

# YouTube Video Data Analysis



AI Boot Camp: Project 1 / Group 4

## Team Members:

1. Mahshid Alinejad
2. Ben Cross
3. Romario Esparza
4. Stephen Martinez

# Project Overview



## **Analysis of Trending YouTube Videos to identify strengths of high traffic content:**

Are you dreaming of launching a successful YouTube channel but feeling overwhelmed by the vastness of the platform? Have you ever wondered what makes a YouTube video go viral?

Cracking the code for creating consistently popular content can feel like a mystery. But there's a secret weapon at your disposal: YouTube trending video analytics. By analyzing these trends, you can gain valuable insights into what viewers are currently engaging with. You can then leverage this knowledge to tailor your content strategy, identify topics with high potential, and ultimately create videos that resonate with your audience and propel your channel towards success.

# Project Overview

## Goals And Questions To Be Addressed:

### Tags

Their significance in being found and presented to the right demographic.

### Publication Timeline

The power of when you publish can impact your views and engagement

### Video Description

The importance of your video description and why it's important.

### Video Title/Video Content Type In Relation To “Going Viral” Within 24hrs

Is there a correlation between the video title/video content type and that video going viral within the first 24hrs after being uploaded?

If so, show what video titles/video content types would be recommended if a user wants a high chance for their video to go viral within 24hrs.

# Project Overview

## Data Collection:

Our dataset is pulled from a source going back to August 2020 for US uploaded videos which is updated once weekly. Within the dataset are 16 unique column labels providing a variety of details to each video such as: Video ID, Title, Published At, Channel ID, Channel Title, Category ID, Trending Date, Tags, View Count, Likes, Dislikes, Comment Count, Thumbnail Link, Comments Disabled, Ratings Disabled, and Description.

## Initial Exploration Process:

Our initial Exploratory Data Analysis (EDA) identified one major challenge. Because the dataset is updated weekly, entries are **added** to the dataset instead of **updated/replaced**. This is not an issue if you wanted to show trends for variables such as “user engagement” for a specific video but could be problematic if you are wanting fixed values for all videos. ie: When asking for the “5 Most Viewed Videos” from the original dataset. The same 2 videos appear multiple times instead of 5 unique videos. The dataset at this time had **267,987 rows**.

## Data Cleanup:

Since we want to present correlations between each unique video, we input code to only allow outputs for any inquiry for the same video once. This now allowed accurate results when a range of data was desired such as “5 Most Viewed Videos” showing the top 5 videos without duplicates titles. This cleanup of data also significantly reduced the overall size of the dataset and simplified the rest of the EDA process. The dataset now has only **48,361 rows**.

# Project Overview

## Approach Taken To Achieve The Goals:

### Tags

Reviewed Tags based on which tags had the most total views

### Publication Timeline

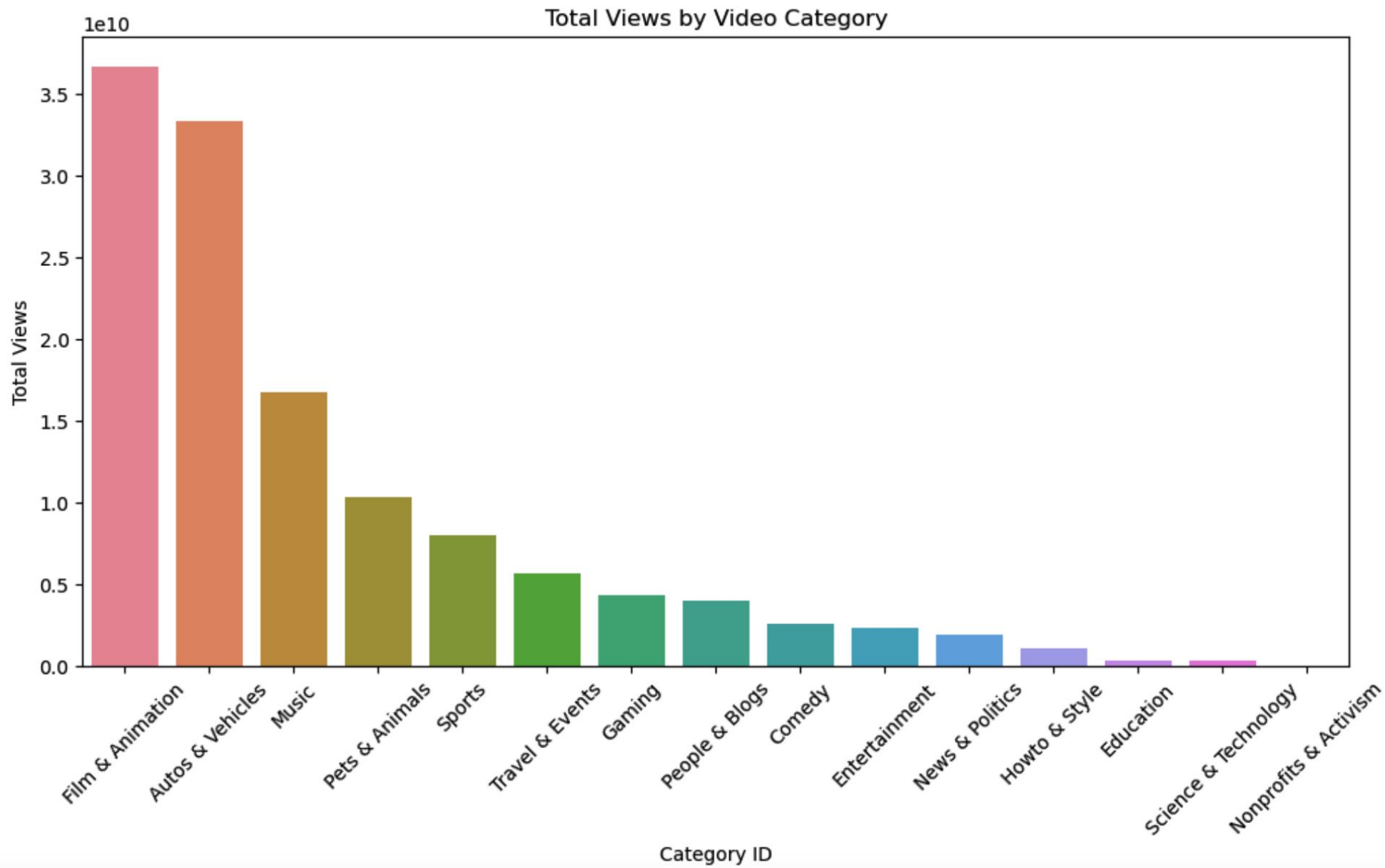
Reviewed the data to find the day and times when videos were published that had the most views

