Youtube Trending Videos Analysis

Background

I am an aspiring YouTuber and my dream is for one of my future videos to be **trending**. In order to do so, I want to get some insights from the past trending videos on YouTube. My friend already helped me to extract the data regarding trending videos, but neither of us have the skill to analyze the data.

Objectives

Provide summary regarding **the common factors** of the trending videos:

1. Specific *topic/category* that the video falls onto.
2. Who uploaded these videos
3. What are written on the *title, tags, or description*
4. the engagement metrics (*views, likes, dislikes, comments*)
5. Relationship of the engagement metrics
6. Time between trending date and upload date
7. How long those videos trending
8. Consecutive Trending Days
9. Segmentation

Datasets

There are 2 datasets you are given for this project:
- **youtube-videos.parquet:** this dataset consists of details of Youtube Trending videos (on a selected period between 2017-2018) with relevant metrics such as Views, Likes and Comments
- **youtube-categories.json:** this dataset consists of details of video categories

Metadata

youtube-videos.parquet: :
- **video_id**: unique ID for a Youtube video, note that this value is NOT unique in this dataset (the same video can be listed more than once on multiple trending_date)
- **trending_date**: when the specific video is trending (in YY.DD.MM format)
- **title**: title of the video (what you see before clicking on a YT video)
- **channel_title**: channel that published the video
- **category_id**: unique ID for the video category (see column id in the other dataset)
- **publish_time**: when the video was published (in YYYY-MM-DDThh:mm:ssZ format)
- **tags**: YouTube tags, separated by delimiter "|"
- **views, likes, dislikes, comment_count**: user engagement metrics towards the video
- **comments_disabled**: whether comment is disabled for the video
- **ratings_disabled**: whether rating system is disabled for the video
- **description**: extra sentences written by the uploader to explain the video content

youtube-categories.json
- **id**: unique ID for the category, refers to category_id on the other dataset
- **title**: title of the given category
- **channel_id**: unique identifier for the channel

Analysis

1. **Categorical Data (Category, and Channels)**
   - Which video category has the largest number of trending videos?
   - Which channels produced most trending videos?
   - Another Categorical Value

2. **Text Data (Titles, Descriptions, and Tags)**
   - Most Common Words, Bigrams, and Trigrams in Videos title
   - Most Common Words, Bigrams, and Trigrams in Description
   - Most Common Words, Bigrams, and Trigrams in Tags

3. **Numerical Data (Engagement Metrics)**
   - Video Comments Distribution
   - Top 5 Trending Videos Based on Number of Comments
   - Video Likes Distribution
   - Top 5 Trending Videos Based on Likes
   - Video Dislikes Distribution
   - Top 5 Trending Videos Based on Dislikes
   - Video Views Distribution
   - Top 5 Trending Videos Based on Number of Views

Analysis

4. **Relationship**
   - Correlation between all engagement metrics

5. **Datetime Data (Trending Date and Published Date)**
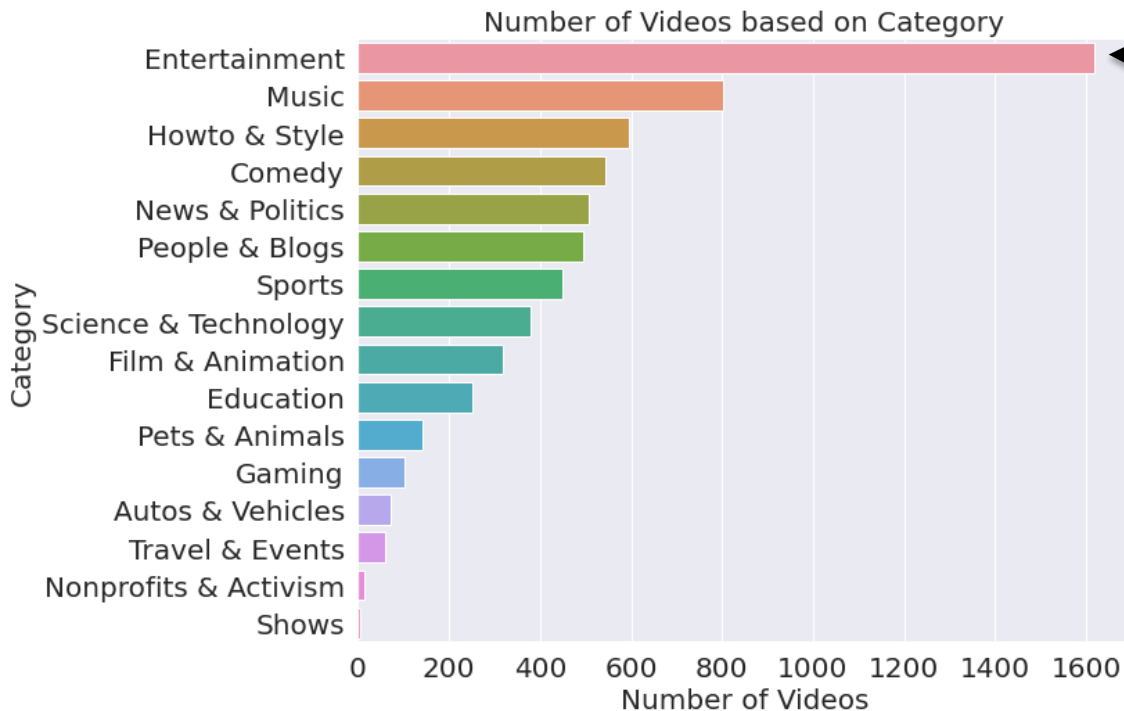   - Number of days between trending and upload
   - How long those videos trending
   - Consecutive Trending Days

6. **Topic Modeling**
   - Topic Group on Titles
   - Topic Group on Tags

# Which video category has the largest number of trending videos?

Number of Videos based on Category

Category	
Entertainment	
Music	
Howto & Style	
Comedy	
News & Politics	
People & Blogs	
Sports	
Science & Technology	
Film & Animation	
Education	
Pets & Animals	
Gaming	
Autos & Vehicles	
Travel & Events	
Nonprofits & Activism	
Shows	

Number of Videos: 0 200 400 600 800 1000 1200 1400 1600

The category that has **the most trending videos** is **entertainment with 1617 videos**, followed by **music with 801 videos** and **style with 595 videos**.

25.46% of all videos are videos with the **entertainment** category

Which channels produced most trending videos?

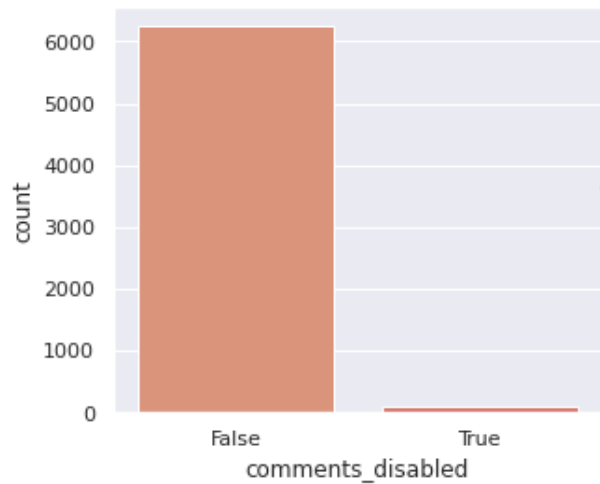## Top 10 Channel with Most Trending Videos
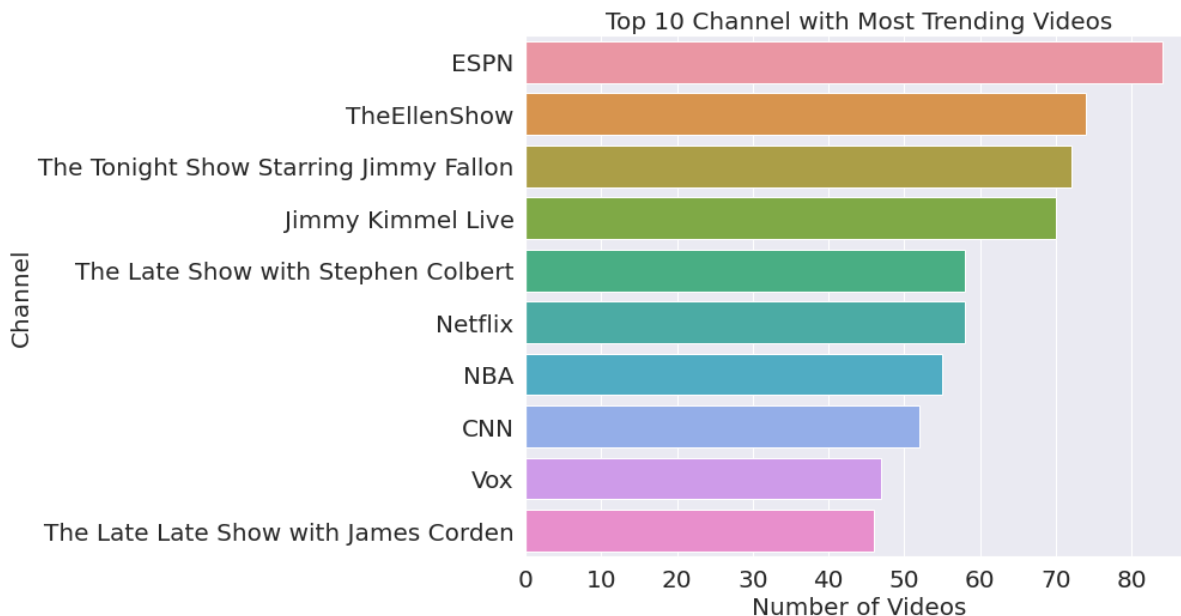
Channel with the Most Trending Videos is **ESPN (Sports) with 84 Videos,** followed by TheEllenhow, and The Tonight Show Starring Jimmy Fallon

