STA 3000

Final Project

Name:Barry Pan

ID:23621605

**Summary**

**Motivation:**

Youtube has become such a big platform for sharing videos, and over the years, it has played a vital role of spreading pop culture through popular videos across the internet web. Only the best of the best videos in Youtube are allowed to be placed in the trending page of Youtube, which signifies which videos are currently popular at the moment. Since Youtube is so big, I want to use the dataset from youtube’s trending page to identify which type of videos tends to be most successful/popular.

**Question:**

I don’t have one exact question, however since the trending page for youtube videos is a good indicator for which type of Youtube videos are successful. I want to identify patterns from the best. Therefore my main questions are which type of successful video tends to get the most views,which type of videos are best received, and what are some factors that can influence its popularity.

**Why do you care about this question:**

I care about this question because being a user who has spent hundreds of hours on this website, I would like to gain more depth of knowledge behind the general trend of videos for youtube. In addition, Youtube is a big platform to share what is currently popular with our “pop-culture”, therefore this analysis would prove to also be useful on catching up on which type of videos tend to be ‘popular’.

**Why should other people care?**

For those who wish to pursue a career in Youtube, this analysis can be useful to identify and guide future content creators on what type of videos they should upload to grow more popular or raise awareness of something. Being successful in Youtube not only might support you financially, but also help people to spread awareness of current day issues through music videos,entertainment,podcasts, and etc.

**Data** **Summary**

Source:

<https://www.kaggle.com/datasnaek/youtube-new#USvideos.csv>

This is a source from kaggle that has extracted data from Youtube’s trending page from 2017 to 2018. It includes many trending videos from other foreign countries, however for relevance; I will only be using the dataset from the US.

**Variables:**

Video id,trending date,title of video,channel title, category of video, published time,tags,views,likes,dislikes, comment count, thumbnail link, comments disabled, ratings disabled, video error or removed, and description.

IMPORTANT:

For this analysis, we will mostly only focus on title of video, category of video,views,likes and comments as the other variables aren’t very good at predicting the popularity of a video.

**Data Manipulation**

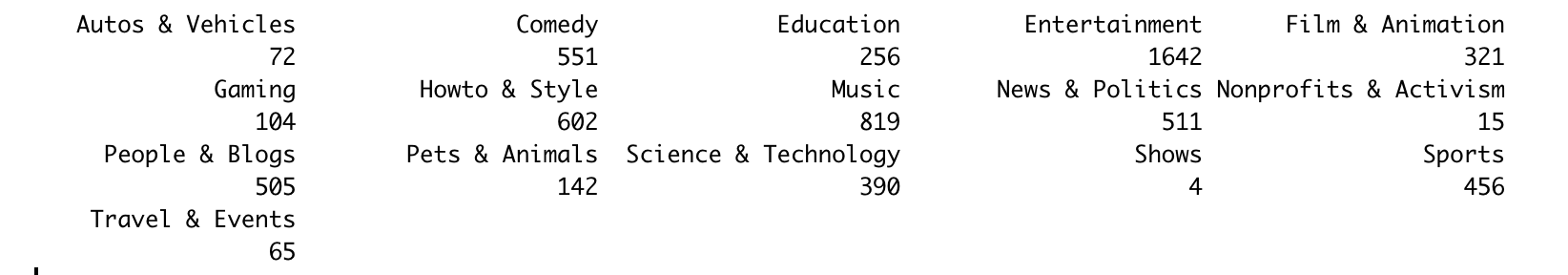
For this dataset, I realized that the same videos would occasionally appear more than once, in order to maintain consistency I had to remove repeating values of video title by using the distinct function. I created a different subset with a distinct function to remove repeating Youtube videos. Therefore the new subset now only holds unique videos and each video only appears once.

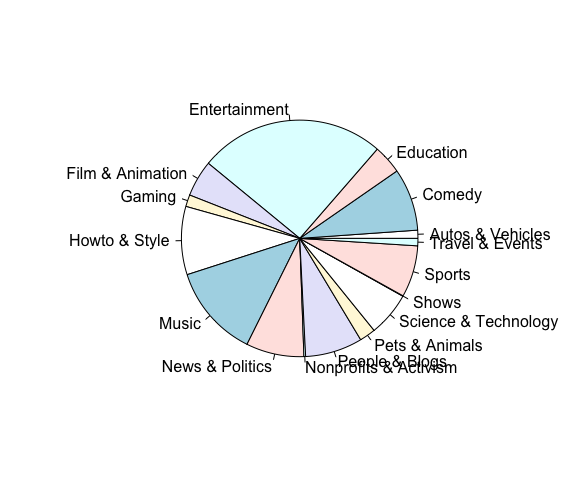
In addition, the dataset category was not listed as a character but as a numeric value. To solve this issue, I had to rename each category id by specifying each numerical value and assigning it to a character value.

