



# **SearchTube Analytica**

## **PROJECT REPORT**

SUBMITTED BY:

Aarushi Sharma

40821003

Priyanshu Sharma

40821088

Rishabh Adhikari

40821092

For the award of the Degree of

Bachelor of Science (B.Sc.)

In Data Analytics

(Delhi Skill and Entrepreneurship University)

Under the guidance of

Ms. Priyanka

BSc. Data Analytics Faculty

B.Sc. Data Analytics

Bhai Parmanand

Shakarpur-II

2021-2024

# CERTIFICATE

We, **Aarushi Sharma(40821003)**, **Priyanshu Sharma(40821088)**, **Rishabh Adhikari(40821092)** hereby certifies that the work which is being presented in this project entitled **SearchTube Analytica** in fulfillment of the requirements for the award of degree of **BACHELOR OF SCIENCE (B.Sc.)** being submitted in **Data Analytics Department**, Delhi Skill and Entrepreneurship University, New Delhi, is an authentic record of my own work carried out under the guidance of **Ms. Priyanka Singh** .The matter presented in this project report has not been submitted either in part or full to any other University or Institute for the award of any degree.

Date :

**Aarushi Sharma**  
**Priyanshu Sharma**  
**Rishabh Adhikari**

Signature of Student

This certified that the above statement made by the student is correct to the best of my knowledge.

**Ms. Priyanka Singh**  
Delhi Skill and Entrepreneurship  
University, New Delhi

Date :

# ABSTRACT

YouTube is a widely popular online video-sharing platform that allows its users to interact with the creators by uploading, viewing, commenting and hitting the like button and bell icon on the vast collection of video content. [1] In this era of Tech-Savvy Generation, YouTube as a platform as emerged out as one of the most influential platforms ,allowing the influencers and screenagers to share their thoughts on wide range of topics-be it education, fashion and grooming ,documentaries, entertainment shows and much more.

With millions of channels covering a wide range of topics, the YouTube content creation business is a quickly expanding and fiercely competitive sector of the economy. New creators must negotiate the difficulties of breaking through and expanding their channels if they are to thrive in this dynamic environment. This calls for the use of comparative analysis of YouTube channels to learn important things about their audience and niche.

By leveraging YouTube's API key to retrieve channel data such as statistics, engagement rates, and video metrics, creators can gain a comprehensive understanding of their own channel as well as their competitors'. This analysis offers several advantages, including improved content quality, increased engagement and audience growth, competitive differentiation, collaboration opportunities, and actionable insights.

Through a thorough examination of successful channels within their niche, creators can identify trends, popular topics, and audience preferences, thus enhancing the quality of their content. Creators may learn which sorts of content lead to greater relationships with their audience and draw in more viewers by analysing metrics like likes, comments, and shares on their channel and the channels of their rivals. Additionally, by examining rival channels, producers may spot market gaps and come up with original content concepts that will set them apart from the competitors [2].

Being mindful of such factors, this project is about integrating a dummy search engine platform for the videos' statistics data and comparative aanalysis of YouTube channels through user provided channel Ids, fetched with the help of YouTube API key.

# ACKNOWLEDGEMENT

We would like to express our sincere gratitude and appreciation to all the individuals and organizations who have contributed to the successful completion of the project "SearchTube Analytica"

First and foremost, we extend our heartfelt thanks to Ms. Priyanka, our project guide, for her valuable guidance, constant support, and motivation throughout the project. Her expertise and insights have been instrumental in shaping our understanding and implementation of the project.

We are grateful to HOD BSc. Data Analytics Dr, Mamata Mittal and other faculty members of the Department of BSc Data Analytics at Delhi Skill and Entrepreneurship University, for providing us with a conducive learning environment and valuable resources that helped us in the development of this project.

We would also like to acknowledge the support and cooperation of our classmates and friends who provided assistance and encouragement whenever needed. Their inputs and discussions have been valuable in shaping the project.

Furthermore, we express our gratitude to the management of Bhai Parmanand Shakarpur-II and Delhi Skill and Entrepreneurship University for providing us with the necessary infrastructure and facilities to carry out this project.

