Identifying Factors that Boost YouTube Video Engagement

YouTube User Engagement Study

01

Background

Background

Growing interest in factors contributing to YouTube success

Existing studies on social media post timing

Aim to assist creators with controllable variables

02

Dataset

"Trending YouTube Video Statistics" dataset from Kaggle

NEXT

Mainly focus on the subset of the United States.

Key features: video title, video category, publish time, views, likes, dislikes, number of comments

NEXT

Data Cleaning and Preparation

Step 1

Eliminated missing values, duplicates, and negative values

Step 2

Applied Winsorization to address outliers in views

Step 3

Generated additional variables: like-percentage and like-dislike ratio

03

Analysis Methodology

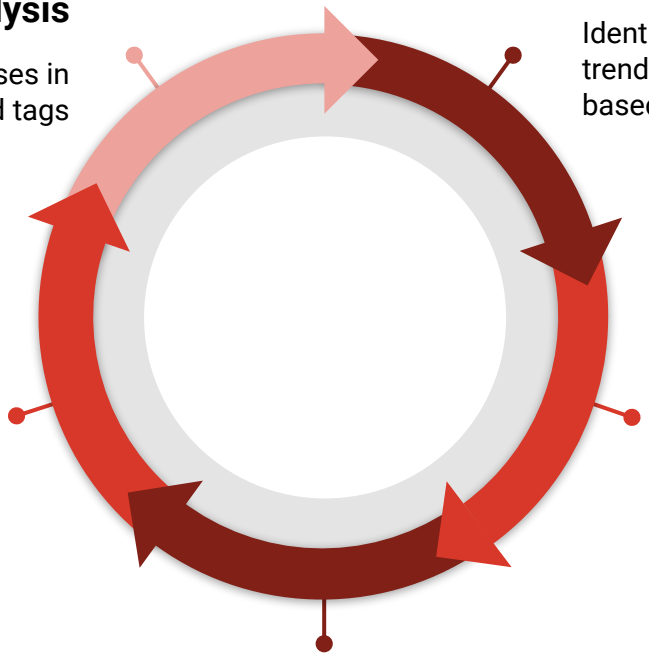
Analysis Overview

Bigram analysis

Identified common phrases in title and tags

Time of day analysis

Identified daily and hourly trends in engagement metrics based on historical data



