University of Tartu

Faculty of Science and Technology

Institute of Computer Science

Ismayil Tahmazov

Understanding popularity on YouTube

Master's Thesis (30 ECTS)

Software Engineering Curriculum

Supervisor:

Rajesh Sharma PhD

Tartu 2020

Abstract:
Understanding popularity on YouTube

YouTube is one of the most popular and largest video sharing websites, which also has induce Social Network features. YouTube has also become one of the biggest promotional and advertisement platform on the internet, thanks to the number of users it attracts on its platforms in terms of content generators and consumers. While some content is popular, some other content fails to attract enough viewers. This thesis investigated the popularity of YouTube videos and channels by analyzing 500,000 videos. We analyzed this data to gain insight into popular Videos on YouTube. In other words, to see what's common among those videos. We tried to analyze the key features of YouTube popularity using metrics like views, likes, dislikes, comments, title, description, and video duration time. In particular, we performed descriptive and exploratory analysis and used techniques like text mining to analyze comments on videos. In addition, we also performed cross-cultural analysis in terms of popular videos to understand which kind of videos are more popular in which geographic locations. Our findings reveal that different types of videos are popular in different places leading to the fact that regional cultures do affect popularity of different videos.

Keywords: Comments, View Count, Likes, Dislikes, User Content Generator, Video Duration, Video title, Video description
CERCS: P170, Computer science, numerical analysis, systems, control

Pealkiri:

YouTube'i videote populaarsuse mõistmine

YouTube on üks kõige suuremaid ja tuntumaid videojagamise veebisaite, millel on ka sotsiaalvõrgustiku tunnused. YouTubest on saanud ka üks suuremaid relkaamplatvorme internetis, tänu kasutajate arvule, kes kasutavad seda platvormi nii sisulooja kui ka sisutarbijana. Kuigi mõned videod on populaarsed, siis mõned teised videod ei suuda piisavalt vaatajaid ligi tõmmata. Käesolevas töös uurisime YouTube'i videote ja kanalite populaarsust, analüüsides selleks 500,000 videot. Me analüüsisime neid andmeid, et saada ülevaadet populaarsetest videotest YouTubes ehk teisisõnu, et aru saada, mis on nendel videotel ühist. Me üritasime analüüsida YouTube'i videote populaarsuse põhjuseid, kasutades selleks mõõdikuid nagu vaatamiste arv, meeldimiste arv, mittemeeldimiste arv, kommentaarid, pealkiri, kirjeldus ja video kestus. Me viisime läbi kirjeldava ja uurimusliku analüüsi ning kasutasime meetodeid nagu tekstikaeve, et analüüsida video kommentaare. Lisaks viisime me läbi kultuuridevahelise analüüsi video populaarsuse mõttes, et saada aru, mis tüüpi videod on populaarsemad mingites geograafilistes asukohtades. Me leidsime, et erinevat tüüpi videod on populaarsed erinevates paikades, mis viis meid järelduseni, et piirkondlikud kultuurid mõjutavad videote populaarsust.

Märksõnad: Comments, View Count, Likes, Dislikes, User Content Generator, Video Duration, Video title, Video description

CERCS: P170, Arvutiteadus, numbriline analüüs, süsteemid, kontroll

Contents

1 Introduction

YouTube, an online social media platform, for hosting videos, created in 2005, has gained billion of users from its inception. The platform also assist in pseudo social networking among user content generators (these content generators could be bloggers, musicians, comedians, video gamers, educators, and students) and content consumers (who watch these videos).
After YouTube acquired by Google the number of professional channels much increased. Especially channels related to music, travel, personal blogs, comedy and gaming, technology. With the increasing of the channels, importance of YouTube in the media market has also increased. Today, even giants like Apple, Microsoft, Google first adversities own products with the help of bloggers on YouTube. Being popular on YouTube is essential for people who want to promote own products or just want to advertise about what they do[24].

Unfortunately, YouTube does not give any global statistics about videos and channel views, likes, comments so that, even simple questions like "How many videos exist on YouTube?" cannot find an answer. YouTube has only one official tool to track personal channels, that is, YouTube Analytics which is not enough to see the full picture of YouTube evolution and popularity metrics.

Before us a YouTube investigated in case of the different sides. But most of these analyses are local investigation of only one channel or certain type of search queries. Before starting our investigation, we also investigated these researches [1] [2] and did some local analysis with the purpose the understand YouTube public data. We did a search query analysis like "Barack Obama", "Trump","Today's trends". We investigated channels like "Imagine Dragon" music channel for the visualize most viewed, liked, commented videos. We found that with local investigation simple queries or channels we can't understand fully YouTube popularity.

The main research goal of this master thesis to understand popularity on YouTube. In particular, we analysed popular channels and videos to investigate about certain characteristics which make them popular. For our investigation, we collected data from the YouTube API (Application Programming Interface) and web scraping tools, by using analytical sampling methods. Our definition of popularity is based on View count. That is the number of views on video. Based on this definition of popularity we selected 500,000 videos. These videos belong to 10 different categories, such as music, education etc.

We investigated the dataset using the following three different research dimensions:

1. **Investigating popular videos:** We choose popular channels and videos from the trending YouTube videos, from the search queries, from external YouTube analysis sites like

[1] https://rpubs.com/nlavee/youtubedata

[2] https://journals.sagepub.com/doi/full/10.1177/1354856517736979

SocialBlade[3].

2. **Cross cultural analysis:** We also explore popular videos in different 10 countries to understand how different cultures have particularly liking towards a particular set of videos.

To perform our analysis we used two different kinds of techniques:

1. **Descriptive Analysis:** We analysed various attributes of videos' data, which can be divided into i) user content generators attributes (Title, Description, Video duration), and ii) videos' consumers metrics (Views, Likes, Dislikes, Comments to understand if there are specific metrics which correlate closely with popularity?

2. **Text Analytics:** We performed Natural Language Processing techniques such as sentiment analysis to understand the emotions of users for popular videos.

Our analysis reveals that popularity of at the same time depends not only on YouTube users but also on user content generators.

1. **Observation 1:** Short video titles and descriptions popular around the users.

2. **Observation 2:** Most of videos average video duration between 5-10 minutes.

3. **Observation 3:** Video popularity not only depends on user content generator or user also depends on :

   (a) Region

   (b) Religion

   (c) Mentality

   (d) Language

   (e) Education

   (f) Life level

   (g) Political situation

Rest of the thesis is organized as follows.

1. In chapter 1 we will give a brief overview of previous related research that has been done regarding YouTube analysis

2. Chapter 3 will focus on describing the dataset.

3. Chapter 4 we are explained our methodology.

[3] https://socialblade.com/

4. An overview of our empirical results is in chapter 5.
5. An overview of our cross-cultural empirical results is in chapter 6.
6. Chapter 7 focuses on our study summary.
7. Finally chapter 8 is for describing our overall contribution and proposes some topics for future research.

2 Related work

This section reviews literature related to the popularity of YouTube videos, channels, and YouTube in general. Many research papers focused on problem understanding YouTube's popularity from various perspectives [23]. Researchers analyzed YouTube in perspective of spam content [4], view count analysis [12], content analysis [17], travel advertises aspect analysis in YouTube videos [37], health investigation using YouTube as data source [45], illegal uploads analysis [11] and geography related analysis [8, 33]. All these analyses show that for YouTube video popularity not only depends on high values of views and likes.

2.1 Understanding YouTube platform

YouTube is the world's third most-trafficked website, with more than a billion unique users a month. YouTube is unmatched as a video content streaming and distribution site, which can effectively be combined with other social media platforms [46]. YouTube channels also allow for a lot of flexibility and provide ways to improve identity across platforms. Personalized recommendations [10, 16, 39, 41] are a key method in today's data-rich environment for extracting information and exploring content. Combined with pure query and browsing, they allow users to access the information in an effective and rewarding sort of way while facing a huge amount of information. YouTube provides some unique opportunities for user content generators and recommendations as to the biggest and most successful online video community with vast amounts of user-generated content [16].YouTube is a website for video sharing with a Social Network [7, 15, 31, 32, 44] features that allows users to upload content to own personal YouTube Channel that features videos of user choice – made by users or others [41]. YouTube offers comment threads that user content generator can control on own channel and posts, and a tracker that lets the user keep track of own posts [44] .

2.2 Understanding popularity of YouTube

During the last decade, YouTube videos evaluated with the high speed [6]. Uploaded video to YouTube during these years increased few times and new video category types created like, gaming and travel. On YouTube, both popular and unpopular videos' views, likes, and upload also increased. This related to the increasing globalization and education level during the last years. Today in every second thousand video uploading to YouTube and millions of live streams on YouTube. After the introduction of smartphones [21] number of videos which uploaded to YouTube increased 3 times. New videos uploads and views progress is different. During the last 10 years uploads increased several times, but video viewers did not increase in the same way. YouTube have more videos than users several times. From the literature reviews, we observed that views the channels related to People and Blog, Gaming, Music, Entertainment categories increased by 6 times during 2006-2016. YouTube now control a big part of music industry [9, 18]. Additionally, the gaming industry has gained more and more popularity. Every day we can find new games in the online game catalogs. With the popularity of this new trend game videos also started to be popular [26].

One of the most important changes was the Partner Program, which started in 2007. YouTube has leveraged the most significant idea in the web space by creating a way to encourage content creators to get paid for their work. YouTube started the inclusion of advertisements in the same year[1]. In 2009, all of a sudden users have facility view High Definition (HD) videos on YouTube, indicating a major content opportunity improvement in quality. In 2010, YouTube moved away from the Adobe Flash Player [35] that was previously used to view videos on browsers (by 2015, they declared that HTML5 was the default playback for most browsers). That wasn't a YouTube's only big improvement made in 2010. While they had experimented with live streaming in 2009 [34], native live streaming technology was launched in 2010. Today we are using YouTube like TV, we can see live news[5], listen to the radio, watch webinars in real-time, and so on. For that investigates YouTube only like a simple video sharing platform not correct [28].

2.3 Recent researches

User-generated content on social networks have a big impact on our daily life [14, 41]. YouTube also become one of the biggest education websites on the internet[38]. Many people spend most of our daily life on social media platforms like YouTube [25]. Authors in the [25] investigated why people like to watch YouTube. For that, they have created questionnaires and ask questions like why you like spending time on YouTube at different university students. They found that people like to spend on YouTube because of different reasons like entertainment, to participate in the discussion, to show other encouragement, and so on. From these answers, one can conclude that people from YouTube want only one thing to "have a good time", for that most popular YouTube categories are music, entertainment, and gaming. Politics and social networks have a close relationship with each other and YouTube is not an exception. This was observed that in the selection of United States in 2017 [19]. YouTube video audience choices analyzed by genders in the article [43].YouTube video thumbnails first think after title users see in the video, for those thumbnails should have some attractive images [22]. For that understand between YouTube video content, title, description, and thumbnail relationship very important for video popularity. YouTube has several customer categories, one of the largest categories is about content for kids. In [2] the authors investigated the collective behavior of a user for this category and what kind of videos they like. They used techniques like sentiment analysis and classification. Additional to that also did some advertisement and audience analysis. The result of the research follows: Different age groups, geographic locations like different contents, and popularity of video depend on thumbnails, faces in the video. Another perspective of YouTube video analysis is to do it by Geographic location [8]. As we can understand in different countries, people like different things. In [8] authors analyzed the relationship between geography and video categories, time frame analyses of the views, social factors which impact to views of videos. However compare of [8] paper and this thesis we are mainly focused only video duration, title, description and video categories analysis.

View Count is a very important metric on YouTube if we can find a relationship with the like count, dislike count, duration of videos we can use that for the popularity formula. YouTube comments have a big impact on video popularity, but very important to classify this comment for understanding the content of this comment. Like news, forwarding to others like video, negative or positive comments [29]. In [13], authors analyzed the relationship between view Count and comments, uploaded frequency of videos, and other aspects. From this, we can understand that view Count is an important metric for measuring popularity, but other metrics are also important. Like Count, dislike Count and the relationship between these metrics with the view count very important for the see the full picture of channel popularity [17, 27]. In this thesis we are also analysed these metrics to understand popularity.

3 Data description

We gathered data directly from the **YouTube API**[4] using the **"tuber"**[5] library and combined YouTube video pages crawling with Selenium. The **"tuber"** is a special library in **R created to use the YouTube API**. The Tuber package provides an excellent kit for collecting comments, views, likes, or general material from any YouTube videos. This package is designed without entering a particular channel authorizing page to explore data. The purpose of this library is to provide a simple automated summary to evaluate fundamental channel statistics, based on its massive functionality. The extracted data set from YouTube with the "tuber" library and web crawling are shown in Table 1. The full data set description shown of the "tuber" library in Table 2:

Video Count	Size	Comment Count	Time Frame
500 K	600Mb	8 million	2009-2020

Table 1: Full dataset

We now give a brief introduction to YouTube data and discuss a few important features. In our dataset we have two type of YouTube metrics :

1. **Metrics which depends on users:** After watching a video, users can give feedback about it. Users can give like or dislike and some comments to the video. These feedbacks can be measured using popularity metrics like **commnentCount, viewCount, likeCount, dislikeCount**. In Table 2 we can see an example data set from tuber library which we used in our investigation.

