

YouTube Videos Trending Analysis

Presented By: Abdurrahman **Syafiq** Izharuddin



Table of Contents

01. Problem Objective

02. Videos Analysis

03. Segmentation

04. Summary

Problem Objective

Problem

A YouTuber wants one of his videos to be trending. For that, the client wants to know what kind of video information can be trending based on data from videos that have been trending before.

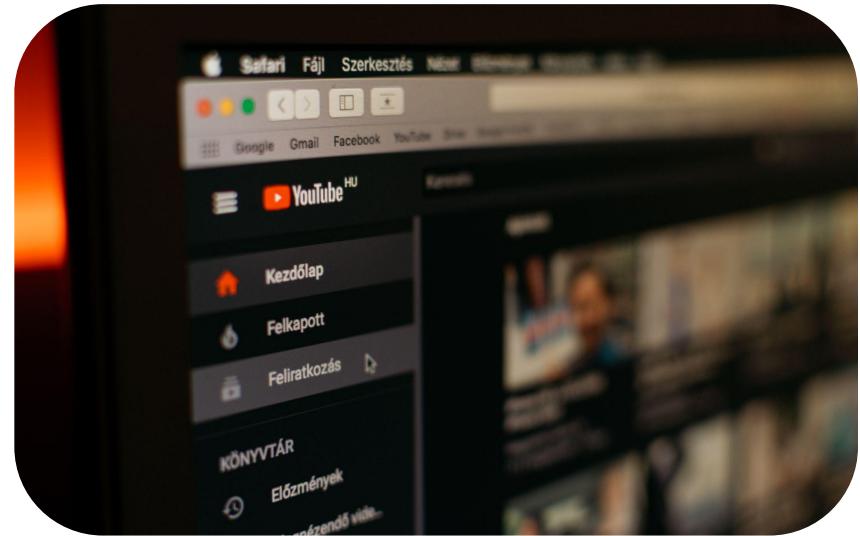


Photo by [Szabo Viktor](#) on [Unsplash](#)

Objective

Identify if there are **common factors of trending videos**, such as:

- Specific *topic/category* that the video falls onto
- Who uploaded these videos
- What are written on the *title* or *tags*
- The engagement metrics
- Video upload time

Knowing the segmentation of videos based on engagement metrics

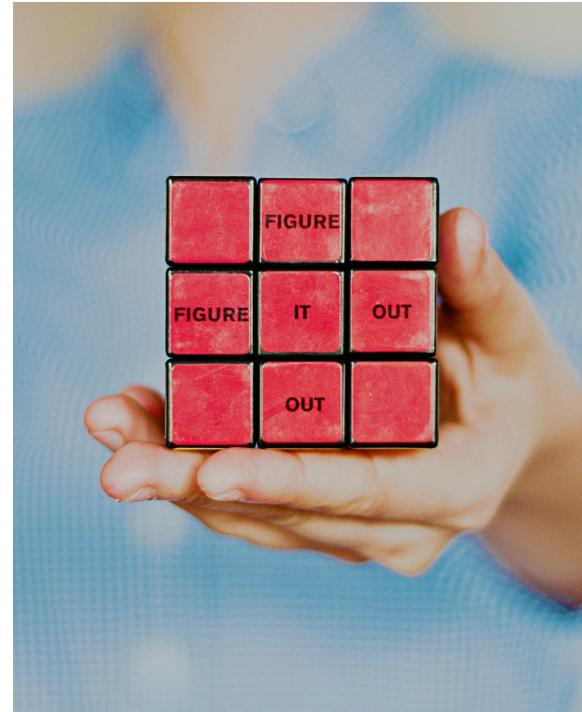
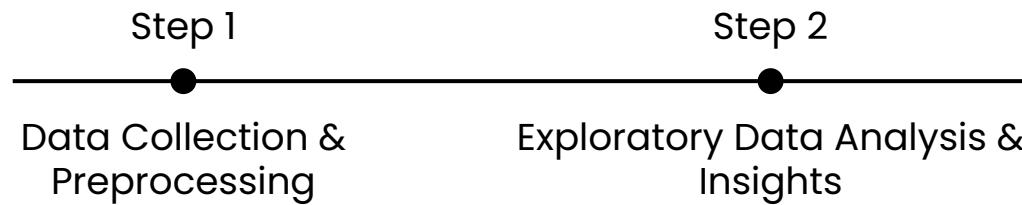


Photo by [Karla Hernandez](#) on [Unsplash](#)

Videos Analysis

Analysis Step



```
youtube_videos.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40949 entries, 0 to 40948
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   video_id         40949 non-null   object  
 1   trending_date    40949 non-null   object  
 2   title             40949 non-null   object  
 3   channel_title    40949 non-null   object  
 4   category_id      40949 non-null   int64  
 5   publish_time     40949 non-null   object  
 6   tags              40949 non-null   object  
 7   views             40949 non-null   int64  
 8   likes             40949 non-null   int64  
 9   dislikes          40949 non-null   int64  
 10  comment_count    40949 non-null   int64  
 11  thumbnail_link   40949 non-null   object  
 12  comments_disabled 40949 non-null   bool   
 13  ratings_disabled 40949 non-null   bool   
 14  video_error_or_removed 40949 non-null   bool  
 15  description       40379 non-null   object  
dtypes: bool(3), int64(5), object(8)
```

```
youtube_categories.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32 entries, 0 to 31
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   id               32 non-null    object  
 1   title            32 non-null    object  
 2   channel_id       32 non-null    object  
dtypes: object(3)
```

- From `youtube_videos` dataframe there are has an incorrect data type, such as `trending_date`, `category_id`, and `publish_time`.
 - There is a `null/missing value` in the `description` column.
-

From 40,949 Rows Data

1.39%

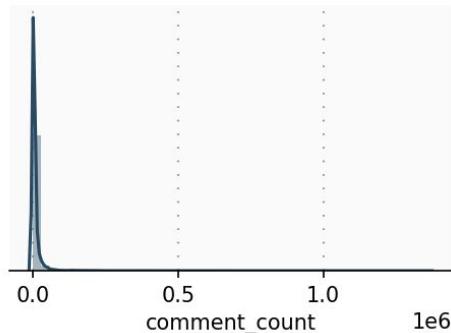
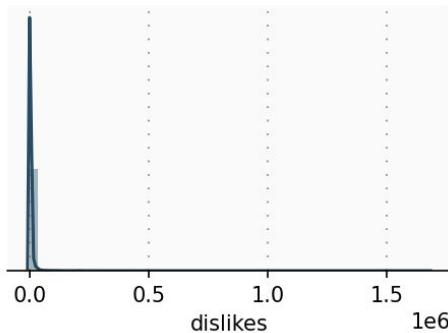
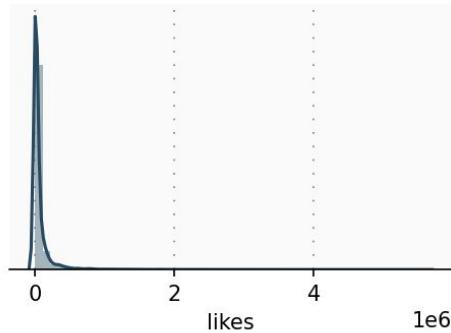
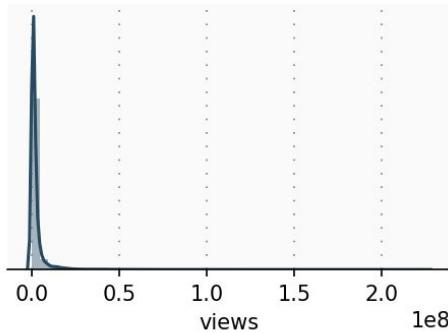
Missing Values

48 Rows

Duplicated

Distribution of Numeric Variables

The distribution of numerical variables has a positively skewed distribution.

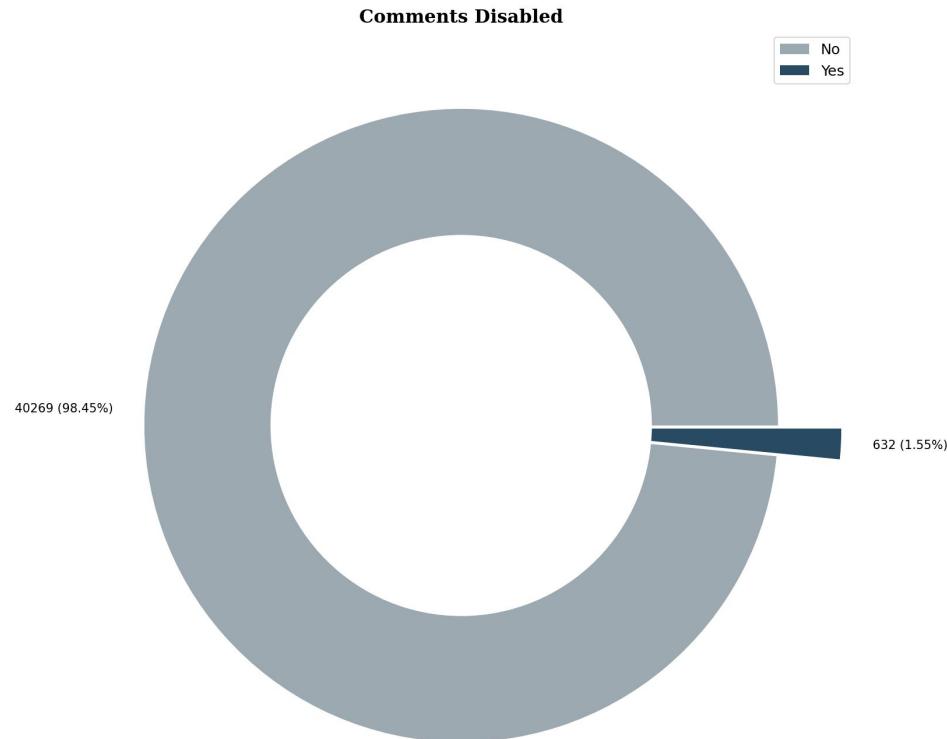


	mean	50%
views	2.360678e+06	681064.0
likes	7.427173e+04	18069.0
dislikes	3.711722e+03	630.0
comment_count	8.448567e+03	1855.0

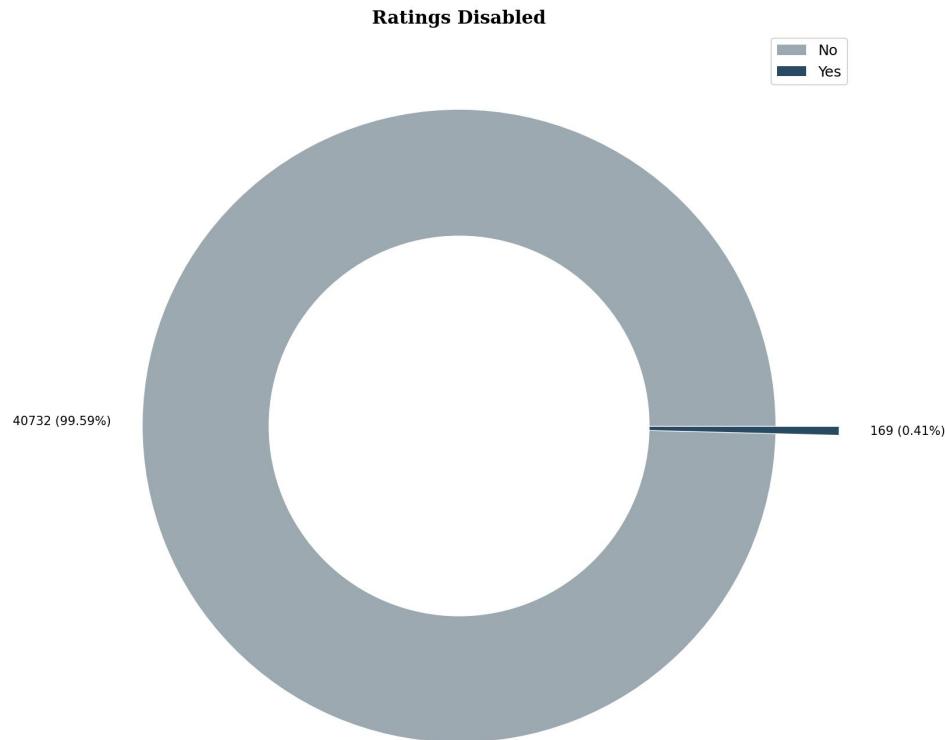
- Positively Skewed has a mean value that is greater than the median value.
- This shows that there are some videos that have very high scores for these features.

	first	last
trending_date	2017-11-14 00:00:00	2018-06-14 00:00:00
publish_time	2006-07-23 08:24:11+00:00	2018-06-14 01:31:53+00:00

- Trending videos on the dataset occur from November 14, 2017 – June 14, 2018 or 7 Months.
 - However, trending videos uploaded from 23 July 2006 – 14 June 2018.
-



- Only 1.55% of videos have comments disabled.



