**ANALYSIS ON YOUTUBE TRENDING VIDEOS**

**PROJECT BATCH – 3**

**TECHNICAL SKILLING - 2**

|  |  |
| --- | --- |
| **NAME** | **ROLL NO.** |
| T VENKATA SAI SATHVIK | 2010030361 |
| JAIDEEP SHARMA | 2010030374 |
| NIHAL AGARWAL | 2010030413 |
| SHAIK ABDUL SHAAN | 2010030153 |

ABSTRACT

In this modern era, many users are addicted to YouTube which is available both as application and through website. In detail, YouTube is a content sharing platform through the means of video in any format which allows users to watch or share any specific video with anyone. Usually there are three genres of content, that goes trending on YouTube, ranging from educational, entertainment and informative. Generally, when it comes to the entertainment genre there are users who would like to see gaming, music, short videos and many more. Students also frequently use YouTube for their study purposes to establish a good secure knowledge. YouTube is not limited to a particular age group, caste, creed or country, anyone can use it throughout the world. We provide analysis through data science to determine videos which are trending based on number of likes, views, and comments.

**LITERATURE REVIEW**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **AUTHOR** | **TITLE** | **PUBLISHED SOURCE** | **METHODS** | **FINDINGS** |
| [Johanes Fernandes Andry](https://ieeexplore.ieee.org/author/37088964712), [Stefan Azriel Reynaldo](https://ieeexplore.ieee.org/author/37089199629), [Kevin Christianto](https://ieeexplore.ieee.org/author/37089202437), [Francka Sakti Lee](https://ieeexplore.ieee.org/author/37089203555), [Julia Loisa](https://ieeexplore.ieee.org/author/37089202538),  [Aman Budi Manduro](https://ieeexplore.ieee.org/author/37089204146). | Algorithm of Trending Videos on YouTube Analysis using Classification, Association and Clustering. | [2021 International Conference on Data and Software Engineering (ICoDSE)](https://ieeexplore.ieee.org/xpl/conhome/9648407/proceeding), 3-4 Nov. 2021, DOI:  [10.1109/ICoDSE53690.2021.9648486](https://doi.org/10.1109/ICoDSE53690.2021.9648486). | Authors has been carried out with existing data mining software to meet the results, find out that rating given by YouTube for getting to the trending list are “views”, “likes”, dislikes”, “comments”. | Finding the highest positive rated video by analysing the different factors of the video such as views, likes, comments and dislikes by the data mining, k means algorithm, classification and clustering. |
| Swati Gayakwad, Rajas Patankar, Dashrath Mane. | Analysis on YouTube Trending Videos | [International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 08 | Aug 2020 .](https://www.irjet.net/archives/V7/i8/IRJET-V7I8732.pdf) e-ISSN: 2395-0056, p-ISSN: 2395-0072. | Analysis for this paper is done by using 3 different ways of analysis. They analyzed the basic statistics of YouTube trending videos by downloading data through YouTube API, then differentiated trending and non trending. | They used three algorithms to analyse the trending videos and non trending videos such as classification, regression, and clustering and similarity matching. Even different factors were analysed. They analysed the best time to upload videos. |
| Sana Khanam, Safdar Tanweer, Syed Sibtain Khalid. | Youtube Trending Videos: Boosting Machine Learning Results Using Exploratory Data Analysis. | The Computer Journal, bxab142, <https://doi.org/10.1093/comjnl/bxab142>, Published:   20 October 2021. | We present our analysis by measuring, mining, analyzing and comparing key aspects of time-series YouTube data with respect to its view and audience response statistics from 40 000 trending YouTube videos collected over 205 days. | We have performed an exploratory data analysis (EDA) on all its aspect to get data insights and used statistics to find similarities between them to understand viewing pattern of different video categories.  We also compare and observe the variation of activity over time with the nature of the event that affects the quality of our analysis. |
| [Iman Barjasteh](https://www.researchgate.net/profile/Iman-Barjasteh), [Ying Liu](https://www.researchgate.net/profile/Ying-Liu-295), [Hayder Radha](https://www.researchgate.net/profile/Hayder-Radha). | Trending Videos: Measurement and Analysis. | Cornell University, <https://arxiv.org/abs/1409.7733>. September 2014. | The study is based on collecting and monitoring high-resolution time-series of the viewership and related statistics of more than 8,000 YouTube videos over an aggregate period of nine months. | They employed Granger Causality (GC) with significance testing to conduct this analysis. Unlike traditional correlation measures, our directional-relationship analysis provides a deeper insight onto the viewership pattern of different categories of trending videos. |
| Aakash Ashok Niture Supervisor: Mr. Pierpaolo Dondio | Predictive analysis of YouTube trending videos using Machine Learning | <https://esource.dbs.ie/bitstream/handle/10788/4260/msc_niture_aa_2021.pdf?sequence=1&isAllowed=y>. : 11/01/2021 | Since trending video statistics consists of number of Views, Likes, Dislikes and Comment counts, the research performed Linear regression model of Machine Learning for predictive analysis of number of views for YouTube trending videos  Achieving maximum accuracy of 62.53%. | In addition, the study performs a comparative analysis of a number of classification models namely Random Forest, SVM, Decision Tree, Logistic Regression and Gaussian Naïve Bayes, to determine which model suits better for predicting the number of days a video will take to get trending from its upload time and the number of days a video will trend on the trending list. |
| Muhammad Nihal Hussain, Serpil Tokdemir, Samer Al-khateeb, Kiran Kumar Bandeli, and Nitin Agarwal | Understanding Digital Ethnography: Socio-computational Analysis of Trending YouTube Videos | <http://sbp-brims.org/2018/proceedings/papers/latebreaking_papers/LB_14.pdf>. | The original dataset from Kaggle has the following attributes: URL of the video, video ID, title of the video, title of the channel that published the video, category in which the video belongs to, number of views, number of likes, number of dislikes, number of comments the video received at the time data was collected, and the date the video was trending | In the second phase, we enhanced the dataset obtained from Kaggle by adding the description of the video, date the channel was created, and the number of subscribers of the channel, using YouTube API. It is a common practice among prominent YouTubers to associate their various social media accounts with their YouTube channel. |

**GITHUB SETUP**

A screenshot of a computer screen

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

Graphical user interface, text

Description automatically generatedA screenshot of a computer

Description automatically generated with low confidenceGraphical user interface, application

Description automatically generatedA screenshot of a computer

Description automatically generated with medium confidence

**DATASET COLLECTION**

* We had collected the dataset needed for our project using Kaggle notebooks and downloaded the dataset which we want to work on.
* We had 10 datasets for different countries, but we had chose to work on the dataset for India and US.
* We will perform trending video analysis on the benchmark dataset and filter out the results.
* The data is already scraped and cleaned using pandas and Jupyter notebooks using NLTK library.
* Here is the link for the pre-processed dataset. https://github.com/mitchelljy/Trending-YouTube-Scraper

**DATASET TOOLS USED**

* To scrape the dataset they have used different tools and libraries that gives us a clean dataset with a maximum precision.
* They used YouTube API KEY to the key identifies your project and provides API access, quota, and reports.
* The different modules for the data extraction and cleaning, except the requests module.
* They had assigned 10 country codes for 10 datasets to filter out the data given in it.
* Running the script using all the technique helps us in cleaning the data.

**DATASET COLLECTION**

Graphical user interface, application, table, Excel

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

**GITHUB COMMITS**

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~

$ pwd

/c/Users/SAI SATHVIK

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~

$ git config --global user.name

sathviksai002

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~

$ git config --global user.email

sathviksai002@gmail.com

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~

$ git clone https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis.git

Cloning into 'Youtube\_Video\_Trending\_Analysis'...

remote: Enumerating objects: 4, done.

remote: Counting objects: 100% (4/4), done.

remote: Compressing objects: 100% (3/3), done.

remote: Total 4 (delta 0), reused 0 (delta 0), pack-reused 0

Receiving objects: 100% (4/4), done.

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~

$ cd Youtube\_Video\_Trending\_Analysis

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (main)

$ pwd

/c/Users/SAI SATHVIK/Youtube\_Video\_Trending\_Analysis

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (main)

$ git branch teamlead\_saisathvik

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (main)

$ git branch

\* main

teamlead\_saisathvik

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (main)

$ git checkout teamlead\_saisathvik

Switched to branch 'teamlead\_saisathvik'

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (teamlead\_saisathvik)

$ git clone https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis.git

Cloning into 'Youtube\_Video\_Trending\_Analysis'...

remote: Enumerating objects: 4, done.

remote: Counting objects: 100% (4/4), done.

remote: Compressing objects: 100% (3/3), done.

remote: Total 4 (delta 0), reused 0 (delta 0), pack-reused 0

Receiving objects: 100% (4/4), done.

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (teamlead\_saisathvik)

$ git add .

warning: adding embedded git repository: Youtube\_Video\_Trending\_Analysis

hint: You've added another git repository inside your current repository.

hint: Clones of the outer repository will not contain the contents of

hint: the embedded repository and will not know how to obtain it.

hint: If you meant to add a submodule, use:

hint:

hint: git submodule add <url> Youtube\_Video\_Trending\_Analysis

hint:

hint: If you added this path by mistake, you can remove it from the

hint: index with:

hint:

hint: git rm --cached Youtube\_Video\_Trending\_Analysis

hint:

hint: See "git help submodule" for more information.

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (teamlead\_saisathvik)

$ git commit -m "commited by teamlead\_saisathvik"

[teamlead\_saisathvik d2238d8] commited by teamlead\_saisathvik

1 file changed, 1 insertion(+)

create mode 160000 Youtube\_Video\_Trending\_Analysis

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (teamlead\_saisathvik)

$ git push origin teamlead\_saisathvik

Enumerating objects: 3, done.