6 out of 10 channels in the entertainment category, such as TheEllenhow, The Tonight Show Starring Jimmy Fallon, Jimmi Kimmel Live, etc.
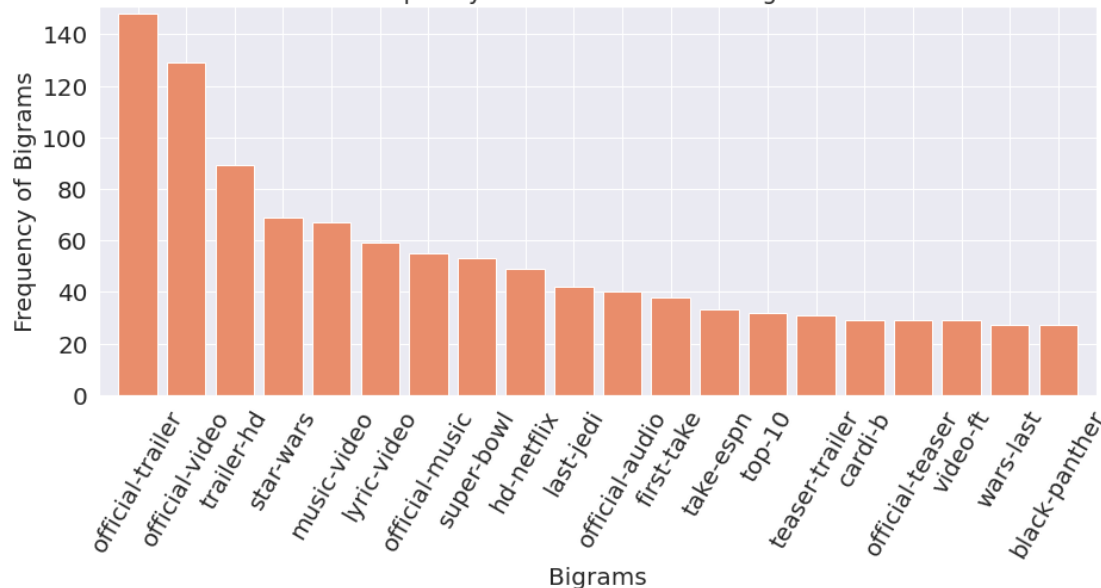
Another Categorical Value

Most Common Words in Videos title

Top 100 Most Common Words

Frequency of 20 Most Common Words

Frequency of Words

Words: official, 2018, video, trailer, vs, 2017, new, ft, live, first, hd, 2, audio, music, day, game, show, makeup, 1, star

Most Common Bigrams in Videos title

Top 50 Most Common Bigrams

iphone-x honest-trailers top-10 part-1
first-look lyric-video lebron-james official-lyric
teaser-trailer black-panther
star-wars golden-globes
cardi-b falcon-heavy music-video
graham-norton kevin-hart

official-trailer

hd-netflix camila-cabello new-york
avengers-infinity season-2 official-audio
bon-appétit take-espn
official-video behind-scenes
norton-show super-bowl
game-highlights shawn-mendes
bbc-news trailer-hd vanity-fair
last-jedi official-teaser
bowl-commercial nbc-news official-music
infinity-war american-idol makeup-tutorial
first-time justice-league
music-awards video-ft plays-night first-take wars-last

Frequency of 20 Most Common Bigrams

Most Common Bigrams in Videos title

Top 50 Most Common Bigrams

Frequency of 20 Most Common Trigrams
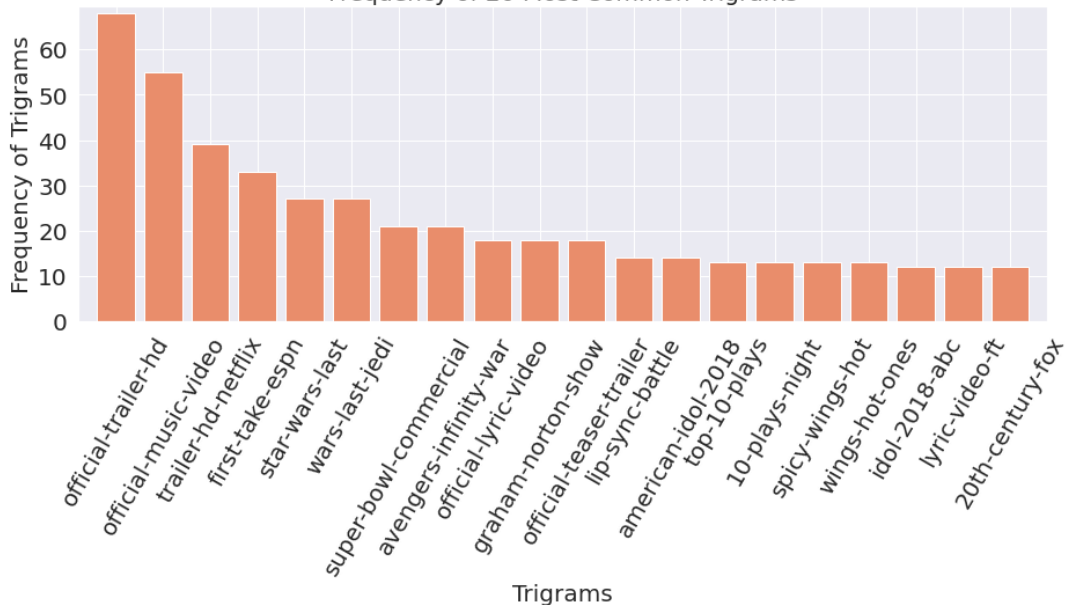
Based on Bigrams and Trigrams, we can conclude that the most combinations of words focus on **film trailers and music videos**, for example, the films **Star Wars: The Last Jedi and Avengers: Infinity Wars**.

Most Common Words in Videos Description

Top 100 Most Common Words

Frequency of 20 Most Common Words

Most Common Bigrams in Videos Description
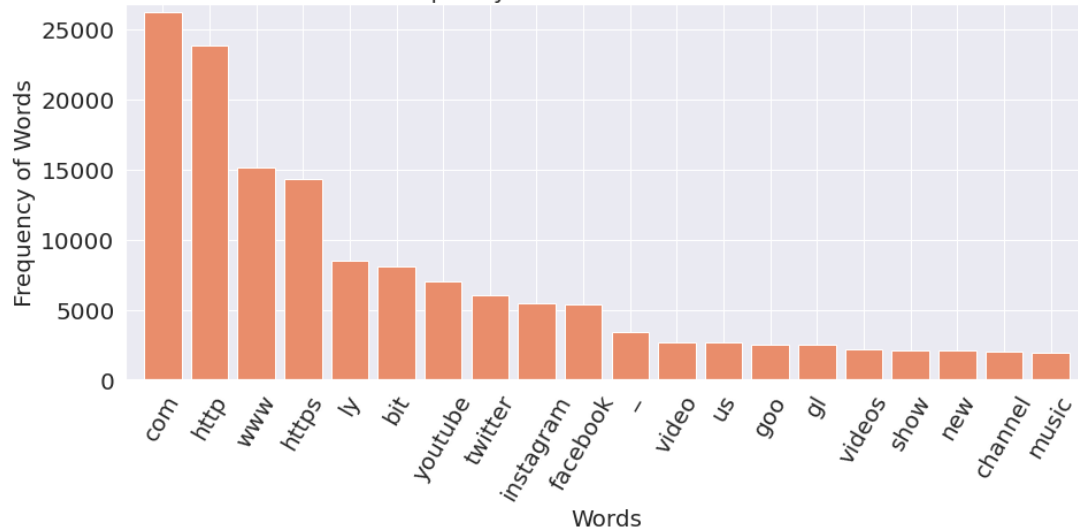
Top 50 Most Common Bigrams

Frequency of 20 Most Common Bigrams
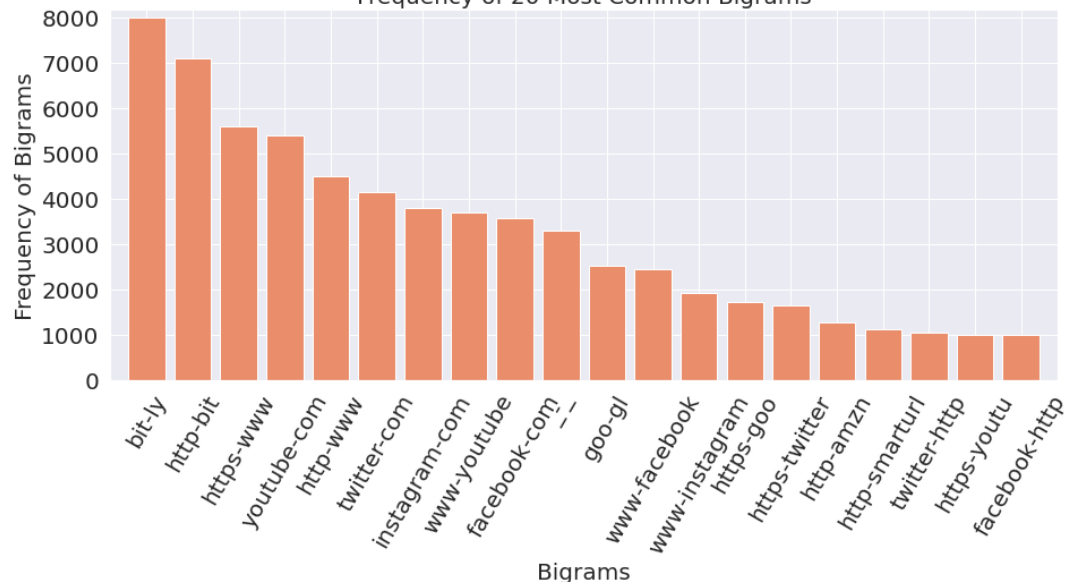
In the video description, most users include their **social media links such as Google, Facebook, Instagram, and Twitter or other things that require links**, for example download links and product links.

