JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Youtube Dashboard Usingn Streamlit

Shaikh Mohd Ashfaque ,Ansari Mohd Naqeeb, Affan Ansari , Daanyal Parbulkar, Khan Shoeb Raza

Abstract: Our dashboard will have 5 windows:

1) Aggregate video analysis 2) Individual video analysis 3) Growth of channel in future 4) Top 10 Most Viewed, Liked, Subscribed, Commented and Shared Videos 5) Top 25 words which gives most and least attraction. In the aggregate video analysis, it will show performance of videos in the last 6 months. In aggregate video analysis it will have parameters: Views, Likes, Subscribers, Shares, Comments added, RPM(USD), Average% viewed Avg duration sec, Engagement Ratio, Views/sub gained. RPM. Revenue Per Mille (RPM) is a metric that represents how much money you have earned per 1,000 video views. RPM is based on several revenue sources including: Ads, Channel memberships, YouTube Premium revenue, Super Chat and Super Stickers. In the individual video analysis there two parameters: 1) To which country the viewers belong to 2) Percentile comparison of the current video with respect to other videos for viewers country we will be using line chart and for percentile comparison we will be using a scatterplot chart. All the positive insights will be having green color and all the negative insights will be having red color.

Introduction:

Given the massive popularity of the Internet, the production of knowledge and its modes of circulation and consumption, along with the media in which it is objectified, are showing evidence of transformation. These

modes cause tension, not only in the process of production, but also in the modern institutions, which have historically been in charge of (re)producing it. Furthermore, there are many questions and tensions surrounding formal education1 , in particular, concerning the strategies associated to the production and transmission of knowledge as well as the actors involved in this process. On the other hand, both informal and nonformal2 education are gaining ground in these processes of production and transmission and becoming acknowledged spaces for this purpose. In this scenario, digital technologies in general, and platforms in particular, take the centre stage. YouTube seems to be playing a leading role in this context, although this role is surely to expire soon. Nonetheless, given its relevance, it is worth exploring how the students from low-income sectors who attend technical schools access, use and perceive it in order to identify its usage and how they build an interaction between the needs arising from extracurricular concerns and those arising from formal education. YouTube is an exclusive platform intended to produce, store, share, visualize and assess online audio-visual content. Basically, "the site takes user generated content to fill their catalogue, generate user traffic and attract advertisers"). In addition, the AVC is produced by amateurs and professionals alike with various goals that range from "simply" sharing a production to its monetization.

YouTube is not only serving as an entertainment platform for the films and the television industry but it has also emerged as a learning platform for many students. Content creators on YouTube, commonly known as "YouTubers" are pushing their content every single day and hour to be relevant to their audiences. It is a known fact that the YouTube algorithm is not a publicly available code and it is kept private for most of the time. Also, YouTube has shared any intuition about what factors are considered

for a video to be in trending section which leaves its audience in dilemma about posting videos. The trending section is where some videos are listed as trending and it is designed in such a way that every user of this platform checks it once in a while. This dashboard aims to present the trending section of YouTube for India in terms of the trends and observations which could be helpful for creators to push their content to more audiences.

I. Review of Literature:

YouTube is one of the most popular platforms and a lot of research has been done on it. Despite its importance, YouTube trending videos analysis has yet to be thoroughly investigated. Trending video analysis still has a lot of room for improvement. Ouyang, Li, and Li investigated the prediction of internet video popularity. The Dashboard was divided into five tasks in this study: Aggregate video analysis, Individual video analysis, Growth of channel in future, Top 10 Most Viewed, Liked, Subscribed, Commented and Shared Videos, Top 10 words which gives most attraction. With a core set of variables and numerous categorization algorithms, researchers first anticipate the future popularity levels of videos. The study then used sophisticated regression models to forecast the number of views based on the popularity levels. The impact of meta-data elements such as title, tag, thumbnail, and description on the popularity and trendiness of YouTube videos was investigated by Hoiles, Aprem, and Krishnamurthy. Their study used a variety of Machine Learning algorithms to predict the YouTube video popularity based on the video's meta features as well as other factors like the number of subscribers. The study used a views ratio per category to provide a simplified output of views, likes, dislikes, and comments scatter plot. Using pre-processing analysis, the thesis aids in understanding the value of these attributes.

II. METHODOLOGY:

Python script was created to fetch the data from the YouTube Data API. The YouTube data API has a clean interface to obtain any type of data from their platform. It authenticates via API keys. The key was generated by registering an account under Google Cloud Platform (GCP), then creating a project and then enabling the YouTube API under that project.

The data returned by the YouTube API is of the nested JSON format that further needs data wrangling. Also, one issue we had to deal with was the paginated results. The JSON results were paginated with each page having token for the next one. Here is one of the JSON output for a video .

Data wrangling usually entails converting and mapping data from one raw format to another in order to facilitate data consumption. The data should not have any null, special or unrecognized characters.

