

TRENDING YOUTUBE VIDEO STATISTICS

AAKANKSHA TASGAONKAR

AMOGH MAHESH

NANJESH GOWDA

TANVI GAWADE

LOUIS TAN

MSIS graduate students

California State University, Los Angeles

Abstract: This project will demonstrate the usage of cutting-edge technology like Elastic Search and Kibana. We will apply the knowledge learned during the lecture, extensive researches and development of Logstash to import the data and use Kibana for visualizations of procured data. Goal of this analysis is to include an insight into YouTube statistics such as the trending videos, the most liked/viewed categories, trending YouTube channels based on these categories. This also includes a well-planned and mapped statistical analysis of the data over a given period.

URL: <https://www.kaggle.com/datasnaek/youtube-new>

Dataset Size: 199 MB

File Format: CSV

Number of Files: 10

Server's OS: Windows OS

Memory Size: Elastic search- 931.5 GB; Kibana- 1.4 GB

CPU Speed: 1.9 GHz

1. Introduction

YouTube (the world-famous video sharing website) maintains a list of the top trending videos on the platform. According to Variety magazine, "To determine the year's top-trending videos, YouTube uses a combination of factors including measuring users' interactions (number of views, shares, comments and likes). This dataset is a daily record of the top trending YouTube videos. The dataset includes several months (and counting) of data on daily trending YouTube videos. Data is included for the US, GB, DE, CA, and FR regions (USA, Great Britain, Germany, Canada, and France, respectively), with up to 200 listed trending videos per day. We would analyze this data to get insights into YouTube trending videos, to see what is common between these videos. Those insights might also be used by people who want to increase popularity of their videos on YouTube. Goal of analysis would include insights such as which YouTube channels have the largest number of trending videos, which video category (e.g. Entertainment, Gaming, Comedy, etc.) is liked and voted by people with respect to each country, how are views, likes, dislikes, comment count, title length and other attributes correlate with each other, which video remained the most on the trending-videos list, When are the most number of view of public achieved during the day (which would include time), Statistical analysis over time.

2. Data Processing

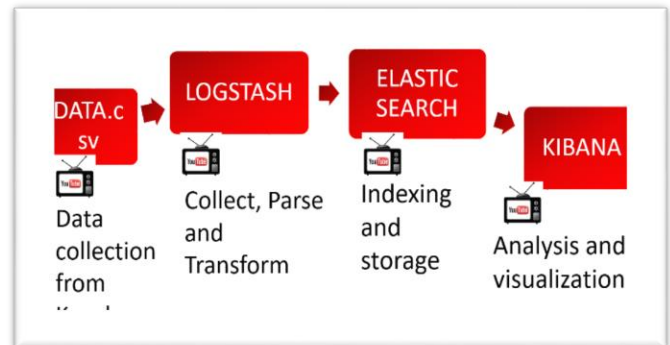


Figure 1: Data Processing

- Dataset downloaded from Kaggle.com in the comma separated format.
- Part of Elasticsearch stack i.e. Logstash to import, parse and transform the data into the elasticsearch cluster.
- Elastic Search for mapping and indexing data formats.
- Kibana for analysis, visualizations and dashboards.

3. Data Visualizations

3.1 Trend of Views in a Timeframe/ Monthly basis

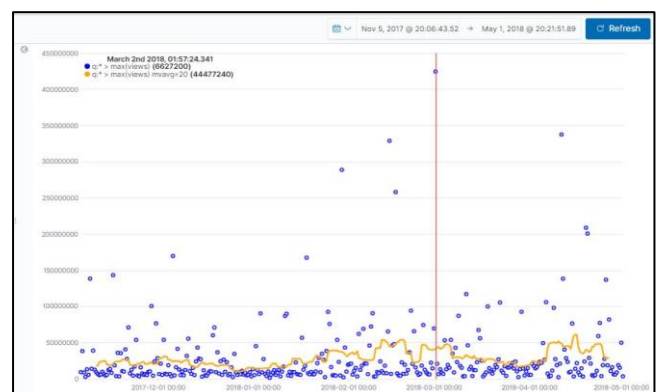


Figure 2: Trend of Views in a Time Frame.

The above time series graph depicts the trend analysis pattern of number of public viewer who are active on various Youtube channels on a monthly basis. The timeline is segregated on a monthly basis to have an overview of how

actively are the audience viewing youtube videos. We can analyze how many users are viewing youtube channels in a particular time frame. From the graph it is evident that heavy traffic was clocked by Youtube in the months of March and April.

3.2 Trend of Views in a Timeframe/ Weekly basis

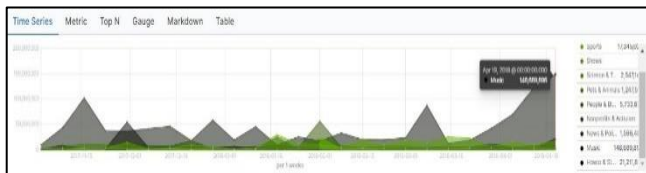


Figure 2: Highest Views for Specific Time Frame - Per Week

We now need to dig in deeper to know what the public are viewing specifically on the youtube channels especially during the timeframe with maximum number of views i.e. March 2018 – April 2018. In the above time series representation, we understand the Youtube category that is most widely preferred by the users on a weekly time frame. The graph shows that Music is the most preferred category amongst the users and it is observed to be in a greater trend in the months of March and April. Viewing of the Music category on the youtube channel is well distributed and is more in number compared to all other video categories or genre.