**Data Summary**

Number of videos:6455

Number of categories in trending page:





So far based on the chart, it is visible that youtube video under the category of entertainment holds the spot of being the most popular in the trending page at 1642 videos and category music coming in second with 809. This means entertainment videos tend to be featured in trending.

**Linear model:**

I will be running a simple linear model that showcases the correlation and relationship between likes,comments and views. I used comments and likes as variables because I suspect that they have a positive correlation with the views a video would get.

Input:Likes

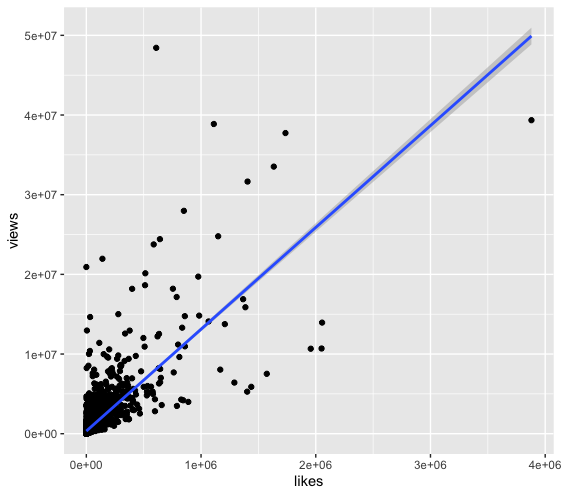
Output:Views

Coefficient:

-Intercept:3.320e+05

-Likes:1.278e+01 p-value:2e-16

Since p value is2e-16 and below .05 and the coefficient is 1.278e+01, we can say that likes is significantly related to the amount of views a video gets and has a positive relationship. The coefficient means that each unit increase in likes will also increase views by 1.278e+01.



Input:Comments

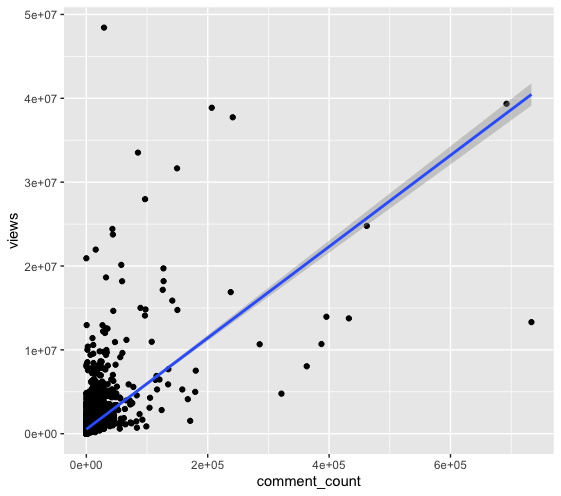
Output: Views

Coefficient:

-Intercept:5.281e+05

-Comments:5.446e+01 p-value:2e-16

Since p value is2e-16 and below .05 and the coefficient is 5.446e+01, we can say that comments are significantly related to the amount of views a video gets and has a positive relationship. The coefficient means that each unit increase in comments will also increase views by 5.446e+01.



Variable: Views Variable: Likes Variable:Comment\_Count

Median : 274060 Median : 8011 Median : 922.0

Mean : 772513 Mean : 34469 Mean : 4488.5

Max. : 48431654 Max. :3880071 Max. :733373.0

Sd : 7394114 Sd : 228885.3 Sd :37430.49

Var :5.467292e+13 Var :52388498047 Var :1401041357

**TOP 15 IN VIEWS:**

**Video Names:**

[1]"Maluma - El Préstamo (Official Video)"

[2] "BTS (방탄소년단) 'FAKE LOVE' Official MV"

[3] "TWICE What is Love? M/V"

[4] "Marvel Studios' Avengers: Infinity War Official Trailer"

[5] "Ed Sheeran - Perfect (Official Music Video)"

[6] "Childish Gambino - This Is America (Official Video)"

[7] "VENOM - Official Trailer (HD)"

[8] "YouTube Rewind: The Shape of 2017 | #YouTubeRewind"

[9] "Drake - Nice For What"

[10] "Sanju | Official Trailer | Ranbir Kapoor | Rajkumar Hirani | Releasing on 29th June"

