Title of the project: YouTube Statistics  
Vertical industry: Entertainment Industry  
Business process: YouTube

Contents

[1.Executive Summary: 3](#_Toc90245938)

[1.1.Background: 4](#_Toc90245939)

[1.2.Objectives: 4](#_Toc90245940)

[1.3. Description of Data: 4](#_Toc90245941)

[1.4.Data Preparation Report: 6](#_Toc90245942)

[2. Problem Statement: 7](#_Toc90245943)

[2.1.Trending Videos and Views: 8](#_Toc90245944)

[2.2.Trending Vs Non-Trending Videos: 8](#_Toc90245945)

[2.3.Channel Ranking: 9](#_Toc90245946)

[2.4.Trending Categories: 9](#_Toc90245947)

[2.5.Trending Videos Published Month: 9](#_Toc90245948)

[2.6.Attributes and Videos: 10](#_Toc90245949)

[2.6.Tableau Dashboard: Tableau Public 11](#_Toc90245950)

[3. Tableau Storyboard: 11](#_Toc90245951)

[3.1.Problem Statement: 12](#_Toc90245952)

[3.2.Purposes/Questions/Application: 12](#_Toc90245953)

[3.3.Conclusion: 12](#_Toc90245954)

[3.4.Recommendations: 12](#_Toc90245955)

# 1.Executive Summary:

YouTube is known as the world-famous online video sharing website and social media platform owned by Google. It was launched on February 14, 2005. YouTube uses different combinations of factors to determine the trending videos and the most common combination being the user interactions in the form of likes, comments, shares and views. Trending videos help viewers to get an understanding and awareness about what is happening in YouTube and in the world.

Trends are dependent on various factors and few trends are predictable which are most commonly any music videos of famous artists or any new movie trailer. However, there are certain trends which are surfaced due to interesting content. Different countries have different set of trending videos based on region. The list of trending videos is updated at a frequency of 15 minutes and the video priorities may change from top to bottom or being in the same position as well.

In this project, United States trending videos would be analysed which constitutes of 40949 videos. Considering that the videos might be trending on multiple days and the same video may appear on the list consecutively means that 40949 videos are not the unique videos. Among the 40949 videos, we have 6351 unique videos. Unique videos are found out in R using the unique function on video\_id.

We have video\_id being the unique value for the video. We have the trending date, title of the video, channel title, video category, published time and the tags associated with the video. As mentioned earlier, we see the trending\_date which means the date on which the video was in trending status. The same video might appear multiple times with different trending date because the video maybe trending on another date as well with different likes, comments, shares. We also have the number of views, likes, dislikes and comments. We can see if the comments and ratings being disabled for the videos.

YouTube should consider these statistics to understand what are the major factors which are involved for the video to be in trending status. YouTube has given multiple factors being responsible for the video to be trending status and we can narrow the filter to minimal categories which would help the video/content creators to ensure that their video is being reached out to wide audience. These are the three major targeted population for whom this analysis would be helpful:

* A Youtuber who is a video/content creator as his profession
* A researcher who would be deriving insights from the behavioural pattern of YouTube trending videos.
* An advertiser, who can capture the opportunity to showcase his advertisements on the videos which would go in trending status once they are published.
* Companies and businesses which would rely on this analysis to advertise and sponsor appropriate channels at the right time and improvise their growth and performance.

Apart from the customary analysis, we would understand the factors to help the YouTube channels in increasing their subscribers count which would eventually benefit the individual or an organization monetarily and improvise their social media presence.

## Background:

YouTube statistical analysis would help us in analysing, finding, measuring and comparing the key attributes of YouTube trending videos. This would enable the growth prospects of YouTube channels by increasing their subscriber count which would eventually channelize the monitory benefits of the channel and improvise their social media presence.

These are different factors for analysing the trending videos:

* Understanding the difference between trending and non-trending videos.
* Analysing the video categories to analyse the most trending category.
* Analysing the basic attributes of trending videos
* Analysing the time and days of week when the video is being published and when the video is being moved to trending status.

## Objectives:

* Statistical analysis of trending YouTube videos over time.
* Analysis of numbers of views and the understanding of views to become the videos to be trending.
* Understanding of attributes which are correlated and the connection between the attributes which are likes, dislikes, title length and other attributes.
* YouTube channels with the greatest number of trending videos.
* Video categories based on priority which have the greatest number of trending videos.
* Understanding the time when the videos are being published and identifying the trend when the videos are published during the week and the timings.

## 1.3. Description of Data:

**Data Source:**

URL: <https://www.kaggle.com/datasnaek/youtube-new>

Format: csv

**Data Description Report:**

The dataset used for the analysis consists of trending videos from 2017-2018 which are for United States region. The dataset constitutes of 40950 rows and 16 columns.