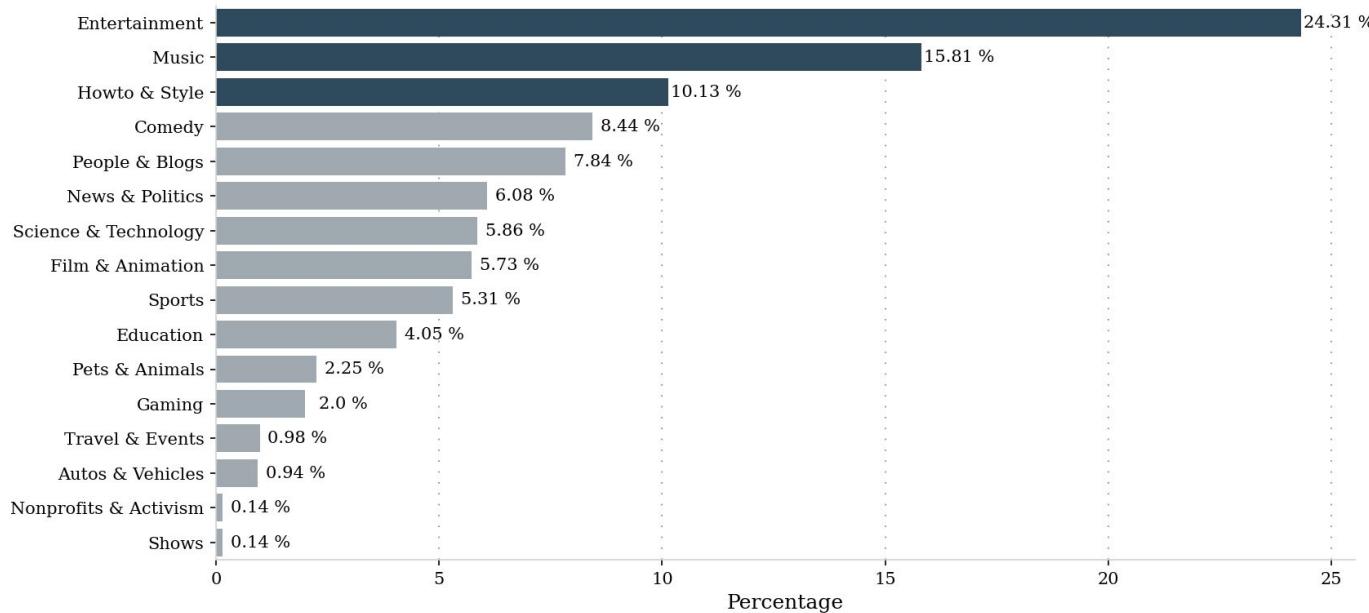
- Only 0.41% of videos have ratings disabled.

	count	unique		top	freq
video_id	40901	6351		j4KvrAUjn6c	29
video_title	40901	6455	Mission: Impossible - Fallout (2018) - Official...		29
channel_title	40901	2207		ESPN	202
category_id	40901	16		24	9944
tags	40901	6055		[none]	1534
description	40332	6901	► Listen LIVE: http://power1051fm.com/\n► Face...		58
category_title	40901	16		Entertainment	9944
channel_id	40901	1	UCBR8-60-B28hp2BmDPdntcQ		40901

- In *video_id* there are 40,901 data contained in the dataset.
- However, only 6,351 unique values.

Percentage of Videos Trending by Title Category

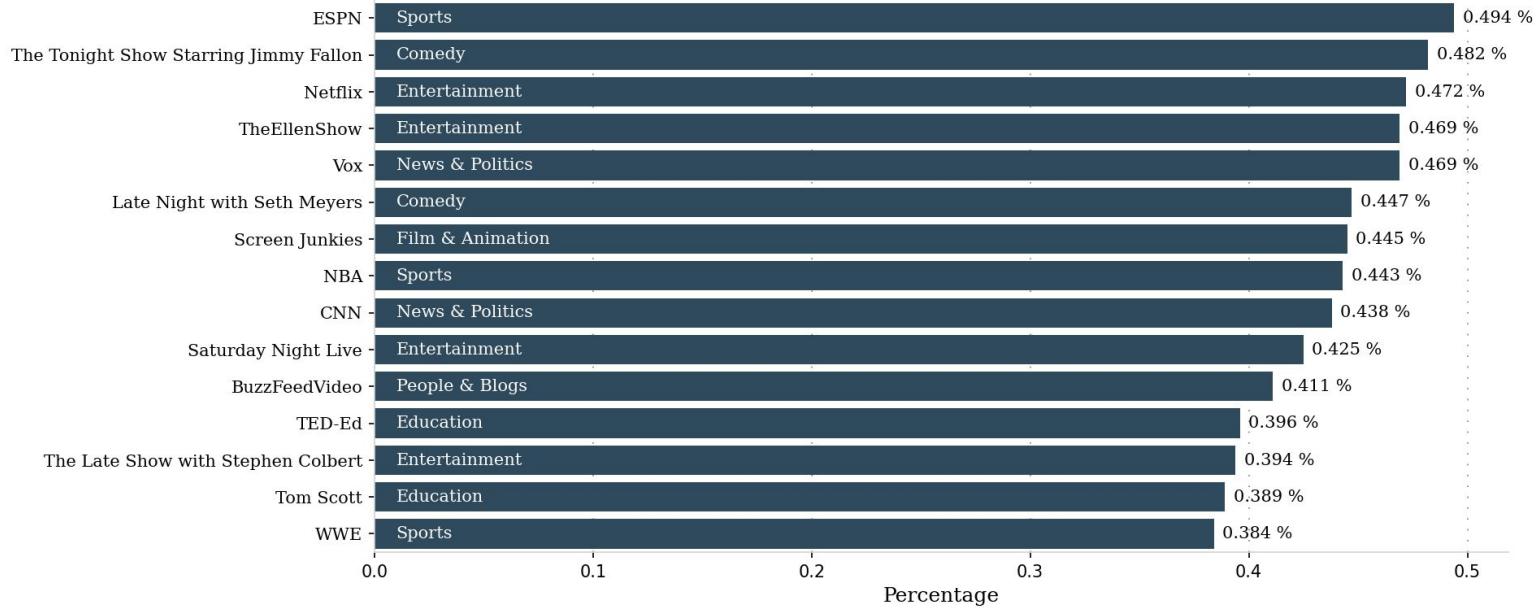
Of the 16 title categories, Entertainment, Music, and Howto & Style had trending videos of more than 10% in November 2017 - June 2018.



Trending videos can have more than one trending date.

Percentage of Videos Trending by Channel Title

Of the 2,207 channels, the top 15 channels with the highest percentage of trending videos, the entertainment title category has the most title categories.

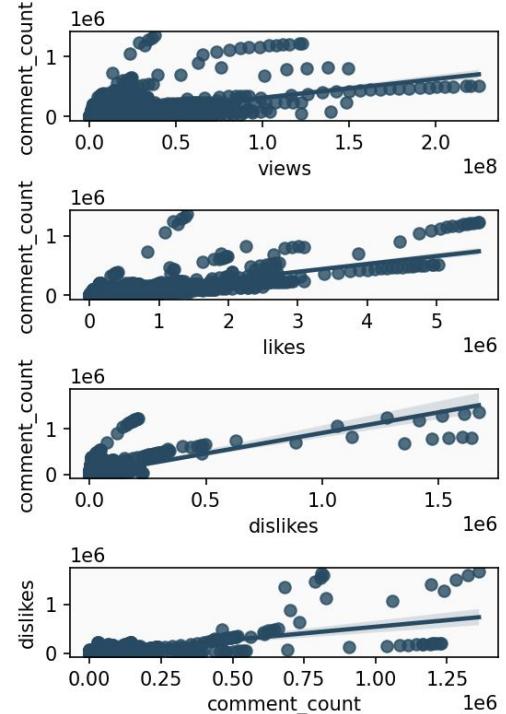
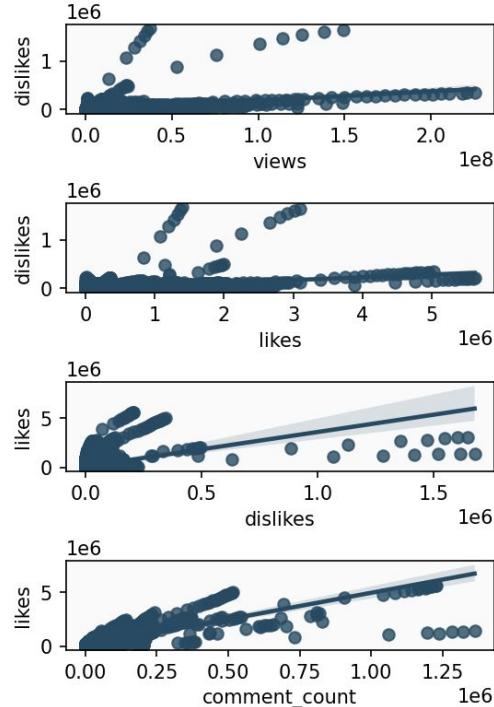
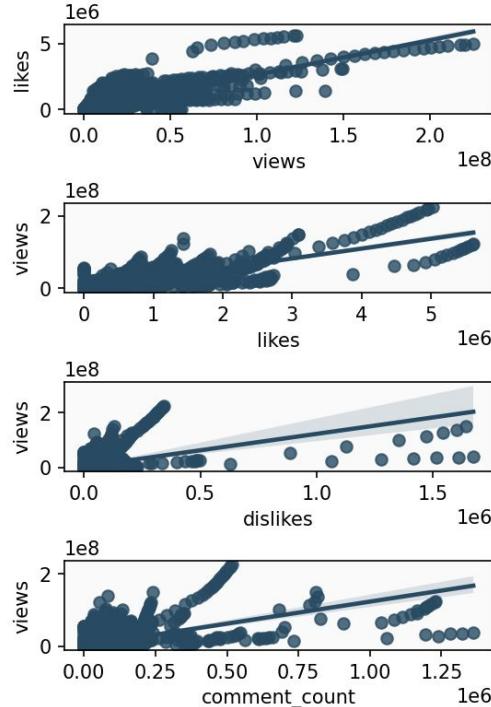


