1. Abstraction -

YouTube has become a prominent media that has tons of video content in different categories. YouTube is now pulling in over 1.8 billion users every day. YouTube having thousands of trending videos in different countries every day. a lot of people are making a good amount of income from their YouTube channels. Since it became a highly influencing social media, people are using this platform as a powerful tool for sharing their thoughts and many more. Here we are analyzing US trending videos dataset in order to find correlations between trending videos’ properties. By doing this we can come up with some common features in all YouTube trending videos in the US, using this output information from our analysis we can give some meaningful suggestions to the people who are uploading videos on YouTube. We believe by applying our findings, a YouTuber can increase the view count of the video.

1. Keywords. You Tube, Analysis, Correlation
2. Introduction

YouTube was launched in 2005 with the purpose of help people and share videos with a global audience [. It was founded by Chad Hurley, Steve Chen, and Jawed Karim, who ran the company from a (de rigueur) small office above a small restaurant in a small Californian city. Afterwards it became as the world's most popular online video site, and almost 5 billion videos are watched now on YouTube every single day. Via YouTube platform, people started to create a video-sharing website on which users could upload, share, and view videos. As a result of this sharing video content has become a cultural phenomenon in current world between kids, teenagers and so and so. And latest statistics show that the traffic to or from YouTube accounts for over 20% of the total web traffic and 10% of the whole Internet traffic . As reported by Alexa, the web traffic monitoring service owned by Amazon, YouTube is the second most popular website globally with over 300 hours of videos uploaded every minute and 5 billion videos watched every single day [3]. Having started in 2005, YouTube has well developed into a leading online video-sharing destination. The millions of video clips on YouTube represent a wide range spectrum of user interests including those of educators, scholars and researchers.

Each You Tube video is belong many specific attributes like title, publish date, channel name, tags etc. and each video is categorized in a specific category which is make easy to find and watch for audience of their desires and aspects. In 2009 there was a research which was conducted to analyze the online video viewership of the US Internet users. That study could find that 38% watches educational videos, 50% of adults in the US tend to watch funny videos, 32% watch TV shows or movies, and 20% watch political videos . In one hand YouTube is one of perfect website for audience to do everything as they wish. In the other hand it is one of business paradise for those who have different kind of skills and seeking to do new things to world. At present most people able to use You Tube as their main earning path or business and even wherever you are in the world anybody is able to create your own YouTube channel and can become a content creator. It is one of outstanding advantage offered to skilled persons by You Tube. In addition to the ease of uploading nearly any kind of video content, viewers or the audience are able to interact with the video content by liking or disliking a video, commenting on a video, commenting on a comment, liking or disliking a comment, or posting a video response. Comments on the videos can be used to understand audience’s reactions to important issues or toward videos. Audience’s feedback is called comments and can be used to mine implicit knowledge about viewers, regions, videos’ content, categories, and community interests .As well as You Tube has enabled feature called hash tags and it is one of quick method that can reach out relevant views fast.it makes a video more popular in between worldwide viewers and finally they become as trending videos while making more views and watch hours.

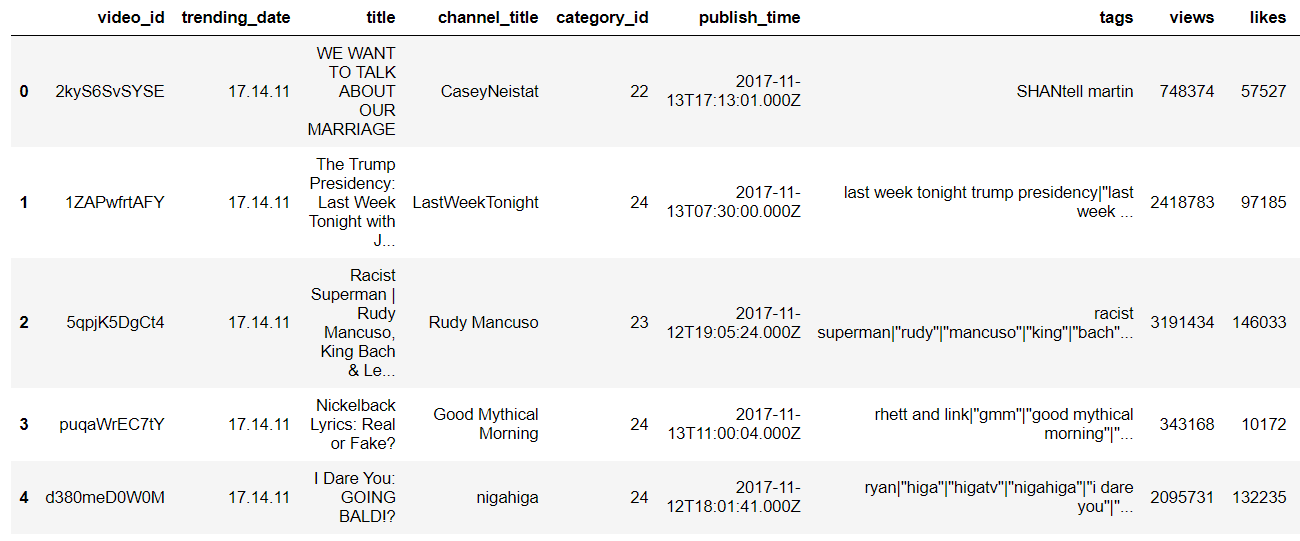
1. Motivation

In this paper we have focused to figure out statistics of You Tube trending videos in US. And we have made data analyzing process to identify how to make a video as a trending You Tube video and how trending video depends on likes, dislikes, comments and user views. Here as the data source USvideos.csv data set has been taken from Kaggle and it consists 40950 rows and 16 columns regarding You Tube videos. Let see our data set and Video id, Trending date, title, channel, category, publish date, tags, views, likes, dislikes, views, comments, thumbnail link, comments disabled, ratings disabled, video error or removed and description are the column names in our data set.at initial stage it consist duplicate values, null values and need to clean before start to analyzing process. For better analyzing it is required each video category id and its relevant name.