Counting objects: 100% (3/3), done.

Delta compression using up to 8 threads

Compressing objects: 100% (2/2), done.

Writing objects: 100% (2/2), 334 bytes | 334.00 KiB/s, done.

Total 2 (delta 0), reused 0 (delta 0), pack-reused 0

remote:

remote: Create a pull request for 'teamlead\_saisathvik' on GitHub by visiting:

remote: https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis/pull/new/teamlead\_saisathvik

remote:

To https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis.git

\* [new branch] teamlead\_saisathvik -> teamlead\_saisathvik

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (teamlead\_saisathvik)

$ git branch member\_jaideepsharma

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (teamlead\_saisathvik)

$ git branch member\_nihalagarwal

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (teamlead\_saisathvik)

$ git branch member\_shaikabdulshaan

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (teamlead\_saisathvik)

$ git checkout member\_jaideepsharma

Switched to branch 'member\_jaideepsharma'

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_jaideepsharma)

$ git push origin member\_jaideepsharma

Total 0 (delta 0), reused 0 (delta 0), pack-reused 0

remote:

remote: Create a pull request for 'member\_jaideepsharma' on GitHub by visiting:

remote: https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis/pull/new/member\_jaideepsharma

remote:

To https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis.git

\* [new branch] member\_jaideepsharma -> member\_jaideepsharma

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_jaideepsharma)

$ git clone https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis.git

fatal: destination path 'Youtube\_Video\_Trending\_Analysis' already exists and is not an empty directory.

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_jaideepsharma)

$ git add .

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_jaideepsharma)

$ git push origin member\_jaideepsharma

Everything up-to-date

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_jaideepsharma)

$ git checkout member\_nihalagarwal

Switched to branch 'member\_nihalagarwal'

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_nihalagarwal)

$ git clone https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis.git

fatal: destination path 'Youtube\_Video\_Trending\_Analysis' already exists and is not an empty directory.

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_nihalagarwal)

$ git add .

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_nihalagarwal)

$ git push origin member\_nihalagarwal

Total 0 (delta 0), reused 0 (delta 0), pack-reused 0

remote:

remote: Create a pull request for 'member\_nihalagarwal' on GitHub by visiting:

remote: https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis/pull/new/member\_nihalagarwal

remote:

To https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis.git

\* [new branch] member\_nihalagarwal -> member\_nihalagarwal

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_nihalagarwal)

$ git checkout member\_shaikabdulshaan

Switched to branch 'member\_shaikabdulshaan'

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_shaikabdulshaan)

$ git clone https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis.git

fatal: destination path 'Youtube\_Video\_Trending\_Analysis' already exists and is not an empty directory.

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_shaikabdulshaan)

$ git add .

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_shaikabdulshaan)

$ git push origin member\_shaikabdulshaan

Total 0 (delta 0), reused 0 (delta 0), pack-reused 0

remote:

remote: Create a pull request for 'member\_shaikabdulshaan' on GitHub by visiting:

remote: https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis/pull/new/member\_shaikabdulshaan

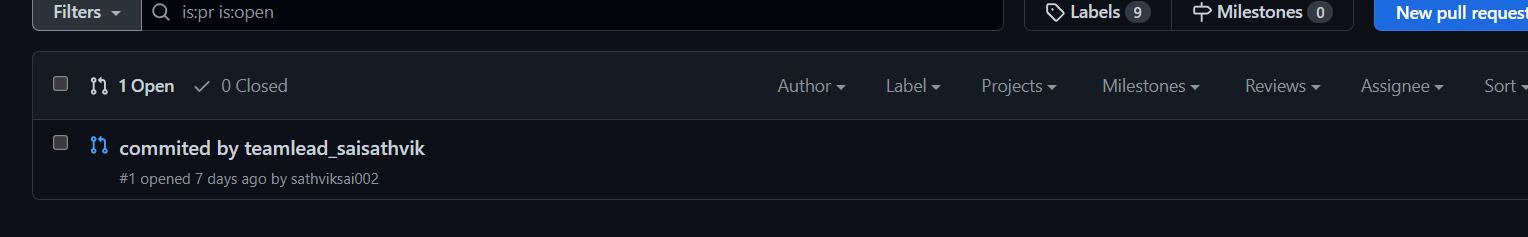
remote:

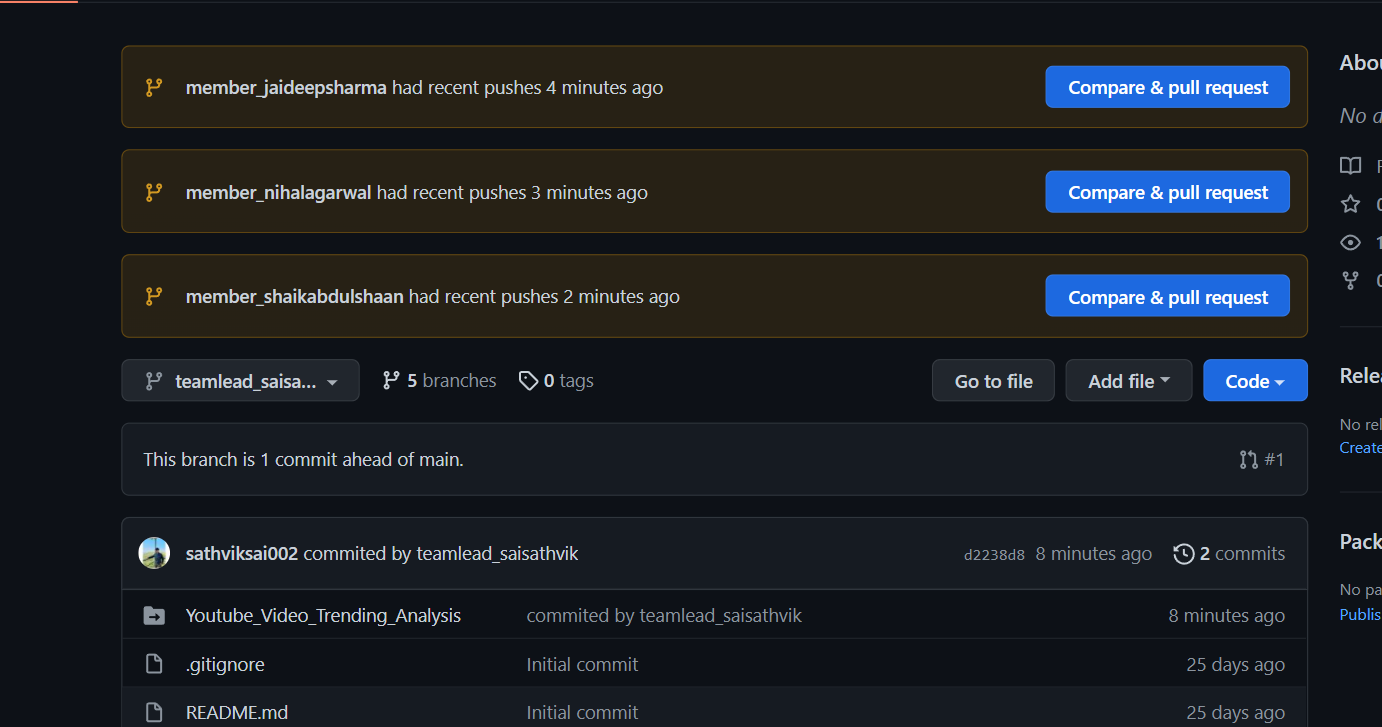
To https://github.com/sathviksai002/Youtube\_Video\_Trending\_Analysis.git

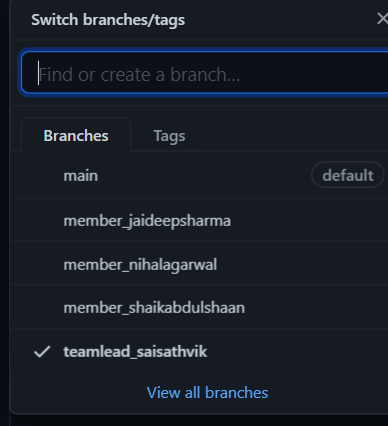
\* [new branch] member\_shaikabdulshaan -> member\_shaikabdulshaan

SAI SATHVIK@LAPTOP-MUIAFN27 MINGW64 ~/Youtube\_Video\_Trending\_Analysis (member\_shaikabdulshaan)

