



US Trending YouTube Video Statistics

Exploratory Data
Analysis and Text
Mining

By Liz Xiong, Yajing Zhou, Kanchan Gyani, Jingjing Huang

01

**Introduction
and Questions**

02

**Data Analytics
Technique**

03

**Findings
and Insights**

04

Conclusions

Introduction

The project is designed to help us uncover the insights behind YouTube trending videos

The original dataset is scraped from YouTube API with the most relevant information on the trending videos

According to Variety magazine, “To determine the year’s top-trending videos, YouTube uses a combination of factors including measuring users interactions (number of views, shares, comments, and likes).”



Youtube

Questions to be explored

- Correlations between the number of views, number of likes, number of dislikes, and number of comments
- Factors that affect how popular a YouTube video will be ; Effects of the publish time on the performance of a video
- Insights from the text analysis of variables like titles, descriptions and tags

Data Analytics Technique



Data Cleaning

Transform data type, impute missing value & manipulate time

Exploratory Data Analysis

Correlation analysis & descriptive analysis

Text Mining

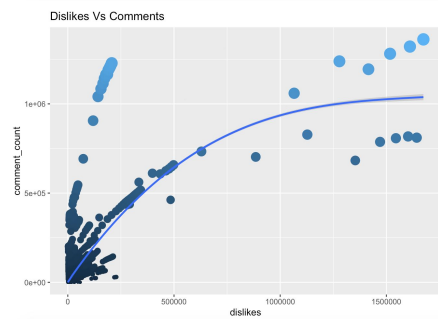
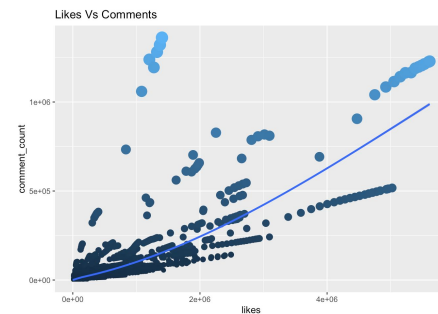
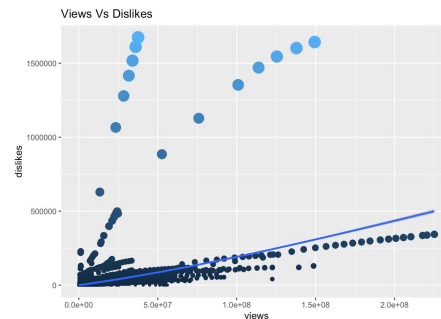
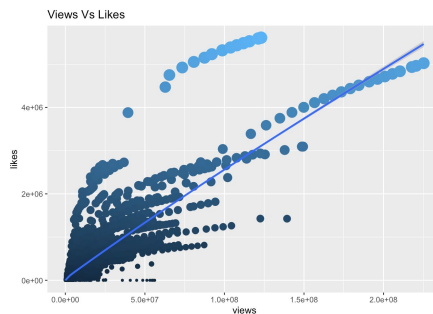
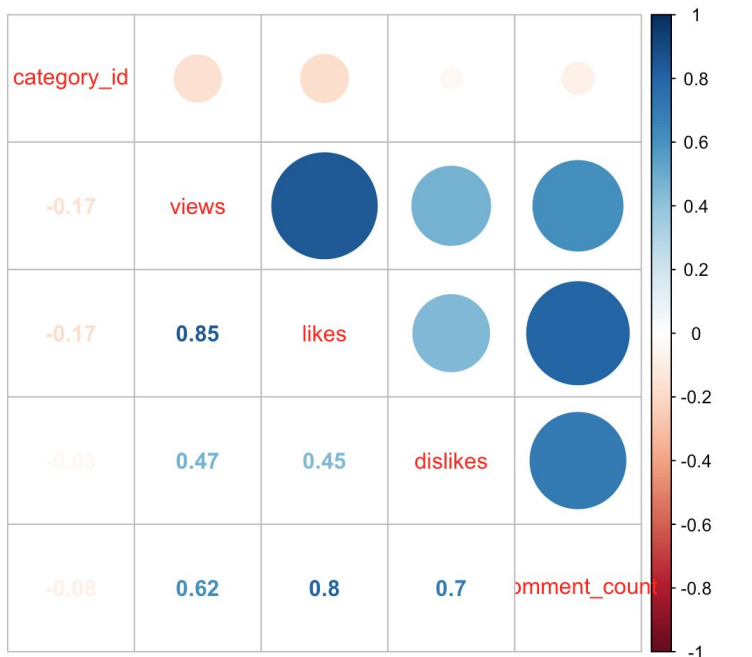
Frequency & Sentiment Analysis