During the analysis we identified correlation between views, likes, dislikes and comments. Later it has well illustrated how the relationships behave between above mentioned attributes for each You tube video. Afterwards analyzing results, we have made diagrams and graphs to emphasize our goal and get final decision and predictions. This research will help to convey the idea for those who are seeking to be youtuber or those who are interesting in You Tube how to make your own You Tube video as a one of leading trending Video in You Tube. The results which have been generated by us via this analyze can be used for business planning and strategy by people who do focus You Tube as high commercial manner.

1. Data processing analysing -

We selected a well-structured dataset for this research from keggle, Our dataset has been titled as Trending Youtube Videos there are hundreds of videos which was trending during that particular day and also they have separated these datasets with different regions for analysis purpose we have selected US youtube trending videos dataset so that we can be focused on a single region. When we are looking at the dataset we only have data from 2017 November to 2018 June unfortunately. Apart from this, the dataset contains 40950 rows and 16 columns. Columns names are video\_id, trendin\_date, title, channel\_title, category\_id, publish\_time, tags, views, likes, dislikes, comments\_count, thumbnail\_link, comments\_disabled, ratings\_disabled, video\_error\_or\_removed, descriptions. Each column contains different types of data, some of the columns’ data we are using for our analysis purpose some of them are not.



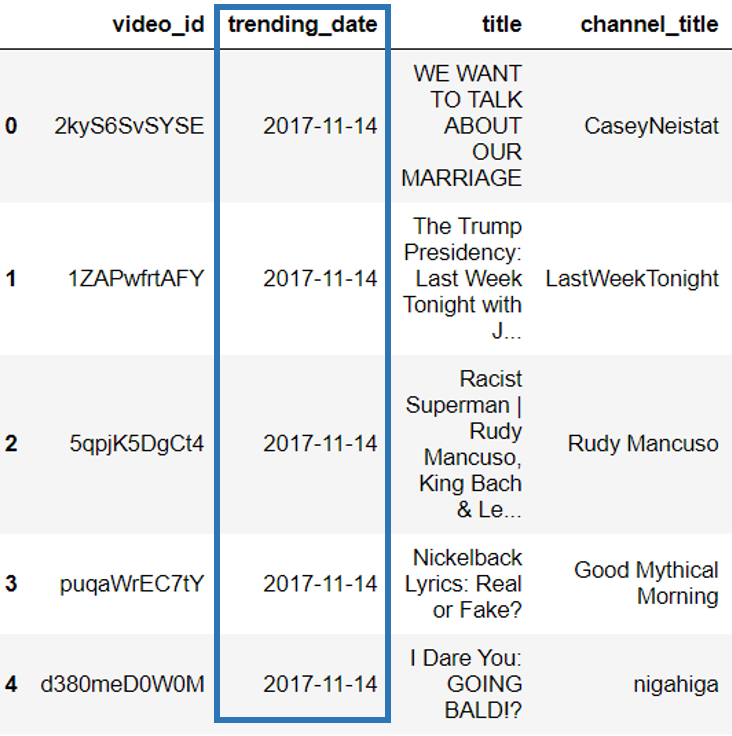


This is our dataset in fig- showing. In this dataset, we have identified some issues here we listing those issues

1. Trending\_date column contains a date which not well-formatted
2. Category\_id column contains different category id but not the category name
3. We don’t need some of these columns for our analysis
4. We need some extra columns of data for further analysis.

We cleaned our dataset in order to overcome these issues we have mentioned above

1. We have to format our trending date column as a pandas DateTime formate. For this, we have used *pandas.to\_datetime()*, here our output shows in fig-



1. Category\_id column only contains videos’ category ids only for analysis purpose we need category name and we have to map that name with this id all over the dataset. For this purpose, we have created a new CSV file which contains all the category then after we have mapped that file with our dataset.