We created the logic to fetch the data but now the question was how to manage all the files and load them at the same time. We used glob to find all the filenames with the extension ".csv" and loaded them one by one in master Pandas Data Frame. So now every time there is new csv file in the data loading folder, that file contents will be loaded automatically into the master Data Frame.

After loading this Data Frame, there were several cleaning and memory reduction processes which were applied on this Data Frame. These steps include:

Reducing memory usage: All the integer data types are inferred as int64 which is not an efficient approach to store a value as small as one digit. It affects speed of execution for further data analysis. Therefore, these also need to be in the correct integer range.

Filling null values. These are the rows whose values are null. We can see that null values are denoted by NaN. These null values are replaced with an empty string, in case of string type column and 0, in case of integer column.

The date columns should be in datetime format used by Pandas. It helps in accessing the specific date attributes. Also, the time zone returned by the YouTube is in the UTC.

Title and description cleaning for emojis.

Populating the category data for the videos

The publishing date is converted to Pandas datetime format and then the time zone conversion takes place. We also removed the emojis from the channel title and description. Calculating the title length of the video

A dashboard for analytics is a set of widgets that enable quick data viewing. A dashboard is a reporting tool that allows website analysts to measure numerous data such as online conversions, visitors, and page views to

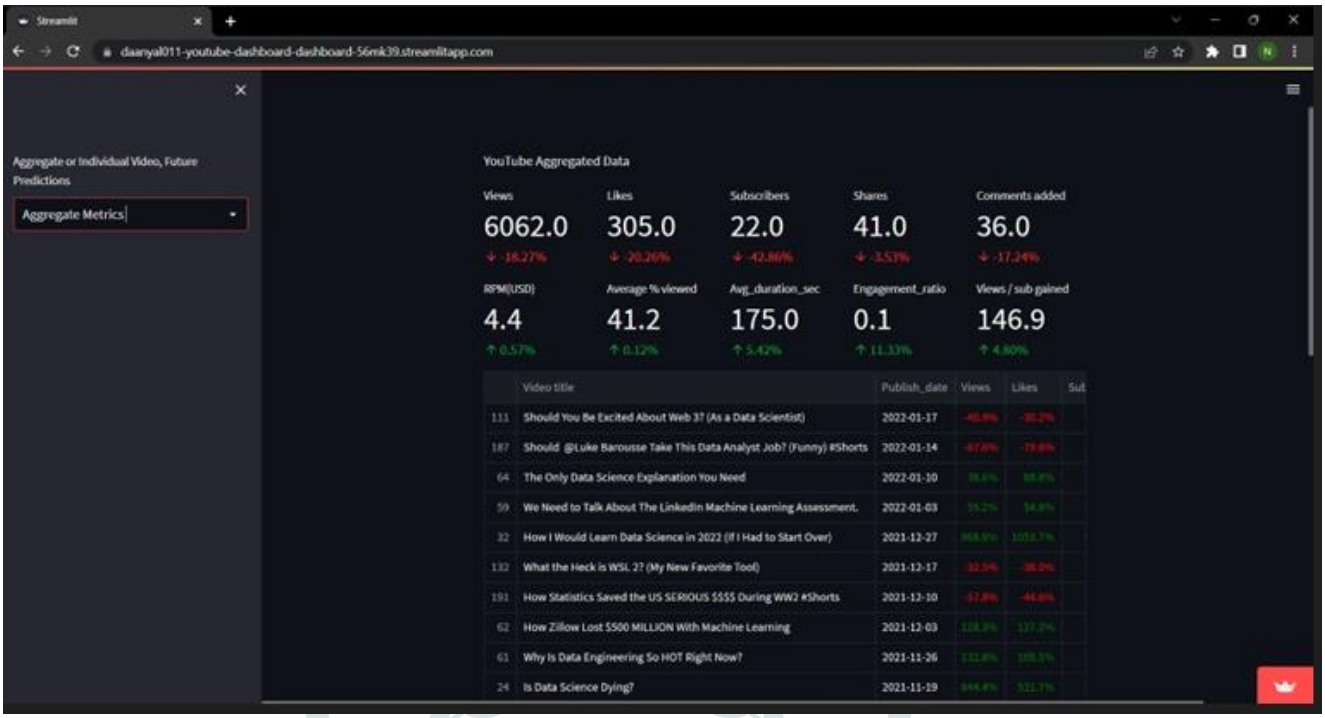
conveniently monitor the operation of a website. The analytics dashboard is flexible, powerful, and simple to use, which is one of its best advantages.

We created a dashboard that is interactive in nature and presents the charts in the form of visualizations. The visualizations are created keeping in mind that an average person can understand the context of the findings by just looking at the graphs.

