Analyzing Hashtag Lifespans and Their Influence on Post Performance

Maryam Zad
DS340H

Introduction

Hashtags shape content discovery on short-form video platforms, but not all hashtags behave the same. This matters especially for #SAHM (*Stay-at-Home Mom*), a large and active TikTok community, where hashtags act not only as discovery tools but also as identity markers that structure how creators join, signal, and sustain shared conversations. Using a comprehensive dataset of all TikTok posts tagged with #SAHM, this project examines how different hashtag lifecycle patterns operate within this community. Understanding these temporal dynamics can reveal how #SAHM trends emerge and how creators use different types of tags to drive engagement.

Research Question:

How do different hashtag life cycles within the #SAHM community influence engagement and reflect shifting patterns in content creation?

Data & Methods

Data Sources

- *Hashtag Metdadata Dataset:* Derived from the #SAHM data set. Includes each hashtags total frequency, number of unique users, first and last encountered timestamps, and number of unique active days.
- *Final Sample (Creators):* A subset of 1,000 long-term creators, selected based on posting longevity and total number of posts within the #SAHM dataset, resulting in 597,071 posts used for analysis.

Data Preparation:

- Filtered the full #SAHM dataset to include only posts from the 10,000 selected long-term creators.
- Exploded each post's hashtag list so every post–hashtag pair appeared as a separate row.
- Merged hashtag metadata into the filtered posts to attach lifecycle features (duration, unique active days, frequency) to each hashtag.

Hashtag Classification

- Calculated each hashtag's duration (days between first and last appearance) and active unique days.
- Used percentile-based thresholds (25th and 75th) to categorize hashtags into three lifecycle groups:
    - Trendy: short lifespan, low number of active days
    - Moderate: mid range lifespan and activity
    - Persistent: long lifespan, high number of active days
- Assigned each hashtag a categorical label based on these lifecycle measures.

Examples of Hashtag Types

- Persistent: *fyp, momsoftiktok, momlife, momtok, sahmlife*
- Moderate: *sahmsoftiktok, sahmtok, capcut, babytok,, momsunder25*
- Trendy: *creatorsearchinsights, onthisday,, over25club, happyeaster, memecut*

Modeling Approach

Model 1 - Decision Tree Classifier

- Trained on hashtag lifecycle features (unique active days, duration, and frequency).
- Predicts whether a hashtag is trendy, moderate or persistent based on these temporal characteristics.

Model 2 - Engagement Regression

- Aggregated post-level data to compute the proportion of trendy, moderate, and persistent hashtags used in each post.
- Regressed log-transformed views on these proportions to measure their relationship with engagement.
- Included user fixed effects to account for creator-level differences in popularity and posting style.

Exploratory Data Analysis

Correlation analysis showed that the proportion of trendy hashtags used by a creator was negatively associated with average views (r = -0.008692), while persistent hashtags were positively associated (r = 0.020742). Moderate hashtags showed a negative relationship with engagement (r = -0.019706).

Scatterplots of hashtag usage versus engagement revealed similar patterns across creators. Most users were heavily clustered at low average view levels for both trendy and persistent hashtag proportions, reflecting the skewed distribution of views in the dataset. Each plot included a small number of high-engagement outliers, but their hashtag patterns were inconsistent: some used more trendy tags, while others relied on persistent ones. Overall, the results suggest that while hashtag behavior shows weak directional relationships with engagement, it is not a primary driver of visibility.

