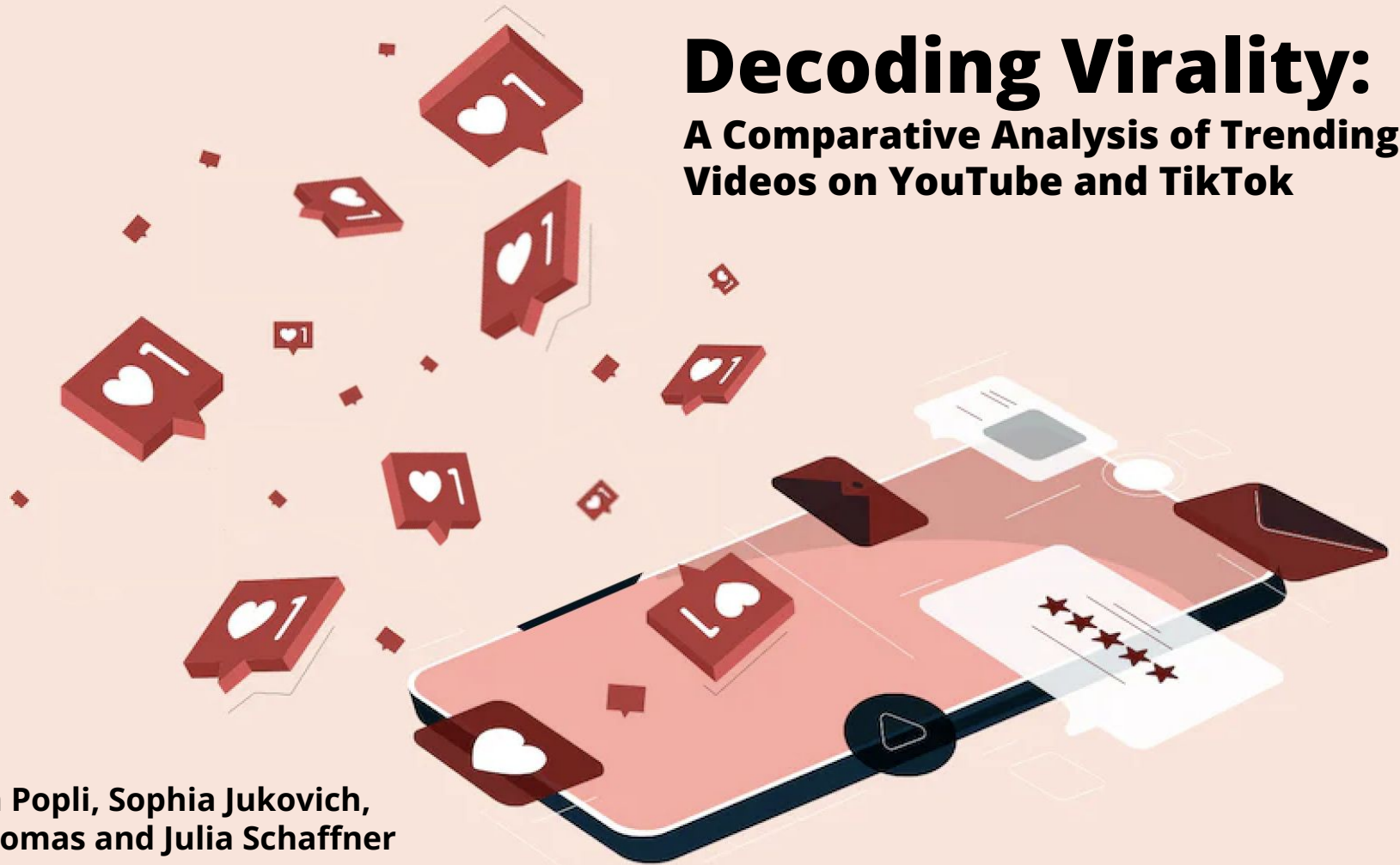


Decoding Virality:

A Comparative Analysis of Trending Videos on YouTube and TikTok



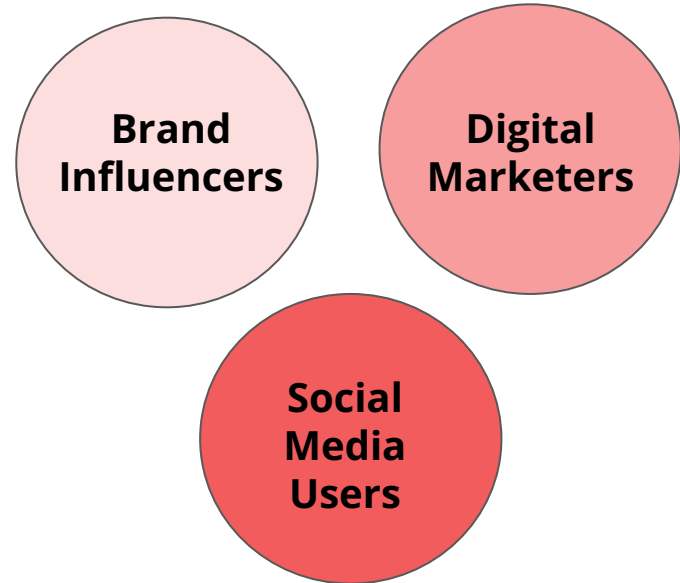
**By: Varun Popli, Sophia Jukovich,
Tatum Thomas and Julia Schaffner**

Motivation & Stakeholders

Our Motivation

- Social media influencing has emerged as an occupation in the digital age
 - Achieving this level of success requires individuals to 'go viral'
- We wanted to understand what factors contribute to content becoming viral on social media
 - We wanted to analyze our favorite video platforms, TikTok & Youtube, to look at the elements that contribute to content's viral success

The Stakeholders



Data Set & Source

YouTube

- Daily records of the top trending YouTube videos in the United States
- Kaggle
- Last Updated: 2019

TikTok

- Records of different trending videos on TikTok across the app
- GitHub
- Last Updated: 2020

About the Data Sets / Cleaning

- 1 40379 Observations in YouTube, 41675 Observations in TikTok
- 2 Cleaned data sets by using `drop.na()` to drop the rows with missing values
- 3 Dropped rows included missing values from descriptions (YouTube) and songs (TikTok)

Question #1:

What are the Characteristics of Virality?

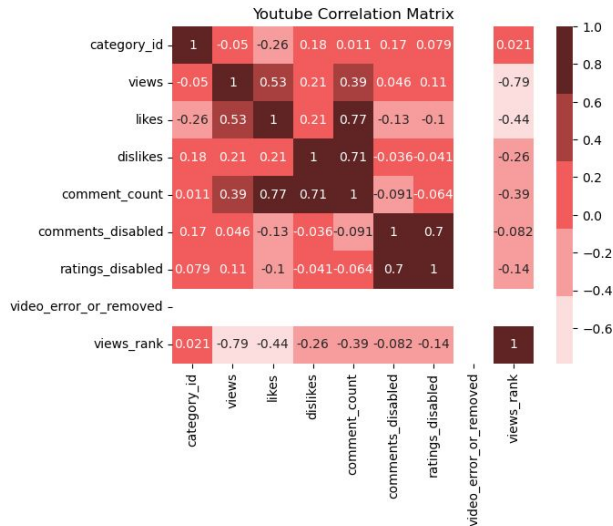
Correlation of YouTube Variables



Methodology

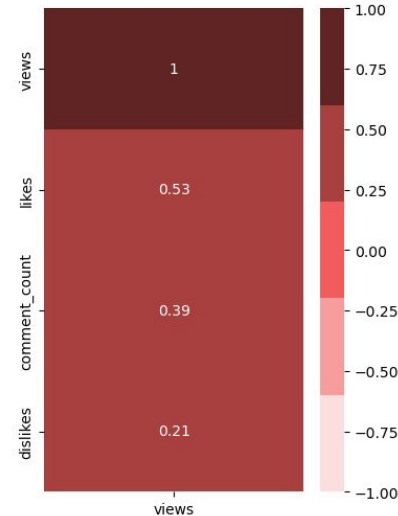
- Correlation Analysis of Numeric Variables
 - Filtered 100 most viewed videos by dropping duplicates, ranking by views, and sorting
 - Seaborn, built in functions for corr and organization
 - Visualizations of all numeric variables and a more focused one for views vs highly correlated variables
- Statistical Analysis of Views
 - Seaborn, Matplotlib, SciPy (for slope calculations)