Most Common Words in Videos Tags

Top 100 Most Common Tags

Frequency of 20 Most Common Tags
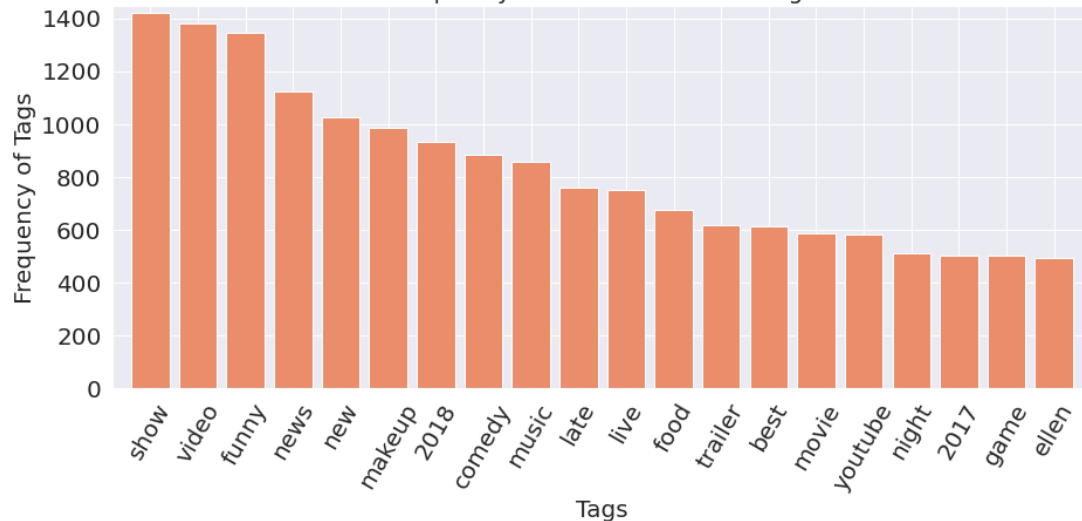
most tags are show, video and funny

Most Common Bigrams in Videos Tags

Top 50 Most Common Bigrams

Frequency of 25 Most Common Bigrams
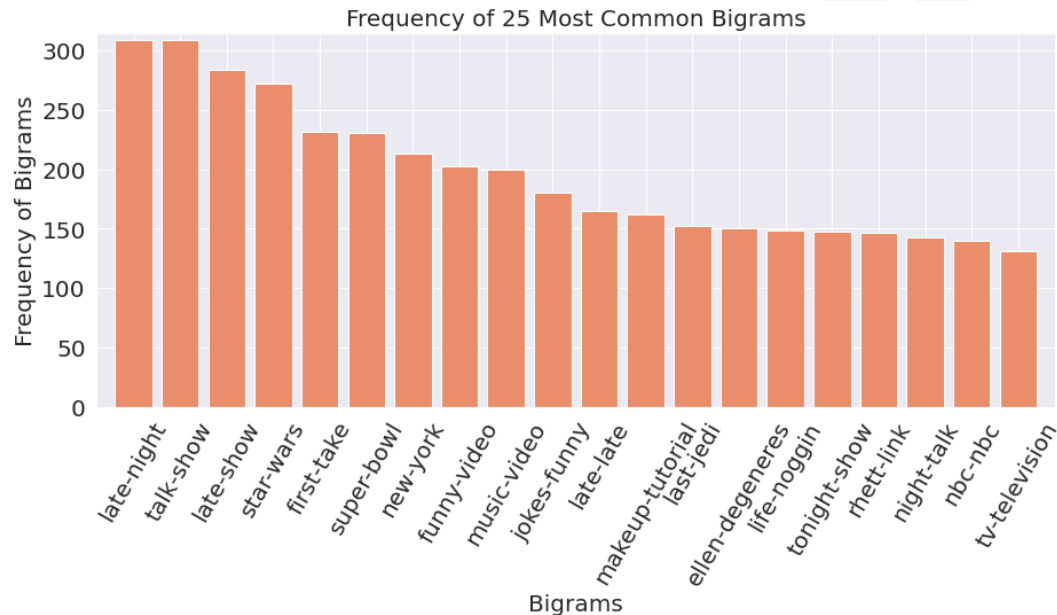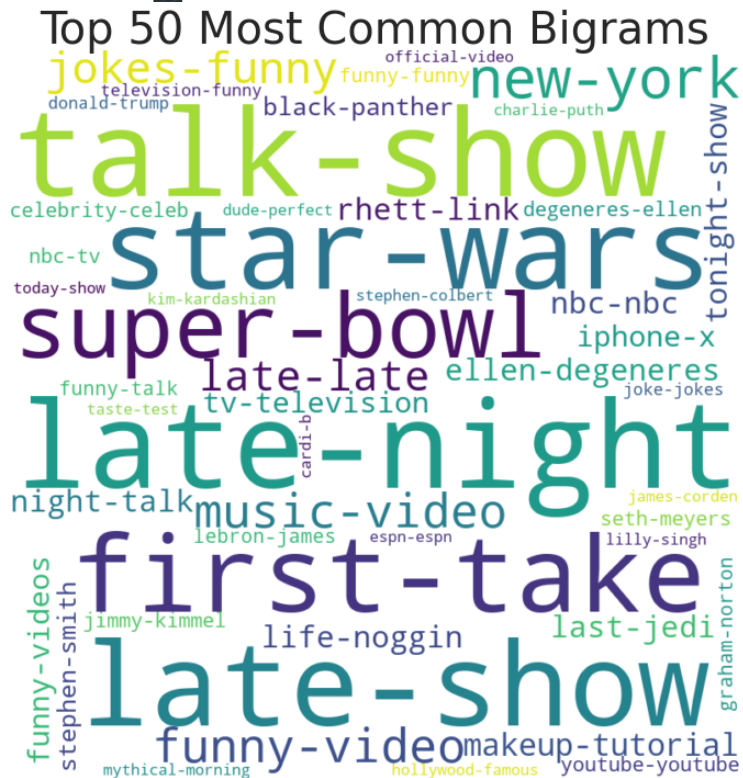
Video Comments Distribution

Comments Count Distribution

Comments Distribution by Category
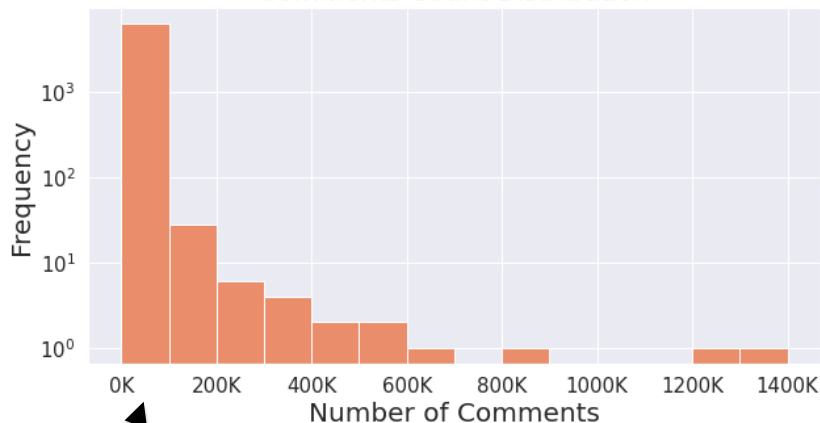
- From the distribution we can see that the **majority of trending videos have less than 100k comments. 99.27%** of trending videos have less than 100k comments and **90%** of trending videos have less than 11,775 comments.

- **Gaming is the category with the highest average number of comments** followed by Music, Comedy, and Shows.
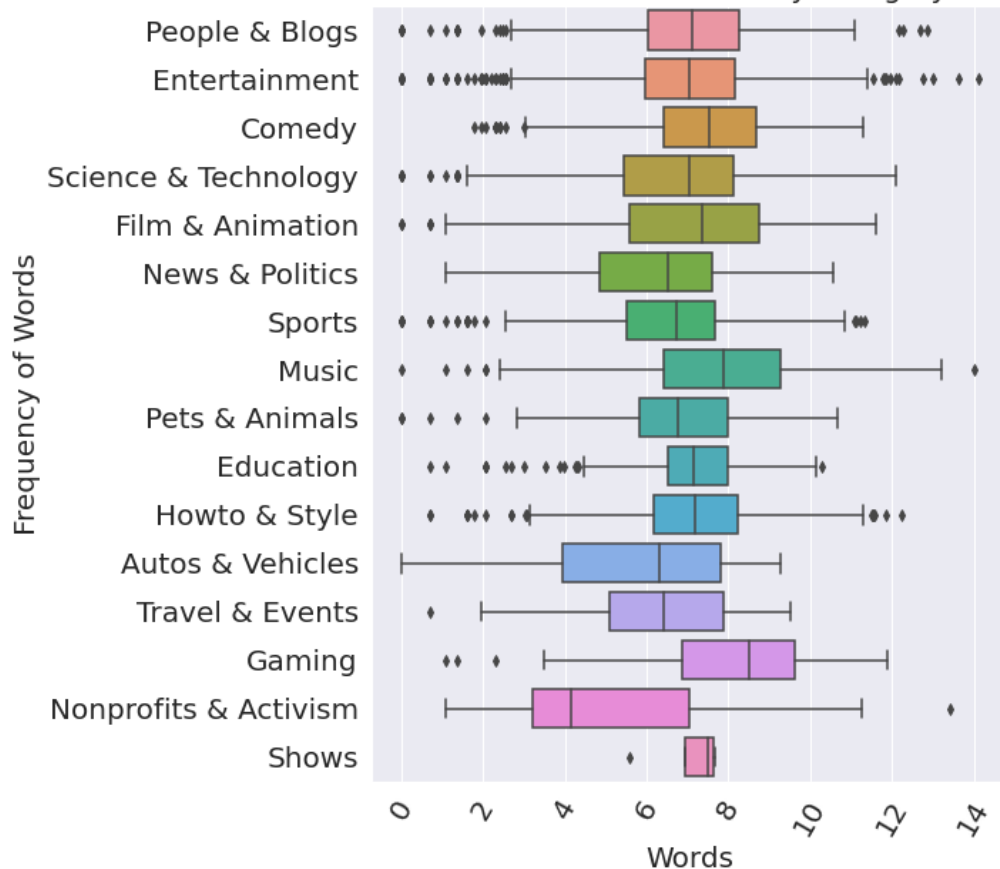
Top 5 Trending Videos Based on Number of Comments

Top 5 Trending Videos Based on Number of Comments

Title	

So Sorry.

