YouTube Trending

Sogra Memon, Megan McGlashan, Bincy Narath

Opportunity

- **Trending on YouTube could be a good representative of:**
  - What most people like watching in each country

- **Marketing opportunities:**
  - Content creators
  - Advertisers
  - Brand Managers

Purpose

Regional Analysis > >

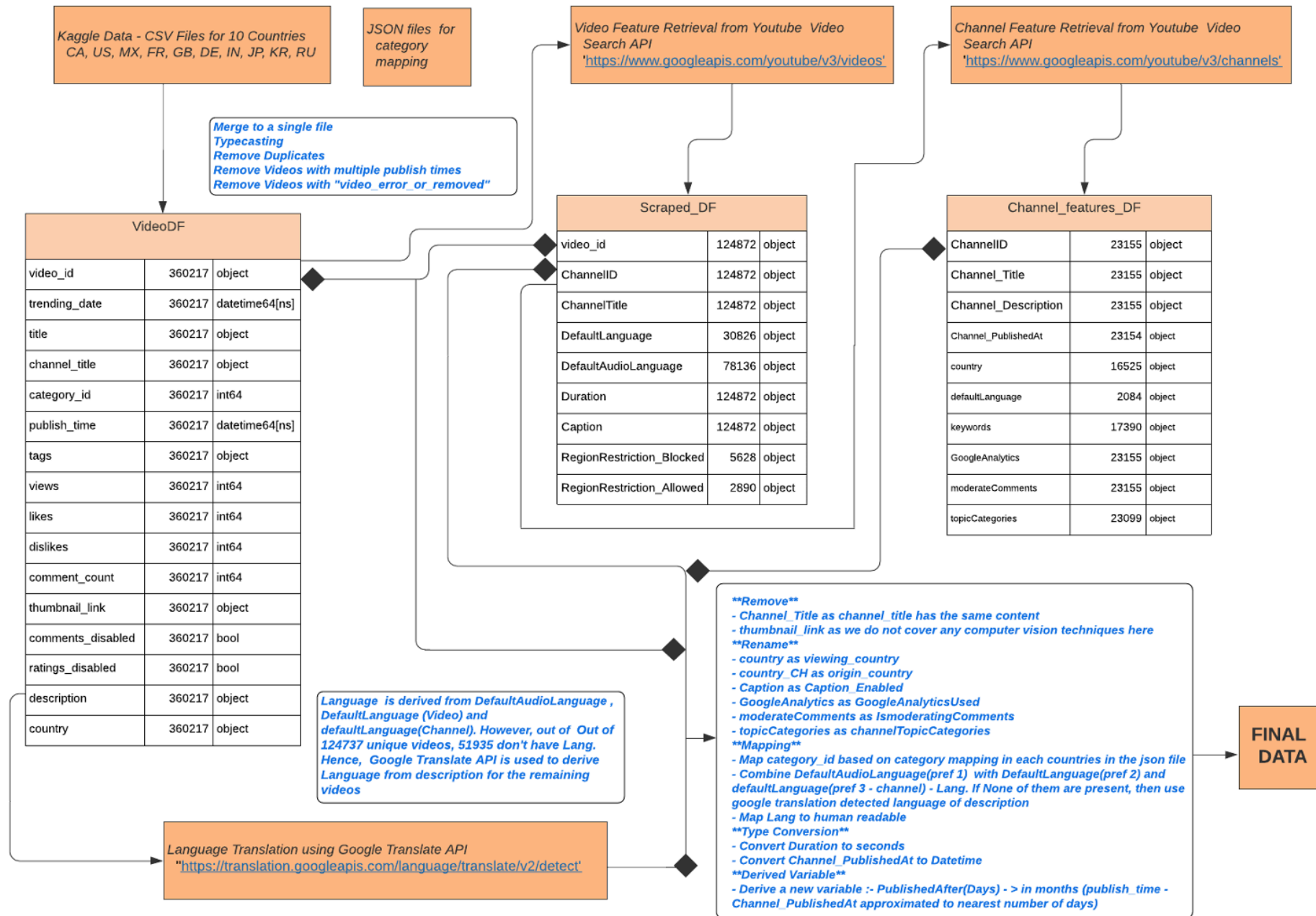How does trending in different countries impact one another?

Popularity Analysis > >

What impacts how long videos trend on YouTube?
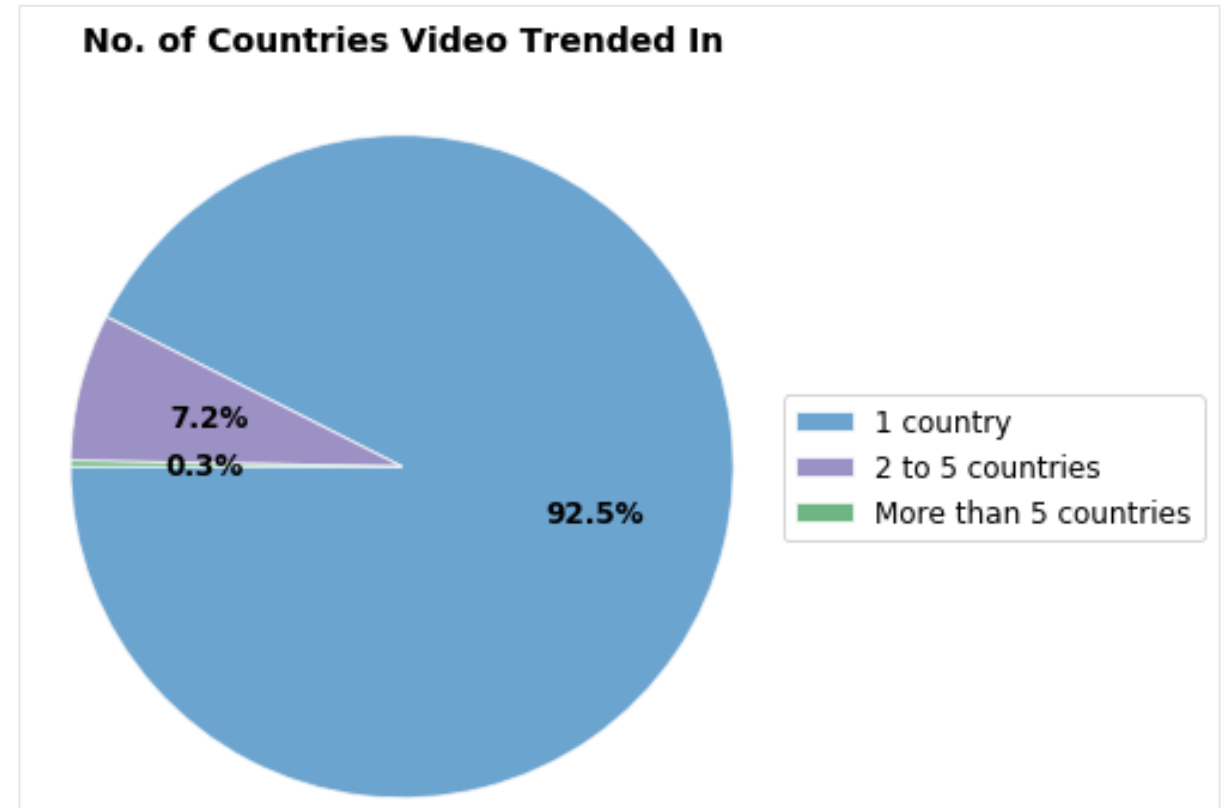
Data Sources