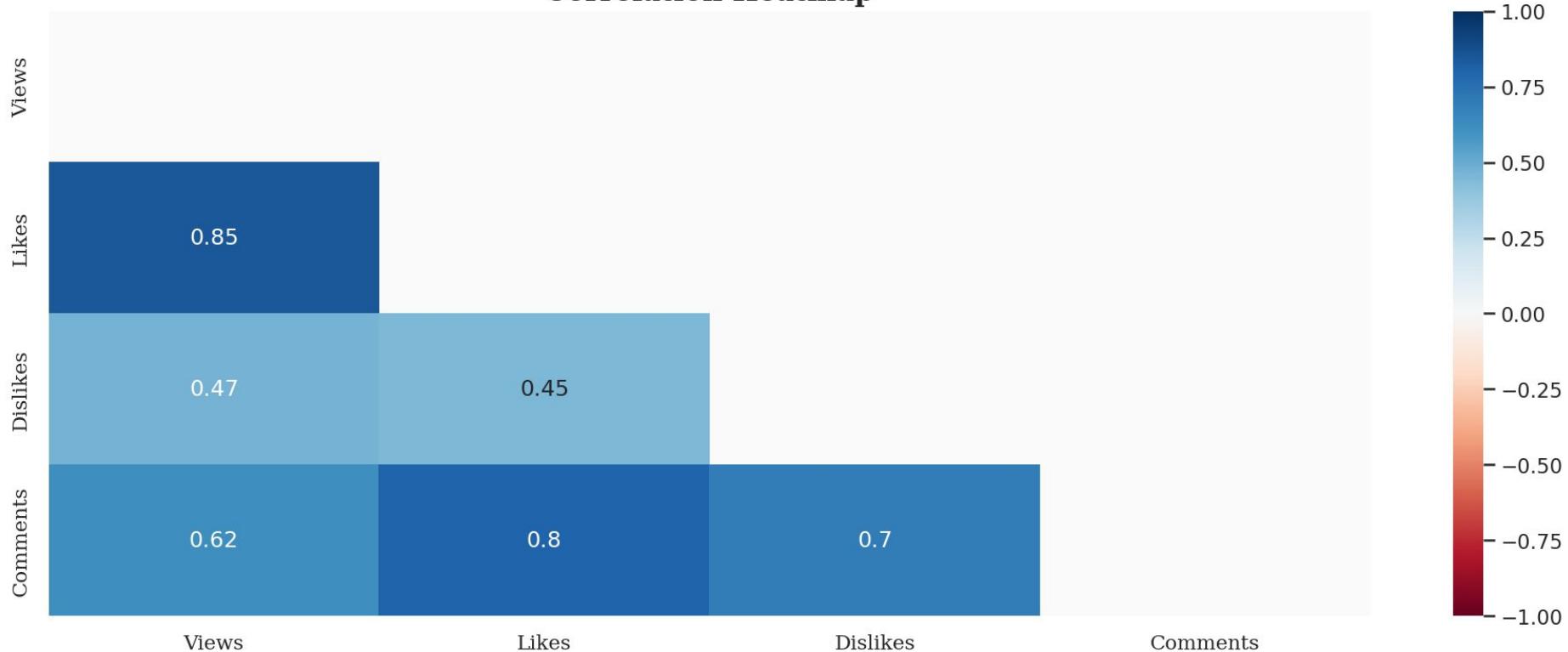
The top 15 channels with the most number of trending videos, most of them have other media or platforms.

Correlation Between Numeric Variables

The correlation between the number of views, likes, dislikes, and comments has a positive correlation.

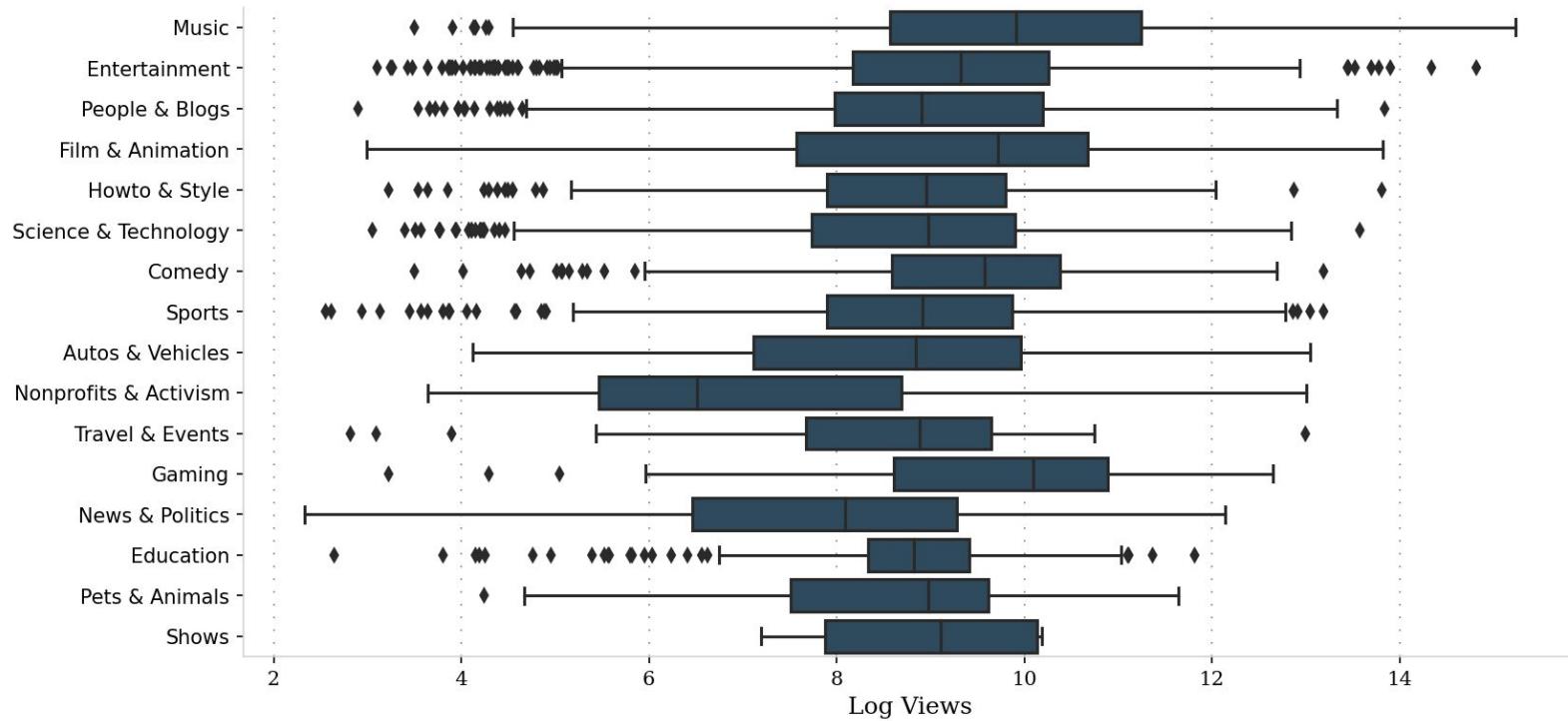
The correlation between the number of views & likes and the number of likes & comments has a high correlation.



Correlation Heatmap

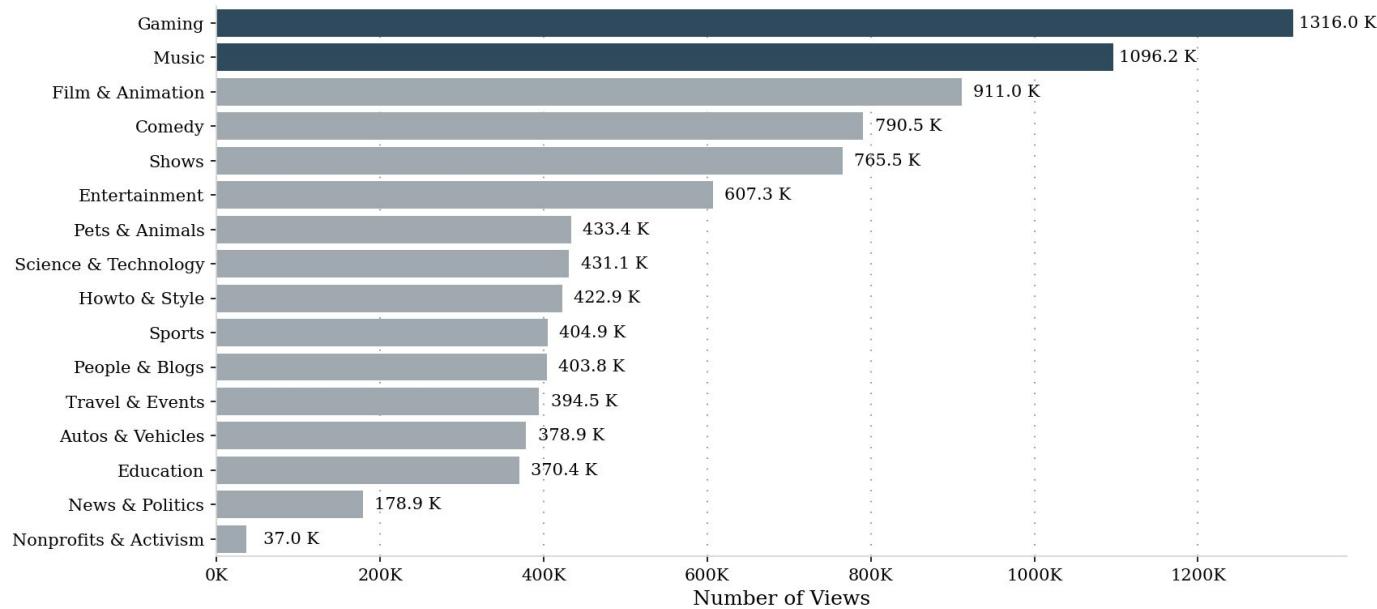
Distribution of Views by Title Category

It can be seen that almost all title categories have some videos that have very high or low Views count.



Number of Views by Title Category

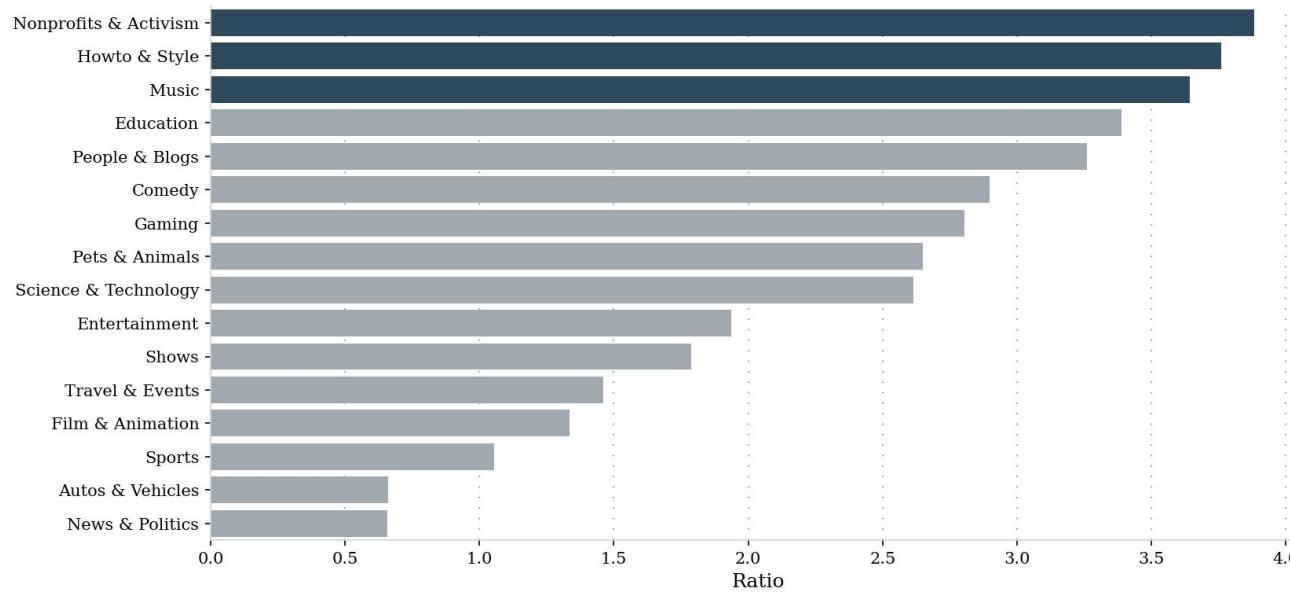
The Gaming and Music title category, has a median value of over 1 million viewers for each video.



