



TikTok User Engagement Dataset

About

A dataset detailing user engagement on TikTok, including metrics like view counts, like counts, shares, comments, and hashtags.

Creators: We fused data from Kaggle/Github respectively

Content: CSV

Location: <https://github.com/INFO-201-Fall-2023-Final/final-projects-TonySUW.git>

Date: NOV 27 2023

Methodology

Data Collection Methods: This dataset was collected from publicly available TikTok posts, focusing on user engagement metrics.

Data Processing and Cleaning: We sample videos based on similar number of likes in both datasets so that we can learn the elements that people like based on popular tiktok popular videos, such as the bgm of the video, the textual content, and so on.

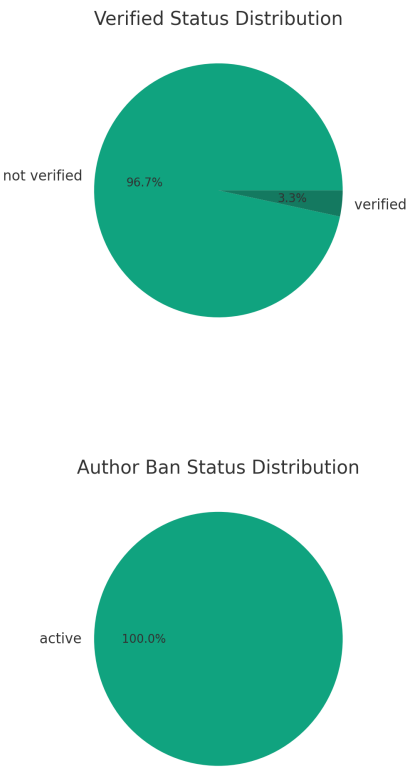
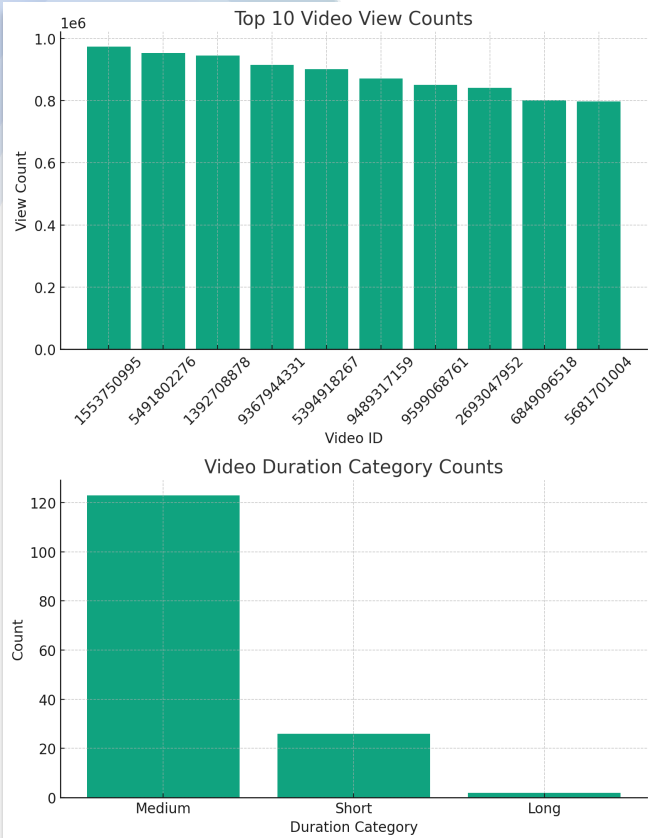
Dataset Structure

Number of Records: Total 151 rows after consolidation of data

Data Fields: Description of each column, such as video ID, video duration, transcription text, view count, like count, comment count, hashtags, etc.

Format: CSV

Here is a rough chart based on the data



Bar Chart of Top 10 Video View Counts: This chart shows the view counts of the top 10 videos, identified by their video IDs.

Pie Chart of Verified Status Distribution: This chart illustrates the proportion of videos that are verified versus those that are not.

Bar Chart of Video Duration Categories: This chart displays the number of videos in each duration category (e.g., Short, Medium, Long).

Pie Chart of Author Ban Status Distribution: This chart shows the proportion of videos from authors who are either active or banned.



Data Subject Description

In the context of growing social media, the short-form video app TikTok has become a global phenomenon, attracting millions of users with its addictive content. Our data-driven research focuses on exploring TikTok's addictive behavior, particularly the complex relationship between message quality and system quality. This work not only responds to a general problem that affects a wide range of people, but also provides a unique perspective on our understanding of how social media affects users' daily lives. Our main goal is to shed light on the addictive trends associated with TikTok, particularly from the perspective of system and information quality, and to identify factors that enhance the platform's appeal and possible negative impacts. Through in-depth analysis of user engagement patterns and content trends, we will explore how information and system quality influence addictive behaviors and consider comparisons of TikTok with other social media platforms to provide a more comprehensive framework for scholars and policymakers. Our study challenges the conventional view that social media addiction is caused only by interesting content, using data evidence to demonstrate the impact of system quality. Overall, our study not only provides unique insights into the TikTok phenomenon, but also highlights a general societal issue, aiming to help readers fully understand the dynamics of addiction and think about the role of social media in their lives by incorporating real-life cases and statistical analysis.

Usage and Applications

Intended Use: This data can help businesses and content creator to understand what is trending at the moment as a way to create their content, for example, businesses can also get this data to decide what they use to advertise their products, and creators can know what content catches the eye of users.

Potential Applications: Examining the Impact of Topic Tagging and Studying Content Virality

Ethical Considerations and Human Rights

Privacy and Confidentiality: Discuss the steps taken to anonymize user data and ensure the privacy of TikTok users.

Consent and Data Collection Ethics: Since the data is public, detail the ethical considerations involved in using such data.

Non-Discrimination: Ensure that the dataset and its potential applications do not lead to discrimination against any group.

Other Human Rights Principles: Discuss how the dataset aligns with transparency, accountability, and respect for human dignity.

Management

How to manage data at rest and in transit?

Internally

Will the dataset be distributed to any other individuals/third parties (e.g., companies, institutions, organizations)? If yes, when and how will the dataset be distributed (e.g. github, zip on website, etc.)?

Yes, Github

Will the dataset be updated ?

NO

Is there a mechanism for individuals to request removal of their personal information?

Not applicable.

Related Literature and References

TikTok User Engagement Data, a dataset authored by Yakhyojon, has data on users of published videos.

top_users_vids and sug_users_vids_all are two datasets from Github's Tiktok_Famous, whose contributors include Ivan Tran, Madison Kohls, Kaushik Naresh, and Isha, with data on the popular video tiktok used by background music, and what the hashtags are