# **YouTube Channel Data Analysis Report**

**INTRODUCTION**

YouTube has started in 2005 and has grown to become the second largest search engine in the world which processes more than 3 billion searches per month. It is, however, generally a myth how the YouTube algorithm works, what makes a video get views and be recommended over another. In fact, YouTube has one of the largest scales and most sophisticated industrial recommendation systems in existence. For new content creators, it is a challenge to understand why a video gets video and others do not. There are many "myths" around the success of a YouTube video, for example if the video has more likes or comments, or if the video is of a certain duration. It is also worth experimenting and looking for "trends" in the topics that YouTube channels are covering in a certain niche.

Having recently stepping into the content creation world with a new YouTube channel on data analytics and data science, I decided to gain some insights on this topic which might be useful for other new content creators. The scope of this small project is limited to educational channels, and I will not consider other niches (that might have a different characteristics and audience base). Therefore, in this project will explore the statistics of around 40 most successful educational YouTube channels.

# **Objective:**

* Analysing video data and verifying different common “myths” about what makes a video do Ill on YouTube
* Do the number of likes and comments matter for a video to get more views?
* Does the video duration matter for views and interaction (likes)?
* Does the title length matter for views?
* How many tags do best performing videos have? What are the common tags among these videos? Across all the creators, I take into consideration as to how often do they upload new videos, on which day of the week and which month?
* Which channel posts more? Total Videos versus Channel.
* Explore the trending topics using Natural Language Processing (NLP) techniques
* Which popular topics are being covered in the videos (e.g., using word cloud for video titles)?
* Analysing the count of video uploads before and after COVID-19

# **Dataset:**

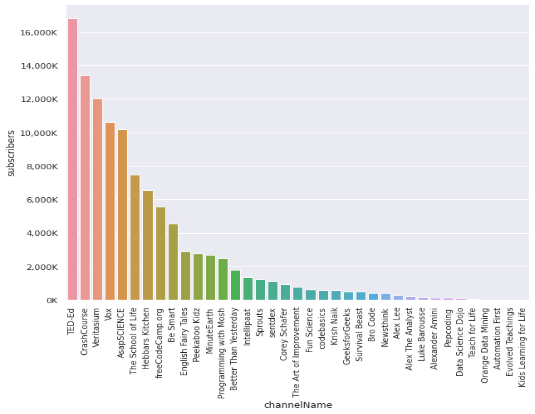
As this project particularly focuses on educational channels, there are not many readily available datasets online suitable for this purpose. The YouTube Data API document explains basic concepts of YouTube and of the API itself. It also provides an overview of the different functions that the API supports. The API is intended for developers who want to write applications that interact with YouTube. The default quota allocation for each application is 10,000 units per day.

I requested an authorization credential (API key). Afterwards, I enabled YouTube API for our application, so that I can send API requests to YouTube API services. Then, I went on YouTube and checked the channel ID of each of the channels that I would like to include in project (using their URLs). Then I created the functions for getting the channel statistics via the API.

# **Application Programming Interface (API):**

APIs can generate massive amounts of value both internally and externally. Managing and processing data is one of the crucial factors in business management, and every company has built IT systems. As the size of the data is increasing exponentially, however, there is a limitation to cope with all the data through the traditional IT system. In this sense, applying API can be a solution with better efficiency and security. It can break down barriers between systems, which enables simplifying work processes, inter-cooperation between organizations and higher protection of data. To work with API, you need to get an authorized key first. It’s for obtaining an authorized key to be connected with API. API Providers, YouTube in this case, don’t merely provide their service without control. To preserve this interface and manage the users, they offer a unique access key to each user. With this, I can connect to the interface of the application

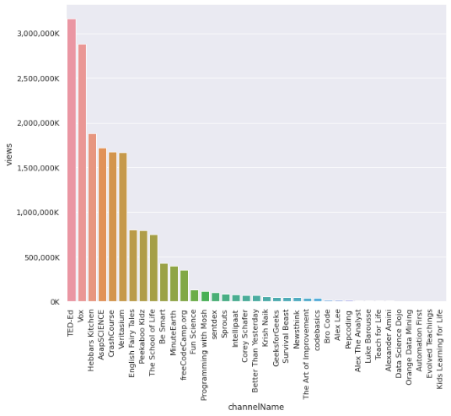
# **Subscribers Vs Channels:**



Subscribers are the key to a Channel. The number of Subscribers per Channel will show us how popular the channel is when compared to the other channels. I will get a fair idea of how the number of subscribers often correlates with the amount of influence a channel has. The graph shows the order of channels based on the subscriber count. The top two channels with the most subscribers are TED-Ed and Crash Course. The channels with the least subscribers are Evolved Teachings and Kids Learning for Life.

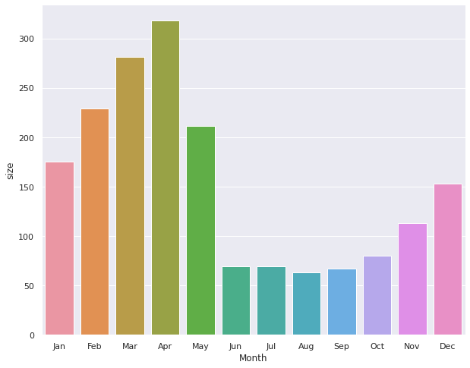
YouTube streamlines channels display for its channel subscribers, when it comes to displaying videos from subscribed channels by prioritizing the channels, they interact most with.