[4]https://developers.google.com/youtube/v3
[5]https://cran.r-project.org/web/packages/tuber/vignettes/tuber-ex.html

Name of tuber response.	explanation
video_id	YouTube video id
publishedAt	video publishing date
title	video title
description	video description
thumbnails.default.url	default thumbnails image url
thumbnails.default.width	default thumbnail's image width
thumbnails.default.height	default thumbnail's image height
thumbnails.medium.url	medium view thumbnail image url
thumbnails.medium.width	medium view thumbnail image width
thumbnails.medium.height	medium view thumbnail image height
thumbnails.high.url	full screen view thumbnail image url
thumbnails.high.width	full screen view thumbnail image width
thumbnails.high.height	full screen view thumbnail image height
channelTitle	Video channel title
liveBroadcastContent	Video live broadcasting info
likeCount	video like count
dislikeCount	video dislike count
commentCount	video comment count
favoriteCount	video favorite count
videoLength	video duration
comments	video comments

Table 2: Simple data set from the Tuber library [40]

In this study, we focus a few metrics, provided by YouTube, like count, dislike count, and the number of comments(Table 3)

Metrics	Minimum	Average	Maximum	Short Description
viewCount	2M	3.5M	1B	number of views
likeCount	0	26308	10683228	number of likes
dislikeCount	0	840	511921	number of dislikes
commentCount	0	1088	309570	number of comments

Table 3: Used metrics

2. **Metrics which only depends on user content generators:**

   All these metrics can reflect video popularity. But additional metrics which user gives after watching a video, some metrics like **video duration, video title and description length** user content generators operate these during the creation of video content. We took into account these metrics also.

For the study, we selected randomly 10 different video categories (figure 1) from our YouTube dataset.

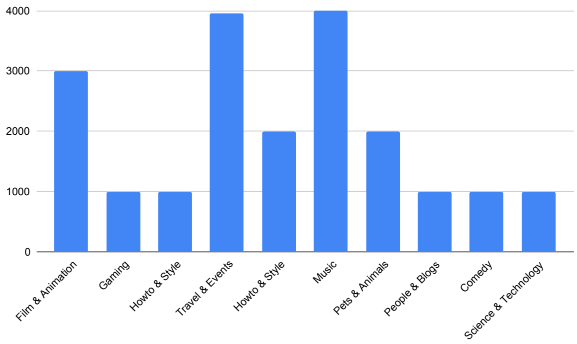


Figure 1: Investigated video categories

From our investigation, we have gathered more than 600 Megabyte YouTube video information datasets these take more than six months in total. (see Figure 2). Mostly our dataset includes Entertainment, How to Style and People, and Blogs video content. Less uploaded videos to the YouTube website in Science and Technology, Gaming, and Music video categories.
You can see our full data set overview in Figure 2:

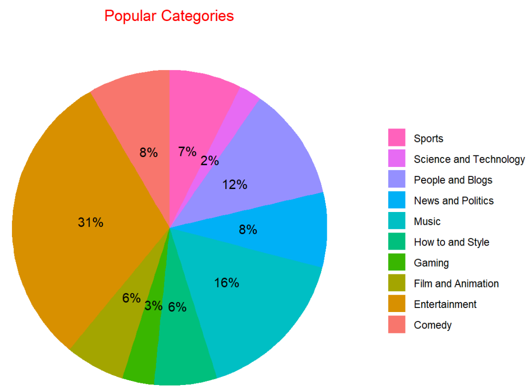


Figure 2: Popular Video categories

But in case of popularity, these percentages are different, for example, Music and Gaming videos more popular than others and YouTube content generators gain more money in these categories than others in total. In YouTube API every video category has own special id in the API response which we can see in Table 4. The video category id must be defined when operating with the YouTube Data API and in the video upload process. With the help of this category ids, we classified the whole dataset to the small parts for the find correlation between YouTube metrics in different categories.

3.1 Data Collection

3.1.1 YouTube API Crawling

YouTube has an API for developers and researchers. We have used this for direct access to the public data of YouTube videos. YouTube API[6] is a part of the Google API[7] and for access to YouTube API, we should use the Google API. For the use of the Google API, we need API id and key. For that, we are using the google console[8]. With the help of this API, we can gather or upload data from any programming language include R and python which we used in our code scripts.

3.1.2 Selenium Crawling

Selenium[9] is a web browser automation tool initially designed to simplify web applications for testing purposes. This is also used for many other uses, such as automating web-based admin tasks, communicating with none API systems, and also Web Crawling. For the crawling, we used python and R programming language. With the help of selenium crawling, we managed to gather YouTube data without any limitations(with the YouTube API v3 we have limitation 10.000 queries in a day) . And also as a web crawling tool, we used Octoparse[10] web crawling. With the help of this tool, we gathered YouTube data without any coding.

[6]https://developers.google.com/youtube/v3
[7]https://developers.google.com/apis-explorer
[8]https://console.cloud.google.com/
[9]https://www.selenium.dev/
[10]https://www.octoparse.com/

id	category name	details
1	Film& Animation	Film and Animation categories include new film trailers, film related videos, film scenes, songs, animation, videos, short stories on cartoons and so on.
2	Autos& Vehicles	The Autos and Vehicles category includes car related video, automotive bikes, unique vehicles and much more.
10	Music	Music category contains all sorts of music related videos, songs etc.
15	Pets& Animals	Category Pets and Animals contain humorous and beautiful videos about pets and animals.
17	Sports	The category Sports contains videos about sports and sports events.
19	Travel& Events	The category Travel and Events includes travel tip videos , travel locations, event videos, event tips and many more for travel and events.
20	Gaming	Games category contains games' videos, game tricks, hints, game reviews and more.
22	People& blogs	Category includes people videos, lifestyle, blogs, website promotion, opinions and more about people and blogs.
23	Comedy	Comedy category contains various types of comedy videos, it is stand up video, short film, animated film, and other amusing humor videos.
24	Entertainment	Entertainment category includes entertainment videos, sometimes this category choosing because not certain which type video it is .
25	News& Politics	Category News and Politics that extend or include news or policy material.
26	How to Style	Videos that have guides, trends, fashions, tips on designs.
27	Education	Education category contains educational videos, if video is for demonstration or information and does not suit in other categories, it can be included in the category Education.
28	Science& Technology	Videos related to science , new discoveries, gadgets smartphones, laptops
29	Nonprofits & Activism	Videos related with nonprofit or activism

Table 4: YouTube Videos Category ids [42]

4 Methodology

In this section, we describe our methodology for analysing the popularity on the YouTube platform.

4.1 Explanation of thesis

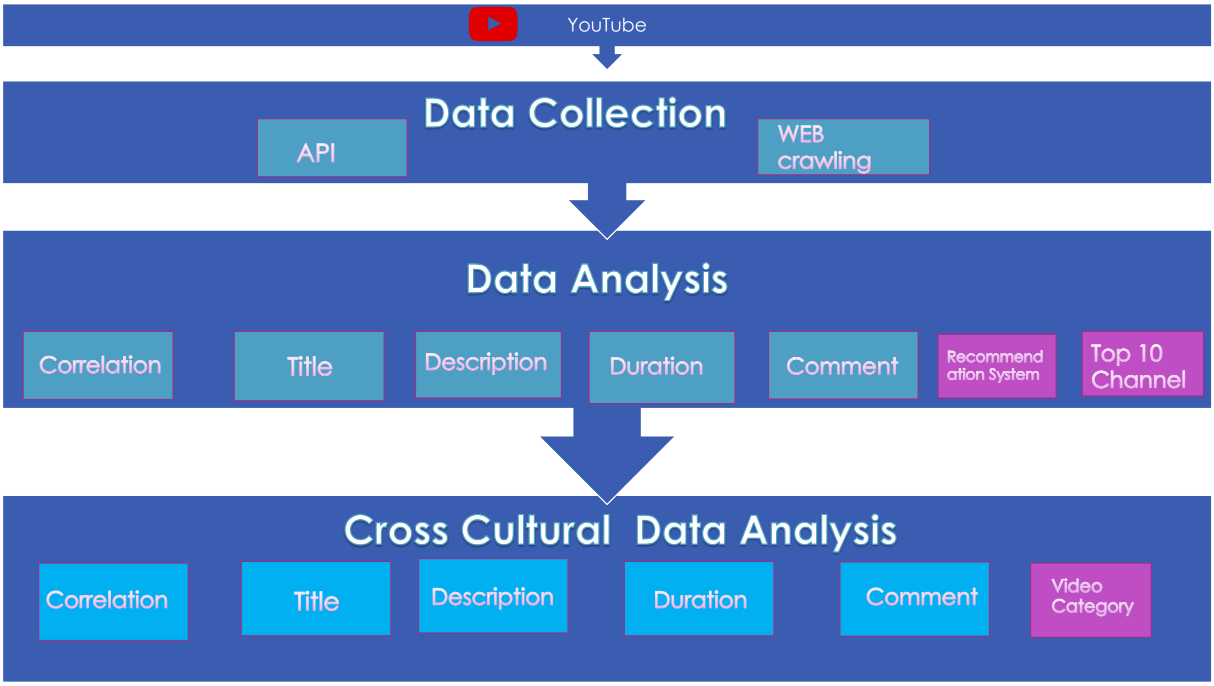


Figure 3: Thesis Evolution

Our methodology is shown in Figure 3.

1. **Data Collection:**
   Firstly, we first performed dataset gathering through
   (a) API Crawling
   (b) Web Crawling with selenium[11]

   See Chapter 3 for more details.

2. Data Analysis
   (a) Analysis of relationship YouTube metrics between each other (Correlation Analysis)
   (b) Investigation of titles' keywords and optimal length (Video Title Analysis)

[11]https://medium.com/@igorzabukovec/automate-web-crawling-with-selenium-python-part-1-85113660de96

(c) Investigation of description' keywords and optimal length(Video Description Analysis)

(d) Investigation of optimal length of video duration in the different YouTube video categories (Video Duration Analysis)

(e) Sentiment analysis of YouTube video comments in the different categories(Comment Analysis)

(f) 10 most popular channel analysis to find general similarities(Top 10 Channel Analysis)

3. Cross Cultural Data Analysis

   (a) Cross Cultural analysis of relationship YouTube metrics between each other (Correlation Analysis)

   (b) Cross Cultural investigation of titles' keywords and optimal length (Video Title Analysis)

   (c) Cross Cultural investigation of description' keywords and optimal length(Video Description Analysis)

   (d) Cross Cultural investigation of optimal length of video duration in the different YouTube video categories (Video Duration Analysis)

   (e) Cross Cultural sentiment analysis of YouTube video comments in the different categories(Comment Analysis)

   (f) Cross Cultural YouTube popular video categories analysis in case of likes, dislikes, comments (Video Category Analysis)

In next sections, we explain the different stages in detail.

5 Data Analysis

5.1 Correlation between YouTube Metrics

Metrics which user content generator can manage(Video category id, title, video duration length, and description lengths) to have a relationship with the main YouTube metrics(views, likes, dislikes, and comment count). We can see this correlation between our whole data set in figure 4.

1. From figure 4 we can say that title and description lengths have a negative correlation with the video views, likes, and dislikes. This means that when title and description length is longer video is not popular.

2. And from figure 4 we explored interesting relationship YouTube video category id have a positive correlation with the description and title length. From this relationship, we can say that with the high number of category id video titles and description length are also longer. For example, in the Film and Animation category YouTube videos (category id 1) titles and description lengths less than the Gaming category YouTube videos (category id 20). From this relationship, we can say that with the increasing of the category id videos will be less popular, as views, likes, dislikes and comments count was decreased. Full list of videos category id table 4.

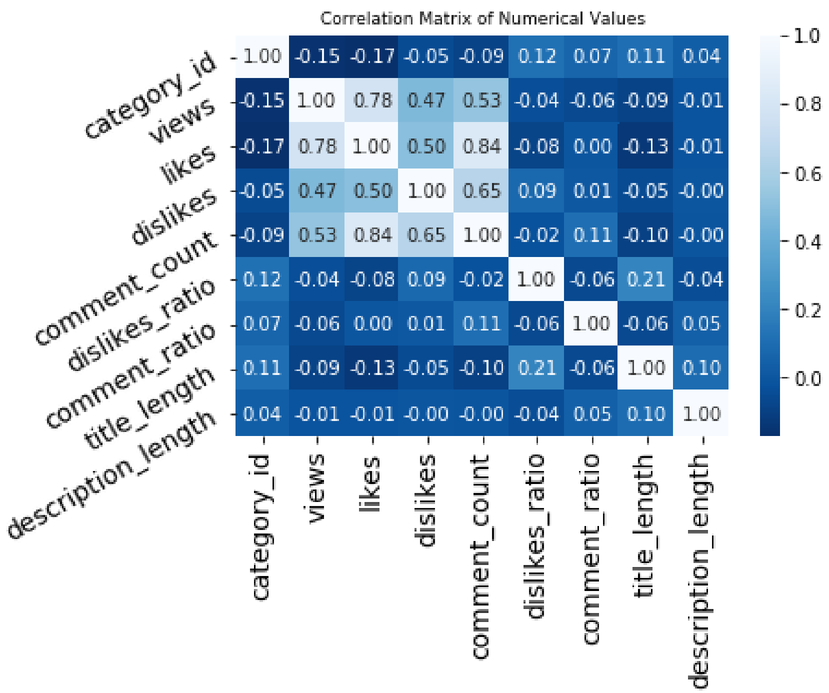


Figure 4: Correlation between YouTube metrics

The correlation between YouTube metrics is different for various YouTube video categories. We can see that in figure 5. For example, title length has a positive correlation in comedy and travel YouTube video categories. These mean that *users like longer titles in comedy and travel video categories*. We see the same relationship for the description length also in the comedy and travel YouTube video category. But for the other analyzed categories in figure 5 title and description length have a negative correlation with the other YouTube metrics. In short, **titles and descriptions with less length have more chances to be popular in these YouTube video categories**.

