

**Subject: Data Visualization**

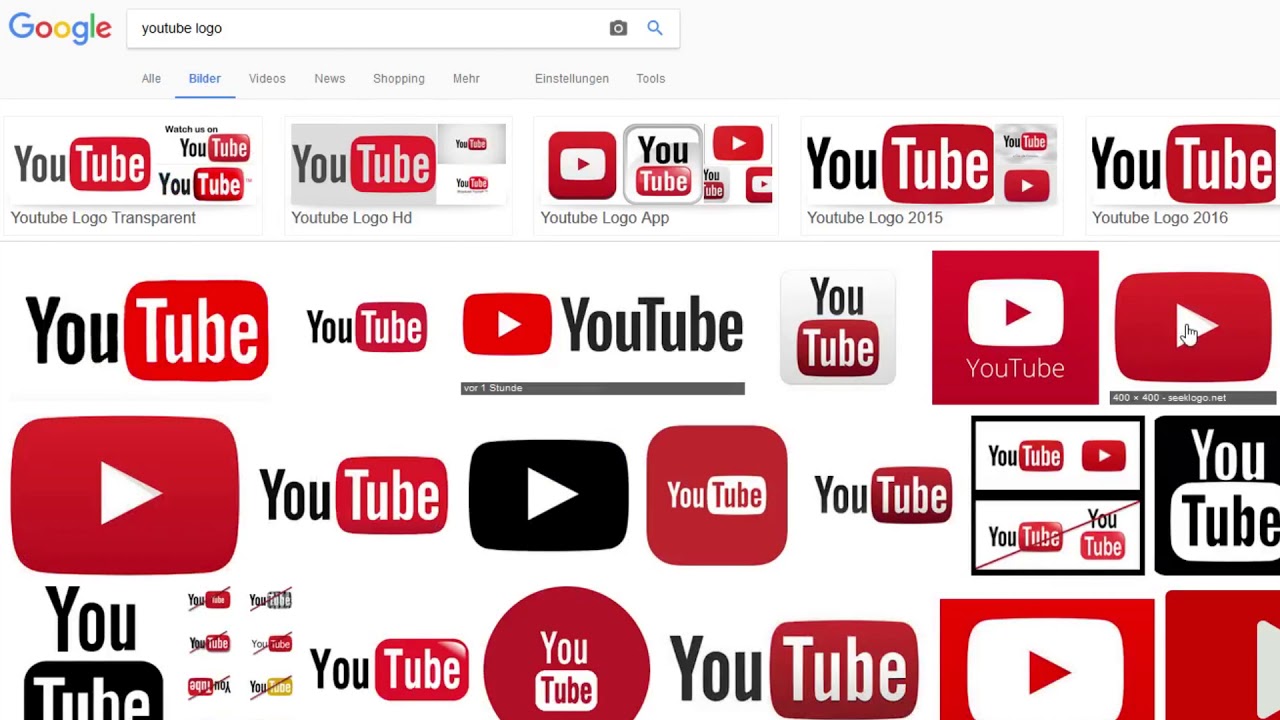
**Lecturer Name: Basel Magableh**

**Assignment 1: CA-2 Report**

A] Group Members :

|  |  |
| --- | --- |
| Student Name(s) | Student ID number (email) |
| 1. Pradeep Kamble | 10511829 [10511829@mydbs.ie] |
| 1. Nikhil Gharge | 10517429 [10517429@mydbs.ie] |
| 1. Pradeep Koli | 10390761 [10390761@mydbs.ie] |
| 1. Aparna Samboji | 10519676 [10519676@mydbs.ie] |

**Dataset: Indian You Tube Video Trending Statistics**



This dataset includes several months (and counting) of the data on daily trending YouTube videos. We have taken Indian YouTube Video Trending dataset. In this dataset all the videos which were trending in year 2017 and 2018. There are following features which are present in the dataset.