BTS (방탄소년단) 'FAKE LOVE' Official MV

YouTube Rewind: The Shape of 2017 | #YouTubeRewind

Suicide: Be Here Tomorrow.

BTS (방탄소년단) 'MIC Drop (Steve Aoki Remix)' Official MV

0K   200K   400K   600K   800K   1000K   1200K   1400K
Number of Comments
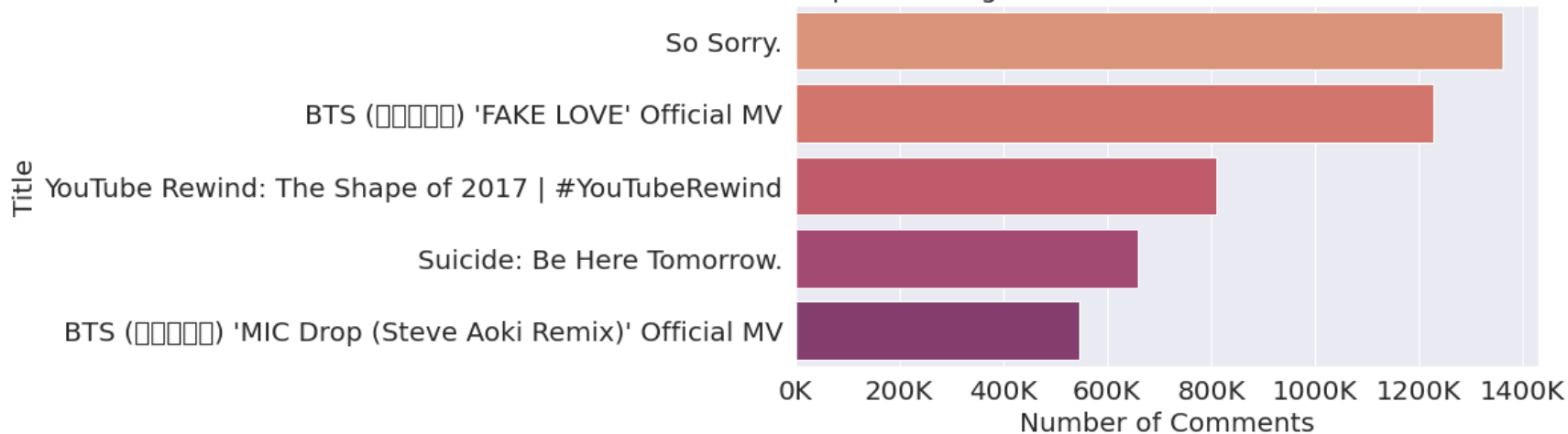
We can notice that **only two videos that have more than 1 million comments**, BTS 'Fake Love' Music Video and So Sorry from Logan Paul.

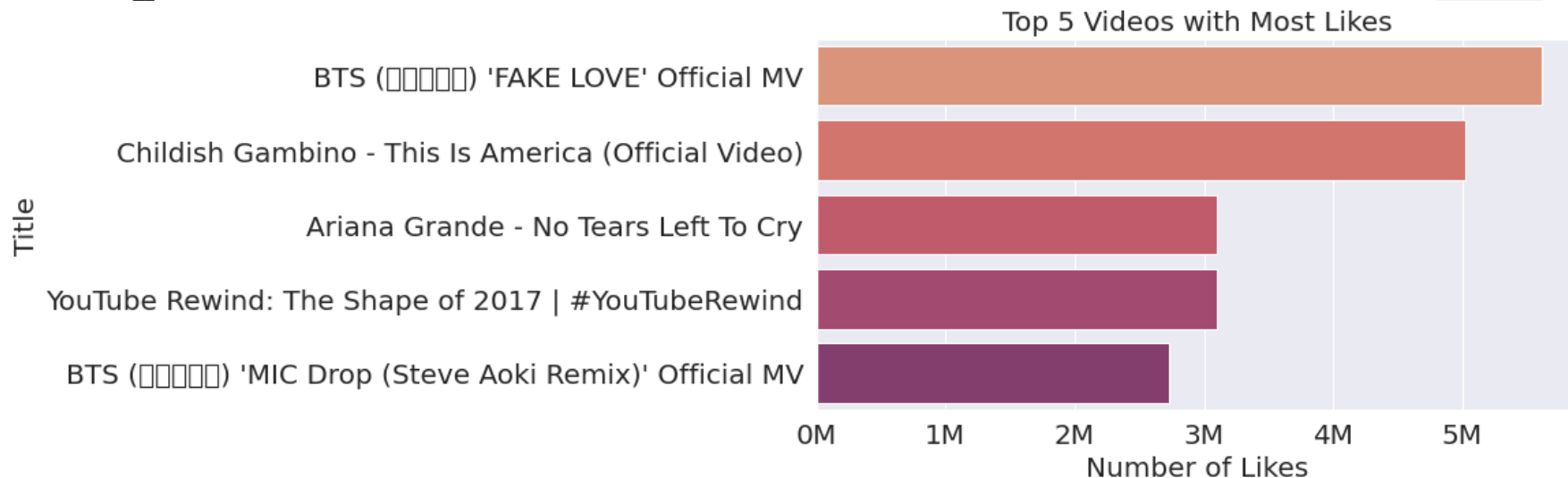# Video Likes Distribution

Likes Distribution

Likes Distribution by Category
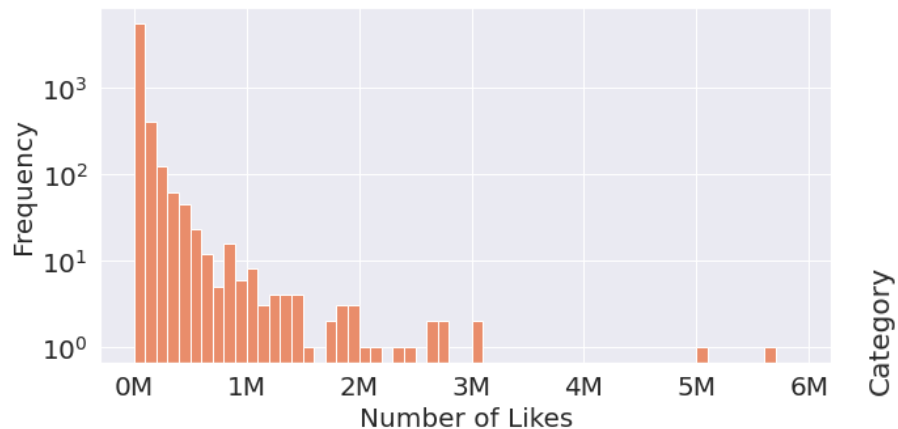
- **88.33%** of trending videos have less than 100k likes