### Video Description

Short, Medium, or Long bins created based on number of characters in video description and analyzing the code to see which had had higher likelihood of capturing more views

### Video Title/Video Content Type In Relation To “Going Viral” Within 24hrs

Used Categorical data to decide what type of User Engagement(s) to use in example. After we reduced the dataset to show only the view count for each video within 24hrs after its upload. Next we listed the videos by the highest value view count and assigned them a “Video Content Type”. This then allowed us to show the top 10 videos that went viral the fastest and show the correlation between Title/Video Content Type.



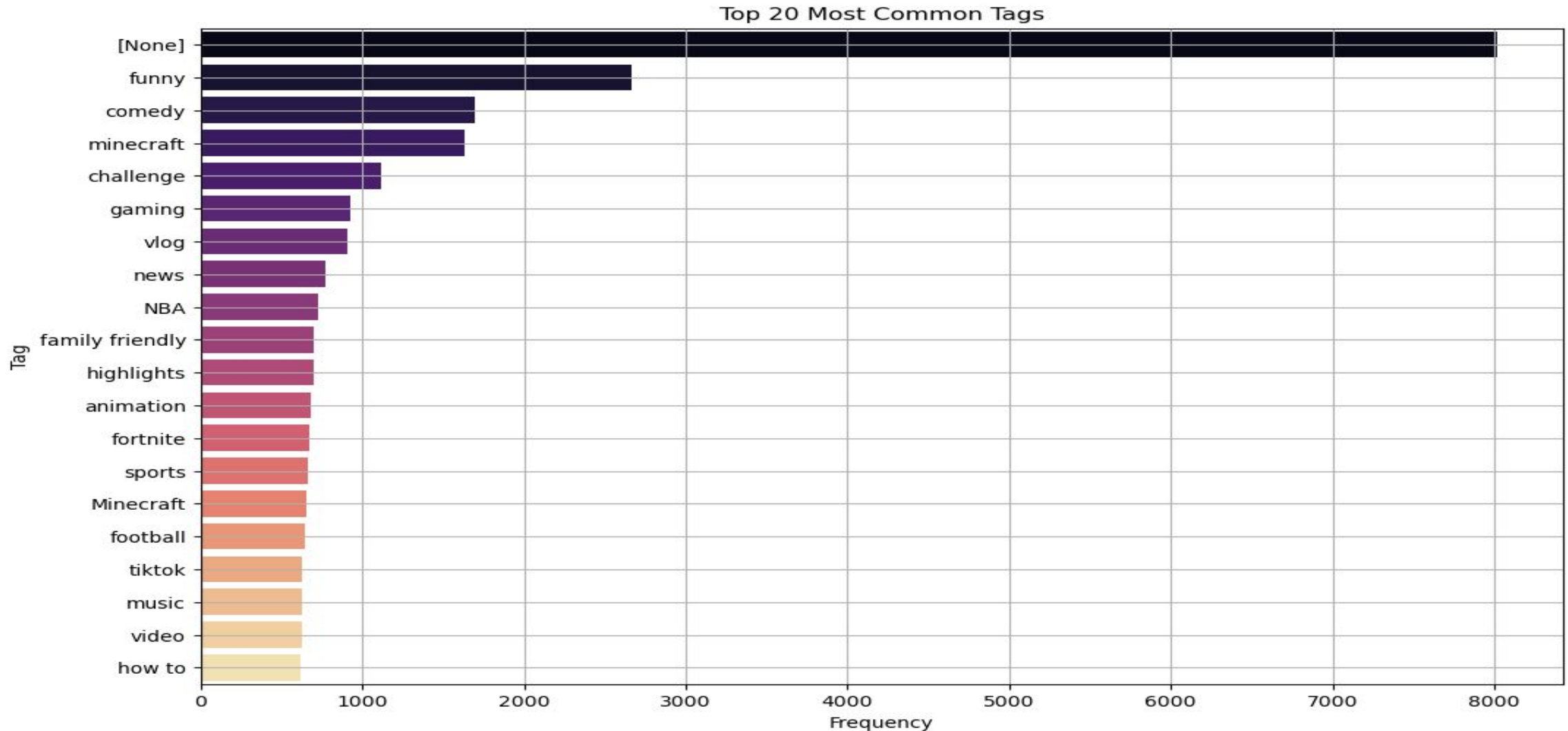
# How YouTube Tags Impact The Discoverability Of A Video

