

- *TikEngage* -

TikTok User Engagement Rate Analysis and Prediction Report

Problem statement:

The exponential rise of TikTok has established it as the leading social media platform, particularly among young consumers, who devote a significant portion of their time to the app. With approximately 1.06 billion active users as of 2023, of which 83% have contributed at least one video, TikTok offers unparalleled potential for content creators and brands to engage with an extensive audience. For content creators, TikTok represents a gateway to reach a vast and diverse audience, presenting an enticing opportunity for content to go viral and gain popularity. The platform has become a dream job for the younger generation, providing an accessible route to influencer status, accompanied by financial gains and a supportive following. Simultaneously, for businesses targeting younger demographics, TikTok serves as a valuable and expansive marketing platform. Its unique marketing channel offers unparalleled viral potential, surpassing what was previously available on other social media platforms. Through TikTok, businesses can showcase product tutorials, participate in challenges, capitalize on trends, entertain their audience, and thereby drive product sales, generate leads, and enhance brand awareness.

To maximize the impact of posted videos, it is essential to delve into TikTok's video metadata and discern the patterns that contribute to popularity. Analyzing this data allows content creators and brands to gain valuable insights into potential interactions a video may receive even before it is posted. Strategic decisions concerning hashtags, background music selection, and optimal upload timing further amplify the spread and interaction of the video. Moreover, by leveraging our experience and methodologies on TikTok, we can extrapolate valuable lessons that may be applied to other social media platforms, enhancing their content strategies and user engagement as well.

Background information:

While various metrics exist to assess success on TikTok, including likes, shares, comments and follower counts, the paramount metric for comprehensively and thoroughly evaluating video performance is the user engagement rate. This metric serves as the industry standard for assessing the impact of posted content. User engagement rate is calculated by the sum of engagement values (likes, shares, and comments) divided by the total number of followers for the account. Notably, TikTok posts boast the highest engagement rate per post compared to all other social networks. The platform's user-friendly creative tools, combined with the surging popularity of short videos and the potential for content to go viral, all contribute to encouraging user interactions with uploaded content.

Despite the abundance of user engagement analyses on other social media platforms such as Instagram, Twitter, or Facebook, TikTok has garnered comparatively fewer in-depth analyses. Existing analyses primarily focus on elements like the choice of background music or hashtags. In contrast, the present analysis, as detailed below, encompasses all aspects of video metadata and leverages a comprehensive model that considers all relevant features. By taking a more holistic approach, we aim to uncover deeper insights into user engagement on TikTok and provide valuable methodology applicable to a wider array of social media platforms.

Data Source:

The TikTok video data utilized for this analysis was collected using the TikTok Research API (unofficial version) on November 7th, 2020. The data collection initiative was initiated during a hackathon project undertaken by a group of data scientists, namely Ivan Tran, Kaushik Naresh, Isha Shah, and Madison Kohls. The detailed process and code employed for data collection were made publicly available on their GitHub repository. In a deliberate effort to ensure comprehensive representation, the data scientists randomly selected TikTok users, encompassing both influencers and regular users in their dataset.

The datasets are formatted as CSV files and comprise an extensive collection of more than 95,000 videos, collected from over 1,700 TikTok users hailing from more than 10 different countries. Each data entry constitutes the metadata of a TikTok video, encompassing vital information such as the video ID, upload time, username, a list of hashtags, background sound titles, video length, number of likes, shares, comments, views, and followers' information, among other relevant details. For a comprehensive understanding of the dataset's attributes, a data dictionary has been provided in the data preprocessing Jupyter notebook, accessible through my GitHub page.