The difference in the previous graph is because this graph uses the median value to determine the average number of views in each category title.

Views - Likes Ratio by Title Category

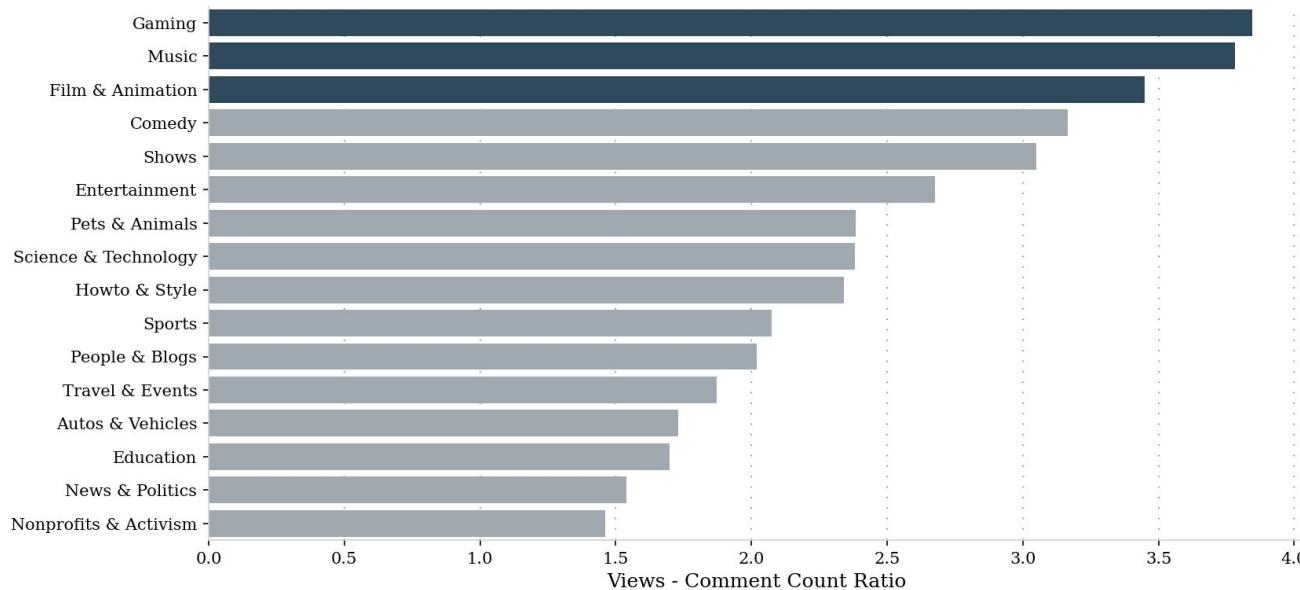
The category titles Nonprofits & Activism, Howto & Style, and Music became the top three with the highest ratio.



The Like Ratio describes the percentage of the number of likes from the number of viewers who watch videos in that category.

Views - Comment Count Ratio by Title Category

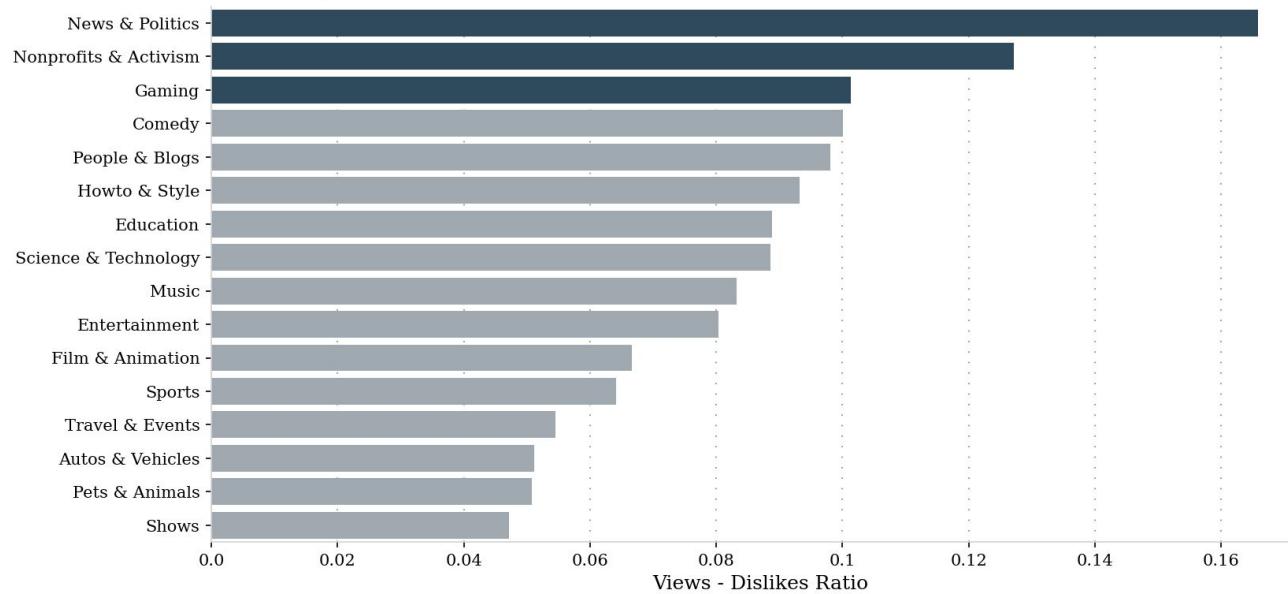
The category titles Gaming, Music, and Film & Animation became the top three with the highest ratio.



Only the Music title category which is in the top three from the previous title category is included in the top 3 likes ratio.

Views - Dislikes Ratio by Title Category

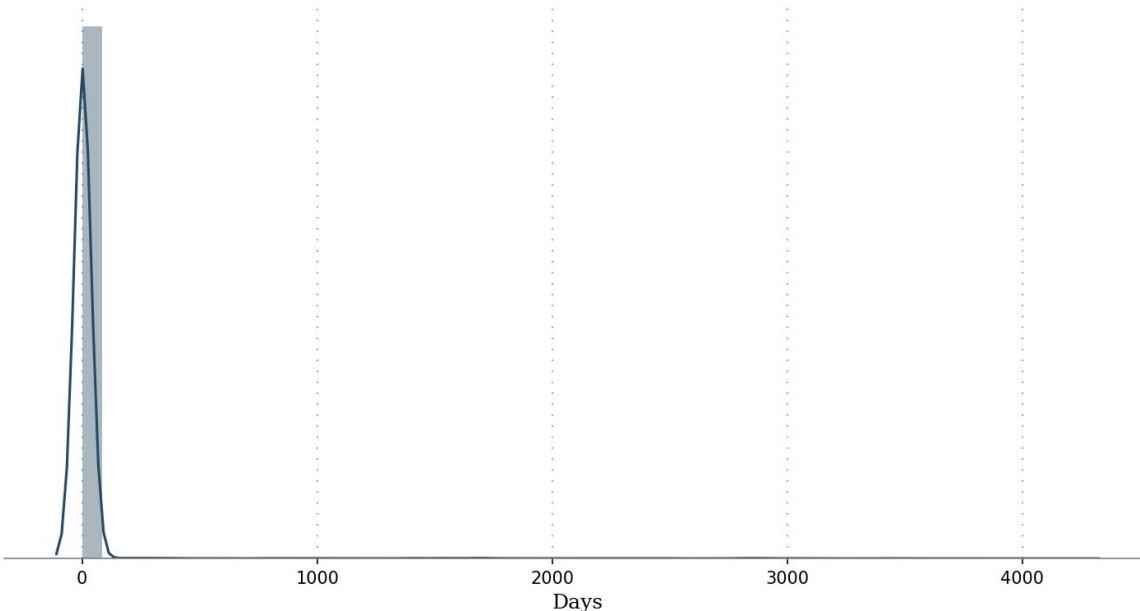
The category titles News & politics, Nonprofits & Activism, and Gaming became the top three with the highest ratio.



The Nonprofits & Activism title category previously had a high likes ratio, but has a high three dislikes ratio as well. For gaming, previously it had a high comment ratio, but it had a fairly high dislikes ratio as well.

Distribution of Days from Uploaded to Trending

There are videos that take a long time to become trending.

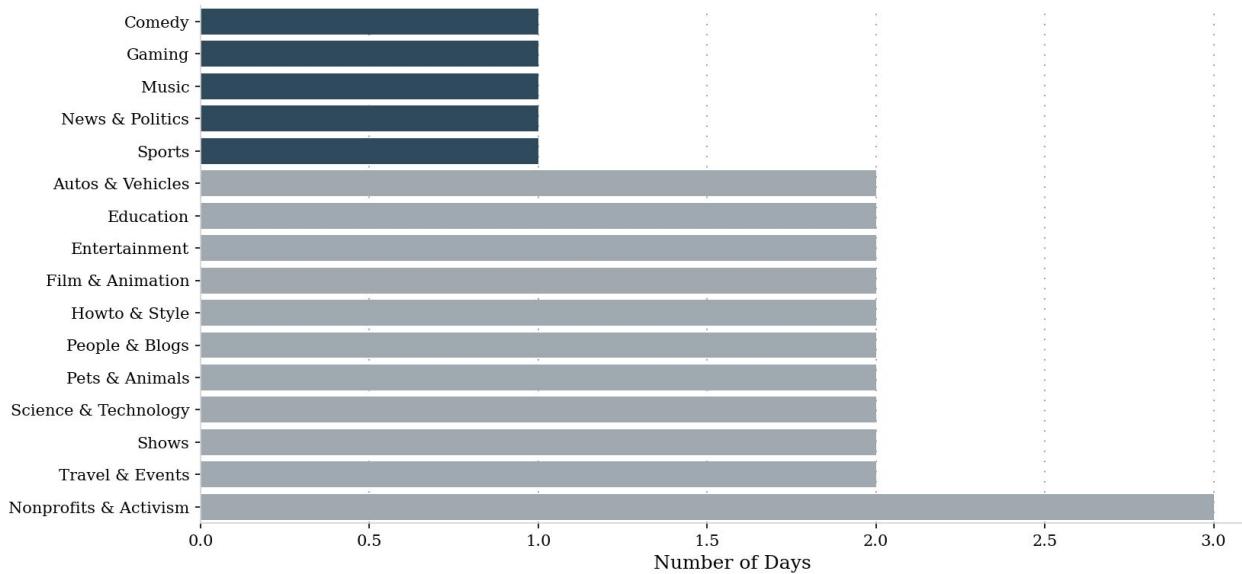


- Will use the median value to find out the average time it takes videos to trend for each title category.