Numeric Variables



Views vs. Most Correlated Variables

Characteristics of Virality Correlating with Views



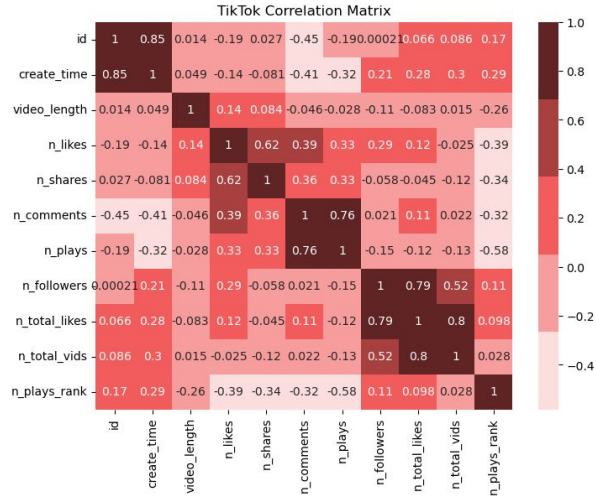
Takeaways

- The strongest correlation with number of 'views' was number of 'likes' with a correlation of 0.53
- Views, likes, comment_count, and dislikes all correlate with each other meaning that these variables help with virality of a video

Correlation of TikTok Variables

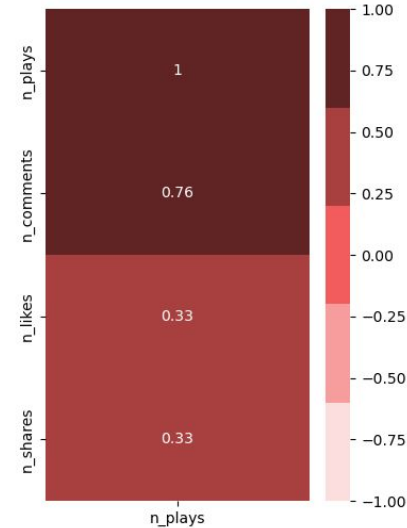


Numeric Variables



Views vs. Most Correlated Variables

Characteristics of Virality Correlating with Number of Plays



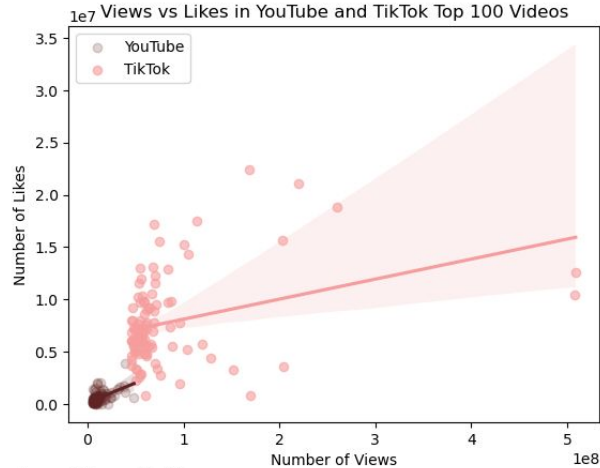
Takeaways

- The strongest correlation with number of 'n_plays' is number of 'n_comments' with a correlation of 0.76
- There is a correlation between n_plays, n_likes, n_comments, and n_shares meaning there is a factor that correlates with virality within these variables