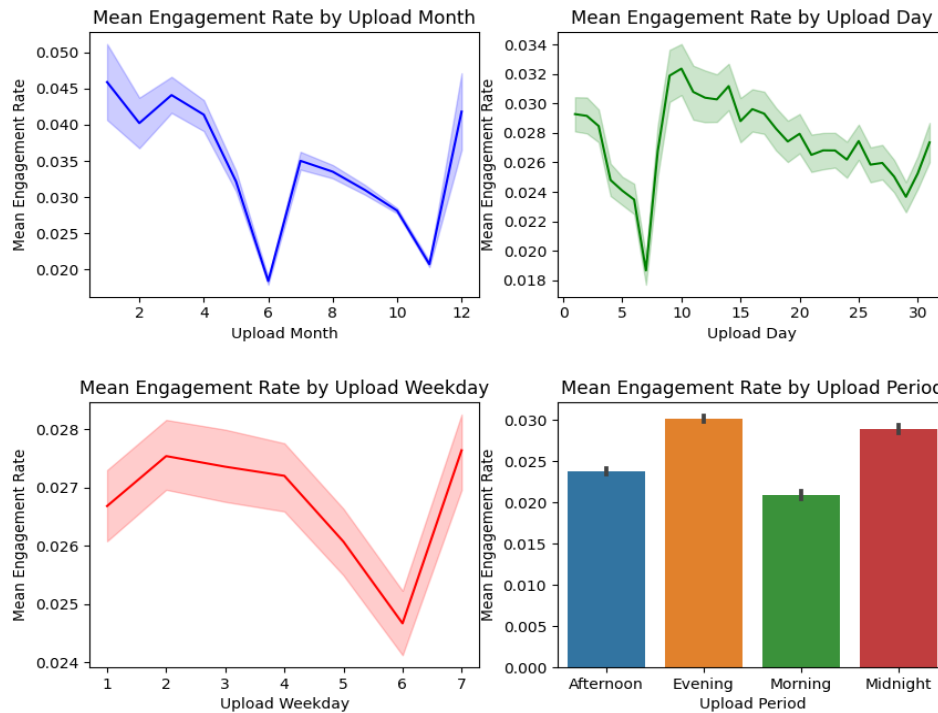
Data Preprocessing and Feature Engineering:

The original dataset underwent preprocessing to handle duplication, missing values, and inappropriate data types. New features, such as uploading year, month, weekday, and period, along with the engagement rate, were introduced to enhance machine learning understanding.

In pursuit of a comprehensive understanding of TikTok user engagement, we conducted an extensive exploratory data analysis (EDA) to extract valuable insights, discern trends, and comprehend the data distribution of TikTok videos. This rigorous examination was undertaken through a combination of summary statistics, data visualizations, and the investigation of temporal patterns. The culmination of this approach empowers content creators and marketers on TikTok with data-driven decision-making capabilities, propelling their content strategies towards increased user engagement and enhanced brand visibility.

Key findings:

Temporal Patterns Driving Engagement:



The analysis reveals essential temporal patterns that impact user engagement with videos. January emerges as the most engaging month, while June and November record the lowest engagement rates. Additionally, videos uploaded during the middle days of the month (10th to 15th) receive higher engagement

compared to early or late days. Furthermore, Sunday stands out as the day with the highest engagement rate, despite having fewer video uploads, while Saturday experiences the lowest engagement rates despite higher upload numbers. Moreover, the evening hours demonstrate the highest user engagement, followed by the midnight, afternoon, and morning periods. Understanding these temporal preferences can inform content creators' strategies to optimize user engagement and interaction.

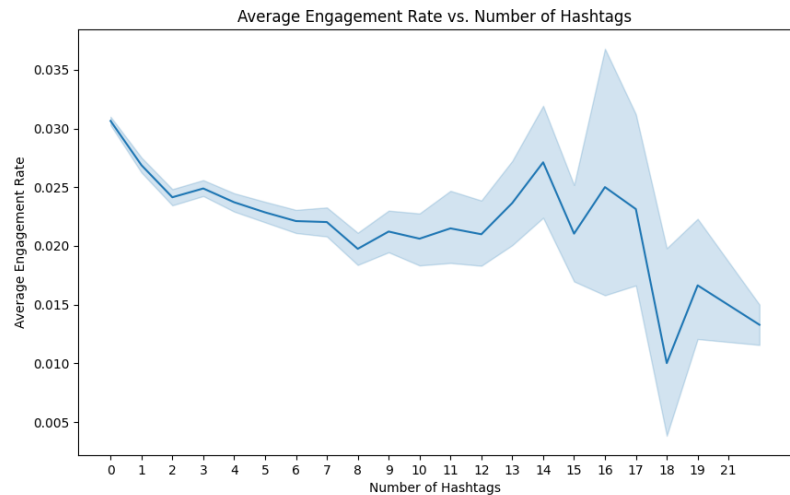
Hashtag Strategies and Engagement:

The data analysis reveals an intriguing relationship between the number of hashtags used and the engagement rate of videos. Surprisingly, videos with 0 hashtags achieve the highest engagement rate, while those with 1 to 10 hashtags display relatively similar engagement rates. However, as the number of hashtags exceeds 11, the engagement rate exhibits significant fluctuations with a considerable amount of variance.

These findings suggest that there is no straightforward linear correlation between the number of hashtags and engagement rate. Instead, the optimal number of hashtags for maximizing

engagement remains elusive due to the observed trends. Nevertheless, based on the data, a prudent choice for content creators would be to use hashtags within the range of 0 to 5. This range appears to strike a balance between attracting potential viewers through hashtags and avoiding excessive usage, which may lead to unpredictable and fluctuating engagement rates.