YouTube tags can certainly impact the discoverability of a video, but they're just one factor among many. Tags help YouTube's algorithm understand what your video is about and can influence where it appears in search and recommendations. However, having tags alone doesn't guarantee more views.

Our data represents that top industries in Youtube channel are gaming, sports and News. Companies with products used in these industries can use Youtube as their digital marketing platform and thrive the best.

The more engagement involves things such as leaving the comment section for example always open. Channels with higher frequency of posting get more views as well.

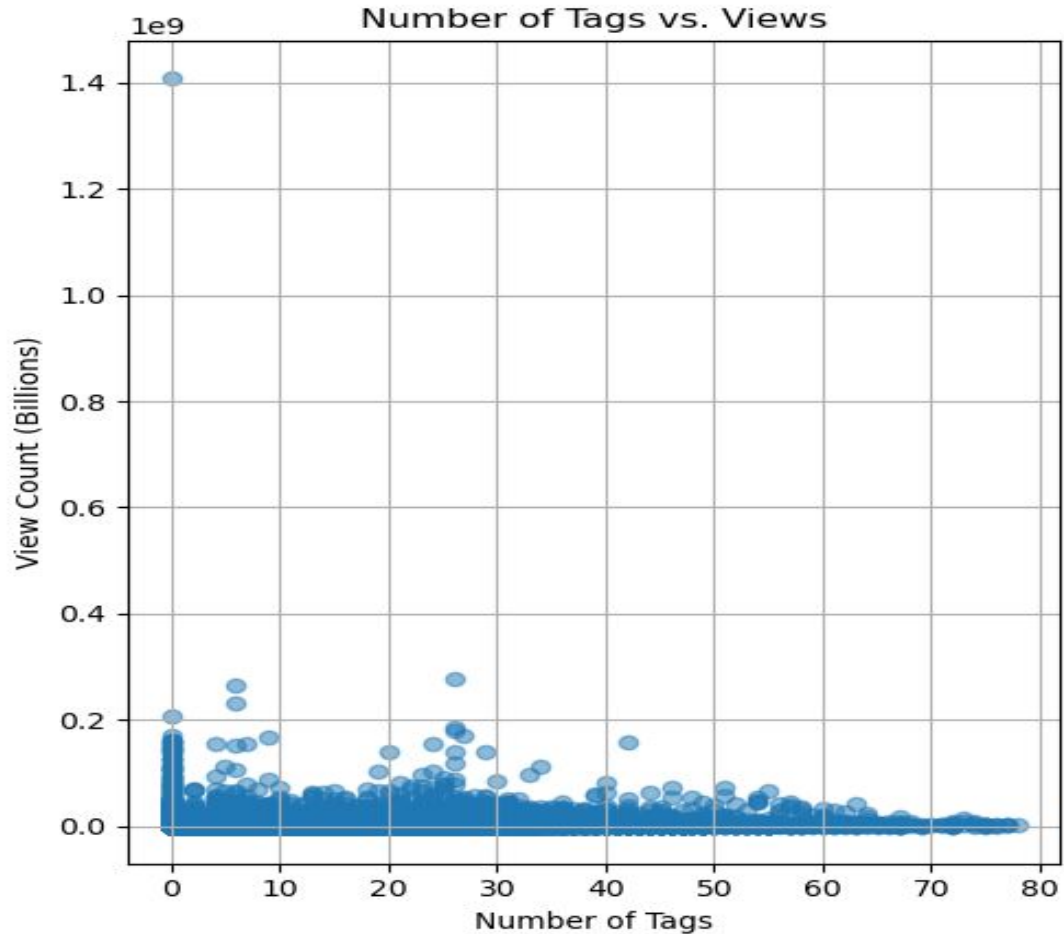
# How YouTube Tags Impact The Discoverability Of A Video