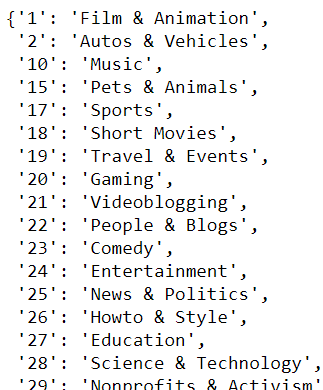
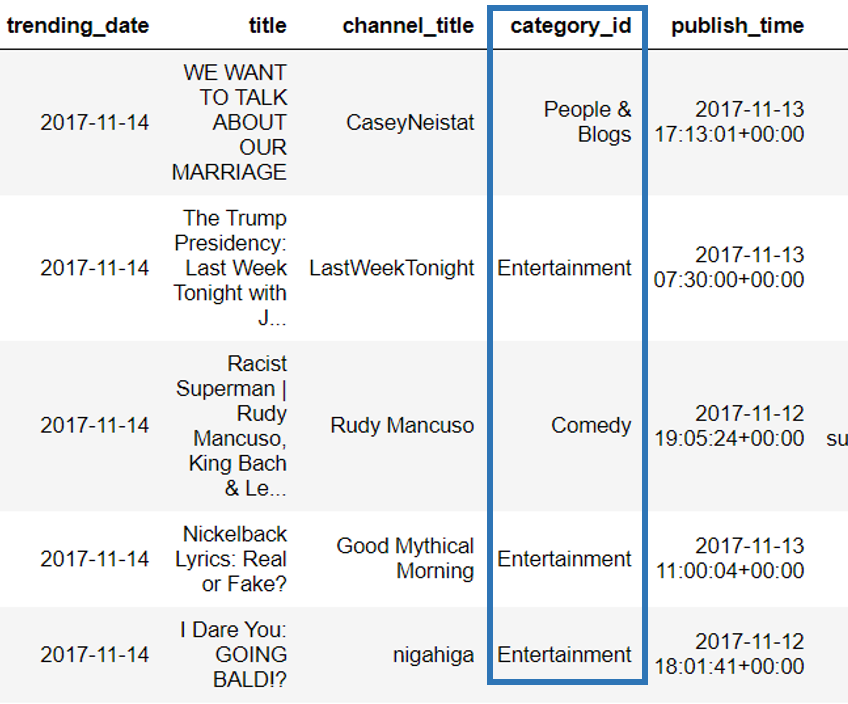
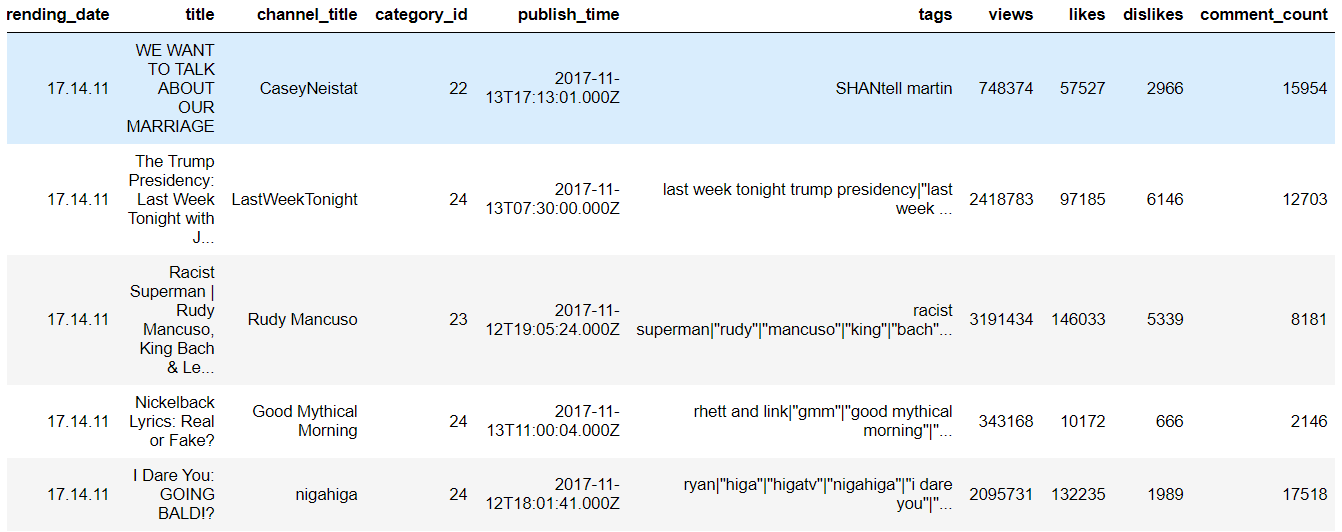


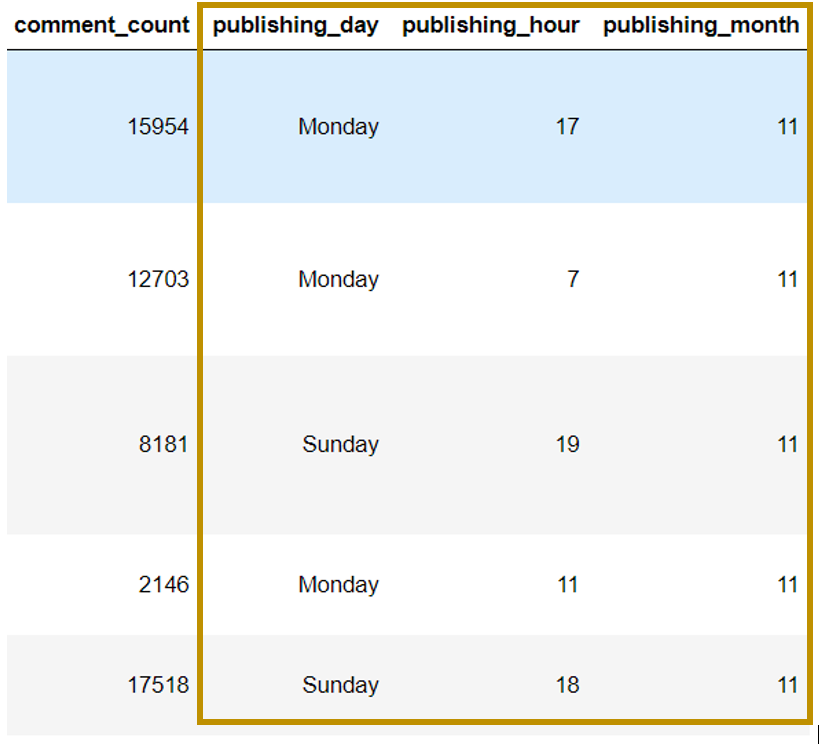
fig - shows our category id and name dataset after we mapped with our dataset it show like this in fig-



