comments, Engagement_Level). We confirmed there were **no missing values** in any column (each had 5000 non-null entries). All posts have unique IDs, so there are no duplicate records. We standardized text fields for consistency – for instance, hashtags were lowercased and the '#' symbol was removed to facilitate grouping (e.g., "#Fitness" became "fitness"). This

allowed accurate counting of occurrences of each hashtag. We created two new numeric fields: **Total_Engagement = Likes + Shares + Comments** for each post, and **Engagement_Rate = (Total_Engagement / Views) * 100** (percentage of viewers who engaged). These derived metrics help quantify overall impact and efficiency of engagement.

Handling Outliers: Given the nature of viral content, we expected some extreme outliers (posts with extraordinarily high engagement relative to views, etc.). We identified a number of outliers in the engagement rate distribution (e.g. ~11.9% of posts had unusually high engagement_rate values). Notably, one TikTok post in the dataset had only ~1.3k views but over 318k likes and 36k comments – an apparent anomaly yielding an engagement rate of **28,174%**. Such values indicate either exceptional user behavior or data inconsistency. We chose to **retain outliers** in the analysis rather than remove them, as they are genuine instances of "extreme virality" or data quirks that are themselves informative. However, when appropriate we applied log-transforms to engagement metrics to visualize distributions without extreme skew.

Analysis Tools: We used Python (pandas for data manipulation, seaborn/matplotlib for plotting) to perform the analysis. Summary statistics were generated to understand central tendencies and variability. We computed frequency distributions for categorical features (platform, content type, hashtag, region) to see which categories dominated the viral content. We plotted histograms for each engagement metric (views, likes, shares, comments, etc.) in both linear and log scales to examine their distribution shapes. We also aggregated the data by platform, content type, and

Godfred Asamoah

Global Social Media Virality: Unpacking What Makes Content Go Viral

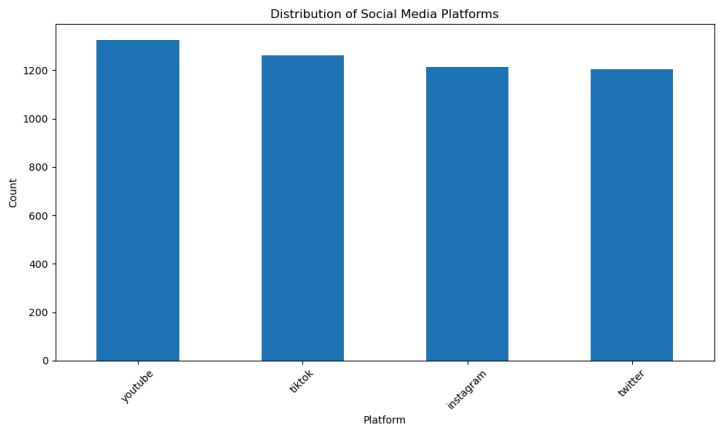
Dr. Meenakshi Nerolu

04/29/2025

hashtag to calculate average engagement metrics, enabling comparisons via bar charts and heatmaps. This combination of statistical description and visualization provides both high-level overviews and granular insights.

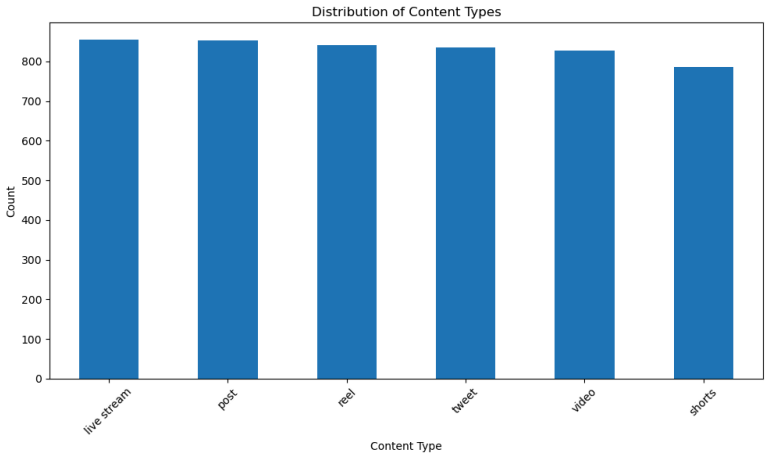
Storytelling (Key Findings)