Dictionary:

* Video\_id: Unique id for each YouTube video
* Trending\_date: Date on which the video was in trending status
* Title: YouTube video title
* Channel\_title: Video channel of the YouTube video
* Category\_id: Categories are divided based on the table below:

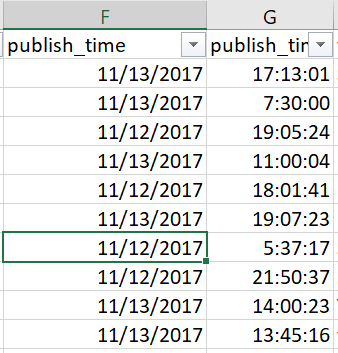
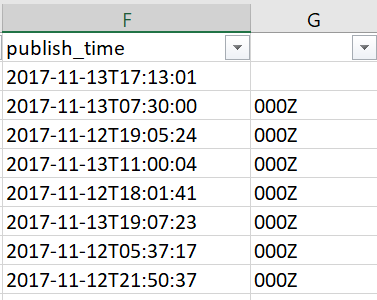
|  |  |
| --- | --- |
| **Category\_id** | **title** |
| 1 | Film & Animation |
| 2 | Autos & Vehicles |
| 10 | Music |
| 15 | Pets & Animals |
| 17 | Sports |
| 18 | Short Movies |
| 19 | Travel & Events |
| 20 | Gaming |
| 21 | Videoblogging |
| 22 | People & Blogs |
| 23 | Comedy |
| 24 | Entertainment |
| 25 | News & Politics |
| 26 | Howto & Style |
| 27 | Education |
| 28 | Science & Technology |
| 29 | Nonprofits & Activism |
| 30 | Movies |
| 31 | Anime/Animation |
| 32 | Action/Adventure |
| 33 | Classics |
| 34 | Comedy |
| 35 | Documentary |
| 36 | Drama |
| 37 | Family |
| 38 | Foreign |
| 39 | Horror |
| 40 | Sci-Fi/Fantasy |
| 41 | Thriller |
| 42 | Shorts |
| 43 | Shows |
| 44 | Trailers |

* Publish\_time: Time when the video is being published
* Tags: Video tags which are used to give context about the YouTube video
* Views: Number of views on the video
* Likes: Number of likes for the video
* Dislikes: Number of dislikes for the video
* Comment\_count: Number of comments on the video
* Thumbnail\_link: Quick snapshot of the YouTube video
* Comments\_disabled: If the comments are disabled for the video. If they’re disabled, then it’s true, otherwise, its mentioned as false.
* Ratings\_disabled: If the ratings are disabled for the video. If they’re disabled, then its true, otherwise, its mentioned as false.
* Video\_error\_or\_removed: If there are any errors while browsing the video or if the video is removed. Its false if none of the mentioned scenarios have pertained.

## Data Preparation Report:

Creating new attributes:

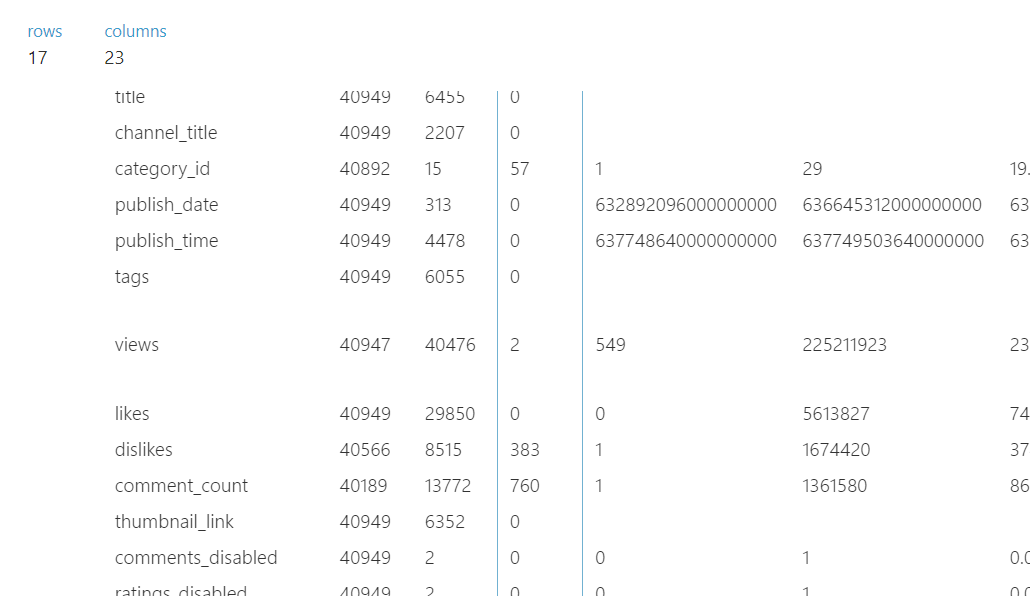
* Modified published\_time and removed 000Z by using delimited function on excel.
* Created additional column named as published\_time



**Clean Missing Data:**

We are using Azure ML Studio for data cleaning and preparation:

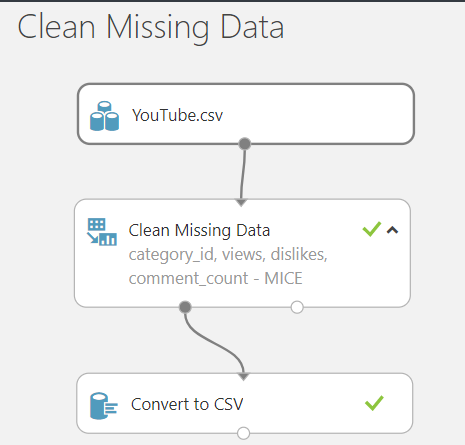
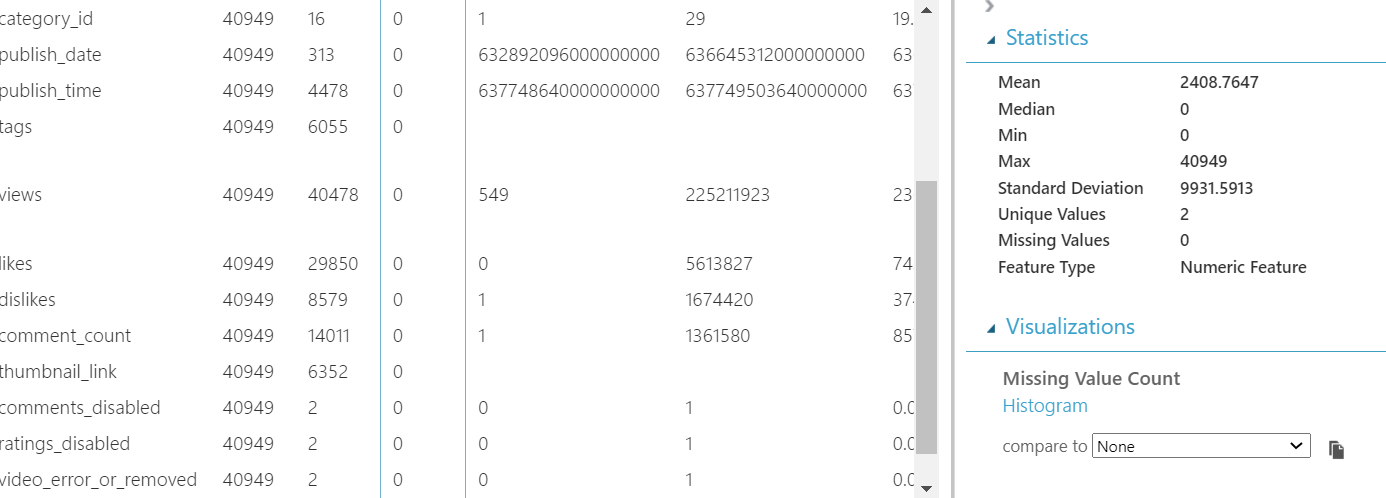
|  |  |
| --- | --- |
| **Attribute** | **Missing Value Count** |
| Category\_id | 57 |
| Views | 2 |
| dislikes | 383 |
| comment\_count | 760 |



**Reason for missing data:**

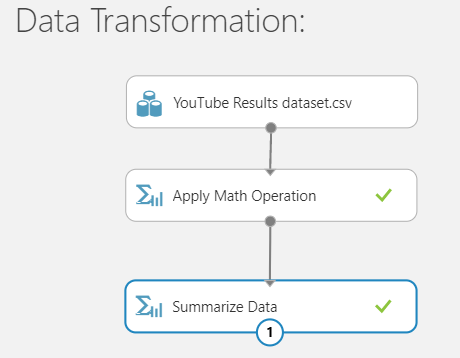
MCAR: Missing completely at random. We are using imputation method ‘MICE’ to fill the missing values with number of iterations being 5. We are generating the missing value indicator column to identify the output results of Missing values.

Once the values are imputed, we do not see any missing values:



Data Transformation: We have transformed the data using apply math pill by using the function Ln+1 to reduce the skewness.

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Original Skewness** | **Ln+1 Skew** |
| Views | 12.240222 | -0.39699 |
| dislikes | 40.194854 | -0.14248 |
| comment\_count | 19.770989 | -0.40025 |
| likes | 10.923768 | -0.9778 |



# 2. Problem Statement:

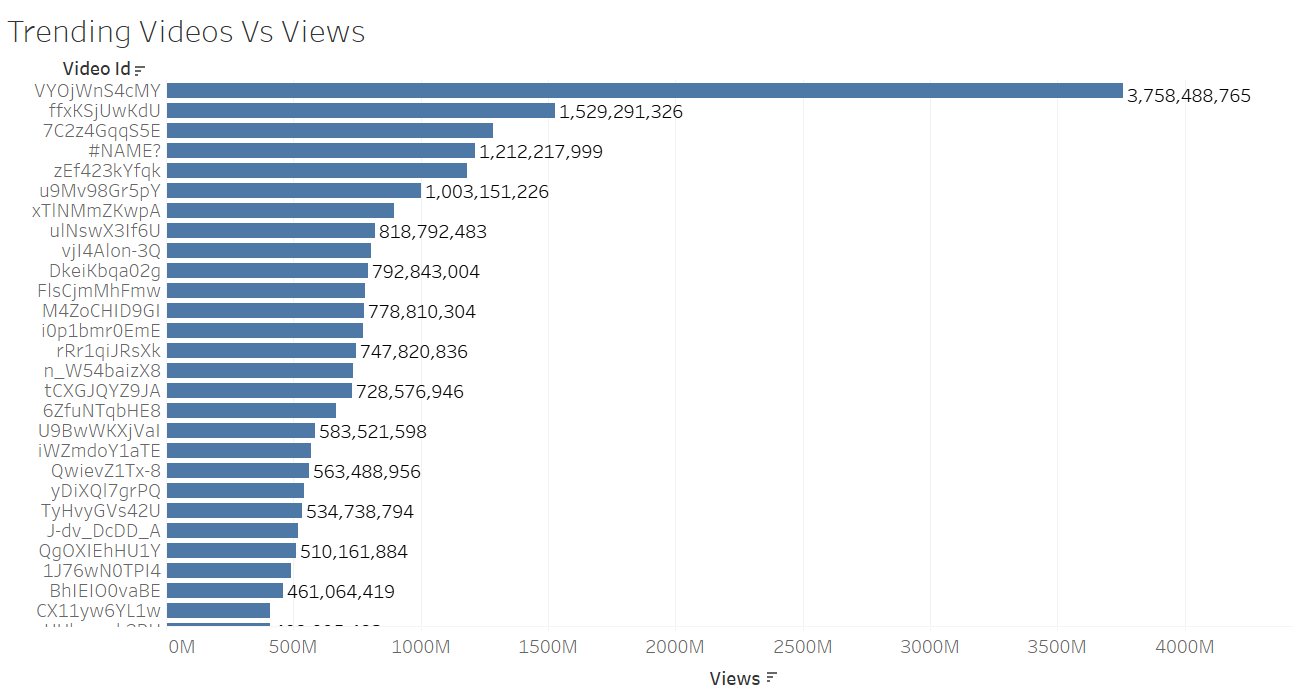
YouTube is known as the world-famous online video sharing website and social media platform owned by Google. YouTube uses different combinations of factors to determine the trending videos and the most common combination being the user interactions in the form of likes, comments, shares and views. Trending videos help viewers to get an understanding and awareness about what is happening in YouTube and in the world.

