

You**Tube** Trending Videos

Cesar F. Pardede

Table of Contents

1. The Data
2. Exploratory Data Analysis
3. Statistical Methods
4. Challenges

The Data

- 200 top trending videos on YouTube
- 6 months (14 Nov 2017 - 14 June 2018)
- 10 regions
 - CA, DE, FR, GB, IN, JP, KR, MX, RU, US
 - ~40,000 observations each (JP ~ 20,000)
- ~380,000 total observations
- <https://www.kaggle.com/datasnaek/youtube-new>

The Data

- 16 variables
 - Trending date
 - Video ID
 - Title
 - Channel
 - Category ID
 - Publishing time and date
 - Tags
 - Views
 - Likes
 - Dislikes
 - Comment count
 - Thumbnail link
 - Comments disabled (T/F)
 - Ratings disable (T/F)
 - Video error or removed (T/F)
 - Description
- Additional variables
 - Region
 - Separate publishing time and date

Ideas

Possible/previous uses

- Sentiment analysis
- Trends by region
- Trends over time
- What makes a trending video?
 - Were there videos with high view counts that did not trend?
 - Days since publishing?

Survivorship Bias

Survivorship bias or survival bias is the logical error of concentrating on the people or things that made it past some selection process and overlooking those that did not, typically because of their lack of visibility. This can lead to false conclusions in several different ways.

- Wikipedia article for Survivorship bias

- Airplane example

Exploratory Data Analysis

There exists videos with as few as 117 views, 0 likes, 0 dislikes, and 0 comments.

This indicates trending videos are selected by something other than just these variables, potentially:

- Reach/relevance
- Tags
- Channel properties (subscribers)
- Shares
- Other

Exploratory Data Analysis

- The entertainment category has twice as many trending videos as the next most trending category (People & Blogs)
- 2.3% of globally trending videos have comments disabled
- 2% of globally trending video have ratings disabled

Exploratory Data Analysis

- Most notable videos by region: CA

	Views	Likes	Dislikes	Comments
Title	YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS	YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS
Channel	YouTube Spotlight	ibighit	YouTube Spotlight	ibighit
Count	137,843,120	5,053,338	1,602,383	1,114,800

Exploratory Data Analysis

- Most notable videos by region: DE

	Views	Likes	Dislikes	Comments
Title	YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS	YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS
Channel	YouTube Spotlight	ibighit	YouTube Spotlight	ibighit
Count	113,876,217	4,924,056	1,470,386	1,084,435

Exploratory Data Analysis

- Most notable videos by region: FR

	Views	Likes	Dislikes	Comments
Title	YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS	YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS
Channel	YouTube Spotlight	ibighit	YouTube Spotlight	ibighit
Count	100,911,567	4,750,254	1,353,661	1,040,912

Exploratory Data Analysis

- Most notable videos by region: GB

	Views	Likes	Dislikes	Comments
Title	Nicky Jam x J. Balvin - X (EQUIS)	'FAKE LOVE' Official MV by BTS	So Sorry.	
Channel	NickyJamTV	ibighit	Logan Paul Vlogs	
Count	424,538,912	5,613,827	1,944,971	1,626,501

Exploratory Data Analysis

- Most notable videos by region: IN

	Views	Likes	Dislikes	Comments
Title	YouTube Rewind: The Shape of 2017			
Channel	YouTube Spotlight			
Count	125,432,237	2,912,710	1,545,017	827,755

Exploratory Data Analysis

- Most notable videos by region: JP

	Views	Likes	Dislikes	Comments
Title	'FAKE LOVE' Official MV by BTS			
Channel	ibighit			
Count	62,796,390	4,470,923	119,053	905,925

Exploratory Data Analysis

- Most notable videos by region: KR

	Views	Likes	Dislikes	Comments
Title	YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS	YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS
Channel	YouTube Spotlight	ibighit	YouTube Spotlight	ibighit
Count	113,876,217	5,150,839	1,470,388	1,142,274

Exploratory Data Analysis

- Most notable videos by region: MX

	Views	Likes	Dislikes	Comments
Title	YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS	YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS
Channel	YouTube Spotlight	ibighit	YouTube Spotlight	ibighit
Count	100,912,384	4,470,923	1,353,667	905,925

Exploratory Data Analysis

- Most notable videos by region: RU

	Views	Likes	Dislikes	Comments
Title	'FAKE LOVE' Official MV by BTS		YouTube Rewind: The Shape of 2017	'FAKE LOVE' Official MV by BTS
Channel	ibighit		YouTube Spotlight	ibighit
Count	62,796,390	4,470,923	884,967	905,925

Exploratory Data Analysis

- Most notable videos by region: US

	Views	Likes	Dislikes	Comments
Title	Childish Gambino - This Is America	'FAKE LOVE' Official MV by BTS	So Sorry.	
Channel	ChildishGambinoVEVO	ibighit	Logan Paul Vlogs	
Count	225,211,923	5,613,827	1,674,420	1,361,580