Title length analysis

Examined relationship between title length and like-to-view ratio

Sentimental analysis

Exploring video title's sentiment impact on video engagement

Predictive analysis

Utilized a SARIMAX model to forecast the best time frame to attract audience

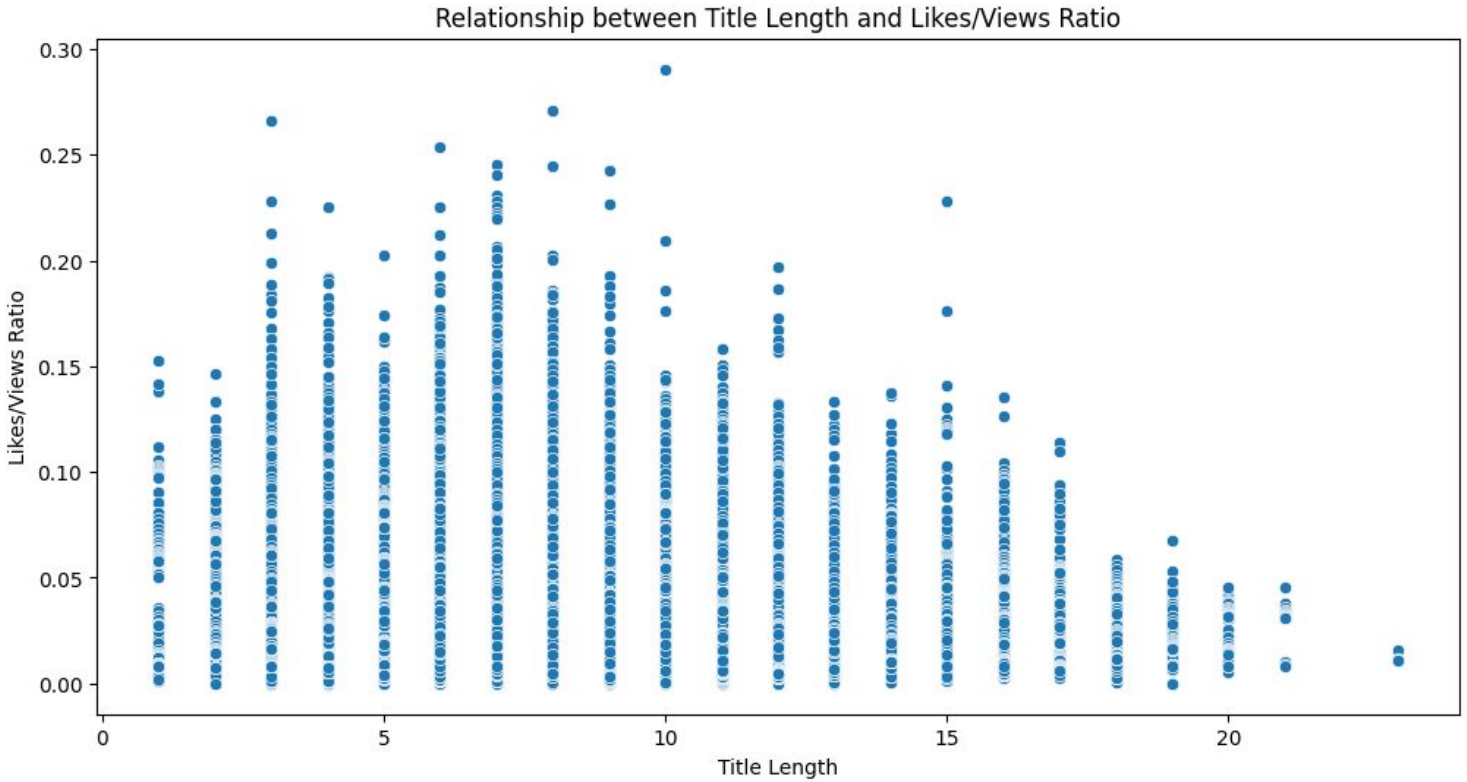
04

Observations

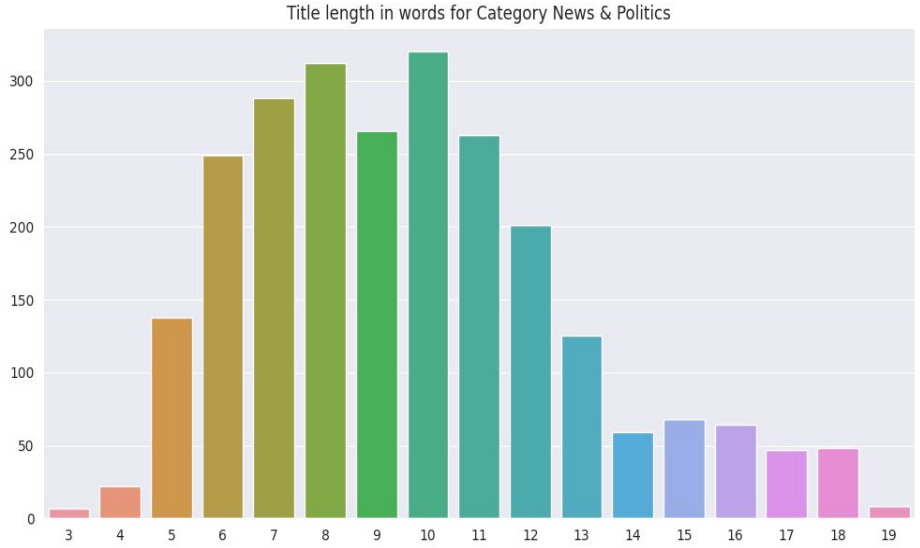
Bigram Analysis - Education

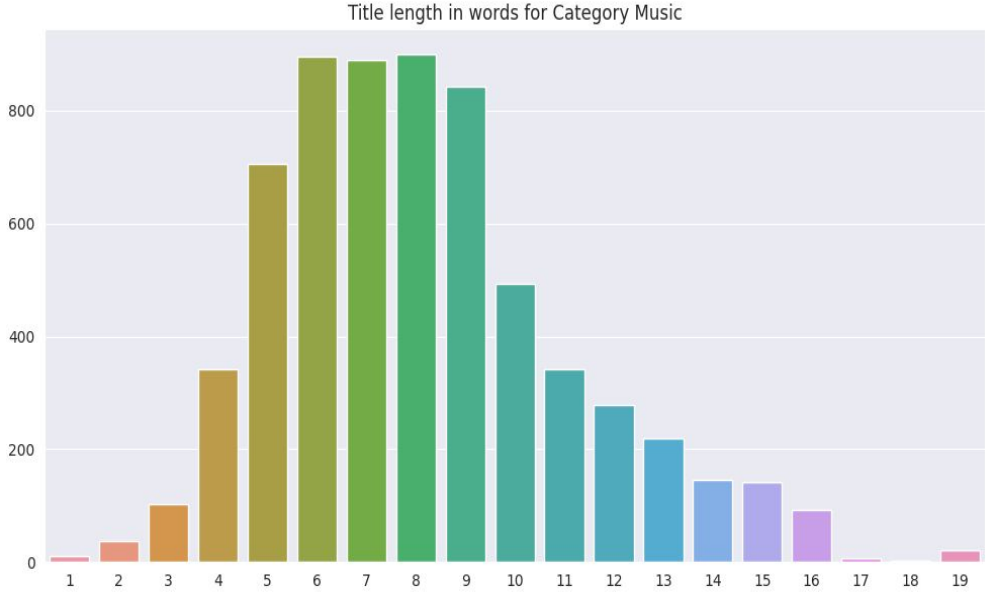