Figure 1: Distribution of viral posts by platform. The dataset includes four major platforms. YouTube had the highest share of viral posts (1324 out of 5000, about 26.5%), followed by TikTok (1260 posts, ~25.2%), Instagram (1212 posts, ~24.2%), and Twitter (1204 posts, ~24.1%). This near-even distribution suggests that viral content is not confined to a single platform; ***all four platforms contribute significantly*** *to the viral phenomena captured in this data. The slightly higher count for YouTube might reflect its broad global user base and the longevity of video content on the platform. Meanwhile, Twitter has the fewest viral*



posts in this sample, which could relate to its fast-paced feed where trends can be short-lived.

Figure 2: Distribution of content types among viral posts. The viral posts encompass a variety of content formats, indicating that virality isn't limited to one style of content. Live Streams (855 posts, 17.1%) and standard Posts (853, 17.1%) were the most common formats, closely followed by Reels (841, 16.8%), Tweets (836, 16.7%), Videos (828, 16.6%), and Shorts (787, 15.7%). The counts are relatively balanced, implying that ***any***



content type can go viral *under the right conditions. Notably, short-form videos (e.g., Shorts) appear slightly less frequently in our viral dataset, yet as we will see, they punch above their weight in engagement. This balance across formats underscores that creators on all mediums – be it quick Tweets or longer live videos – have potential for virality.*

Godfred Asamoah

Global Social Media Virality: Unpacking What Makes Content Go Viral

Dr. Meenakshi Nerolu

04/29/2025

Figure 3: Top 10 hashtags by number of viral posts. Viral content often coalesces around popular themes denoted by hashtags. Here, ***#Fitness*** *leads with 536 viral posts (10.7% of the dataset), followed by* ***#Education*** *(525 posts) and* ***#Challenge*** *(507 posts). Comedy (#Comedy, 505) and dance trends (#Dance, 496) are also highly represented, reflecting the appeal of entertainment content. Tech-related (#Tech, 491) and fashion (#Fashion, 487) content show up frequently as well. Interestingly,* ***#Viral*** *itself appears in 481 posts, suggesting a meta trend where tagging content as "viral" is common. The prevalence of these hashtags indicates that tapping into popular topics or challenges can boost a post's viral potential – indeed, using the right trending hashtag can expand content reach and attract new followers*[*influencity.com*](influencity.com)*. It's worth noting that while #Fitness has the most viral posts by volume, it is not the top in engagement rate (as discussed later), whereas #Viral-tagged posts, despite being fewer in number, tend to achieve exceptionally high engagement.*

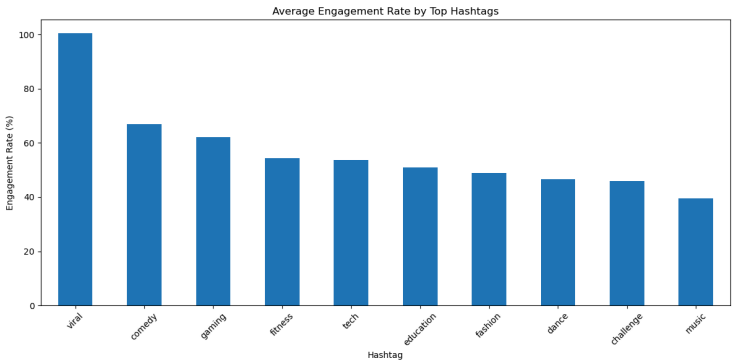


Figure 4: ***Average engagement rate (%) by platform*** *for viral posts. We observe significant differences in how audiences engage on each platform. Instagram leads with*