Data Modeling

Kaggle Data - CSV Files for 10 Countries *CA, US, MX, FR, GB, DE, IN, JP, KR, RU*

JSON files for category mapping

Video Feature Retrieval from Youtube Video Search API 'https://www.googleapis.com/youtube/v3/videos'

Channel Feature Retrieval from Youtube Video Search API 'https://www.googleapis.com/youtube/v3/channels'

Merge to a single file
Typecasting
Remove Duplicates
Remove Videos with multiple publish times
Remove Videos with "video_error_or_removed"

VideoDF

video_id	360217	object
trending_date	360217	datetime64[ns]
title	360217	object
channel_title	360217	object
category_id	360217	int64
publish_time	360217	datetime64[ns]
tags	360217	object
views	360217	int64
likes	360217	int64
dislikes	360217	int64
comment_count	360217	int64
thumbnail_link	360217	object
comments_disabled	360217	bool
ratings_disabled	360217	bool
description	360217	object
country	360217	object

Scraped_DF

video_id	124872	object
ChannelID	124872	object
ChannelTitle	124872	object
DefaultLanguage	30826	object
DefaultAudioLanguage	78136	object
Duration	124872	object
Caption	124872	object
RegionRestriction_Blocked	5628	object
RegionRestriction_Allowed	2890	object

Channel_features_DF

ChannelID	23155	object
Channel_Title	23155	object
Channel_Description	23155	object
Channel_PublishedAt	23154	object
country	16525	object
defaultLanguage	2084	object
keywords	17390	object
GoogleAnalytics	23155	object
moderateComments	23155	object
topicCategories	23099	object

Language is derived from DefaultAudioLanguage, DefaultLanguage (Video) and defaultLanguage(Channel). However, out of Out of 124737 unique videos, 51935 don't have Lang. Hence, Google Translate API is used to derive Language from description for the remaining videos

Language Translation using Google Translate API "https://translation.googleapis.com/language/translate/v2/detect'

Remove
- *Channel_Title as channel_title has the same content*
- *thumbnail_link as we do not cover any computer vision techniques here*
Rename
- *country as viewing_country*
- *country_CH as origin_country*
- *Caption as Caption_Enabled*
- *GoogleAnalytics as GoogleAnalyticsUsed*
- *moderateComments as IsmoderatingComments*
- *topicCategories as channelTopicCategories*
Mapping
- *Map category_id based on category mapping in each countries in the json file*
- *Combine DefaultAudioLanguage(pref 1) with DefaultLanguage(pref 2) and defaultLanguage(pref 3 - channel) - Lang. If None of them are present, then use google translation detected language of description*
- *Map Lang to human readable*
Type Conversion
- *Convert Duration to seconds*
- *Convert Channel_PublishedAt to Datetime*
Derived Variable
- *Derive a new variable :- PublishedAfter(Days) - > in months (publish_time - Channel_PublishedAt approximated to nearest number of days)*

FINAL DATA

Regional Analysis

Videos Trending in Multiple Countries

Videos rarely trended in multiple > > countries

No. of Countries Video Trended In

- 1 country
- 2 to 5 countries
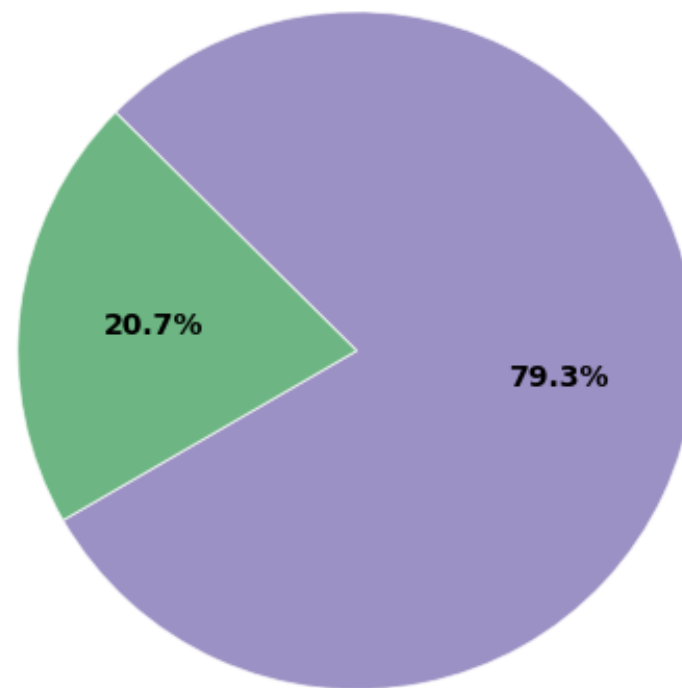- More than 5 countries

7.2%
0.3%
92.5%

Videos Trending in Multiple Countries

Among the few videos that trend in multiple countries

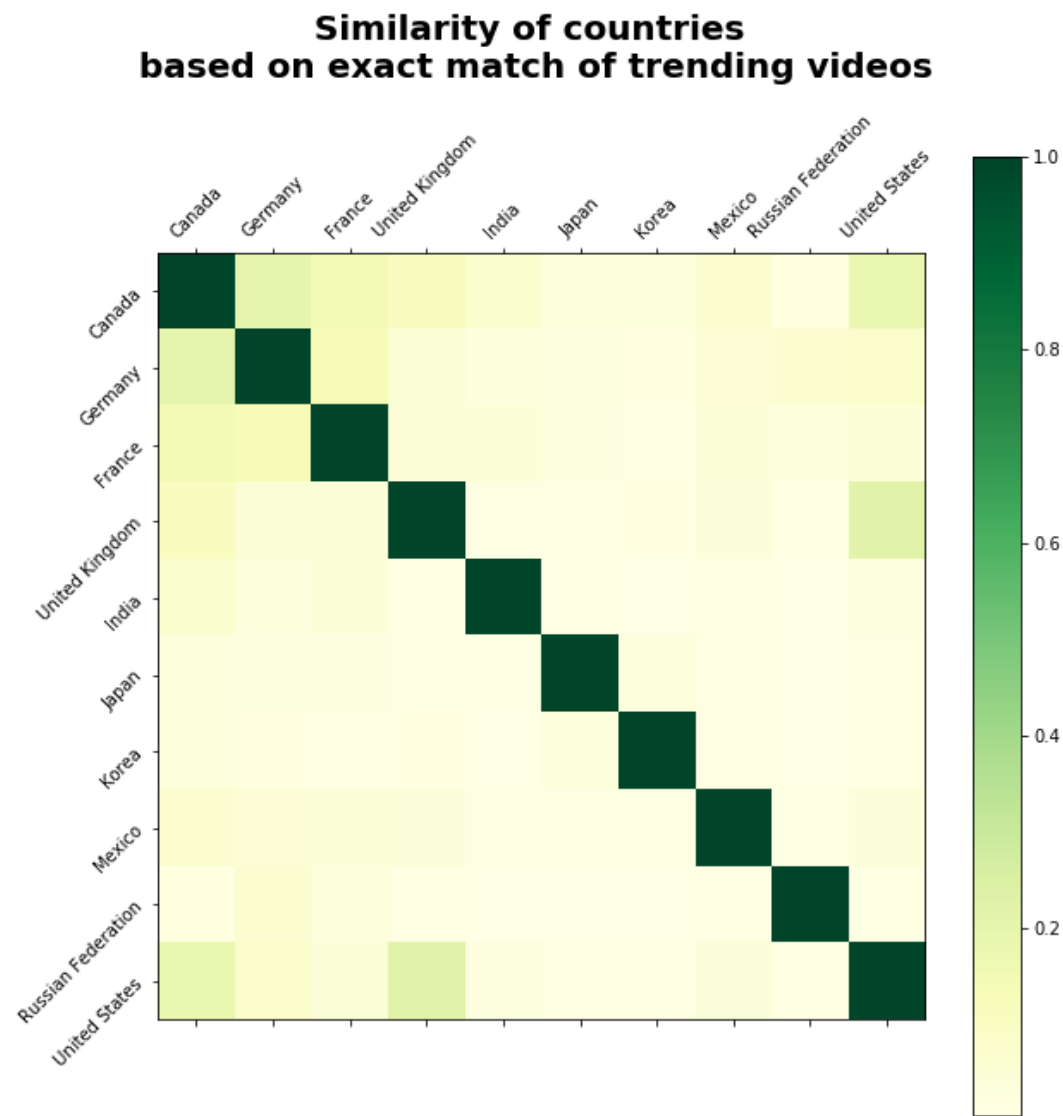**> > The vast majority trend for the first time on the same day across countries**