3.3 Trend of Views in a Timeframe/Daily basis

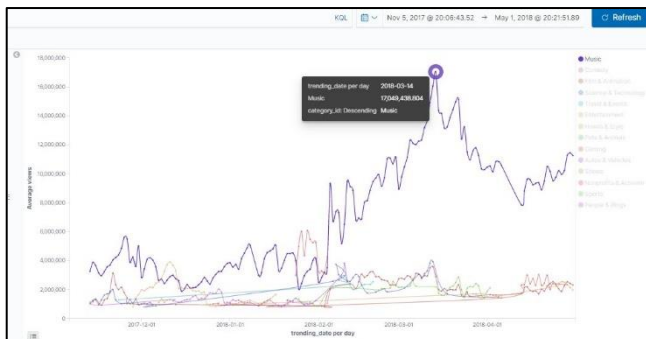


Figure 3: Trending Categories – per Day.

Furthermore analysis was performed to see what the daily trend of these categories looks like. This graphical representation gives insight about the categories that are trending on a daily basis. The above analysis depicts that Music is the most preferred Youtube category on a daily basis, which means that majority of the users in a particular time frame used Youtube to watch Music Videos. Category Music depicts a greater potential in attracting more number of viewers and we can determine the likelihood of this category to grow in the upcoming future. Other notable categories such as Science & Technology, Film and Animation has also caught the public's eye on a daily basis.

3.4 Trending Youtube Channels



Figure 8: Favourite Youtube Channels.

We now drilled to analyze that Music category was the most viewed category amongst all video channels streamed over Youtube which further necessitated the need to analyze which Youtube channels were the most liked based on the categories analyzed earlier. The above tag cloud representation shows the frequency of liking a Youtube Channels. The greater the size means most favourite and vice-versa. We now understand that the audience like what they view on these particular channels. It can be observed that Childish Gambino Vevo is the most favourite Youtube channel following Drake Vevo, Ariana Grande so on and so forth among the users which was further researched and found that Childish Gambino's "This Is America" became the 31st song to debut at number one on the US Billboard Hot 100, becoming both Gambino's first number one and top ten single in the country. It has also topped the charts in Australia, Canada, and New Zealand. The song won in all four of its nominated categories at the 61st Annual Grammy Awards: Record of the Year, Song of the Year, Best Rap/Sung Performance and Best Music Video. This made Gambino the first hip-hop artist to win Record of the Year and Song of the Year, and "This Is America" the first rap song to win these awards.

3.5 Worldwide Youtube Category Likings



Figure 4: Heat Map showing Country-Wise Category Liking.

Above analysis depicts the country-wise preference of a particular Youtube category using the heat map. From the above analysis it is clear that the audiences in the European nations have a greater affinity towards the music categories,

following which are the nations like United States, South Korea, Canada. Entertainment holds a higher place after the music videos amongst these Nations. Furthermore nations such as the Europe and United Kingdom are the leading countries in the category Nonprofit and Activism which is a special program streamed on Youtube. The YouTube Nonprofit Program helps nonprofits connect with supporters, volunteers, and donors. It allows any qualified not-for-profit group to increase donations and reach other important organizational goals through video and their YouTube Channels.

3.6 Geospatial analysis of Comment Counts/Country

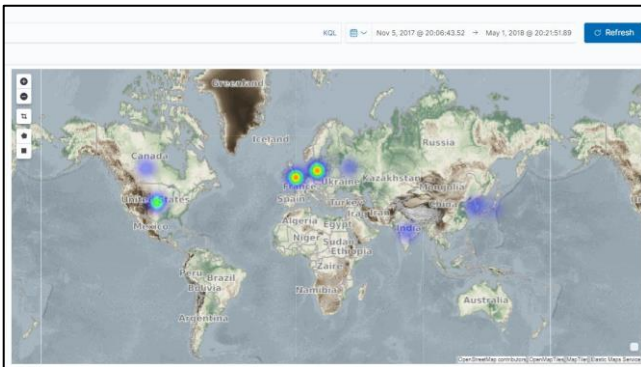


Figure 7: Heat Map showing Comment Count per Country.

The above geographical heat map represents the number of comments per location. From this representation we can find out the location having maximum number of users who comment on Youtube videos. It can be seen that in countries like United Kingdom, France and Germany the users have commented more number of times on the Youtube videos. Whereas the trend gets weaker as we come across countries like India and Japan.

4. Summary

- We successfully used the cutting edge technology learned in class such as Elastic Search and Kibana.
- Youtube videos depicted a great trend in viewing and liking in the months of March and April 2018.
- During the trending timeframe i.e. Nov 2017 to May 2018, music was the most preferred Youtube category on a daily, weekly and monthly basis.
- European Nations and United Kingdom contributed for the maximum viewership in categories such as Music and Entertainment which also comprised maximum active users who commented on the Youtube Videos.
- European Nations and United Kingdom were the most active nations who contributed the most on the programs – Nonprofit and Activism.
- Childish Gambino-Vevo was the most favourite Youtube channel among the users during the year 2018 which also won the Song award of the year.

4. Github URL

<https://github.com/tanvigawade/ElasticSearch-Kibana>

5. Reference

- <https://www.elastic.co/webinars/elastic-maps-for-geospatial-analysis?baymax=rtp&storm=ribbon-3&elektra=products-maps&iesrc=ctr>
- <https://www.elastic.co/guide/en/elasticsearch/reference/master/keyword.html>