Godfred Asamoah

Global Social Media Virality: Unpacking What Makes Content Go Viral

Dr. Meenakshi Nerolu

04/29/2025

an average engagement rate of ~66.2%, meaning a sizable fraction of viewers interact (like, share, or comment) on viral Instagram posts. TikTok is a close second at ~63.9%, aligning with the notion that TikTok's algorithm and community encourage frequent interaction (likes, shares of short videos, etc.). In contrast, YouTube (50.6%) and Twitter (46.5%) show lower engagement rates on viral posts. This is consistent with general social media patterns: ***"Instagram and TikTok tend to see higher engagement rates, while Twitter… usually [has] lower numbers."***influencity.com *The more visual, immersive nature of Instagram/TikTok likely prompts users to engage more (double-tapping, commenting on creative videos), whereas Twitter's virality often comes from retweets without as much direct engagement per view. These findings underscore that the same viral status does not equate to the same level of interaction across platforms – the platform's format and audience behavior moderate engagement intensity.*

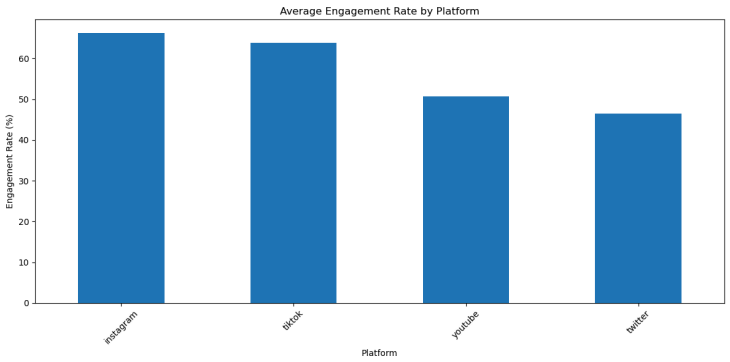


Figure 5: ***Average engagement rate (%) by content type*** *for viral posts. Short videos (****Shorts****) stand out with the highest engagement rate, about 80.6% on average – an extraordinary figure indicating that viral short clips often convert viewers into active*

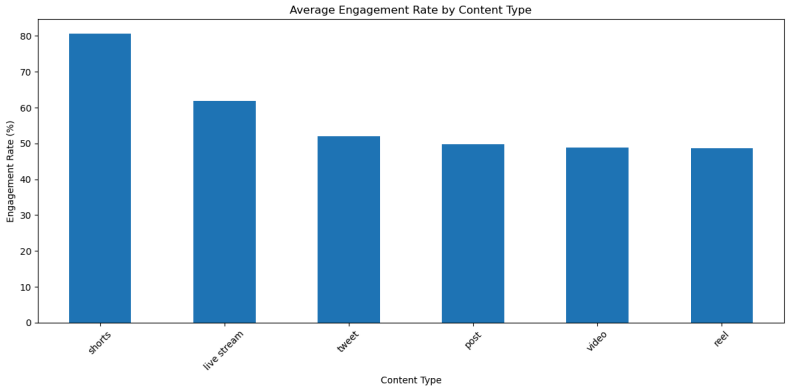
Godfred Asamoah

Global Social Media Virality: Unpacking What Makes Content Go Viral

Dr. Meenakshi Nerolu

04/29/2025

engagers. This supports the idea that bite-sized video content is highly effective at prompting reactions (likely due to being concise and easily shareable)[theshareofvoice.com](theshareofvoice.com). Live Streams also show strong engagement (~61.9%),



possibly because viewers of live content are highly invested in real-time interaction (commenting during streams, etc.). Traditional text-based Tweets and standard image/text Posts have moderate engagement rates (~52.1% and 49.8%, respectively), while regular Videos and Reels hover just below 50% engagement. It is fascinating that ***short-form videos (Shorts) outperform even Reels*** *– despite both being brief video formats, the way they are presented on platforms (YouTube Shorts vs. Instagram Reels) might influence interaction. Overall, the content format clearly influences how deeply audiences engage: formats that are immersive or time-sensitive (live or quickly consumed video) yield a larger share of viewers who respond or share. This suggests creators aiming for virality should tailor their format to maximize engagement – for instance, leveraging short, catchy videos or interactive live sessions to trigger more viewer interaction.*