Whether videos that trended in multiple countries first did so on the same day or different days

20.7%

79.3%

First Trended on the Same Day
First Trended on Different Days (all or some)

Similarity in Trending

Canada watches similar videos to:

> US

> UK

> France

> Germany

Similarity of countries
based on exact match of trending videos

Source of Content

Canada and the **UK** generate little trending content, rather most of their trending videos originate from the **US**.

Known Sources of Videos by Country(%)
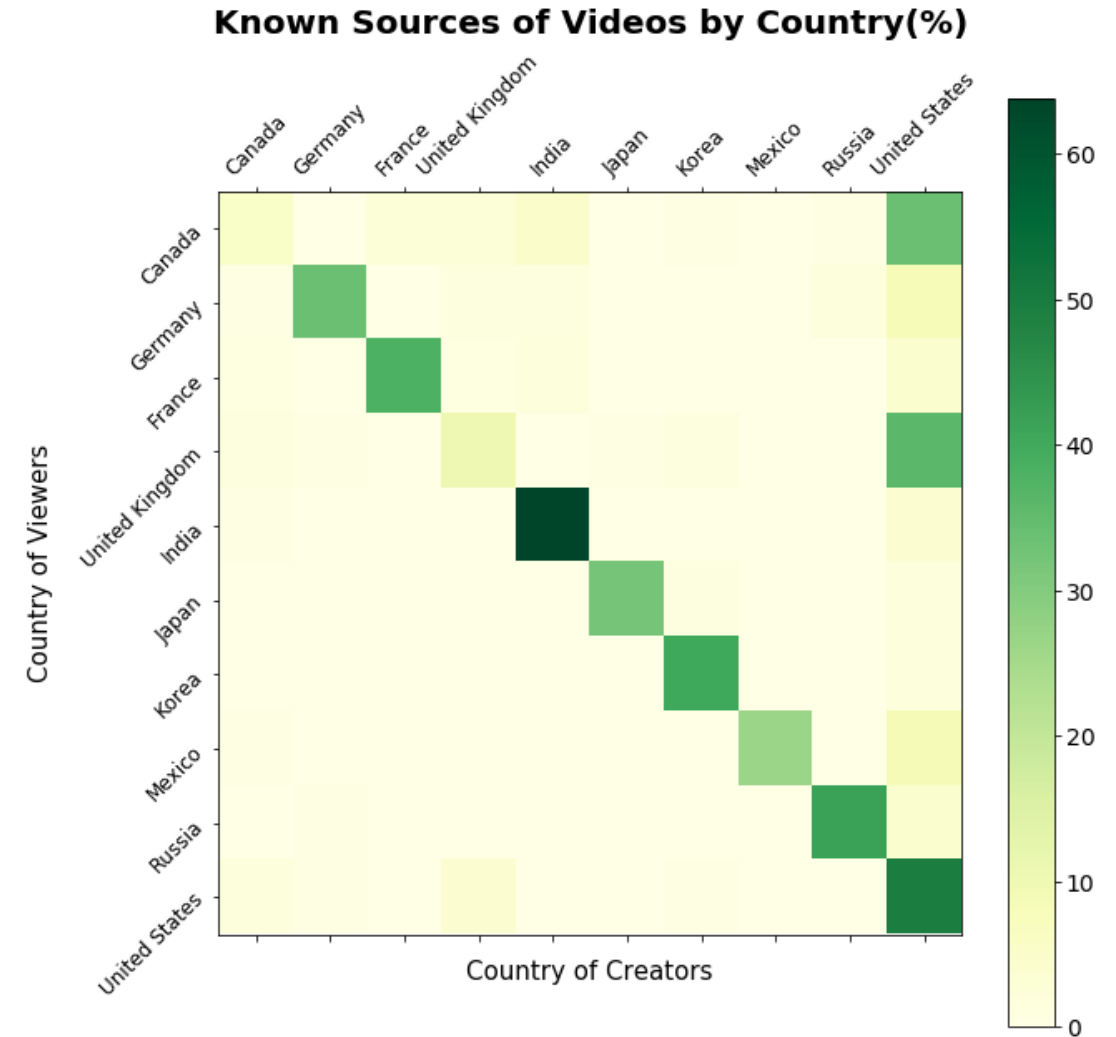
Categories

Entertainment is the most common category in all countries except the UK and Russia.

Trending Videos Visual: Countries & Categories
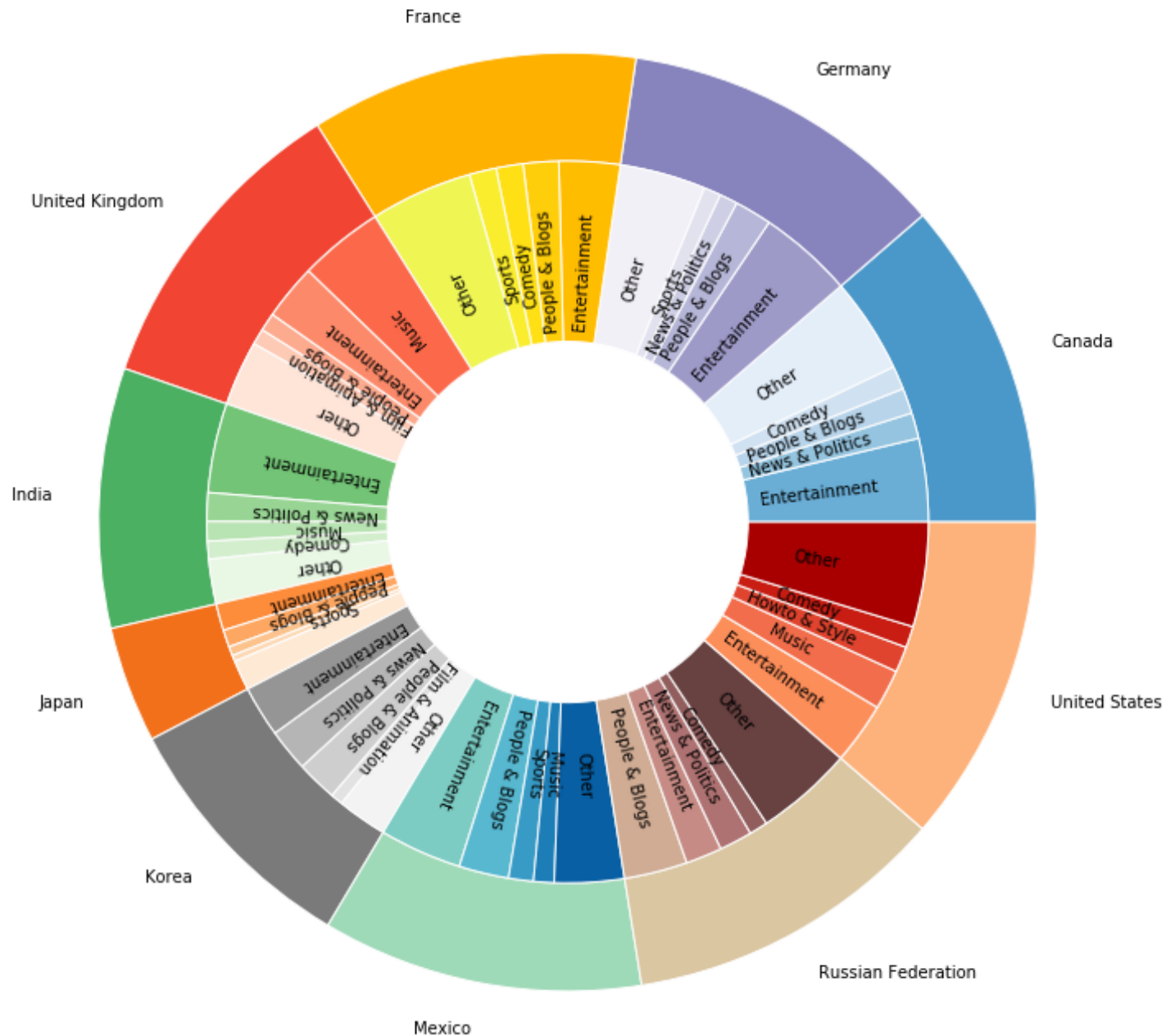
Number of Days Videos Trended

