

Unraveling the Success of TikTok Videos: A Data Driven Approach

Jared Brotamonte

Department of Mathematics and Statistics, 801 Osborn Dr, Flagstaff, AZ 86011

Abstract

As social media platforms like TikTok continue to shape our digital landscape, understanding the mechanics of content virality becomes increasingly crucial. This project aims to address this need, paving the way for more informed and effective content strategies on TikTok. This project explores the dynamics of TikTok video popularity using data collected from the platform. The data was gathered using a web scraping tool called Octoparse, a tool used by many well-known companies including Sony, Audi, and even universities like Purdue University. User-related variables include the username, bio, total likes received, number of followers, and number of people the user follows. Video-related variables comprise of the videos likes, comments, shares, saves, hashtags used, the sound used, caption, publication date, and length of the video. The primary objective data mining of these variables to discern their impact on a video's popularity. This analysis will provide insights into the characteristics of successful TikTok videos, offering understanding of content virality on TikTok. The project outcomes could contribute to social media analytics and digital marketing and provide actionable insights for content creators and marketers, enabling them to strategize their content more effectively. By understanding the factors that drive video popularity, one could optimize content to reach a broader audience and achieve higher engagement. As social media platforms like TikTok continue to shape our digital landscape, understanding the mechanics of content virality becomes increasingly crucial. This project aims to address this need, paving the way for more informed and effective content strategies on TikTok.

Introduction

In the digital age, social media platforms have become a significant part of our daily lives, influencing our behaviors, opinions, and even shaping global trends. Among these platforms, TikTok has become a one of the main forms of social media known for captivating users worldwide with its short-form video content. As TikTok continues to grow in popularity, understanding the mechanics of content virality on this platform becomes increasingly important for those who want to utilize TikTok's platform for their own gain.