Last but not least, we would like to thank our families for their constant support, understanding, and encouragement throughout the duration of this project. Their belief in us and their words of motivation have been a driving force behind our efforts.

In conclusion, we would like to acknowledge the collective efforts and contributions of all the individuals and organizations mentioned above, as without their support, this project would not have been possible.

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overview .....	1
<b>2</b>	<b>System Analysis</b>	<b>2</b>
2.1	Purpose .....	2
2.1.1	Existing System .....	2
2.1.2	Proposed System .....	3
2.2	Problem definition.....	3
2.3	Feasibility Study .....	3
2.3.1	Technical Feasibility .....	3
2.3.2	Economic Feasibility .....	3
2.3.3	Operational Feasibility .....	3
2.4	Objective of the study .....	3
<b>3</b>	<b>System Requirement</b>	<b>4</b>
<b>4</b>	<b>Methodology</b>	<b>5</b>
<b>5</b>	<b>Design Document</b>	<b>7</b>
5.1	Purpose.....	7
5.2	Scope .....	7
5.3	Overview .....	8
5.4	Data Design.....	8
5.5	ER Diagram.....	10
5.6	Cardinality.....	10
<b>6</b>	<b>Development of the System</b>	<b>11</b>
<b>7</b>	<b>Outcomes</b>	
<b>8</b>	<b>Future Scope</b>	
<b>9</b>	<b>References</b>	

# LIST OF FIGURES AND TABLES

Figure 4. 1: Methodology of SearchTube Analytica .....	13
Figure 5.1: ER Diagram .....	17
Table 5.1: ER Diagram Cardinalities .....	18
Figure 6.1: General Demonstration of Working of API key.....	19
Figure 6.2: Backend Overview of SearchTube Analytica .....	22
Figure 6.3: Home Page of SearchTube Analytica .....	23
Figure 6.4: Instructions page of SearchTube Analytica .....	23
Figure 6.5: Discover Section of SearchTube Analytica .....	24
Figure 6.6: Analysis section of SearchTube Analytica .....	24
Figure 6.7: Combining the back end and front-end sides. ....	25
Figure 7.1: Bar Chart for Subscriber Count .....	26
Figure 7.2: Pie Chart for View Count Proportion .....	27
Figure 7.3: Box Plot for View Count .....	28
Figure 7.4 :Scatter Plot for View and Like Count .....	28
Figure 7.5: Line Chart for Engagement(views) .....	29
Figure 7.6: Line Chart for Engagement(Likes) .....	30
Figure 7.7: Line Chart for Engagement(Comments).....	30
Figure 7.8: Histogram for Description length and View Count .....	31
Figure 7.9: Stacked Column Chart for video publishing dates .....	31
Figure 7.9: Stacked Column Chart for video publishing dates .....	32
Figure 7.10: Word Cloud for Description .....	32
Figure 7.11: Word Cloud for Tags.....	32
Figure 7.11: Word Cloud for Tags.....	33
Figure 7.12: Pie Chart for tags proportion .....	33
Figure 7.13: Histogram for Content Duration .....	34
Figure 7.14: Scatter plot for Tags and View Count.....	35
Figure 7.15: Scatter Plot for Title length and View Count.....	35
Figure 7.16: Scatter Plot for Duration and View Count.....	36

# CHAPTER 1

## 1. Introduction

YouTube being the second largest search engine with more than 60 billion hours' worth of video views per month is a large contributor to bandwidth consumption both on consumer premises as well as businesses. [1]

The YouTube content creation market is a growing and highly competitive industry. There are now millions of channels on the platform, covering a wide range of topics and interests. However, with so much competition on the platform, it can be challenging for new creators to break through and grow their channels. Hence, YouTube channels comparative analysis is important to keep a track on your channel analysis as well as your' s competitor channel.