These statistics are important in analysing the factors and attributes which are important for the video to be in trending status. Major objectives for the project are listed below:

* Statistical analysis of trending YouTube videos over time: This would help us in understanding the video categories which are most trending over a period of time and analyse the type of videos to be released by the business or to invest in advertising for the businesses.
* Analysis of numbers of views and the understanding of views to become the videos to be trending: Analysing the views with a basic threshold to understand if most of the trending videos have any pattern of views on the videos.
* Understanding of attributes which are correlated and the connection between the attributes which are likes, dislikes, title length and other attributes: Understanding the correlation between all the attributes.
* YouTube channels with the greatest number of trending videos: Analysing the channels to check if the trending videos are from popular channels or if the channels belong to an individual content creator.
* Video categories based on priority which have the greatest number of trending videos: Analyse the trending videos categories to get an overview of people who are interested in any particular domain would help the business in launching their advertisement or videos in that particular category which is being watched by the majority of the population.
* Understanding the time when the videos are being published and identifying the trend when the videos are published during the week and the timings: Analyse the published month to see if holidays and other events affect the trending video statistics.
* Companies and businesses which would rely on this analysis to advertise and sponsor appropriate channels at the right time and improvise their growth and performance which happens due to increase in subscriber count if the video is being reached to wide array of population.

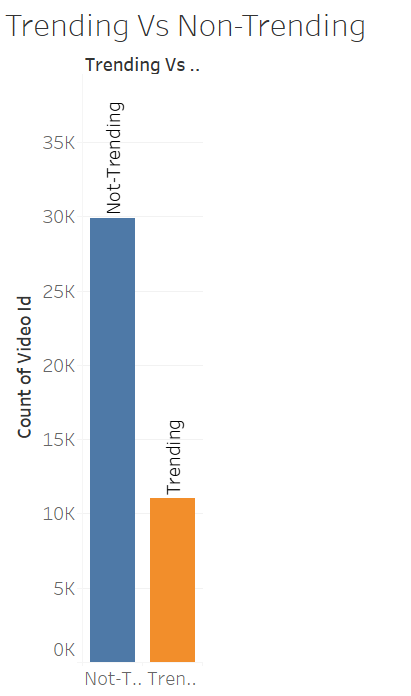
Considering these major objectives, any advertiser and company would rely on this statistical analysis to understand the descriptive and diagnostic information and would be able to predict the growth statistics for future. Analysing YouTube’s performance with analytics would be the most appropriate way in understanding the future prospective of any business or individual.

## 2.1. Trending Videos and Views:



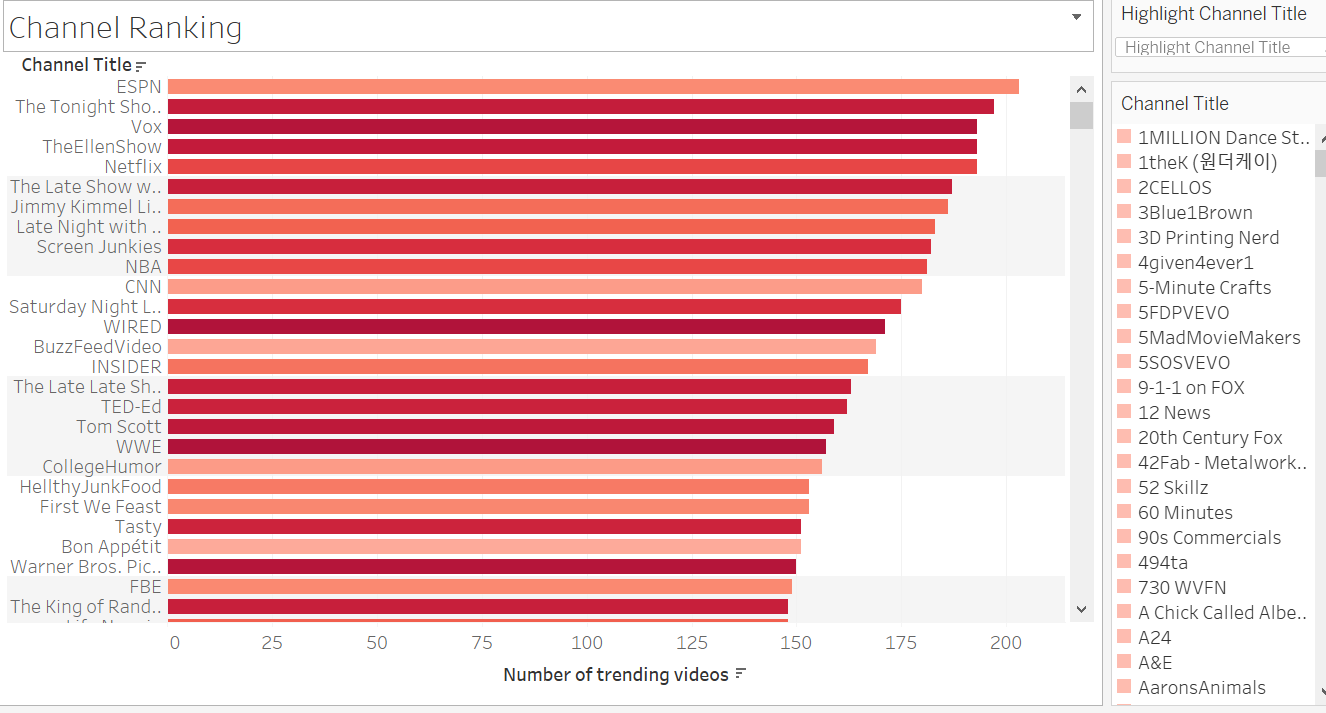
The trending video which has the highest number of views is “Childish Gambino – This is America” with 3,758,488,765 views which has the video ID: VYOjWnS4cMY. The video was published on 5/6/2018 and the video went to trending status on 5/8/2018 which is within two days from the published date. We see that the second most trending video was published on 20th April 2018 and the video went in trending status on 21st April 2018 which is within a day and the video was in trending status consecutively for 20 days. The least trending video has 748 views which was published on 01st Jan 2018 and it went on trending status on 18th Jan 2018. This analysis clearly shows that the trending status of the video is dependent on number of views for the video and number of times the video has been in trending status consecutively which eventually increased the views.