Video typical
trend for longer
>>

in the **UK** and
the **US**

Days Videos Trended by Country

Country	Single Day	2 to 5 Days	More than 5 Days
United Kingdom (N=3,266)	7	17	75
United States (N=6,346)	11	37	51
Korea (N=15,855)	39	60	
India (N=16,302)	47	50	1
Canada (N=24,406)	59	39	
Germany (N=29,613)	72	27	
France (N=30,559)	75	24	
Mexico (N=33,492)	83	16	
Russia (N=34,268)	84	15	
Japan (N=12,901)	88	11	

Percentage of Videos
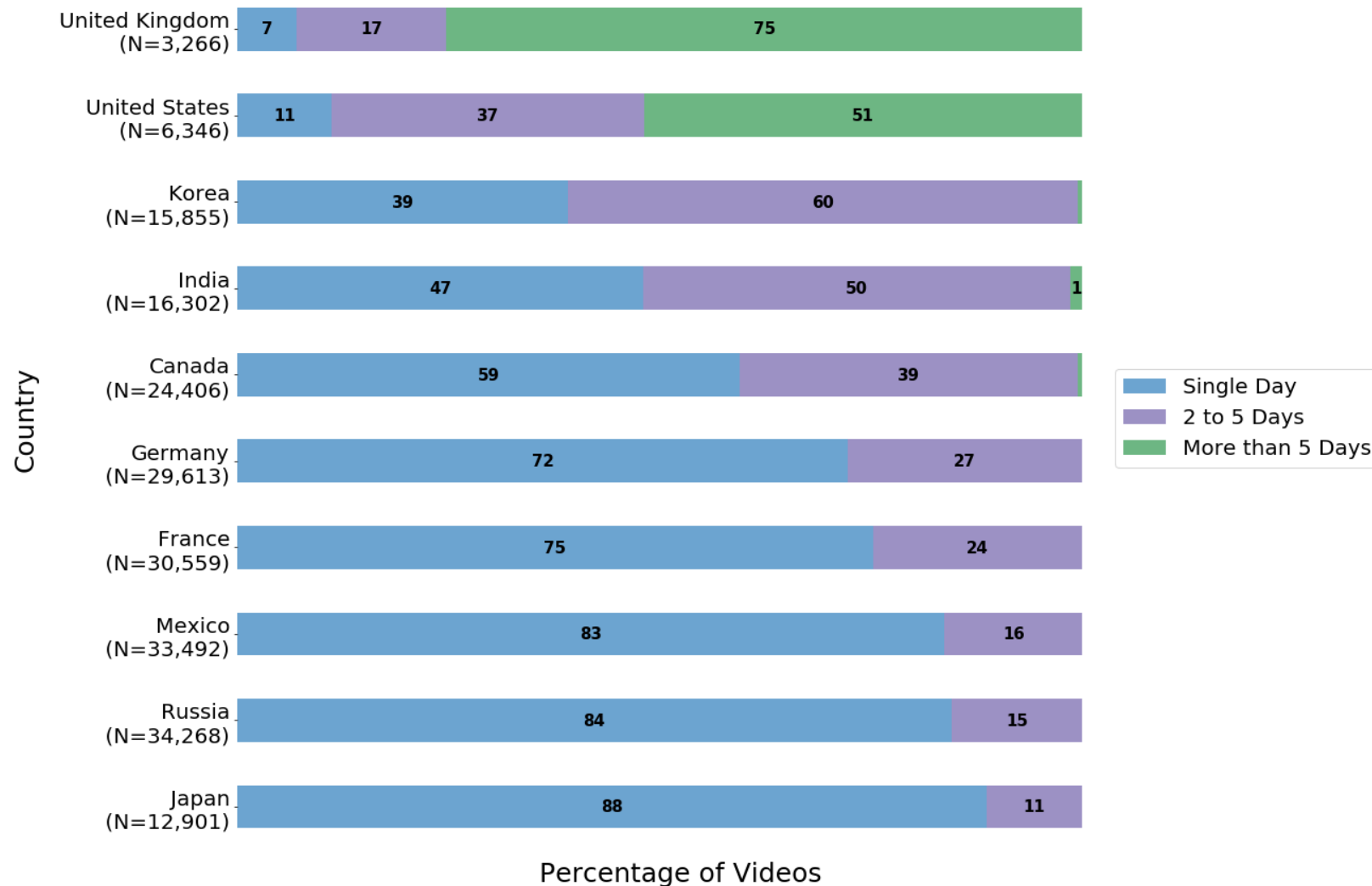
Popularity Analysis

Popularity Prediction

Popularity ⇔ Total Trending (in Days)
Correlation (Continuous Trending, Total Trending) = **0.9869**
99.2% of times there was No gap in trending!