1. Finally, we have removed unwanted columns which are not needed for our analysis purpose, these the list of columns we have removed thumbnail\_link, comments\_disabled, ratings\_disabled, video\_error\_or\_removed, description. But we didn’t remove video\_id, title, channel\_title, tags columns for our future analysis purpose. After removing all these the dataset showing like this in fig-



1. Apart from these columns which already came up with dataset we have created some new columns with other existing data for our further analysis purpose, these columns have been created whenever they needed for the particular analysis purpose. These are columns we have created publishing\_day, publishing\_hour and pulishing\_month. Here in this fig- shows those created columns.



1. Methodology - I am doing

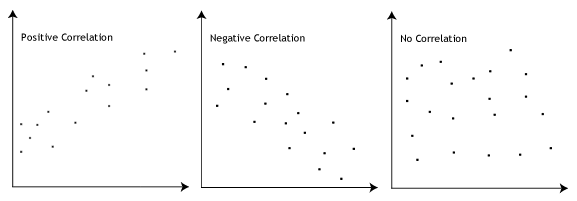
* correlations / let’s take any 2 variables
* Descriptive analysis
* Predictive analysis

In this paper, we are conducting a study on trending YouTube videos’ properties in order to find some meaningful insights.

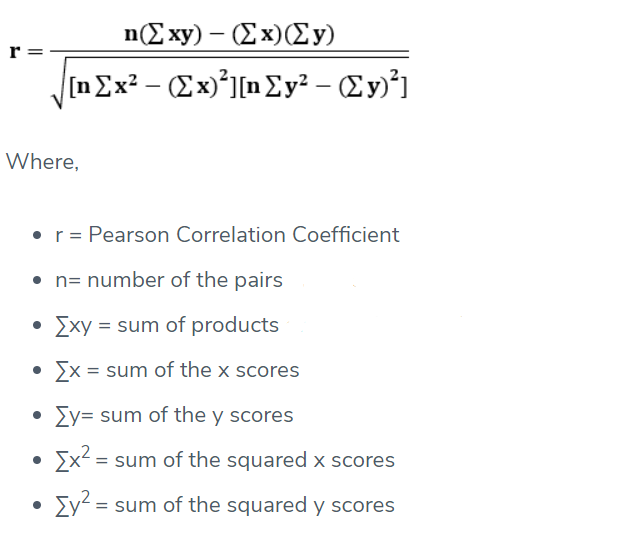
We have used the Pearson Correlation Coefficient for finding the correlations between videos’ views, like, dislike and comments\_cout.

What is the Pearson Correlation Coefficient

Pearson Correlation coefficient used to find the linear correlations or association between two variables, it can take a range of values from +1 to -1 if it’s 0 then there is no relationship between them. A value greater than 0 indicates a positive association and a value less than 0 indicates a negative association. It indicates the strength of the relationship.



Here we are finding Pearson correlations coefficient between these views, like, dislike and comments\_cout property of training videos so that we can identify an impact on one variable in another using this below formula.

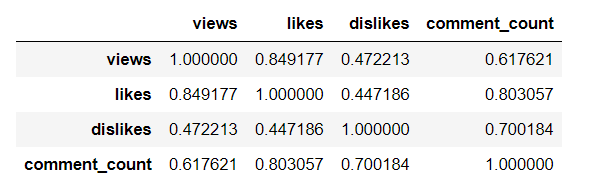


Apart from this correlations coefficient analysis, we can have some different analysis as well for finding more insights so this purpose.