## 2.2. Trending Vs Non-Trending Videos:

We have created a calculated field to understand the trending status of the videos by categorizing the video based on the number of likes which are being received by the video. If the likes are more than 50000, those videos are considered to be in trending status and if not, the videos are in non-trending status when compared with the trending videos.

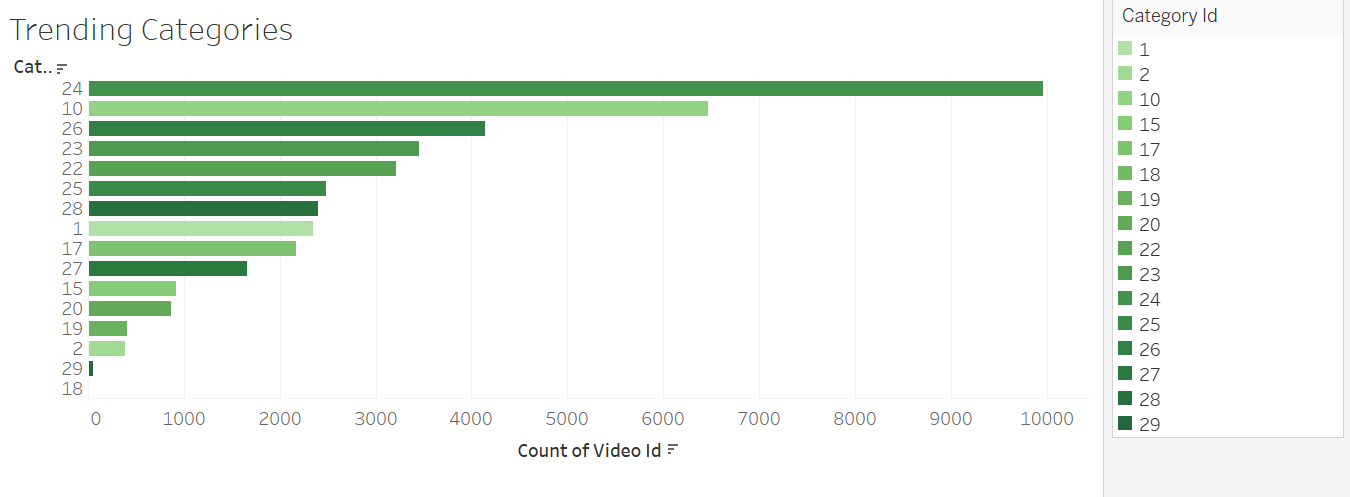
We can conclude that if the video is in trending status with most number of likes, then there are more chances that the video would be shown in trending status consecutively for the future dates as well.

## 2.3. Channel Ranking:



Most number of trending videos has been posted by ESPN with 203 trending videos and the top three channels have been the popular channels which have the greatest number of trending videos. The channel subscribers increase if there are a greater number of trending videos from a specific channel which in turn would increase their chance of being in trending videos most of the times. We see that the channel which has the least number of trending videos is a local channel which has least number of subscribers.

