Milestone 4: Final Draft

*Dataset:* [*YouTube Trending Page*](https://www.kaggle.com/rsrishav/youtube-trending-video-dataset?select=US_youtube_trending_data.csv)

D. Asoh, M. Kornely, L. O’Brien, B. Paddock, S. Peterson

**Introduction**

Our group wanted to explore the challenge of creating content and remaining prevalent in the digital media age. Specifically, we wanted to dive deeper into how to establish a platform on the crowded, social media giant of our generation, YouTube.

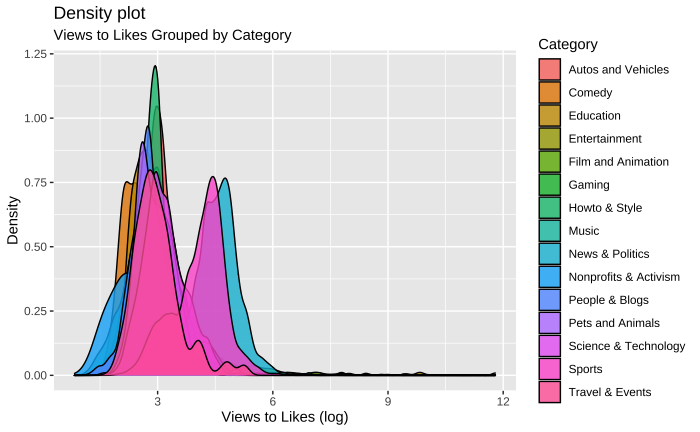
It’s known that many of the most popular creators frequently have content that appears on the trending page. Here, one’s content is showcased to all of the platform’s users, and a channel’s viewership and interactions can reach new heights. So with this analysis, we want to help guide aspiring creators to better understand the current YouTube landscape, such as what categories are popular and the time and day trending content is posted. We went about identifying a few initial areas of uncertainty that content creators come across throughout the course of their careers. One of the largest such areas has to do with trying to understand YouTube's perplexity of YouTube’s fabled algorithm (The Guardian). We are also inspired by Fairuza Alief’s analysis of YouTube data, where they walk through a well-prepared project focused on YouTube statistics being used to ‘gauge consumer opinion for marketing decisions’, to provide a better understanding of a dataset similar to ours. This analysis was very useful in providing a sense of direction for our project.

Our dataset was found on Kaggle with the creator of the dataset using the Youtube API described [here](https://developers.google.com/youtube/v3) to collect data on the Trending Page. Although this dataset is updated daily, we only based our analysis on our initial download (dates are from August 3rd, 2020 to March 11th, 2022) for consistency.

## 

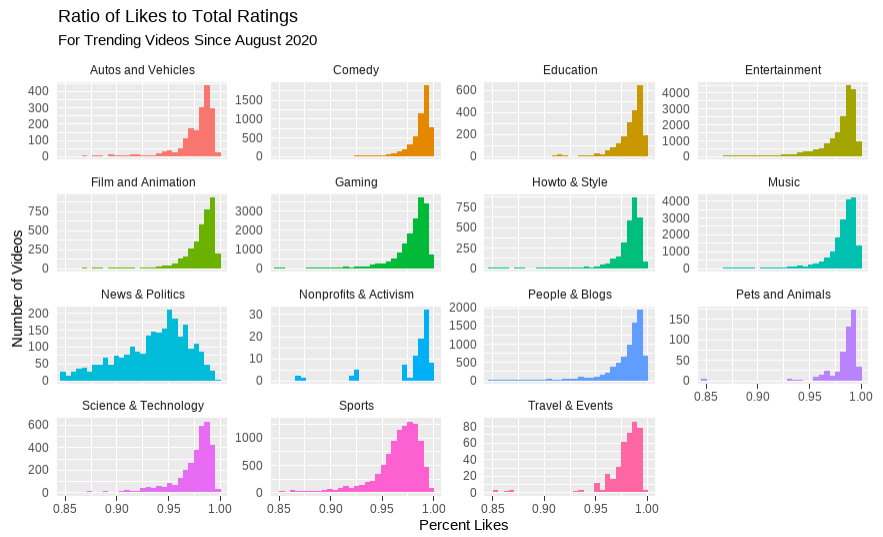
## How does YouTube weigh popularity with the quality of content being shown?

The trending page is generally populated with videos that are considered ‘popular’. The most basic factor that we thought would influence a video’s popularity would be the number of views and positive feedback (likes) a video gets. To answer this question, we created a plot of the density of the views to likes ratio for each trending video colored by category and a plot of the total number of videos vs the ratio of likes to total ratings for each category.



