Analysis on YouTube Trending Videos

Swati Gayakwad[1], Rajas Patankar[2], Dashrath Mane[3]

[1]PG Student, Vivekanand Education Society's Institute of Technology, Dept. of MCA, Mumbai, India
[2]PG Student, Vivekanand Education Society's Institute of Technology, Dept. of MCA Mumbai, India
[3]Assistant Professor, Vivekanand Education Society's Institute of Technology, Dept. of MCA, Mumbai, India

---***---

Abstract - *Online video streaming platforms are heavily used nowadays. Websites such as YouTube offers content creators a great platform to share their knowledge, ideas and interesting information to their viewers. For a video to reach to maximum people, YouTube offers a trending page on website that shows videos which are trending at that particular time. Other than few viral videos that achieve high view count which are predictable to end up in trending section, rest of the videos cannot be predicted. Corporate companies are using social media for improving their businesses, the data mining and analysis are very important in these days. This paper deals with analysis of YouTube Data on Trending Videos. The analysis is done using user features such as Views, Comments, Likes, and Dislikes. Analysis can be performed using algorithms like Linear Regression, classification and other Machine learning models and python libraries like pandas, matplot library to classify the YouTube Data and obtain useful information.*

*Key Words***: *Data Analysis, YouTube, Trending, Python, Linear Regression, Big data, Data mining, Machine Learning, popular, social media, Views, pandas*

1. INTRODUCTION

YouTube is a widely used and famous online video platform in the world today. YouTube has a list of trending videos that is updated constantly. Analyzing these trending videos can give content creators greater perspective and knowledge for increasing their popularity and brand of their channels. Companies and businesses using social media and digital platform can also use this analysis to boost their growth in business by publishing videos or sponsoring appropriate channels at right time.

This paper can help in finding, measuring, analyzing, and comparing key aspects of YouTube trending videos.

1.1 Objectives

- To understand characteristics that makes a video end up in trending section on YouTube.

- To grasp knowledge about YouTube's algorithm and extract key aspects that can be used to improve growth of YouTube channels.

- Provide content creators and other digital businesses with information about features required to make a video that trends and reaches to maximum people.

- To accelerate a channels growth and increase their subscriber count by making trending videos.

- By increasing subscriber count and views this can help you youtuber's to earn more money, sponsorships and opportunities to excel in their social media carrier

2. METHODS OF ANALYSIS

Analysis for this paper is done by using 3 different ways of analysis.

2.1 Analyzing the Basic Statistics of Trending Section videos.

One of the methods is by obtaining all the data associated with trending videos for example statistics such as number of views for each trending videos, time taken for video to end up in trending section, number of likes, comments and description for that videos etc. These statistics will help in understanding basic conditions for a video to appear in trending section of YouTube page. This data can be extracted by using YouTube API provided by YouTube. It offers lot of services such as downloading video content, description, titles, thumbnails and other basic statistics**.** [1]

2.2 Studying the Difference between Trending and Non Trending Videos.

Comparing the statistics of both, trending and non-trending videos on YouTube can be another method for performing this analysis. By comparing certain attributes such as number of views, likes, comments, shares count for each hour or time interval after the video is uploaded, for both the type of videos, can help in understanding how a particular video navigates through YouTube's algorithm steps and ends up in trending section while other does not appear in trending section. For this type of analysis equal amount of data about trending and non-trending videos will make the results more accurate and form better conclusion. [1]

2.3 Analyzing Youtubers who upload Most Trending Videos.

The third method of analysis is by studying particular youtubers or content creators that post most trending videos in a particular region or in a category. By reviewing their performance statistics, we can also form some assumptions that can provide us results and information about trending videos. This youtubers might have certain practices that they adopt while publishing their videos on YouTube. This study of their behavior can help us in obtaining crucial facts about what interest audience in watching their videos and also about YouTube algorithm that make these videos appear in trending section. [1]

By using these 3 methods of analysis and integrating their results, we can form much more firm conclusions that we can draw, as each of these methods have some pros and cons with their use.