- **Music and Gaming is the category with the highest average number of likes**

Top 5 Trending Videos Based on Number of Likes
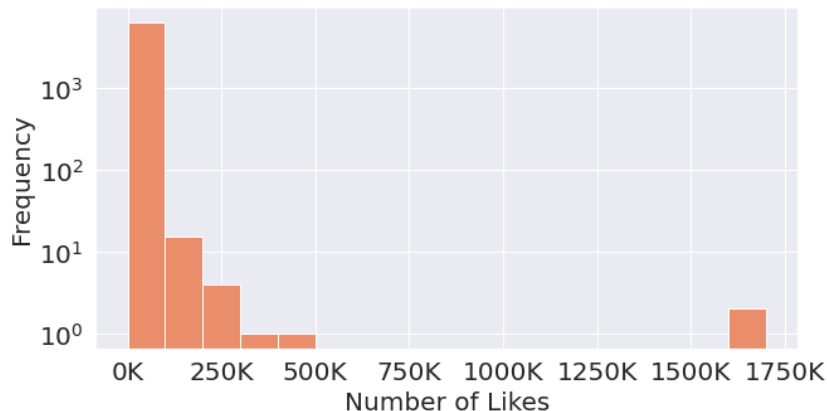
Top 5 Videos with Most Likes
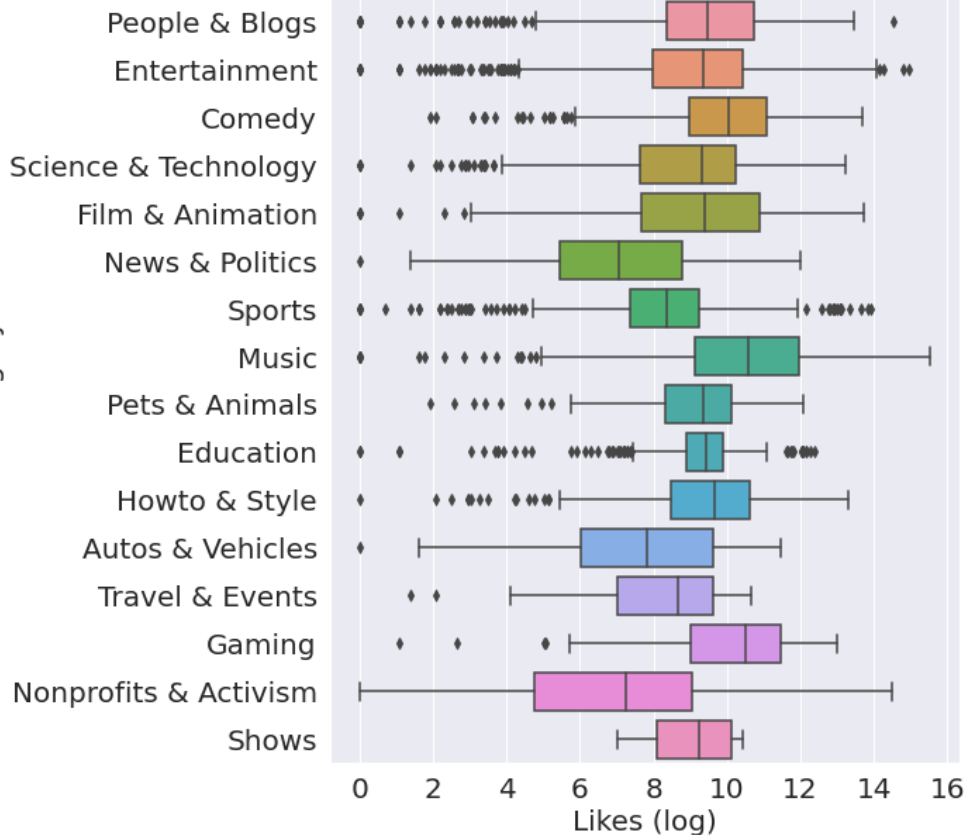
Videos with **largest number of likes is BTS 'Fake Love' Music Video followed by Childish Gambino 'This Is America' Music Video with over 5 million likes.**
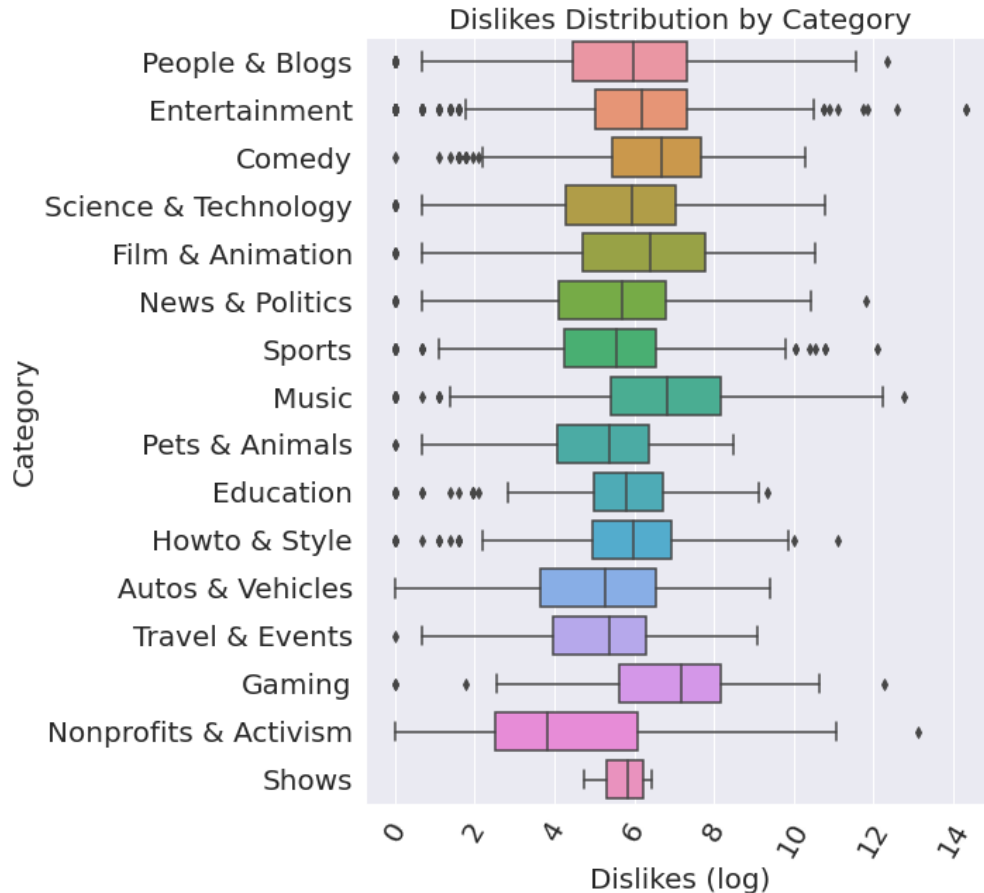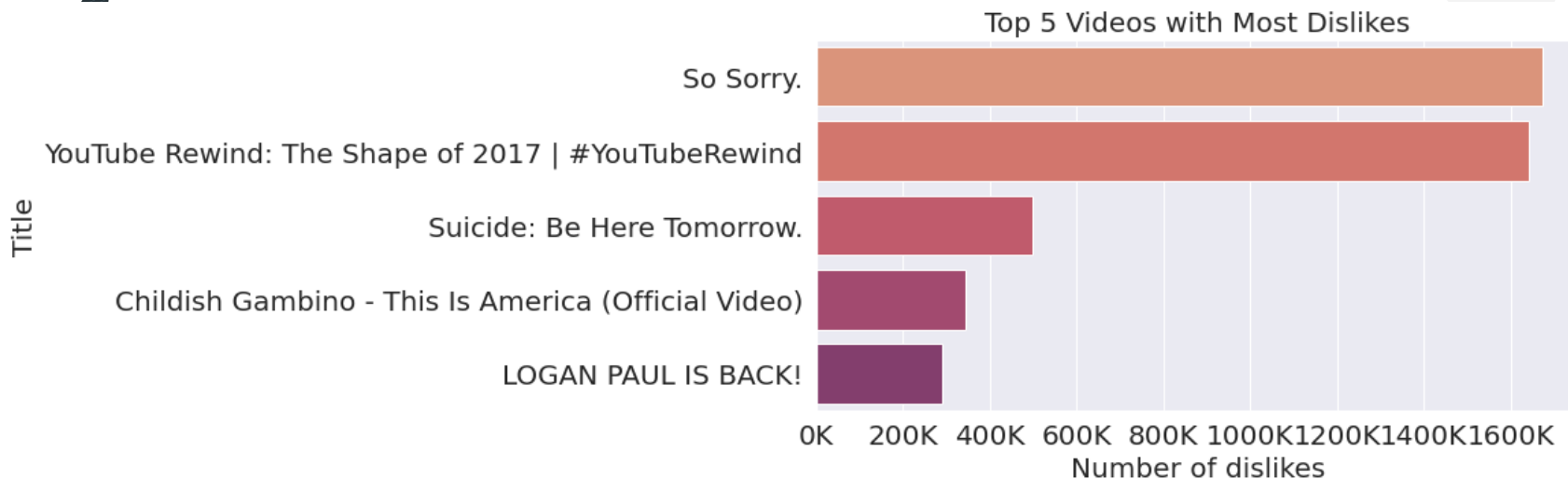
Video Dislikes Distribution

Dislikes Distribution

Dislikes Distribution by Category

- **95.68%** of trending videos have less than 10k dislikes

Top 5 Trending Videos Based on Number of Dislikes
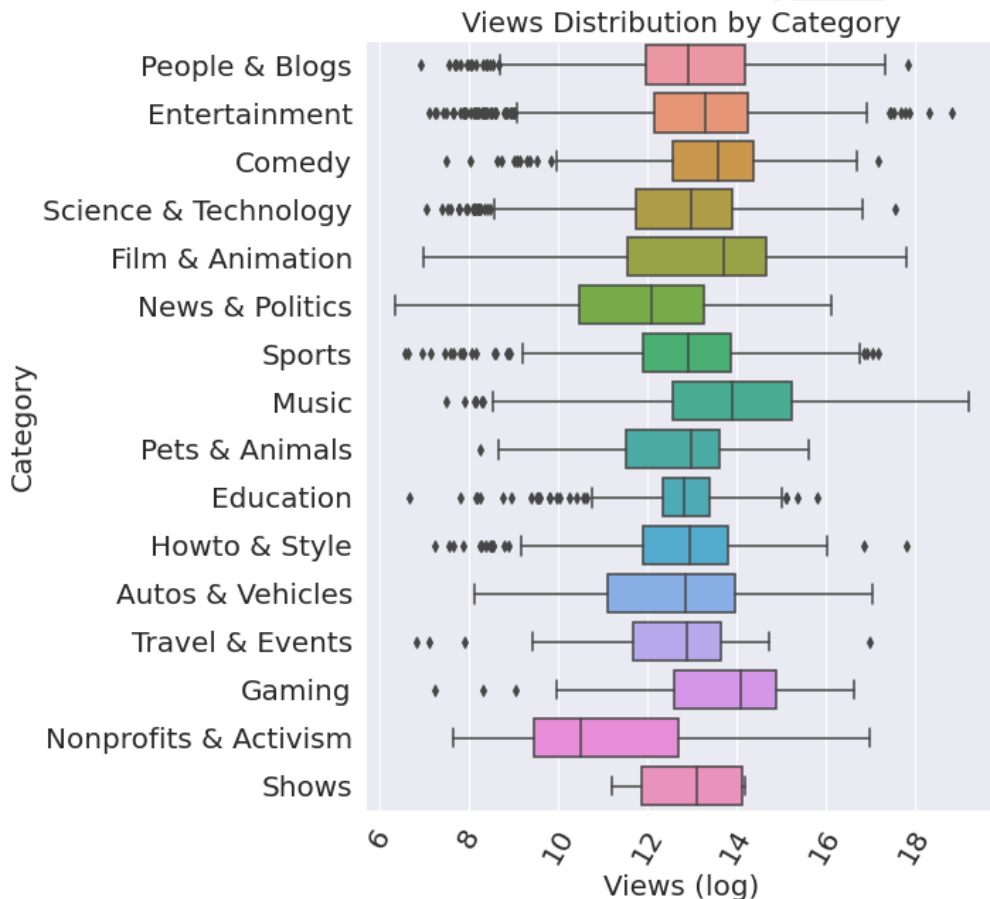