Comparison of Most Correlated Variables

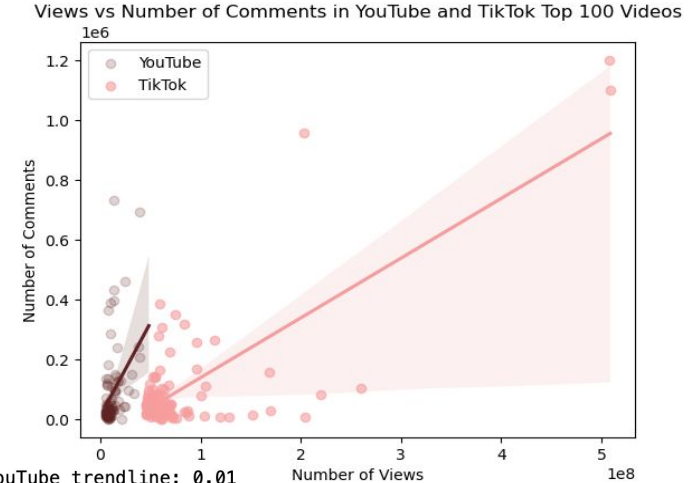


Views and Likes



Slope of YouTube trendline: 0.04
Slope of TikTok trendline: 0.02

Views and Comments



Slope of YouTube trendline: 0.01
Slope of TikTok trendline: 0.002

Takeaways

- Views and likes have a stronger positive correlation in the YouTube dataset because the trendline is steeper
- Views and comments have a stronger positive correlation in the YouTube dataset because the trendline is steeper
- In both graphs, the 95% confidence interval is larger for TikTok meaning the 100 most viewed videos in TikTok have a greater distribution in both likes and comments
- Both graphs match with heatmaps, the scatterplot shows the range of correlation with individual videos and overall trendline when compared with the other dataset

Question #2:

How does the public's interaction correlate with content's number of views?

Methodology

Likes and Comments vs. Views

- Determine which interaction metrics are most associated with views using a pairplot → likes and comments
- Create scatterplots for likes and comments against views on both platforms

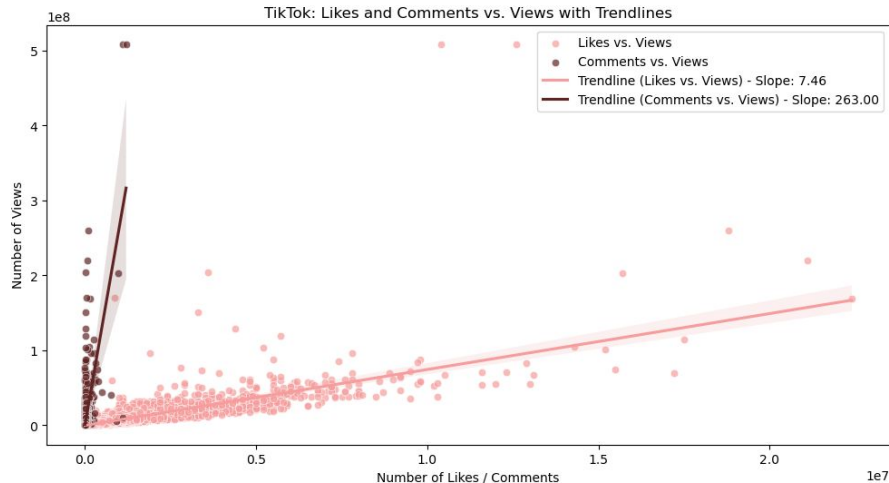
Like-to-Comment Ratios

- Use a logarithmic transformation to display like-to-comment ratios in boxplots for each platform
- Bin the like-to-comment ratio column
- Create barplots to visualize the relationship between the ratios and views