Furthermore, it is noteworthy that even without strategic hashtag use, videos created spontaneously by regular users can still garner significant engagement. This highlights the organic and user-driven nature of the platform, where engaging content can thrive regardless of a meticulously crafted hashtag strategy.



Prominent Hashtag Trends:

The analysis of top hashtags on TikTok highlights their critical role in content promotion and audience targeting. Notably, hashtags such as #fyp and #foryou emerged as dominant contenders, indicating their widespread usage to garner more views on TikTok videos. Additionally, the presence of geographically specific hashtags like #tiktokindia suggests the active participation of Indian users on the platform.



Certain hashtags, such as #trending, #duet, #dance, #comedy, #funny, #halloween, and #love, represent distinct video categories, effectively capturing audience attention and catering to specific interests.

Furthermore, sponsored content hashtags like #ad and #bangenergy underscore the platform's potential for brand marketing and the inclusion of advertisements on TikTok.

The insights gained from analyzing these top hashtags offer valuable information on the motivations behind their usage, encompassing the pursuit of attention, geographic associations, category targeting, and the incorporation of sponsored content. This understanding equips content creators, marketers, and platform users with valuable knowledge on prevalent trends

and practices, enabling them to strategically utilize hashtags to maximize engagement and effectively reach their desired audience.

Impact of background music:

The analysis of top-performing videos based on engagement rate reveals five prominent song titles that have consistently garnered high user engagement. These song titles are "Crying in the Club," "We Are the Champions," "@JacksonWang 創作的原聲" (original sound created by Jackson Wang), "Say a Little Prayer" (Glee Cast Version), and "Original Sound - Advocate Bhardwaj." These videos have achieved an average engagement rate exceeding 10%, indicating their widespread appeal and resonance with the TikTok audience. Content creators may find these song choices particularly effective in enhancing user engagement and overall video performance on the platform.

Modeling and results:

Among the five trained and tuned machine learning models on TikTok video features, the XGBoost (eXtreme Gradient Boosting) Regressor model has demonstrated the highest level of performance. This powerful model offers users the capability to gain insights into their potential engagement rate by adjusting various factors, including the number of hashtags, upload time and period, as well as basic information related to their account, such as total followers and total videos posted. By utilizing this model, users can effectively analyze and understand the impact of these parameters on their video engagement, empowering them to optimize their content strategy and enhance their overall presence on the platform.

Finding and conclusions:

The insights obtained from our comprehensive data analysis present content creators and marketers on TikTok with a strategic roadmap. By comprehending temporal patterns, they can adopt tailored posting strategies, while hashtag and background music analysis empowers optimized discoverability and audience targeting. Leveraging these valuable insights enables businesses to capitalize on TikTok's exceptional user engagement potential, establishing a robust brand presence and fostering meaningful connections with their target audience. As TikTok continues to evolve, the application of data-driven insights will prove instrumental in driving innovation, ensuring relevance, and optimizing engagement in this ever-changing social media landscape.

For the next phase of analysis, we propose two key steps to further enhance our understanding. Firstly, exploring the content of hashtags and conducting segment analysis can provide deeper insights into user preferences and interests, allowing for more targeted content creation. Secondly, incorporating real-time video data from TikTok and utilizing time-series models for

forecasting have potential to significantly improve the accuracy and thoroughness of our predictions, enabling content creators and businesses to make more informed and data-driven decisions. These additional analyses will contribute to an even more robust and effective model, further enhancing the value of our insights in the dynamic and rapidly evolving TikTok environment.

Appendix:

My GitHub:

https://github.com/tianlan8/TikTok_Engagement_Rate_Analysis_Prediction

Original data source:

https://github.com/datares/TikTok_Famous/tree/main

Reference:

<Machine Learning Yearning> by Andrew Ng

<https://www.kaggle.com/code/julienjta/tiktok-popularity-prediction-stackblend-etc/notebook#2.-Prediction-of-the-popularity>

<https://www.kaggle.com/code/antoniosabatini/tiktok-popularity-track-eda-ml-models/notebook#notebook-container>

https://www.kaggle.com/code/dataranch/tiktok-popular-songs-feature-importance/notebook#**Overall-Feature-Importance**

https://github.com/datares/TikTok_Famous/tree/main/Analysis/TikTok%20Videos

<https://www.kaggle.com/code/vbradculbertson/tiktok-engagement-forecasting-and-analysis/notebook#Random-Forest>

<https://www.kaggle.com/code/erikvdven/tiktok-some-python-magic-in-a-notebook/notebook#Some-first-Analysis-%F0%9F%93%88>

https://www.researchgate.net/figure/Learning-curves-can-be-used-to-examine-the-behavior-of-a-neural-network-model-during_fig9_349898718