Bigram Analysis

Title Length Analysis



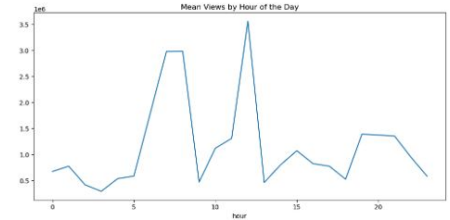
Title Length Analysis

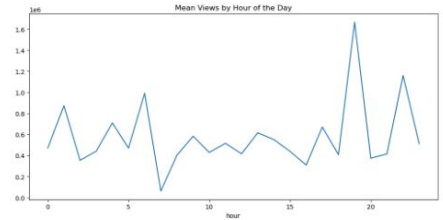


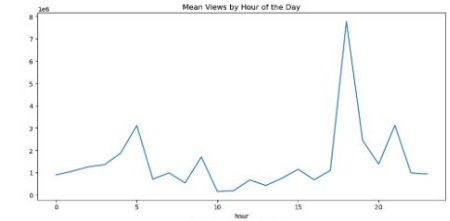


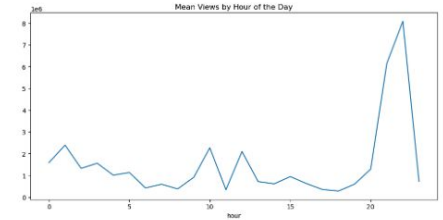
Publish Time of Day Analysis

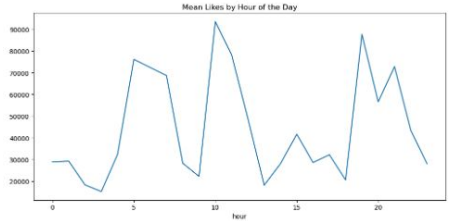
- Daily peaks in views, likes, and comments across all categories
- Peak times vary among different categories