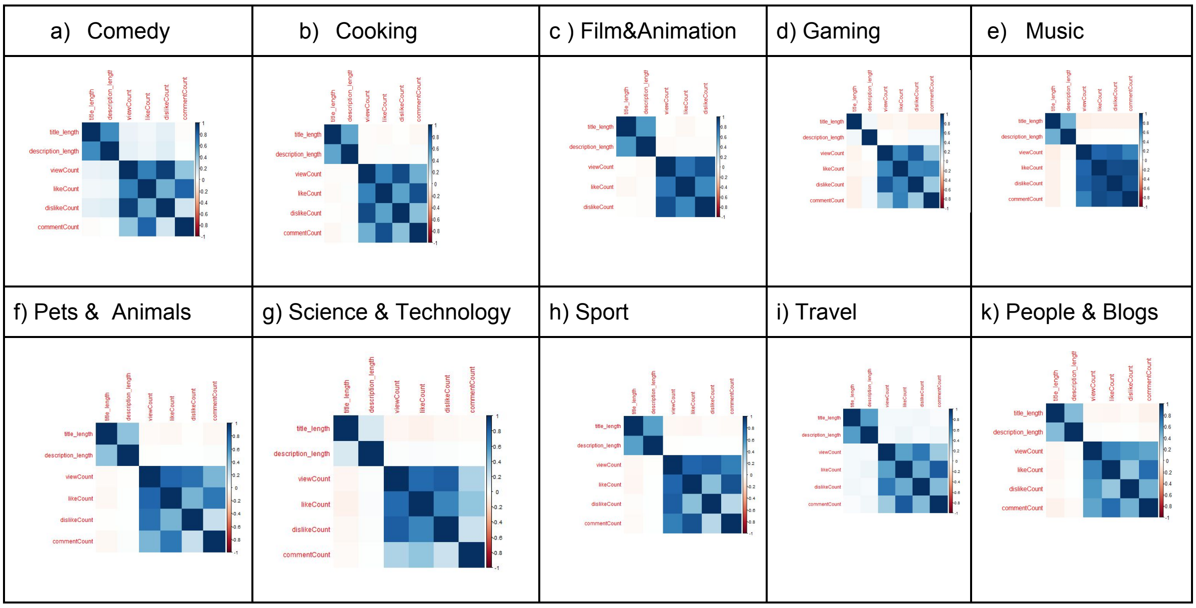


Figure 5: Correlation Matrix in the popular YouTube categories

5.2 Title Analysis

For most content creators that post videos to YouTube, one of the main goals is to be higher on the suggested YouTube ranking system. But, the fact is that 70 percent of YouTube videos are discovered through searching. That's why the keywords used for YouTube videos are highly important to achieve better user-friendliness.

1. The first aspect people look for when they're searching for a video is a title that will show them the answer they're looking for. But it's not the only purpose that the title of YouTube platform.

2. Videos' titles also lets YouTube realize what the video is about and this video which

search quires results should be add. That's why the recognition and use of keywords in the title is one of the easiest ways to automate videos' popularity.

3. In figure 7 we can see the categories videos and most used words in the title. Every category on YouTube has special most used keywords, for example, in a Comedy category most used words: " stand up"," funny ", " top", "movie"," sketch" and so on. When content creators upload videos to YouTube, in the title at least should behave one popular keyword, as the YouTube searching algorithm work with these keywords in the title.

- ***What is a description of the YouTube video?***

  A video title is a piece of metadata that helps a lot of YouTube during a search. If the video title optimized correctly during Google or YouTube search generated content will be appearing frequently.

- ***Why this is important?***

  Video title first thing the user sees during the YouTube search, for that this is very important. Like users, YouTube also during search first check video titles for the results. YouTube divides words and expressions in the titles for the use of it as keywords.

- ***What is it best length for a YouTube video title?***

  Video title length is very important because long titles will not be read by users. For that user content generators mostly using short titles, around 50 characters like in the figure 6. The length of titles can be changed for the different countries, different video categories, live or standard videos.

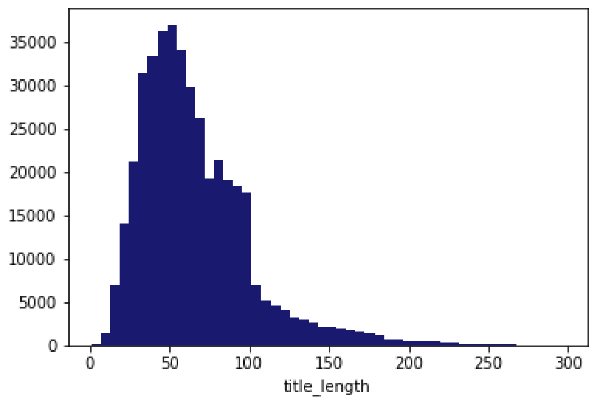


Figure 6: Title Length

- ***Which words more popular in the titles and which one user content generators should be used?***

  Titles using during the search queries, for that using correct keywords very important during the creation of YouTube video content. Write the correct title very important for video popularity. For example, in comedy, YouTube video category most using keywords is a comedy, sketch, stand up, for the cooking YouTube video category cook recipe, food, for Film and Animation movie, full, HD, official, trailer.

  In figure 7 we can see the full list of words and categories which mostly using during by user content generators. In 7 we created a detailed description of frequently used words in titles for the different YouTube video categories. We have used techniques like the word cloud and most used word bar plot. All of these keywords first thing YouTube search and recommendation algorithm use during the search.

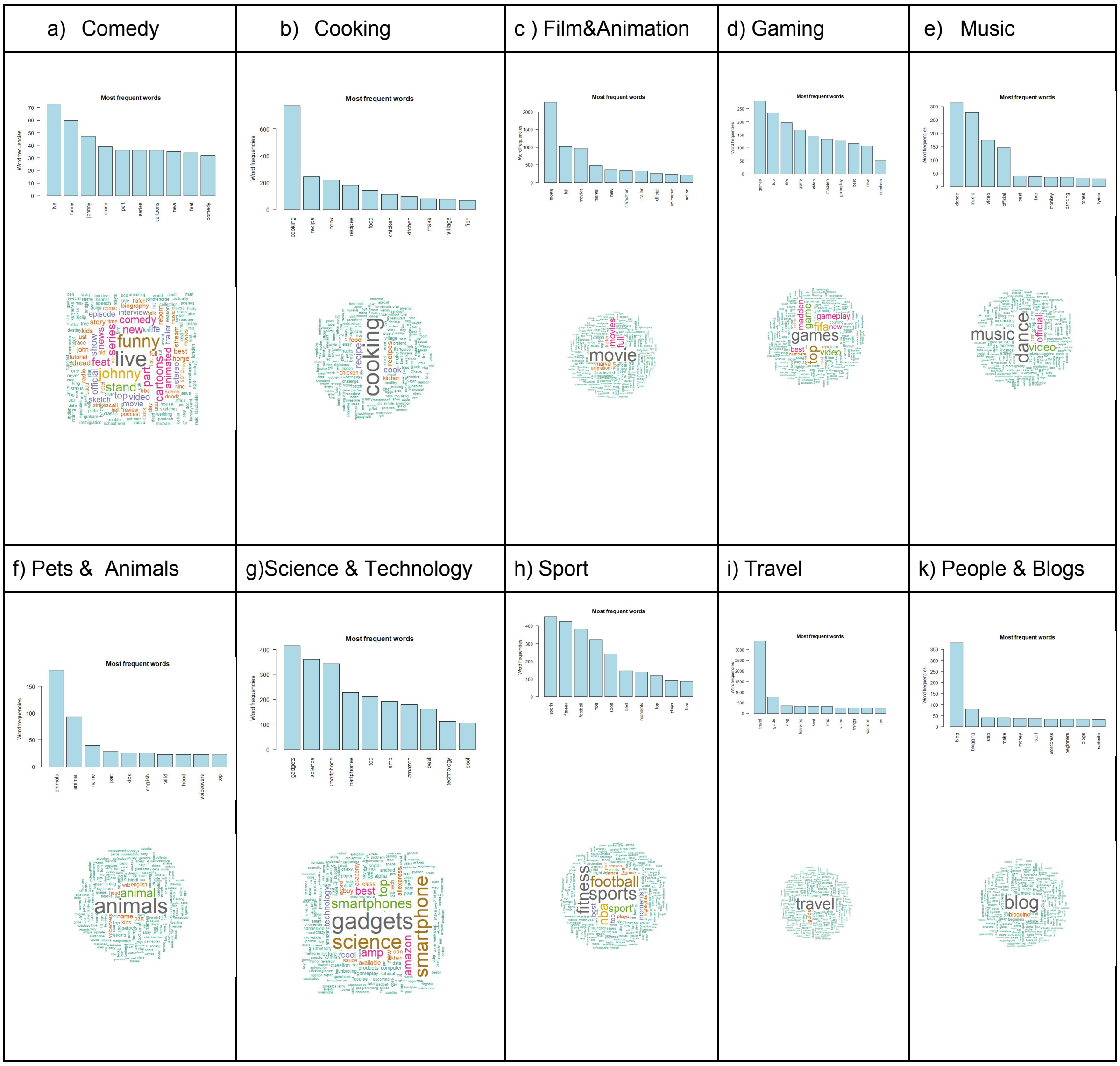


Figure 7: Most used words in titles

5.3 Video Description Analysis

- ***What is a description of the YouTube video?***

  First of all description is a short text content of YouTube video which help YouTube user understand the content of the video in a few seconds. Later, video description is a piece of metadata that helps a lot of YouTube fully understand a video's actual content. Excellencies-optimized explanations will contribute to increased YouTube search ratings.

- ***What is it best length for the YouTube video description?***

  Description length one of the important metrics of the YouTube videos. Yes, a detailed description of video content is important but, when content generators use too many words no one will not be reading it. This will annoy users, for that optimal video description length is important. From figure 8 we see that most videos have less than 2000 characters in the description. It can be concluded from this description is better, if user content generators can explain video content in the few sentences, this will help users in understanding video and will increase chance popularity of the video.

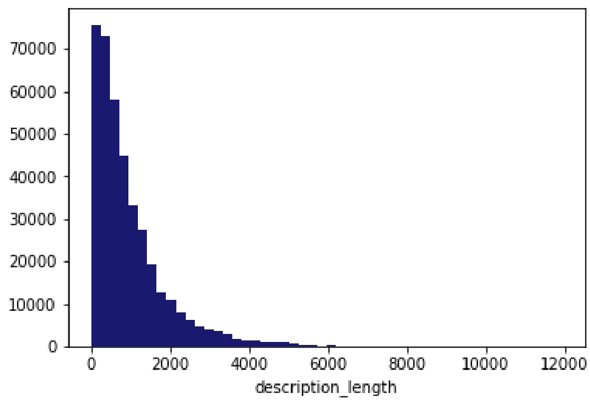


Figure 8: Description Length

- ***Why this is important?***

  1. Description of third thing users sees in the YouTube video content. For that this very important thing.
  2. Description using in the search of the YouTube queries. Description divides into small parts like words and expressions and YouTube use these like keywords in the search.
  3. Using the correct words very important if the user content generator wants to video appears in lists of recommended videos and first appears during the search queries [36].

- ***Which keywords user content generators should use in the video description?***

  Every YouTube video category has own popular keywords we can see that in figure 9.

  1. For example, if user content generators' video related to the Science and Technology most used words in the description is the technology, gadgets, amazon, top, best, and so on. As we say before all words in the description using as a keywords by YouTube search algorithm.
  2. In the comedy, the most used words are the one, best, actor, show. With the help of most used words in these categories, we can understand the best description properties for popularity. Most of the user content generators add to the description link to subscribe, other related videos from their channel for the advertise own products. This is the best practice for the increase of subscribers counts and channel popularity.
  3. In the Sports video category, popular words in the description are the football, NBA, fitness, season, sport, game, dance. These words are the most used keywords in the search Sports video category.
  4. In the Pets&Animals, popular words in the description are animal, funny, pet, children, kids, top, and so on. These related to that when users search, they are mostly using keywords like funny, pet, kid, and pet words to find needed results.