Visualization

Histogram, word cloud, scatter plot & heatmap

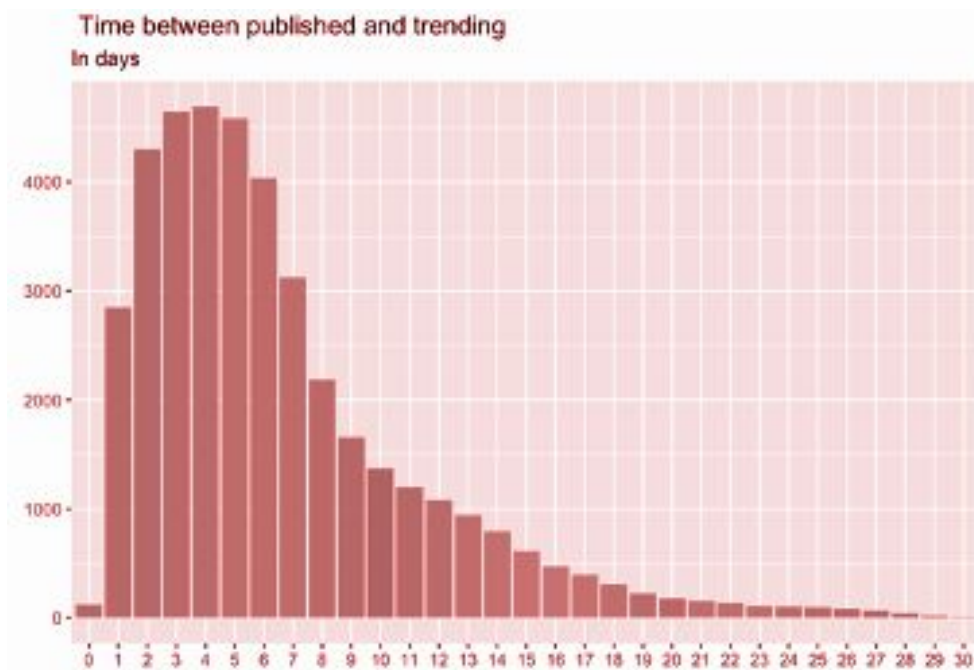
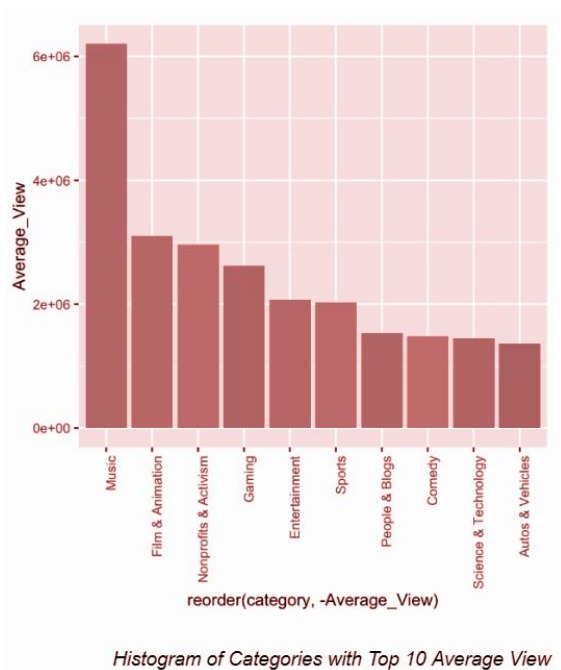
Exploratory Data Analysis

- High correlation between views and likes, likes and comment_count, dislikes and comment_count.
- Videos with higher views tend to have more likes or more dislikes, and videos with more likes or dislikes tend to have more comments.



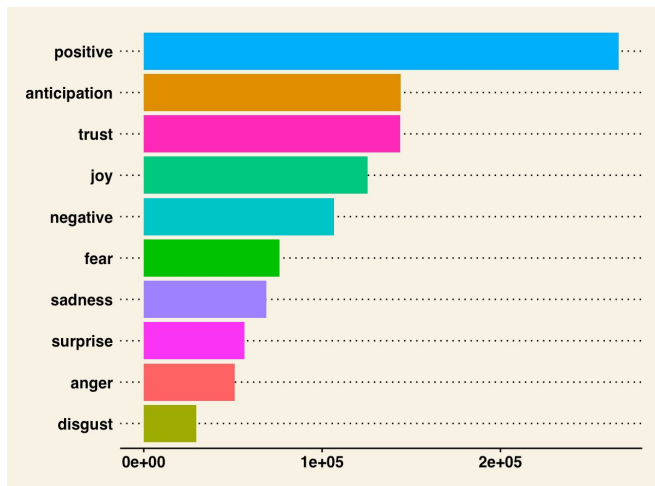
Exploratory Data Analysis

- Music and Film & Animation videos have the highest average view.
- A video rarely becomes trending on the day it is published. 2-6 days is the golden time.