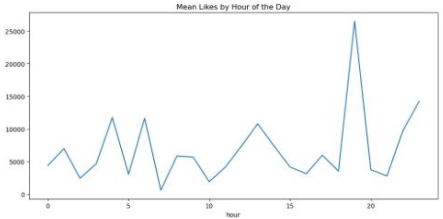


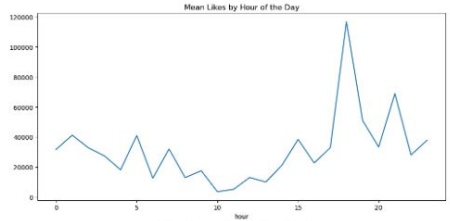


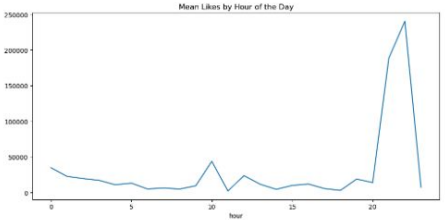


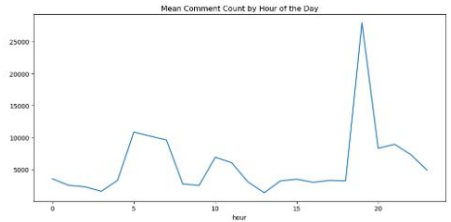


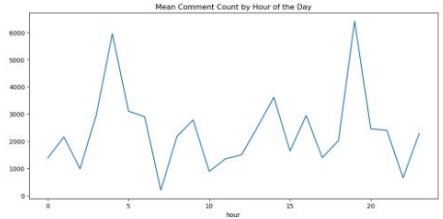


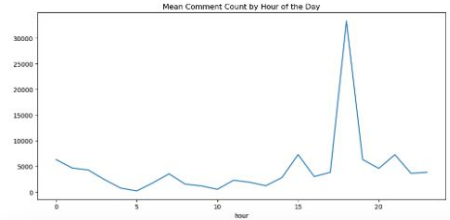


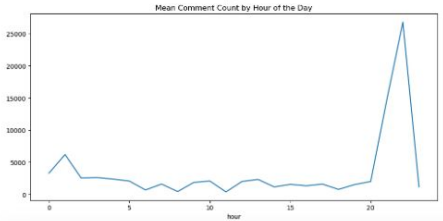






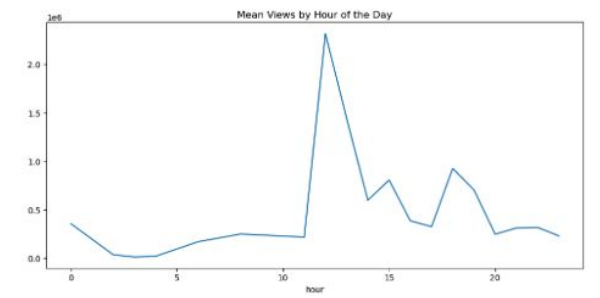


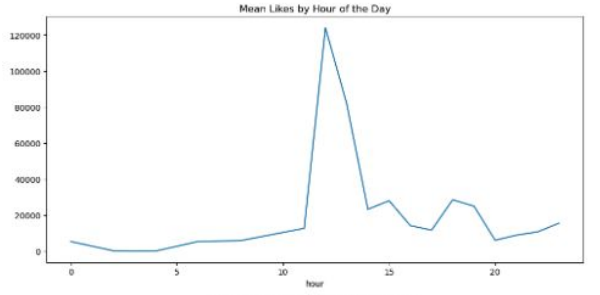


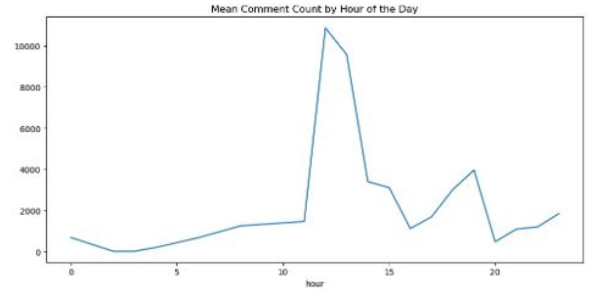


Publish Time of Day Analysis - Education