3. ABOUT TRENDING VIDEOS ON YOUTUBE

Trending page in YouTube helps viewers to see what's happening in their region, which event or videos are people watching the most. Videos with wide range of viewers that find the video interesting are shown in Trending section. Like a new movie trailer or new song from known artist, some Trending videos are predictable. Other videos are surprising, and based on time, event and day of that video. Trending displays same list of Trending videos to all users in each country, it does not show any personalized videos. The list of videos in Trending section updates for every 15 minutes. [2] With each update, videos can stay in the same position in the list, Move up or down. Users can't pay to have their videos to shown in trending results that means trending videos are never sponsored.

4. DATASETS AND DATA CLEANING

Dataset used for analysis are of 2017-2018 and 2019. In this analysis, all trending videos for the year of 2017-2018 were analyzed (more than 40,000 videos) and in 2019 (more than 70,000 videos). To check similarity between those videos, have to analyze these data to get perception of trending videos.

The Dataset that we have used is obtained from internet. We will analyze USA trending videos. We also have dataset of 2000 non trending videos for comparing it with 2000 trending videos.
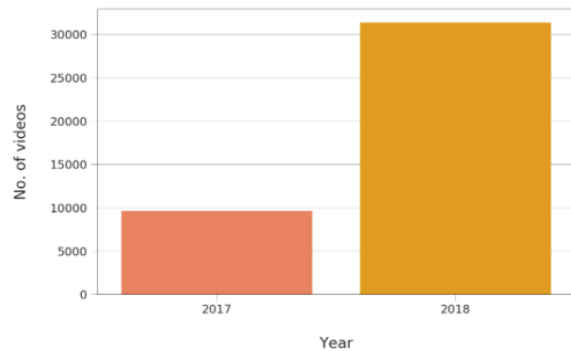

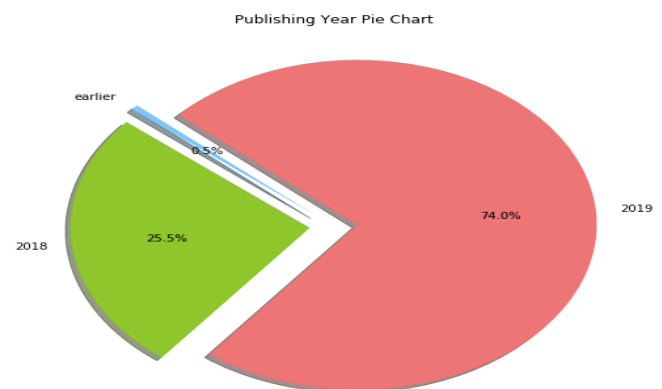**Chart -1**: Analysis of publishing videos in 2017-2018

Chart -2: Analysis of publishing videos in 2018-2019

For any Analysis we need very clean and accurate data that do not have lot of redundancies and missing the value hence it data cleaning is very necessary step in the analysis.

Some python libraries for cleaning and formatting the data are pandas, NumPy, Dora, Datacleaner, Pretty Pandas, and Arrow etc. [3]

Some steps in Data cleaning with pandas are:

- Dropping unwanted columns
- Changing index of a data frames.
- Tidying up fields in data (making values in column of similar type)
- Renaming columns and skipping rows.

5. ALGORITHMS

Some of the algorithms that can be used for analysis are:

5.1 Classification

It is a process of creating classes that represent users and use cases. Class Probability Estimation tries to predict how to classify each single individual data asset. Based on the specific question to be answered, classes are created. Other qualities of the asset are evaluated in making the prediction. [4]

5.2 Regression

The most commonly-used forecasting method is the Regression method. Regression can be confused with classification methods because to predict an outcome, the process of using known values is the same. It is a model used to predict continuous value. Linear regression, Logistic Regression, Polynomial Regression are some types of regression. [4]

5.3 Clustering

Clustering is a technique which involves grouping of data points. We can use this algorithm for classification of each datapoint into a specific group. The datapoints which are grouped will have similar features while those are in different groups will have highly dissimilar features. We need to identify how much data is linearly correlated to each other hence to find correlation some analysis is required. [4]

5.4 Similarity Matching

Similarity Matching looks for correlation of attributes in order to recognize similarities between individuals. If two customers or products are similar in certain ways, it's reasonable to predict that they will be similar in other ways as well. This can be used to find customers for targeted marketing campaigns or for managing the company's image with targeted online ads. The similarity profile will reflect characteristics that will have a bearing on the question at hand, so it could include attributes such as age and purchase history.