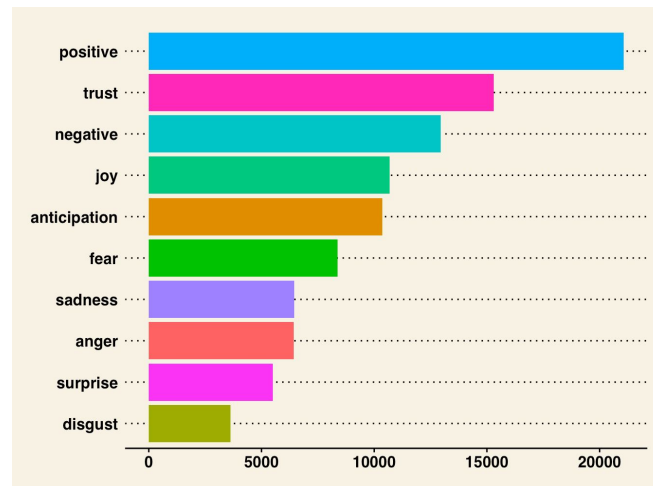
Text Mining

Positive words are mostly used in descriptions and titles, however the negative words stands out too



NRC Lexicon - description

Topmost emotions - positive, anticipation, trust and joy



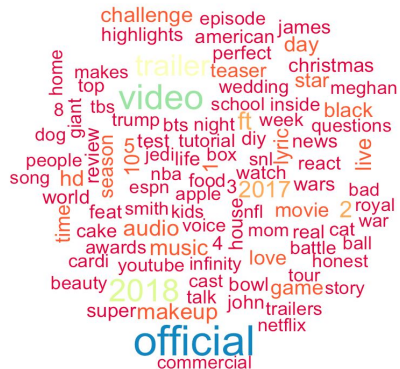
NRC Lexicon - title

Topmost emotions - positive, trust, joy and anticipation

Word Cloud

- To increase the visibility of their videos, Youtube video uploaders should use common words
- Leverage popular tags to make increase the chances of turning the video trending

positive



Title

Most are 'Official' videos
2018 has more trending videos
than 2017



negative

Title 'bing' Lexicon

Love, perfect, beauty, awards - top
positive words

Negative words do not stand out in
proportion to the positive

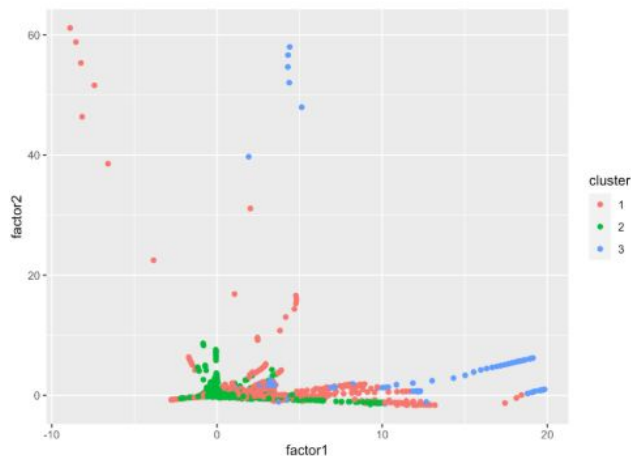


Tags

#video, #funny, #music -
Most frequently used tags

Clustering

- Memory Error during calculation of distances while performing Hierarchical Clustering
- K-means Clustering performed on numeric variables with 3 clusters identified from within sum of square and ratio of sum of square
- Cluster 3 videos(categorized non-profit) have highest avg. views but lowest likes-dislikes ratio



2-D Cluster Plot

##	k_segments	views	likes	dislikes	comment_count
## 1	1	23960331	651853.38	36387.74	68908.13
## 2	2	1368381	47813.78	2057.70	5490.26
## 3	3	108444347	2875072.48	228910.44	381255.40

Cluster Table

Conclusion

- 'Views' is the most important factor defining trending videos. Irrespective of *like* or *dislike*, higher views eventually bring more comments and make the video trending
- The analysis shows trending videos with both *positive* and *negative* emotions
- Most trending videos belonged to official channels or were part of the entertainment industry that uploaded official trailers to Youtube
- The comments section for most of the trending positive videos was disabled and those with negative sentiment were flagged as removed
- A video rarely becomes trending on the day it is published. 2-6 days is the golden time
- The upload time is fairly spread out throughout the weekdays and hours in a day