- Certain top bigrams correlate with content nature and user engagement patterns





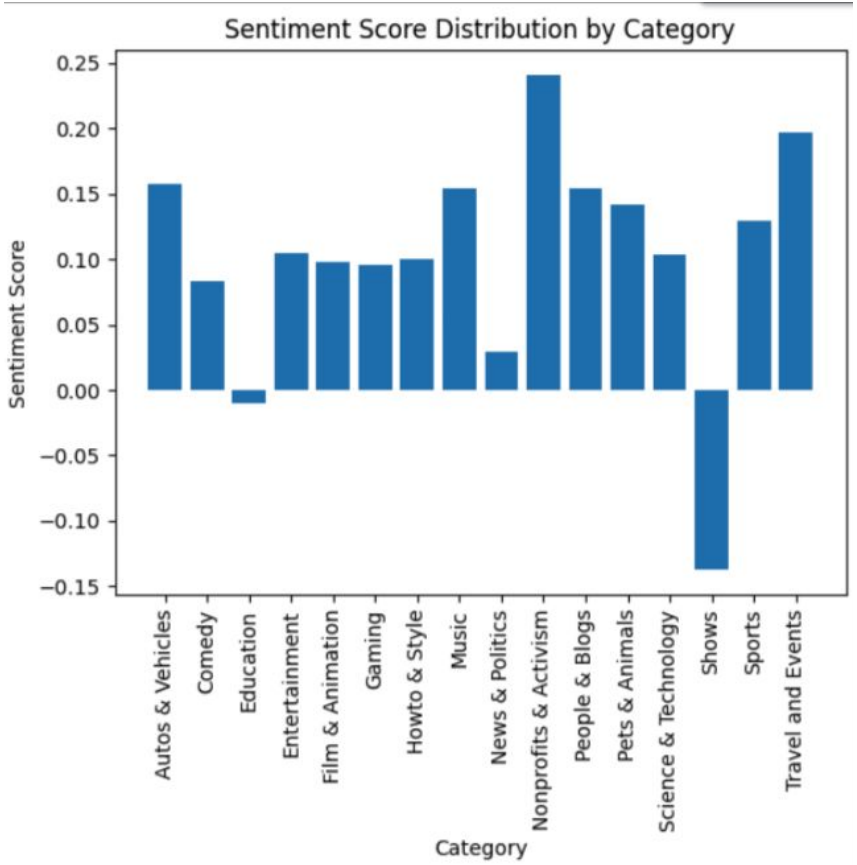


Predictive Analysis Using Time Series

Time Series:
Leveraging a SARIMAX model, we predicted the views, likes, and comments for different days and times for the upcoming week, aggregating the predictions by time block (morning, afternoon, and night). This allowed us to determine the optimal time to post content based on the maximum predicted value for each metric in different time block.