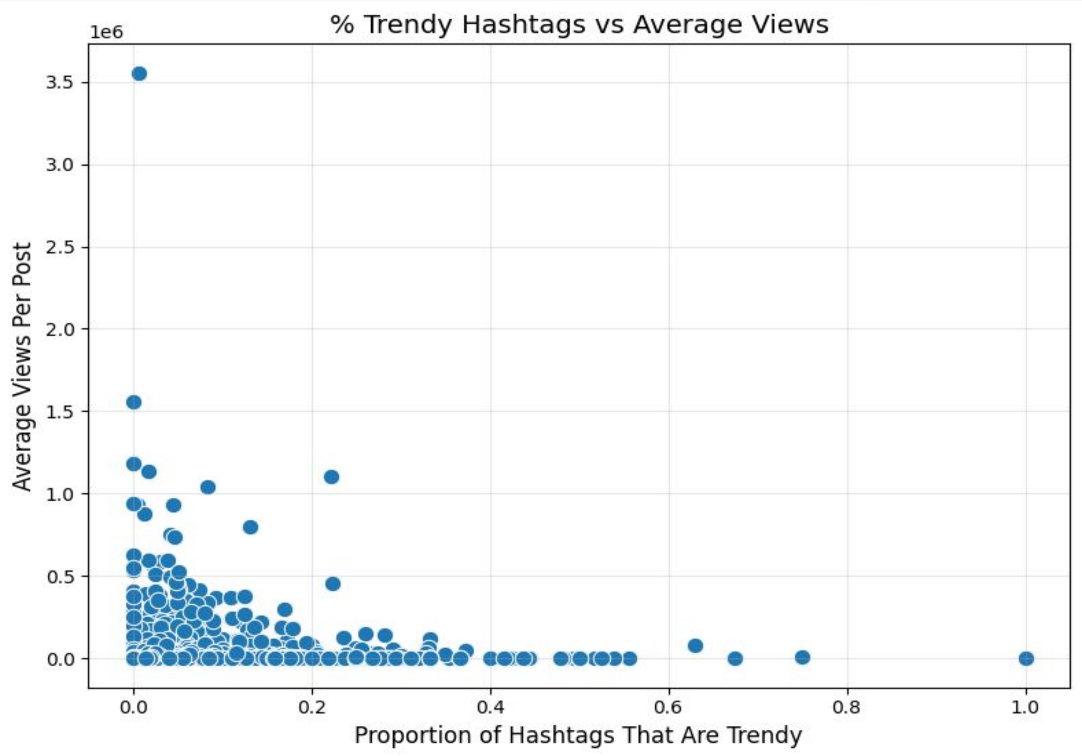

Figure 1. *Relationship between the proportion of trendy hashtags used by each creator and their average views per post. A slight negative pattern is visible, but most creators cluster at low view levels, indicating limited influence of trendy tags on engagement.*

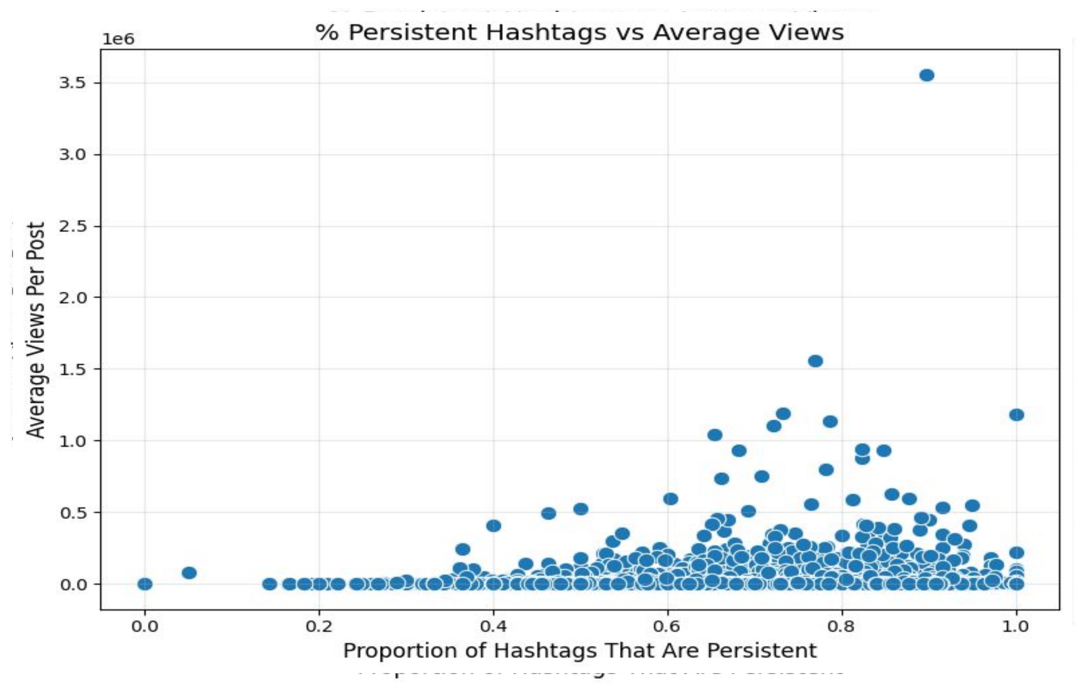

Figure 2. *Relationship between the proportion of persistent hashtags used by each creator and their average views per post. Most creators cluster at low view levels, with a slight positive trend but no strong predictive pattern.*