1. Video ID: Video Id of the video
2. Trending Date: The date when video is on trending
3. Title: Title of the video
4. Channel Title: Title of the channel
5. Category ID: Category id of the video
6. Publish Time: Time of video Publishing
7. Tags: Tags on the video
8. Views: Number of views
9. Likes: number of likes
10. Dislikes: number of dislikes
11. Comment Count: total number of comments
12. Thumbnail Link: Link of the thumbnail
13. Comments Disabled: the comments which are disabled
14. Ratings Disabled: the ratings which are disabled
15. Description: the description of trending video

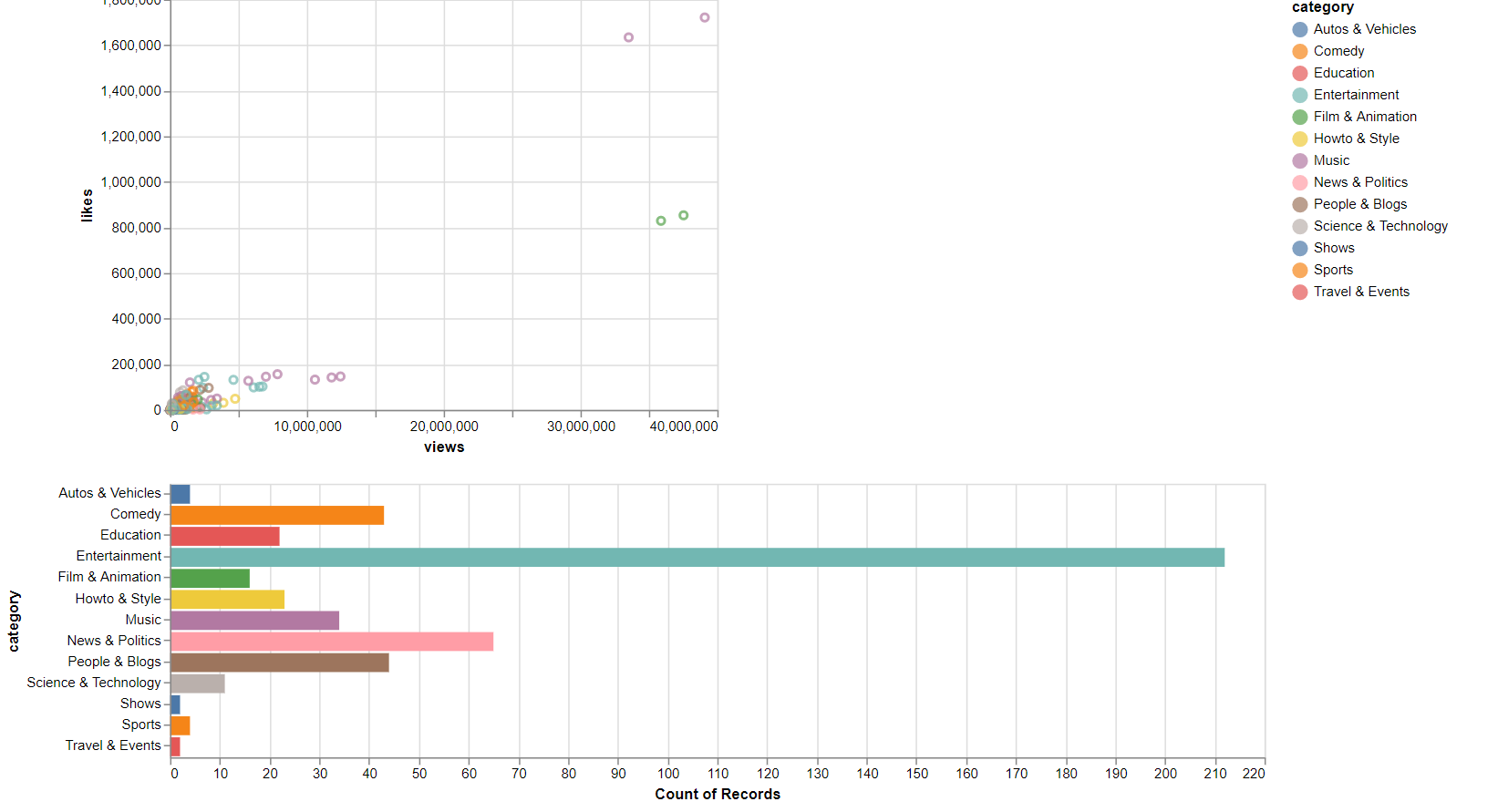
**Initial Analysis Questions:**

1. How many views do our trending videos have? Do most of them have a large number of views? Is having a large number of views required for a video to become trending?
2. Which video remained the most on the trending-videos list?
3. How many trending videos contain a fully capitalized word in their titles?
4. What are the lengths of trending video titles? Is this length related to the video becoming trendy?
5. How are views, likes, dislikes, comment count, title length, and other attributes correlate with (relate to) each other? How are they connected?