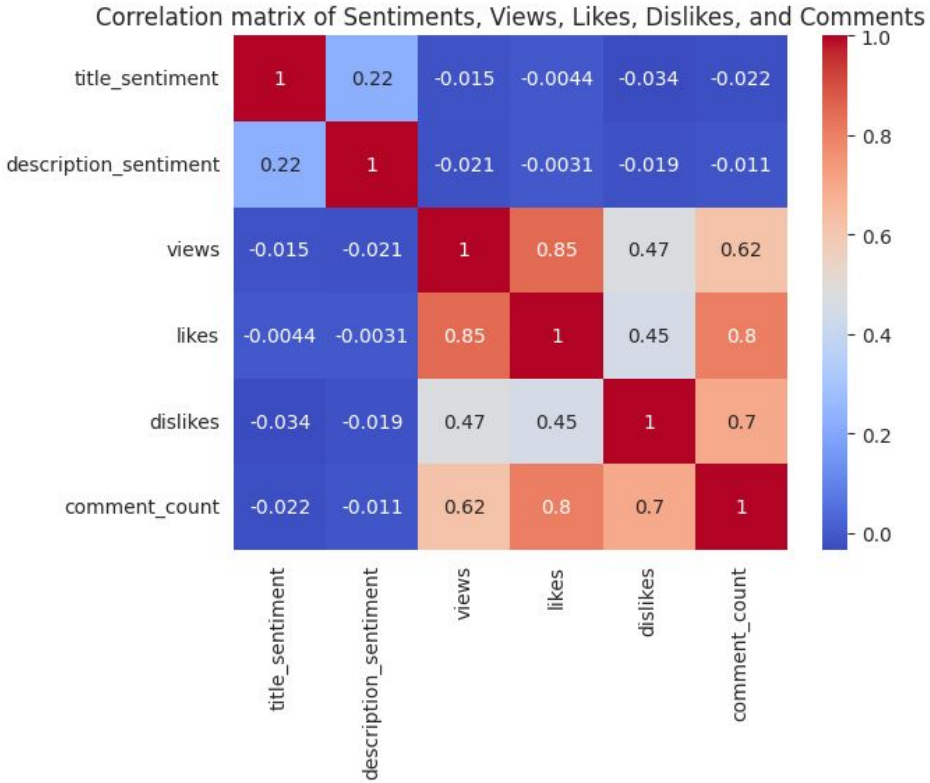
Predictive Result:
The best time to post for maximum views is 1 PM in the afternoon, for maximum likes is 12 AM at midnight, and similarly for maximum comments is 12 AM at midnight.

Sentiment Analysis on Title & Title Description



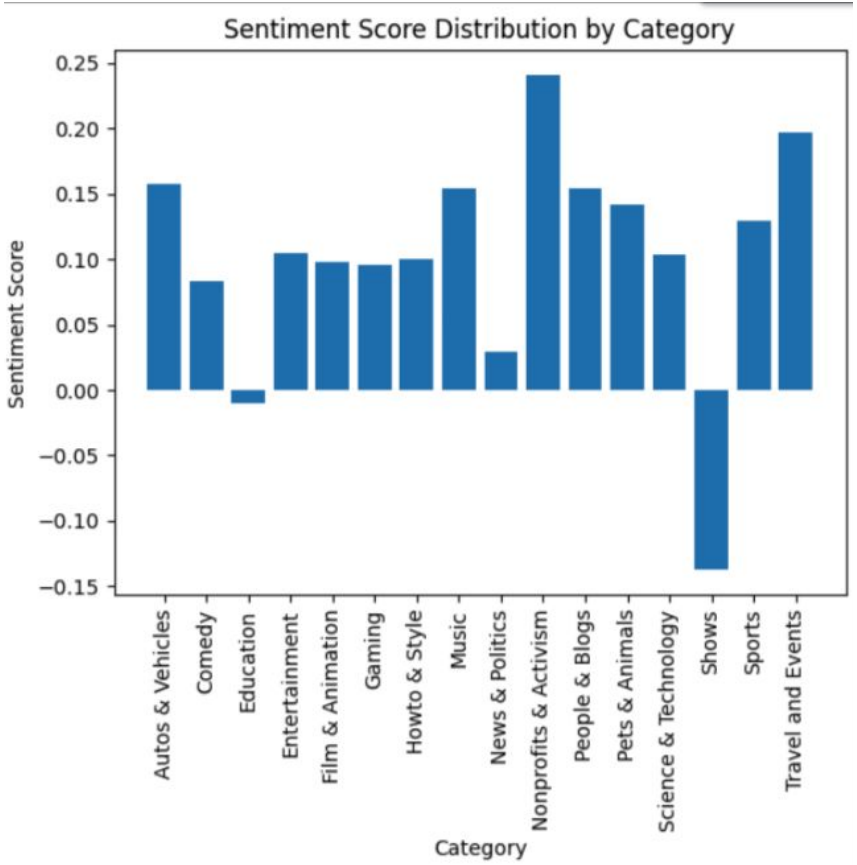
- a python library **TextBlob** to generate sentiment analysis.
- In addition, we ran **confusion matrix in heatmaps** to analyze if any correlation exists between likes and title sentiments and description_sentiments