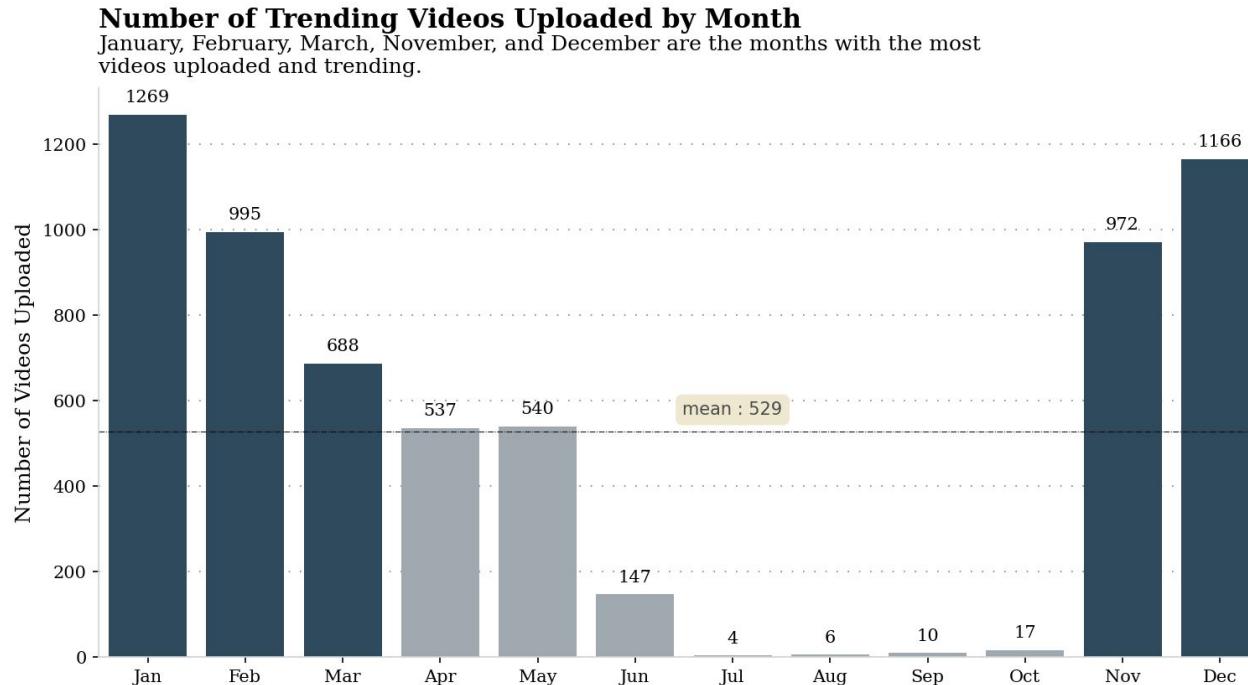
	count	mean	std	min	25%	50%	75%	max
publish_to_tranding	6351.0	22.522280	209.963296	0.0	1.0	2.0	3.0	4215.0

Number of Days for The Video to be Trending by Title Category

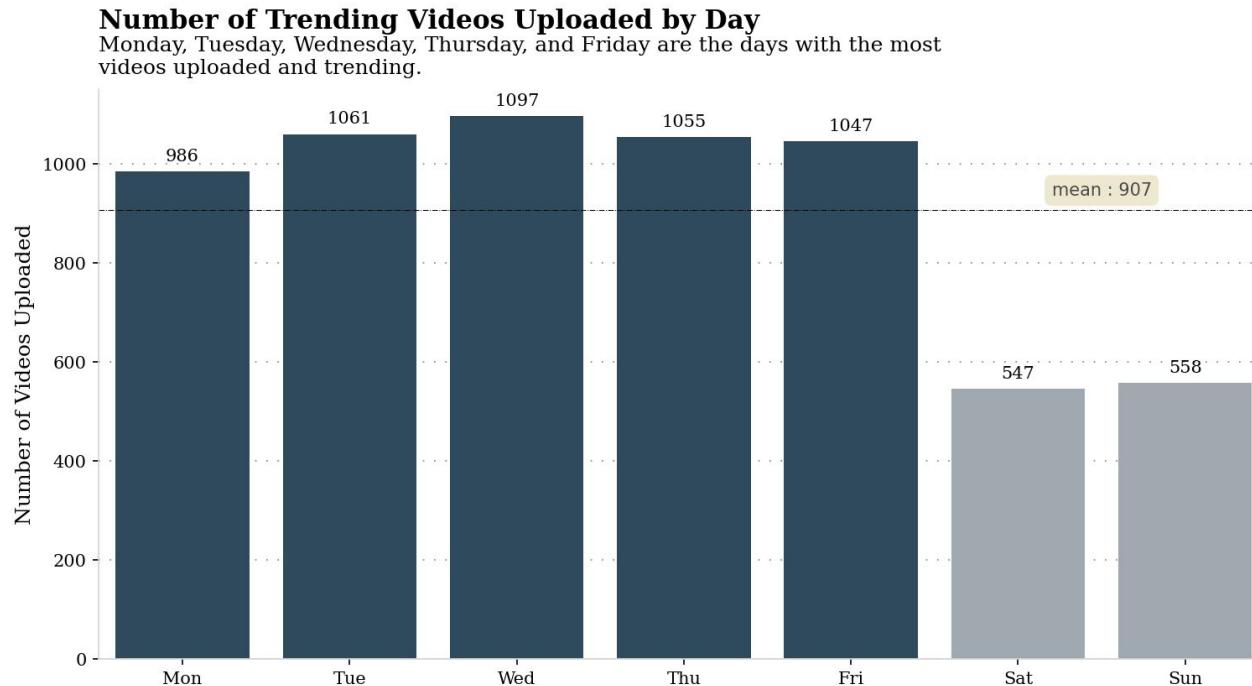
The category titles Comedy, Gaming, Music, News & Politics, and Sport have a relatively fast time to become trending.



The music category, which previously had the second largest number of trending videos, the second largest number of views, was included in the top three, the ratio of likes and the highest number of comments, and a relatively low dislikes ratio, became one of the categories with the fastest time to become trending.



The end and the beginning of the year are the months with the highest number of uploaded videos.



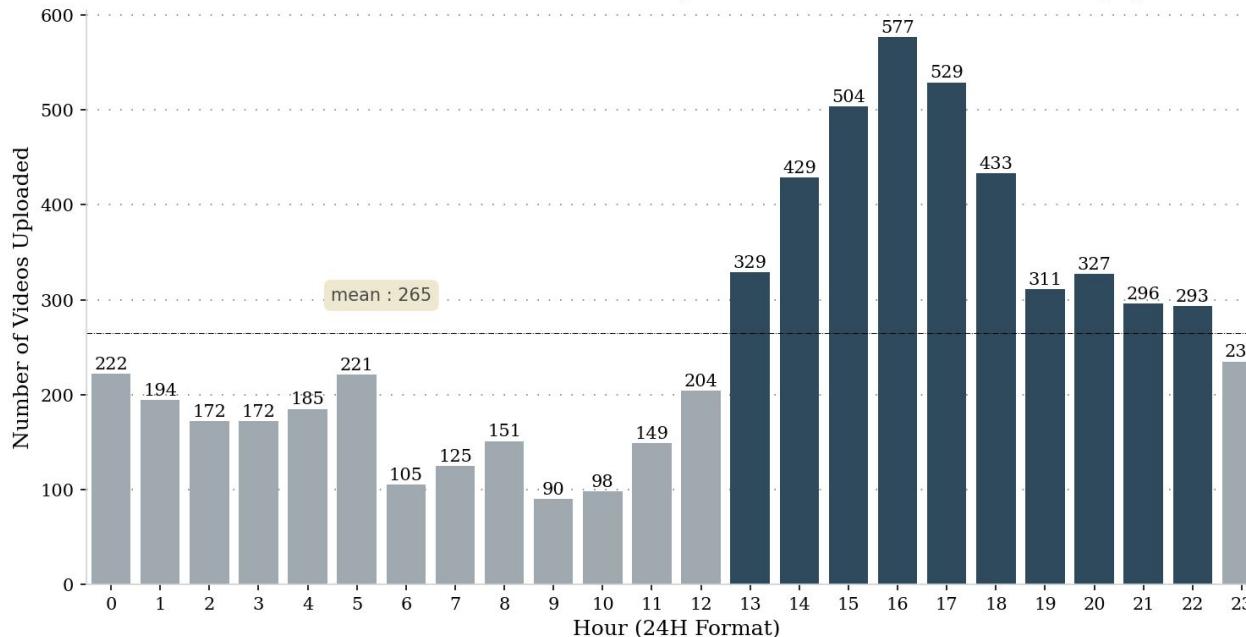
When it comes to YouTube viewing behavior, mobile is a lot like TV: The world watches at home, during prime time, and on horizontally oriented screens^[1]. Therefore, Weekdays are the time when the most videos are uploaded so that videos can be enjoyed during weekends.

[1] thinkwithgoogle.com/data-collections/youtube-stats-video-consumption-trends/

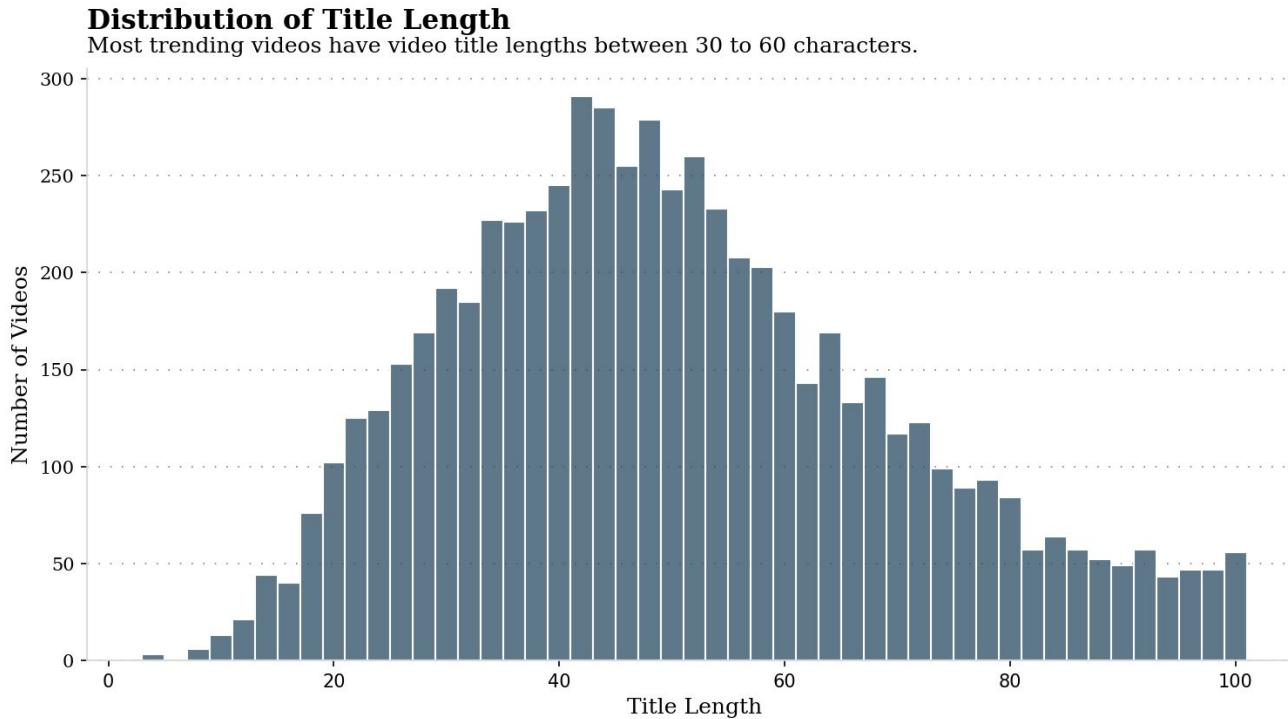
Number of Trending Videos Uploaded by Hour

Around 13 to 22 is the best time to upload with the number of trending videos more than the average value.