6. What are the most common words in trending video titles?
7. Which YouTube channels have the largest number of trending videos?
8. Which video category (e.g. Entertainment, Gaming, Comedy, etc.) has the largest number of trending videos?
9. When we’re trending videos published? On which days of the week? At which times of the day?

**Discoveries and insights:**

Our analysis starts with plots of individual variables to assess distributions and data quality. As we progress, we build up interactive views for our analysis questions and Summarize Indian YouTube video trending statistics.

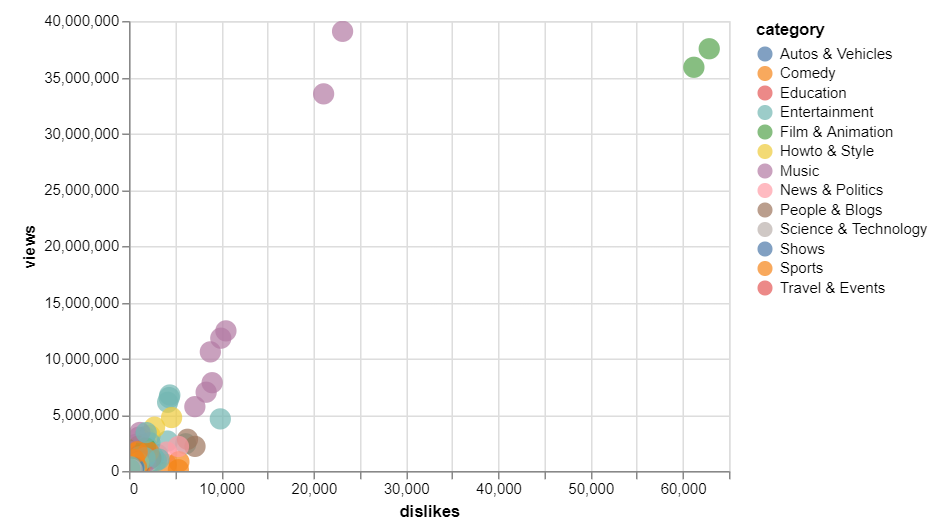
1. **Point Graph with Histogram as per category with numbers of views and likes:**



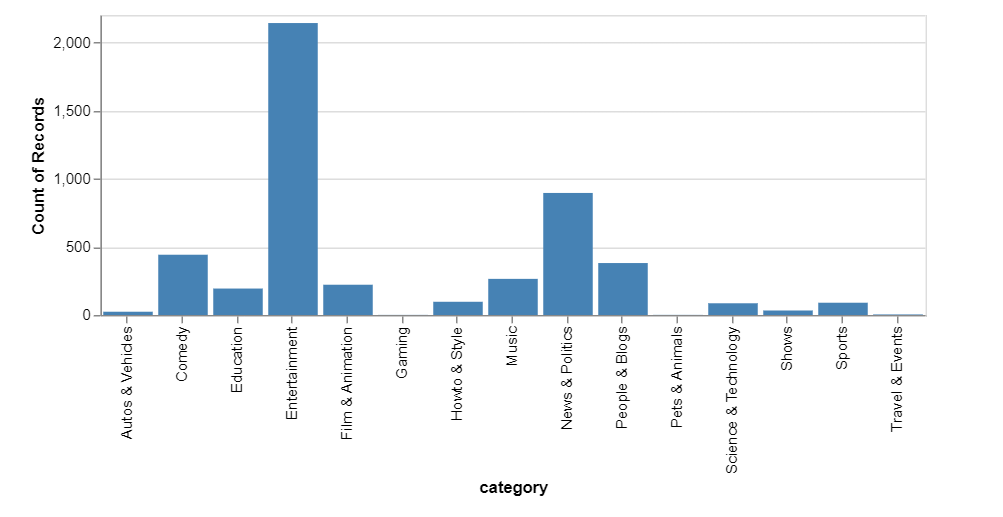
The Point graph shows relation between views and likes. On x-axis we take views and on y-axis we take likes. We mentioned the legend of category which includes category of You tube videos. From Point graph we can easily determine that music category have a greater number of views and likes. The second most is film and animation graph having more number of views and likes.