Top 5 Videos with Most Dislikes

Videos with **largest number of likes is So Sorry from Logan Paul followed by Youtube Rewind 2017.**

Video Views Distribution

Views Distribution

Views Distribution by Category
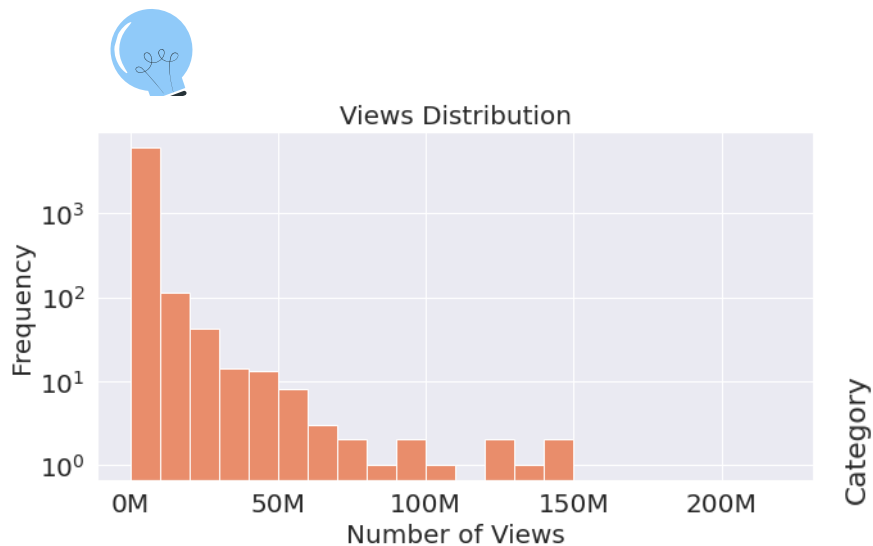
- **92.89%** of trending videos have **less than 5 million views** and **80%** of the videos have less than **1.88 Million views.**
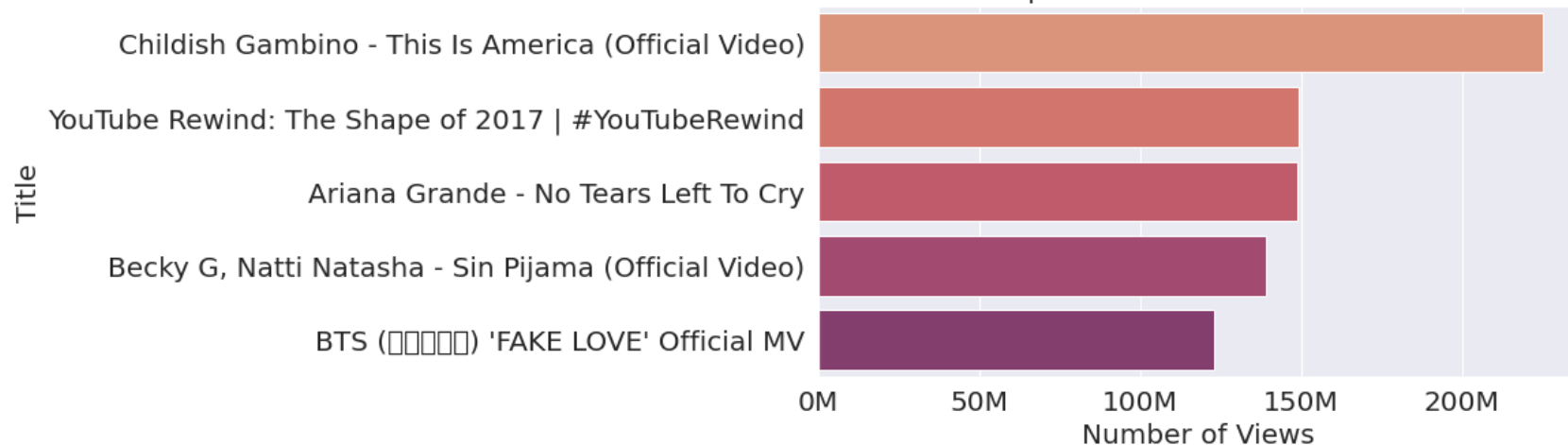
- **Gaming is the category with the highest average number of likes, followed by Music**

Top 5 Trending Videos Based on Number of Dislikes
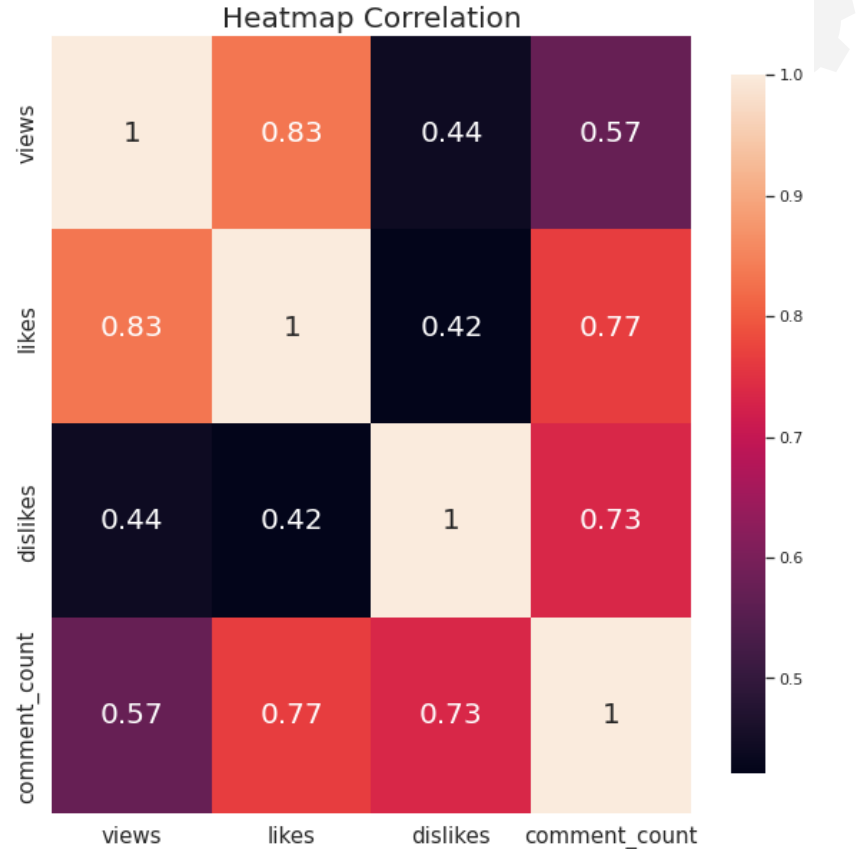
Top 5 Videos with Most Views

Videos with **largest number of views is Childish Gambino 'This Is America' Music Video, with 225 Million Views**

Relationship of the Engagement Metrics

- based on the heatmap, **all correlations from numerical data are positive**, **Likes and views** have the largest correlation value of **0.83.**


Heatmap Correlation

Number of days between trending and upload

	upload_to_trending
count	6351.000000
mean	22.522280
std	209.963296
min	0.000000
25%	1.000000
50%	2.000000
75%	3.000000
max	4215.000000

Distribution of Number of days between trending and upload

- **73.51%** of videos become trending in less than 3 days

- only **1.88%** of videos are trending on the same day as when they were uploaded