[11] "周杰倫 Jay Chou【不愛我就拉倒 If You Don't Love Me, It's Fine】Official MV 發燒影片華語地區蟬聯第一 美國直衝前二"

[12] "To Our Daughter"

[13] "Migos, Nicki Minaj, Cardi B - MotorSport (Official)"

[14] "Marvel Studios' Avengers: Infinity War - Official Trailer"

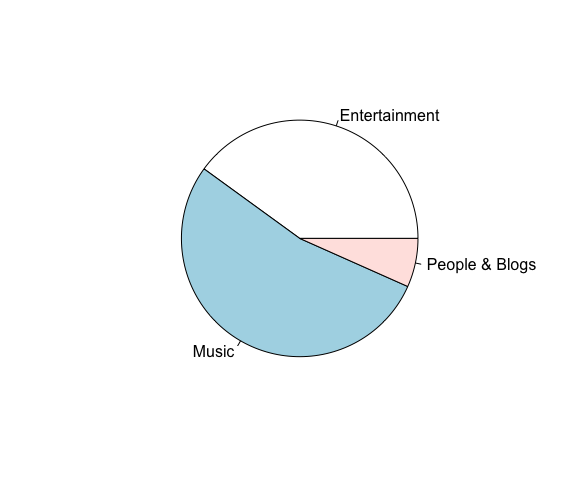
[15] "Sanju | Official Teaser | Ranbir Kapoor | Rajkumar Hirani"

**Categories in views:**

-Music:8

-Entertainment:6

-People & Blogs: 1

****

This time based on the chart it seems that category music is most popular in terms of highest views in trending. This is a heavy indicator that music videos tend to garner the highest views.

**TOP 15 IN LIKES:**

[1] "BTS (방탄소년단) 'FAKE LOVE' Official MV"

[2] "BTS (방탄소년단) 'MIC Drop (Steve Aoki Remix)' Official MV"

[3] "j-hope 'Daydream (백일몽)' MV"

[4] "BTS (방탄소년단) LOVE YOURSELF 轉 Tear 'Singularity' Comeback Trailer"

[5] "Marvel Studios' Avengers: Infinity War Official Trailer"

[6] "Ed Sheeran - Perfect (Official Music Video)"

[7] "BTS (방탄소년단) 'Euphoria : Theme of LOVE YOURSELF 起 Wonder'"

[8] "BTS (방탄소년단) 'FAKE LOVE' Official MV (Extended ver.)"

[9] "Childish Gambino - This Is America (Official Video)"

[10] "j-hope 'Airplane' MV"

[11] "Ariana Grande - No Tears Left To Cry"

[12] "we broke up"

[13] "BTS (방탄소년단) 'FAKE LOVE' Official Teaser 2"

[14] "LOGAN PAUL IS BACK!"

[15] "Suicide: Be Here Tomorrow."

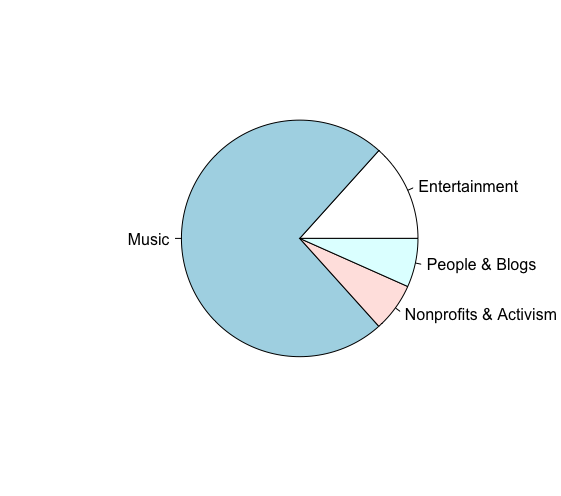
**Categories in likes:**

-Music:16

-Entertainment:2

-People & Blogs:1

-Nonprofit & Activism:1



It seems like the music category tends to attract the most likes which makes music videos one of the best received videos.

**TOP 15 IN COMMENTS:**

**Video Names:**

[1] "So Sorry."

[2] "BTS (방탄소년단) 'FAKE LOVE' Official MV"

[3] "YouTube Rewind: The Shape of 2017 | #YouTubeRewind"

[4] "LOGAN PAUL IS BACK!"

[5] "BTS (방탄소년단) 'MIC Drop (Steve Aoki Remix)' Official MV"

[6] "j-hope 'Daydream (백일몽)' MV"

[7] "Suicide: Be Here Tomorrow."