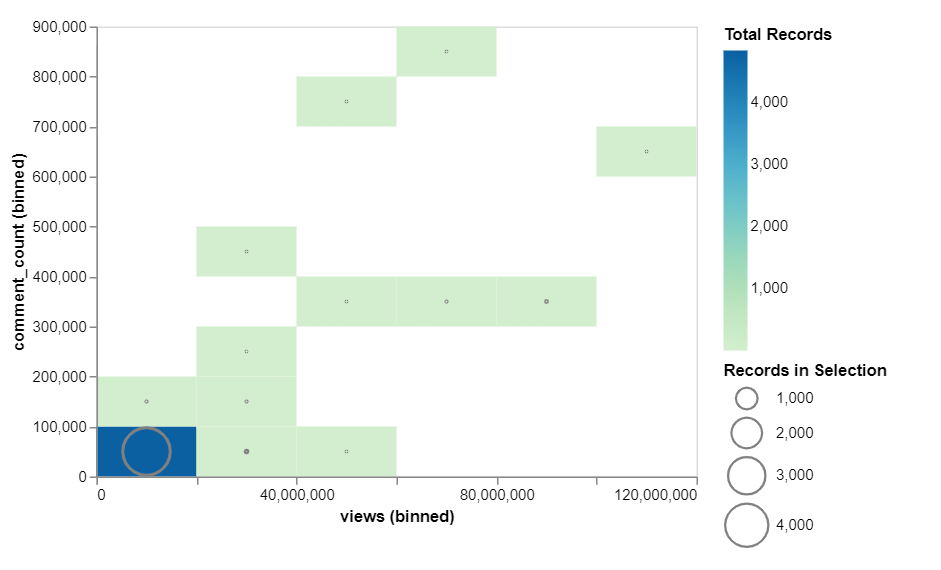
The histogram shows relation between category and count of records. On x-axis we take count of records and on y-axis we take category. From histogram we can easily conclude that Entertainment having a greater number of videos which are on trending. On opposite Travel and Events have less number of videos. Comedy and People blogs have almost same number of videos.

### Point Graph as per category with numbers of views and dislikes



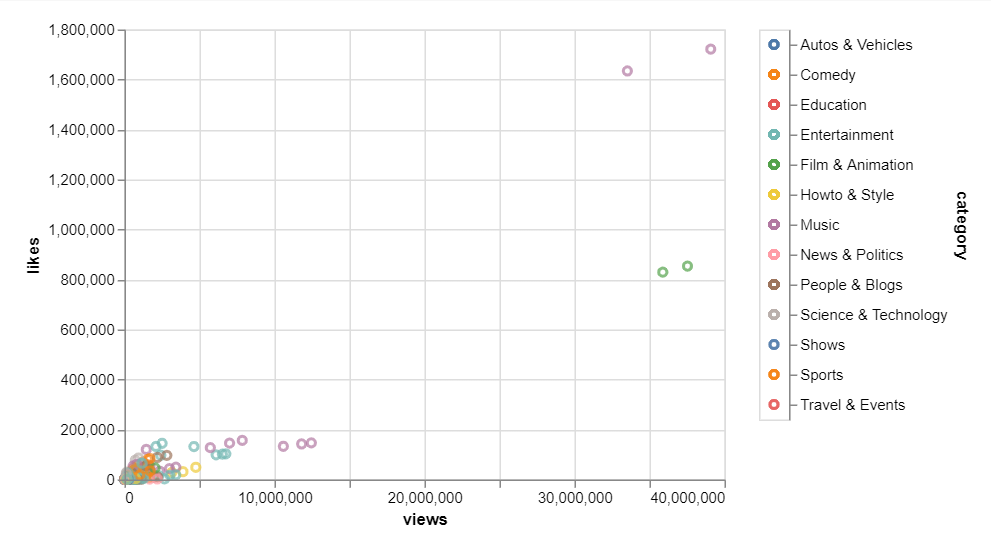
The Point Graph shows the relation between numbers of views and dislikes as per category. On x-axis we take dislikes and on y-axis we take views. From above graph we can easily determine that film and animation having YRF channel with 37533416 views and 62858 dislikes. On the other hand, Set India having less number of views i.e.160501 and 108 dislikes.