III. BLOCK DIAGRAM:

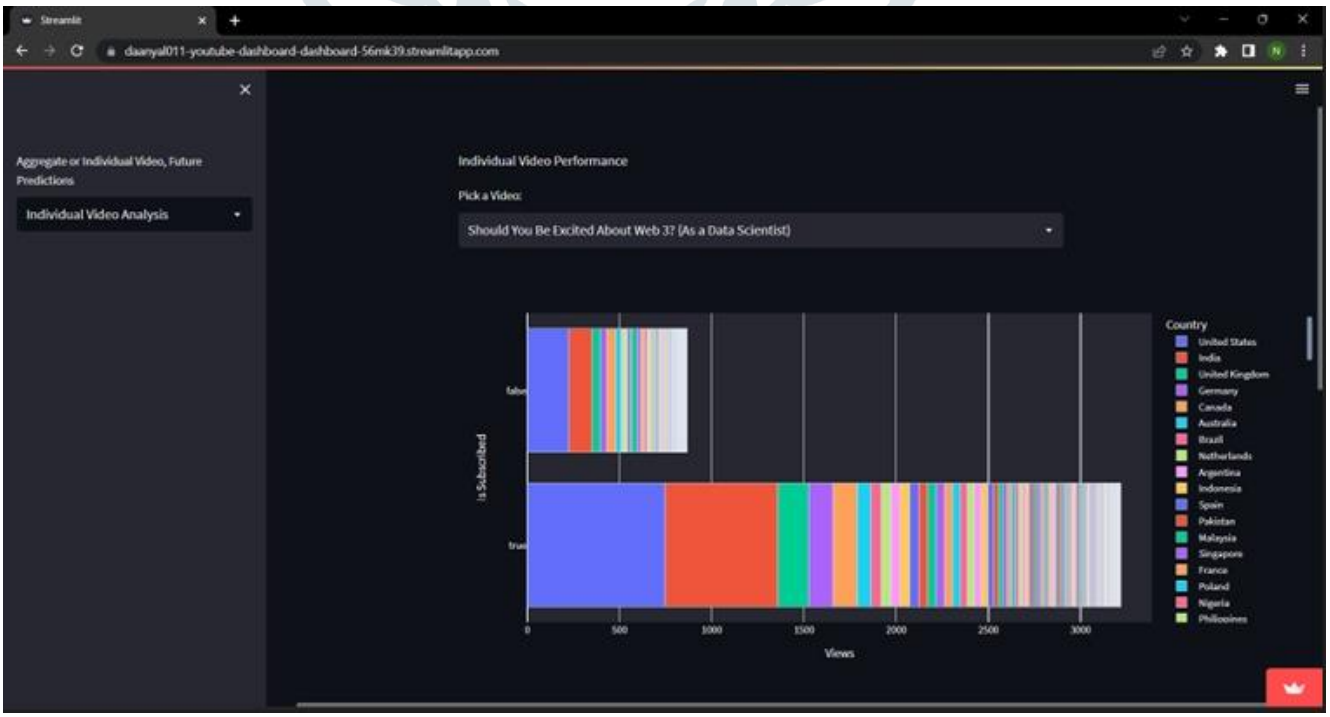
IV. RESULT:

1. Aggregate metrics

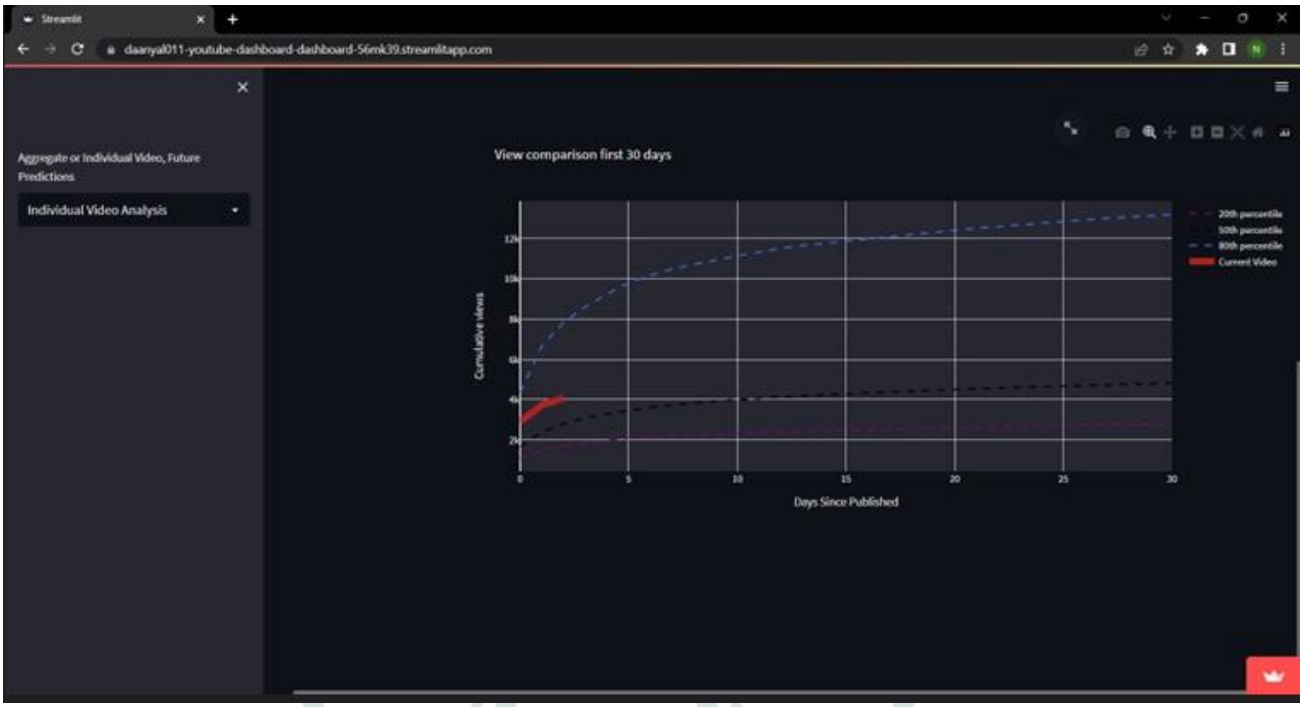


It shows the aggregate metrics of the videos for the past six months. That same parameters are shown in the form of pivot table for each Individual video also.

2. Individual Video Analysis

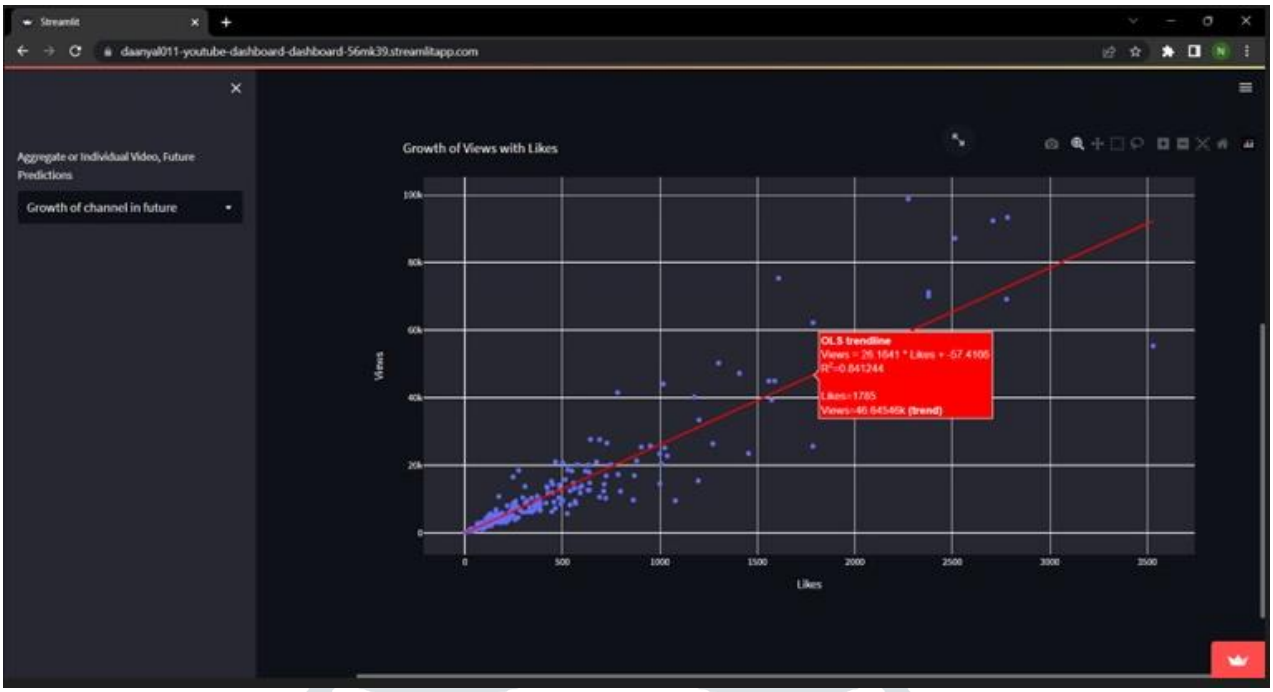


It shows which country gave him the most views in the first 30 days since uploading the video and how many of the viewers were subscribers and non-subscribers.