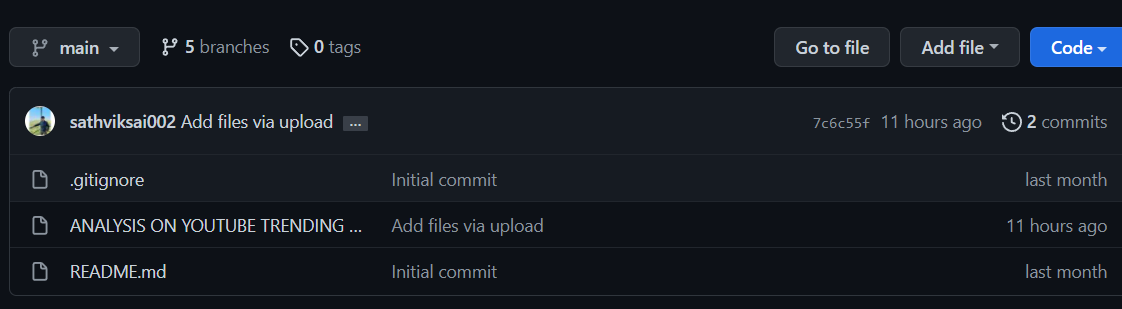
$

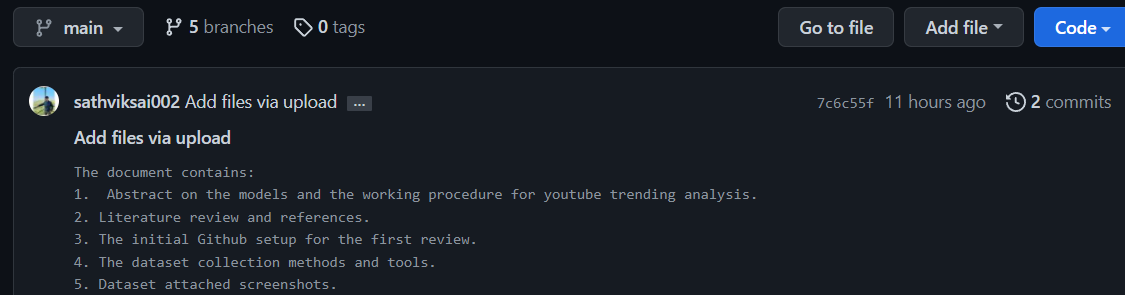










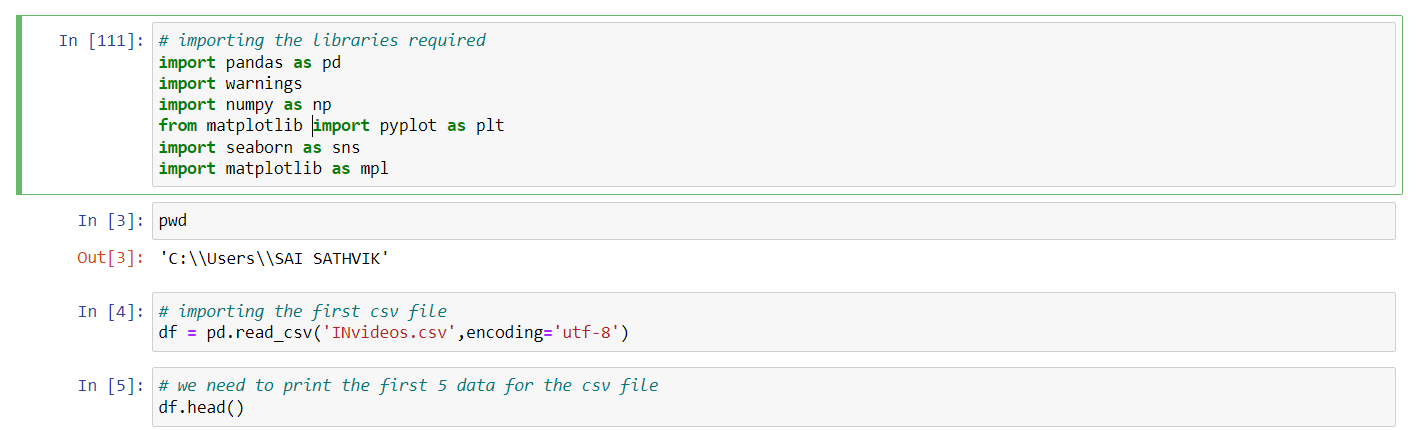


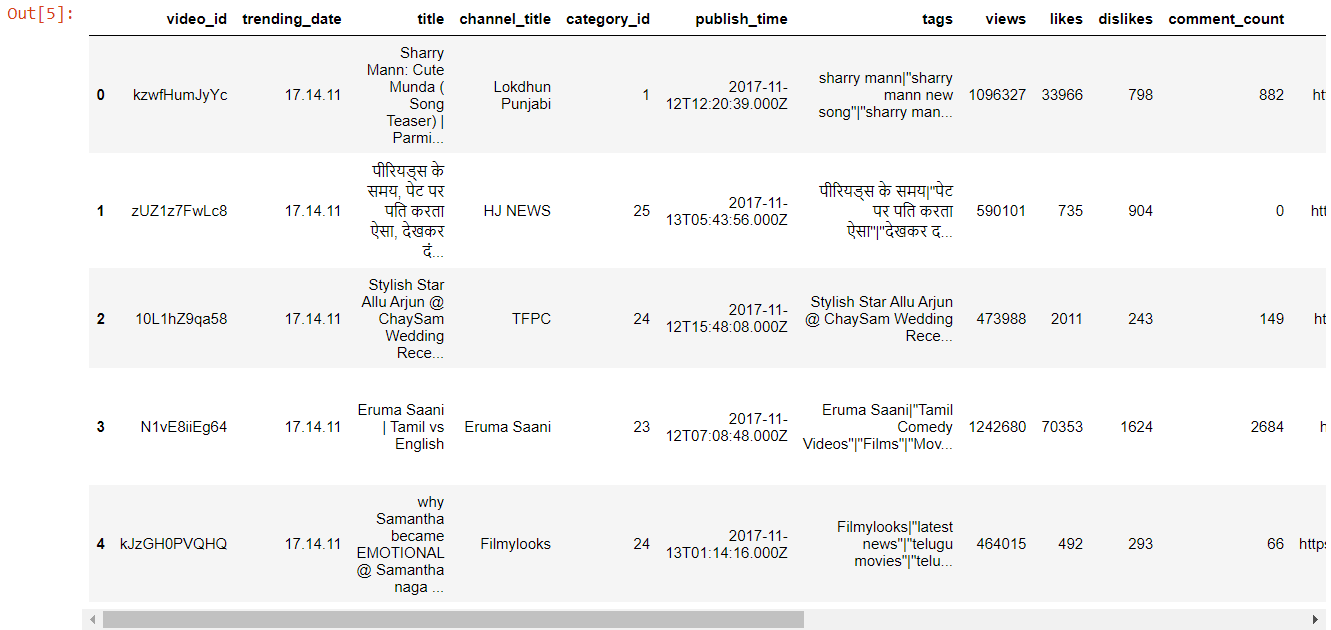
IMPLEMENTATION FLOW CHART

Diagram

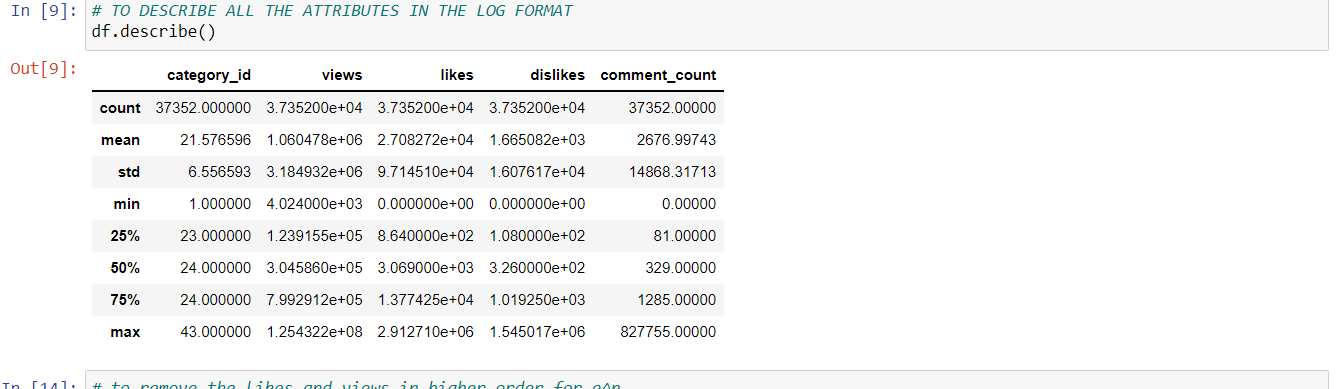
Description automatically generated

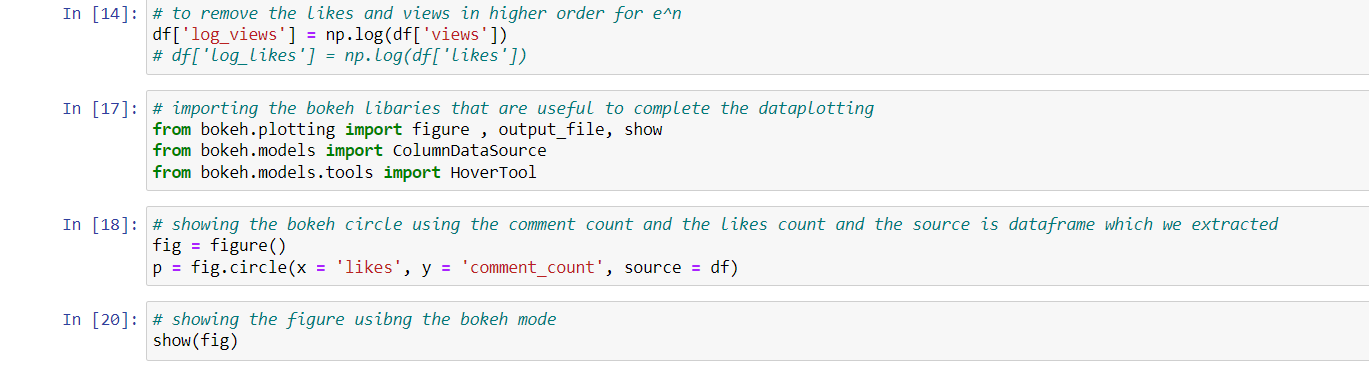
**WORK PROGRESS**

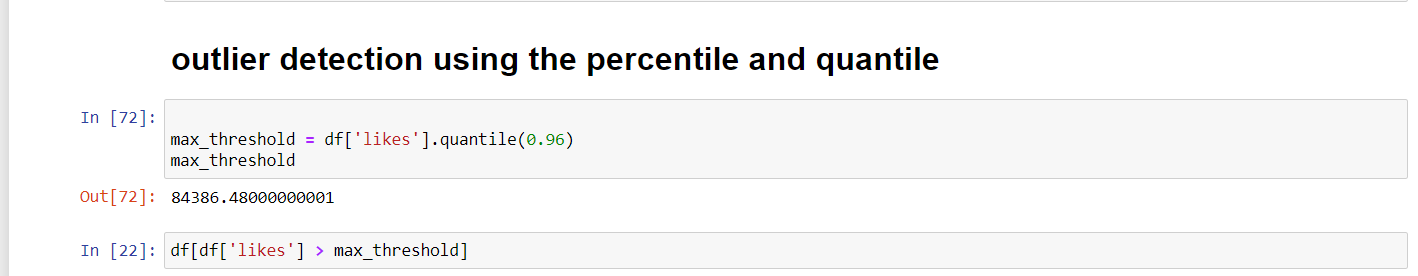


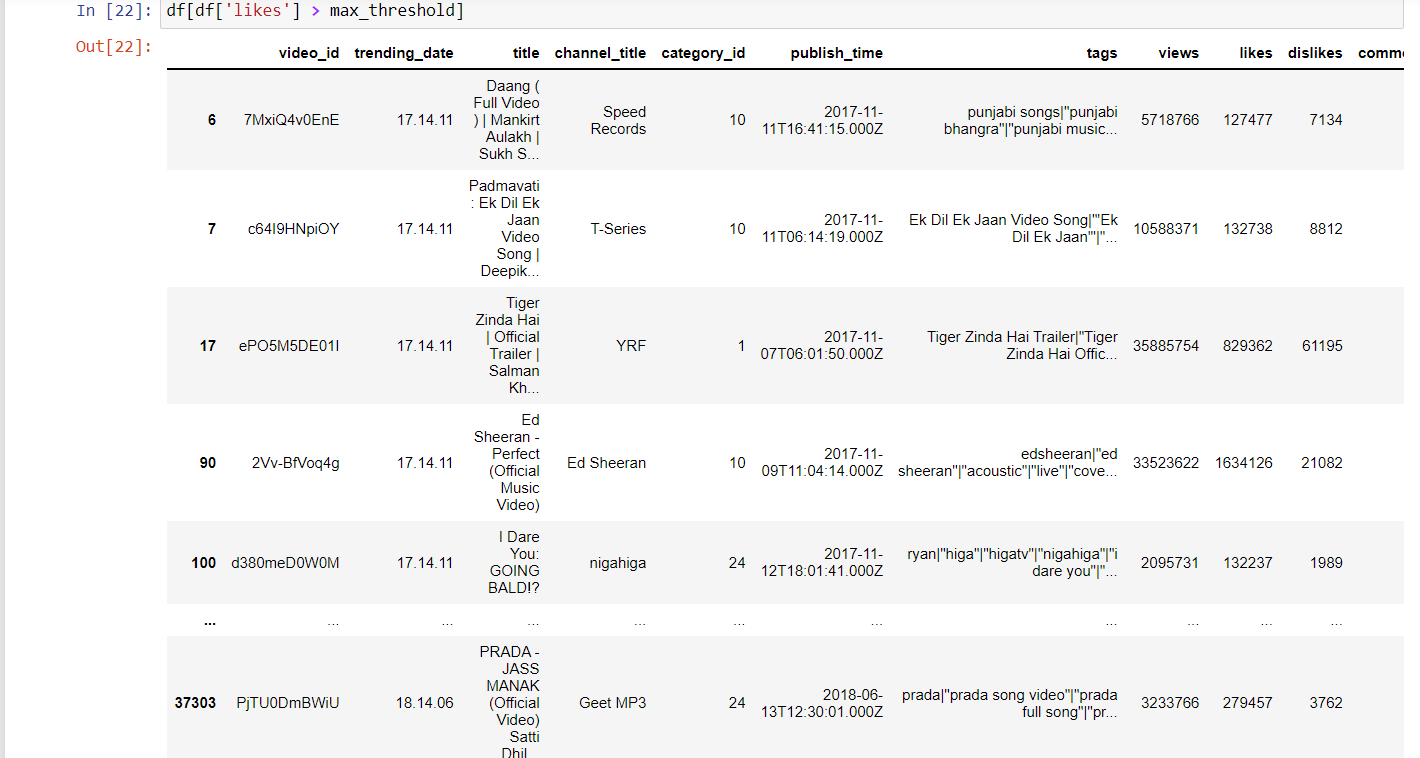