*Figure 1.*

*In this visualization, it is apparent that most categories have trending videos with a view to likes ratio of about 2.5 to 3.5 on a log scale or views to like the ratio of 12.18249 to 33.11545. However, there were a few outlier categories that had a view-to-like ratio over 54.59815. Those categories are “News & Politics” and “Science & Technology”. This suggests that videos in these two categories may have a harder time getting on the trending page.*

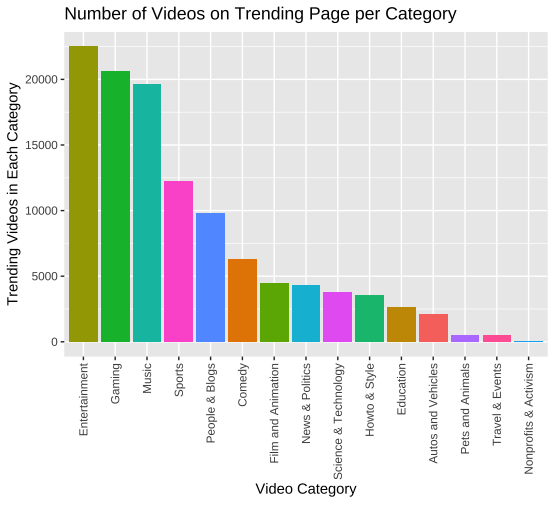


*Figure 2.*

*In this visualization, it is apparent that trending videos in each category have between a 95-100 percent likes to total ratings (likes + dislikes) ratio. However, the “News & Politics” category has far more trending videos that are less than this ratio. This suggests that videos in this category don’t need to be well received by people to become trending since people have various political views which may cause them to dislike videos that disagree with their views.*

## Which category of videos has the most success getting on the Trending Page?

Potential YouTube content creators have a daunting task; What kind of videos should they make to appear on the trending page? Understanding which categories of videos are generally more popular can help creators avoid frustration when a video isn’t doing as well as expected, as well as make sound decisions when deciding the types of content to create. The visualizations below can give valuable insight into this question:



*Figure 3.*

*In this visualization, it is apparent that YouTube videos in the categories Entertainment, Gaming, and Music have out-number the other categories by a fair amount since August 2020. While this is useful information, it would help creators to understand why these videos are appearing on the trending page so much.*

## Another way we can examine why certain categories are more popular than others is by looking at the total views that each category receives.

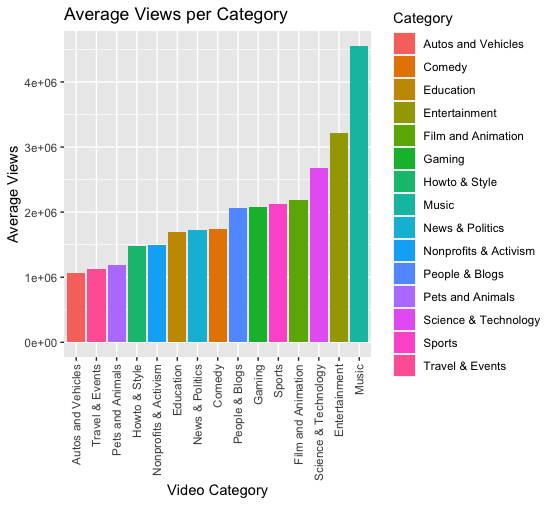
## 

*Figure 4.*

*In this visualization, it is obvious that Music and Entertainment YouTube videos are the only categories that reach more than 200 million views. One important thing to note is that music videos are often watched multiple times by people who enjoy the song, which likely contributes to their increased views. Most videos of other categories are only watched a single time by a viewer*.

## 

To supplement the visualization above, we wanted to include a visualization that showed the average views per video in a category on the trending page. While we initially wanted to do a violin plot or box plot for Figure 4, we ran into issues due to the immense amount of outliers in the data (i.e videos that go very viral), causing the boxplots to be impossible to view.



*Figure 5.*

*In this visualization, it is easy to see Music Youtube Videos have the highest average views per video, logging in at 4 million views, with Entertainment at 3 million views.*

To help creators further understand why certain categories are more likely to appear on the Trending Page, we will look into the time of day videos of each category are posted next.