YouTube analysis is a thorough process that involves analysing various aspects of YouTube videos, channels, and the platform itself to gain valuable insights and improve performance. Studying video metrics such as views, likes, dislikes, comments, watch time, CTR, and average view duration is critical. These metrics can be used to determine how well a video is performing, the level of audience engagement, and the overall impact it has on viewers. Channel metrics like the number of subscribers, subscriber growth rate, views per channel, and engagement rate provide an overview of a channel's popularity, its growth trajectory, and the level of loyalty exhibited by its audience.

YouTube analysis tools available on the market have many complex features, charge a fee, and provide limited analysis. This project aims to provide a comparative perspective to the analysis, which is easily accessible to the audience via a search engine web app.

### 1.1 Overview

The project, "SearchTube Analytica" is about unleashing the potential of the YouTube channel ,discovering trends, analysing competitors and charting the journey with the cutting-edge analysis and search engine web app.

# CHAPTER 2

## 2. System Analysis

### 2.1 Purpose

To this date, it is difficult to find high-level statistics on YouTube that paint a fair picture of the platform in its entirety . [2]The project's goal is to create a user-friendly web application that allows users and content creators to analyze and compare various YouTube channels in real time and for free. This app will provide detailed metrics such as views, likes, and subscriber count, allowing users to gain insights into channel performance and growth. The application's highlight will be its ability to enable easy side-by-side comparisons of multiple channels, allowing users to make data-driven decisions. Overall, the project's goal is to provide a free and easy-to-use solution for YouTube channel analysis and comparison.

#### 2.1.1 Existing System

It is possible for creators and businesses to analyse their channel's performance, audience engagement, and video data using a variety of methods and technologies that are being used for YouTube analytics. [1] Here are a few well-liked choices: [5]

**1. YouTube Analytics :** YouTube offers creators a thorough analytics dashboard. It provides information on viewers' demographics, traffic sources, views, watch time, engagement indicators, and more.

**2. Google Analytics :**Although not specifically designed for YouTube, Google Analytics may be combined with your YouTube channel to collect more detailed information.

**3. Social Blade :** It offers real-time YouTube channel statistics, including subscriber count, video views, engagement metrics, and estimated earnings.

**4. TubeBuddy :** It offers data on video performance, demographics, and competitor insights.



### 2.1.2 Proposed System

The proposed system is a web application that offers free and easy-to-use tools for analysing YouTube channels to users and content creators. The application retrieves live information about various channels via API keys, providing real-time insights into their performance metrics, audience engagement, and growth trends. The ability to conduct comparative analysis among selected channels is one of the system's key features. This enables content creators to manage their channels more effectively by identifying successful strategies, benchmarking against competitors, and identifying areas for improvement. It also helps users looking for specific content discover channels that consistently deliver high-quality content in their areas of interest.

This system promotes data-driven decision-making, improves content strategy development, and increases audience engagement by removing any financial barriers and providing comprehensive analysis tools. Its goal is to provide a user-friendly interface that simplifies the analysis process and assists both creators and users in their pursuit of YouTube success and content discovery.

### 2.2 Problem Definition

The system aims to analyze YouTube channels, understand competition, evaluate performance, identify opportunities, and provide users with relevant content efficiently, saving time and effort.

### 2.3 Feasibility Study

The feasibility of the project can be studied on the basis of three aspects:

**2.3.1 Technical Feasibility:** The project is deployed using the existing hardware/software that can support the algorithms that we will be implementing in our project.

**2.3.2 Economical Feasibility:** The model doesn't demand any financial aid.

**2.3.3 Operational Feasibility:** The project is implemented in a way that integrates a search engine, called the Discover section and an Analysis Section on a web app -SearchTube Analytica, where user can find the YouTube Channels he/she queries for and can perform comparative analysis (recommended for Content Creators)

## **2.4 Objective of the study**

Time is one of the most valuable resources in today's competitive market, and manually searching through and comparing various channels, whether for personal or professional usage, takes time. This topic came to light as a result of the need for a way to study more and save time in this area. This application's main goal is to assist users and content producers in saving time and knowing more about their preferred YouTube channels as well as others.