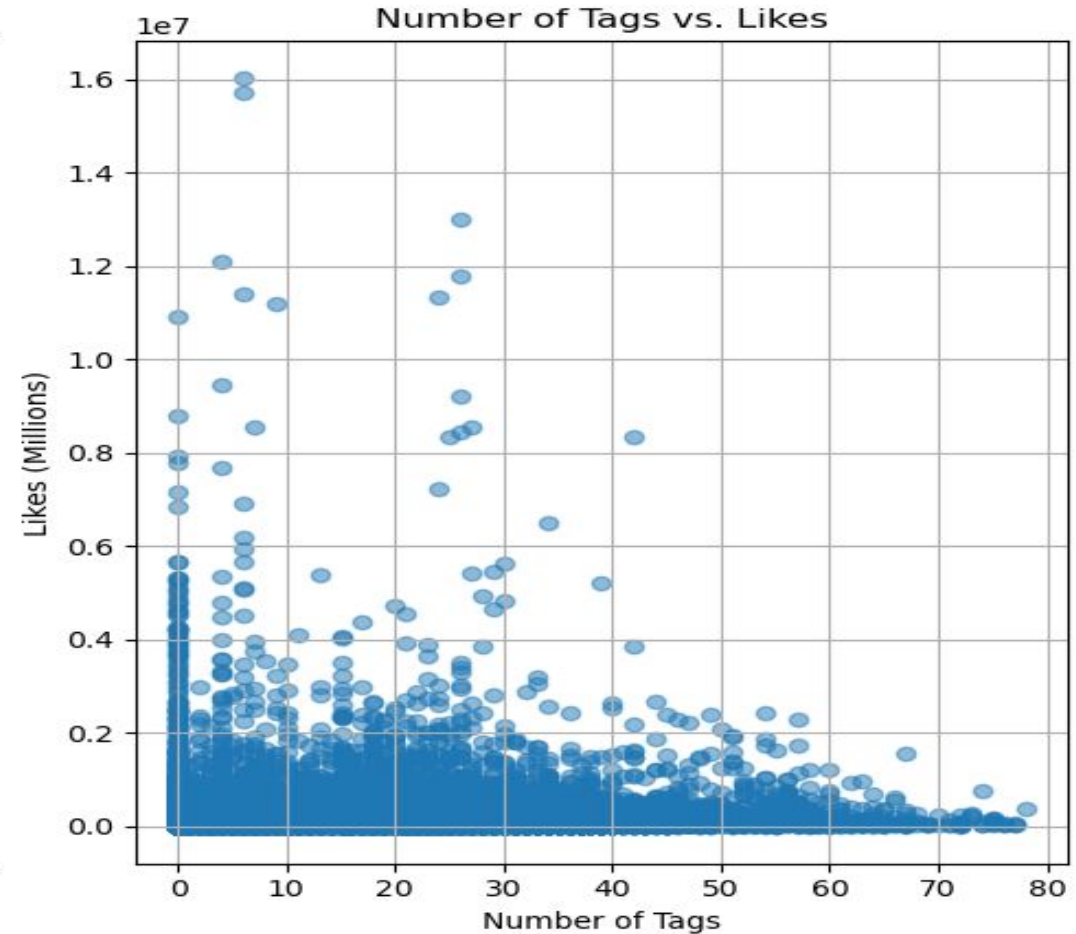


# How YouTube Tags Impact The Discoverability Of A Video

## Data set relation



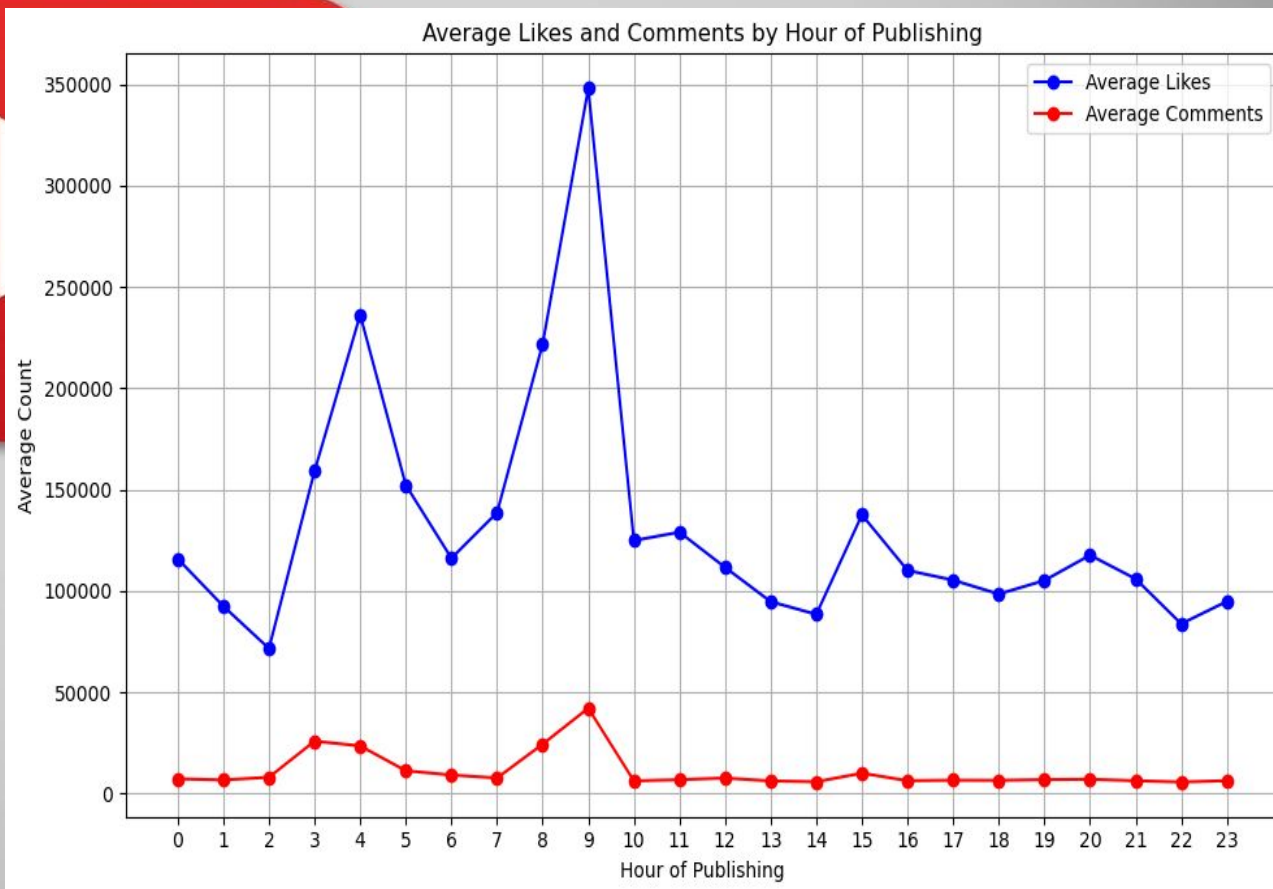