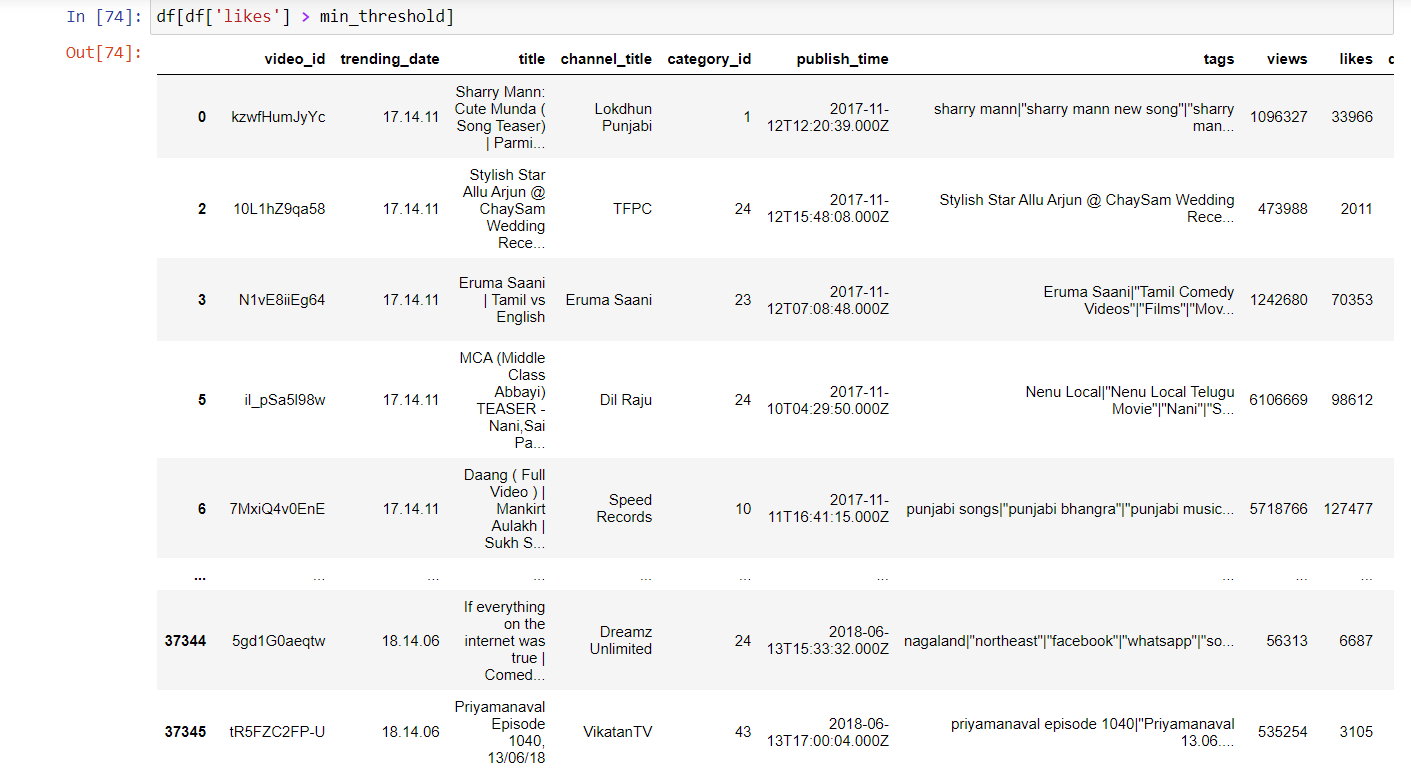


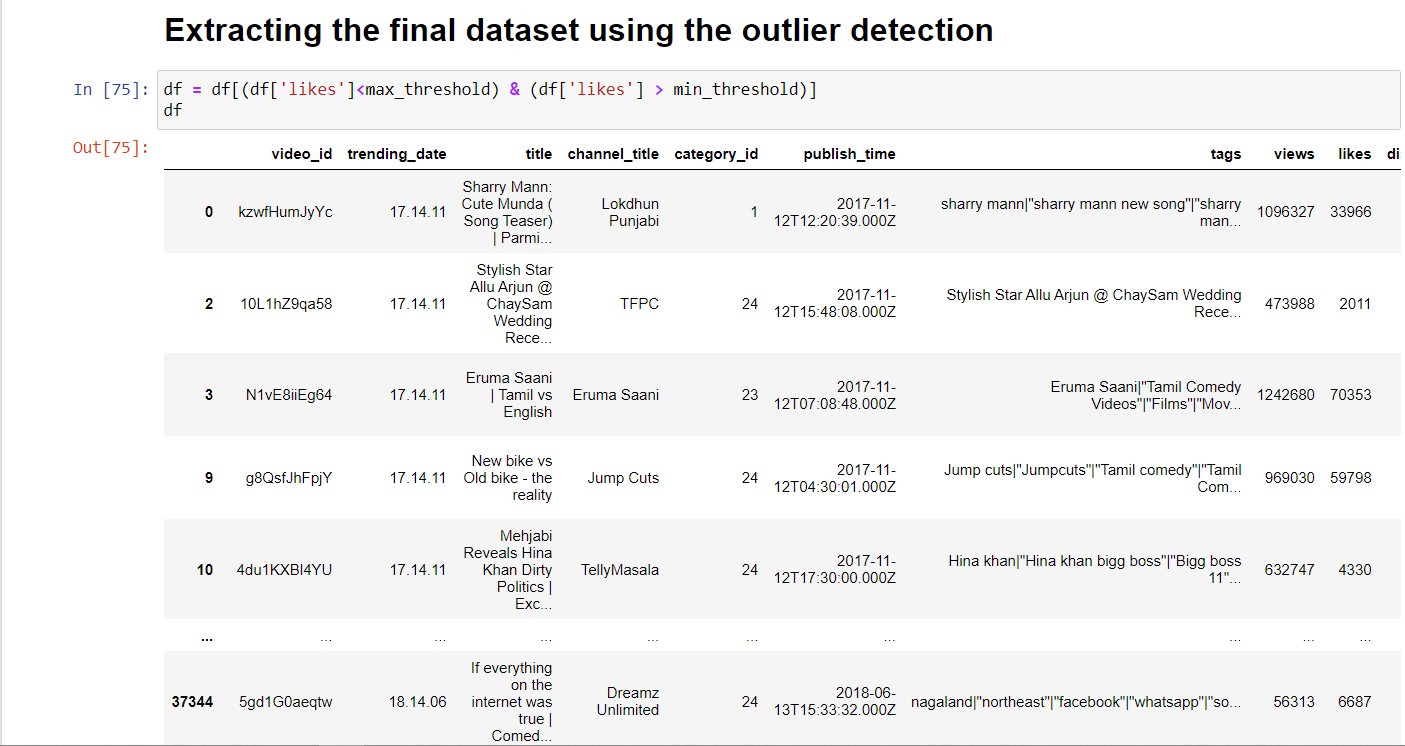








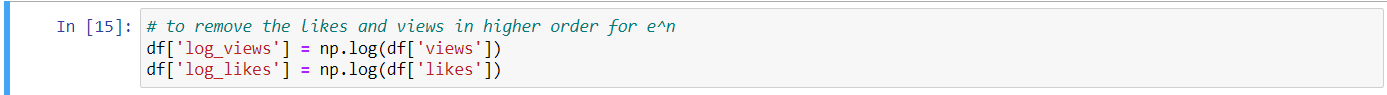




**After detecting the outliers and extracting the new dataset from it and we loaded it into the Jupyter notebooks.**

****

**changing the number of views and likes in the format of the log to get the right graph.**

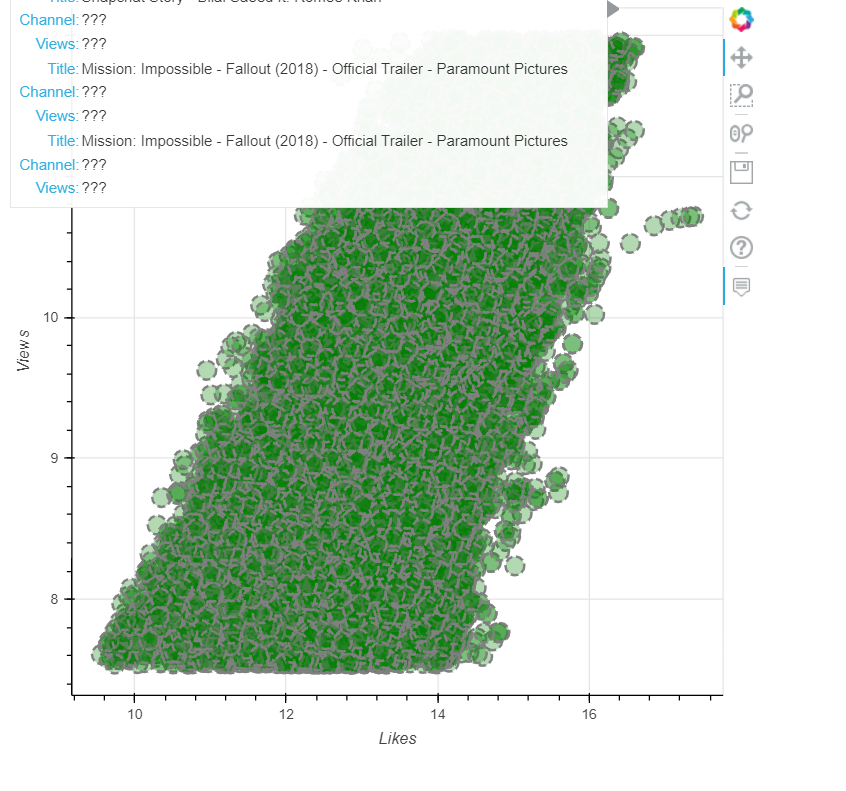
