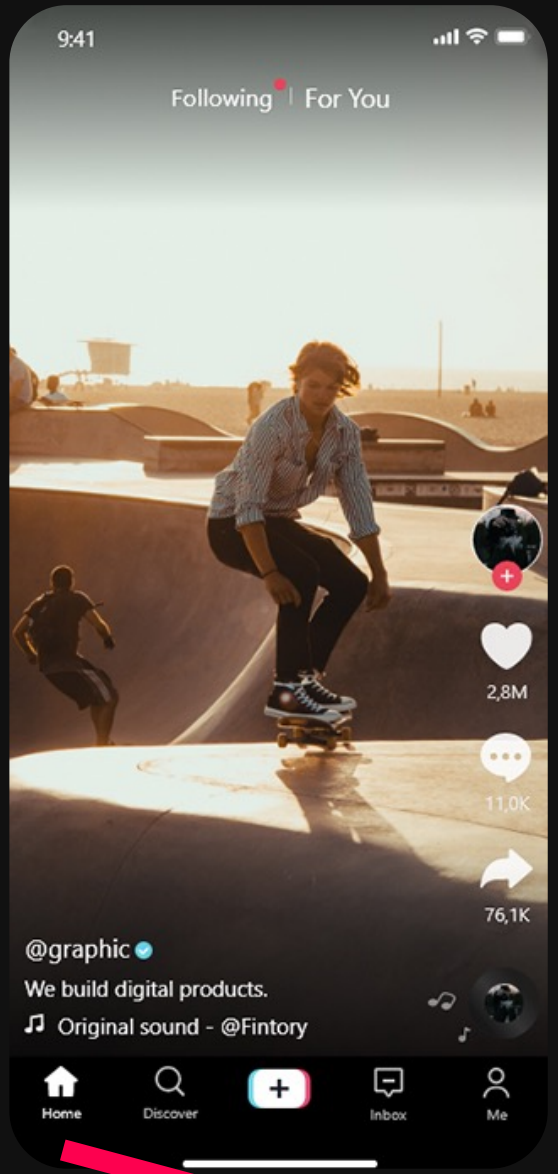
TikTok

User Engagement Analysis & Prediction

Tian Lan

"TikTok has 1.06 Billion active users worldwide"