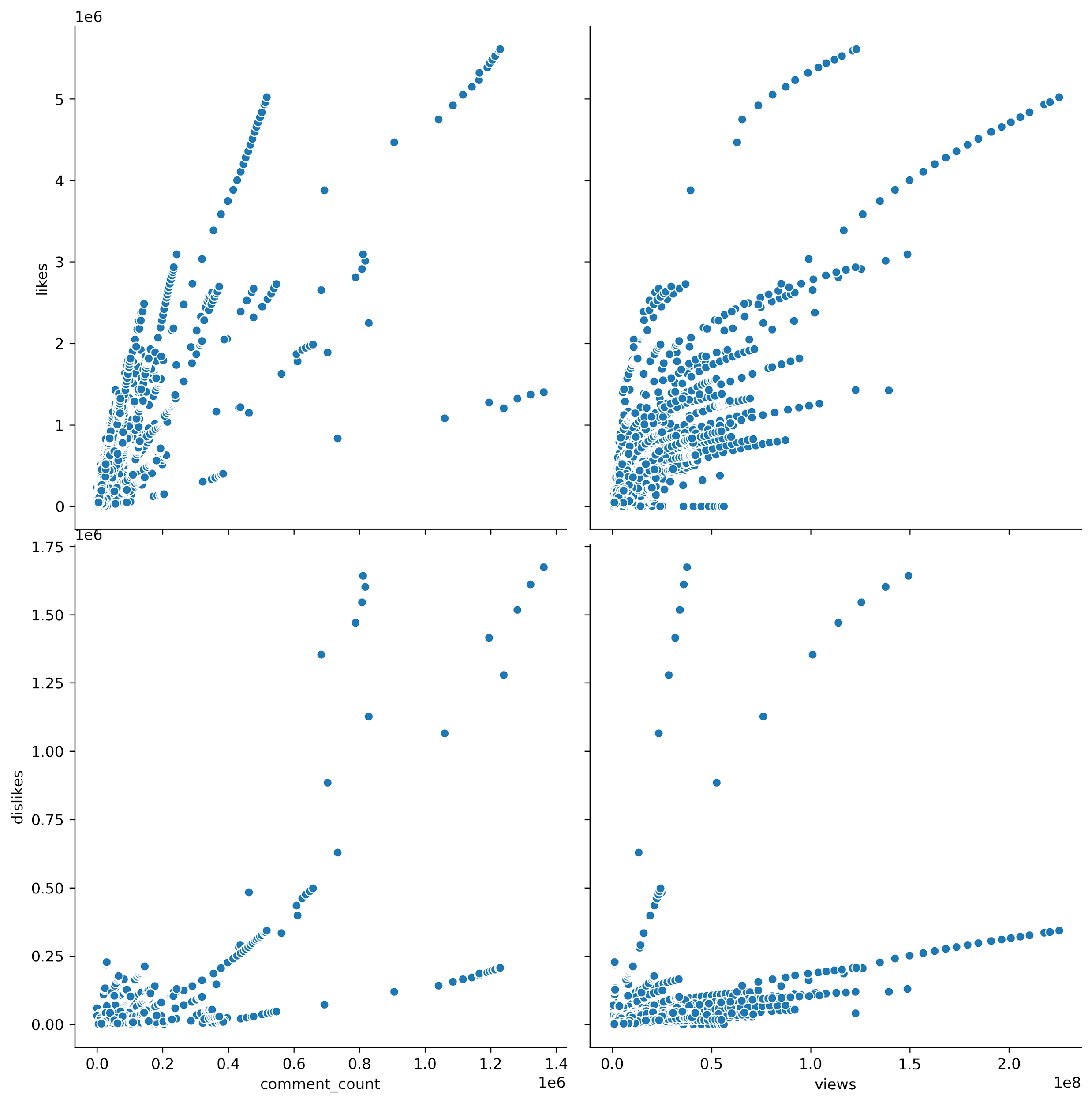
* trending videos’ category with its views and likes
* Which day of the week most of the trending videos had been uploaded
* Which time in a day most of the trending videos had been uploaded

1. Results and Conclusion

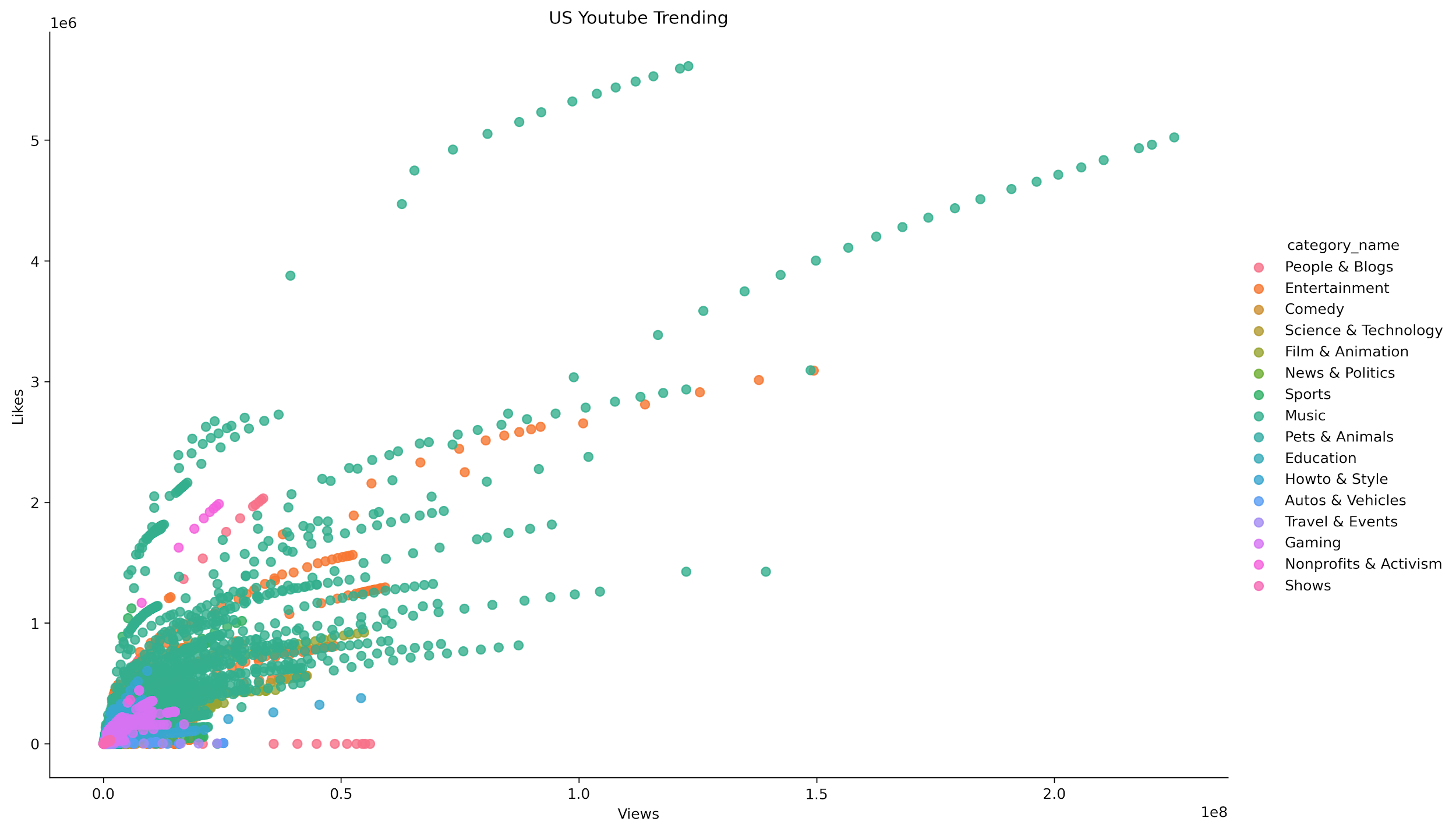
We found the Pearson correlations coefficient between columns views, like, dislike and comments\_cout here our output shows in fig -