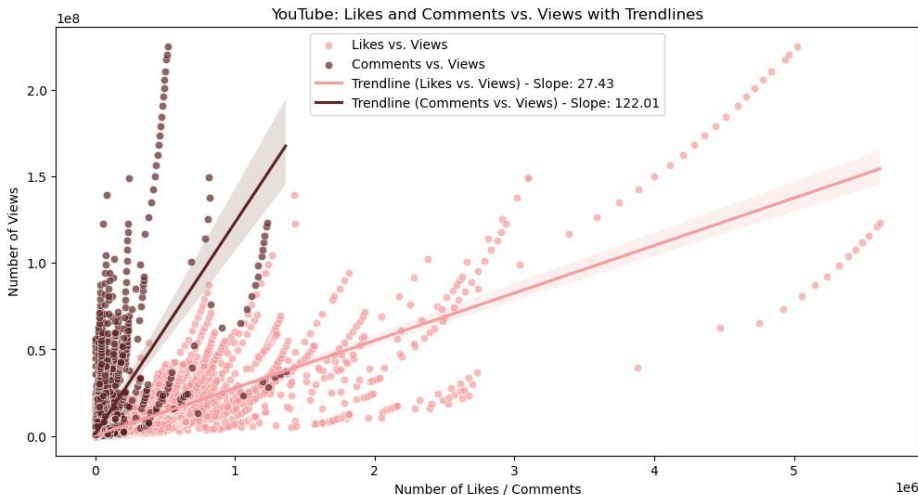
Likes and Comments vs. Views



TikTok



YouTube



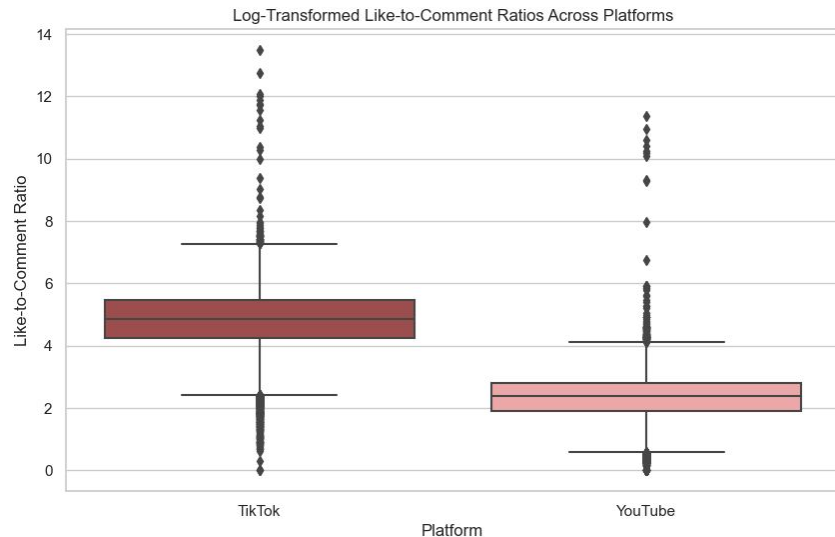
Takeaways

- There is a similar relationship across both platforms for likes/comments and views: a much steeper slope for comments with views compared to likes with views
- On both platforms, videos tend to increase in views more with every comment, while likes have a positive but weaker relationship.
- The difference between each metric's relationship with views is less dramatic on YouTube compared to TikTok.

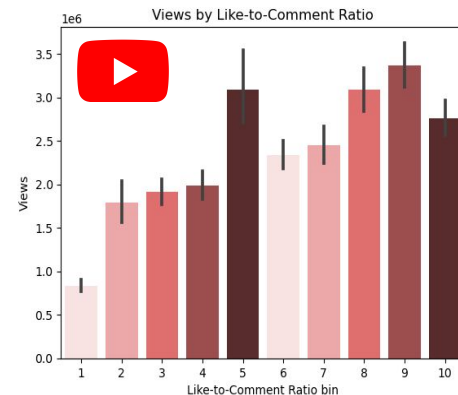
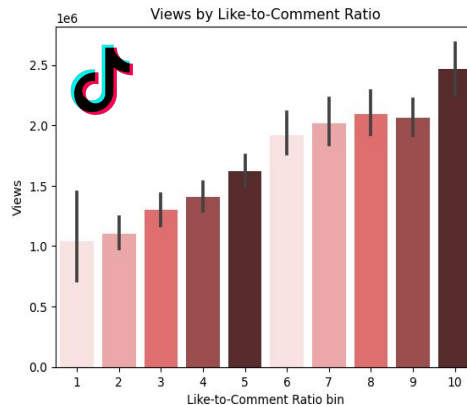
Like-to-Comment Ratios



Log Transformation for Ratios



Binned Ratios Against Views



Takeaways

- The higher median on the log-transformed scale indicates that, on average, TikTok posts tend to have higher like-to-comment ratios compared to YouTube posts.
- On TikTok, there is a more direct relationship between like-to-comment ratios and views compared to YouTube. However, both platforms show a general trend of increasing views on average as like-to-comment ratios increase.

Question #3:

What factors shorten the ‘time of virality’ for content on the platforms?