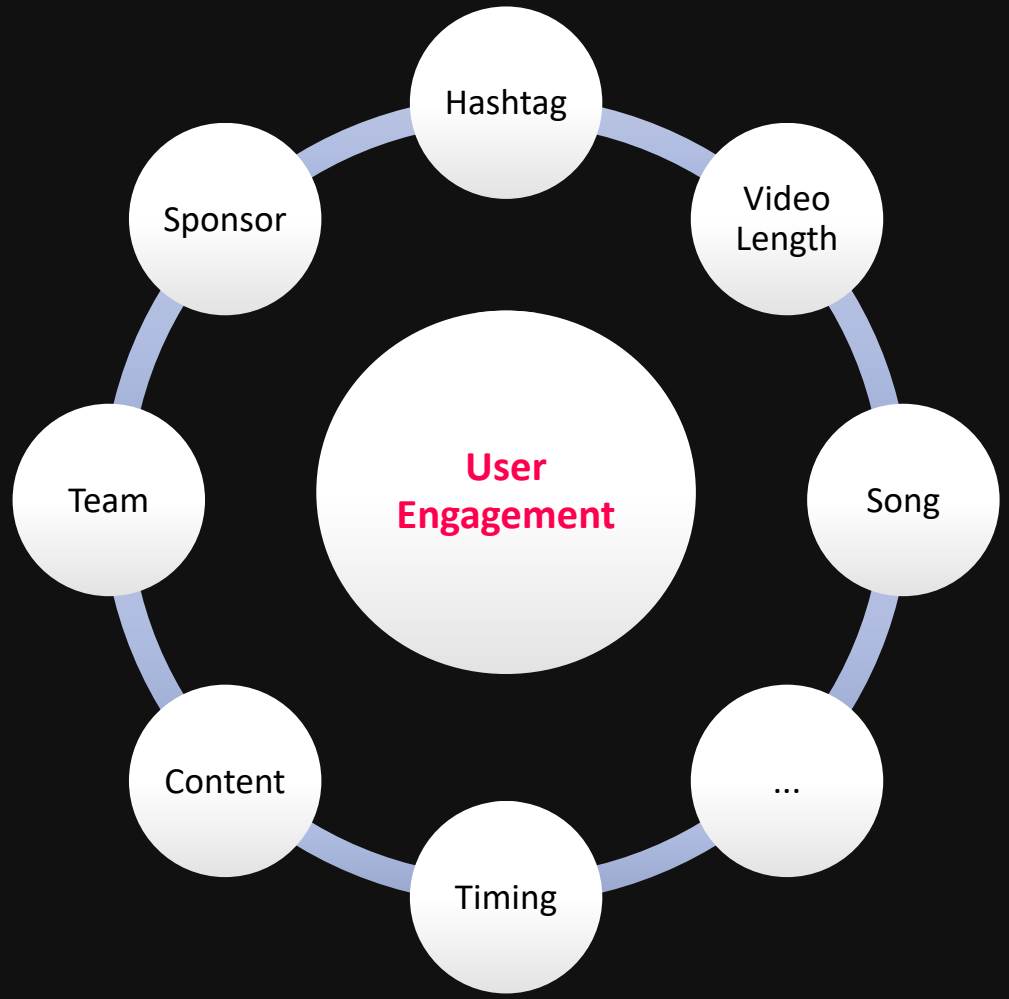


Want to *go viral*?

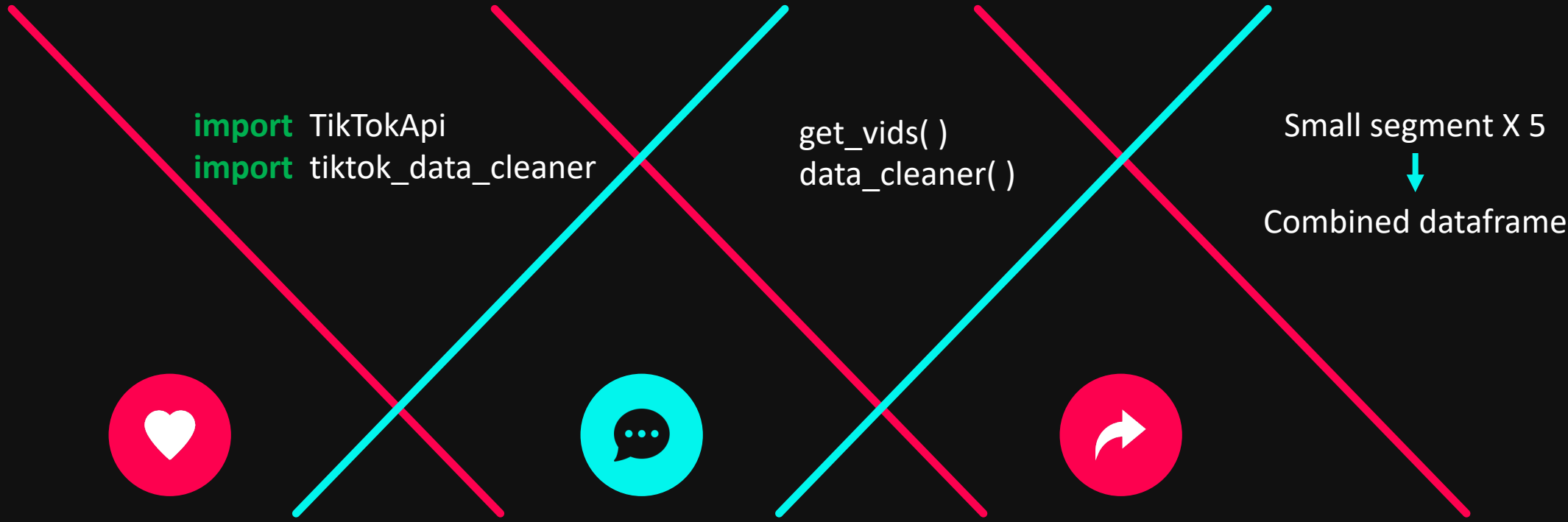
2.8M

11.0K

76.1K



TikTok Research API (unofficial version)



Data Description

Column Name	Description
ID	Video identification number
Create Time	Unix datetime for the upload of the video to the TikTok app
User	Creator username
Hashtags	Hash keywords applied to the video description to influence TikTok algorithm
Song Title	Sound applied to the video
Length	Length of the video in seconds
Likes & Shares & Comments & Views	Number of Likes & Shares & Comments & Views the video received from other users
Followers	Number of TikTok users who follow the creator's account
Total Likes	Total likes from other users on all creator's videos
Total Videos	Total number of videos uploaded by the creator

Each datapoint: TikTok video metadata

Some New Features:

- Number of hashtags
- Bag of words for hashtags
- Number of times a song be used
- Total Engagement
- Days of trends
- …