Popularity Prediction Problem
1Day Trending > 'JustTrending'
2 to 5 Days Trending > 'Popular'
6 to 14 Days Trending > 'UlraPopular'
>14 Days Trending > 'Phenomenon'

Total Trending Vs Continuous Trending
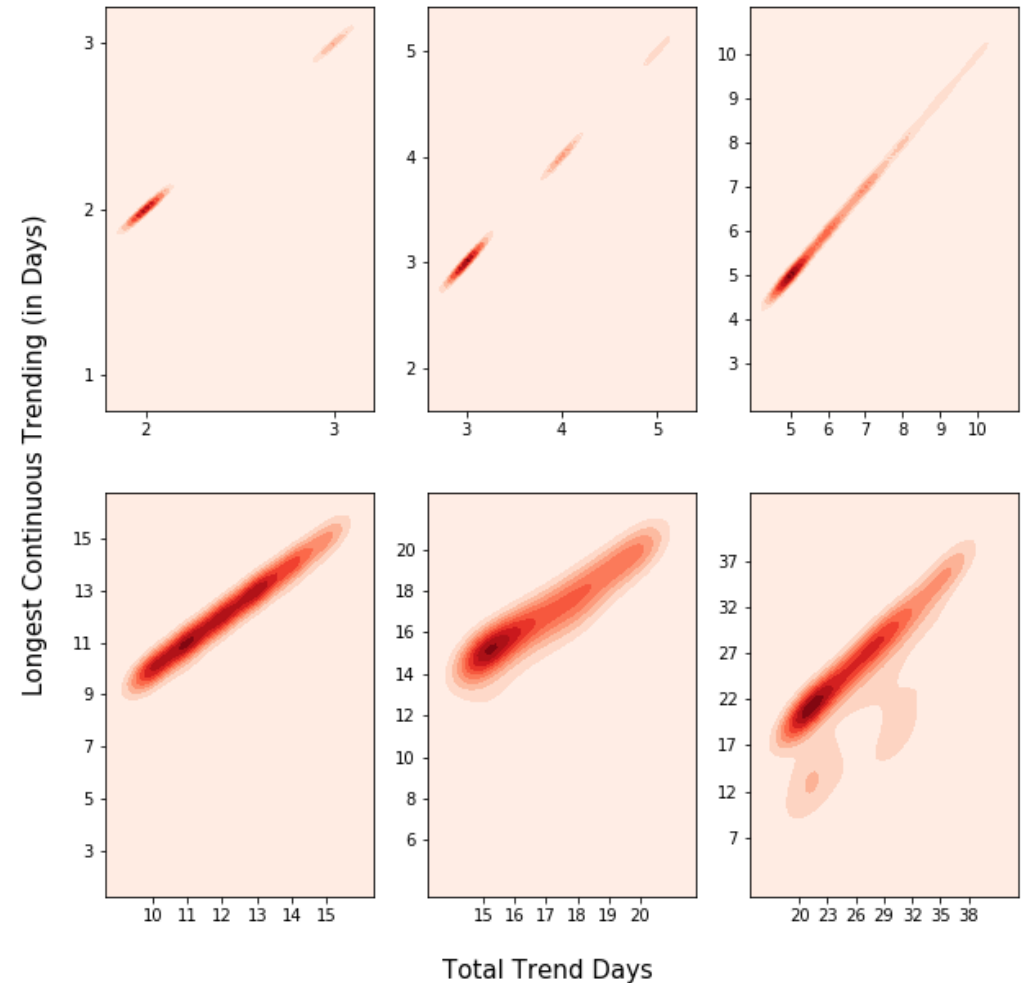
Impact of Features on Popularity: Statistical Approach