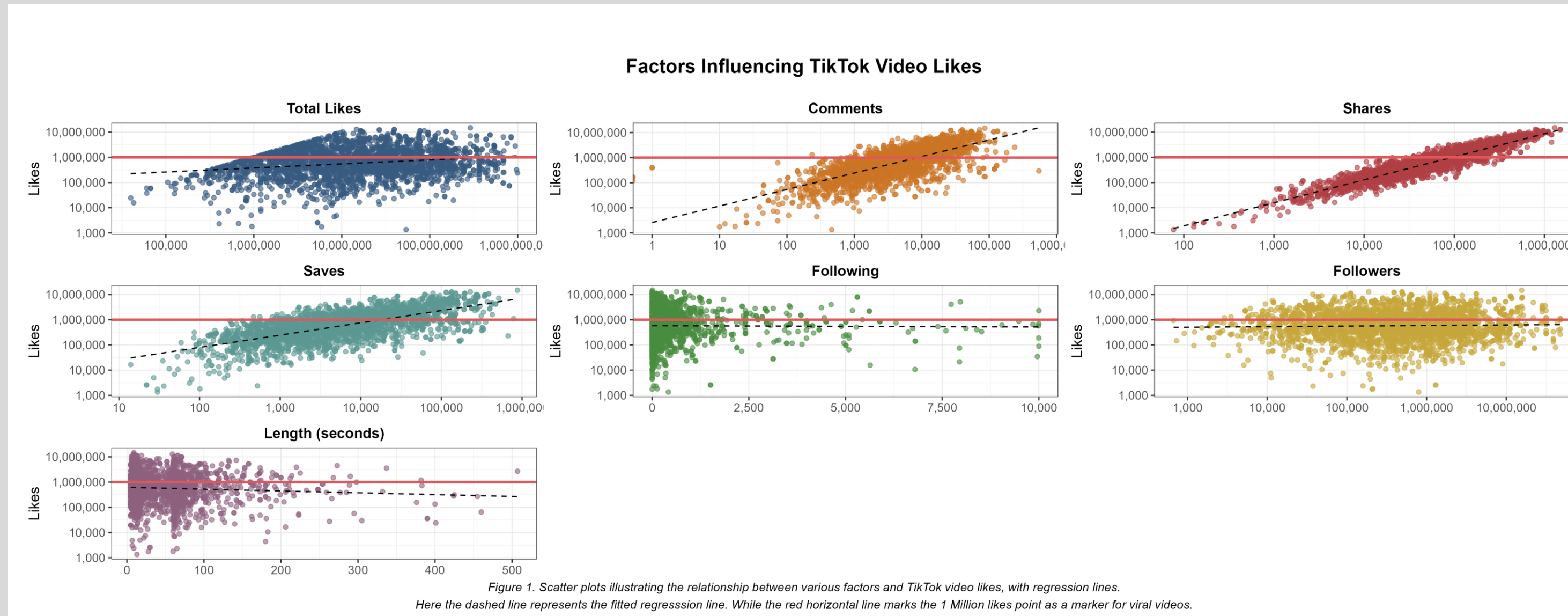
Methods and Data

The data for this project is collected using Octoparse, a powerful web scraping tool. The process involves setting up scraping tasks in Octoparse, specifying the TikTok URLs to scrape, and defining the data fields (such as username, bio, total likes, etc.) to extract. The reliability of the data collected is ensured through the fact that the data is scraped directly from TikTok's website. While Octoparse itself is viewed as reliable due to its ease of use, ability to handle complex scraping tasks, and excellent customer support. Although there is no documentation on the backend of Octoparse and how it works, it's reputation among well-known companies, like Sony and Audi, as well as universities, such as the University of Texas and Cornell University help attest to its reliability. The data itself consists of 2828 entries with a total of 12 different variables used for this project.

Table 1. Variable Descriptions

Variable	Description	Variable	Description
username	Username of the TaTaTok creator	saves	Number of saves of the TaTaTok video
nickname	Nickname of the TaTaTok creator	following	Number of accounts the TaTaTok creator is following
hashtags	Hashtags used in the TaTaTok video	followers	Number of followers of the TaTaTok creator
likes	Number of likes on the TaTaTok video	total_likes	Total number of likes on all videos of the TaTaTok creator
comments	Number of comments on the TaTaTok video	date	Date when the TaTaTok video was created
shares	Number of shares of the TaTaTok video	length	Length of the TaTaTok video in seconds

Correlation between Likes and other Variables



The scatter plots illustrate the relationship between various factors and TikTok video likes:

- Positive trends observed for total likes, comments, shares, saves, and followers
 - Suggests videos from accounts with higher engagement metrics tend to get more likes
- Strength of correlations varies:
 - Total likes, comments, shares, and saves exhibit stronger positive associations with video likes
 - Number of accounts followed shows weaker correlation
- Relationship between video length and likes appears weaker
 - Indicates video duration may not be a major determinant of like counts

Overall, the plots highlight the importance of building an engaged audience and creating content that resonates with viewers, as reflected by metrics like comments, shares, and saves, in driving higher like counts on TikTok videos.