Exploratory Data Analysis

- Most notable videos by region: Global

	Views	Likes	Dislikes	Comments
Title	Nicky Jam x J. Balvin - X (EQUIS)	'FAKE LOVE' Official MV by BTS	So Sorry.	
Channel	NickyJamTV	ibighit	Logan Paul Vlogs	
Count	424,538,912	5,613,827	1,944,971	1,626,501

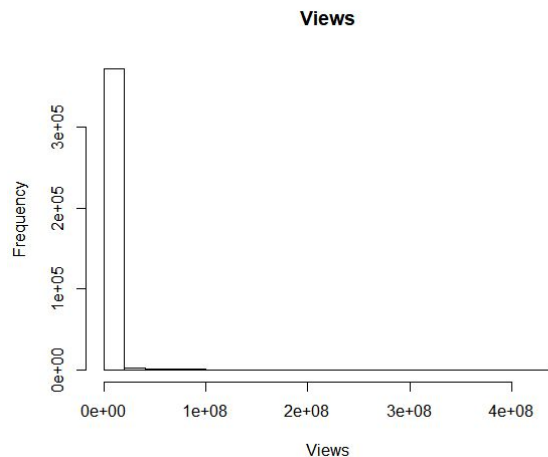
Exploratory Data Analysis

Quantiles of views, likes, dislikes, and comment count

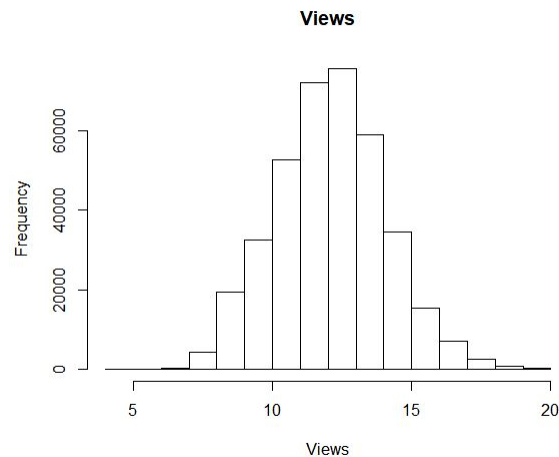
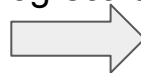
	Views	Likes	Dislikes	Comments
Minimum	117	0	0	0
Median	177,370	3,446	179	511
Mean	1,326,568	37,884	2,126	4,254
95%	4,479,059	162,042	6,561	16,263

Exploratory Data Analysis

Distribution of views, likes, dislikes, and comment count

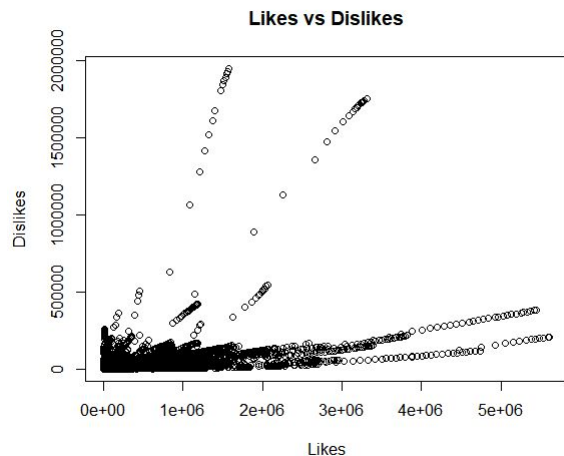


log-scale

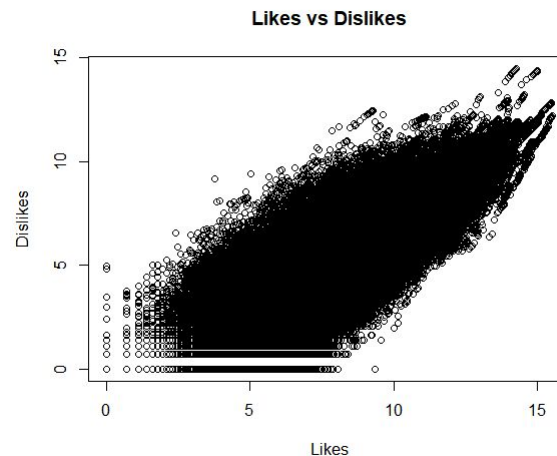
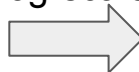


Exploratory Data Analysis

Distribution of views, likes, dislikes, and comment count



log-scale



Exploratory Data Analysis

95% of videos go trending within 14 days of being published. 50% of videos start trending 1 day of being published.

One video started trending on 5 FEB 2018, 4,215 days after being published. That video was “Budweiser - Original Whazzup?”