A comparison between high-view and low-view creators shows no meaningful differences in hashtag composition. Both groups devote about 5% of hashtags to trendy tags (4.93% vs. 4.98%), while persistent hashtags dominate for both, with only a slight increase among low-view creators (73.85% vs. 73.42%). Moderate hashtag usage is similarly close, with high-view creators using marginally more (21.65% vs. 21.17%). Overall, these near-identical patterns indicate that hashtag composition does not meaningfully distinguish high- from low-engagement creators in this dataset.

Results

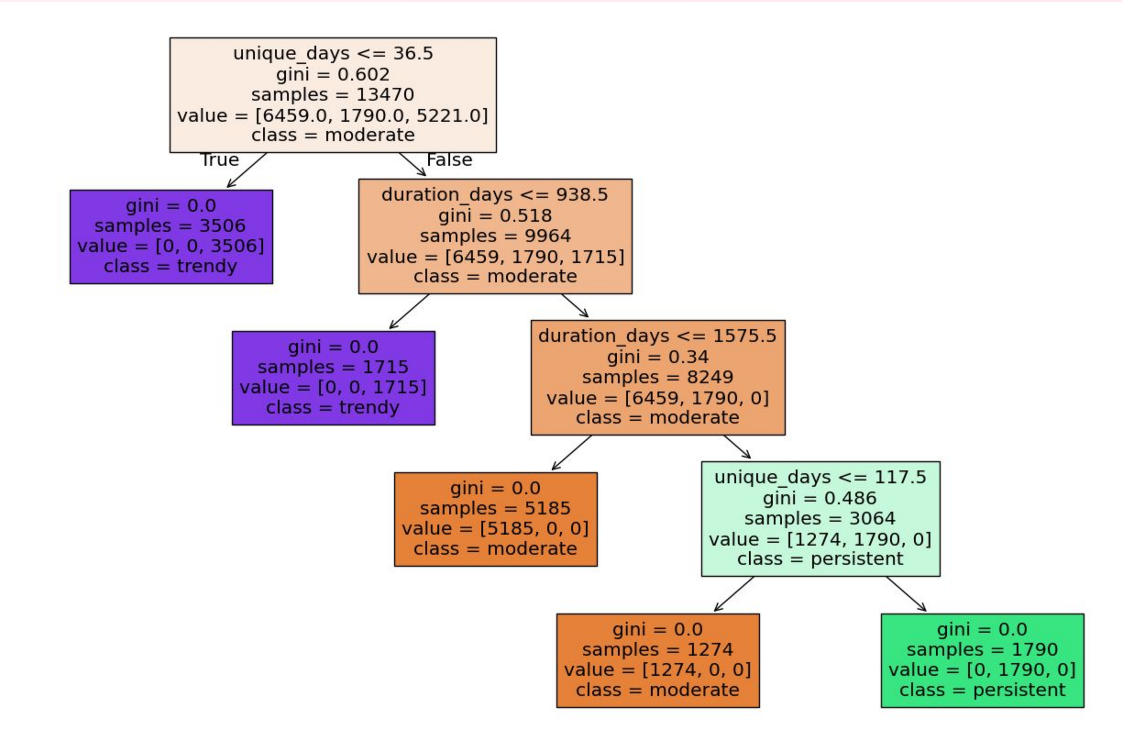

Figure 3. *Decision tree classification of hashtag types using unique active days and total duration. Short-lived hashtags are classified as trendy, long-lived and widely used tags as persistent, and intermediate patterns as moderate.*