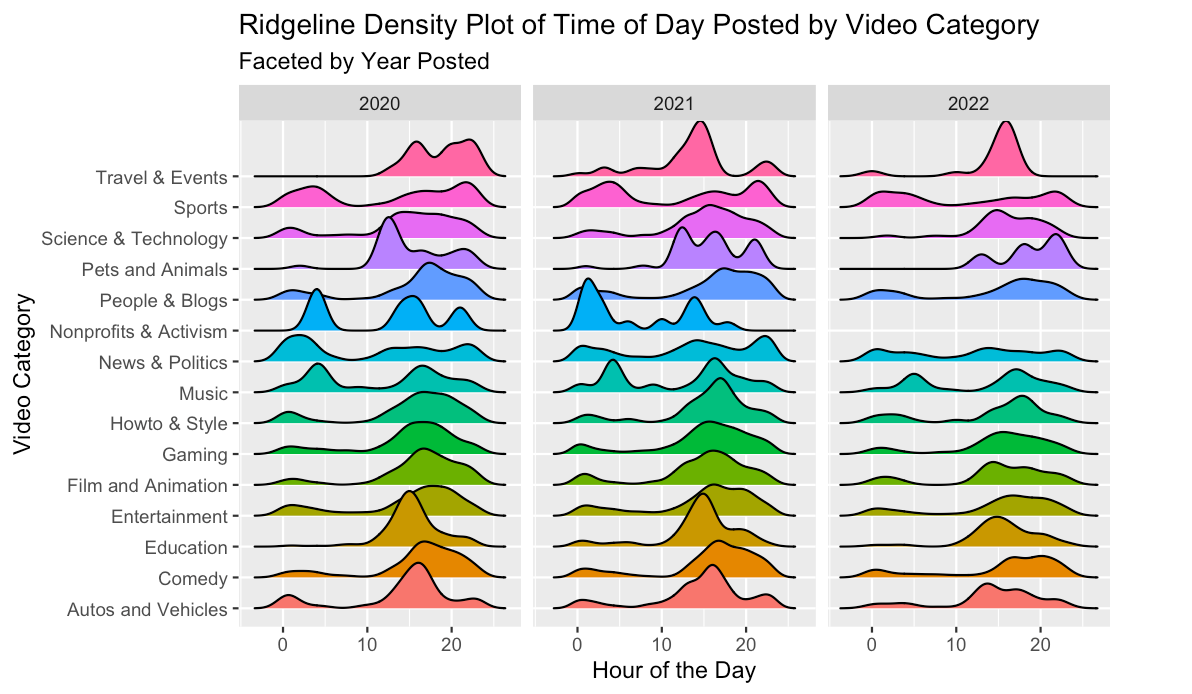
**Does the time of day post affect getting on the trending page?**

## 

*Figure 6.*

*According to the figure above, Music YouTube videos are posted around noon, while Entertainment and Gaming videos are posted around 2:30 p.m.- 3:00 p.m. These times make sense for why they would be popular because it is the middle of the day when people are on lunch breaks, or almost done with school/work.*

Using a ridgeline plot to examine the densities for each hour of the day (i.e Hour 10 being 10:00 AM and Hour 20 being 8:00 PM), we wanted to explore the density of the posting times for each category to help creators understand the popular posting times for each category[[1]](#footnote-1).