Sentiment Analysis on Title & Title Description



- No direct influence from title sentiments over views and likes
- Still crucial to have a more appealing and positive title and description overall

Sentiment Analysis on Title & Title Description

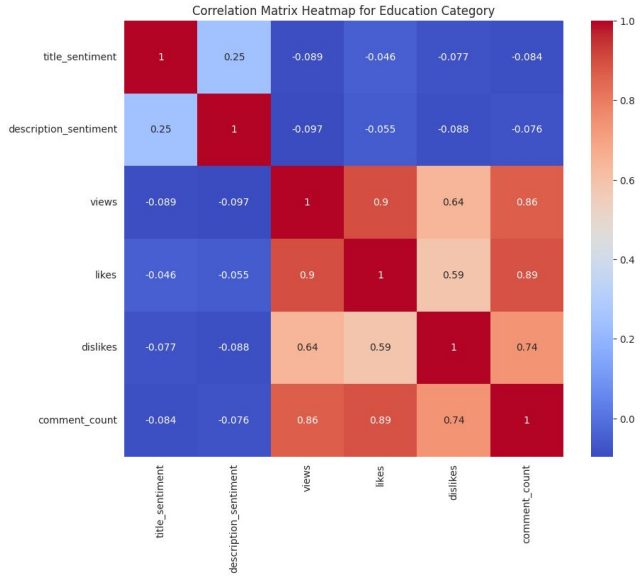


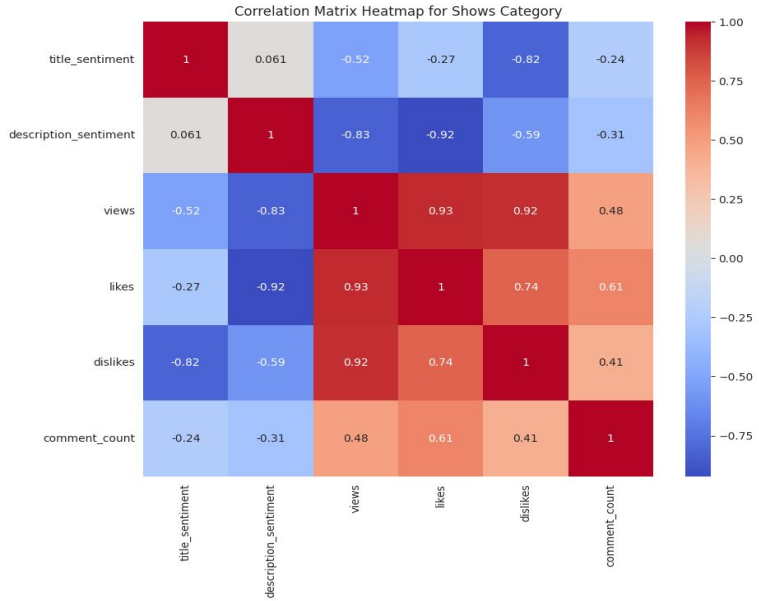
	title	category	title_sentiment
26612	Meet World ' Terrifying Caterpillar	Education	-1.0

- Negative sentiments in ‘Shows’ and ‘Categories’
- Mostly negative sentiment words are adjectives to exaggerate in order to catch user attention, such as words like “Terrifying, deadliest”

Sentiment Analysis on Title & Title Description

- In 'Shows' and 'Education' categories, exaggerate Words like “terrifying” might stimulate a polarized opinion in users, thus we see sentiments negatively correlates to both likes and dislikes
- More neutral or negative sentiment might stimulate more engagement.