6. ANALYSIS ON VIDEO LENGTH

With the help of python and other libraries we can easily count the string length of each title for trending videos and estimate what is an average title length for a trending video. [5]

Below graph shows title character length for each count of videos, where most videos have title length between 30 and 60 character approx. Different graphs such as bar graph and scatter plot can be used to understand relationship between title length and view.
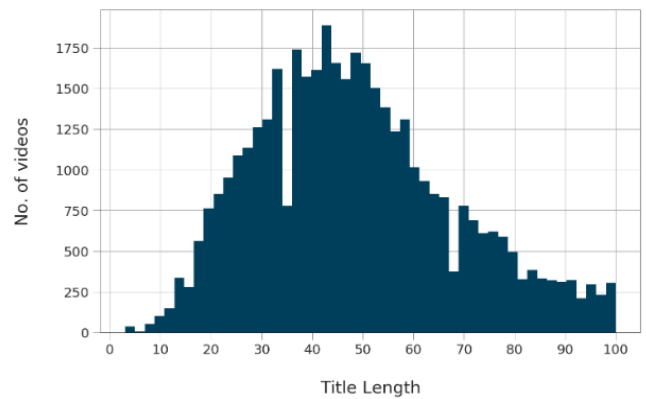
Chart -3: Title character length for each count of videos in bar graph
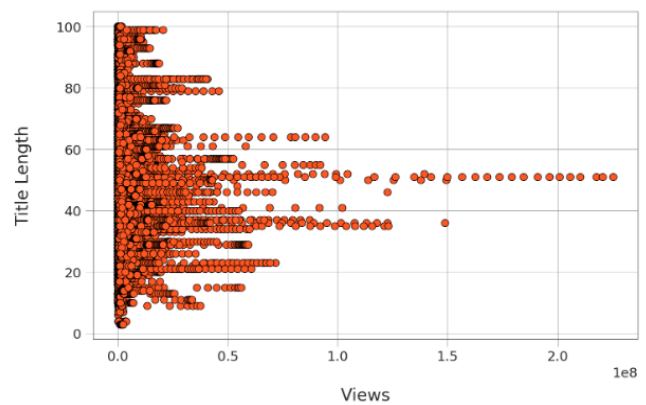
Chart -4: Title length character for each view count in scatter plot

7. ANALYSIS ON CHANNELS

Just by counting channels and trending videos number we can get which channels produce most trending videos. This can be done using the using functions from panda's libraries on data frame containing the number of times the channel was on trending section.

Sum of count for each channel is calculated and then plotted in graph with the help of python's matplot library. [5]

Chart -5: Shows the channels which have largest number of trending videos

8. ANALYSIS ON CATEGORY

Machine learning algorithms such as classification can be used to classify the categories of videos with their number of trending videos count to understand which category has more trending videos. [5]