****

**Having the tooltips and adding the external factors and specifications for the glyphs and graph plots.**

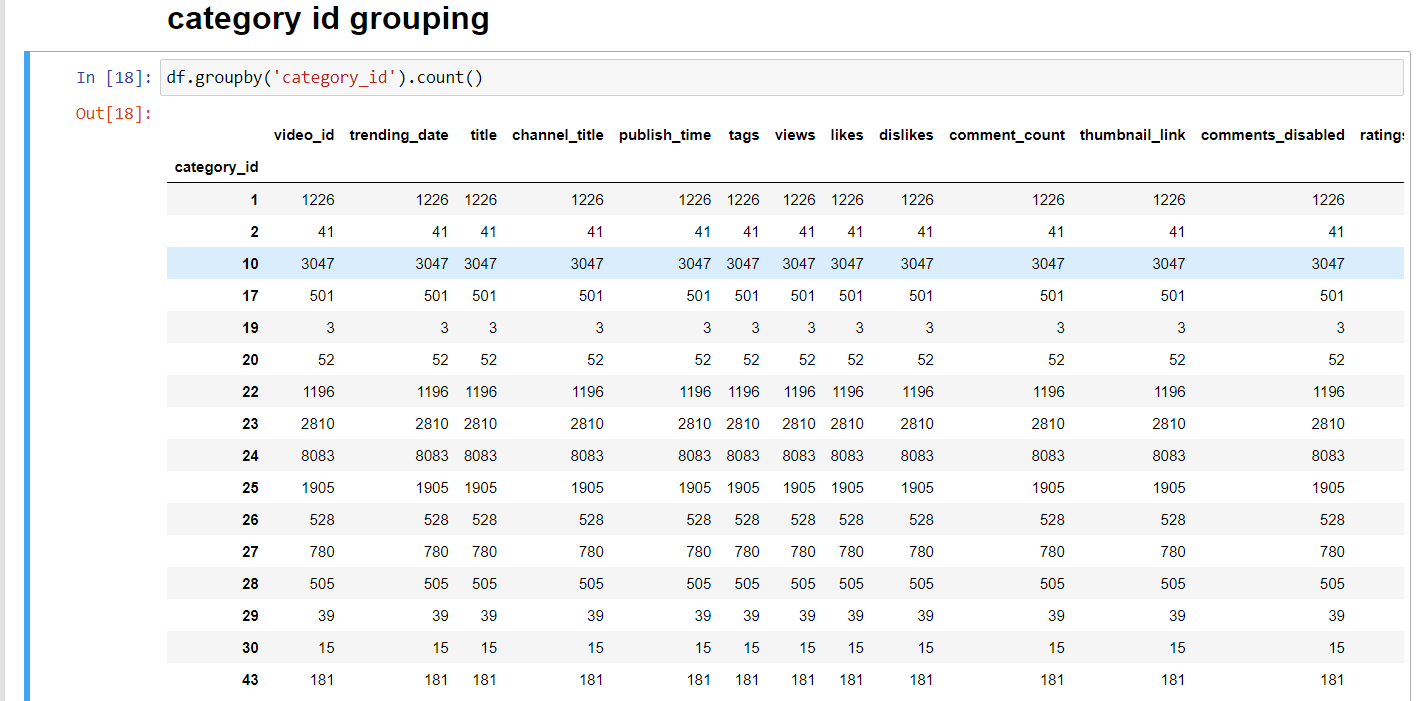
****

**Showing the figure after plotting it through the log likes and log views.**

****

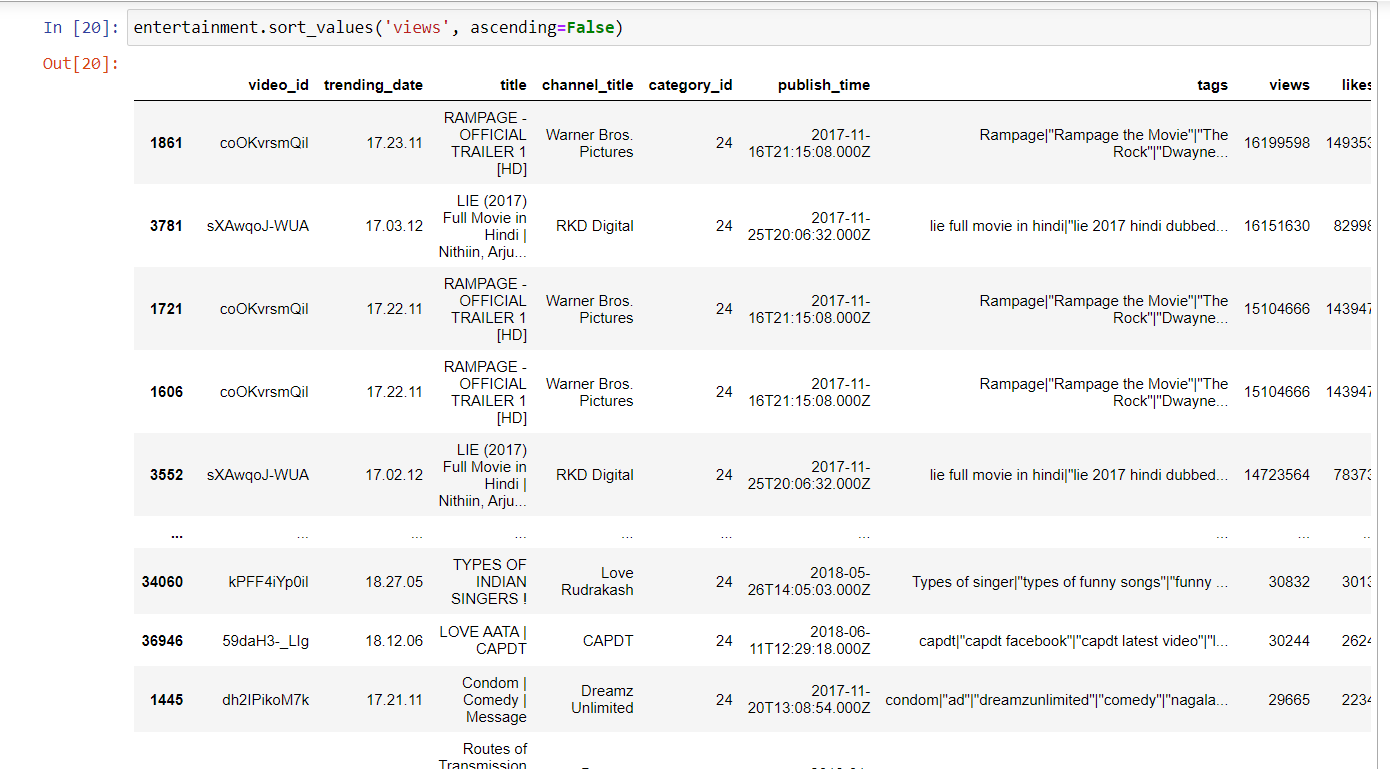
****

**grouping the videos using the category wise of the YouTube videos.**

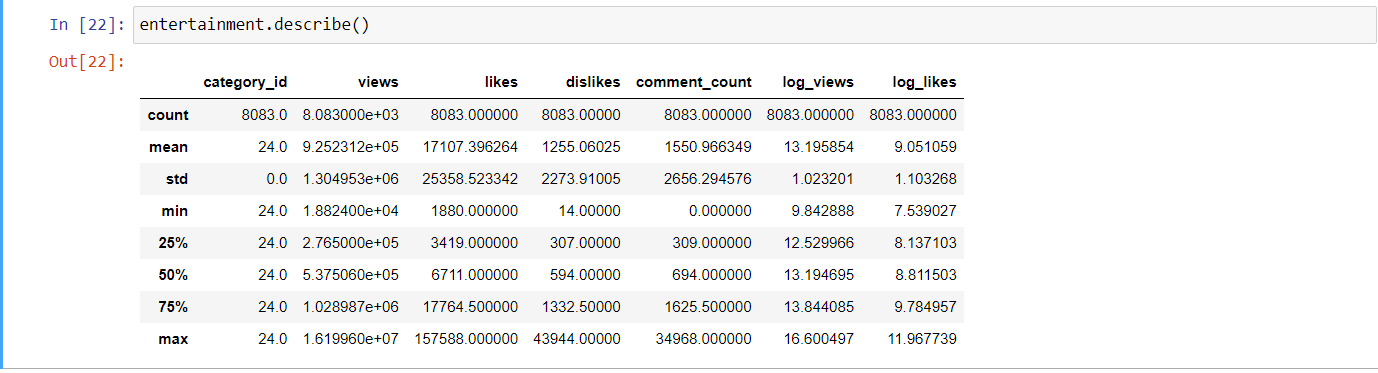
****

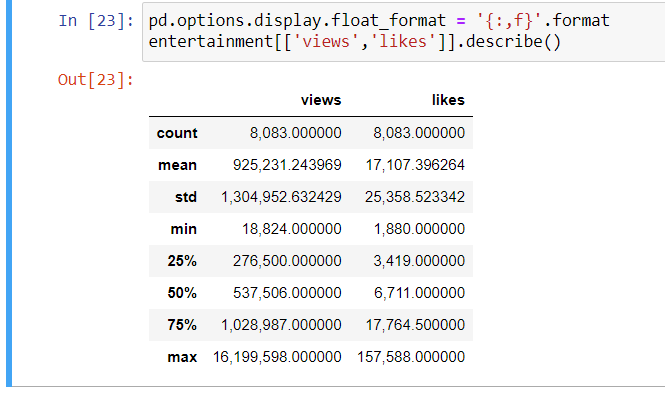