Target: Engagement Rate

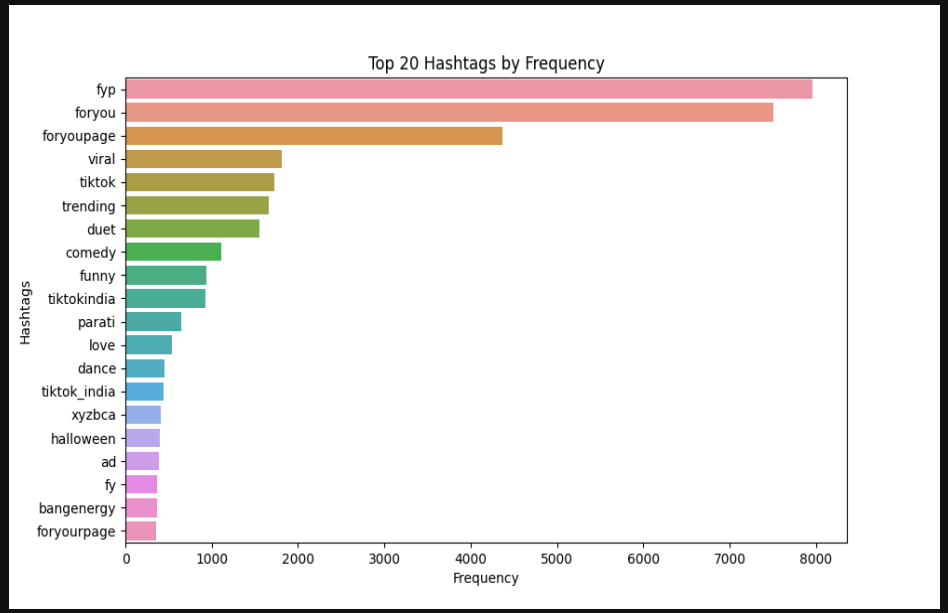
(Social Media Industry Standard)

Data Preprocessing

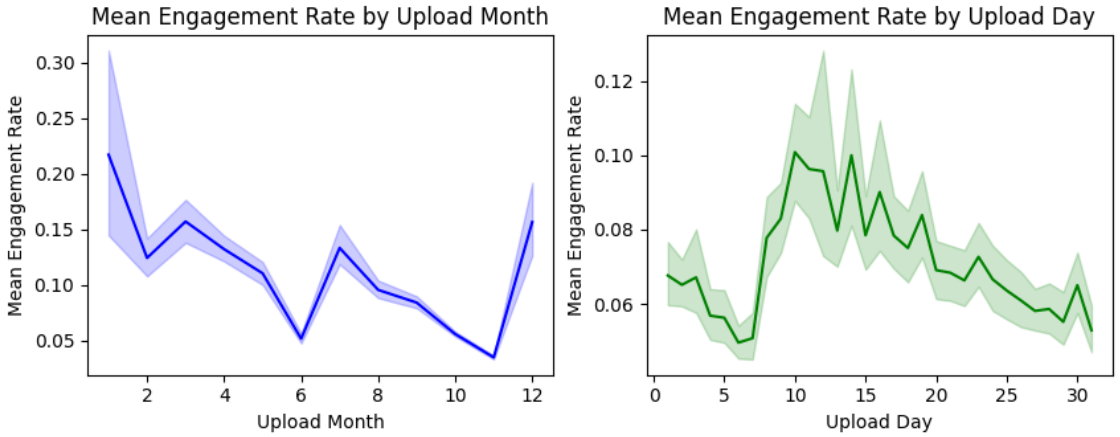
Original Shape: 95,963 rows X 13 columns

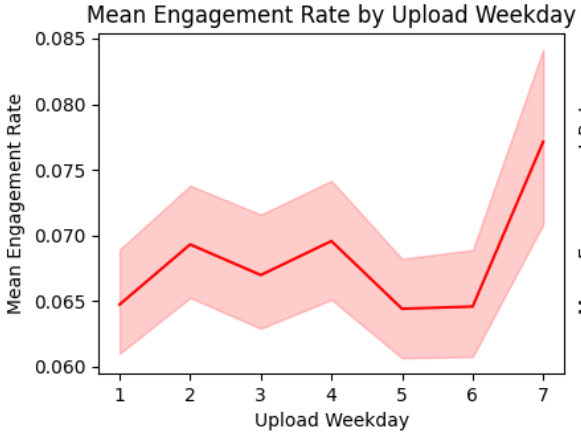
- Duplication: 43,119 rows
- Missing value: 42 NaN value in Song column
- Refine data type
- Redundant columns
- New calculated features

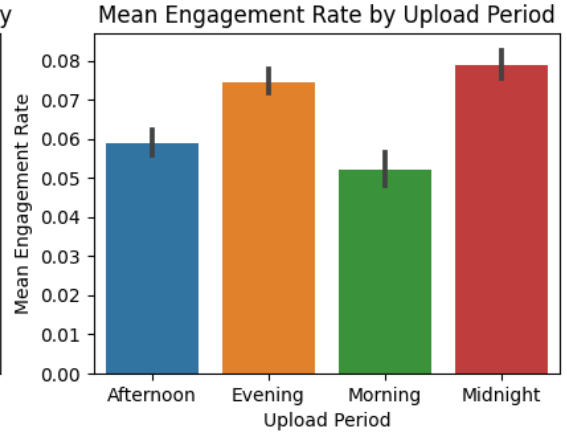
Cleaned dataframe: (52,844, 20)