Chart -6: Shows the channels which have largest number of trending videos

9. ANALYSIS ON KEYWORDS

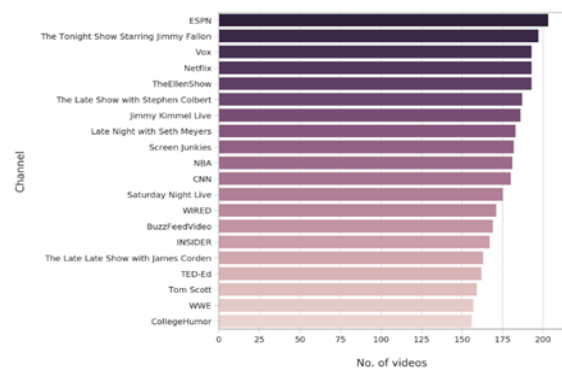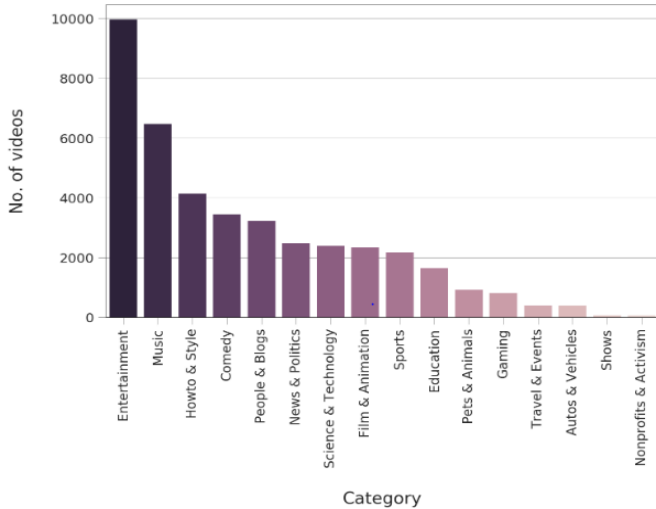With the help of python and pandas library we can get the most common words in video titles after removing the usual structuring words such as articles, prepositions and combining words such as a, an, the, but etc.

Ignoring words like "the" and "of", we observe that the symbols "|" and "-" occurred several times in trending video titles 11452 to 10663 times. The common words in Trending video titles we noticed are "Trailer", "How","2018" and "Videos" each occurred in 1613-1901 video titles.

Below diagram represents a word cloud to visualize most common words in titles of our Trending Videos. The more common the words is, will be bigger in font size. This can be done using python libraries such as matplotlib and word count library.

Fig -1: Shows most used keywords in trending videos

10. COMMENTS, LIKES AND RATINGS

Correlation represents as value between +1, 0 and -1 where +1 denotes highly positive correlation, -1 denotes highly negative correlation and 0 denotes No correlation. The highly positively correlated which has correlation value as-Between Views and Likes: 0.85, Likes and Comments: 0.80, Dislikes and Comments: 0.70.There is some positive correlation between Views and Dislikes, Views and Comments, Likes and Dislikes. Some Results obtained by observing comments like ratings are we see only 2% of Trending Videos prevented users from commenting. We see that only 169 trending videos out of 40949 prevented users from commenting. Videos have both comments and ratings disabled. So there are just 106 trending videos that have both comments and ratings disabled.

11. ANALYSIS ON BEST TIME TO UPLOAD VIDEOS

By extracting publishing time for every video with videos count for every time slot we can understand best time for a trending video to get uploaded and published. Hence by using algorithm like regression we can predict that a video published in a given time slot having certain set of features can trend or not.

Some summaries after doing analysis on day and time are the Best day and time to upload video on YouTube is in midweek afternoons. Best hours are: On Monday, Tuesday and Wednesday from 2pm to 4pm and on Thursday and Friday from 12 pm till 3 pm. After collecting data from 120 million views of total 1300 channels, we can know on weekends i.e Saturday and Sunday the best time to upload videos is between 9 to 11 am. To post videos on YouTube from all the days of week, Thursday and Friday are best. Most YouTube viewers are children's or adults. The viewership extends on Thursday, increases on Saturday and then decreases on Sunday afternoon. In weekdays, the best time to upload videos on YouTube is between 2pm and 4pm.You may think it's early but YouTube traffic is high in the evening between 7pm and 10pm.

However, the reason for uploading videos earlier is just as Google indexing blog content. For videos, YouTube does the same, allow your video to be indexed and show up in searched results by uploading your videos a few hours earlier. You need to post your YouTube videos between 10 to 11am because in the afternoon many people are active on YouTube. To know only best time to upload YouTube videos is not enough to make millions of views for your videos.