  We observed that if user content generators can predict popular words for every category popularity of video or channel can be increased. For the manage with all these keywords in the description user, content generators can use YouTube Analytics like programs for the predict and use popular words in the certain YouTube video category.

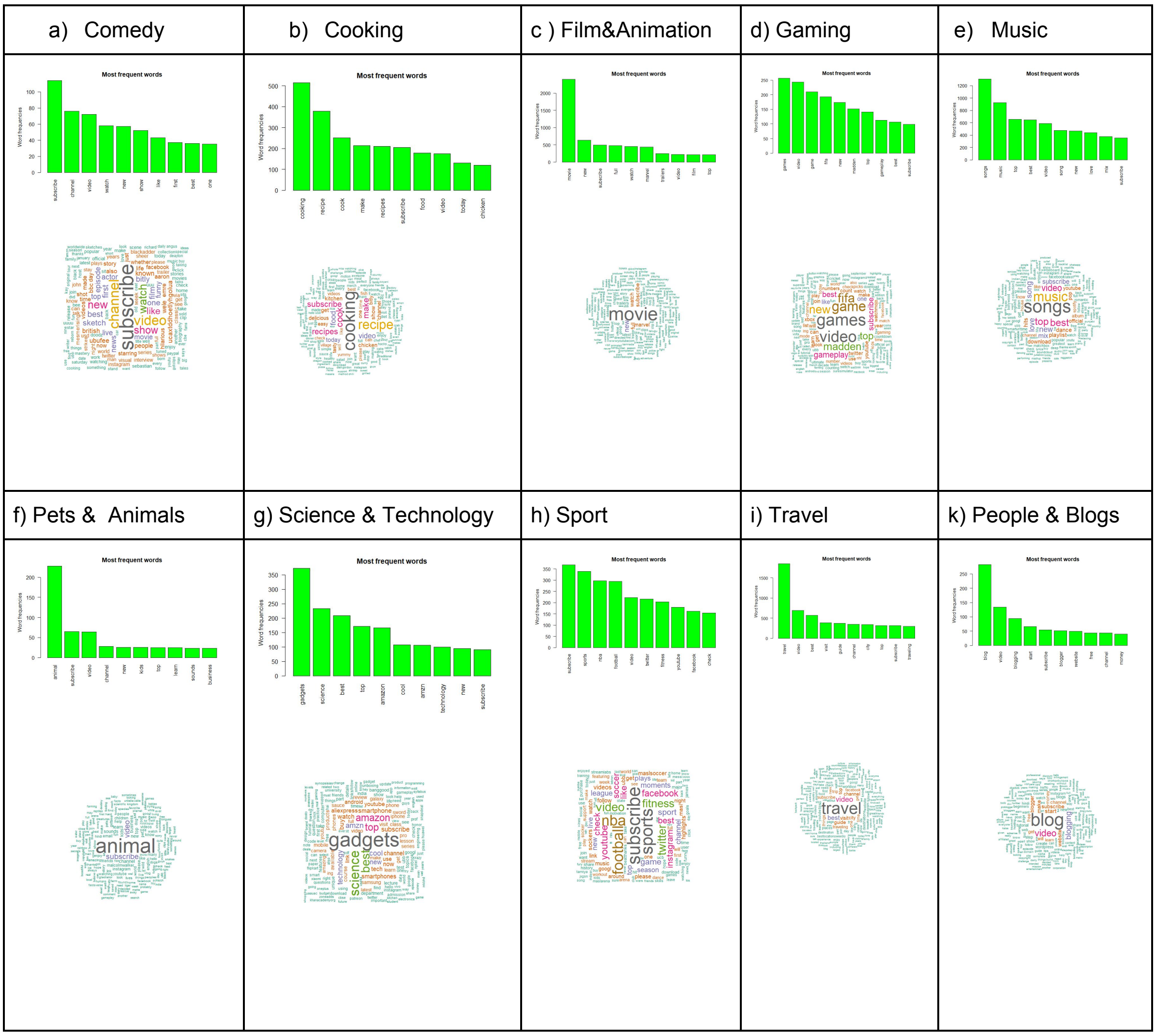


Figure 9: Video description Analysis analysis

5.4 Video duration analysis

A big difference between traditional media web providers and YouTube is the length of YouTube videos' posts. Although most typical sites contain a small to a medium number of long videos, YouTube videos usually have a length of between 5 minutes and 2 hours and are mainly made up of short videos as shown in figure 10. YouTube has some 15 minute limit during the first video uploads. For that on YouTube, during the search, we see videos mostly short videos. During the analysis, we found that with increasing video length video quality decreases. Rendering and editing of long-length videos take a lot of time in the modern personal computers, for that video length more than 45 minutes videos not very popular in the around YouTube users.

YouTube statistics show that the best duration for a video max is 15 minutes. However, this

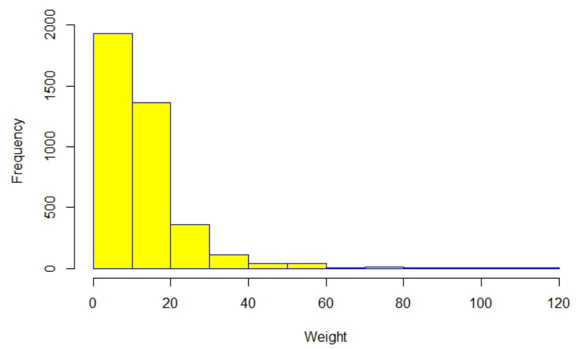


Figure 10: YouTube videos' length from dataset

depends on the video category. For example, for music content the best duration is 3-4 minutes, for a cooking is 10 minutes. Figure 11 displays the distribution of video length for the top ten most popular categories. Most of these categories' video duration distribution change very rapidly. People&Blogs(5-20 minutes) and Science&Technology(5-15 minutes) videos duration evenly distributed. From that we can understand 2 thing:

1. Equally duration distributed video categories captured all media in that certain category
2. Short videos are popular most of categories.

If we look at more accurately we can see that for every category optimal video length is different. For example:

1. In the Film and Animation, the optimal video duration is between 5 and 10 minutes. This related to that most popular thing in this category is movie trailers and short films.
2. The comedy video category has more video duration than others for approximately 10-15 minutes. The reason for that is comedy sketches, in which the minimum duration of this is 10, maximum is 2 hours.

3. The shortest video duration is How to Style videos. This video includes content related to fashion, how a do something, or fix it in a few minutes. This kind of short video category's duration smaller than others.

4. From figure 11 we can say that all videos' duration is short, around 5-10 minutes, From this investigation, we can say that short videos more popular than longer ones.

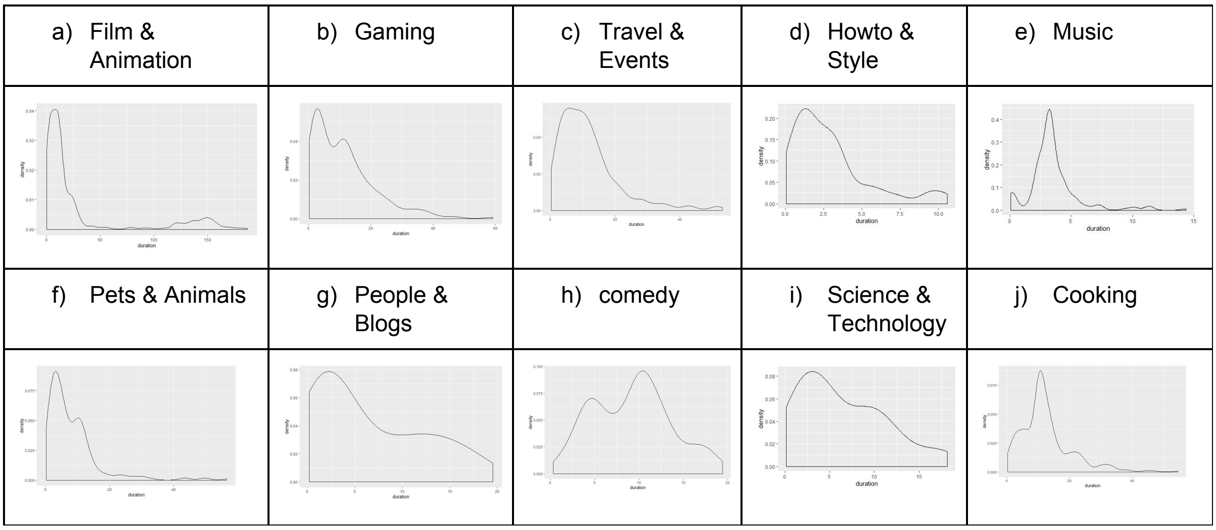


Figure 11: YouTube video duration Analysis plot in the different categories

5.5 Comment Analysis

Many web platforms used share nontextual content such as videos, images, animations that allow users to add to the comments. YouTube is the most popular of them [3]. For the investigation of YouTube video popularity, one of the important metrics is video comments. The best tool for investigation YouTube comments is sentiment analysis[20]. In sentiment analysis, we gathered comments from videos and tried to investigate it. What people are thinking about the videos, authors. Sentiments describe the audience rate about the video. Analysis of sentiments challenging for the limitation of sentiments algorithms on the internet. Most of the words in YouTube videos comments not an English or modified, unreadable. These kinds of situations create some additional problems and errors during the investigation of sentiments in the comments. YouTube has a sentiment rate for video verification. If total sentiment analysis positive the video will promote to the main page and the recommendation video list. The percentage of positive comments is one of the important popularity metrics. If video comments are mostly positive this means that the video has the chance to be popular (figure 13). As we can see from figure 13, most of the comments are neutral and positive comments are more than negative in our dataset. These related to that our dataset created mainly from popular videos.

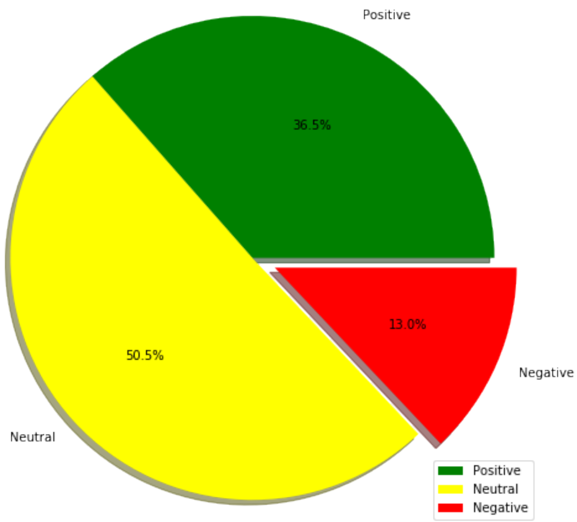


Figure 12: Comments sentiment percentage

Positive and negative percentage of sentiments in the comments are for every video category is different as we can see in figure 13. This mostly related to the content, for example:

1. Gaming videos have more negative comments than other categories in our investigation. This was related to the mainly spam unrelated videos in this category, most of the users not like certain games or gamers, and this impact in the comments.

2. The second unpopular category is in the comments is film and animation, this also related to poor content of movies or trailers.

3. Travel videos have the biggest positive comments percentage in our investigation and negative comments very low almost 5 percent. This related to the popularity of travel videos and most of these video content related to nature and new cities, countries, places. When users watch these kinds of videos no one wants to say something negative.

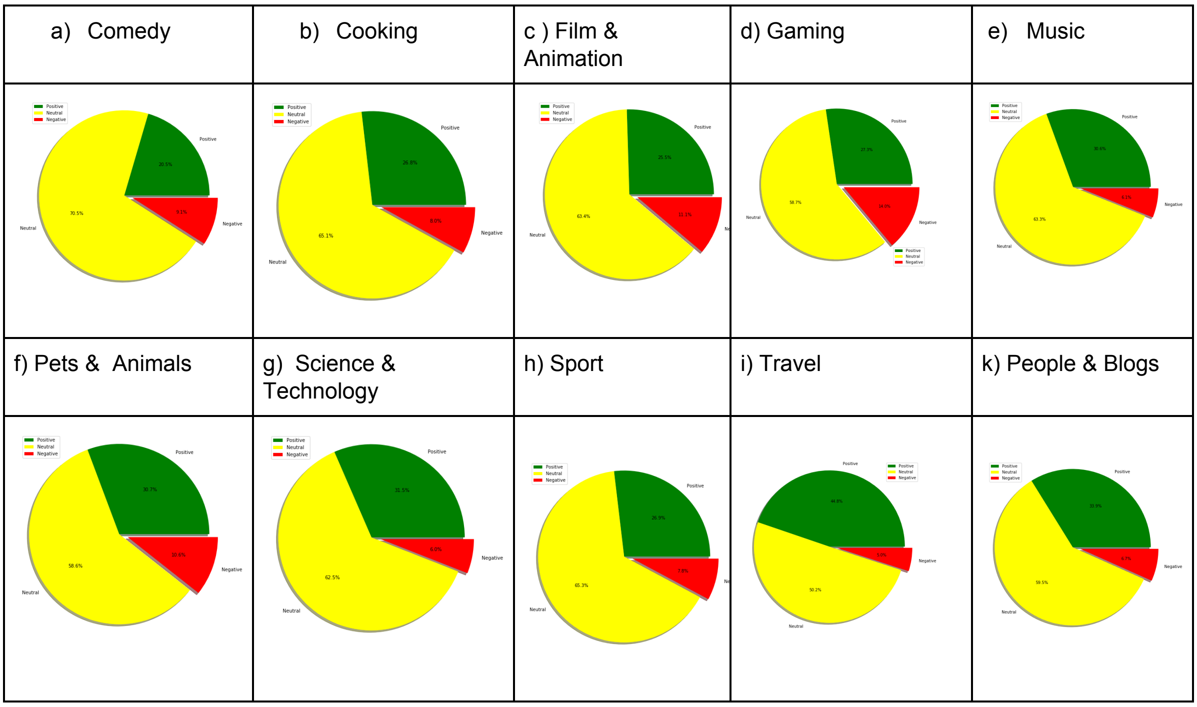


Figure 13: Sentiment percentage of different video category's comments

The next investigation step is world cloud analysis and sentiment analysis. Most used words in video comments are a key feature for the investigation video popularity. Because during search queries on YouTube also taking account of these comments like tags. From the figure 14 we can see different video categories sentiment comments analysis.