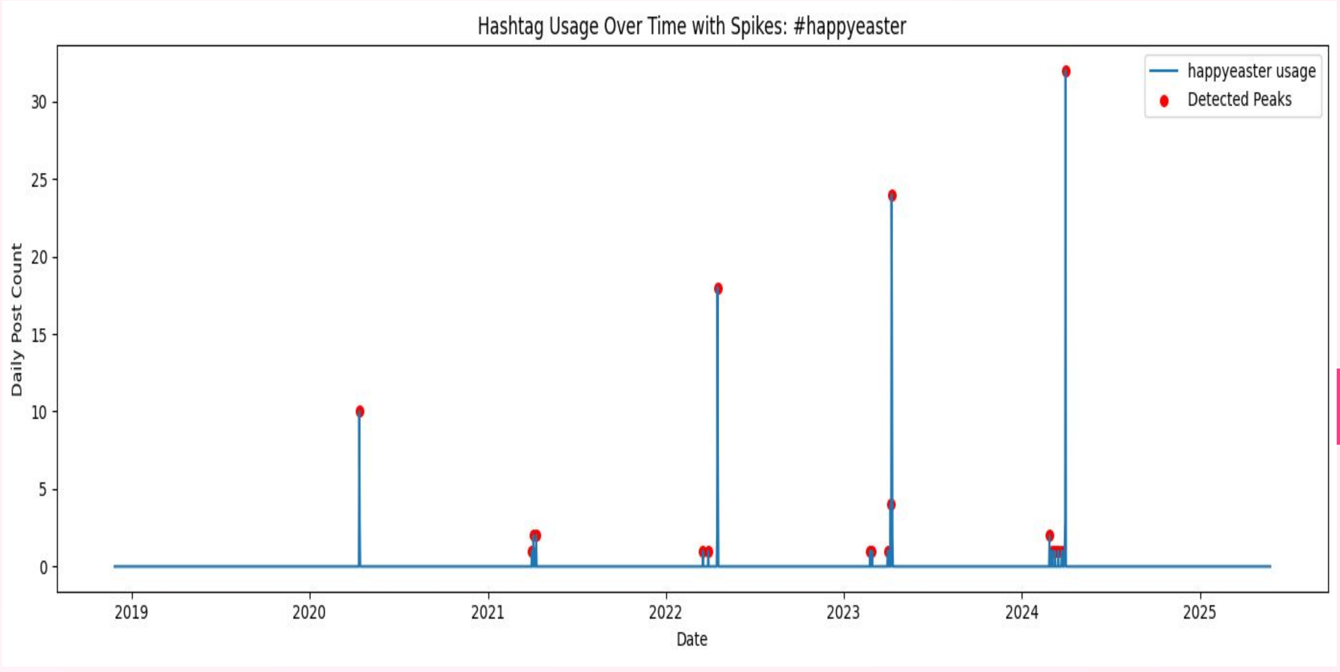

Figure 4. *Daily usage of the hashtag #happyeaster over time, illustrating a trendy hashtag pattern. Usage remains near zero for most of the year but exhibits sharp, short-lived spikes around Easter, reflecting event-driven attention rather than sustained or persistent*