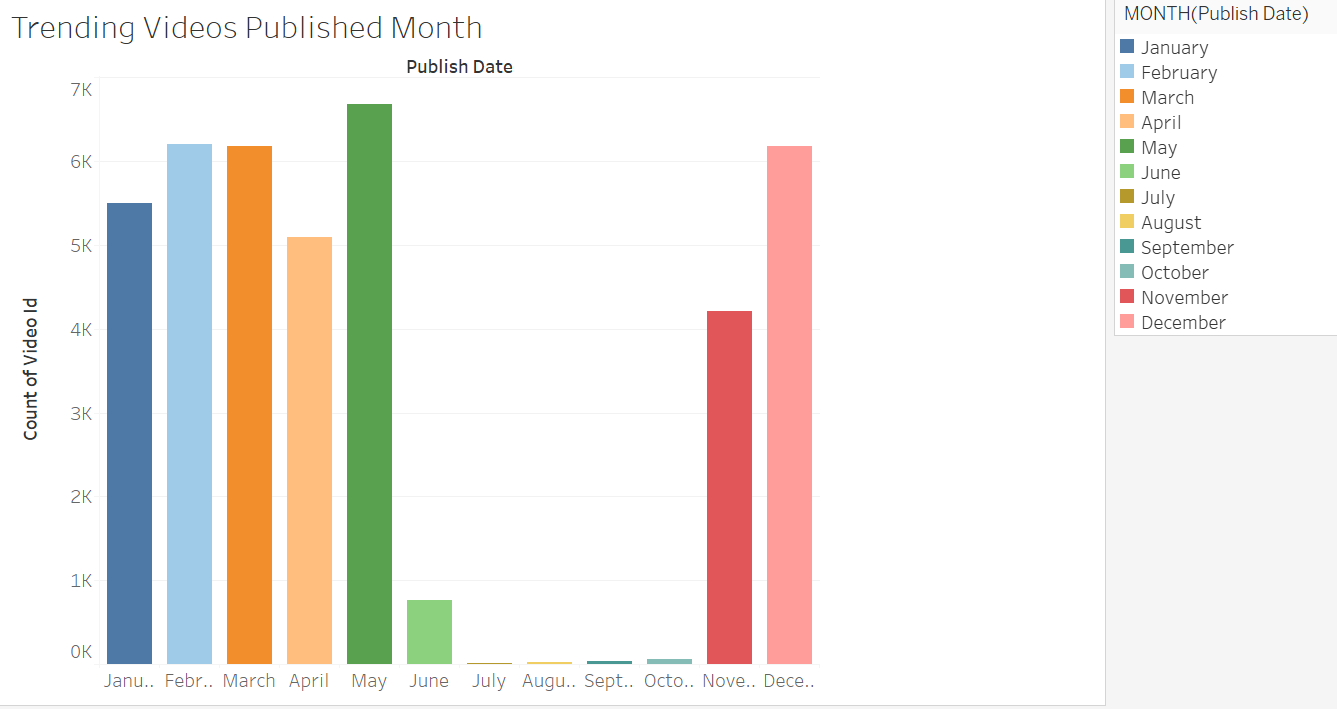
## 2.4. Trending Categories:



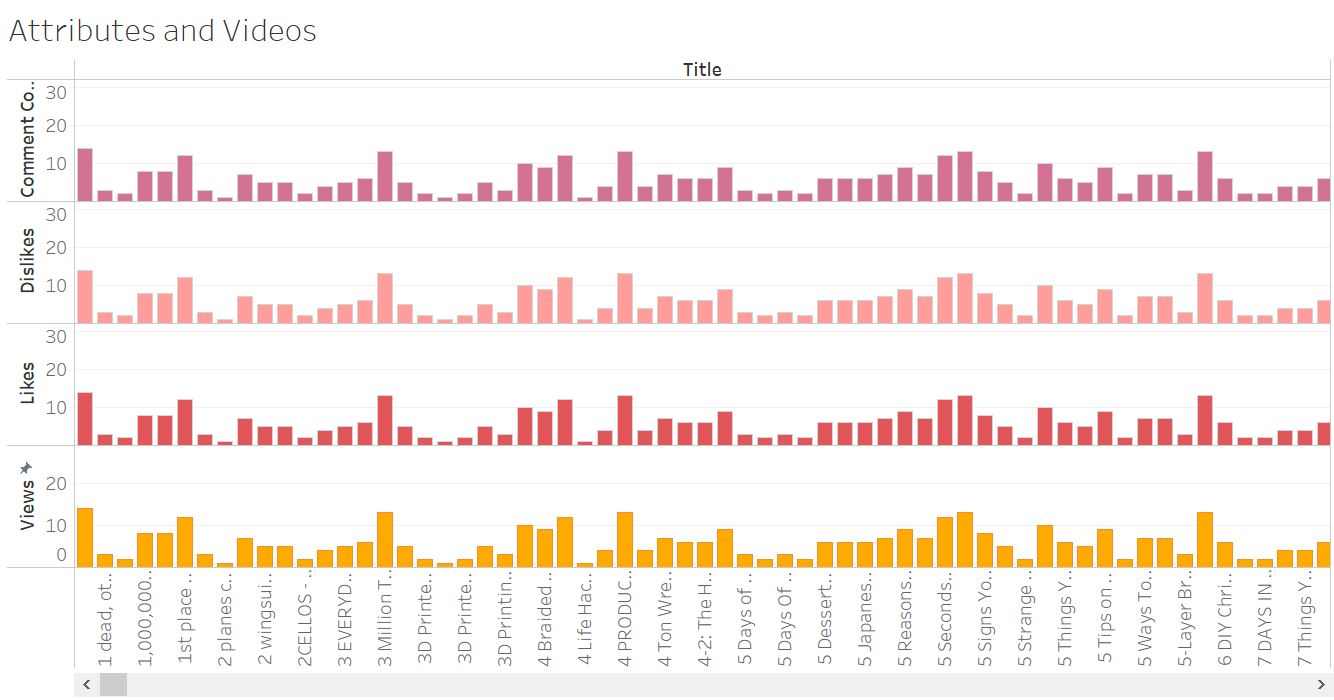
We have defined categories in the data description report wherein each category is given a numeric value. We see that the category id of 24 which is ‘Entertainment’ category has the greatest number of trending videos. We can create targeted videos for the population based on the category which is popular to improvise the video traffic and drive more growth perspectives. Category 18 has the least number of trending videos which is a short movies category.