1. **Interactive chart with cross highlight:**



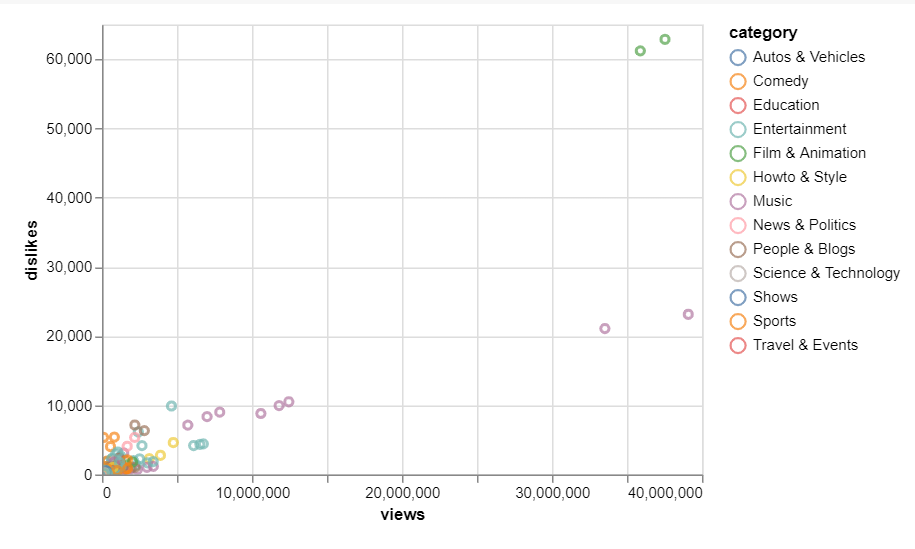
There are Two charts we have shown barplot and cross highlighter. In barolot on x-axis we take category and on y-axis we take count of records. From barplot we can easily determined that Entertainment category have greater number of records. Whereas Travel and events have less number of records. Once we click on the category of bar plot we can see automatically records in selection i.e. how many number of records in each category. In cross highlighter that point shows the how many views per category.