Around 15 o'clock to 17 the hour with the most trending videos, more than 500 videos trending uploaded.



More than 1500 videos were uploaded at around 15 to around 17. Same as in the previous slide, so that the uploaded videos can be enjoyed when the viewer finishes their work or at night.



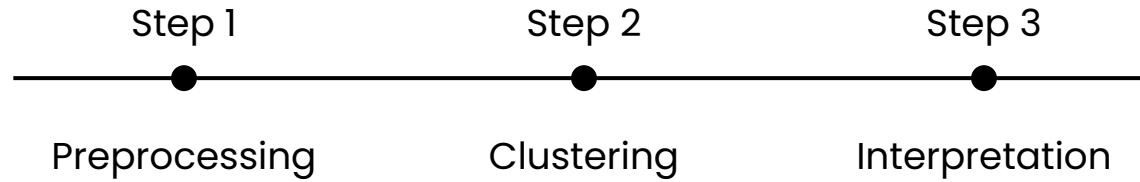
A title with a number of characters between 30 - 60 does not make the title too long or too short.

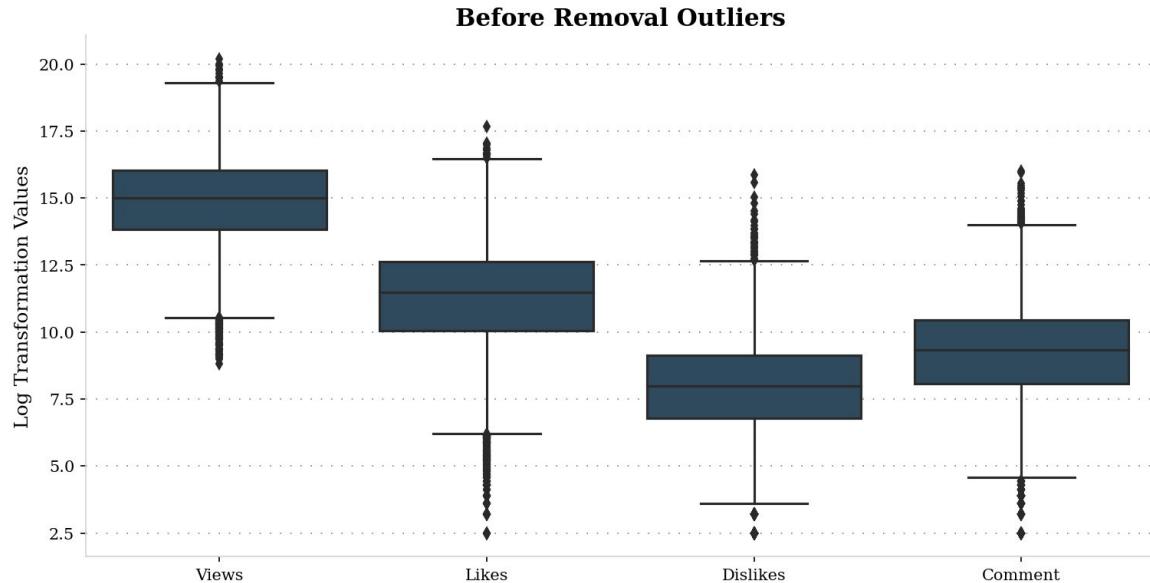


The average trending video uses the symbols ` - ` and ` | ` in the title as a separator and use the words Official, Video, Trailer, New, and Live.

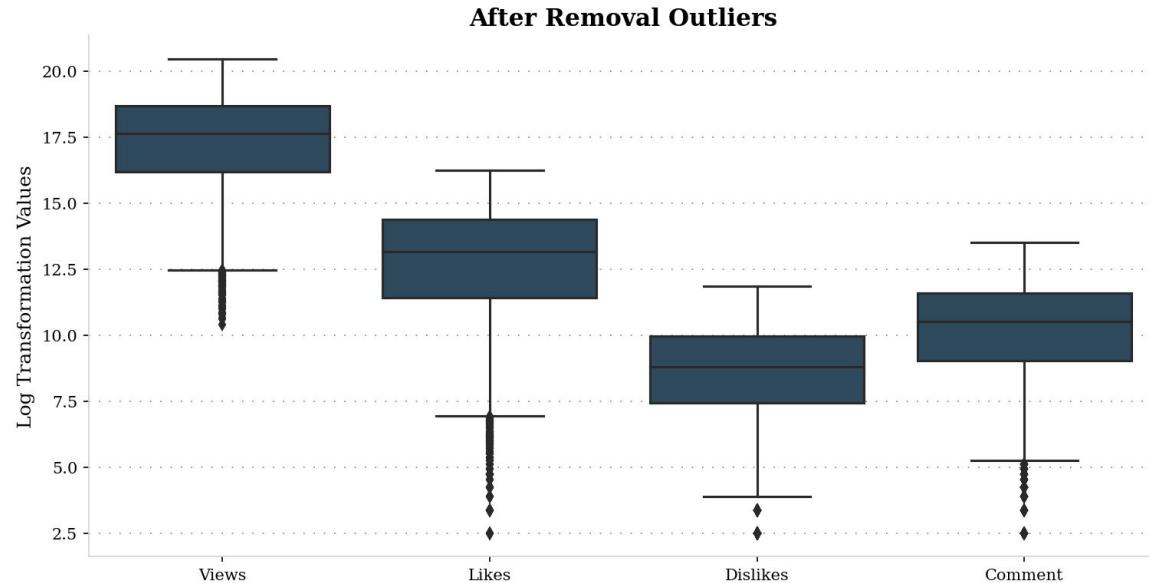
Segmentation

Segmentation Step





- There are outliers for each Numerical feature.
- The clustering to be performed is sensitive to outliers.

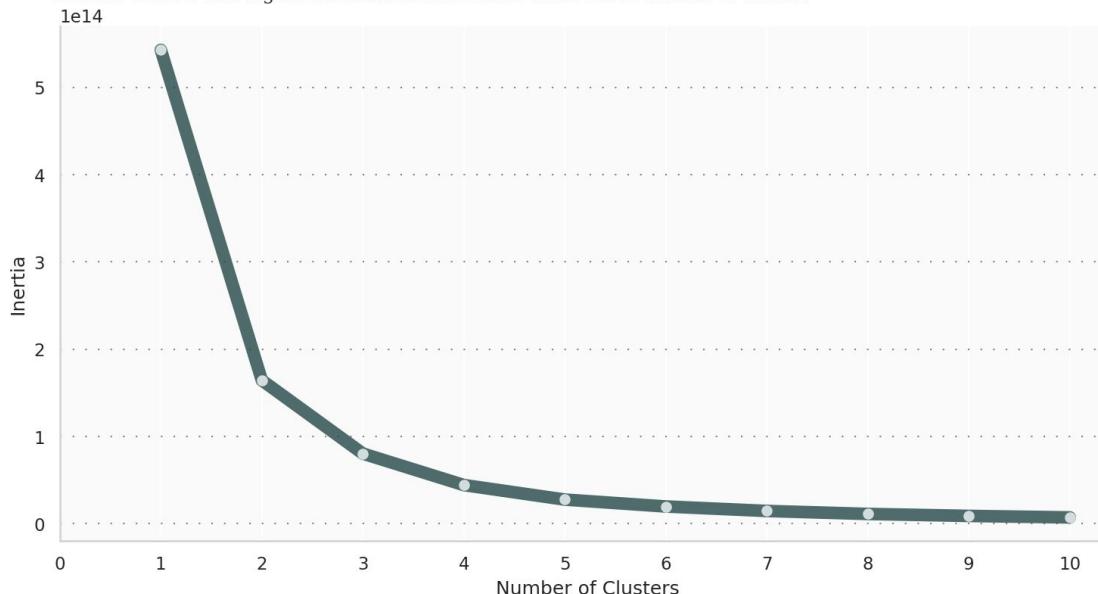


- Outliers looks better than before.
- After removing the outliers, the number of data is 5,059 from 6,351.

Number of rows before removal outliers: 6351
Number of rows after removal outliers: 5059

Elbow Method

Cluster 4 has a less significant inertia difference when a new cluster is added.



- To determine the optimal number of clusters, choose a value of k on the "elbow" i.e. the point after which the distortion/inertia begins to decrease linearly, which is **4**.

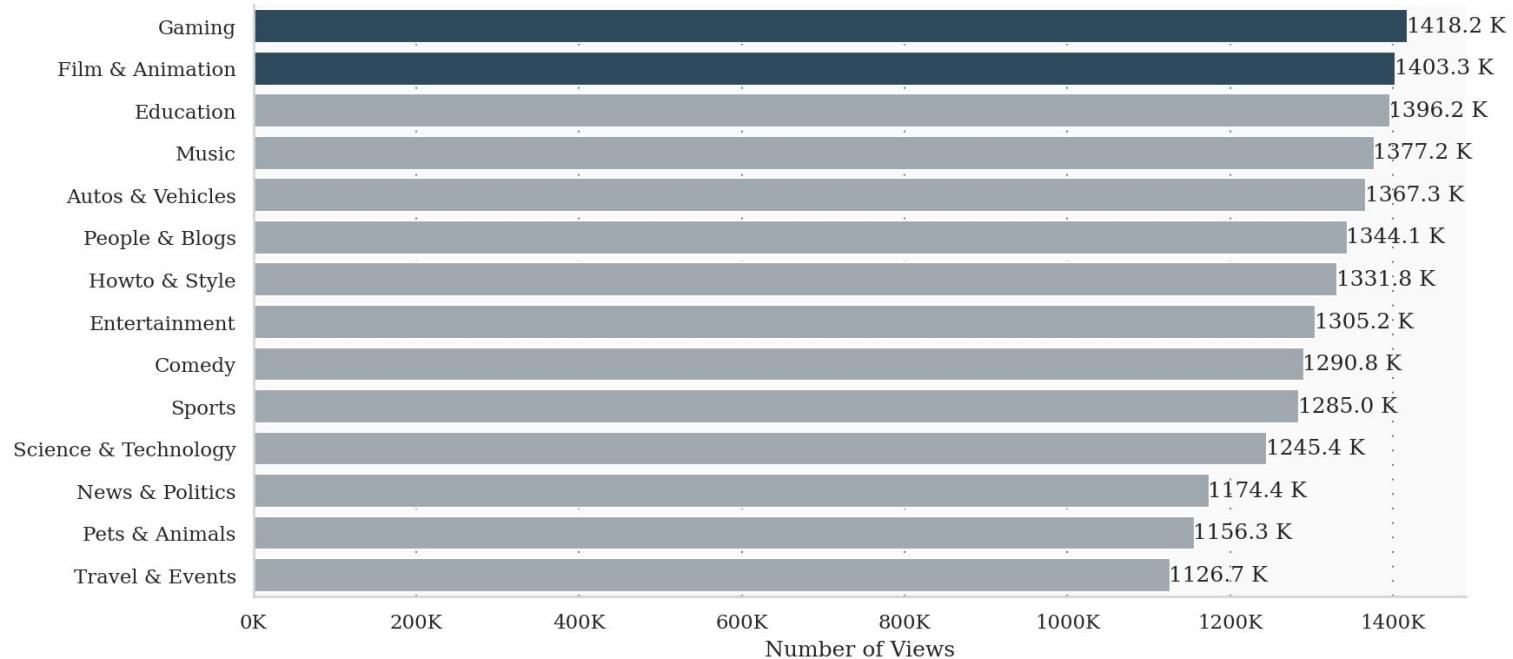
cluster_label	views	likes	dislikes	comment_count
most viewed	1.311680e+06	27755.853333	814.177778	2508.977778
many viewed	7.486628e+05	22517.210445	606.500768	2252.858679
less viewed	3.536247e+05	13103.514388	388.475540	1460.770504
least viewed	8.187838e+04	3433.423559	116.998926	445.630147