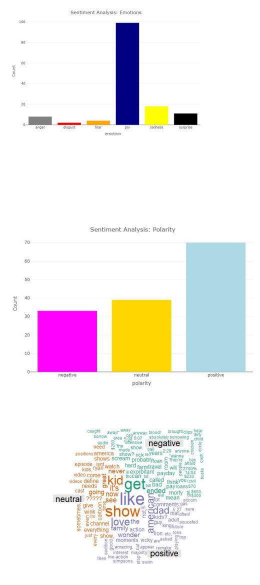
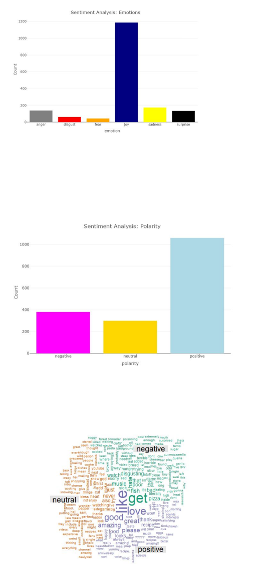
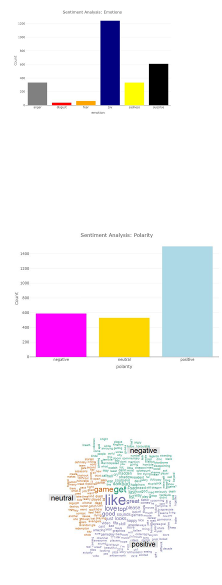
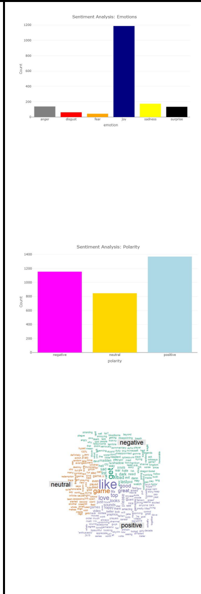
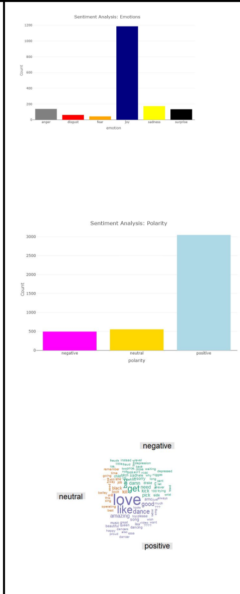
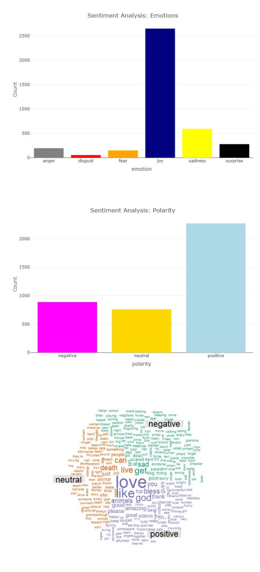
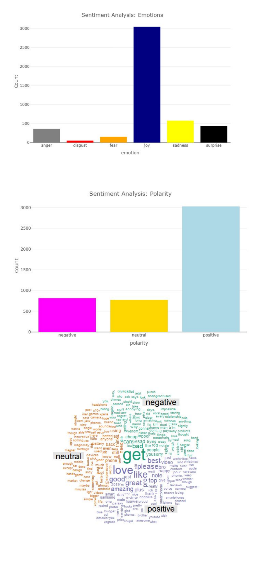
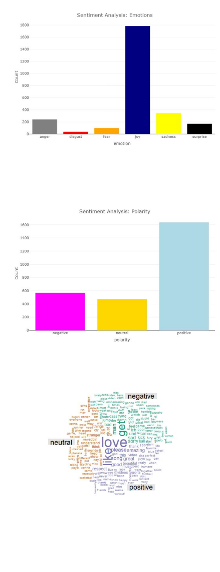
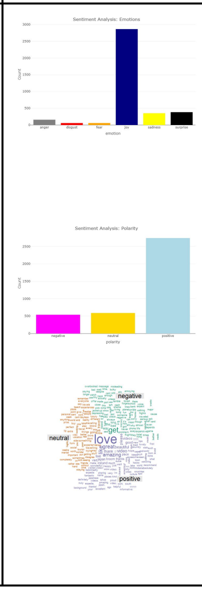
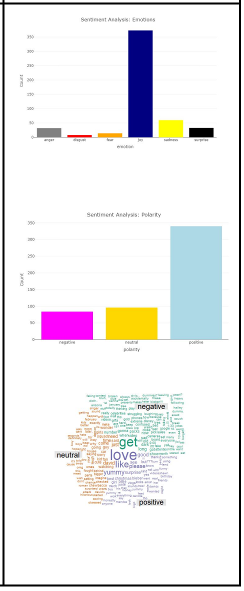
a) Comedy	b) Cooking	c)Film&Animation	d) Gaming	e) Music
				
f) Pets & Animals	**g)Science & Technology**	**h) Sport**	**i) Travel**	**k) People & Blogs**
				

Figure 14: Comment Table

5.6 Recommendation System Analysis

YouTube recommendation systems work principle one of the important aspects of YouTube video popularity analysis. For the title and description of the user-selected video and with the search query of user YouTube creates complicated recommendations network of videos during every act of user. YouTube uses keywords in the title, description, search query as search tags in the YouTube database, and demonstrates results of this search in the recommendations videos part.

To investigating best tags for video popularity, we did deep label analysis. We are picking tags from the recommendation videos from our main data set and created a network like shown in Figure 15.

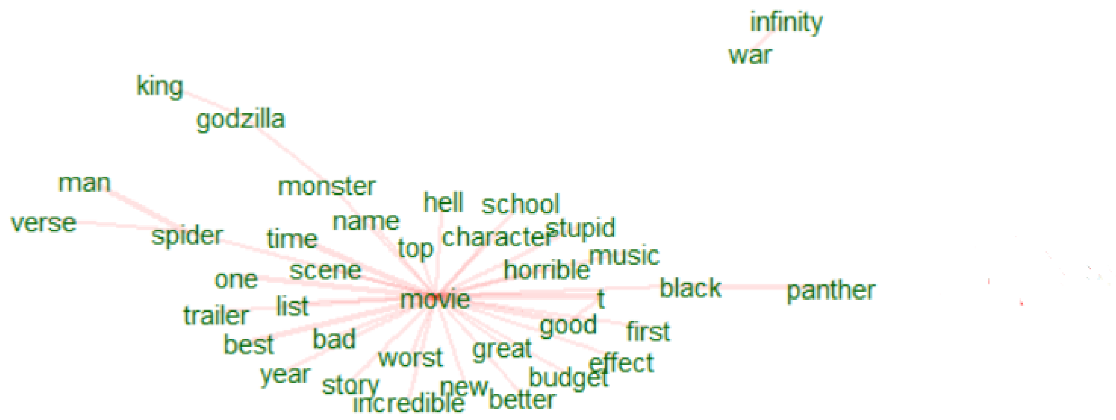


Figure 15: Tag network of "marvel movies" search query

Figure 15 created from search query "marvel movies" and select only the most viewed samples. From figure 15 we can see that the most used tag is "movie", this is the core of the tag of the network. Another secondary tag is "time", "name", "scene" and so on. From analysis like this, we found the best tags for a certain type of category or search query.

On the internet user content generator with the help of sites like YouTube Analytics can find the best keywords and tags for a certain type of videos. With the help of these tags, user content generators can increase several times the views of channels and videos.

5.7 Ten most subscribed channel analysis

Reference [12].

Rank	Channel	subscribers in million	Language	Category
1	T-Series	133	Hindi	Music
2	PewDiePie	103	English	Entertainment
3	Cocomelon - Nursery Rhymes	75	English	Education
4	SET India	67	Hindi	Entertainment
5	5-Minute Crafts	65	English	How-to
6	Canal KondZilla	56	Portuguese	Music
7	WWE	55	English	Sport
8	Justin Bieber	52	English	Music
9	Zee Music Company	51	Hindi	Music
10	Like Nastya	50	Russian	Entertainment

Table 5: Ten most popular channels

Additional to our previous work, we also analyzed the ten most popular channels for investigating the popularity of YouTube videos. In the table 5 showed 10 most popular channels. As you can see the most popular language on YouTube is English, the most popular video category is Music and Entertainment. Also, we analyzed this video duration most of them less than 5 minute and titles length is on average less than 50 characters, description length is on average less than 2000 characters.

We analyzed relationships of the main metrics with each other in these 10 channels in figure 16.

1. All these channels average views of more than 10 million.
2. Most liked and commented are entertainment category channels like "PewDiePie" and "SET India". According to that dislike count is very low, even in the "PewDiePie" close to zero.
3. Music video category channels have less view and more like and comment count.
4. In the education video category like and dislike counts are higher than other channels. In this channel comment count also very low.

[12] https://en.wikipedia.org/wiki/List$_{o}f_{m}ost-subscribed_{Y}ouTube_{c}hannels$

5. The interesting fact that How-to category channels like "5-minute crafts" view counts, like counts and comment counts, not very high but this category in the top 10.

6. Some channels turned off video comments ("Like Nastya" - channel).

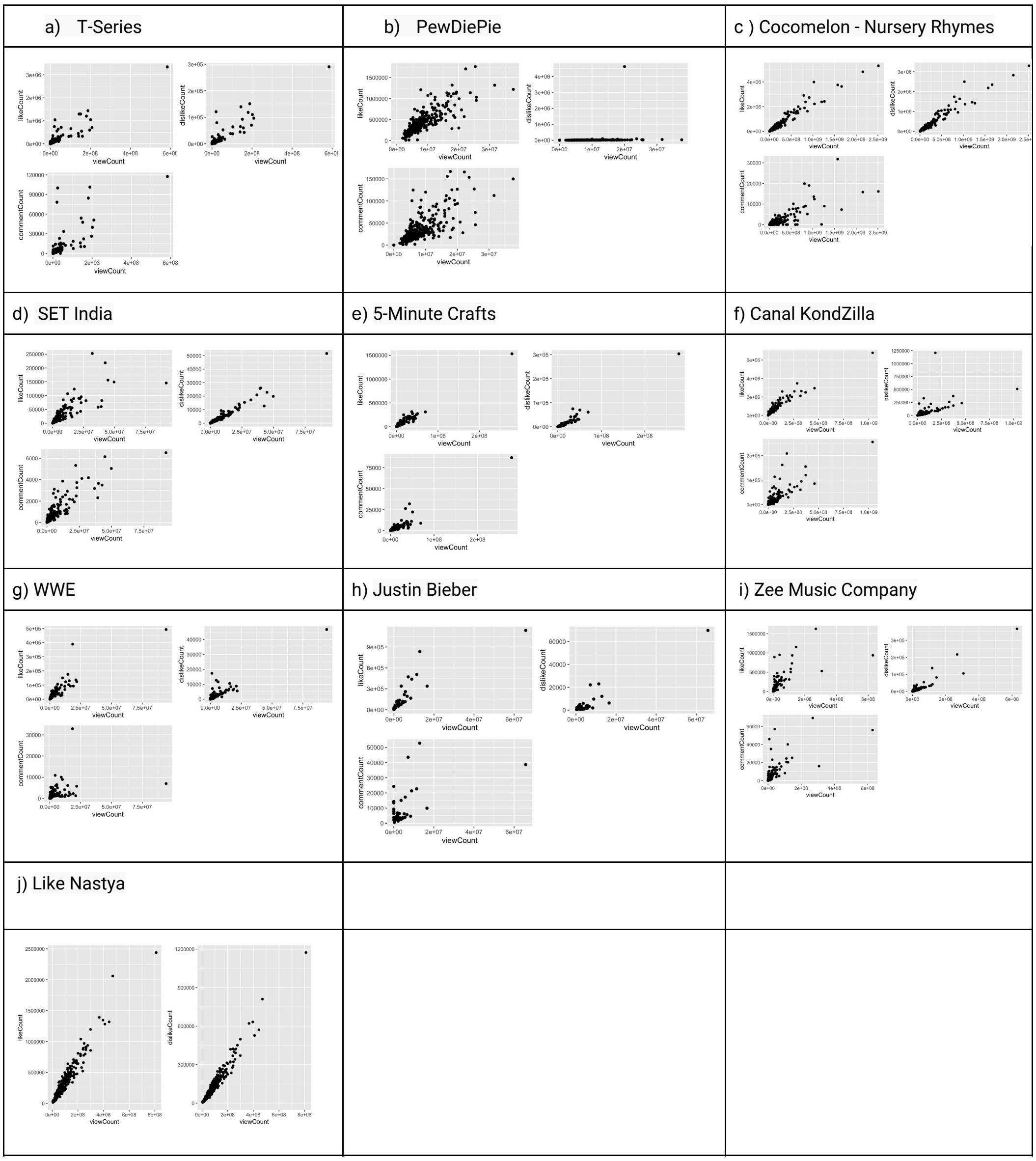


Figure 16: Top 10 Channels' main metrics relationship

6 Cross cultural analysis

Video streaming platforms such as YouTube offer videos to different cultures from around the world and allows us to understand these differences when it comes to video content. Based on a regular listing of the most popular YouTube videos in 10 countries, we examine the viewing of famous videos in countries that differs in the preference categories of content, titles and description length, video duration, relationships between likes, dislikes, comment counts and comments. Cross-cultural videos have a few differences:

1. Cultural differences can depend on:
   (a) Region
   (b) Religion
   (c) Mentality
   (d) Education
   (e) Life level

   For example, if we compare the Russian Federation and the United States, one can conclude that:

   (a) The most popular thing in Russia is the people and blogs YouTube video category figure: 17, plot g).
   (b) But in the United States, this is an entertainment YouTube video category (figure: 17, plot b).

   This was related to the life and education level of these two countries. In the United States people work more than 40 hours in the week and after hard working day most popular thing in the US is entertainment, but in the Russia political situation not a stable and unemployed people amount more than in the United States, for that people and blog category is most popular YouTube video category.

2. Another YouTube video difference in the various countries is the cooking culture as we know that every country has own cooking traditions, service styles, food types, and so on. For that's not easy to understand another country cooking tradition, this is one of the barriers to globalization.

3. Another significant barrier between countries is language as I am multilingual, I saw in my investigations that most of the videos are duplicates and have the same content with the other YouTube videos. The only difference between these videos is languages. Content creators using plagiarism to increase own YouTube video portfolio. We can see these kinds of things in Russian, Turkish video content creators they are mainly using the

United States YouTube videos as a source of plagiarism. The language barrier does not give people the chance to look at other countries' content and escape these poor content videos.

6.1 Cross Cultural Videos' category investigation

We are explored categories in the different countries for the few factors:

1. **Popular video categories around the world**

   During our investigation, we found a few interesting facts:

   (a) In rich countries like the US, Germany, Great Britain, France education YouTube video category videos not very popular, against that in poor countries like India and Mexico education category more popular than others. This big difference related to people's salaries, life conditions, and other aspects. We can see that in figure 17. As we can see from figure 17 that in most countries first, the popular YouTube video category is Entertainment. Of course, this is understandable as YouTube first created as an Entertainment video sharing platform. But if we look again at the figure 17 we can see that Russia and Great Britain Entertainment category are not in the first place. For example, in Russia, the first place is People and Blogs. As we say before this related unstable political situation in Russia. In the Great Britain the most popular video category is music , this related people's video taste , if look at the historical archives we will see that Great Britain been heart of culture for centuries[30].

   (b) As are say before everything depends on people's video taste. From figure 17 we can see that in Canada, India, Russia, and South Korea, one of the most popular categories in News and Politics. We think these related to increased nationalism movements. Sports videos are most popular in all of the countries in our dataset except Germany. In total across these 10 countries, the most popular categories are Entertainment, News and Politics, Music, People and Blogs, How to Style, and Sports.

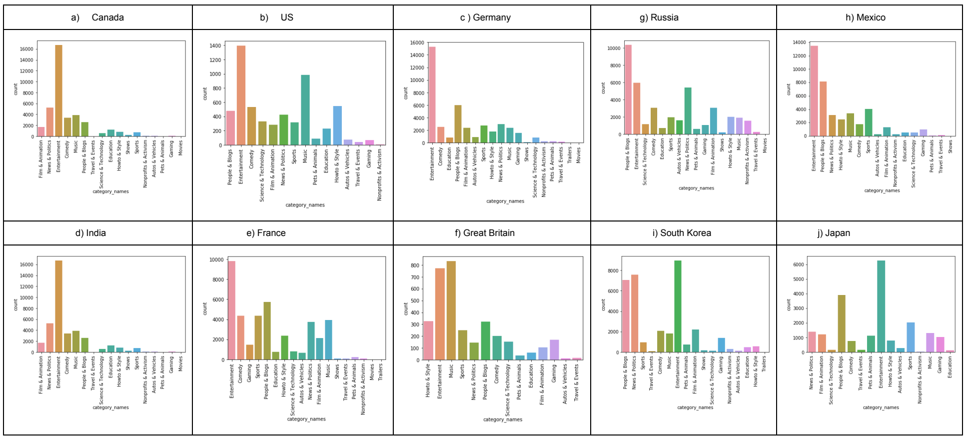


Figure 17: Popular video categories in 10 Countries

2. **Cross cultural most viewed videos' categories**

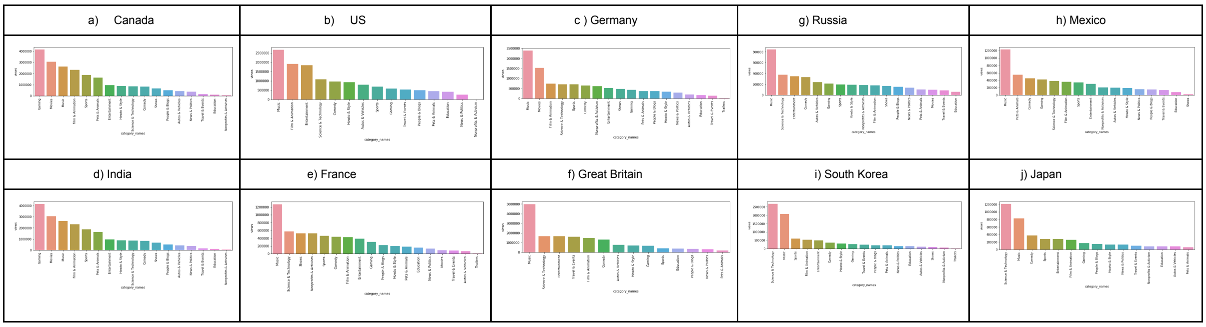


Figure 18: Cross Cultural popular video categories by views

Video views as we say before essential factor of video popularity. This understandable that video views are different for the various countries:

(a) From figure 18 we can say that most viewed video categories worldwide are music, movies, Film & Animation, and Gaming. Music is more popular than almost all countries except in India, Canada, South Korea, and Japan. Music is understandable why popular YouTube music video streaming popular in the last years. All popular singers first upload own songs to YouTube as a test platform, later to Spotify and iTunes.

(b) The second most popular content on YouTube is gaming, this is popular in India and Canada. These related to the industrial revolution in the gaming industry in these countries. If we look at the global statistics of game companies in India and Canada have more startups in the game industry than the United States and Europe. Also, one advantage that in these two countries the English language one of the popular languages.

(c) In South Korea and Japan's most popular content in Science and Technology, from our investigation, we can say that this related to the country's technology level and love to learn new things. Now giants like Samsung, Kia, Hyundai located in South Korea, and Japan located companies like Toyota, Sony, Mitsubishi. All of these companies have a big impact on the country's developing level and users' choices during the YouTube search. Science and Technology are also popular in Germany, Russia, France, Great Britain, and the United States. If we compare these countries' education system and industrial level, these two aspects are almost the same.

(d) **Why developed countries like Germany, the US, Canada, Great Britain Science and Technology video category not popular like in South Korea and Japan?** We found that this related to cultural differences, as these countries mainly use YouTube as an entertainment social platform. YouTube first started as a platform of the Entertainment videos for the English speaking countries. For in the European and American countries YouTube video categories like Entertainment, Film & Animation, Comedy, Music, more popular than categories like Science and Technology, Nonprofit and Activism, Education. In the Eastern countries, YouTube entered late in the previous decade and during this time YouTube analogs already captured the market. For that only Science & Technology and Music, YouTube video categories had a chance to be popular.

(e) In our investigation, we found that the Pets and Animals YouTube video category not very popular, but in Mexico this video category second most-viewed YouTube video category. We think that this related that Mexico is number 2 in the world of the household pets count, also Mexico popular with own farms even more than the United States.

3. **Cross cultural most liked and disliked categories**

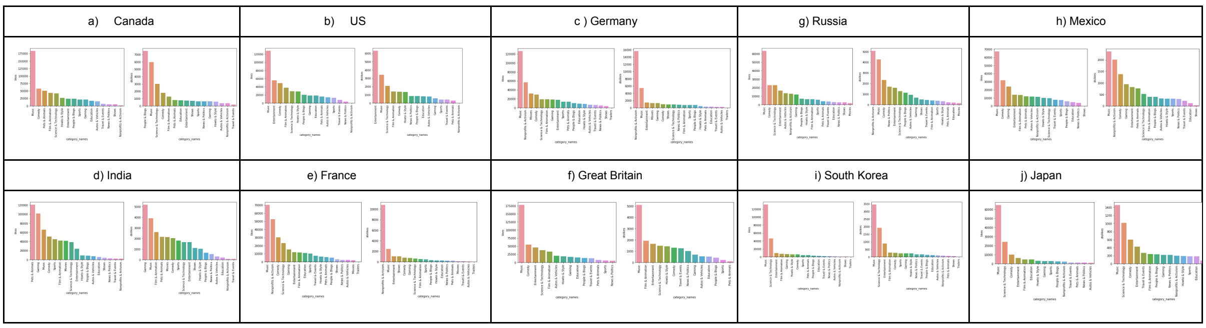


Figure 19: Cross-Cultural video like and dislike relationship with video categories

We have an interesting difference between video views and likes. After analysis of our countries dataset we found interesting relationships:

(a) Some less viewed like Pets and Animals YouTube video category have more likes than others.

(b) From figure 19 we can see that again, most liked the video category is a Music YouTube video category.

(c) In figure 19 we see that the most disliked categories are Nonprofit and Activism this happens in Germany, Russia, France, and Mexico. Unpopularity Nonprofit and Activism in these countries related to religious and cultural differences.

(d) We saw interesting similarities in South Korea and Japan, both of them mostly like and dislike Music and Science and technology. None of the other countries not interesting at all with Science and technology like these two countries.

(e) From our analysis of the most liked and disliked video categories by countries we can say that most of the cases popular and most viewed video categories have more dislikes than less popular video categories. This related to the discussion of these videos in social networks. For that during the analysis take into consideration dislikes only as an unpopularity indicator, not a fully correct.

(f) In the European countries YouTube video categories like News and Politics, Film and Animation, People and Blogs, Travel & Events not very popular. But in the American countries (the US and Canada) Film and Animation, People and Blogs YouTube video categories are popular against European countries.

(g) In the case of the percentage of the video categories South Korea and Japan very unusual. In these countries, popular video categories are only Music and Science & Technology. All other video categories have a very little percentage in the liked and disliked video categories plots. In Eastern countries, YouTube not fully captured the market of media.

4. **Cross cultural most commented videos' categories**

From figure 20 we can see most commented YouTube video categories in the different countries. Comments are one of the important metrics of YouTube. From our investigation of comments in these countries we came to such a conclusion:

(a) Usually most commented YouTube category is Music, but we have an exception in India, Germany, and France.

(b) In France and Germany most discussed YouTube video category is the Nonprofit and Activism video category.

(c) In India are Science and Technology. Science and Technology videos are very popular in India, but the view count is low.