## **Point Graph as per category with numbers of views and likes:**



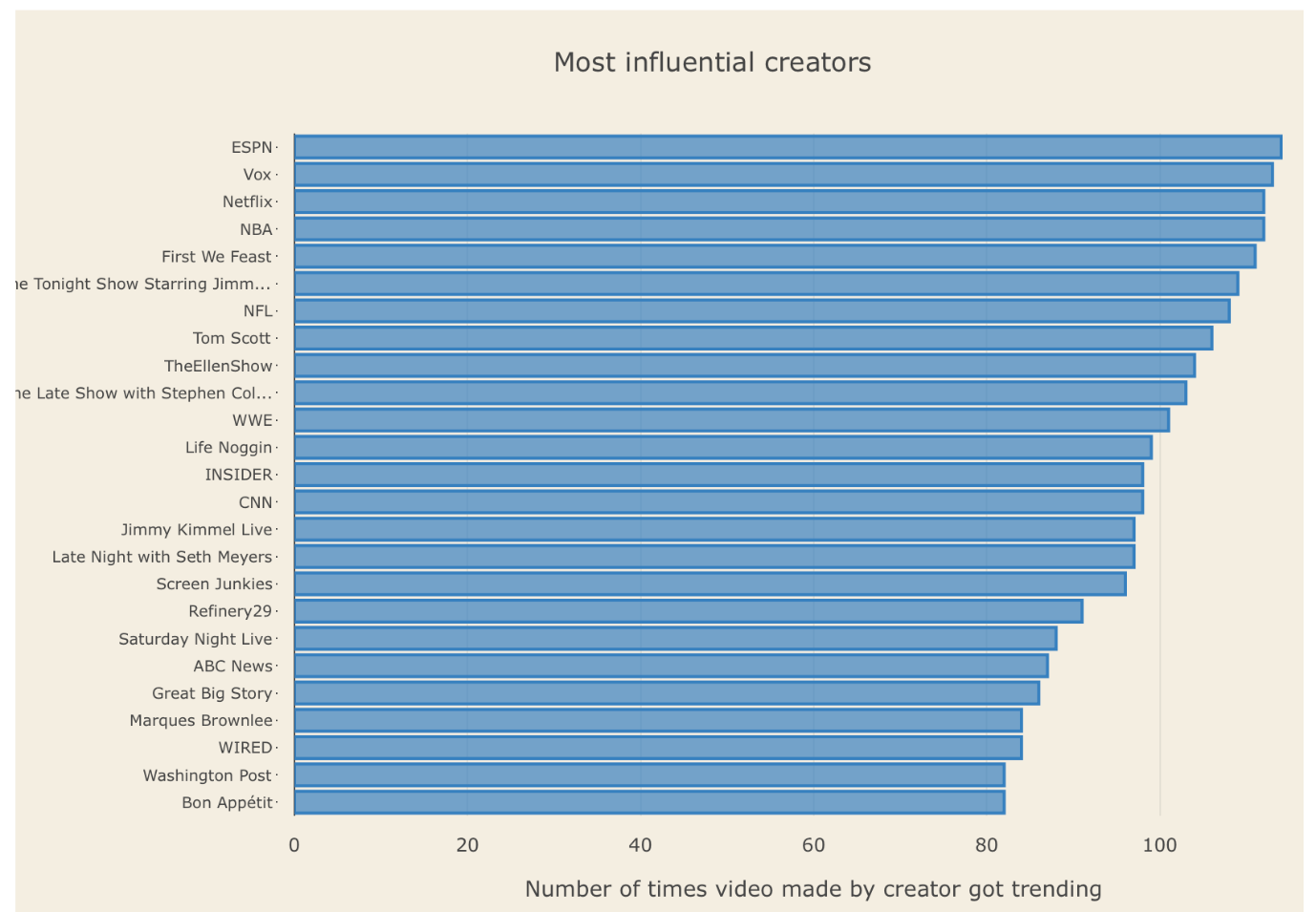
The above graph shows that likes has linear trend with number of views. We can observe that Film & Animation and Music category has the maximum number of views and likes. In contrast other category has very few views and likes. From the data we can say that most viewed videos has relatively more likes among other videos.