Feature importance of categorical variables using chi-square test
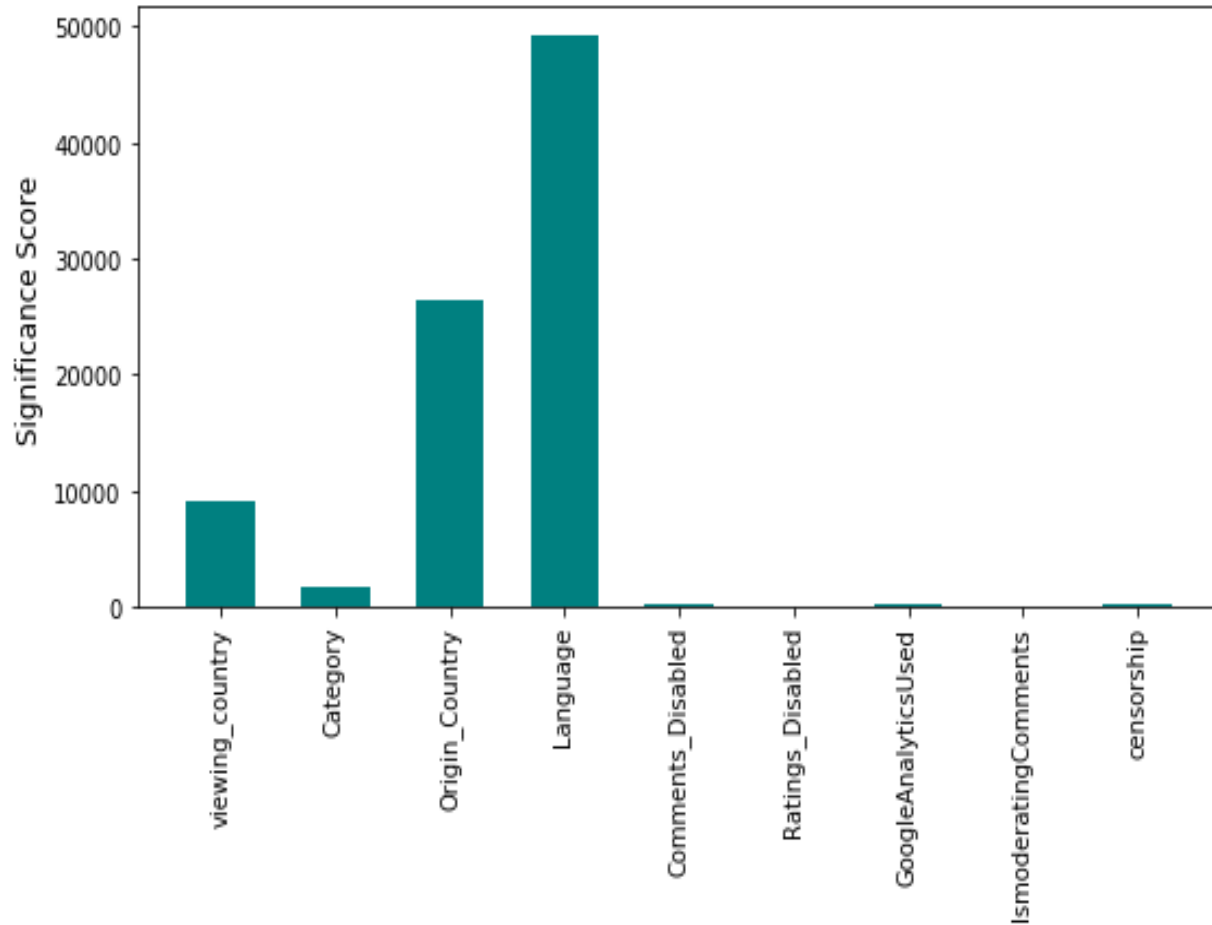
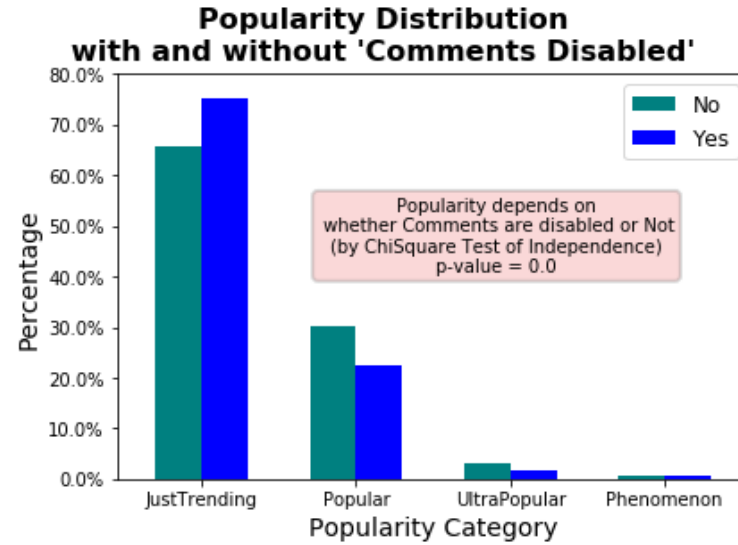Feature importance of numeric variables using ANOVA test
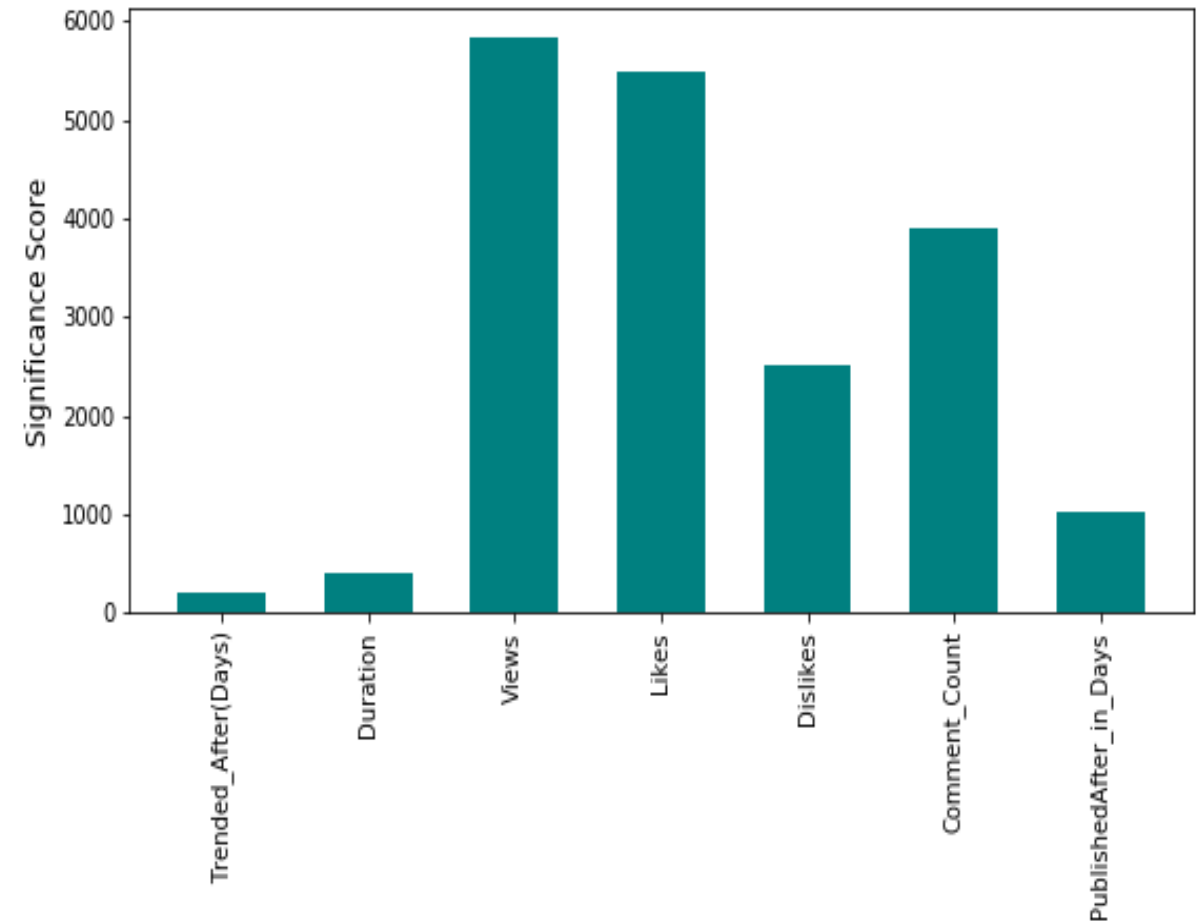
Categories Vs Popularity
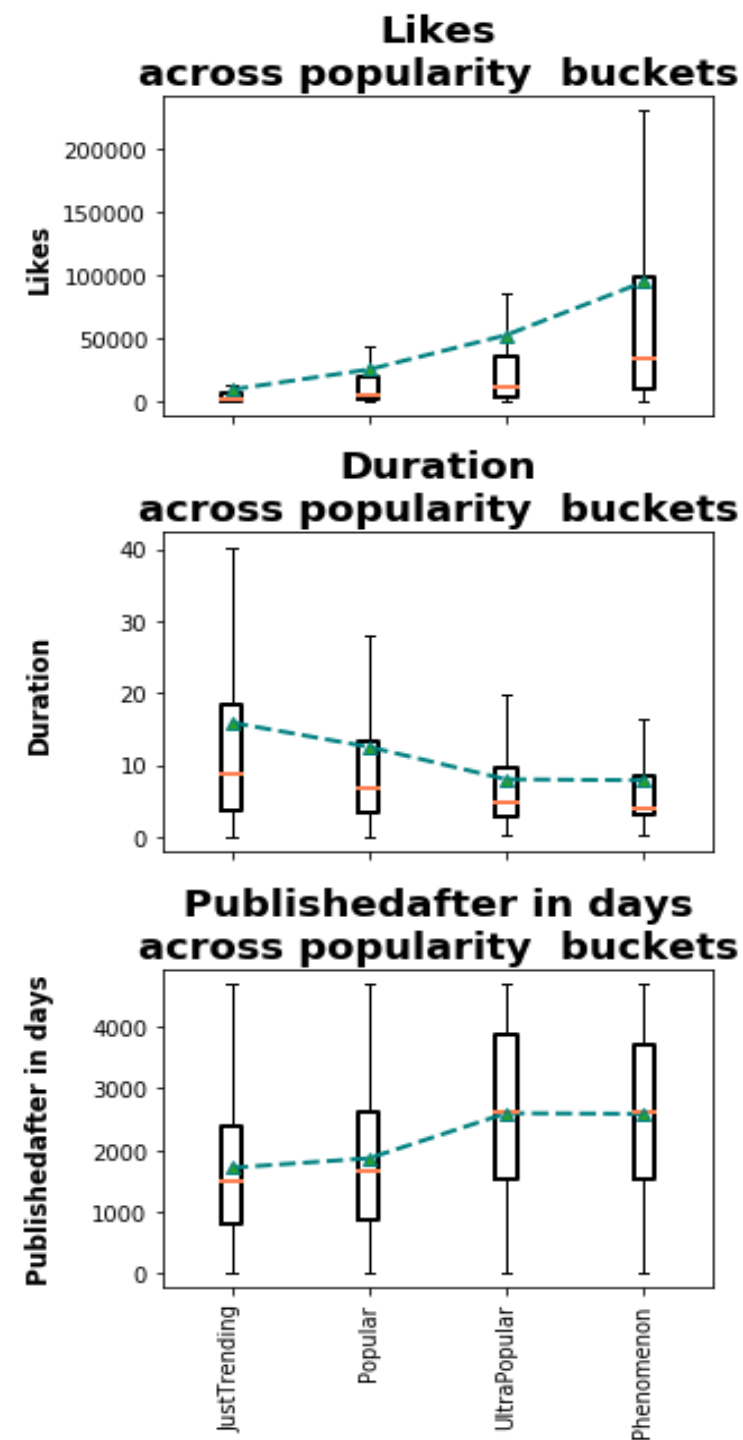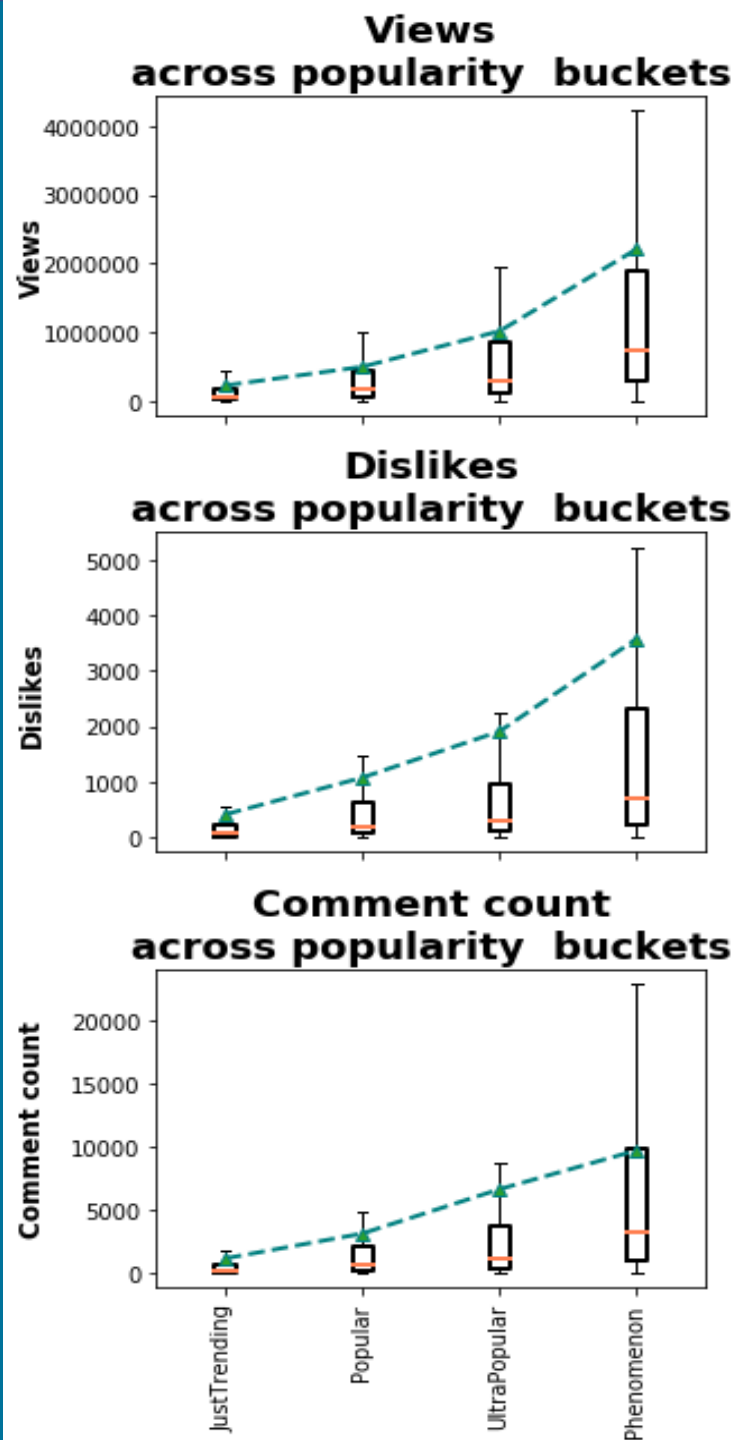