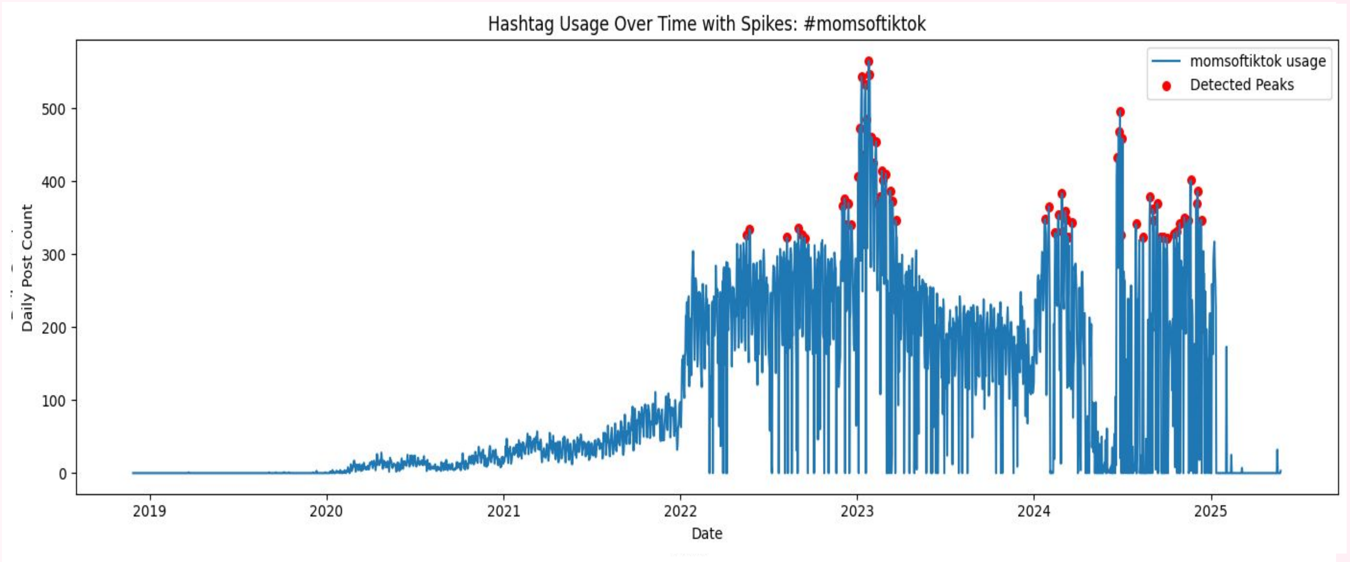

Figure 5. *Daily usage of a persistent hashtag,, showing sustained, repeated activity over an extended period. The frequent peaks reflect ongoing, stable engagement characteristic of persistent tags.*

Regression results indicate that hashtag lifecycle categories have no meaningful relationship with engagement. The baseline OLS model without user fixed effects explains virtually none of the variation in views ($R^2 \approx 0.001$), indicating that hashtag composition alone does not predict visibility. Adding creator fixed effects does not improve explanatory power ($R^2 \approx 0.002$), and all lifecycle coefficients remain statistically insignificant and effectively zero. These results suggest that engagement differences are driven by factors not captured in the model, such as creator- or content-specific characteristics, rather than hashtag lifecycle composition.

Limitations & Future Work

Limitations

- The analysis focuses exclusively on long-term #SAHM creators, so results reflect engagement patterns within this community and may not generalize to other TikTok niches.
- Hashtag categories (trendy, moderate, persistent) were defined using temporal thresholds, which may oversimplify the full complexity of hashtag meaning and usage.
- Important factors influencing engagement, such as algorithmic boosts, video quality, posting time, or audio choice, were not included in the dataset.
- The study relies on post-level metadata rather than content-level features, limiting the ability to interpret why certain hashtags are used or how they function in context.

Future Work

- Extend the analysis to additional creator communities to compare hashtag lifecycle dynamics across content niches.
- Incorporate richer video-level predictors (audio trends, sentiment, posting time, video length) to capture the behavioral drivers of engagement more fully.
- Explore longitudinal or causal designs to test whether using trendy hashtags at strategic moments affects short-term visibility or creator growth.

Conclusion

This project examined whether the lifecycle behavior of hashtags, trendy, moderate, or persistent, predicts engagement within the #SAHM creator community. While descriptive patterns revealed small differences in hashtag use across creators, these relationships were weak and inconsistent overall.

Regression analyses confirmed that hashtag lifecycle categories explain very little variation in views. Adding creator fixed effects did not materially improve explanatory power, and the influence of hashtag type remained negligible, suggesting that hashtag lifecycles do not meaningfully shape engagement. Instead, observed differences in engagement are likely attributable to factors not captured in the model, such as creator- or content-specific characteristics.

Overall, these findings indicate that while hashtag behavior offers insight into how #SAHM creators participate in community discourse, it is not a strong determinant of performance. Engagement appears to be driven more by creator identity, audience alignment, and content style than by the temporal dynamics of hashtag usage.

Acknowledgements

I would like to thank Professor Eni Mustafaraj for her guidance throughout this project. Her support and feedback were invaluable in shaping the direction of this work.

Repo: *https://github.com/DS340H/capstone-public-maryamkzad.git*