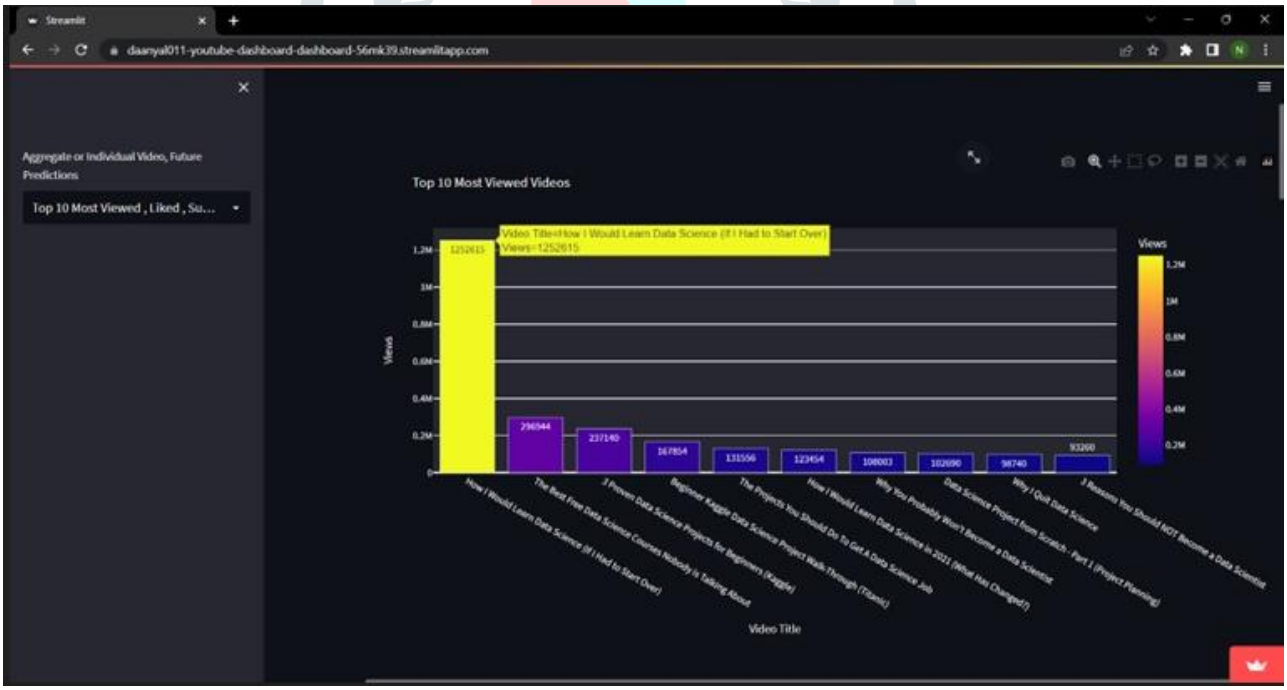
It shows how the video selected from the select box performs in its first 30 days in comparison to other videos in percentile format. We have selected 20 , 50 and 80 percentile