Difference in Days of Trending across Categories

Pets & Animals	Mean = 2, Median = 1
Gaming	Mean = 2, Median = 1
Howto & Style	Mean = 2, Median = 1
News & Politics	Mean = 2, Median = 1
Nonprofits & Activism	Mean = 1, Median = 1
Education	Mean = 2, Median = 1
Entertainment	Mean = 2, Median = 1
Music	Mean = 3, Median = 2
Film & Animation	Mean = 2, Median = 1
Autos & Vehicles	Mean = 1, Median = 1
Comedy	Mean = 2, Median = 1
People & Blogs	Mean = 2, Median = 1
Sports	Mean = 2, Median = 1
Travel & Events	Mean = 2, Median = 1
Science & Technology	Mean = 2, Median = 1
Shows	Mean = 1, Median = 1

Days of Trending

Impact of "Boolean Features" on Popularity
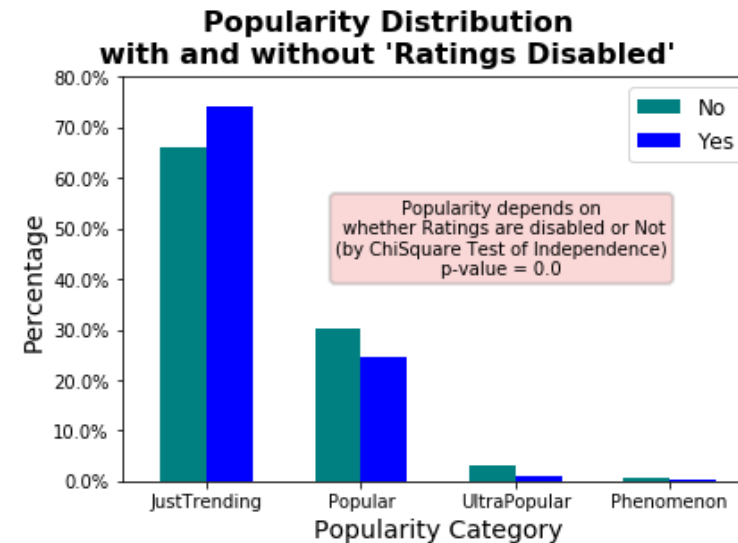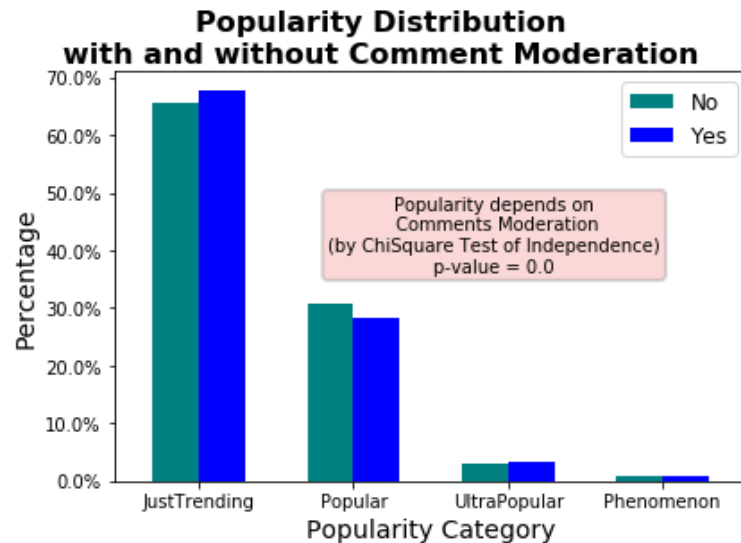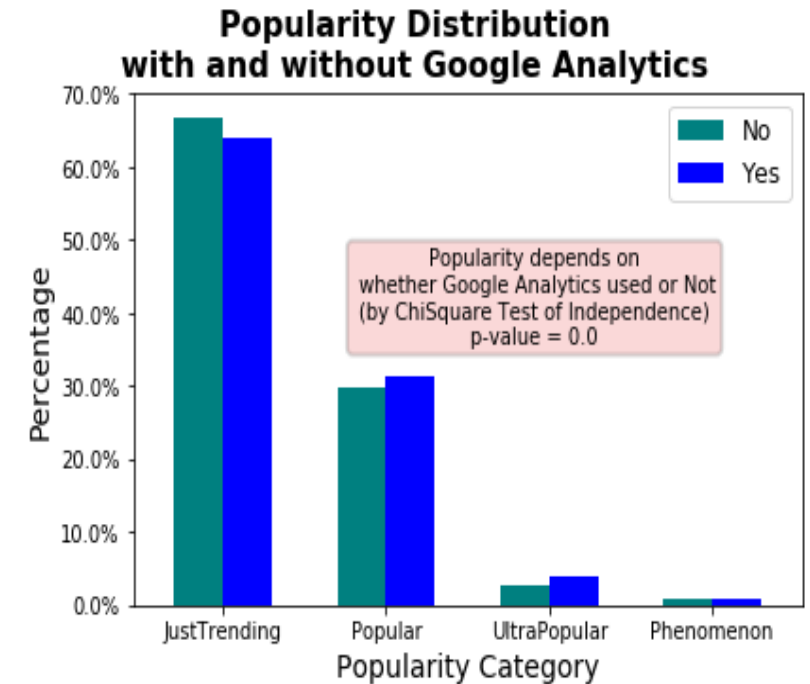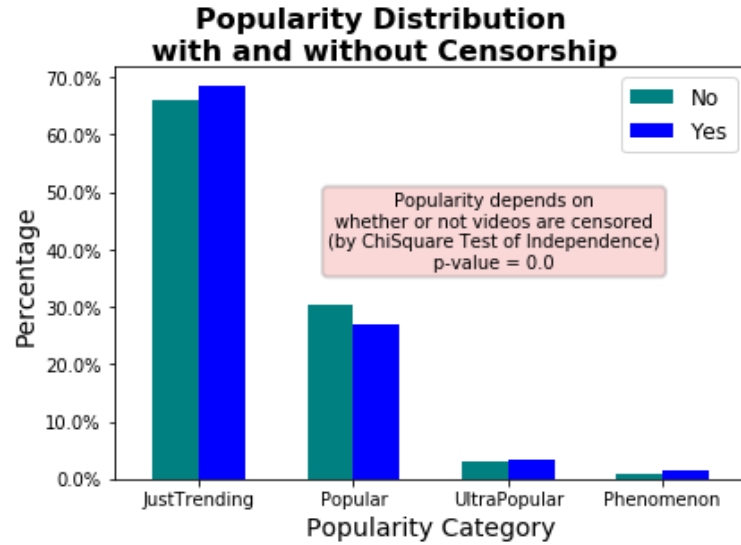
"Numeric Features" Vs Popularity