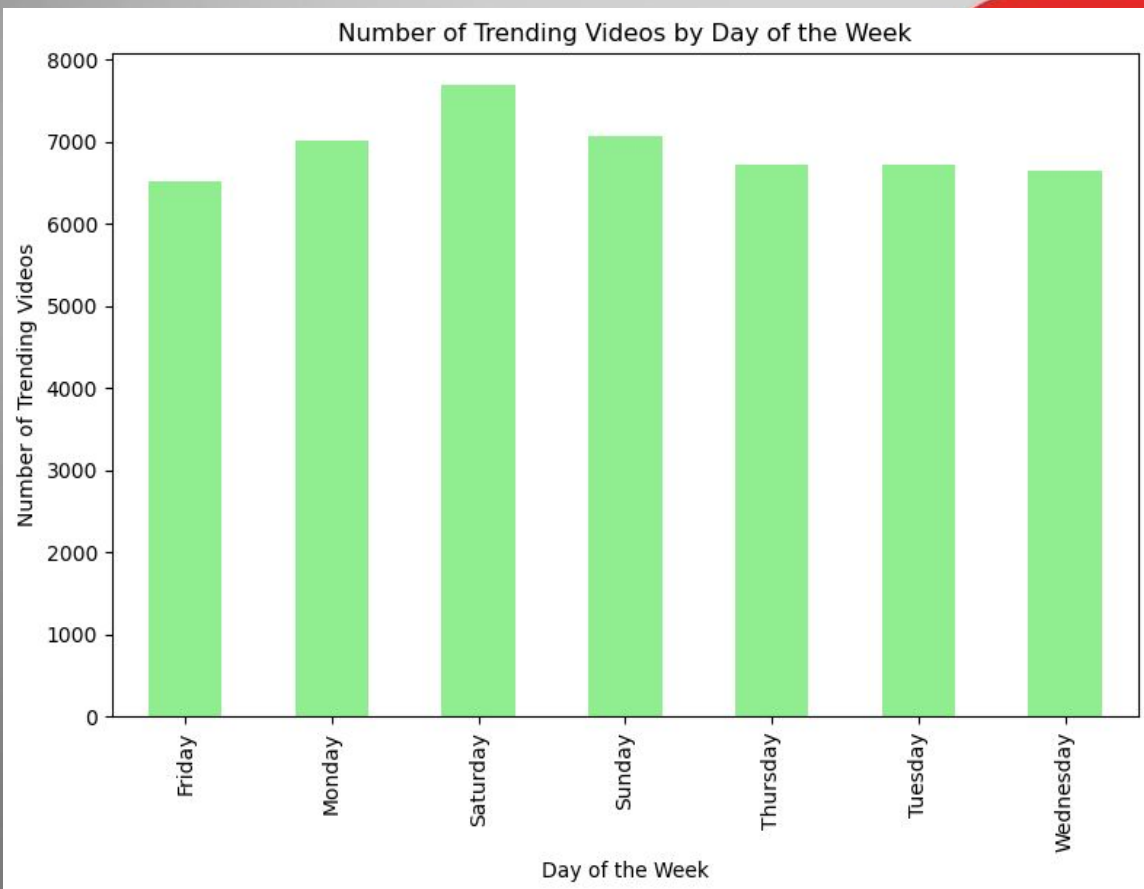
Correlation between number of tags and views:  $-0.0029450952135724453$   
Correlation between number of tags and likes:  $-0.01888141528680547$