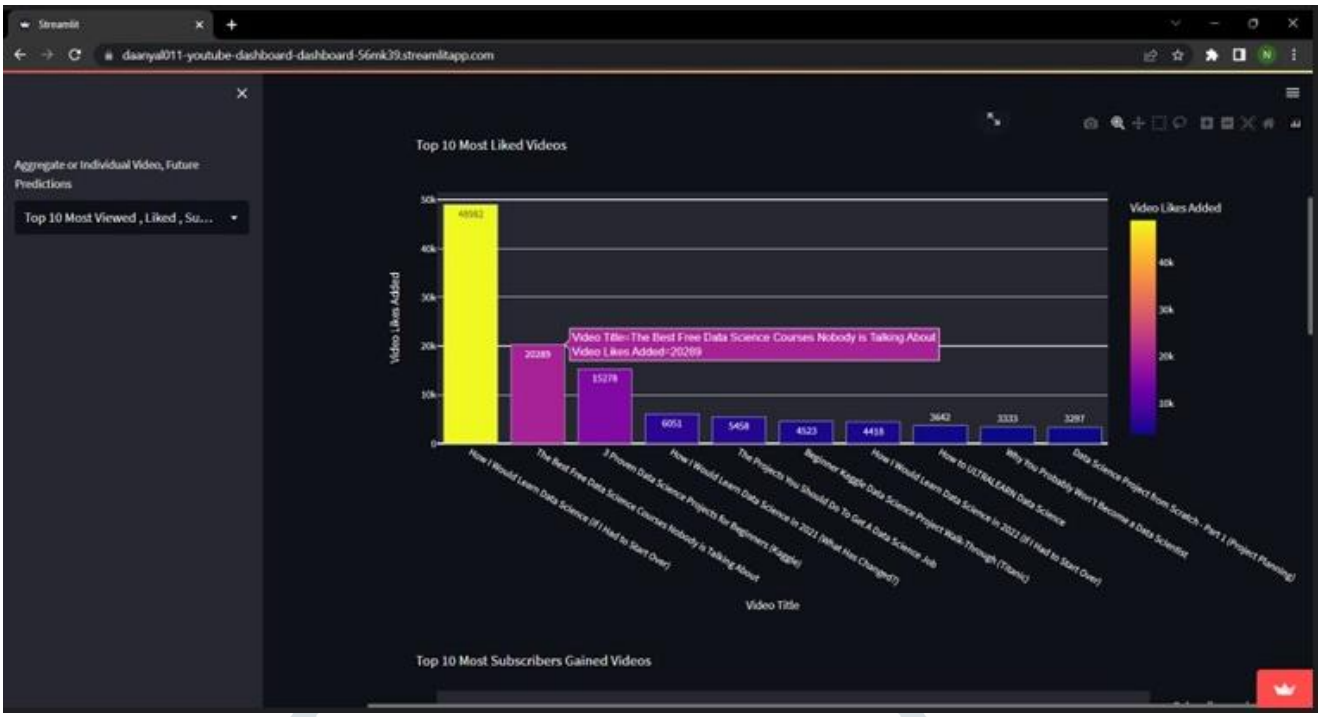
3. Growth of channel in Future

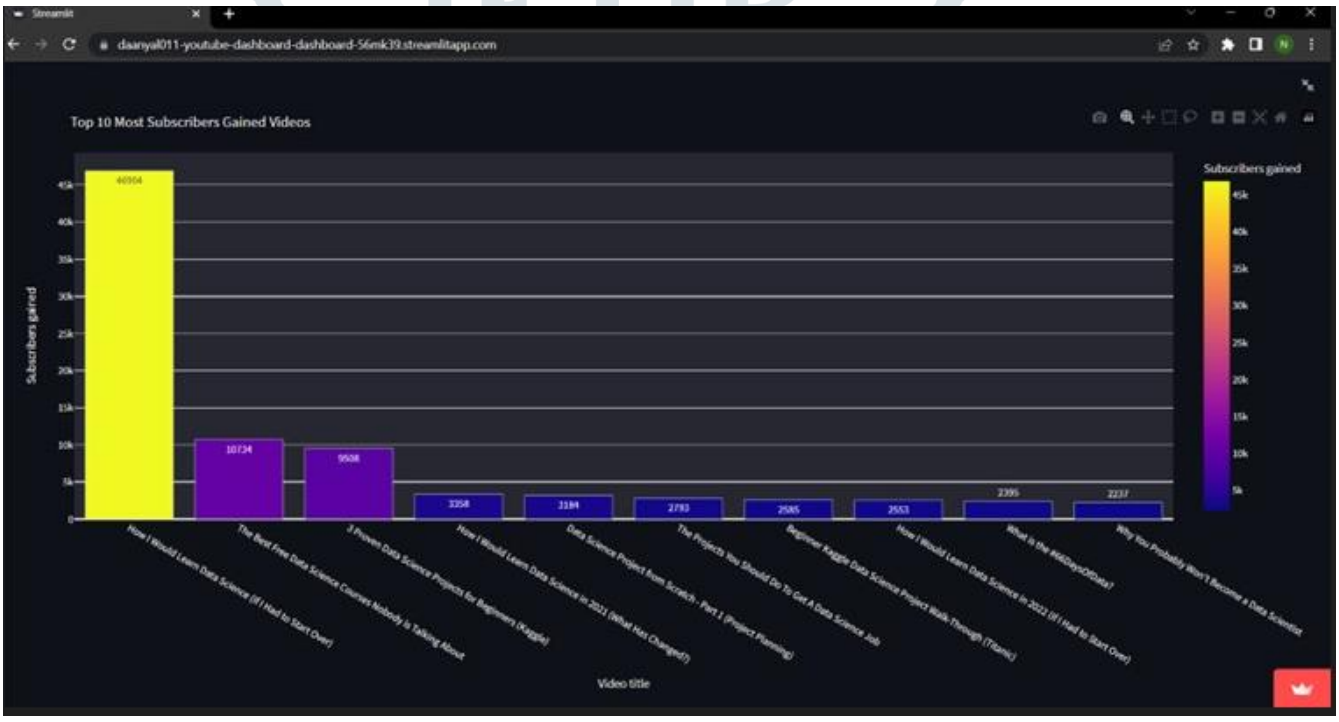


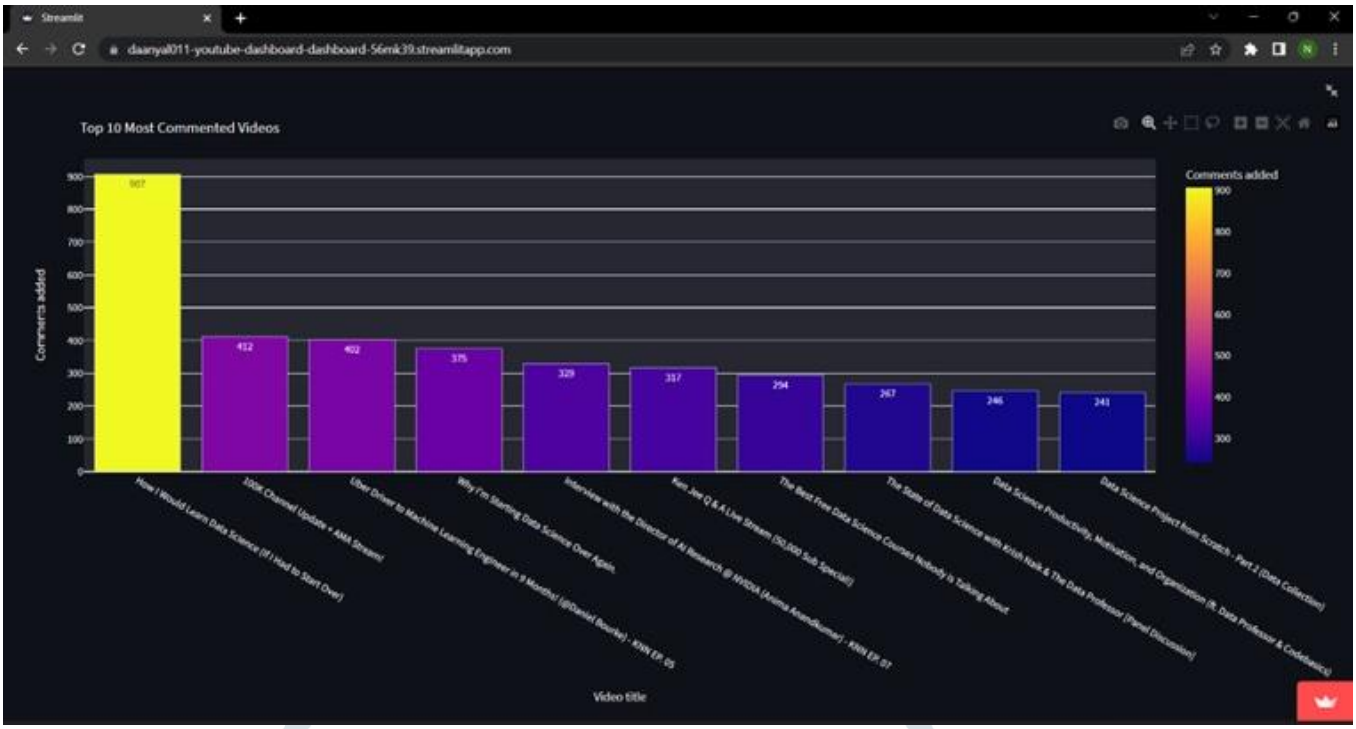
In this we have predicted growth of Shares with likes and growth of views with likes using regression

4. Top 10 most liked, viewed, shared, subscribed and commented videos



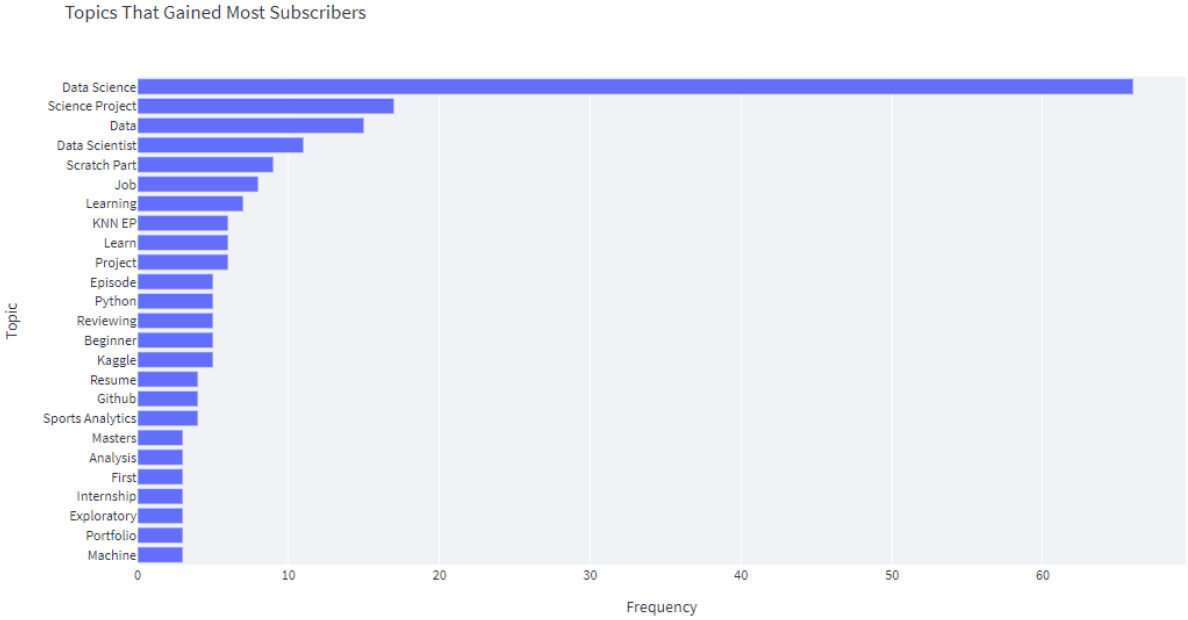

Streamlit
daanyal011-youtube-dashboard-dashboard-56mk39.streamlitapp.com
Aggregate or Individual Video, Future Predictions
Top 10 Most Viewed , Liked , Su...
Top 10 Most Liked Videos
Video Likes Added
Video Title=The Best Free Data Science Courses Nobody is Talking About
Video Likes Added=20289
Video Title
Top 10 Most Subscribers Gained Videos


Streamlit
daanyal011-youtube-dashboard-dashboard-56mk39.streamlitapp.com
Top 10 Most Subscribers Gained Videos
Subscribers gained
Video title