**Here we have the entertainment category as the highest view count and next the music category so we will create other dataframe for it and sort it by view count**

****

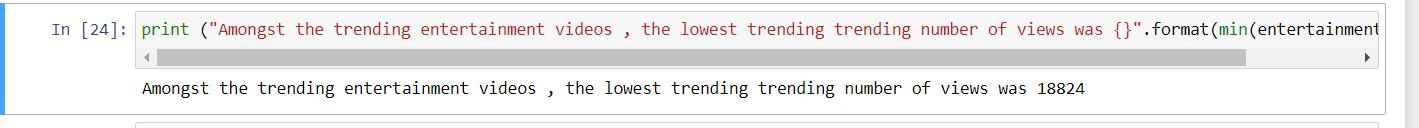
****

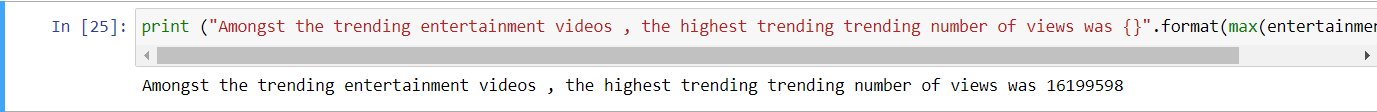
****

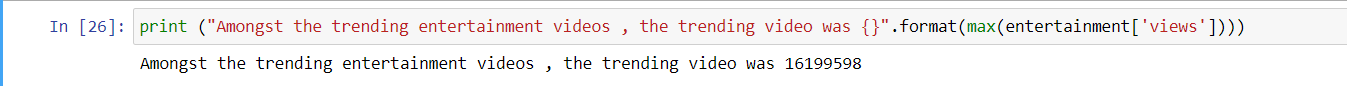
****

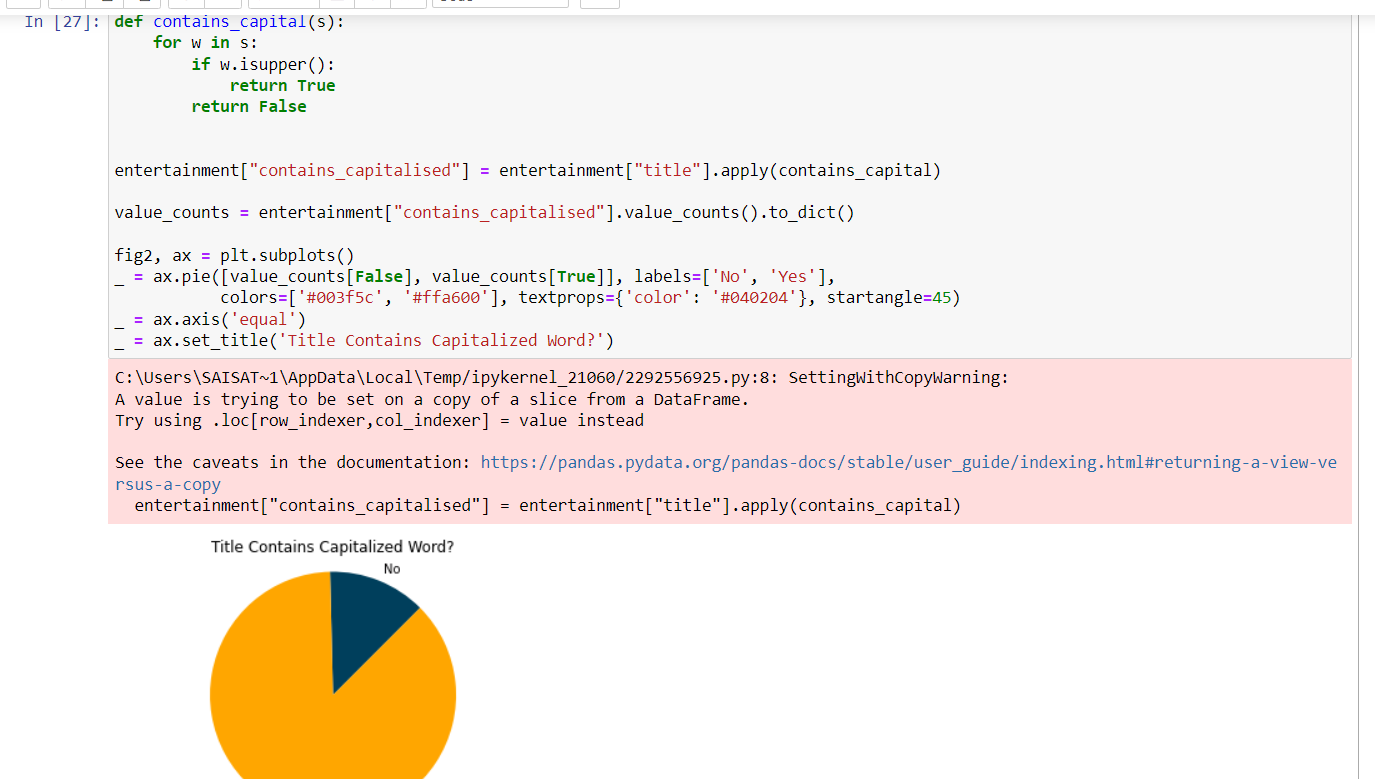
****

# for entertainment category to be in trending you need minimum of 18000 views and 2000 likes, and the highest is 1.6 crore views and 85k likes