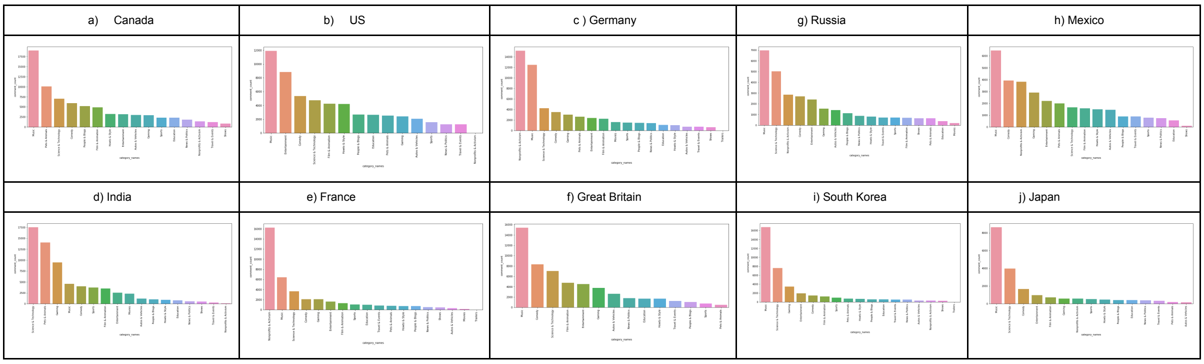


Figure 20: Cross Cultural video comment count vs video categories

6.2 Cross Cultural YouTube videos correlation analysis

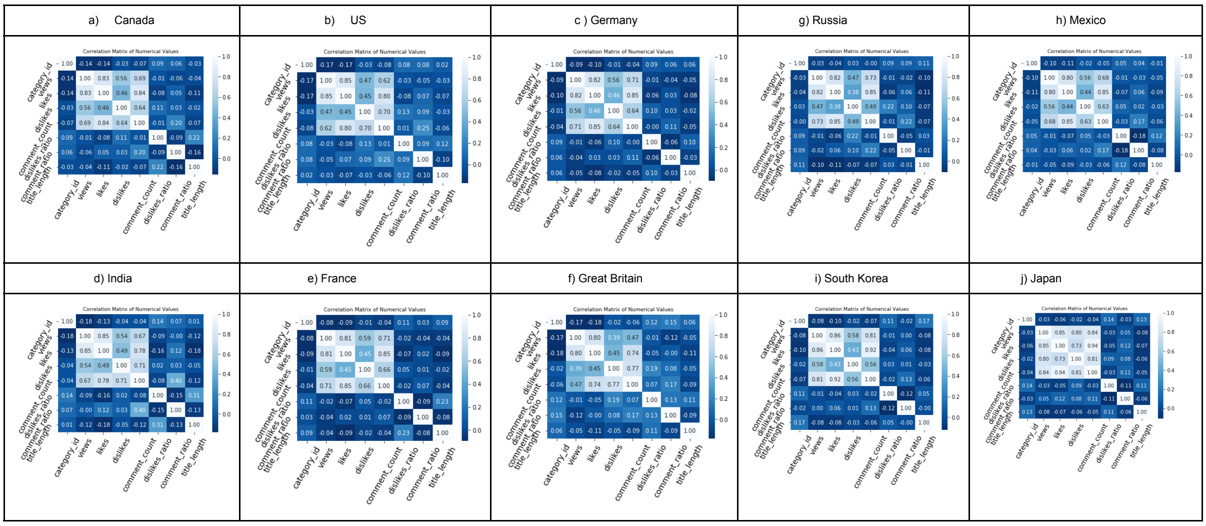


Figure 21: Correlation Matrix of YouTube Metrics in different countries

Additional to our investigation, we did a correlation analysis to find a relationship between video views, likes, dislikes, comment count, video categories, and title length. From figure 21 we can see this impact on YouTube metrics in different countries. We found that:

1. The length of the YouTube video title has a different effect in various countries. For example, in Russia, France, Japan, South Korea, and Great Britain have a big impact on the YouTube video category. In these countries, people like short video titles.

2. YouTube video category id has a positive correlation with the title length. This means that with the increasing category id videos' title also being longer. For example, in the Nonprofits and activism (29) title length longer than the Gaming video category (20). With the YouTube category id and the videos' view-count and likes have a negative relationship. With the big number of category id, the number of views and likes decreases. For example Gaming YouTube video category more popular than Nonprofits and activism YouTube video category.

3. YouTube videos' views and likes, comment count has a positive correlation with the videos' category id. Big number category ids have more comments than others.

Full list YouTube category id in the table 4.

6.3 Cross Cultural videos' Title analysis

From our investigation, we can say that title length difference related to these countries' cultures and education, no one likes to read the big title because they are can understand content video from few words. From our previous investigation on the random YouTube videos, it can be argued that better keep additional information in the video description part, this can help also in during a search and user content generator will not disturb users with big titles. Against that in India and Mexico title length not have a big impact on the video categories. Because in low economically developed countries, people not like to read additional information in the description, if we look title length in India and Mexico they are bigger than other countries. We can see that in figure 22.

During the analysis we found few things:

1. The total title length is around 40-60 characters, except videos from India, South Korea, and Japan (50-100 characters).

2. Most unusual is India video content generators titles bigger than most of them, around 100 characters. India video content creators add more details than needed in the titles, for that if we look at worldwide these videos not very popular in different countries.

3. We found that, YouTube videos title length less than others in developed countries, except in Japan and South Korea. This related to the level of education, with the knowledge level of people. Exception in Japan and South Korea related to cultural differences as they not too much globalized in case culture. Another exception is Mexico (not developed country, but the video title length like is developed countries), this related to the cultural pressure of the United States.

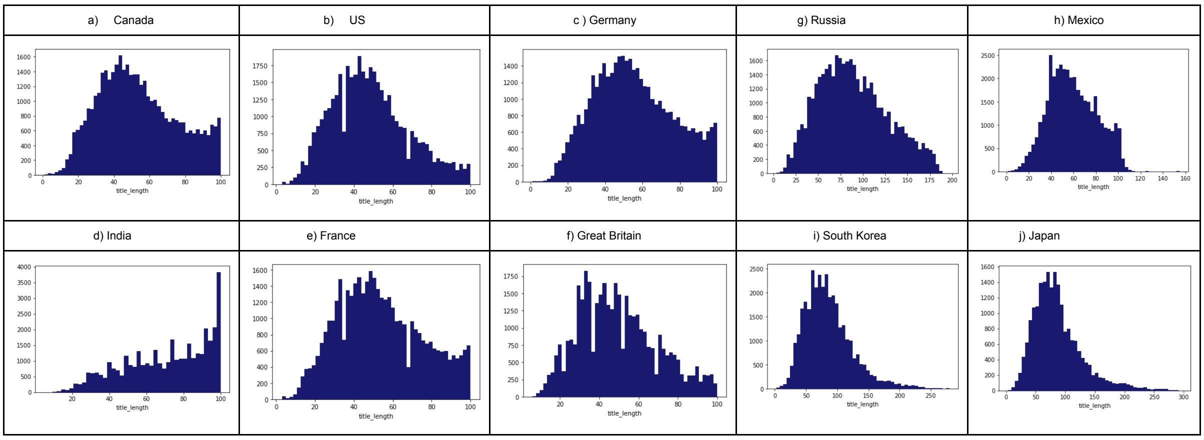


Figure 22: Title length of videos in different countries

6.4 Cross Cultural's videos description's analysis

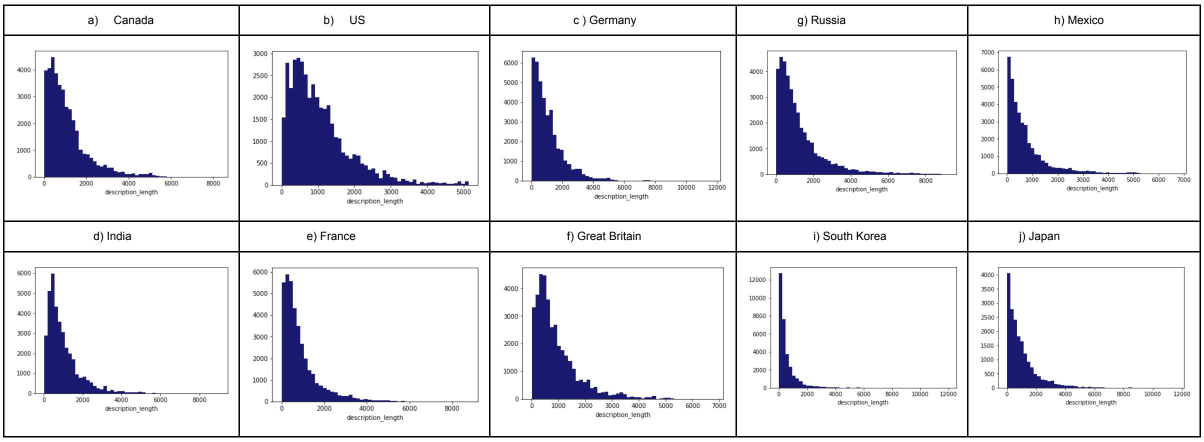


Figure 23: Description length of videos in different countries

One of the important metrics for the user content generator on the YouTube description length of video content. This length can be different in various countries. Between titles' length and descriptions' length have inverse proportionality. We can see that if we compare title length figure 22 and description length figure 23 . For example, in developed countries like Canada and the US description length is long, the title length is short. In India and Mexico description length is shorter than other countries. Like in the title length analysis, South Korea and Japan are an exception in the developed countries' statistics. In these two countries, the description length is shorter. Most of countries descriptions' length less than 2000 Characters.

6.4.1 Cross cultural videos duration analysis

One of the video parameters in which user content generators can control this is the video length (figure 24). Most of the country's video length is between 5 and 10 minutes, except for

Russia and Mexico (15-20 minutes). In India average video duration less than 5 minutes. Video duration has a close relationship with video categories. For example:

1. Film&Animation,People&Blogs,News&Politics and Comedy YouTube video categories usually longer than others. These videos in the first five popular videos in the Russia and Mexico.

2. How to Style&Do, Music, Science&Technology,Entertainment videos usually between 5-10 minutes. These categories usually popular most of the countries.

Also, we see a different distribution of video duration in figure 24. For example, in Canada, the US, Germany, India distribution of videos changes very smoothly. These mean that in these countries video categories popularity distributed very accurately and YouTube captured the whole online media market. But in Russia, South Korea, Great Britain, France, and Japan video duration distribution are not changed smoothly. YouTube only covers a few popular video categories in these countries. From that we can say that, in these countries YouTube have other opponents in the market.

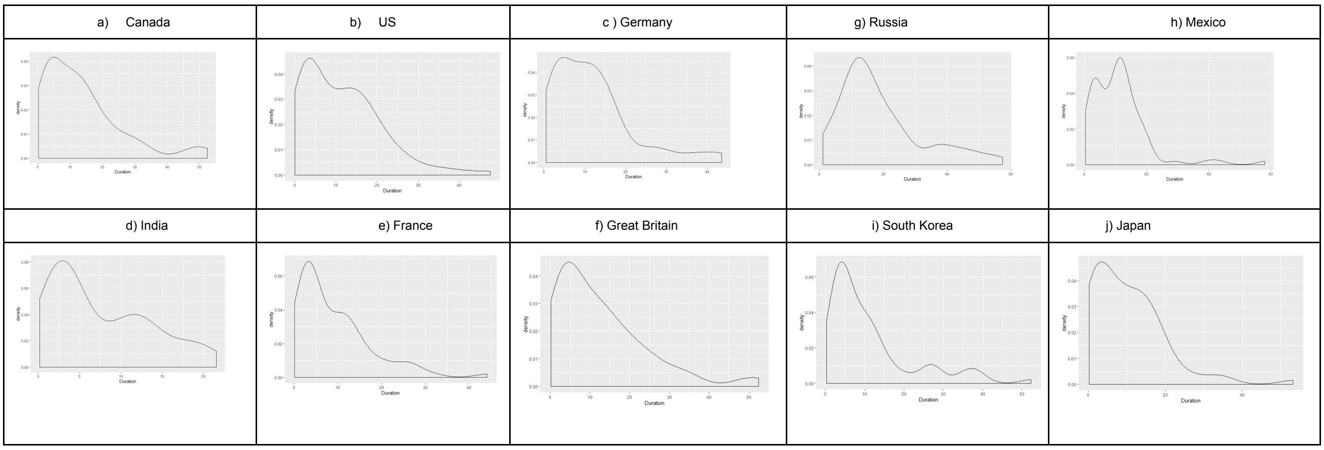


Figure 24: Cross Cultural YouTube video duration analysis

6.5 United States and Great Britain comments analysis

In the investigation comments differences in the various countries, we did sentiment analysis two English speaking countries, the United States and Great Britain(25). In the US positive comments 3 percent more than GB, and negative less 1 percent. These differences related to these two country video category choices, for example, in the US most popular categories are People & Blogs and Comedy, in Great Britain Music and How to Style. Music and How to Style mostly related to the feelings and of these contents negative comments are normal. From the word cloud and most used word plots, we can see that most of the used words in the comments are the same. But in the US some of the words are a little bit rude like niger word which very frequently uses in the US comments.