**5. Point Graph as per category with numbers of views and dislikes:**



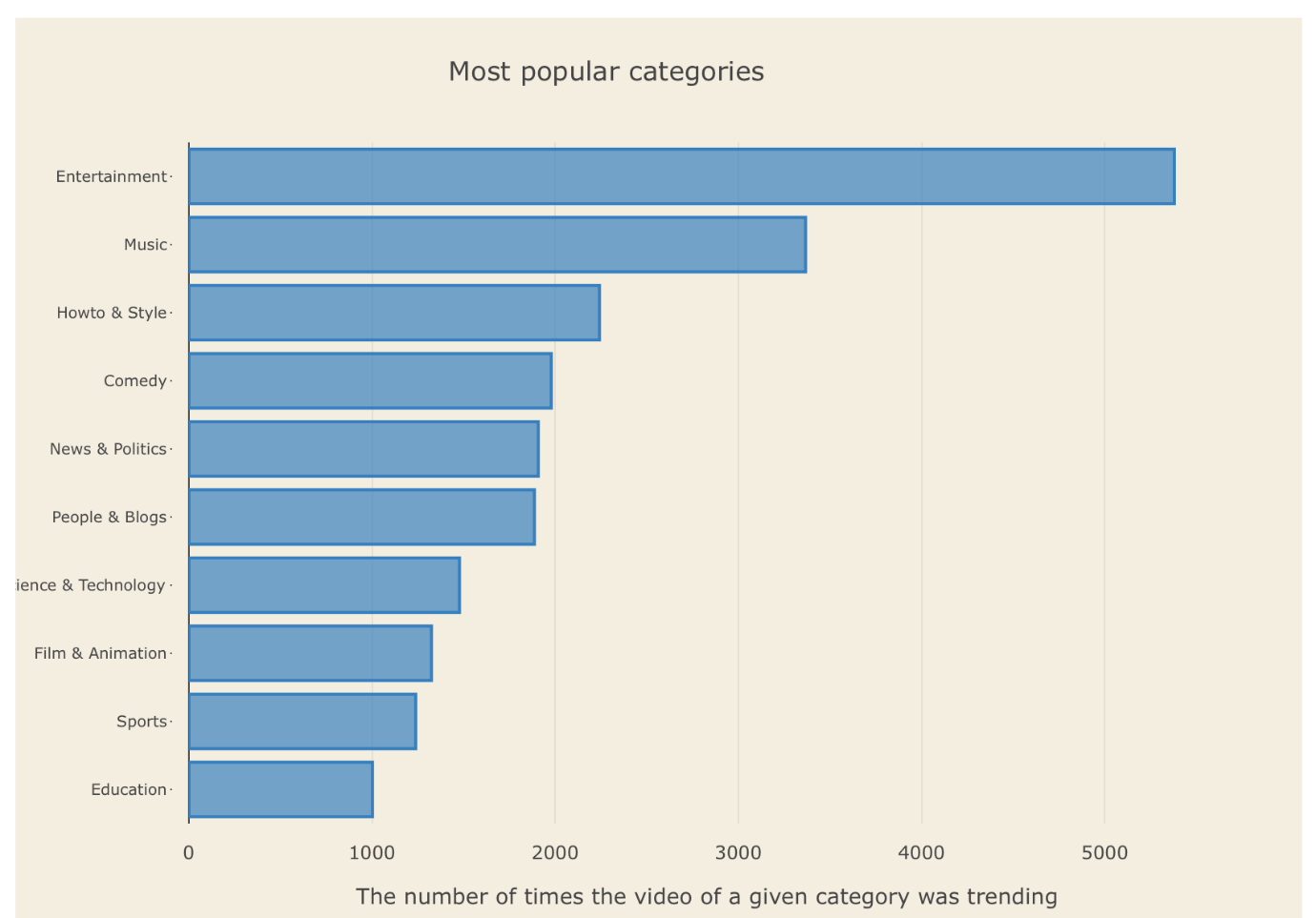
From the graph, we can infer that there is some category which is associated with more dislikes as views increases. Music has the maximum number of views with maximum dislikes whereas Entertainment has the lowest views with dislikes.