How long those videos trending
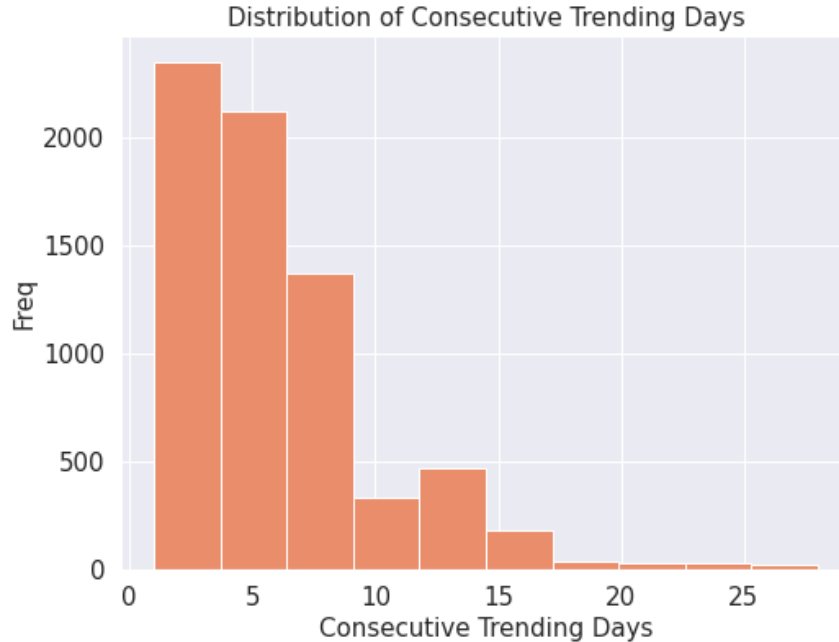
Distribution of Total Trending Days
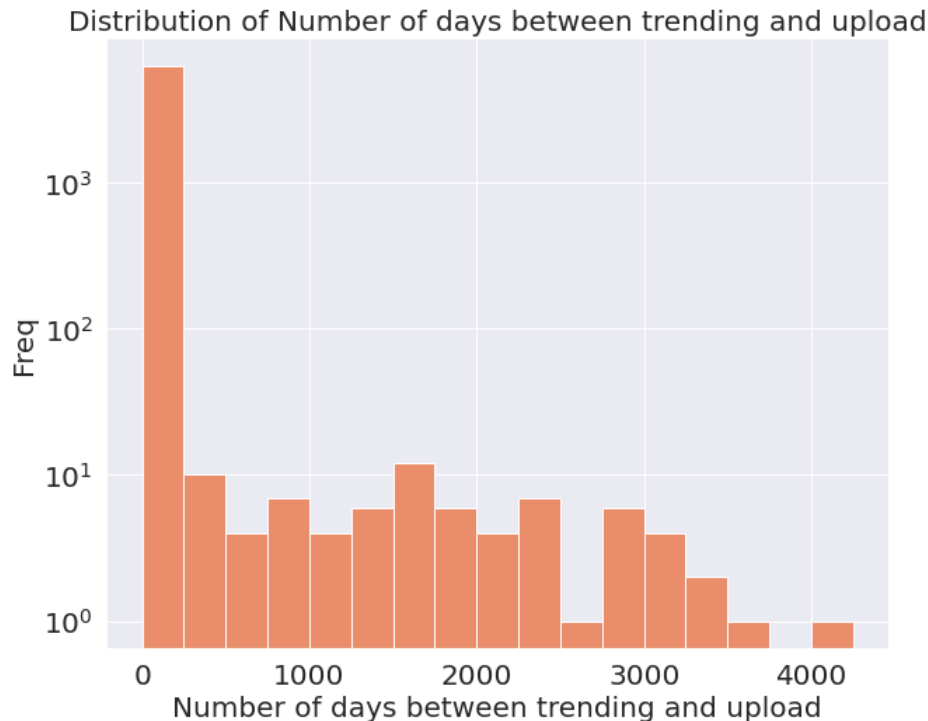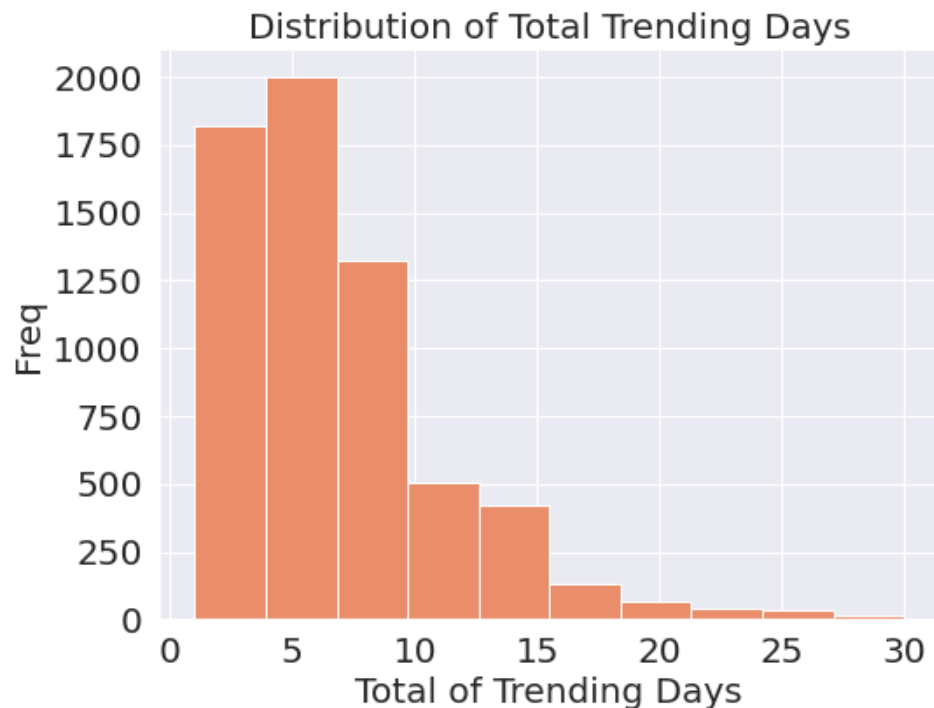
	trending_days
count	6351.000000
mean	6.447646
std	4.642806
min	1.000000
25%	3.000000
50%	6.000000
75%	8.000000
max	30.000000

Consecutive Trending Days

Distribution of Consecutive Trending Days

	consecutive_trending_days
count	6959.000000
mean	5.884322
std	4.367000
min	1.000000
25%	3.000000
50%	5.000000
75%	8.000000
max	28.000000

- **There are 7 trending videos for 28 consecutive days**
- **the average video is only trending for 5-6 days in a row**

Topic Modeling

Cluster of word by topic

Collection of Text Documents

Topic Modeling

Cluster of document by topic
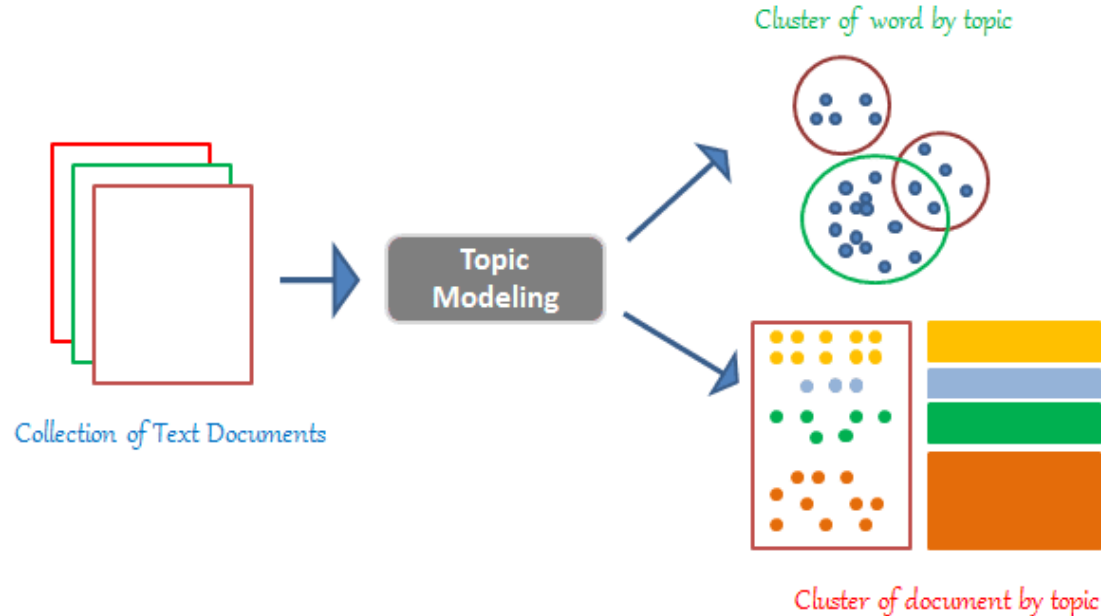
Topic modeling is a type of statistical modeling for discovering the abstract "topics" that occur in a collection of documents. **Latent Dirichlet Allocation** (LDA) is an example of topic model and is used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distributions.
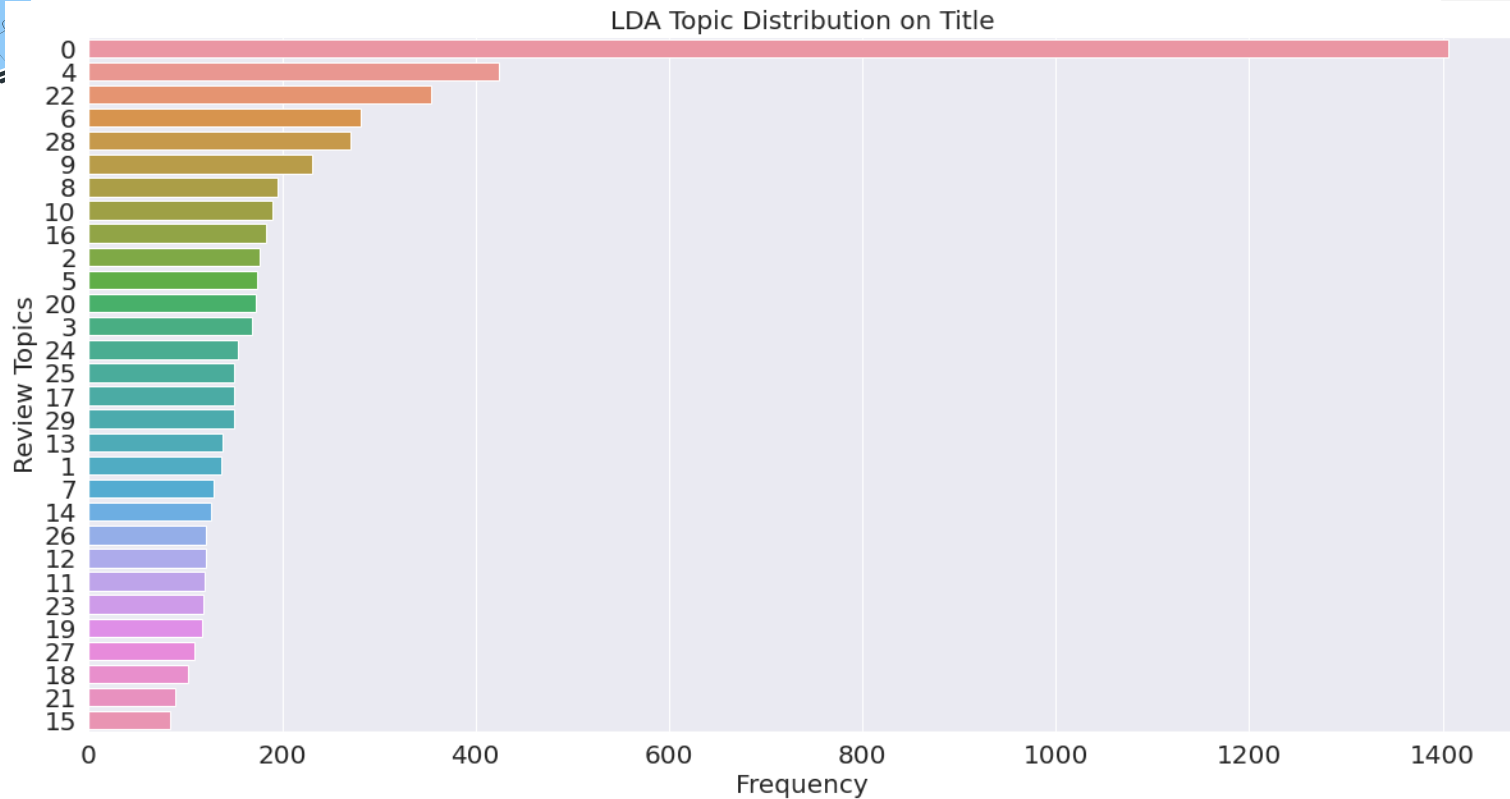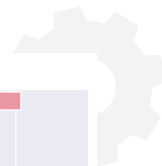
Topic Group for Title

```
Topic 0: new movie 10 year bad team old star man find
Topic 1: go christmas every day first new time year ever never
Topic 2: game black play man come say woman golden panther hair
Topic 3: world test big perfect minute bbc film fall first christmas
Topic 4: video official ft try lyric music part explain tutorial family
Topic 5: star war top house nfl tour episode apple miss avenger
Topic 6: make best diy food battle interview work kevin cut stop
Topic 7: react use song mv iphone google hot make bts live
Topic 8: 2017 full know youtube great tb face like react use
Topic 9: get makeup week name friend car ad bts star perfect
Topic 10: live news watch stephen smith call president moment harry bowl
Topic 11: look night inside red dead high film lip first bbc
Topic 12: like scene secret people behind hit final show get refinery29
Topic 13: take last kid jedi school dance camila secret show star
Topic 14: love cake baby hot kim announcement grace name vs know
Topic 15: voice fire date justin lose cast wire kim national iphone
Topic 16: first time super bowl meet nba everything story girl love
Topic 17: one life thing buy march fox ad second vs bbc
Topic 18: ever chris national find never get hot christmas grace weird
Topic 19: vs real girl question vogue bts tv wire tom answer
Topic 20: show dog open snl white back hart youtube nbc fall
Topic 21: award american music win nbc fall let performance 2017 kim
Topic 22: trailer official hd talk season netflix teaser fan honest cook
Topic 23: espn john break giant shoot theory eat go netflix kevin
Topic 24: challenge highlight review box league grace adam audition food makeup
Topic 25: audio beauty feat official see cardi cover teaser make live
Topic 26: cat commercial car sport ice mom lip bowl challenge get
Topic 27: james face refinery29 holiday michael lebron olympics fall hot fire
Topic 28: 2018 trump wild give special jimmy need want nbc trailer
Topic 29: day home ball car minute use movie dog special let
```