## 2.5. Trending Videos Published Month:

We have defined the months wherein we see most of the published videos which would help us in identifying specific timeframe for releasing videos to generate maximum traffic. We see that most of the videos which are trending are published during May which has most number of holidays and considered to be a break season wherein most of the population is available. Least traffic of published videos is seen during July, august, September and October which is during the mid-year and most of the population is busy during that timeframe.

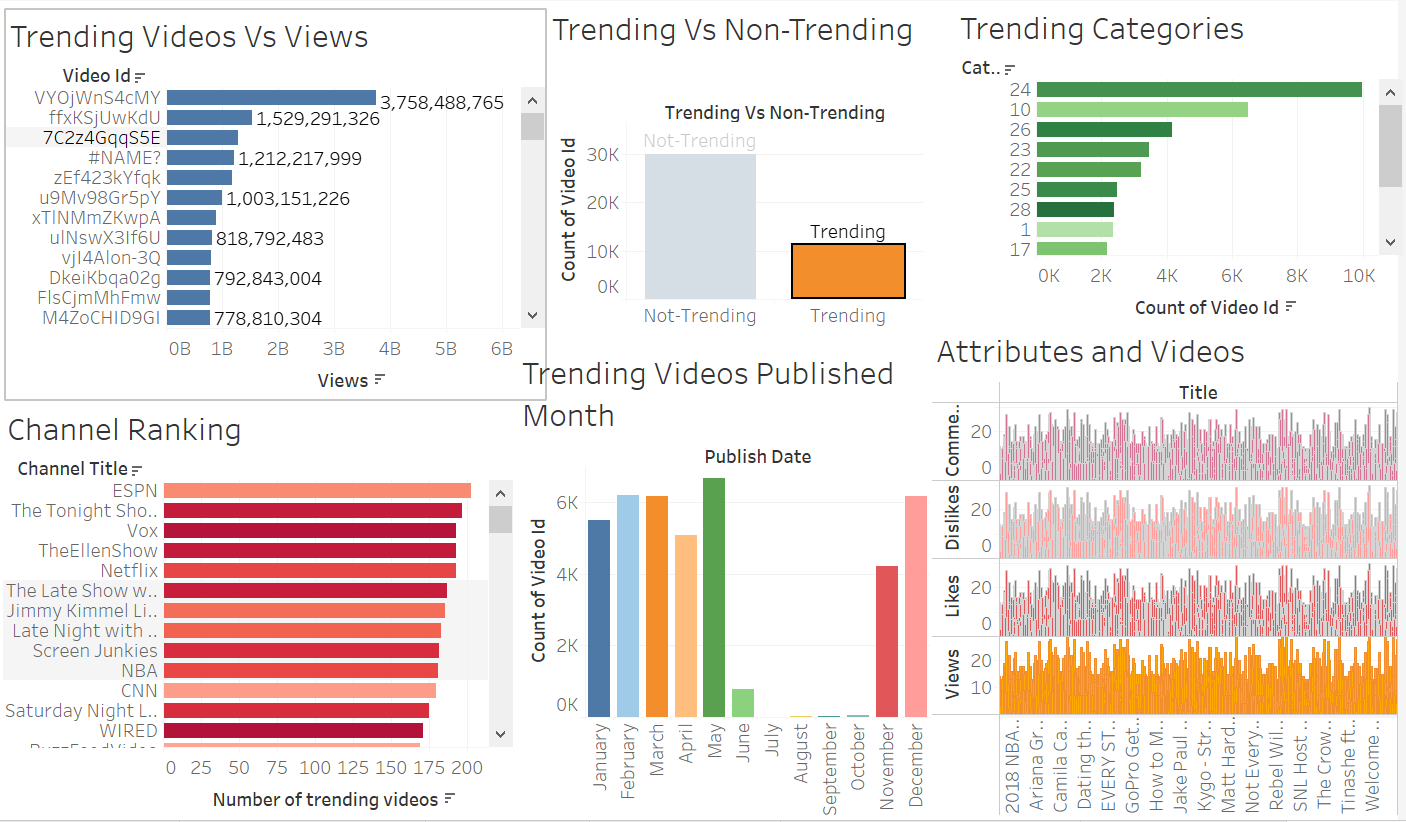


## 2.6. Attributes and Videos:



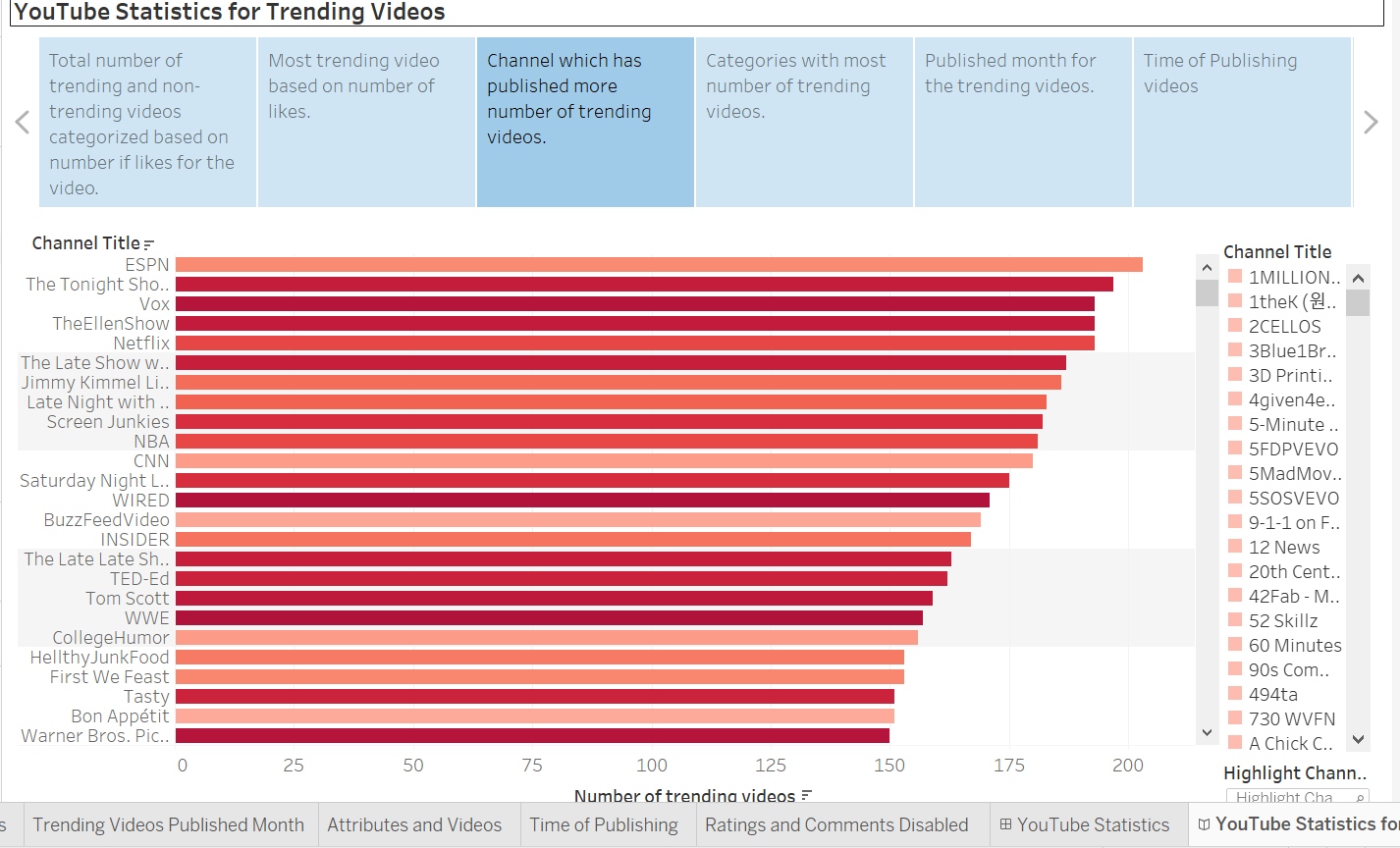
We tried analysing the attributes such as likes, dislikes, views and comment counts which are the most important factors for the video to go in trending status. If we observe closely, we see that all the attributes are parallel for any specific video with very minor difference between the attributes which means that the attributes are correlated. If any video has a greater number of views, that video might have a greater number of likes and comments as well. All these attributes are correlated.

## 2.6.Tableau Dashboard: [Tableau Public](https://public.tableau.com/views/YouTubeStatistics_16393717086050/NumberofViewsfortrendingvideos?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)



# 3. Tableau Storyboard:

Link to Tableau Storyboard: [Tableau Public Storyboard](https://public.tableau.com/views/YouTubeStatisticsStoryboard1/YouTubeStatisticsforTrendingVideos?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)



## 3.1. Problem Statement:

We are analysing the YouTube Statistics for trending videos to gain insights on the attributes which are relevant for the videos to be in trending status. This will help us in analysing the factors which are responsible for the video to be in trending status. Understanding these statistics as an individual content creator would enhance my profile by increasing my subscriber count which eventually would lead to monitory growth.

Another important objective as an advertiser or business entity would be that an advertiser would be able to direct his advertisements to the video which is in trending status to gain maximum profit from the ads. Any business entity would be able to improvise their revenue by reaching out to maximum audience.