# CHAPTER 3

## 3. System Requirements

### Hardware:

- Processor: Intel Core i5 or equivalent
- RAM: 8GB
- Storage: 250GB of available hard disk space/ssd
- Graphics: Any modern integrated graphics card (To speed up the processing speed)

### Software:

- Python (Programming language)
- Python libraries such as pandas, matplotlib.pyplot, seaborn, googlepiclient.discovery, datetime, NumPy
- Other python libraries such as streamlit, isodate, parser, tabulate, plotly, plotly express
- Jupyter Notebook

# CHAPTER 4

## 4. Methodology

**4.1 API Key Creation:** . An API key is a unique identifier that allows access to an API. It is typically used to authenticate and authorize access to the services of an API. The foremost step to retrieve data from YouTube is to have a unique credential i.e., the YouTube API Key. An API key can be easily generated with creating a project and related credentials on Google Developers Console . [3]

**4.2 Data Collection and Creation:** YouTube has its own Database stored in json format. The organization maintains a comprehensive database that stores a vast amount of information related to its platform such as videos, channels, playlists, comments, subscriptions, analytics etc. This massive amount of data is equivalent to a resource, which has its specific methods and endpoints that an external user can use to interact with and retrieve data from the YouTube platform. For an external user to use these functionalities, a user must have a unique identifier i.e., YouTube API Key. For the project, we are fetching the data with the help of an API key.

**4.3 Pre-Processing the Data:** The collected data needs to be processed before it can be used for the analysis. Pre-processing steps included under this project are:

- Converting duration of videos into seconds.
- Converting specific columns for numeric data type.

**4.4 Data enrichment:** Data enrichment is a process that involves combining first-party data collected from internal sources with data collected from other internal sources or third-party external sources. Following are the ways in which the data is enriched under this project:

- Extracting the days on which the videos were uploaded.
- Converting the duration of the videos to seconds.
- Quantifying the number of tags used.
- Computing the length of Description.
- Computing the length of Title.

**4.5 Output Optimization:** The YouTube API keys allows data to be fetched for limited number resources for a single API key. If these resources exceed the limit, then a Quota Error is executed. Therefore, the results to be displayed on SearchTube Analytica need to be optimized so that the API key does not get exhaust in a single execution/iteration. Moreover, to increase the efficiency in fetching out the results, the number of resources to be fetched should be optimized.\

**4.6 Exploratory Data Analysis(EDA) /Comparative Analysis:** It is the summarization of the data(the actual comparative analysis) highlighting the main points often using visualizations. The Analysis section

of SearchTube Analytica displays the comparative analysis of channels according to the user with the help of visualizations such as :

- Bar Chart
- Pie Chart
- Box Plot
- Channel Engagement Charts (line Chart)
- Histograms
- Scatter Plot
- Word Cloud

**4.7 Integration and Front End:** Integrate the outputs/results of Discover and Analysis sections on a web app named SearchTube Analytica so as to make user experience and user interface better. The web app is deployed with the help of Streamlit .