Predicting the Virality of a Video

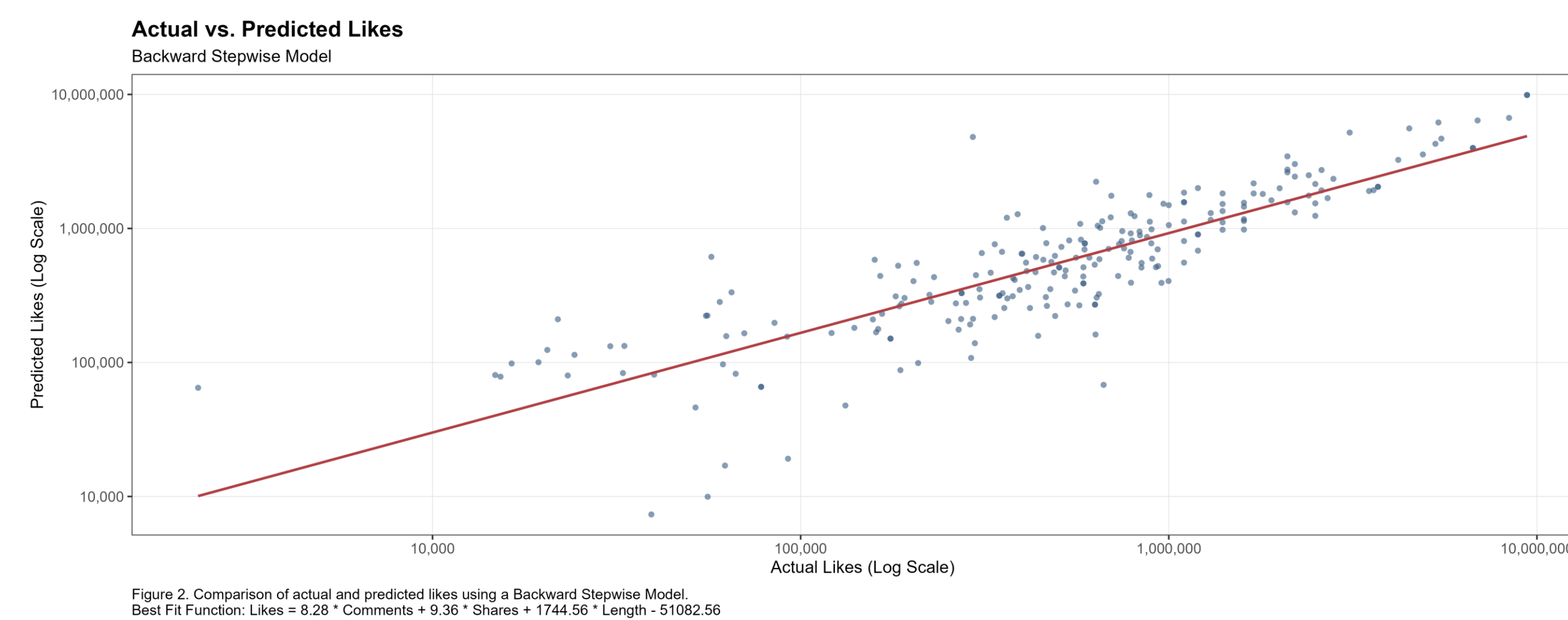


Figure 2. Comparison of actual and predicted likes using a Backward Stepwise Model.
Best Fit Function: Likes = 8.28 * Comments + 9.36 * Shares + 1744.56 * Length - 51082.56

The backward stepwise regression model identifies the most influential predictors of TikTok video likes by iteratively removing insignificant variables from an initial full model. Which in turn helps build a more accurate model to try and predict how many likes a video might get.

The final model includes:

- Comments: Each additional comment is associated with an approximate increase of 8.28 likes.
- Shares: Each additional share corresponds to an approximate increase of 9.36 likes.
- Length: Every additional second in video duration approximately increases likes by 1744.56.

The model excluded the following variables as insignificant predictors:

- Total_Likes, Saves, Following, and Followers

Surprisingly, the model pushes the length of the video one of the most influential predictors of likes. This is contradictory of the what we know from figure 1. This could be a sign of the outliers affecting the models, but more research would need to be done to know.