## 3.2. Purposes/Questions/Application:

We have created tableau storyboard to understand the factors which are necessary for the identification of business solution for YouTube statistics. Please find the detailed analysis below:

* Designated the trending and non-trending videos based on the number of likes received for a YouTube video. If the likes are less than less than 50K, we have considered them as non-trending video and if they are more than the threshold, they are considered as trending videos.
* We have analysed the most trending videos for the given timeframe and the total number of views received for the video to be in trending status.
* Analysed the channel which publishes a greater number of trending videos and if the channel is a popular channel or not based on the number of subscribers.
* Analysed the most trending videos category to understand the public response to video based on various categories. We see that most of the audience are drawn towards entertainment category.
* Analysed the published month for the videos to be in trending status and we see that most of the videos are being published during start and end of the year.
* Most of the videos are being published during evening time when most of the audience is at leisure hour and accordingly the video is being published.
* Overview of likes, dislikes, views and comments are being analysed to understand the correlation between the attributes and we observe that most of the videos have very near relation with the other attribute.
* We tried analysing the videos which has both comments and ratings being disabled and we see that most of the videos do not have their comments or ratings disabled.

## 3.3. Conclusion:

YouTube being an open platform for an individual and for a business entity has huge amount of information and we can understand its statistics to gain maximum benefit from the platform.

## 3.4. Recommendations:

YouTube should consider the trending videos statistics not only based on the repetitive trending videos; it should consider the statistics on a daily basis. Having repetitive videos would eventually lead to one channel being at the top most of the time which is leading to an unhealthy competition.