Methodology

'Time of Virality'

- Defined as length of time for content to 'go viral' from posting date
- Calculated by finding the difference between content's posting and trending dates

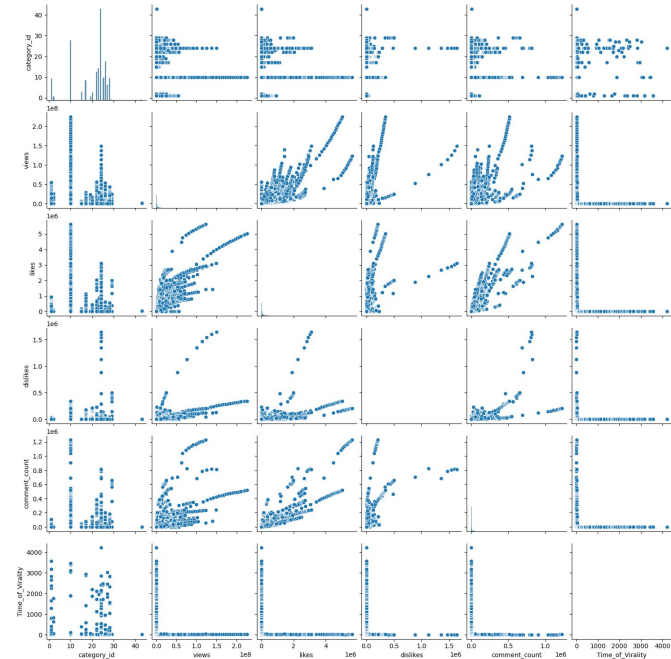
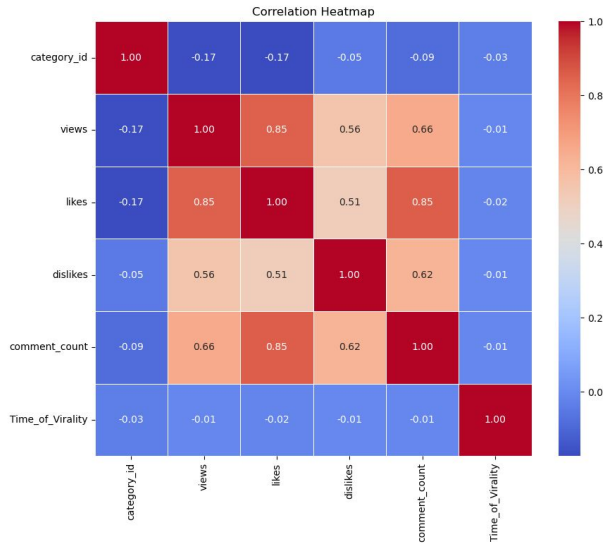
Correlation with 'Time of Virality'

- Parsed data and dropped the few missing data points (in the description column which was non-numerical so had no impact on this question)
- Converted trending date and published date to the same format so they can be mathematically manipulated
- Created new column for 'Time of Virality'
- Sorted data in ascending order by 'Time of Virality'
- Dropped the non-numeric data
- Created correlation heatmap and pairplots to identify relationships between variables

Correlation with 'Time of Virality'



YouTube



Takeaways

- There appears to be no correlation between 'Time of Virality' and the other variables
- No way to game the system and go viral
- It is not as much about time of virality as much as it is about going viral and the degree of virality

Question #4:

How do descriptions *enhance* content's exposure and potential to go viral?

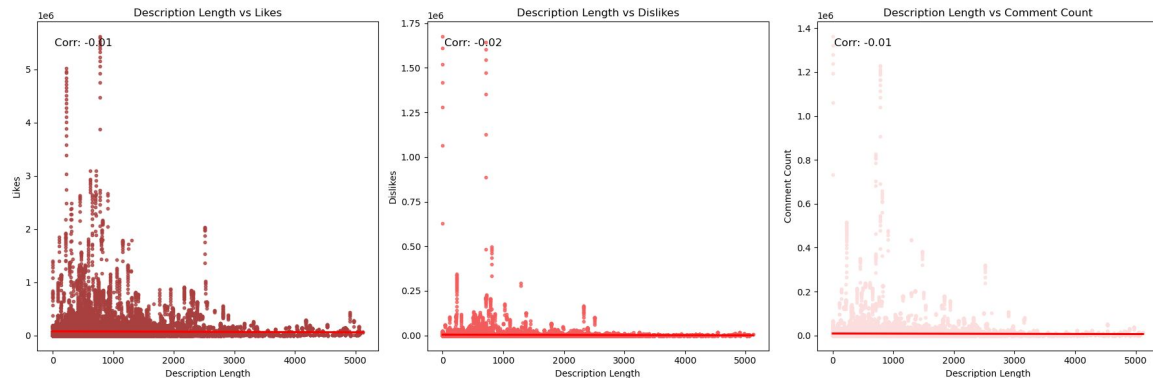
Description Features Vs. Viewer Engagement