1. Most Viewed Cluster with an average **number of viewers > 1 Millions**, Likes > 25K, dislikes > 750, and Comments > 2K.
 2. Many Viewed Cluster with an average **number of viewers > 500K**, Likes > 20K, dislikes > 500, and Comments > 1.5K.
 3. Less Viewed Cluster with an average **number of viewers > 250K**, Likes > 10K, dislikes > 250, and Comments > 1K.
 4. Least Viewed Cluster with an average **number of viewers > 50K**, Likes > 1.5K, dislikes > 50, and Comments > 250.
-

Cluster Most Viewed

Number of Views by Categories Title for Cluster Most Viewed

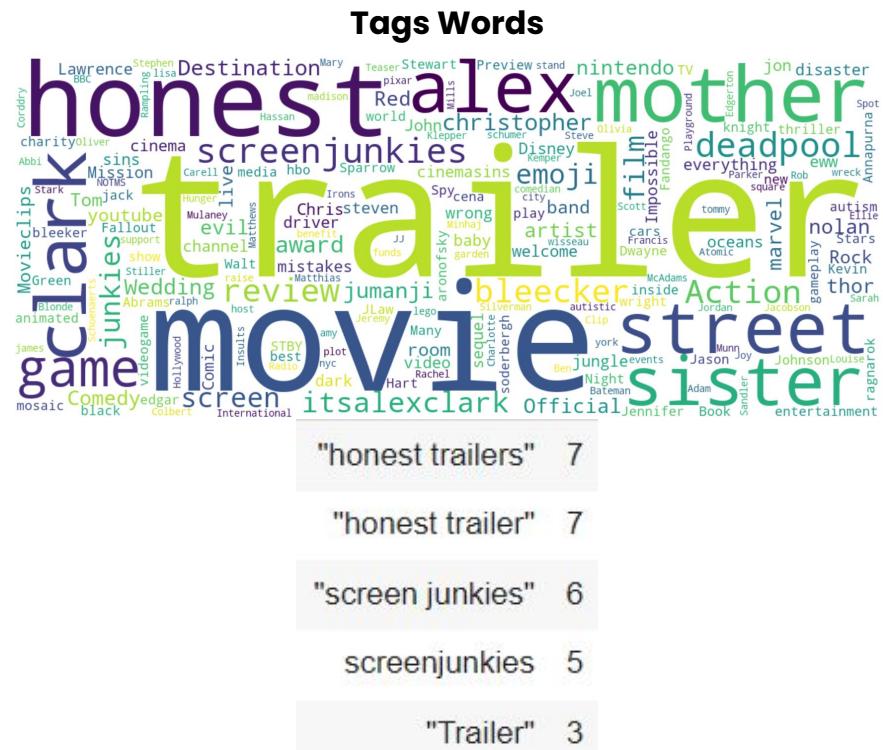
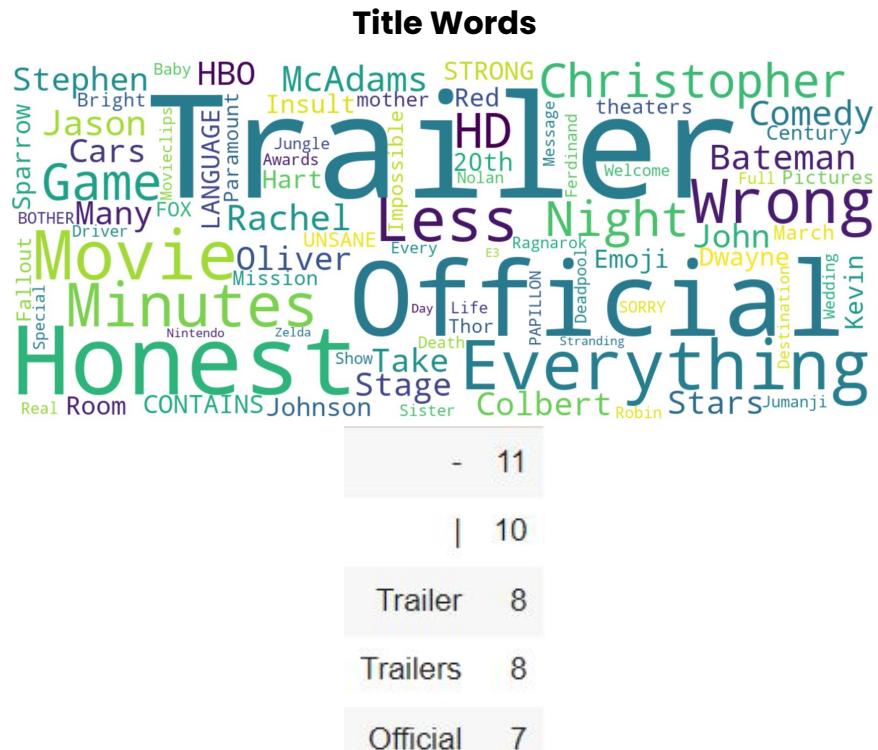
Category titles Gaming and Film & Animation are the row of title categories with the highest number of views in this cluster, more than 1400K viewers.



Data Scientist – Take Home Test

Segmentation Step 3 – Interpretation

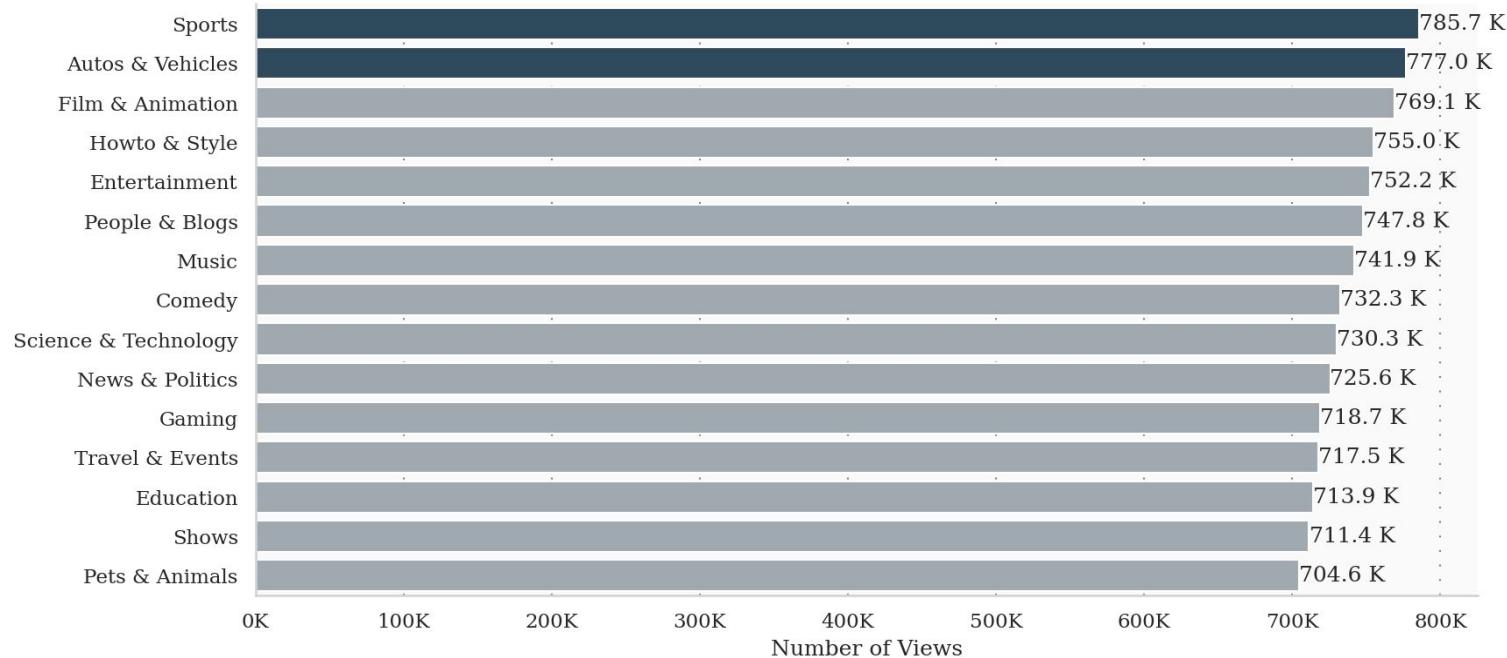
Cluster Most Viewed



Cluster Many Viewed

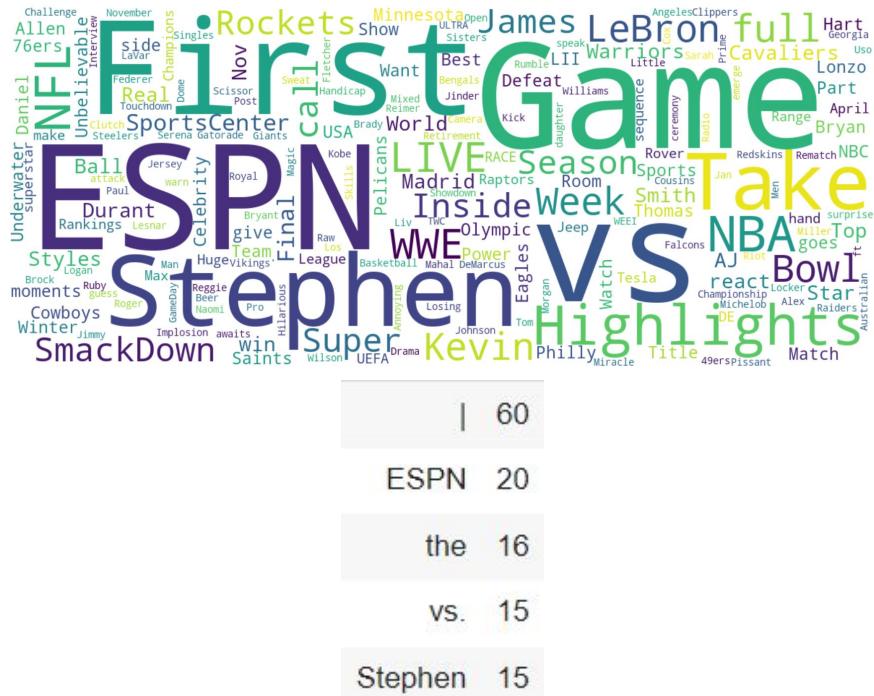
Number of Views by Categories Title for Cluster Many Viewed

Category titles Sports and Autos & Vehicles are the row of title categories with the highest number of views in this cluster, more than 775K viewers.