The impact of Hashtags on Virality

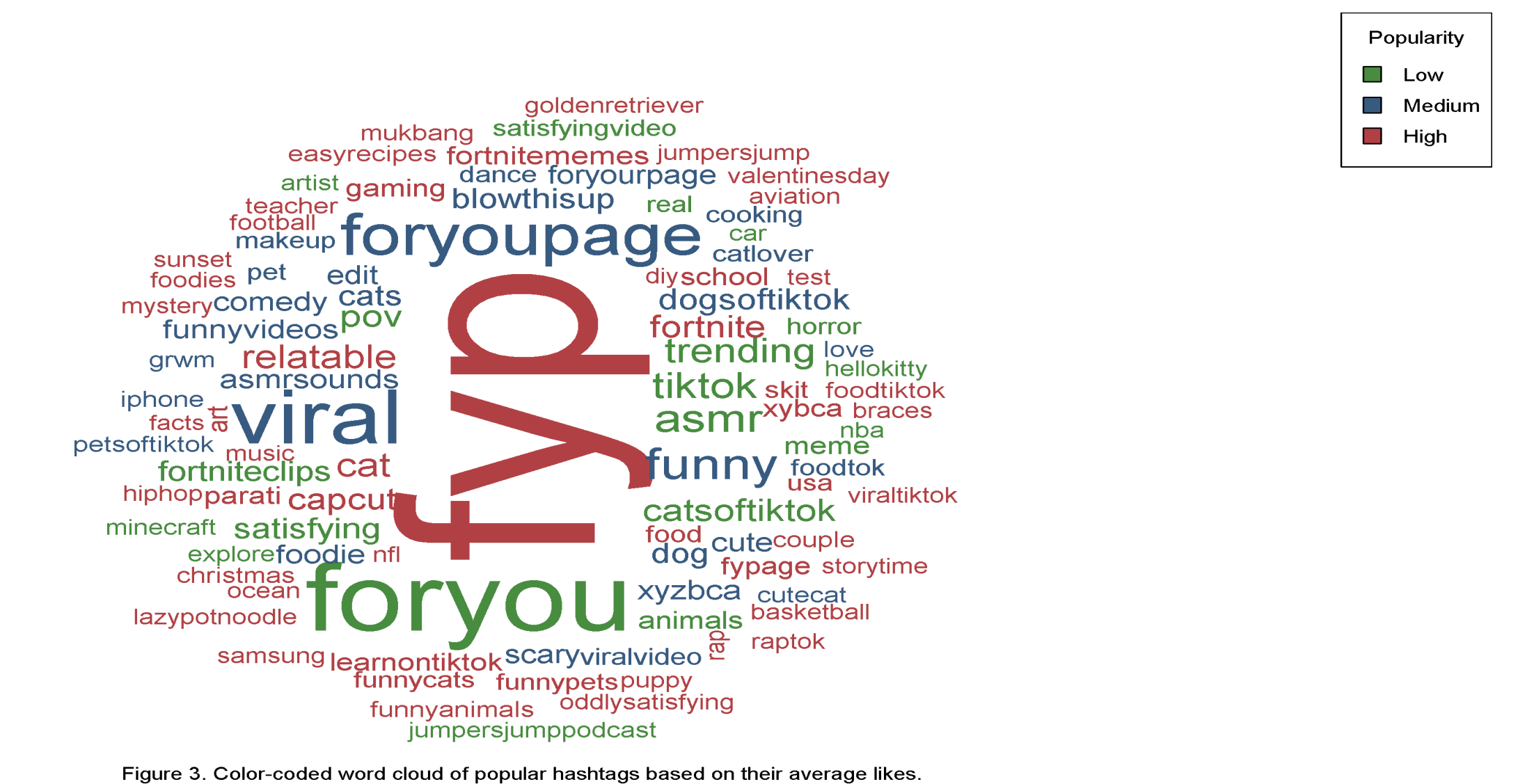


Figure 3. Color-coded word cloud of popular hashtags based on their average likes.

The color-coded word cloud visualizes the frequent hashtags used on TikTok videos, with:

- Size representing hashtag popularity
- Color indicating relative performance in receiving likes

Key insights:

- Prominence of hashtags like "#foryoupage", "#viral", "#fyp" for reaching the coveted "For You" page, highlighting algorithm optimization importance
- Diverse range of hashtags across genres (gaming, hip-hop, cooking, animals, comedy) reflecting TikTok's broad content landscape
- High popularity (red) of niche hashtags like "#lazypotnoodle" and "#fortnitememes", suggest niche content's role in driving engagement
- Varied color coding of trending hashtags, indicating effectiveness depends on factors beyond just using certain hashtags.

Conclusions

In conclusion, this study's analysis of TikTok video data provides valuable insights into the factors driving engagement and success on the platform. The regression models identified comments, shares, and video length as significant predictors of like counts, quantifying their impact and enabling more accurate predictions. Additionally, the color-coded word cloud visualization highlighted the importance of optimizing content for TikTok's algorithm by leveraging popular hashtags like "#foryoupage" and "#viral", while also catering to niche interests through targeted, community-specific hashtags. The findings underscore the need for a balanced strategy that combines algorithm optimization, niche content creation, and continual analysis of hashtag performance to maximize visibility, resonance with audiences, and overall engagement metrics like likes, shares, and comments on the highly competitive TikTok platform.

References

- Octoparse Corporation. (2024). Octoparse (Version 8.6.8) [Software].
 - <https://www.octoparse.com/>
- TikTok Inc. (2024). Data retrieved from February to March of 2024 from the TikTok website.
 - <https://www.tiktok.com/>

Acknowledgements

- Northern Arizona Universities Department of Mathematics and Statistics
- Northern Arizona Universities Office of Undergraduate Research and Creative Activity
- Dr. Robert Buscaglia