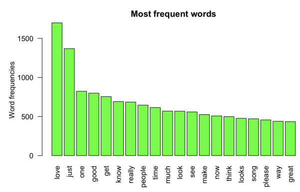
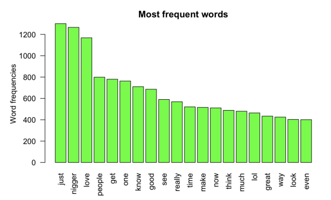
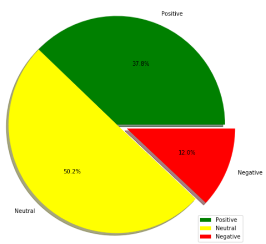
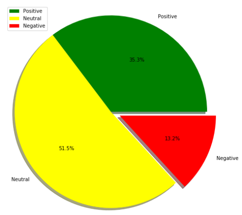
Country	Great Britain	United States
Most used words in comments		
Word Cloud		
Sentiments percentage		

Figure 25: Differences in comments US and GB

7 Summary

During our investigations, we found that popularity on YouTube depends on a few factors like: video duration, the title of the video, a description of the video, user comments, video view count, video likes to count, and dislikes count. All of these YouTube metrics among themselves divide into two categories.

1. **Observation 1.** *The first one depends on the user content generators like:* video duration length, the format of title the video title and description.

2. **Observation 2.** *The second, which doesn't depend on the user content generators only depends on the users like:* video view, like, dislike and comment count.

After analysis of the thousands of the popular videos, we decided that most of the cases for the popularity of the videos YouTube metrics should like in the table 6:

Metric name	Range of the metrics
Views Count	10K - 1M
Like Count	0 - 2K
Dislike Count	4K - 0.5M
Comment Count	0 - 5K
Description Length	1000 - 2000 Characters
Title Length	50 - 100 Characters
Video Duration	5 - 15 minutes

Table 6: YouTube metrics range

After analysis of our YouTube dataset we observed that:

1. Most of popular channels have more than million subscribers.

2. Sentiment analysis of the comments very important for the investigation video popular or not. From the analysis of sentiments of comments , we came to such a conclusion that popular videos' positive comments more that 25 percent most of cases. And not every popular video have only positive comments, some part of comments are negative.

3. After analyses of more than millions of video duration, we can say that the best duration is 10 minutes for the average. But this can change based on different video categories. For example, for music, this is 3-4 minutes, comedy 5 minutes, movies 1 hour, comedy around 45 minutes.

4. Relationships between videos' likes and dislikes counts very necessary for video popularity, as an understandable reason like counts, should be more than dislikes counts.

5. Title and description lengths and words used in there also very important for the video future destiny. Because longer titles and descriptions are not popular around the users.

6. Last and one of the important metrics on YouTube is a video category. Video category one of the important metrics using during search queries of YouTube. User content generators should choose an accurate video category when uploads video to the YouTube network, as some video categories are very similar like Film&Animation and comedy.

7. During our investigation, we found that in the different countries' tastes of video content can be different, for example:

    (a) In Russia, China, Japan, and Latin American countries people like long titles and descriptions.

    (b) But in countries like Germany, the united states, France's short length, and short description length videos are popular.

    (c) Also the same thing in the duration of videos in western countries like Germany, France, and the United states short length videos are popular. All these related people's free time, in economically developed countries no one has time for long videos.

    (d) With the different countries video category popularity also different, for example, developed countries most popular things is a video game, how to style, how to do, entertainment videos.

    (e) But in the third world countries is mainly popular politics videos, education, travel, people, and blogs. And in countries like Japan, South Korea mostly popular is Science and Technology videos. All of these related people's lifestyles, the problems, and current political situations in their own country.

    (f) After appearing corona virus, videos like health, medical advice, personal hygiene more popular than before. This shows that popular trends can change in a few days.

8 Conclusion

YouTube is one of the most popular social media platforms which attract billion users every day. User content generators are posting regularly new videos to remain popular among their followers and to gather new followers. In this thesis, we studied popular YouTube videos, by first collecting videos which have 1M views or 100K following of channels. We analyzed the relationship between key popularity metrics (viewcount, likecount, commentscount, comments, dislikecount, titleLength, descriptionLength, videoLength).

8.1 Challenges

During our investigations we meet with the few problems.

1. First and important of them YouTube API v3 which we used for to get public data have limit 10,000 query in one day. For this limitation, our big dataset scraping codes could not work and we faced regularly crash during the execution of our program. To fix this problem we gathered dataset with a small parts.

2. Data manipulation and cleaning of dataset also took a lot of time, for the size of dataset simple cleaning algorithm some times took hours.

3. COVID 19 and quarantine. We lost a lot of time for the this pandemic.

8.2 Future Directions

This research is a first step in understanding YouTube popularity, which provides the initial foundation for future explorations. This work can be improved using the following directions:

1. We might would like to collect and analyze a bigger data set.

2. We would also like to add more categories in our analysis.

3. We would also like to understand the time series analysis of video's popularity.

References

[1] The evolution of youtube, Feb 2019. URL `https://www.bluearcher.com/blog-item-evolution-of-youtube`.

[2] C. S. Araújo, G. Magno, W. Meira, V. Almeida, P. Hartung, and D. Doneda. Characterizing videos, audience and advertising in youtube channels for kids. In *International Conference on Social Informatics*, pages 341–359. Springer, 2017.

[3] M. Z. Asghar, S. Ahmad, M. Qasim, S. R. Zahra, and F. M. Kundi. Sentihealth: creating health-related sentiment lexicon using hybrid approach. *SpringerPlus*, 5(1):1139, 2016.

[4] N. Ashar, H. Bhatt, and S. Mehta. A framework for detection of video spam on youtube. 2015.

[5] R. Bandari, S. Asur, and B. A. Huberman. The pulse of news in social media: Forecasting popularity. In *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.

[6] M. Bärtl. Youtube channels, uploads and views: A statistical analysis of the past 10 years. *Convergence*, 24(1):16–32, 2018.

[7] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and K. Ross. Video interactions in online video social networks. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 5(4):30, 2009.

[8] A. Brodersen, S. Scellato, and M. Wattenhofer. Youtube around the world: geographic popularity of videos. In *Proceedings of the 21st international conference on World Wide Web*, pages 241–250. ACM, 2012.

[9] C. Cayari. The youtube effect: How youtube has provided new ways to consume, create, and share music. *International Journal of Education & the Arts*, 12(6):n6, 2011.

[10] F. Cena, E. Chiabrando, A. Crevola, M. Deplano, C. Gena, and F. N. Osborne. A proposal for an open local movie recommender. In *PATCH 2013: Personal Access to Cultural Heritage*, pages 1–6. CEUR Workshop Porceedings, 2013.

[11] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 1–14. ACM, 2007.

[12] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. Analyzing the video popularity characteristics of large-scale user generated content systems. *IEEE/ACM Transactions on Networking*, 17(5):1357–1370, Oct 2009. doi: 10.1109/TNET.2008.2011358.

[13] G. Chatzopoulou, C. Sheng, and M. Faloutsos. A first step towards understanding popularity in youtube. In *2010 INFOCOM IEEE Conference on Computer Communications Workshops*, pages 1–6. IEEE, 2010.

[14] C. Chau. Youtube as a participatory culture. *New directions for youth development*, 2010 (128):65–74, 2010.

[15] X. Cheng, C. Dale, and J. Liu. Statistics and social network of youtube videos. In *2008 16th Interntional Workshop on Quality of Service*, pages 229–238. IEEE, 2008.

[16] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296, 2010.

[17] Q. Do. Youtube channel analysis. *RPubs*, 2018.

[18] K. Dougan. "youtube has changed everything"? music faculty, librarians, and their use and perceptions of youtube. *College & Research Libraries*, 75(4):575–589, 2014.

[19] K. English, K. D. Sweetser, and M. Ancu. Youtube-ification of political talk: An examination of persuasion appeals in viral video. *American Behavioral Scientist*, 55(6):733–748, 2011.

[20] R. Feldman. Techniques and applications for sentiment analysis. *Commun. ACM*, 56(4): 82–89, 2013.

[21] A. Finamore, M. Mellia, M. M. Munafò, R. Torres, and S. G. Rao. Youtube everywhere: Impact of device and infrastructure synergies on user experience. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 345–360, 2011.

[22] F. Gelli, T. Uricchio, M. Bertini, A. Del Bimbo, and S.-F. Chang. Image popularity prediction in social media using sentiment and context features. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 907–910, 2015.

[23] F. Giglietto, L. Rossi, and D. Bennato. The open laboratory: Limits and possibilities of using facebook, twitter, and youtube as a research data source. *Journal of technology in human services*, 30(3-4):145–159, 2012.

[24] L. Hamzaoui Essoussi and D. Merunka. Consumers' product evaluations in emerging markets: does country of design, country of manufacture, or brand image matter? *International Marketing Review*, 24(4):409–426, 2007.

[25] P. Haridakis and G. Hanson. Social interaction and co-viewing with youtube: Blending mass communication reception and social connection. *Journal of Broadcasting & Electronic Media*, 53(2):317–335, 2009.

[26] M. Holland. How youtube developed into a successful platform for user-generated content. *Elon journal of undergraduate research in communications*, 7(1), 2016.

[27] A. Hossain. Youtube analysis on 'bangladesh'. *RPubs*, 2018.

[28] G. Kruitbosch and F. Nack. Broadcast yourself on youtube: really? In *Proceedings of the 3rd ACM international workshop on Human-centered computing*, pages 7–10, 2008.

[29] A. Madden, I. Ruthven, and D. McMenemy. A classification scheme for content analyses of youtube video comments. *Journal of documentation*, 2013.

[30] H. Meyer-Delius, Harald, H. Meyer-Delius, and Harald. The 20 types of youtube videos with most views in the uk, Feb 2020. URL `https://blog.printsome.com/top-20-types-of-videos-with-most-views-on-youtube/`.

[31] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42. ACM, 2007.

[32] J. C. Paolillo. Structure and network in the youtube core. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, pages 156–156. IEEE, 2008.

[33] M. Park, J. Park, Y. M. Baek, and M. Macy. Cultural values and cross-cultural video consumption on youtube. *PLoS one*, 12(5), 2017.

[34] K. Pires and G. Simon. Youtube live and twitch: a tour of user-generated live streaming systems. In *Proceedings of the 6th ACM multimedia systems conference*, pages 225–230, 2015.

[35] E. Protalinski. Youtube ditches flash for html5 video by default, Jan 2018. URL `https://venturebeat.com/2015/01/27/youtube-ditches-flash-for-html5-video-by-default/`.

[36] S. Qin, R. Menezes, and M. Silaghi. A recommender system for youtube based on its network of reviewers. In *2010 IEEE Second International Conference on Social Computing*, pages 323–328. IEEE, 2010.

[37] D. Reino, B. Hay, et al. The use of youtube as a tourism marketing tool. 2016.

[38] C. Snelson. Youtube across the disciplines: A review of the literature. *MERLOT Journal of Online Learning and Teaching*, 2011.

[39] A. Sobecki. Service recomendation on wiki-ws platform. 2015.

[40] G. Sood. tuber v0.9.8. URL `https://www.rdocumentation.org/packages/tuber/versions/0.9.8`.

[41] A. Susarla, J.-H. Oh, and Y. Tan. Social networks and the diffusion of user-generated content: Evidence from youtube. *Information Systems Research*, 23(1):23–41, 2012.

[42] P. a. B. TechPostPlus, Vishalsays:, T. F. 23, TechPostPlussays:, P. Author, N. Mughalsays:, T. M. 19, Samiamsays:, T. M. 22, Omsays:, and et al. Youtube video categories list faqs and solutions, Mar 2020. URL `https://techpostplus.com/2019/04/26/youtube-video-categories-list-faqs-and-solutions/`.

[43] A. Vedantham. Making youtube and facebook videos: Gender differences in online video creation among first-year undergraduate students attending a highly selective research university. 2011.

[44] M. Wattenhofer, R. Wattenhofer, and Z. Zhu. The youtube social network. In *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.

[45] D. Williams, S. J. Sullivan, A. G. Schneiders, O. H. Ahmed, H. Lee, A. P. Balasundaram, and P. R. McCrory. Big hits on the small screen: an evaluation of concussion-related videos on youtube. *British journal of sports medicine*, 48(2):107–111, 2014.

[46] L. YouTube. Youtube. *Retrieved*, 27:2011, 2011.

9 Appendices

YouTube data scraping scripts :
https://github.com/IsmayilOfficial/YouTubeScrap

YouTube datasets :
https://github.com/IsmayilOfficial/YouTubeDatasets
https://www.kaggle.com/datasnaek/youtube-new

YouTube Sentiment Analysis scripts:
https://github.com/IsmayilOfficial/YouTubeSentimentAnalysis

Cross Cultural Popularity Analysis scripts:
https://github.com/IsmayilOfficial/crossvideopopularity

YouTube Video Duration Analysis scripts:
https://github.com/IsmayilOfficial/Videoduration

Title Analysis scripts:
https://github.com/IsmayilOfficial/TitleAnalysis

Description Analysis scripts:
https://github.com/IsmayilOfficial/DescriptionAnalysis

Correlation Between YouTube metrics Analysis scripts:
https://github.com/IsmayilOfficial/CorrelationAnalysis

Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Ismayil Tahmazov**,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   Understanding popularity on YouTube ,

   Supervised by Rajesh Sharma

2. I grant the University of Tartu a permit to make the work specified on p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons license CC BY NC, ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive license does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Ismayil Tahmazov
01/05/2020