# **Channels Vs Views:**



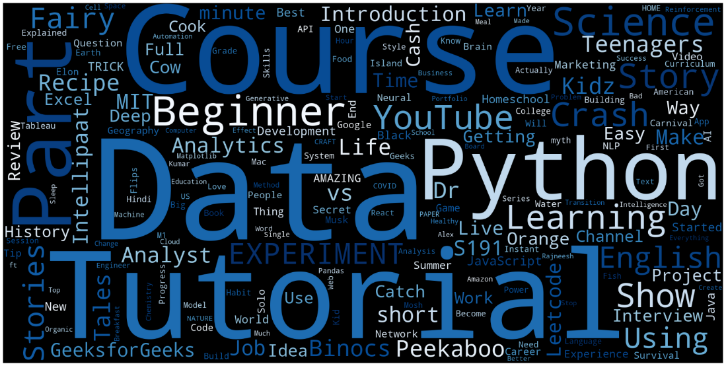
I will investigate the rank of a channel considering the total number of views of a channel. Most of the time, the rank is fair like the subscriber count. Here **Ted Ed** remains the most popular channel considering both subscribers and views. Though the subscribers for the channel **Crash course** are more than for the channel Vox, the vox channel got maximum views. Some channels have more subscribers with less views. Some channels have more views with fewer subscribers. So, the views and subscribers are not always directly proportional.

# **Videos per Month:**



In this graph I are trying to show which Ire the preferred months to post a video on YouTube the best time can get a channel as many as 2 to 5 times more views. It is essential that they post videos intelligently, rather than at any old random time. In general, overall viewership on YouTube dips in May and September. And here I see the same phenomena in terms of educational videos. These are transitional months where most YouTube's users, which are in large part school-age people and young adults, are either finishing or starting school, sports seasons, vacations, and/or doing lots of outdoor activities

# **Word Cloud:**



Natural Language Processing, or NLP (Natural Language Processing) is defined as the automatic manipulation of natural language, like speech and text, by software. This Word Cloud shows the most frequent words like Data, Course, Tutorial etc. used in these video titles. The bigger the word, the more frequent is its occurrence in the video title. Hence the size of the word is directly proportional to its occurrence in a word cloud.

# **Preferred Days to Post:**

Chart, bar chart

Description automatically generated

I see in the graph above, the best time to upload your video to YouTube is during Tuesdays and Wednesdays (weekdays). The reason for this is that videos typically pick up their most views in their first two days. If the channel's second-busiest day is the day before its busiest day, try posting on that day instead. That way, channels benefit from having updated content on the two most active days for your channel.

# **Title length Vs View Count:**

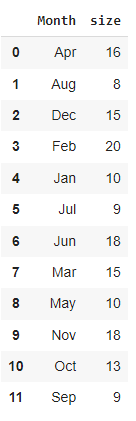
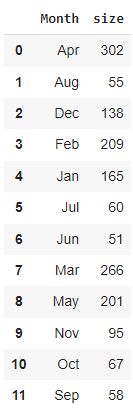
**Chart, scatter chart

Description automatically generated**

There is no clear relationship between title length and views as seen the scatterplot below, but most-viewed videos tend to have average title length of 30-70 characters.

# **Pie-Charts for Pre and Post Covid videos:**

# **Pre\_covid\_videoCount** **Post\_covid\_videoCount**

Chart, pie chart

Description automatically generated Chart, pie chart

Description automatically generated

As per the above pre-covid and post-covid depiction graphs for the number of videos posted throughout all the months, I could see that there is a visible increase in the number of videos posted for channels post-covid.

**Conclusion:**

I took only some features of channel videos for our analysis in this project. But there is more scope in this stream for analyzing the data of channel videos and here are some recommendations for the same:

I can do sentiment analysis on videos based on the comments and feedback on our videos if they are positive, negative, or neutral. Another recommendation that I could produce comparing with other niche areas are visual appeal against explanation through speech and words, making shorts, comparison of organizing the playlists, effective tagging - these help in making the channels more popular.

These days people have truly little time to spare. Short videos get more attention. Channels that gain more attention with shorter videos can have more followers and thereby can be an enormous success. Identifying these market gaps also helps the channel to stay in trend. A frequent and periodic analysis of channels helps in making better quality videos with interesting delivery methods, attracting more viewers and subscribers. Expectations and suggestions from users implemented to make better videos as to how Ill the content is presented, what is missing, what is expected from the viewers, what are suggestions and aspects to be considered while publishing the next video etc.

Some tips like expressive language, visual brand identity conveying authenticity, good color choices throughout the channel, custom YouTube thumbnails that lead to increased click rates, using search engine optimization for channel visibility will help the channel's popularity thereby also increasing its revenue generation. This analysis and knowledge about channels also help in appreciating the efforts of channel owners and encourage others to see trending ways and tips of showcasing their knowledge to their audience.

By: -

Surekha Kandula