Average view count for videos with tags: 2483568.8453644025  
Average view count for videos with no tags: 3428577.530856502  
Videos with no tags tend to get more views.

# Day Of The Week And Median Spans To Trend

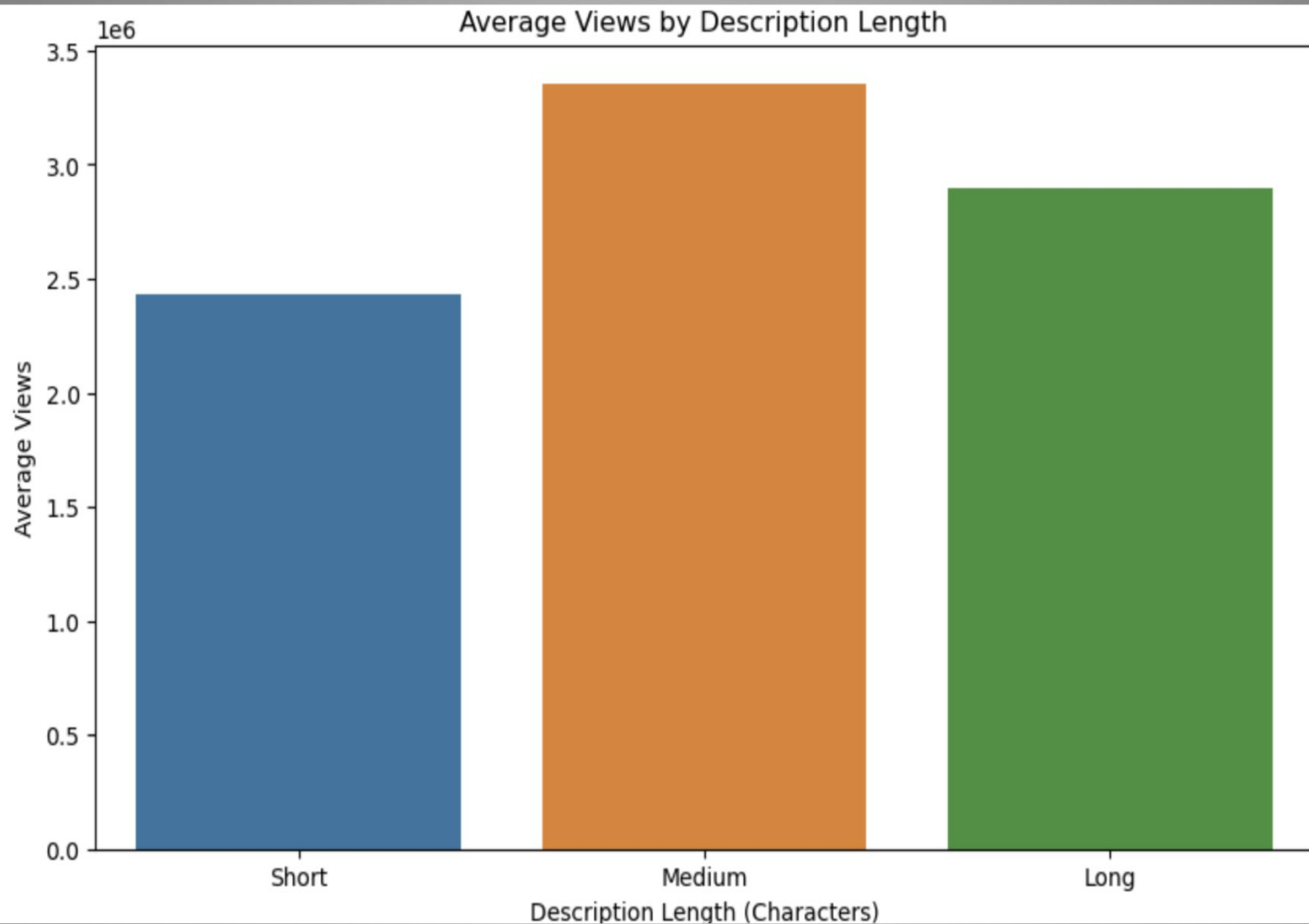
Our data shows that Saturday are the best days to go viral and the Youtube video postings are usually go viral after 9 hrs of publishing. According to our analysis it also takes an average 5 days for videos to trend.