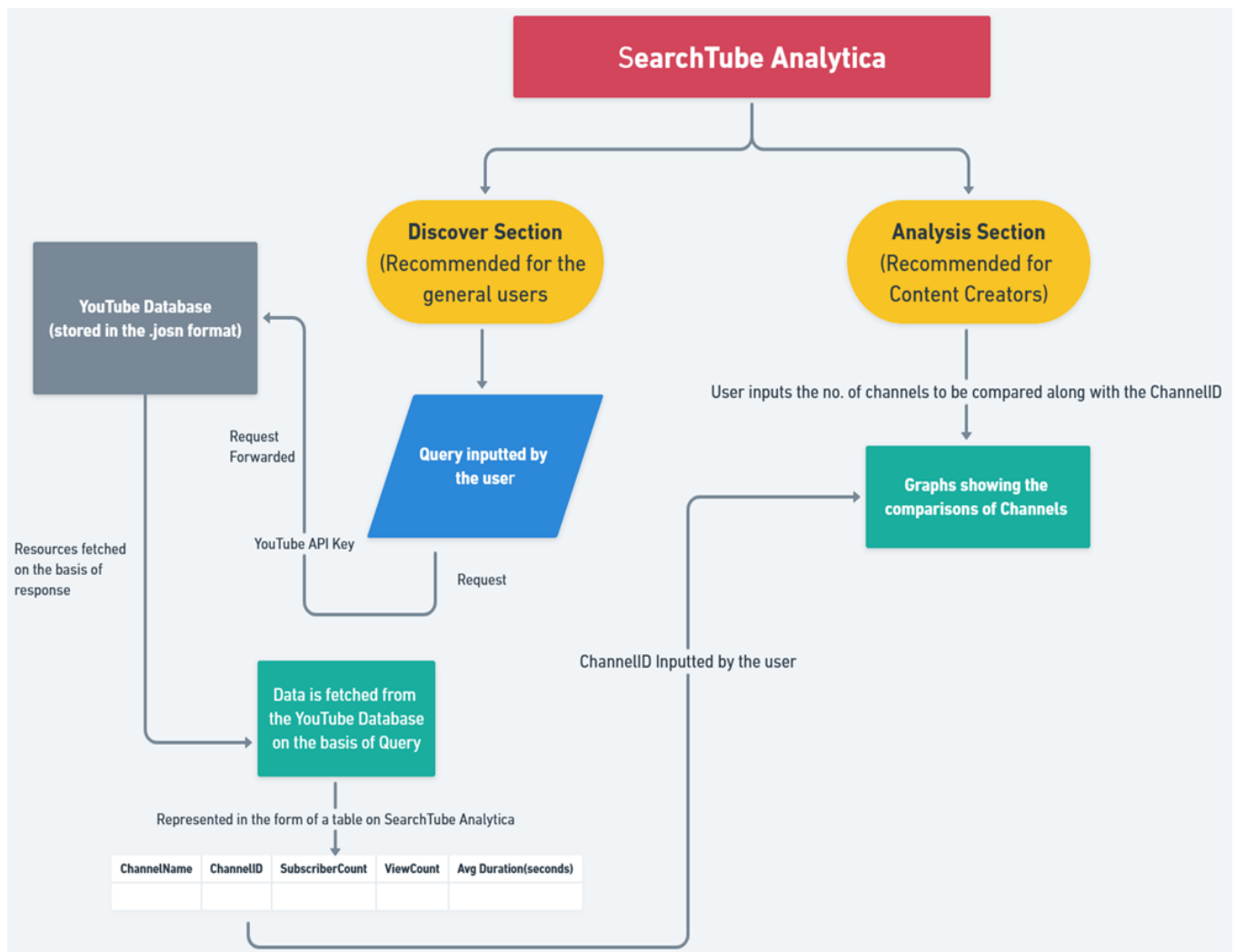


Figure 4. 1: Methodology of SearchTube Analytica

# CHAPTER 5

## 5. Design Document

### 5.1 Purpose:

Purpose of "SearchTube Analytica," is to develop a user-friendly web application that allows users and content creators to analyze and compare various YouTube channels in real-time and for free. The application aims to provide comprehensive metrics and insights into channel performance, audience engagement, and growth trends.

The purpose of the SearchTube Analytica's Discover section, is to save users time when searching for YouTube channels related to user query by automating the process and providing relevant information in a structured format. The code leverages the YouTube Data API to perform the search and retrieve important details about the channels, such as view counts, subscriber counts, average video duration, and average watch time.

By using this, users can quickly obtain an organized list of channels that match their search query without manually visiting each channel and collecting the data individually. This saves time and effort, especially when dealing with a large number of channels.

Moreover, SearchTube Analytica's Analysis section, aims to simplify the analysis process and assist both content creators and users in their pursuit of YouTube success and content discovery. By studying successful channels within their niche and identifying trends, popular topics, and audience preferences, creators can enhance the quality of their content and differentiate themselves from competitors.

### 5.2 Scope

The scope of the "SearchTube Analytica" project encompasses the development of a web application that facilitates the analysis and comparison of YouTube channels. The project aims to provide a range of features and functionalities to users and content creators, enabling them to gain valuable insights into channel performance and audience engagement.

The scope of the project are as follows:

**Efficiency:** The process of searching for YouTube channels and gathering channel data, eliminating the need for manual search and data collection. This saves users considerable time and effort.