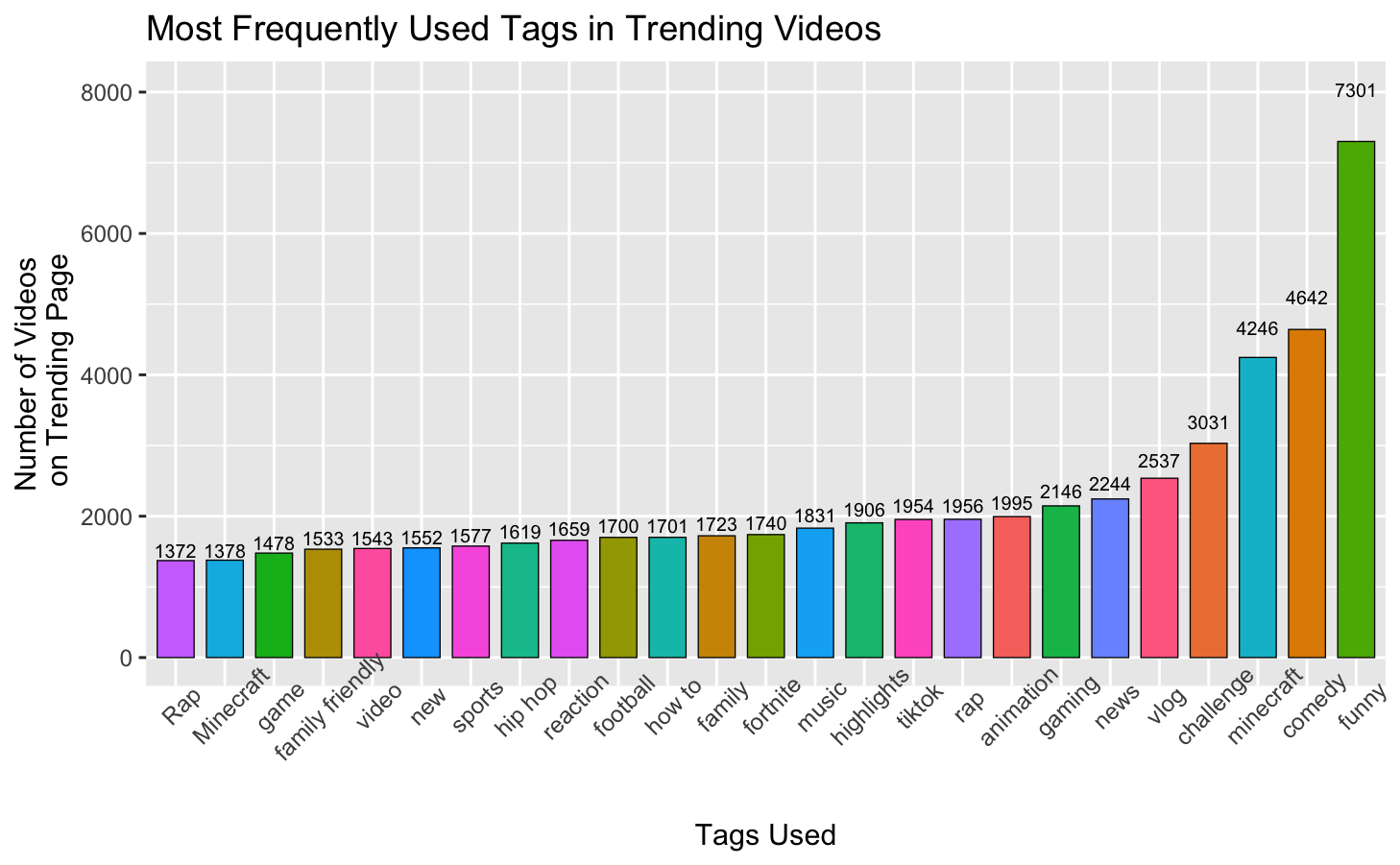
*Figure 7.*

*In this visualization, it is clear that there are peaks present across several different video categories. Across the years, it seems that these peaks tend to stay relatively similar with some amount of deviation. One important takeaway from this visualization is the uniformity of the News and Politics category across the different hours of the day. We inferred this could be due to the unpredictability of breaking news occurring throughout the day. Many of the other peaks seem to align to the end of the workday, around 4:00 to 5:00 PM.*

## 

## What are the most frequent tags included in trending videos?

Alongside likes, dislikes, and categories, another piece of metadata that creators should understand how to appropriately add to videos and help them trend are “tags”. Tags, short for hashtags, are brief descriptive keywords that creators can add to their video descriptions and can be thought of as mini video categories or subjects. Considering all of this, we thought it would be important to visualize what tags appear most frequently on the trending page to prompt creators to use certain tags with their videos and help guide them to come up with other related tags.



*Figure 8.*

*This plot illustrates the most frequent tag for trending videos is the keyword “funny”, which has been used 7300+ times in our dataset of 100,000+ trending videos. Tags like “comedy”, “Minecraft”, and “vlog” were also among some of the more popular descriptors.*

**Conclusion**

While we aren’t able to fully understand each factor that contributes to the popularity of videos posted on YouTube and what lands videos on the Trending Page, we uncovered several valuable insights that could help creators with a lesser following understand how many creators structure their videos and guide them to making sound decisions with their content. We determined that unless a creator is looking to post News videos, they must achieve a likes ratio of around 95% or more. If a creator is looking to get a first breakout video, they might consider changing up their normal content and try their hand at a video in Music, Entertainment, or Gaming, as these categories appear most often on the Trending Page, and generally achieve higher views on average than other types of content. While the time of the day a video is posted may not play the biggest role in determining the popularity of a video, creators might use our insights from above to help them determine when the appropriate time of day to post a video of a certain category might be. For example, we have seen through three years that Gaming videos posted around 7 or 8 PM trend more often than other times of the day. However, a creator may choose to post at a low point in the trend, as it could be that there are simply more videos posted during a peak(therefore more competition). And finally, we uncovered some insights in regards to the tags that are used most frequently on trending videos and offered some suggestions to novice creators who may struggle with coming up with their tags for videos.

**References**

*Lewis, Paul, and Erin McCormick. “How an Ex-Youtube Insider Investigated Its Secret Algorithm.” The Guardian, Guardian News and Media, 2 Feb. 2018, https://www.theguardian.com/technology/2018/feb/02/youtube-algorithm-election-clinton-trump-guillaume-chaslot.*

1. The data we collected did not include any observations for the Nonprofits & Activism category for 2022. [↑](#footnote-ref-1)