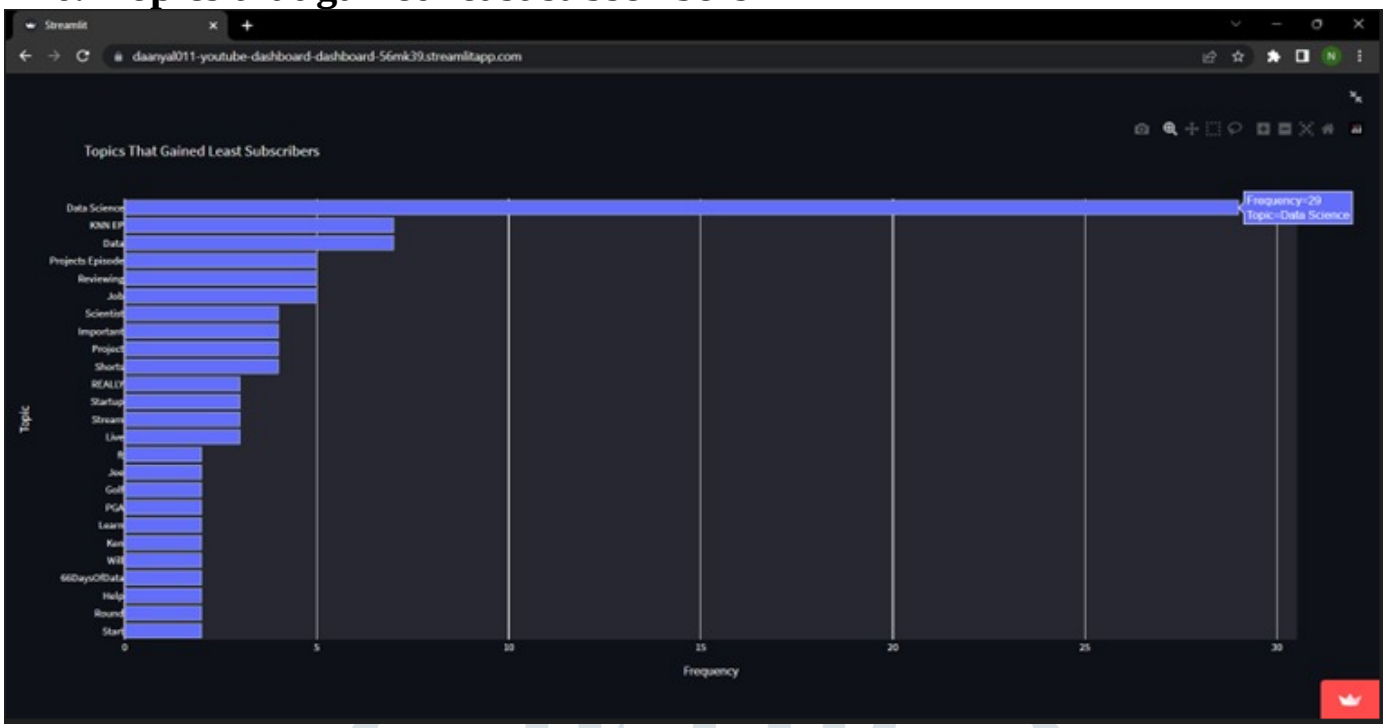
In this we have shown the top 10 most liked, viewed, shared, subscribed and commented videos in the form of bar graph.

5. Topics that gained most subscribers



Top 25 word which gave the most subscribers. We found this out using wordcloud library. We selected a criteria of 100 subscribers, if a topic gives 100 or more subscribers then it is considered for this graph.

6. Topics that gained least subscribers



Top 25 word which gave the least subscribers. We found this out using wordcloud library. We selected a criteria of 20 subscribers, if a topic gives 20 or less subscribers then it is considered for this graph.

Conclusion:

Our findings for measuring, assessing, and comparing essential characteristics of YouTube popular videos were reported in this study. Knowing the optimal time to publish a video to YouTube isn't enough to get millions of views and make your video popular. Other elements to consider are good titles, good thumbnails, video SEO, proper tagging, and the number of subscribers, all of which are important in driving views for your material. Understanding these statistics will aid YouTube in not only developing better video processing algorithms, but also in making judgements for individual youtubers.

References

i. reference/generated/plotly.graph_objects.Layout.html
ii. https://docs.python.org/3/
iii. numpy.org
iv. https://www.kaggle.com/
v. https://pypi.org/project/country-converter/

vi. https://www.geeksforgeeks.org/generating-word-cloud-python/

vii. "Plotly language support", https://en.wikipedia.org/wiki/Plotly

viii. Ouyang, S., Li, C. and Li, X. (2016). A Peek Into the Future: Predicting the Popularity of Online Videos. IEEE Access, 4, pp.3026–3033

ix. Paper refer "YouTube Trending Video Dashboarb"

x. Juhasz, A. (2008), 4 lessons of Youtube for activists. Conference paper delivered at OurMedia, August 2008, Accra, Ghana

xi. https://pandas.pydata.org/

xii. https://streamlit.io/

xiii. https://plotly.com/python-api-