1. **Most influential creators:**



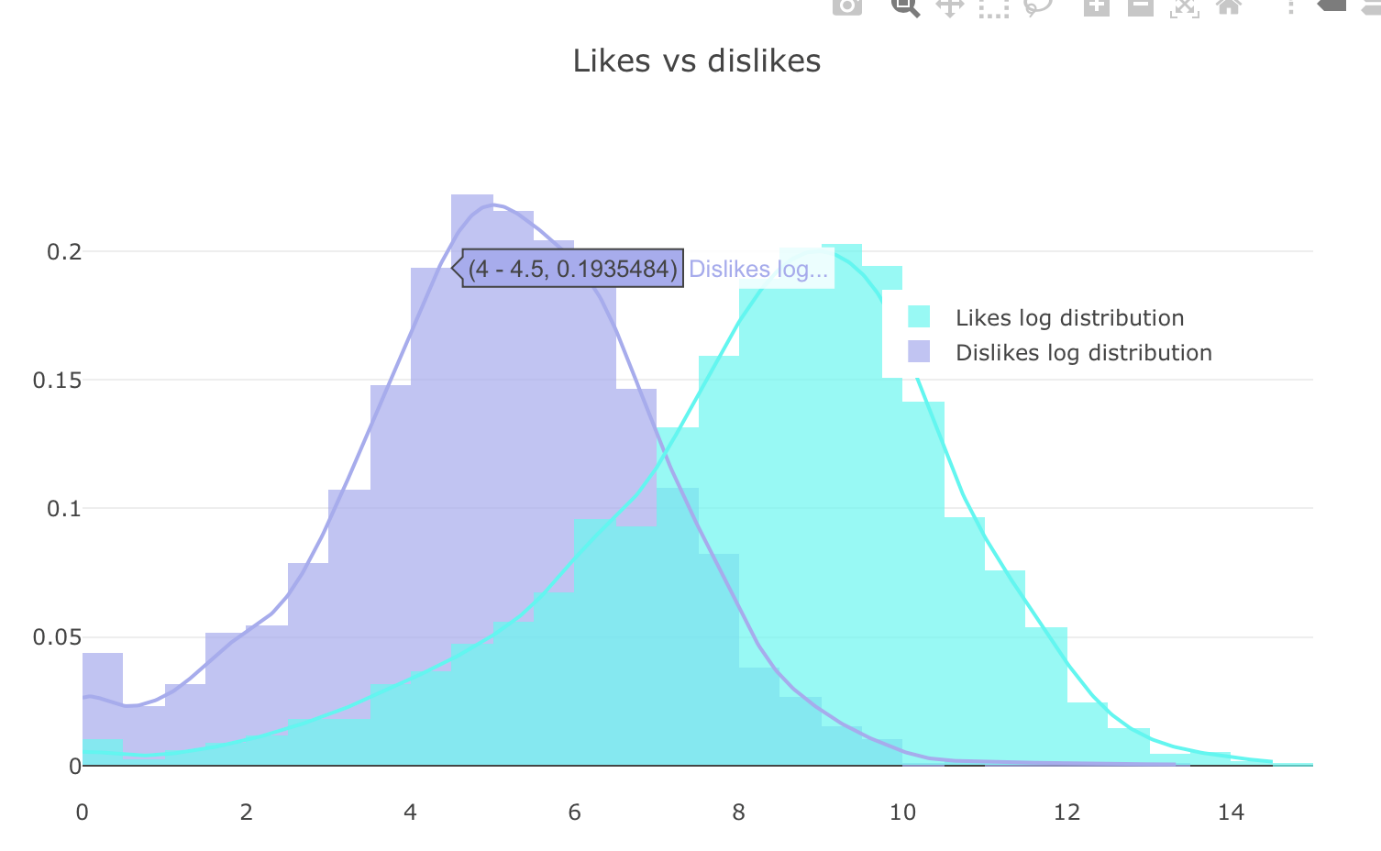
In this bar graph, we have shown top 25 most influential creators. The selection of top creators is based on the number of videos that have been trending over a period of time. As we can see ESPN tops the chart followed by Vox, Netflix, and so on.

1. **Most popular categories:**



In this bar graph, the data represents most popular categories among all of the categories present on YouTube. Graphical representation suggests that Entertainment tops the chart with more than 5000 trending videos followed by music (approx. 3300 trending videos) and so on.

1. **Distplot Distribution:**



In this graphical representation, normal distribution of likes vs dislikes is shown.  As it can be seen in the code we have set the curve top eto normal ignorer to show normal distribution for both the components e.g. like and dislikes.

1. **Word cloud Graph:**

|  |  |
| --- | --- |
|  |  |

1. Most Commonly used tags:

The word cloud on tags used in YouTube dataset shows Big Boss, Tamil, new, news and latests tag.

1. Commonly used words in descriptions:

The above graph shows the word cloud of titles present in YouTube. It shows Trump, Official Trailers, Bigg Boss values.

**Conclusion:**

The main goal behind creating an interactive visualisation was to put-forth our perception by keeping our dataset in mind and give an idea how the graphics and images are seen and interpreted. The targets displayed had a clear insight and determining this becomes easy when the visualisation is explored.

**References:**

**Dataset:** <https://www.kaggle.com/datasnaek/youtube-new#INvideos.csv>

**Colab link:** <https://colab.research.google.com/drive/1FZwaYO6wrtz7YGtVQ2vIhutxxyW90cBc>

**GitHub Link:** <https://github.com/10511829/10511829-Trending_You-tube_video_statistics>