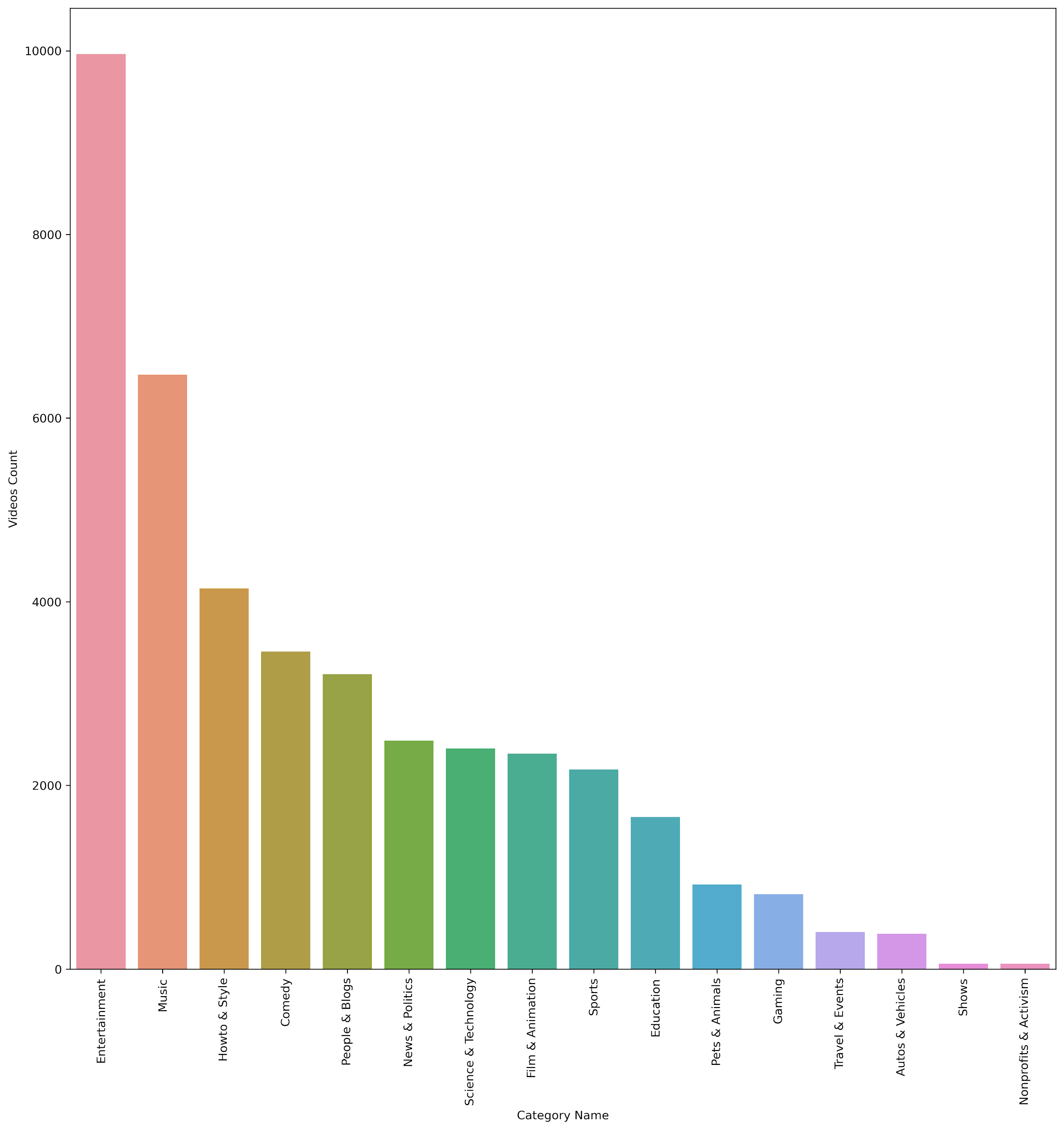


From the above table, views-likes showing the highest Pearson correlations coefficient value, after this views-comments\_count showing second-highest value but views-dislikes showing the least value. So that we can come up with the conclusion likes column having the most impact on views column apart from the other two properties of the video. This result giving a clear view that if any video becomes a trending it mostly depends on how people like that video. If any video satisfying audience then it will become a trending one. Apart from this finding, we can come up with one more, If we look at comment\_count-likes