****

****

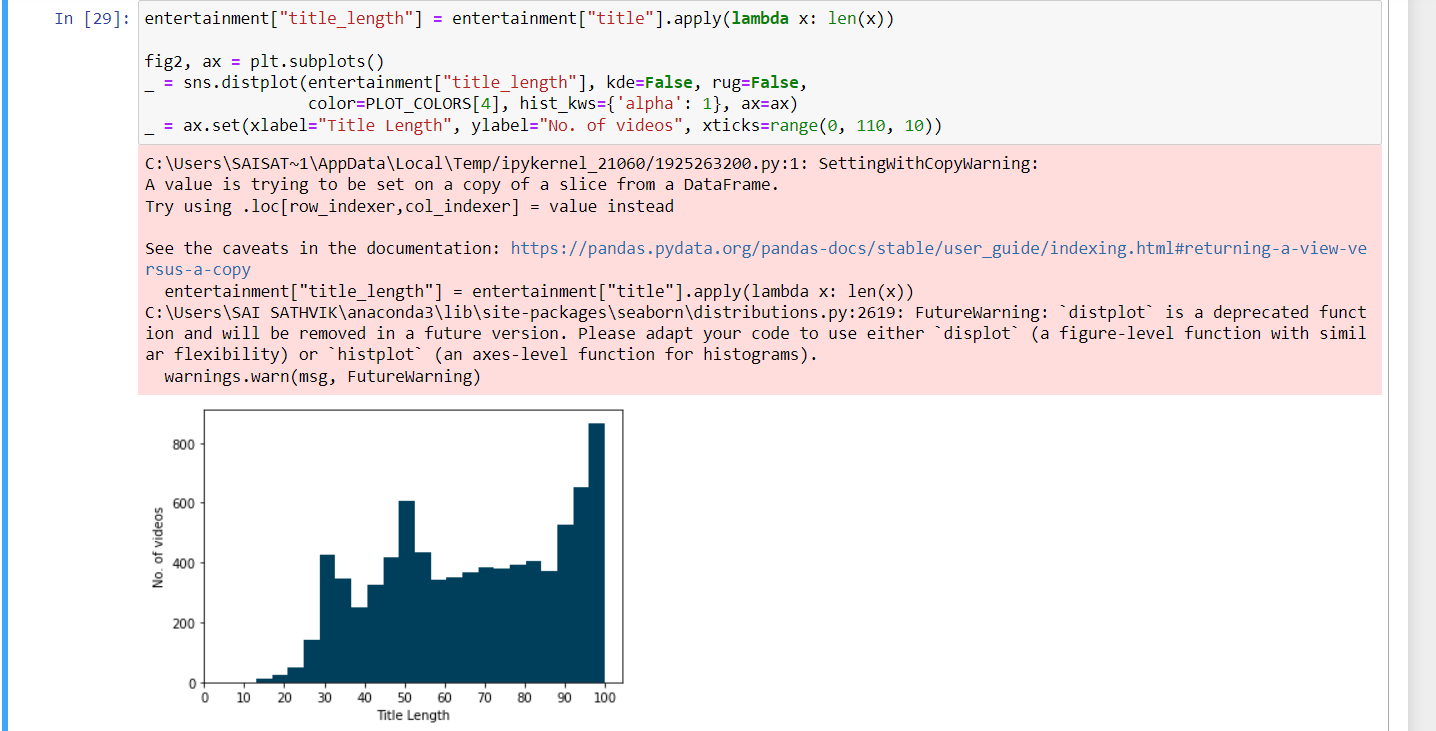
****

****

# from above we can see that almost 80% of the videos has title as capital letter when it is in trending for entertainment category.

# Let’s add another column to our dataset to analyze the length of titles of videos, then plot the title length histogram to get an idea of the length of trending video titles:

****

****

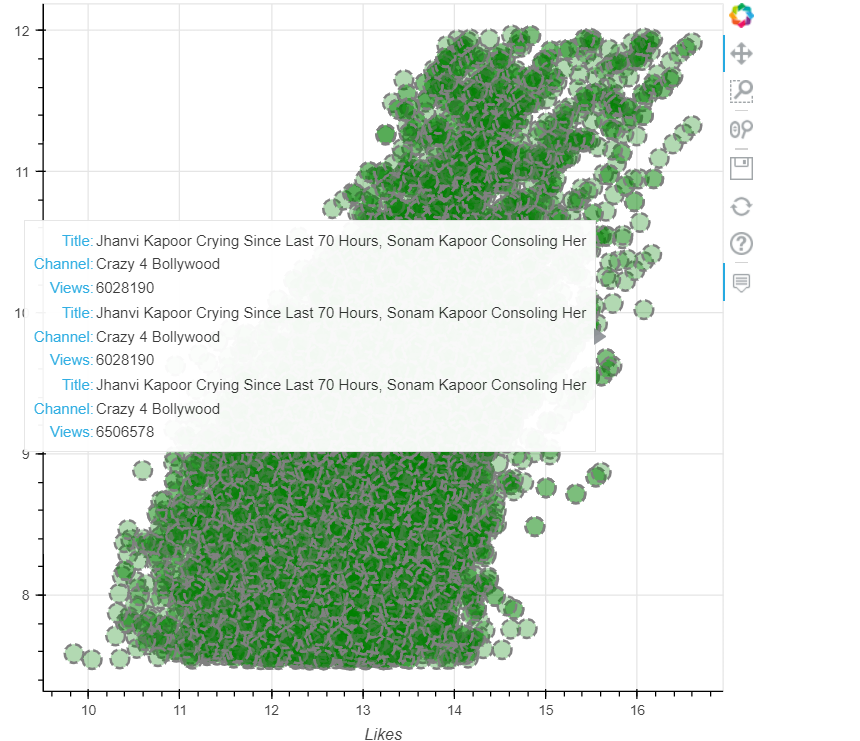
# From the above graph we can see that the more videos have more than or equal to atleast 100 characters to be in the trending list

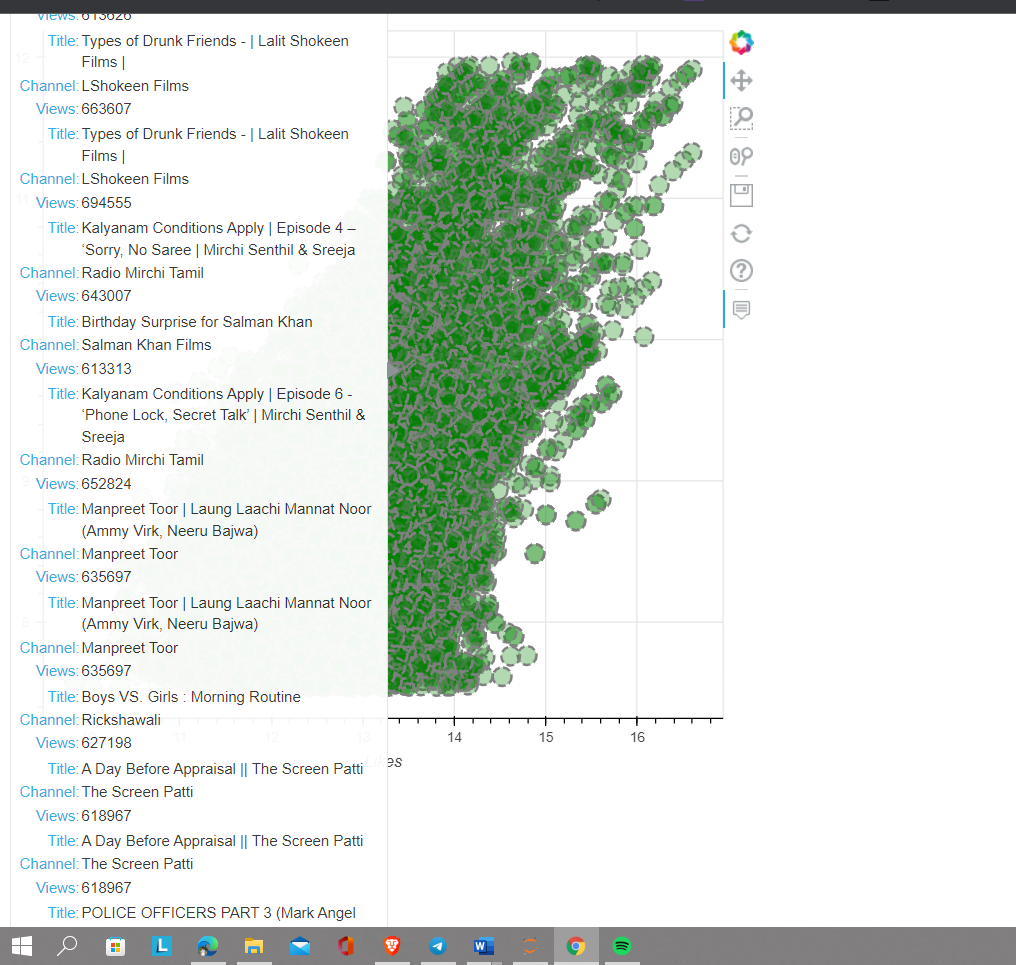
****

# from the above scatter plot we can see that there is no relationship between the length and the views, but if we look keenly there are videos which are more than 80 characters that has 160000000 views, and there are videos which have 100 characeters but still <20000 views