05

Conclusion

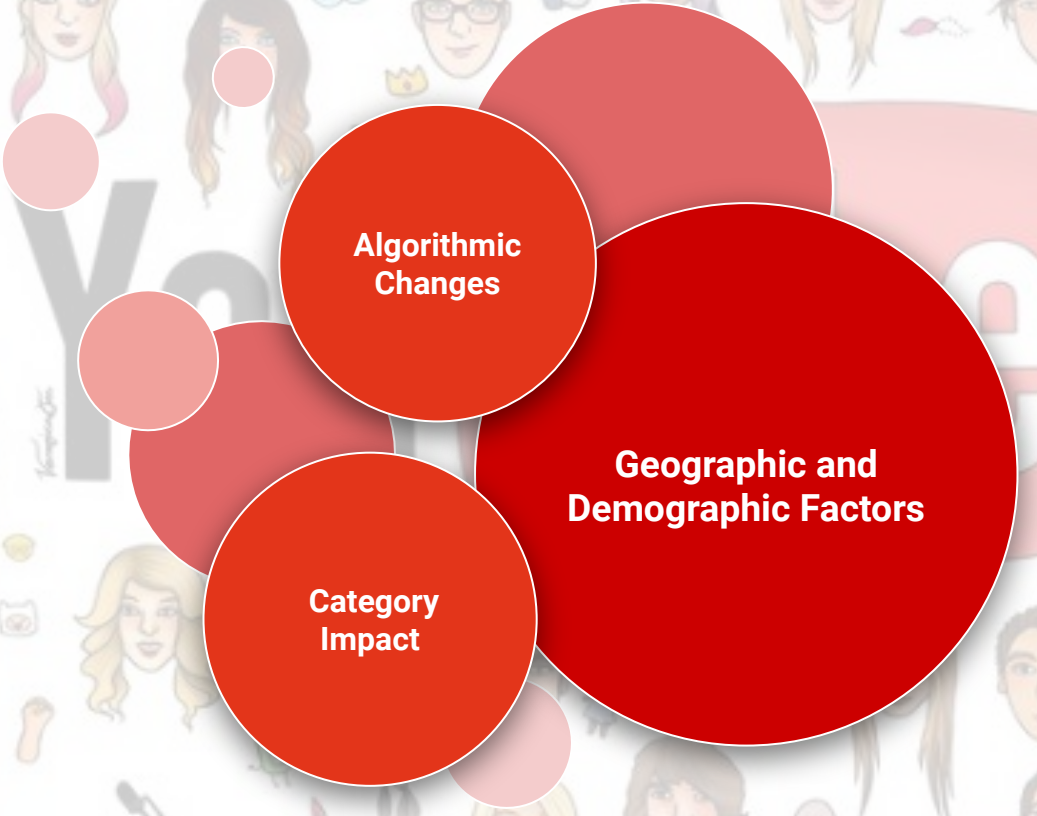
Conclusion

To increase engagement metrics without modifying contents of the videos, YouTubers could utilize the following insights generated from our project:

- Posting Time:
  - Max views: 1 PM in the afternoon
  - Max likes and comments: 12 AM in the midnight
  - Various across categories

- Title and Description:
  - General appealing wording
  - Title length between 6 to 10 words

Discussion: Further potential areas



06

Reference

Reference

- Singh, N., Jaiswal, A., & Singh, T. (2022). Best time to post and review on Facebook and Instagram: Analytical evidence. South Asian Journal of Management. https://doi.org/10.1108/SAJM-09-2022-0059

- Statsmodels Developers. (n.d.). SARIMAX: Statsmodels tsa statespace sarimax. Retrieved from https://www.statsmodels.org/stable/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html

- Brownlee, J. (2018). Forecasting time series data with SARIMA and Python. Machine Learning Mastery. Retrieved from https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, B. E. (2015). Time Series Analysis: Forecasting and Control.

THE END

THANKS