Cluster Many Viewed

Title Words



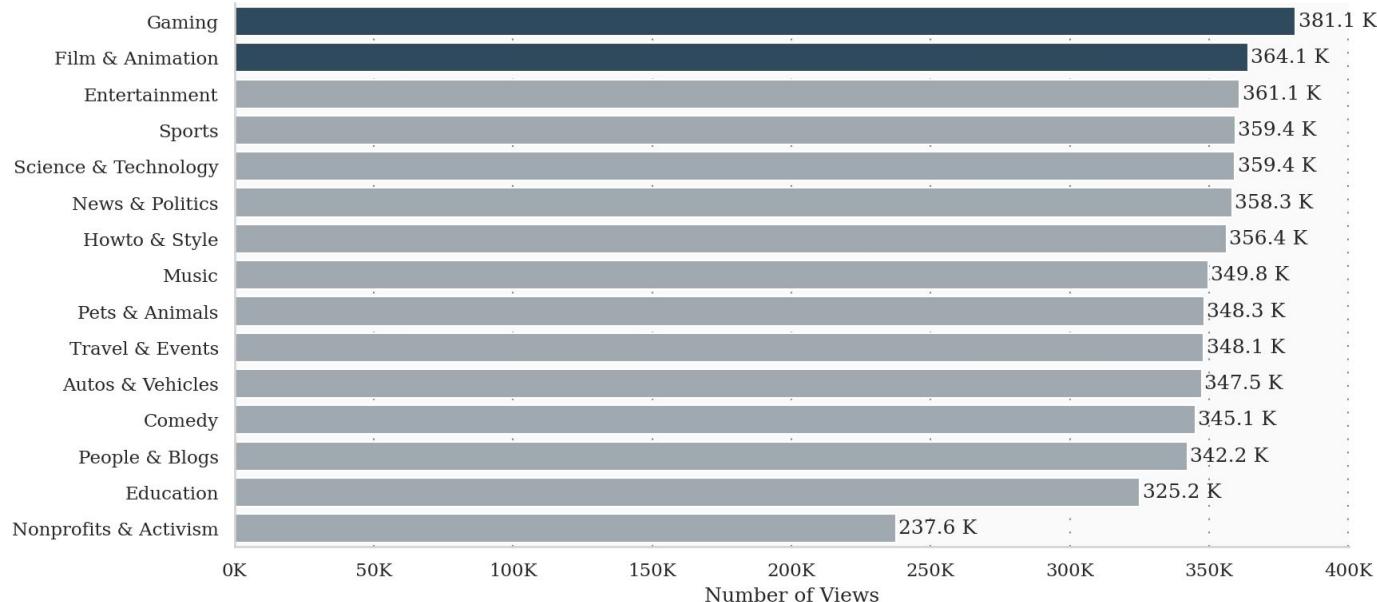
Tags Words



Cluster Less Viewed

Number of Views by Categories Title for Cluster Less Viewed

Category titles Gaming and Film & Animation are the row of title categories with the highest number of views.

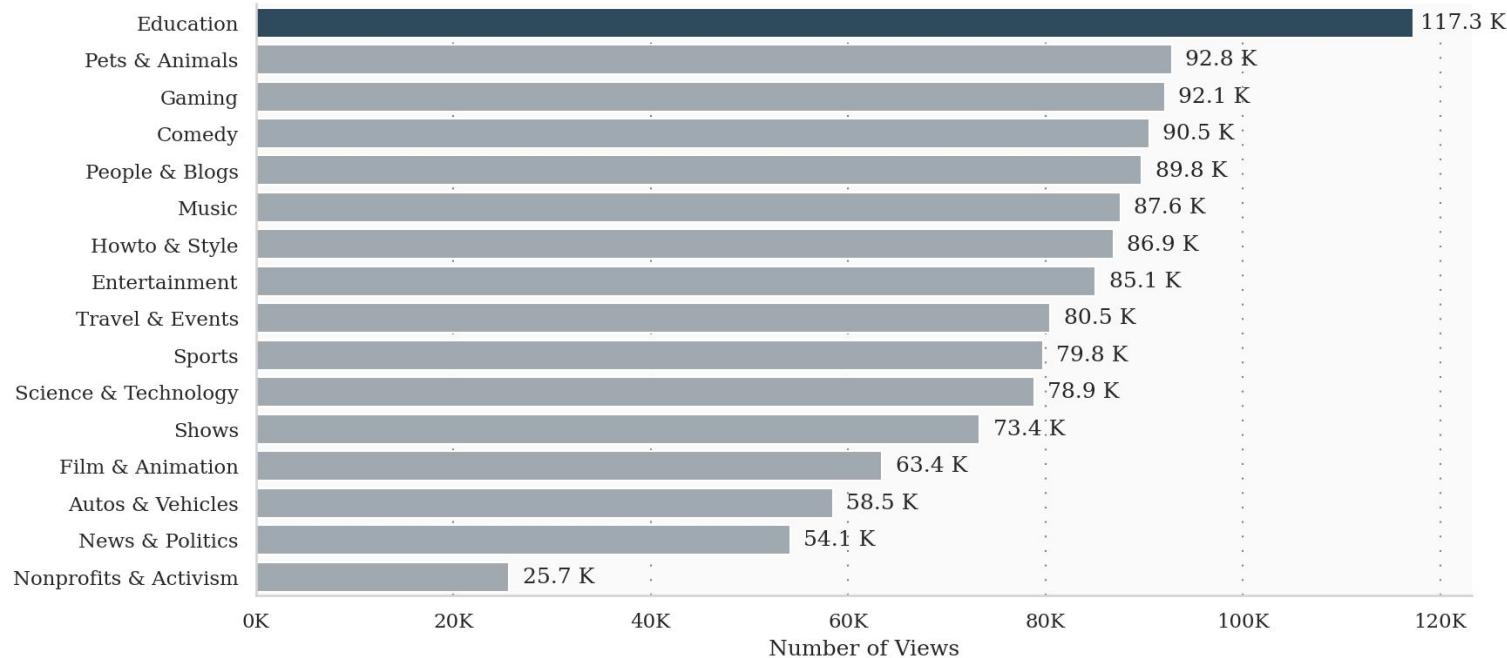


The title category with the most views in the Less Viewed cluster is the same as in the Most Viewed cluster, only differing in the number of views.

Cluster Least Viewed

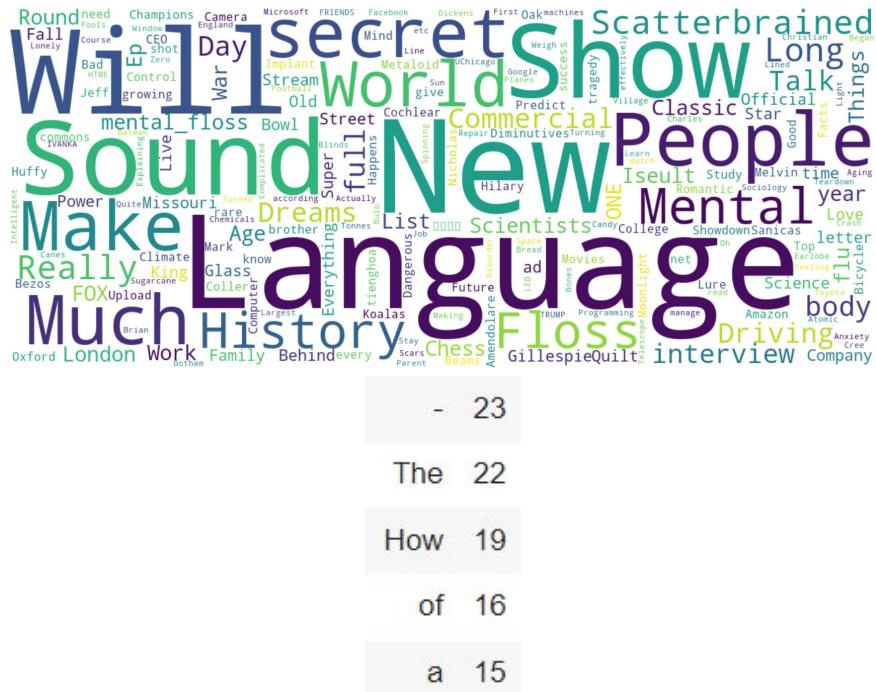
Number of Views by Categories Title for Cluster Least Viewed

Category titles Educations are the row of title categories with the highest number of views, more than 100K viewers.

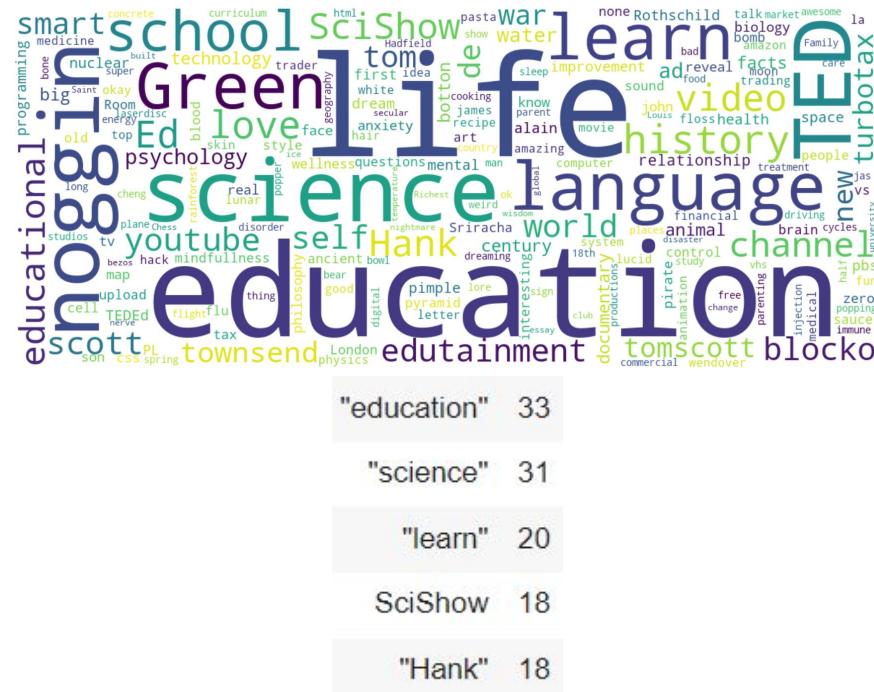


Cluster Least Viewed

Title Words



Tags Words



Summary

Data Scientist - Take Home Test



The **top three title categories with the most trending videos** are **Entertainment, Music, and Howto & Style**.



Music title category has **the second highest** number of views with **more than 1 million views**, **the third highest Views – Likes Ratio** with **more than 3.5%**, the **second highest Views – Comments Ratio** with **more than 3.5%**, and quite low **Views–Dislikes Ratio** with **less than 0.09%**. Also, **became one of the categories with the fastest time to become trending**.



The beginning and end of the year are the most common months for uploading videos. By **uploading on weekdays** and **between around 15.00 o'clock** and **around 17.00 o'clock**.



The number of characters in the **video title** ranges from **30 to 60 at most**, with the **symbol / or - as a separator** and in general **words** like **Official, Video, Trailer, and Live** are used the most.

For Clusters:

1. **Cluster Most Viewed – Gaming** and **Film & Animation** Title Category with **Tags Honest Trailer** and **Screen Junkies**.
2. **Cluster Many Viewed – Sports** and **Autos & Vehicles** Title Category with **Tags ESPN** and **Stephen A.**.
3. **Cluster Least Viewed – Educational** Title Category with **Titles that depend on the content** and **contain question words** like **What, Why, When**, etc.

For each cluster, the **video title uses** the – or | **symbols for the separator**.

Thank You