Chart -7: Trending videos and their time

Chart -8: Number of Videos published per hour

12. ANALYSIS ON VIDEO TAGS

When you upload a video to YouTube, you can add tags to the video. Tags cannot display on the video page but you can see tags by viewing the source code of the page or using browser extensions such as, Tubebuddy and vidIO.

- Trending videos has 21 tags as of average.
- Only 3.5% of trending videos have no tags.
- The minimum number of tags a trending video has is 0.

Once again by using word count library and matplotlib we can create an image consisting of most number of words used in these tags.

Fig -2: Common words in Trending Video Tags

13. ANALYSIS ON VIDEO THUMBNAILS

Apart from video title's, thumbnails plays a very important role in making user to click on video to watch based on the contents in the thumbnails. Using object detection libraries with python such as opencv, imageAIi, tensorflow we can identify objects in thumbnail by estimating their percentage probability and then counting the probabilities for each object such as people, car, Text we can understand what are the best items that can be shown as a thumbnail for a video.

Chart -9: Best items that can be shown as a Thumbnail for a Video.

So a person image appears more than 1,000 times in trending videos thumbnails.

14. VIDEO DURATION

By comparing video duration and view count of trending videos we can understand what the best duration for a trending video is has. Trending videos between 2 to 20 minutes long are highest in rank than videos which are 2 minutes long on average. [6] If they use storytelling

techniques to maintain and build interest throughout the video – generally rank higher than very short videos. So, video duration influences rank ability, but so does the ability to tell a good story. (Study found that 10-16 minute videos are well liked.

15. INTERPRETATION OF SOME KEY RESULTS OF ANALYSIS.

Table -1: Analysis for year 2017-2018

Characteristics	2017-2018
Dataset	It consists data more than 40,000 trending videos.
Video Title Length	Title length between 30 and 60 characters
Largest Category	Entertainment
Largest Channel	Largest Channel is ESPN
Common words in Video Titles	Official, Video, Trailer, How,2018
Views, Likes, Comments	Trending videos of 71% has views less than 1.5 million .Trending videos of 68% has likes less than 40,000.Trending videos of 67% has comments less than 4,000.
Published Day and Time	Best hours are: On Monday, Tuesday and Wednesday from 2pm to 4pm and on Thursday and Friday from 12 pm till 3 pm.
Video appeared on the Trending list	On trending list videos may appear for more than one day.
Correlation of Trending videos	The highly positively correlated which has correlation value as- • Views and Likes: 0.85 • Between Likes and Comments: 0.80 Between
Thumbnails	Not Analyze

Table -2: Analysis for year 2019

Characteristics	2019-2020
Dataset	It consists data more than 70,000 trending videos in dataset.
Video Title Length	Title length between 36 and 64 characters
Largest Category	Entertainment
Largest Channel	Technology (Linus Tech Tips)
Common words in Video Titles	Official,Video,2019,vs,Trailer,music, Game, new, highlights, first, challenge.
Views, Likes, Comments	The video which is most liked ,viewed and commented was BTS song 'Boy

	With Luv'.
Published Day and Time	Publishing trending videos hours were between 4 PM and 6 PM.
Video appeared on the Trending list	And on average, videos takes 1.5 days to become trending. On trending list, 97% of the videos appear for 2 days after publishing video.
Correlation of Trending videos	The highly positively correlated which has correlation value as- • Views and Likes: 0.81 • Comments and Views: 0.61
Thumbnails	A person is the most common object in the thumbnails (Using Object-Detection algorithm).

16. GOOD PRACTICES WHILE PUBLISHING VIDOES ON YOUTUBE.

Some tips that can be obtained after performing the analysis and summary of analysis are:

• Optimize Meta Data for Better Search and Visibility:

Some of the things that algorithm considers are you need to ensure your video title, description & content matches to what user is searching for, to become a high rank in YouTube video searching.[7]

• Use relevant keywords:

Try to make a video by adding clear and understandable keywords to your title and description. So user could be able to read and click it to watch full video.