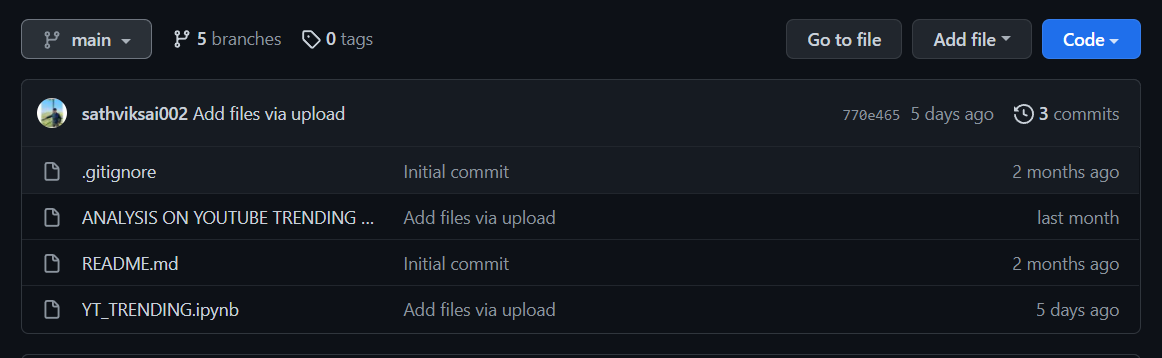
****

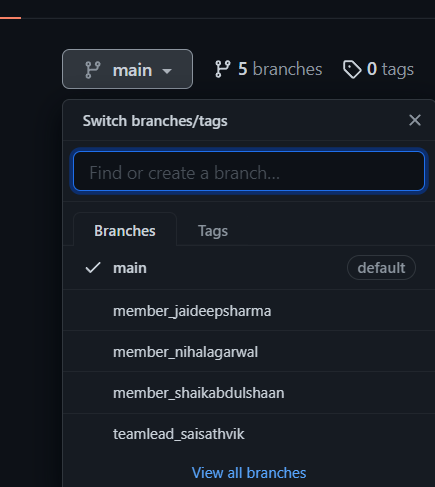
****

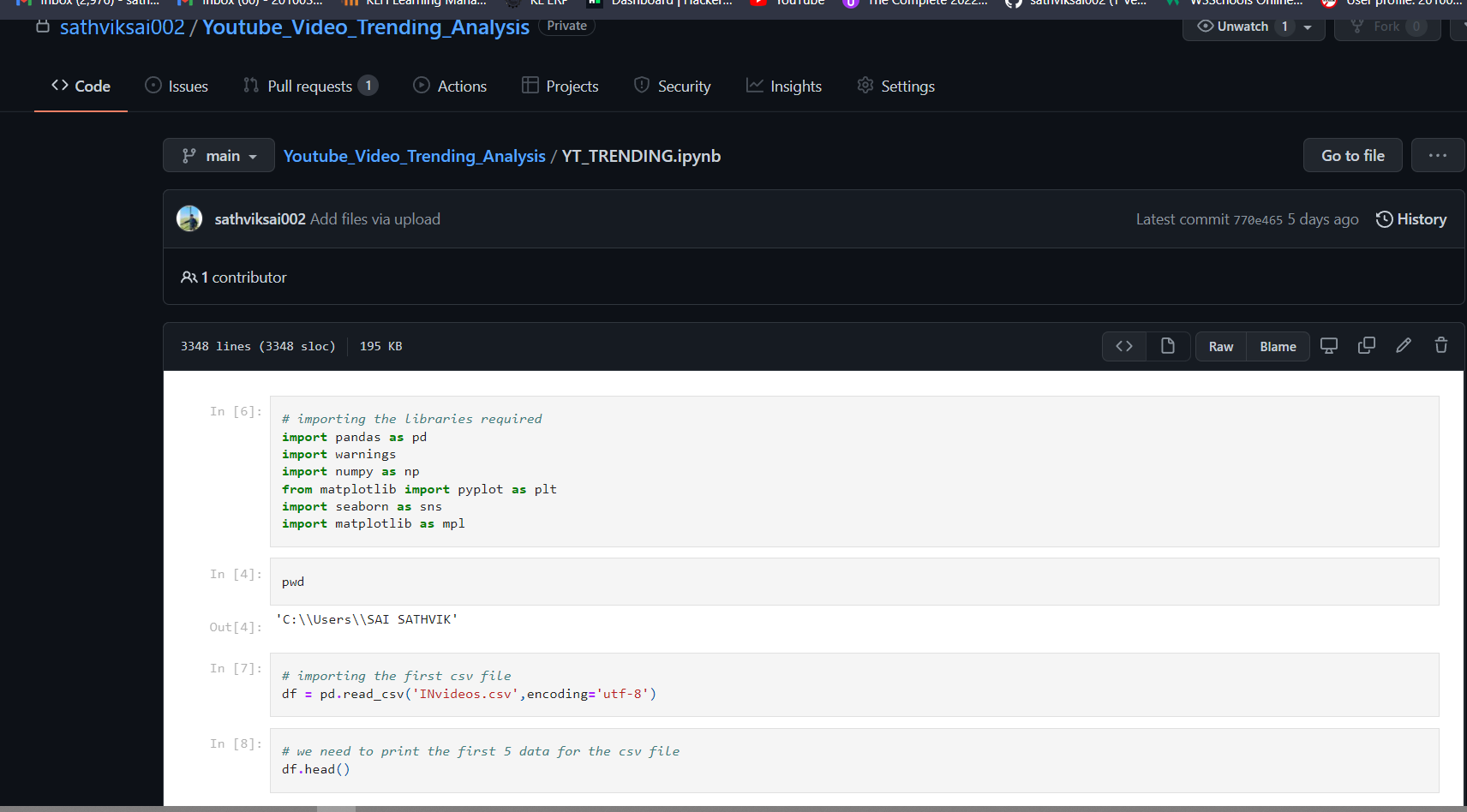
****

****

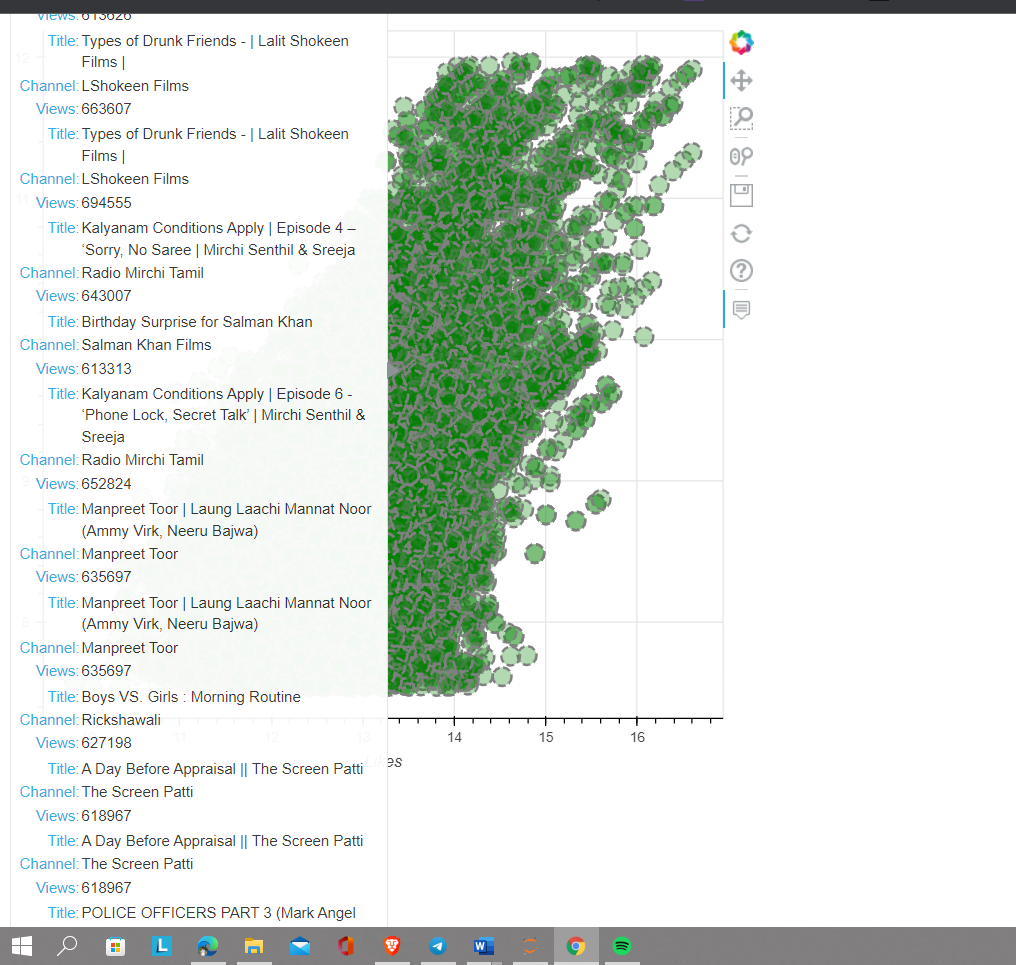
**GITHUB COMMITS:**

****

****

****

**BETA TESTING**

****