Methodology

- Statistical Analysis of Description Lengths
 - Pandas, Seaborn & Numpy for regression analysis
- Text Analysis of Youtube Descriptions
 - Regex-based emoji detection [import re]
 - Defined a function to identify the presence of links
 - Visualized findings with bar plots and histograms
- Text Analysis of TikTok Hashtags
 - Logarithmic transformation to display emoji-in-hashtag-to-likes ratio for TikTok
 - Defined bins for the number of hashtags and created bar plots for these measures

Description Length

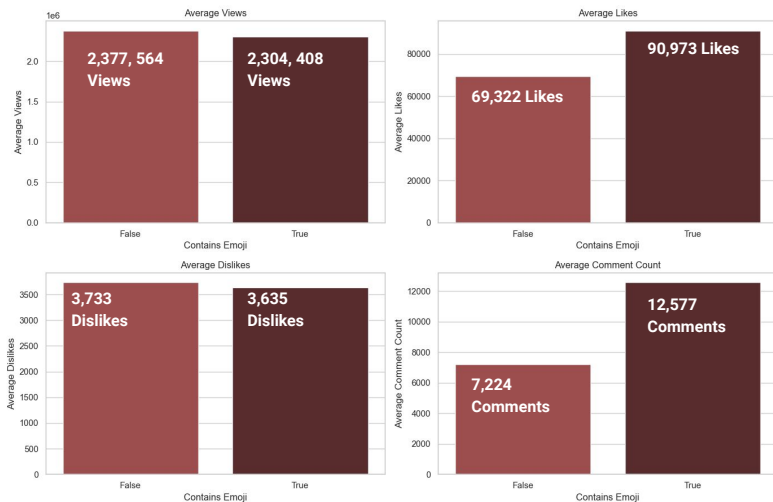


- Weak negative correlation with a slight tendency for the metrics to decrease as the length of a description increases
- Generally a video's description length does not appear as strong predictor of its likes, comments, or dislikes

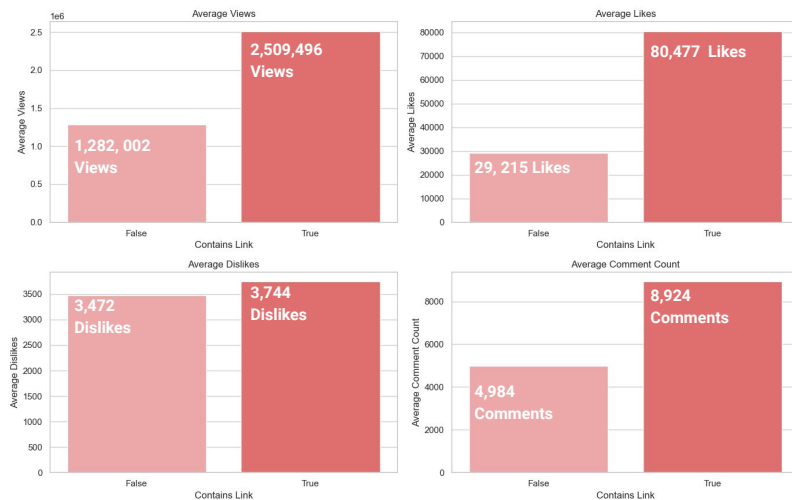
Description Features Vs. Viewer Engagement