Topic with 10 most frequent words

Topic Group for Title

LDA Topic Distribution on Title
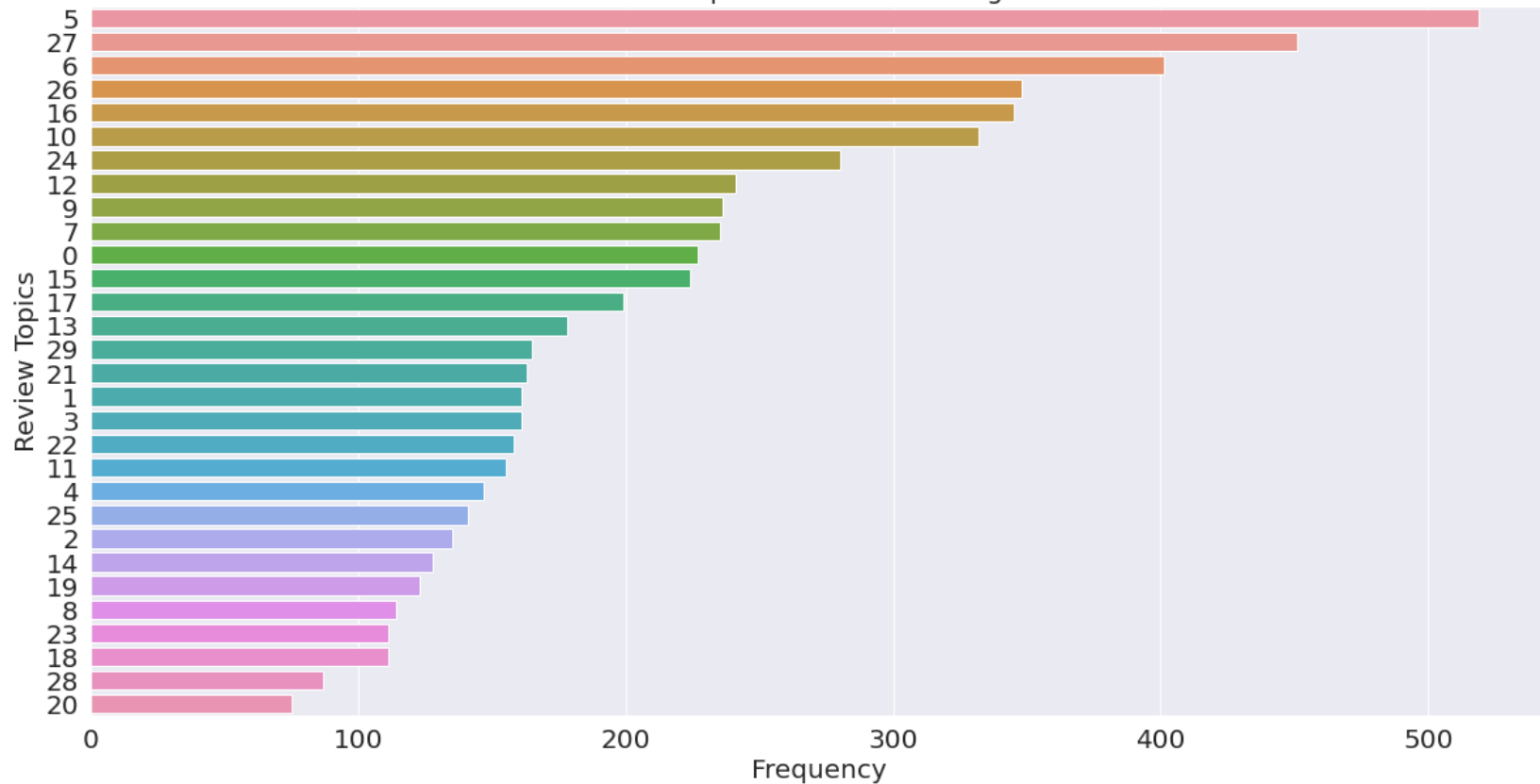
Topic Group for Tags

Topic 0: school high hack ever standup honest learn blue tesla amaze
Topic 1: animal take big king jack documentary great wild fry jr
Topic 2: espn perfect go charlie health bbc nick ted heart way
Topic 3: challenge react kid google liza reaction nfl brother fine know
Topic 4: make diy disney fox year box nintendo homemade world hannah
Topic 5: news smith nba trump basketball game animation live lebron break
Topic 6: interview sport celebrity none football entertainment game tonight magazine soccer
Topic 7: tv television car netflix series original john watch movie online
Topic 8: test voice buzzfeed food team try people taste best eat
Topic 9: 2017 live christmas award red 2018 holiday seth best meyers
Topic 10: makeup tutorial beauty black get hair face pet random video
Topic 11: family girl noggin funny anthony friendly fun boy robot lol
Topic 12: food recipe cook cake water kitchen hot bake world chef
Topic 13: star war last jedi bts cardi mark scene johnson explain
Topic 14: super bowl commercial chocolate green ice ad pizza money power
Topic 15: ellen dog man marvel talent show date fan degeneres panther
Topic 16: movie trailer 2018 film official theory new royal american ryan
Topic 17: life vlog house day home tour alex real fire prank
Topic 18: video cat show nbc today funny amazon kitten cute race
Topic 19: first time olympics woman ball winter 2018 meet evans korea
Topic 20: youtube channel simon education grace steve party audition press reveal
Topic 21: new science one york vs space city video adam travel
Topic 22: fashion style room trick trend secret every work fall like
Topic 23: night live snl sketch chris comedy season episode taylor host
Topic 24: game iphone review apple top 10 art design tech best
Topic 25: baby kardashian kim link good rhett news morning parody mythical
Topic 26: show late funny comedy talk stephen video joke humor celebrity
Topic 27: music video song pop record official play new love lyric
Topic 28: highlight fallon james graham lip full kelly battle bad sing
Topic 29: dance thing wwe kevin street la shot wrestle camera lilly

Topic with 10 most frequent Tags
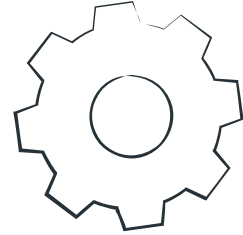
Topic Group for Tags

LDA Topic Distribution on Tags

References

1. https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0
2. https://towardsdatascience.com/preprocessing-text-data-using-python-576206753c28
3. https://towardsdatascience.com/nlp-part-3-exploratory-data-analysis-of-text-data-1caa8ab3f79d
4. https://ammar-alyousfi.com/2020/youtube-trending-videos-analysis-2019-us?src=kgl#trending-videos-and-comments

Thanks!

rahmatmamat48@gmail.com