• Use an Engaging Thumbnail:

Use a custom thumbnail that is peculiar only to your channel so that users can see there is a difference from other auto-generated ones [7].

• Video Translation:

Essential to do video translation by providing subtitles along with the video or some textual information that can help the viewers to understand if a particular individual cannot understand or hear the language spoken in the video. Different people have different rate of absorbing content hence providing some form of textual aide can help the viewer retain more information and appreciate the video. [8]

• Call included for action:

At the beginning of the video or while concluding it is good practice to encourage audience to like, share and subscribe the videos. Also asking users to post comments about what

they feel about the video can increase the audience engagement. This in turn can make the video have higher chances of appearing in trending list, as more people are talking about it. [8]

- Use YouTube Analytics:

YouTube offers platform like YouTube studio to content creators that includes analytics section page from which a particular user can obtain overall statistics and performance of his/her own videos. This statistics can give insights to the youtuber about viewer's behaviors and certain patterns. For example a youtuber can know which age group of audience is watching his/her videos, at what time they watch and from which region or country the viewers are in more numbers. This platform can help users by making content for these specific users. [8]

- Be a consistent Youtuber:

YouTube's algorithm favors channel promotion over single video promotion. By consistently uploading videos and having constant interaction with viewers can increase views and subscriber count of channel and improve the reach of video there by increasing the chances of the video to end up in trending page. [7]

- All Clicks Are NOT Created Equal:

"Long clicks" are more valuable than "short clicks. "Short clicks are called when if a viewer watches your video for only 8 seconds and ends his or her session, then YouTube's algorithm figures that he or she wasn't satisfied with your content. Long click means, if a viewer watches your video for 2 minutes or longer and continues to watch more videos during that session, then YouTube's algorithm figures that he or she was satisfied with your content. [6]

17. CONCLUSIONS

In this paper, we presented our findings for measuring, analyzing, and comparing key aspects of YouTube trending videos. To know only best time to upload a video on YouTube is not enough to generate a millions of views for your Videos to become trend. There are some other factors to considered are Good Titles, Good thumbnails, Video SEO, proper tagging, and the number of subscribers are all factors that is a key generating views for your content. Understanding this Statistics will not only help YouTube to develop better algorithms to process videos but also benefit to make decisions for individual youtubers.

REFERENCES

[1] "(PDF) Trending Videos: Measurement and Analysis." https://www.researchgate.net/publication/266262149_Trending_Videos_Measurement_and_Analysis (accessed Apr. 03, 2020).

[2] "YouTube Trends (GUIDE) ▷ Grow Channel ⬚ Trending Videos." https://keywordtool.io/blog/youtube-trends/ (accessed May 12, 2020).

[3] "Pythonic Data Cleaning with Pandas and NumPy – Real Python." https://realpython.com/python-data-cleaning-numpy-pandas/ (accessed Jul. 27, 2020).

[4] "The Top 10 Machine Learning Algorithms for ML Beginners." https://www.dataquest.io/blog/top-10-machine-learning-algorithms-for-beginners/ (accessed Aug. 27, 2020).

[5] "YouTube Trending Videos Analysis | Kaggle." https://www.kaggle.com/ammar111/youtube-trending-videos-analysis (accessed May 12, 2020).

[6] "YouTube Algorithm: 7 Key Findings You Must Know." https://www.searchenginejournal.com/youtube-algorithm-findings/296291/#close (accessed Jul. 27, 2020).

[7] "6 Tips to Make Your Video Trending on YouTube | The Manifest." https://themanifest.com/digital-marketing/resources/tips-to-make-video-trending-youtube (accessed Jul. 27, 2020).

[8] "How the YouTube Algorithm Works in 2020 – Authors' Guilds." https://www.authorsguilds.com/how-youtube-algorithm-works/ (accessed Jul. 27, 2020).