Virality & Emojis



Virality & Link Inclusion



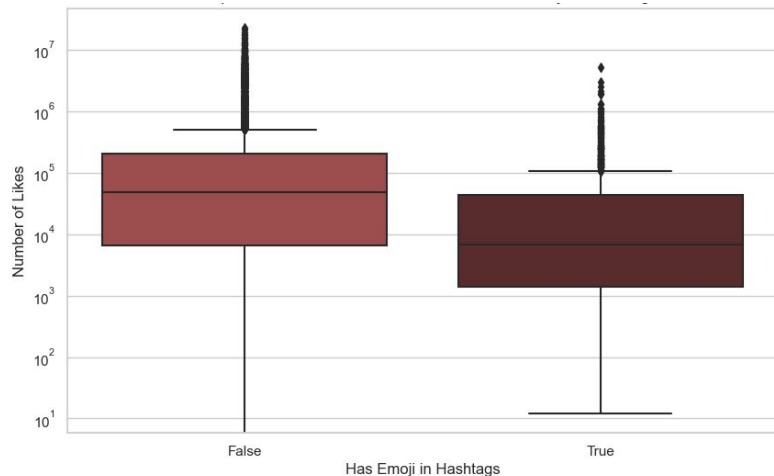
Takeaways

- Video descriptions with emojis do not significantly impact views & dislikes, but lead to notably higher likes and comments
- Videos with links in their description experience an increase in views, likes & comments [dislikes is relatively unchanged]
- Emojis & links both boost engagement, but links are more effective in increasing overall views

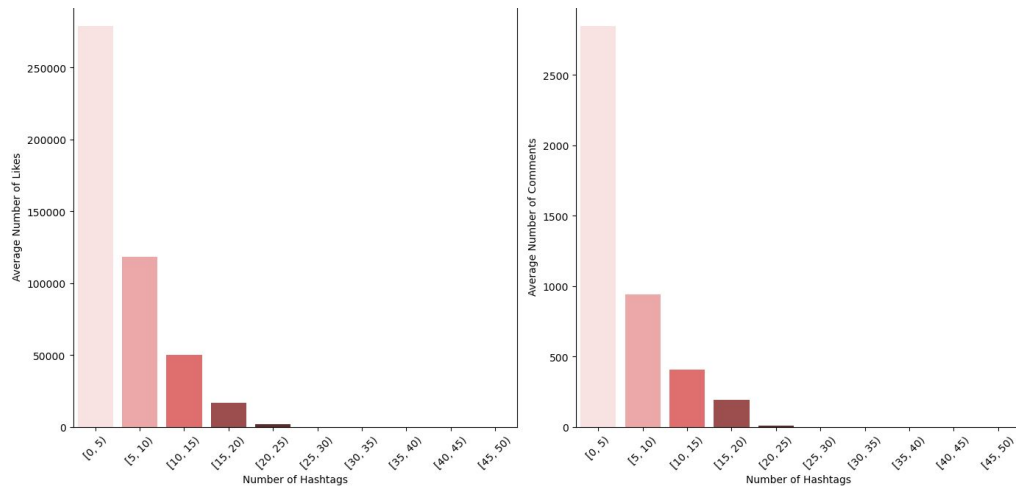
Description Features Vs. Viewer Engagement



Virality & Emojis



Virality & The Number of Hashtags



Takeaways

- In contrast to Youtube, videos with emojis in their hashtags receive fewer likes compared to those without emojis
- Videos that use a relatively small number of hashtags, in the range 0 to 5, are more likely to go viral
- A minimal and targeted approach to hashtag usage may be more effective in harning higher engagement

Final Recommendations!

Our Recommendations

- **Target Comments for Tiktok Virality and Likes for YouTube Virality**
 - *Content should be designed to stimulate comments on Tiktok and likes on YouTube as they have a higher influence on virality*
- **Balance Like-to-Comment Ratio on Tiktok**
 - *Content that has too many comments relative to likes receives less views on average, so content that is designed to stimulate comments should not disregard likes as a factor*
- **Don't focus on going viral fast**
 - *There doesn't appear to be a strong correlation between the highest quality of virality (i.e. highest number of likes, comments, views, etc.) and the shroted 'time of virality,' so focus on creating content that does well in terms of those metrics*
- **Strategize Emoji Use Differently for Each Platform**
 - *Use emojis strategically to increase engagement on Youtube, but be more restrained with emoji use on TikTok, where they may have the opposite effect*

Questions?