# Result/Conclusion Video Description

**Ben Cross**

- Created 3 Bins
  - Short, Medium, Long
    - Short 0-1499 Characters
    - Medium 1500-3399 Characters
    - Long 3400-5000 Characters
- Goal to see what Bin had the highest mean of most Views



# Video Title / Video Content Type In Relation To “Going Viral” Within 24hrs

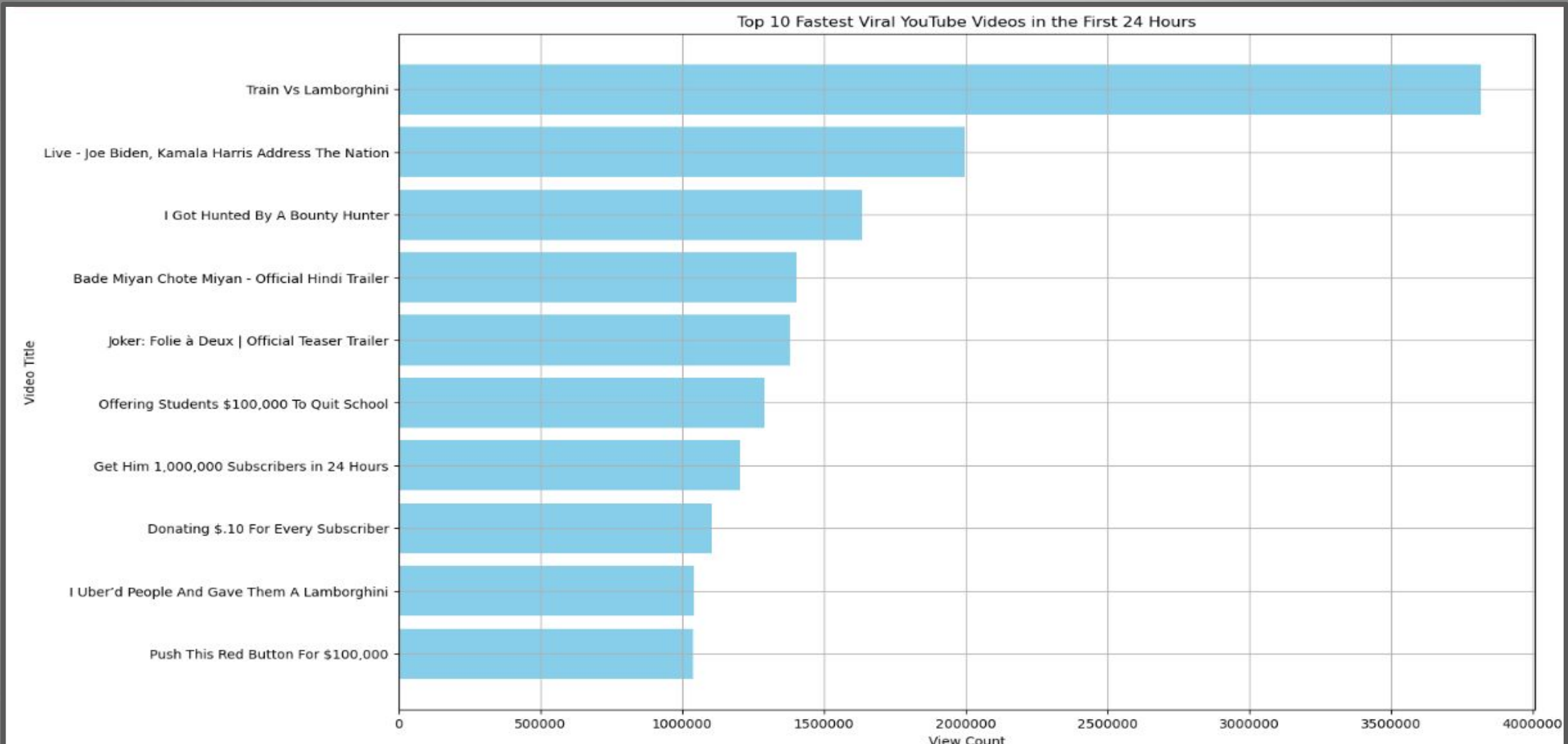
The term “Going Viral” is used when a video receives a large amount of user engagement in a short period of time. This can be from the amount of: views, likes, dislikes, user comments, or any combination of the four.

Our data shows that the top 10 fastest viral videos within 24hrs of being uploaded all have a correlation with their Title/Video Content Type.

As a reminder, the dataset we are working with is a snip of US Youtube data from August 2020 to April 2024.

The “Video Content Type” is generalized when explaining the relationship in the visuals provides in the next slides.