and comment\_count-dislike we can identify both are having a nearly same value it’s indicating that whether people like it or not they are giving their opinion. Apart from this above table, we can come up with this below pairplot.



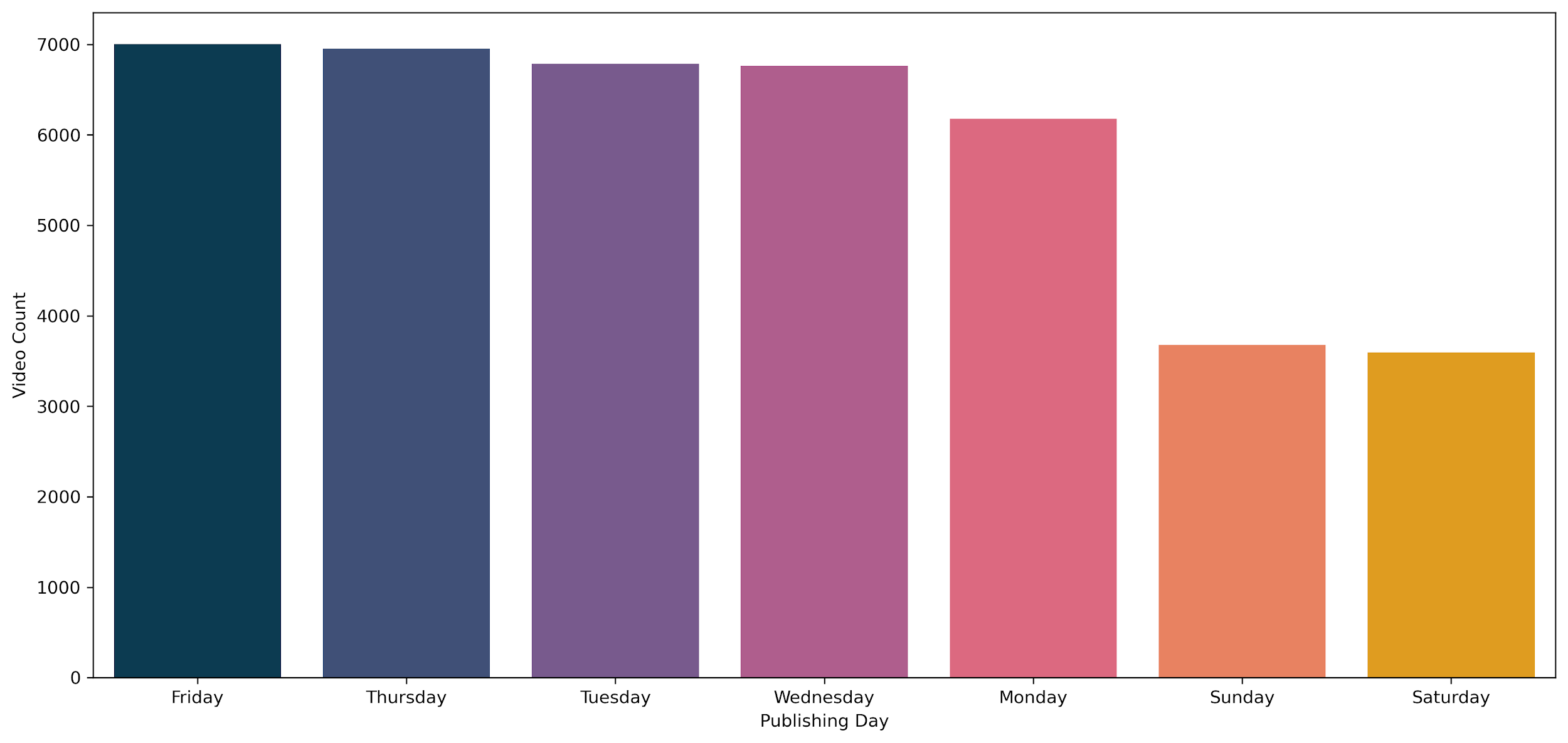
We have analysis the video category with views and likes so that we can get some more details with videos category which category video gets more views and likes we plotted these details in this below diagram.

1. 
2. Above plot, we can see that most people viewed and liked entertainment videos. To compare between different video category we came up with this below chart.