[8] "Melting Every Lipstick From Sephora Together"

[9] "BTS (방탄소년단) LOVE YOURSELF 轉 Tear 'Singularity' Comeback Trailer"

[10] "Marvel Studios' Avengers: Infinity War Official Trailer"

[11] "we broke up"

[12] "TWICE What is Love? M/V"

[13] "BTS (방탄소년단) 'Euphoria : Theme of LOVE YOURSELF 起 Wonder'"

[14] "MOMMY AND DAUGHTER SURPRISE DADDY WITH PREGNANCY ANNOUNCEMENT!!! (SPEECHLESS)"

[15] "NEW WET N WILD 'GOTH-O-GRAPHIC' COLLECTION | HIT OR MISS?!"

**Categories in comments:**

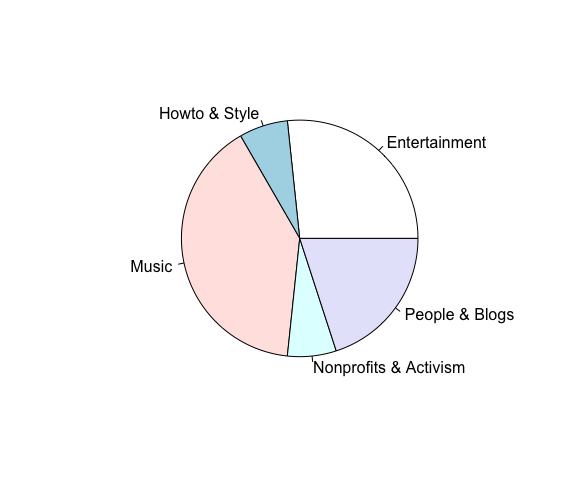
Entertainment:4

Music:6

People & Blogs:3

Howto & Style:1

People & Blogs:1



Once again music videos comes out on top. Comments can popularize a video,however; a video may have a lot of comments due to being “controversial”

**Hypothesis Testing(Two Sample):**

Now based on the previous information we just extracted from basic statistics, it seems that the two most popular categories of videos are music and entertainment.However, so far music seems to dominate the highest viewed videos. So we have a bias that music videos have a higher view count. Using the mean of views from both categories, I want to see if there is a significant difference in mean from zero with a 95% confidence interval in a two sample test. Therefore I will use my null hypothesis as the difference in mean is zero, while my alt hypothesis as the difference in mean is not zero.

Results:

t = -4.4735, df = 1038.5, p-value = 8.543e-06

mean of Entertainment video mean of of Music videos

827875.2 1476112.8

95 percent confidence interval:

-932579.8 -363895.2

From running the hypothesis test, we see that the p value is well below .05, which means the data is significant. Since it is significant, we will have to reject the null hypothesis and go with the alternative that there is a difference in the true mean.

Conclusion:

Since there is a significant difference in mean of views between music videos and entertainment videos, we will have to conclude that music videos are indeed the most successful videos on Youtube. Although entertainment videos are the majority of the videos in trending, their success in views don’t reach the level of those music videos. It is evident music videos have superior views compared to the rest when the mean views of music videos is 1476112.8 and the mean for view of entertainment video is only at 827975.2. Another indicator was when we saw how music video was the majority of the categories of videos when looking at the top 15 highest views on Youtube. One reason for why music videos are the most popular type of videos is due to the fact people often revisit the video that adds on to views. For instance, a person might use Youtube to replay their favorite song but wouldn’t replay an entertainment video. From running the linear model, we can also see that likes and comments also contribute and have a positive relationship with views. Likes have a coefficient 1.278e+01 and comments have a coefficient of 5.446e+01 when in linear relationship with views. This makes sense as Youtube’s algorithm allows the Youtube video to become more recommended to others the more comments and likes a video has. Evidently in the amount of views,comments, and likes; music videos tend to have the most.

If I were to recommend a type of video to post on Youtube, I would recommend people to post music videos as they are the best received,are most popular, and tend to hold the most comments/likes which can further spread awareness for the video.

Limitations: If there was one, I would say it would be if the video was reuploaded or not. When a video is reuploaded due to difficulties, all of the preexisting views are deleted and would have to start anew. This information is vital as it can demonstrate whether another type of video were to be more popular despite having lower views due to their videos being reuploaded. In addition, entertainment videos are notorious for being revoked or reuploaded due to copyright issues.

Questions answered:

Which type of successful video tends to get the most views:Music videos

Which type of videos are best received:Music videos

What are some factors that can influence videos popularity:Comments and likes,