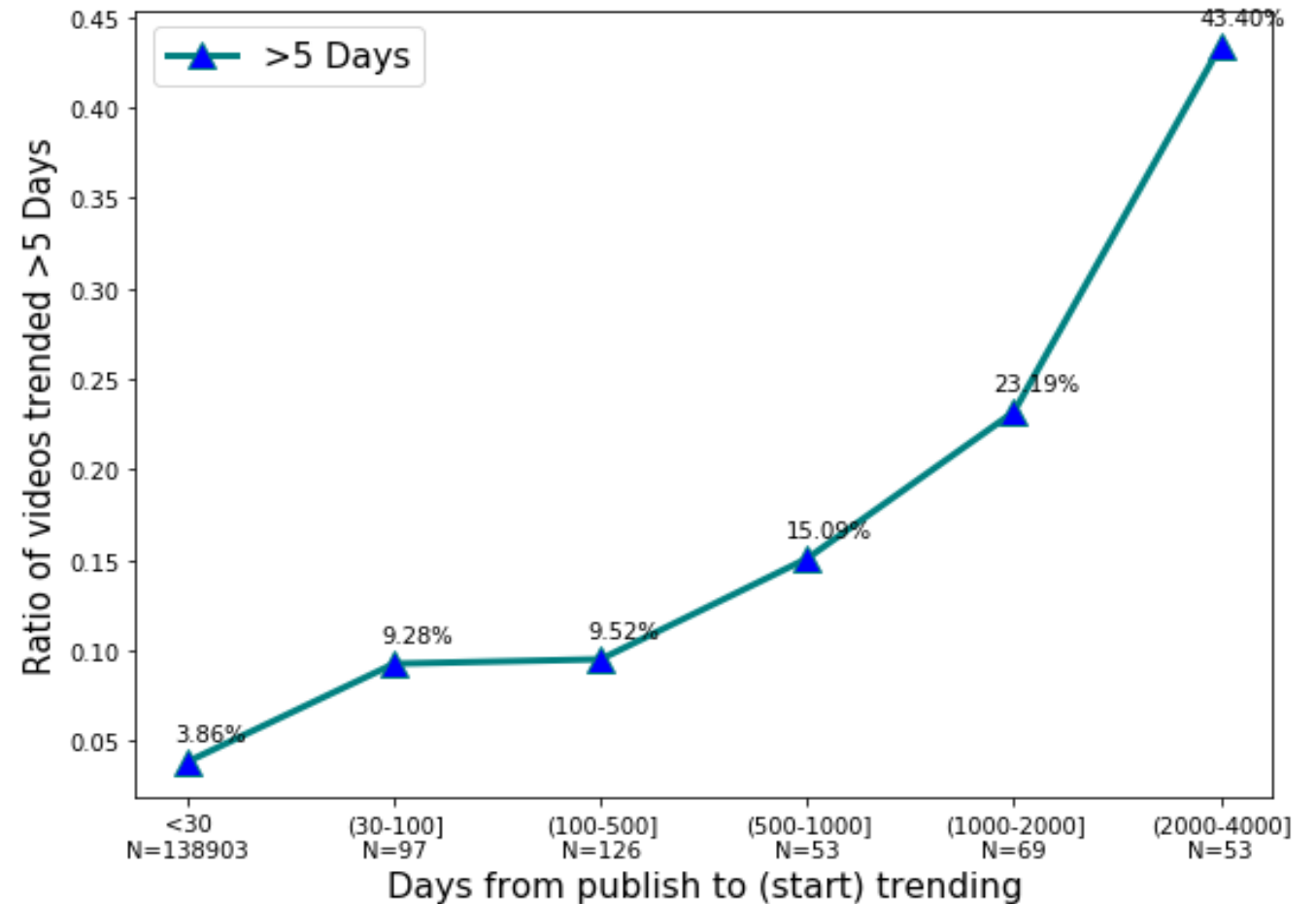"The Late bloom Effect"

Ratio of videos trending >5 Days
Vs
Days taken to start trending

%of trending videos
by "Days to start trending"

Within 1 Month
N=138903

99.7%   0.3%   >1 Month
N=398

Ratio of videos trended >5 Days

43.40%

23.19%

15.09%

9.28%   9.52%

3.86%

<30
N=138903

(30-100]
N=97

(100-500]
N=126

(500-1000]
N=53

(1000-2000]
N=69

(2000-4000]
N=53

Days from publish to (start) trending

What's next ?

Model Building & Future Scope

✔ **Built a base model - Test accuracy - 78.05%**

Interpretability of models

Incorporate Video Description, Title and Tags in the model (NLP)