

# Trending YouTube Videos Analysis

{VEERENDRARAMES.KAKARLA@STONYBROOK.EDU, MOHITH.KURAKULA@STONYBROOK.EDU}



#### Introduction

YouTube allows users to upload videos, rate them, comment on them, add to playlist etc., also it maintains the list of trending videos based on user-interactions(views, shares, likes, comments etc.,). data of top trending You Tube videos and this dataset is obtained using You Tube APIs where the datasets are obtained for different countries. It's interface suggests which version to be chosen based on IP address of the user. Individuals and large companies use this platform to increase their audience.

Different visualization approaches are used to analyse and draw some insights between trending videos for different countries.

#### **DATASET**

Dataset considered for this project includes months of data on daily trending YouTube videos. Data is considered for different countries like US, India, Britain etc., with up to 100 listed trending videos per day. Each countries data is in separate csv file. Dataset is obtained using You Tube APIs where the datasets are obtained for different countries

To use YouTube Data API we need valid API key, also to get data for specific country we need to pass the country name as parameter and call the API to get the data, also there are readily available scripts through which we can get the data directly into csv files. Using those scripts datasets are generated.

# **INSIGHTS**

From Visual Analysis below are some insights we got from the dataset. In this section the analysis is done specifically for USA country dataset:

- 1) Positive Correlation between Views and Likes.
- 2) Entertainment and Music stand out as Top 2 categories.
- 3) Channels 'ESPN' and 'Netflix' are Top 2 channels.
- 4) Videos published on 'Friday' has more chance of getting to Trending List.
- 5) Words 'How to', 'Pop' are more in tags. 6) More trending videos has days diff of 4 between Publish and Trending date. Insights above are varied for each countries.

#### Conclusion

We analysed the dataset of different countries and audience interests do vary from one country to another country like in USA Entertainment and Music stand in Top 2 categories, where as in India it is Entertainment and News & People which stand out as Top 2. Similarly each countries top 2 categories varied.

For all the countries dataset the correlation between likes and views is always positive.

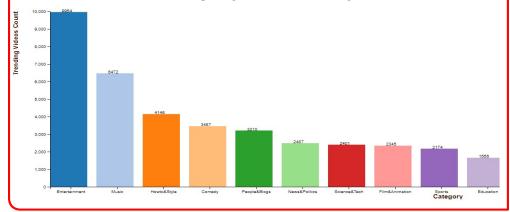
We also made some predictions which include when is the correct time to publish a video and what's the time it takes for a video to get into trending list. From bubble chart we could see that all channel which come under entertainment and music category has more videos in the trending list, also here these insights vary with each country.

### VISUALIZATIONS

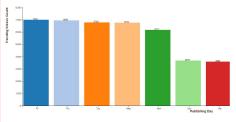
Word Cloud of Tags Used



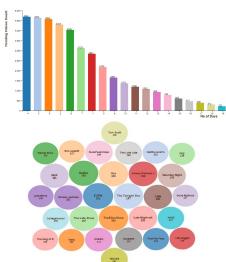
Bar Chart for Top Categories which are Trending in USA



# **PREDICTIONS**



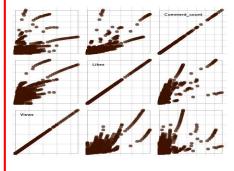
Bar charts b/w count of Trending Videos and Publishing Day, also difference of Publishing Day and Trending Day



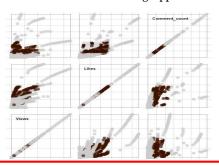
From above visualizations we predict that when the video is from Entertainment category and when uploaded on Friday and the day difference between the publish day and trending day is 4 then the video will most likely trend in USA.

## CORRELATION

Scatter Plot Matrix - Views, Likes, Com-Count



Scatter Plot Matrix –Brushing Applied



# REFERENCES

- Dataset: https://github.com/DataSnaek/Trending-YouTube-
- Scraper https://www.kaggle.com/datasnaek/youtube-new.
- https://bl.ocks.org/mbostock/4063663
- https://bl.ocks.org/alokkshukla/3d6be4be0ef9f6977ec6718b2
  - http://bl.ocks.org/d3noob/8952219
  - https://blog.risingstack.com/d3-js-tutorial-bar-charts-withjavascript/