Godfred Asamoah

Global Social Media Virality: Unpacking What Makes Content Go Viral

Dr. Meenakshi Nerolu

04/29/2025

Beyond these primary findings, our analysis also revealed that posts tagged **#Viral** achieved the highest average engagement rate (over 100%, meaning on average they garnered more total interactions than viewers, which can happen if content is re-shared beyond the initial view count or viewed multiple times by highly engaged users). Other highly engaging hashtags included #Comedy and #Gaming, each with ~60–70% engagement rates, indicating that funny and gaming-related content often spurs disproportionate interaction relative to its reach. In contrast, #Music and #Dance, while common in viral posts, had comparatively lower engagement rates (~40–50%), perhaps because casual viewers enjoy the content passively without always interacting. These nuances highlight that *virality is multifaceted*: a high view count doesn't always mean high engagement, and the topic and format greatly influence how actively the audience participates.

Conclusion

In this study, we unpacked the anatomy of viral social media content through a global lens. The analysis shows that **viral content is a cross-platform phenomenon** – no single network has a monopoly on virality, though engagement dynamics differ from one platform to another. Visual-centric platforms (Instagram, TikTok) yield more interactions per view than text-focused

platforms (Twitter), reflecting inherent differences in user engagement behavior[influencity.com](influencity.com). We also found that content format plays a crucial role: short-form videos drive the most vigorous engagement, leveraging their snappy, attention-grabbing nature[theshareofvoice.com](theshareofvoice.com), while even longer live streams manage to keep viewers actively involved. Importantly, aligning content with trending themes via hashtags remains a **powerful strategy for virality**, as evidenced by the high volume of viral posts around popular tags (e.g., #Fitness, #Education, #Challenge). However, sheer frequency of a hashtag does not guarantee engagement – the tag #Viral, for instance, though less common, corresponded to exceptionally high engagement, suggesting that self-referential viral content can amplify user response.

Overall, what makes content go viral appears to be a synergy of **relevance (trending topics), platform fit, and format effectiveness**. Creators aiming for viral success should craft content that not only rides popular trends (using hashtags strategically) but also fits the engagement style of the target platform and leverages formats that encourage interaction (e.g., videos that prompt viewers to like/share/comment). Our findings align with broader social media observations that highly engaging content often involves storytelling or humor, encourages participation, and is tailored to platform algorithms. While virality can never be guaranteed, understanding these patterns – which platforms foster the most interaction, what content types resonate, and how reach translates to engagement – provides valuable guidance. In a digital landscape where

attention is precious, such insights can help content creators and marketers increase their odds of striking that ever-elusive viral chord.

References

1. Sangiorgio, E., & Di Marco, N. (2020). *Evaluating the effect of viral posts on social media engagement*. Scientific Reports, **10**. DOI: 10.1038/s41598-024-84960-6nature.com.

2. Influencity (2025). *What Is a Good Engagement Rate on Social Media? Benchmarks and Tips*influencity.cominfluencity.com. *(Influencity Blog)*.

3. Tayssir (2023). *Create Viral Content: Secrets for Social Media Success*. The Share of Voice Blogtheshareofvoice.com.

Acknowledgments

The author would like to thank Dr. Nerolu and coursemates for their guidance on this project. Special appreciation goes to the creators of the dataset for making this analysis possible. This

Godfred Asamoah

Global Social Media Virality: Unpacking What Makes Content Go Viral

Dr. Meenakshi Nerolu

04/29/2025

project was conducted as part of an academic program in data analysis, and the support of peers during discussions and code reviews is gratefully acknowledged.