Exploratory Data Analysis

Longest trending videos

	Video title	Days trending (global sum)
By title	Mission: Impossible - Fallout (2018) - Official Trailer	100
By video ID	Childish Gambino - This Is America (Official Video)	92

Statistical Methods and Modeling

PCA

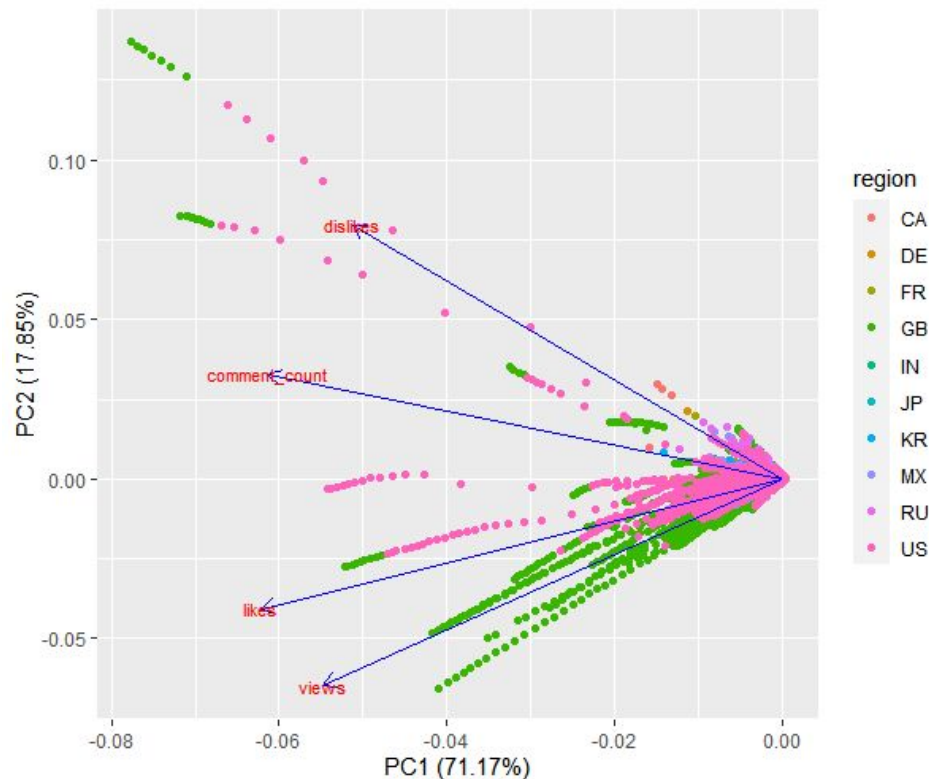
Standard Deviations			
1.6872024	0.8450078	0.6075770	0.2648774

Loadings	PC1	PC2	PC3	PC4
Views	-0.4760740	-0.5631718	-0.5676547	-0.3660044
Likes	-0.5393249	-0.3537354	0.3733230	0.6668057
Dislikes	-0.4449097	0.6905684	-0.4947726	0.2834973
Comments	-0.5334206	0.2842957	0.5418470	-0.5839865

Statistical Methods and Modeling

Plotting PC1 and PC2

- 89% of the variation
- Dislikes and comments are more related than other factors
 - Viewers are more vocal about videos they dislike.
- Views and likes are more related than other factors
 - Obviously, a good video will get more views.



Statistical Methods and Modeling

I tried to train and test SVMs, but ran into problems.

- The data was too large to feed in all together, so I tried subsets split up by region or categories.
- I could train some SVM's but every kernel except 'linear' resulted in a large number of support vectors (around 39k SV's for 40k observations).
- Linear SVM's had 14k SV's.
- Could not predict.
- Could not plot.

Statistical Methods and Modeling

I tried to train and test SVMs, but ran into problems.

- Numerous errors
 - Including: “Error: cannot allocate vector of size 10.0 Gb”
- Each training cycle took 5-10 minutes

Potential Future work

- Parse tags and descriptions
- Neural Network

Challenges

- Unwieldy data size
 - Long computational time
 - Plots took forever to render
 - R sessions aborted
- Limited programming ability
 - Tags and descriptions went unused
- Limited time
 - Some challenges could be overcome with more time

Questions?

Backup Slides

Category ID key:

- | | |
|----------------------|---------------------------|
| 1. Film & Animation | 22. People & Blogs |
| 2. Autos & Vehicles | 23. Comedy* |
| 10. Music | 24. Entertainment |
| 15. Pets & Animals | 25. News & Politics |
| 17. Sports | 26. Howto & Style |
| 18. Short Movies-+= | 27. Education |
| 19. Travel & Events | 28. Science & Technology |
| 20. Gaming | 29. Nonprofits & Activism |
| 21. Videoblogging-+= | |
| 22. People & Blogs | |

Additional categories:

- | | |
|-----------------------|---------------------|
| 30. Movies | 38. Foreign- |
| 31. Anime/Animation- | 39. Horror- |
| 32. Action/Adventure- | 40. Sci-Fi/Fantasy- |
| 33. Classics- | 41. Thriller- |
| 34. Comedy*- | 42. Shorts- |
| 35. Documentary- | 43. Shows |
| 36. Drama- | |
| 37. Family- | |
| 44. Trailers | |

32 Total id's

- category ids don't appear

* Comedy listed twice

+ Categories don't show up in personal YouTube editor

= Short Movies and Videoblogging categories possibly deprecated?