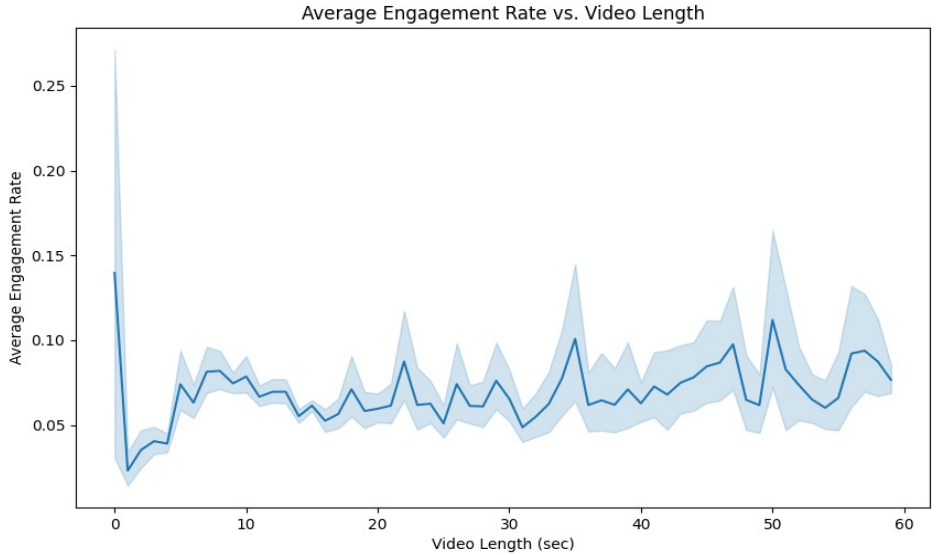


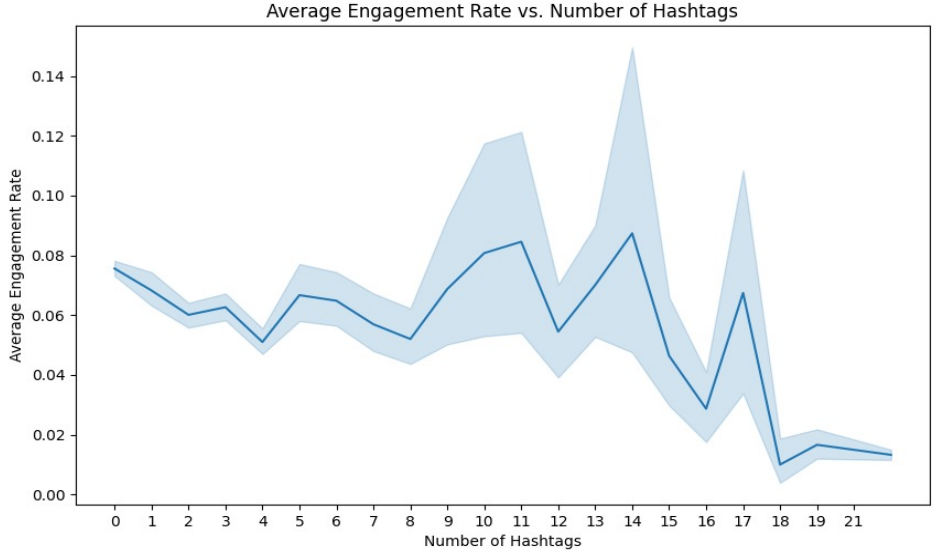
EDA & Insights









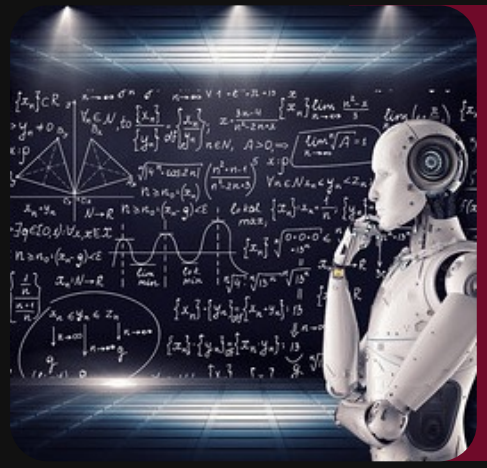


Next Step

Models:

- Random Forest Regression
- Gradient Boosting Regression
- XGBoost & LightGBM

- Real-time data
- SARAMAX or ARIMA modeling
- Time series forecasting



- Compare the performance of models
- Best way to present the model

- Visualization
- Presentation
- Report

Thanks!