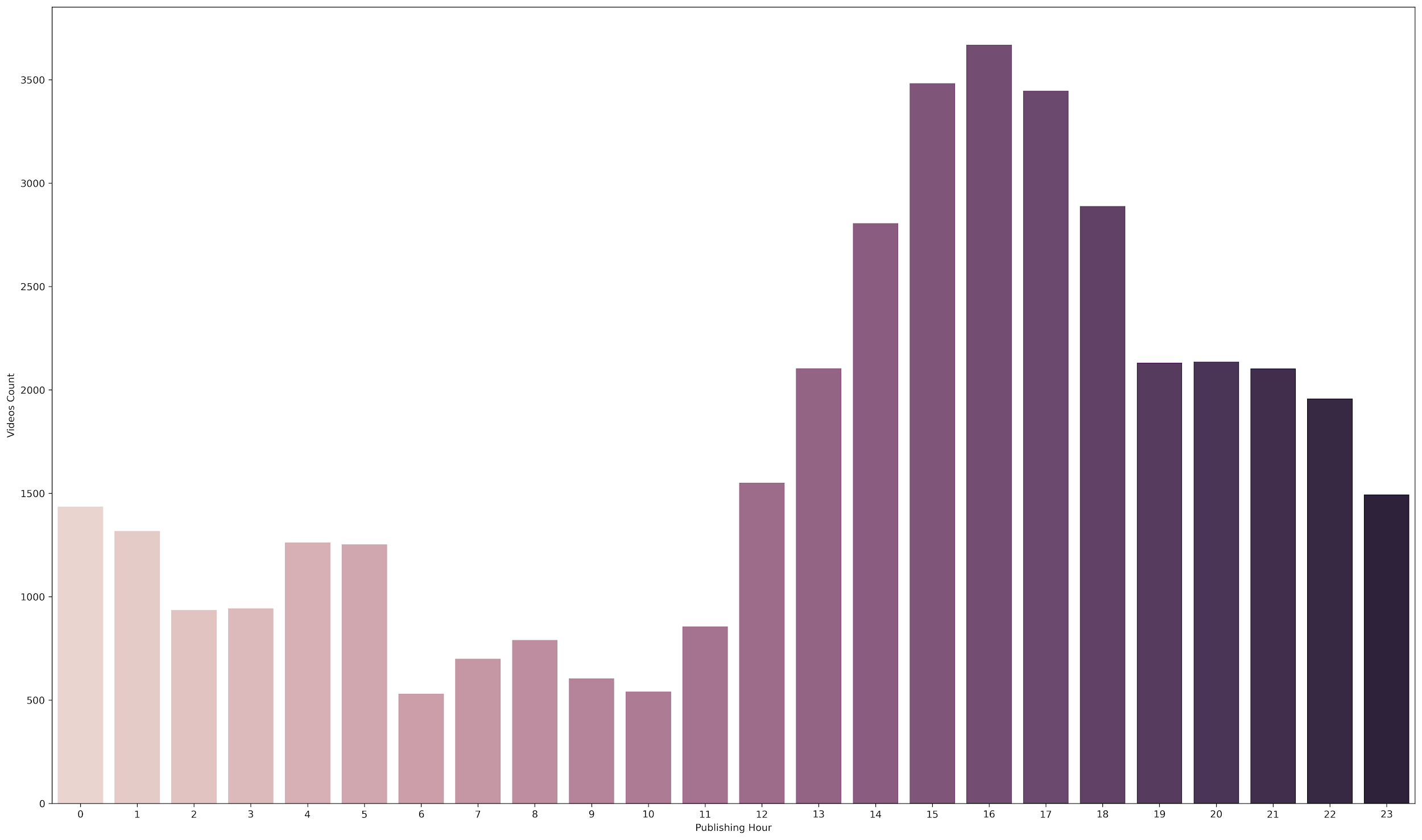


From the above chart, we can see the Entertainment category having the highest trending videos. Here we have noticed one more thing is entertainment having more than four times of education videos.

Furthermore, we came up with some different analysis parts, we have find each video published day of the week and get the total count for each day and plot it on a bar chart, here the below image showing the figure



From this chart, we can see that most of the trending videos have been uploaded on Friday and the least amount of video has been uploaded on Saturday. There might be some different reasons for this insight Friday people getting off from work and having more free time during weekends so they are watching and spending more time on YouTube. After this, we came up with another analysis that we found the total count of publishing hour and we plotted into a bar chart.



From the above bar chart, we can clearly identify most of the trending videos have been uploaded during 3 PM to 5 PM.

All these above analysis and finding are telling a lot of information about US YouTube trending videos. After all these findings we can make a good decision about how to make our video as a trending video. Since YouTube became a good source of income and influencing, media people are making a ton of videos every single day. We can help a YouTuber to make their video a trending one we can tell them you have to upload a video on this day, this time and this category. Apart from this, we can say a video became a trending one when people like that video it’s all about audience satisfaction.

1. Acknowledgement

We would like to express my gratitude to my primary supervisors Dr. Shehan Perera and to Dr. Uthayasanker Thayasivam for their support and guidance to carry out this analytic study. We would also like to thank our friends and family who supported us and offered deep insight into the study.

1. Future work

In future, we can conduct and analyze more depth using tags, descriptions and thumbnail columns. So that we can find the most used tags on trending videos for each video category this will really help us, like this we can conduct further analysis on descriptions and thumbnails as well.

1. References