# Video Title / Video Content Type In Relation To “Going Viral” Within 24hrs





# Video Title / Video Content Type In Relation To “Going Viral” Within 24hrs

title	view_count
Train Vs Lamborghini	38173956
Live: Joe Biden, Kamala Harris Address The Nat...	19983734
I Got Hunted By A Bounty Hunter	16372557
Bade Miyan Chote Miyan - OFFICIAL HINDI TRAILE...	14061617
Joker: Folie à Deux   Official Teaser Trailer	13797158
Offering Students \$100,000 To Quit School	12899180
Get Him 1,000,000 Subscribers in 24 Hours	12035592
Donating \$.10 For Every Subscriber	11038284
I Uber'd People And Gave Them A Lamborghini	10436587
Push This Red Button For \$100,000	10379834

← This shows the top 10 fastest viral videos within 24hrs after upload listed by highest View Count and their Title.

We then grouped the videos into “Video Content Type” based off their Title.

Rank By Views	Video Title	Video Content Type
1	Train Vs Lamborghini	High Excitement/Risk
2	Joe Biden, Kamala Harris Address The Nation	Political
3	I Got Hunted By A Bounty Hunter	High Excitement/Risk
4	Bade Miyan Chote Miyan	Movie Trailer
5	Joker: Folie à Deux   Official Teaser Trailer	Movie Trailer
6	Offering Students \$100,000 To Quit School	High Excitement/Risk
7	Get Him 1,000,000 Subscribers in 24 Hours	Unique Fundraiser
8	Donating \$.10 For Every Subscriber	Unique Fundraiser
9	I Uber'd People And Gave Them A Lamborghini	High Excitement/Risk
10	Push This Red Button For \$100,000	High Excitement/Risk

# Video Title / Video Content Type In Relation To “Going Viral” Within 24hrs

Videos that tend to go viral the fastest based on total number of views within 24hrs after being uploaded fall within 4 “Video Content Types”:

Rank By Views	Video Title	Video Content Type
1	Train Vs Lamborghini	High Excitement/Risk
3	I Got Hunted By A Bounty Hunter	High Excitement/Risk
6	Offering Students \$100,000 To Quit School	High Excitement/Risk
9	I Uber'd People And Gave Them A Lamborghini	High Excitement/Risk
10	Push This Red Button For \$100,000	High Excitement/Risk

Rank By Views	Video Title	Video Content Type
4	Bade Miyan Chote Miyan	Movie Trailer
5	Joker: Folie à Deux   Official Teaser Trailer	Movie Trailer

Rank By Views	Video Title	Video Content Type
7	Get Him 1,000,000 Subscribers in 24 Hours	Unique Fundraiser
8	Donating \$.10 For Every Subscriber	Unique Fundraiser

Rank By Views	Video Title	Video Content Type
2	Joe Biden, Kamala Harris Address The Nation	Political

Note: The Video Content Type of “Political” could be considered an outlier and removed as it is from the 2020 Presidential Election and not recurring throughout like the other Video Content Types.



# Video Title / Video Content Type In Relation To “Going Viral” Within 24hrs

## In Conclusion:

The data shows of **US** Youtube videos from **August 2020 to April 2024** that **50% of the Top 10, including the #1 fastest viral video**, are all **High Excitement/Risk** related.

If you wanted to create a video that would have a high chance of going viral within 24hrs after being uploaded, the video Title/Video Content Type should be **High Excitement/Risk**.

# Summary

## **Room for Growth**

Using simple techniques you can improve the traffic that your YouTube Channel or Video will receive.

- Using Tags, Publication Timelines, Description Length, and Viral Trends/Titles you can pull views to your content.
- There are multiple ways to increase your views but these are the first 4 areas we suggest you focus.

## Problems Encountered

### **Some challenge we encountered during this analysis:**

- Time zones of when videos were published and/or trending are not defined within the data set.
- Super Viral “Gangnam Style” Videos skewing the data
- Review of the Data led to intrigue in additional avenues within the data.
  - The Scope had an opportunity to get extremely massive and complex
- A need to dive deeper to find the methods to get videos gradual wins (i.e. 1000-5000 views). A strong area to be as a Local Business.
- The dataset source we used is a “improved version” of another dataset source that is a “improved version” of another dataset which is pulled from a master original source. The master original source is extremely massive in file size (multiple Terabytes) and requires special access we don’t have.

# Future Considerations



## Some additional efforts and information we would like to do in the future with this dataset analysis:

- What hour of the day is best to publish based on your geolocation and your viewers?  
Better Defined
- What's the impact of a Thumbnail on a YouTube video?
- How can we dissect the data to get hyper focused demographics that will engage with your product, service or "community" feature?
- What is the best way to time the Market? (Aligning with something happening in real time or coming future)
  - Example: Paris Olympics are around the Corner, could we create a series of Comedic Company Olympics that brought attention to a Brand on the wave of the Olympics?
- Find the Correlation between Video Length and Engagement Metrics