**YouTube Channel Analysis:** The application allows users to analyze their own YouTube channels by retrieving and presenting various metrics such as views, likes, comments, subscriber count, engagement rate, and more. Users can track the performance of their channels over time and identify areas for improvement.

**Comparative Analysis:** One of the main highlights of the project is the ability to conduct side-by-side comparisons of multiple YouTube channels. Users can select specific channels and compare their metrics, engagement rates, growth trends, and other relevant data. This feature enables users to benchmark their channel's performance against competitors and identify strategies for success.

**Trend Analysis:** The project aims to provide insights into trends within specific niches or topics on YouTube. By analysing successful channels within a niche, users can identify popular content themes, audience preferences, and emerging trends. This information can be used to enhance content strategy and stay ahead in the dynamic YouTube landscape.

**User-Friendly Interface:** The web application is designed to be user-friendly, providing a seamless and intuitive experience for users. The interface will be accessible and easy to navigate, allowing users to retrieve and visualize channel data effortlessly. The goal is to simplify the analysis process and make it accessible to users with varying levels of technical expertise.

**Real-Time Data Retrieval:** The project utilizes the YouTube API key to retrieve live data about channels, ensuring that users have access to up-to-date information. Real-time data retrieval enables users to monitor channel performance in real-time and make informed decisions based on the latest insights.

**Data Visualization:** The project incorporates various data visualization techniques such as bar charts, pie charts, box plots, line charts, histograms, scatter plots, and word clouds. These visualizations aid in the interpretation and understanding of channel metrics, engagement patterns, and comparative analysis.

**Feasibility :** The application doesn't require any additional fee for access.

### 5.3 Overview :

This Web application uses the YouTube API to find and analyze YouTube channels related to “query” of the user. It retrieves channel data such as view count, subscriber count, average watch time, and average duration, presenting it in a structured format to save users time and effort. [5]

This web application offers free and user friendly tools for analyzing and comparing YouTube channels. It leverages YouTube’s API key to provide real-time insights into channel performance and growth. By conducting comparative analysis and studying metrics like views, likes, and subscriber count, users can make data-driven decisions to improve their own channels and stay ahead of competitors.

The application aims to enhance content quality, increase engagement, and foster audience growth. With its intuitive interface and comprehensive analysis tools, "SearchTube Analytica" is designed to support both content creators and users in their pursuit of YouTube success and discovery.

### 5.4 Data Design:

YouTube maintains and organizes comprehensive data, consisting of videos, likes, views, comments, channel IDs, description, tags and other meta data. An activity resource contains information about an action that a particular channel, or user, has taken on YouTube. The actions reported in activity feeds include rating a video, sharing a video, marking a video as a favorite, uploading a video, and so forth. Each activity resource identifies the type of action, the channel associated with the action, and the resource(s) associated with the action, such as the video that was rated or uploaded. [1]

YouTube has its own Database stored in json format. This massive amount of data is equivalent to a resource, which has its specific methods and endpoints that an external user can use to interact with and retrieve data from the YouTube platform. For an external user to use these functionalities, a user must have a unique identifier i.e., YouTube API Key. For the project, we are fetching the data with the help of an API key

In the context of the YouTube API, "resource," "item," and "snippet" are terms that refer to different aspects of the data structure and organization within the API.

**Resource:** A resource in the YouTube API represents an entity or object that you can interact with or retrieve data about. Resources include items like videos, playlists, channels, comments, and more. Each type of resource has its own unique set of properties and information associated with it.

**Item:** In the YouTube API, an "item" typically refers to an individual instance of a resource. For example, if you're retrieving a list of videos from a channel, each video in the list would be considered an item. Similarly, if you're retrieving comments on a video, each comment would be an item.



Snippet: The "snippet" is a specific part of the resource or item that contains a subset of information about it. It provides basic details and metadata for the resource without going into exhaustive detail. The snippet typically includes information like the title, description, thumbnails, publication date, and other relevant data that provides a summary or snapshot of the resource.

### 5.5 ER Diagram:

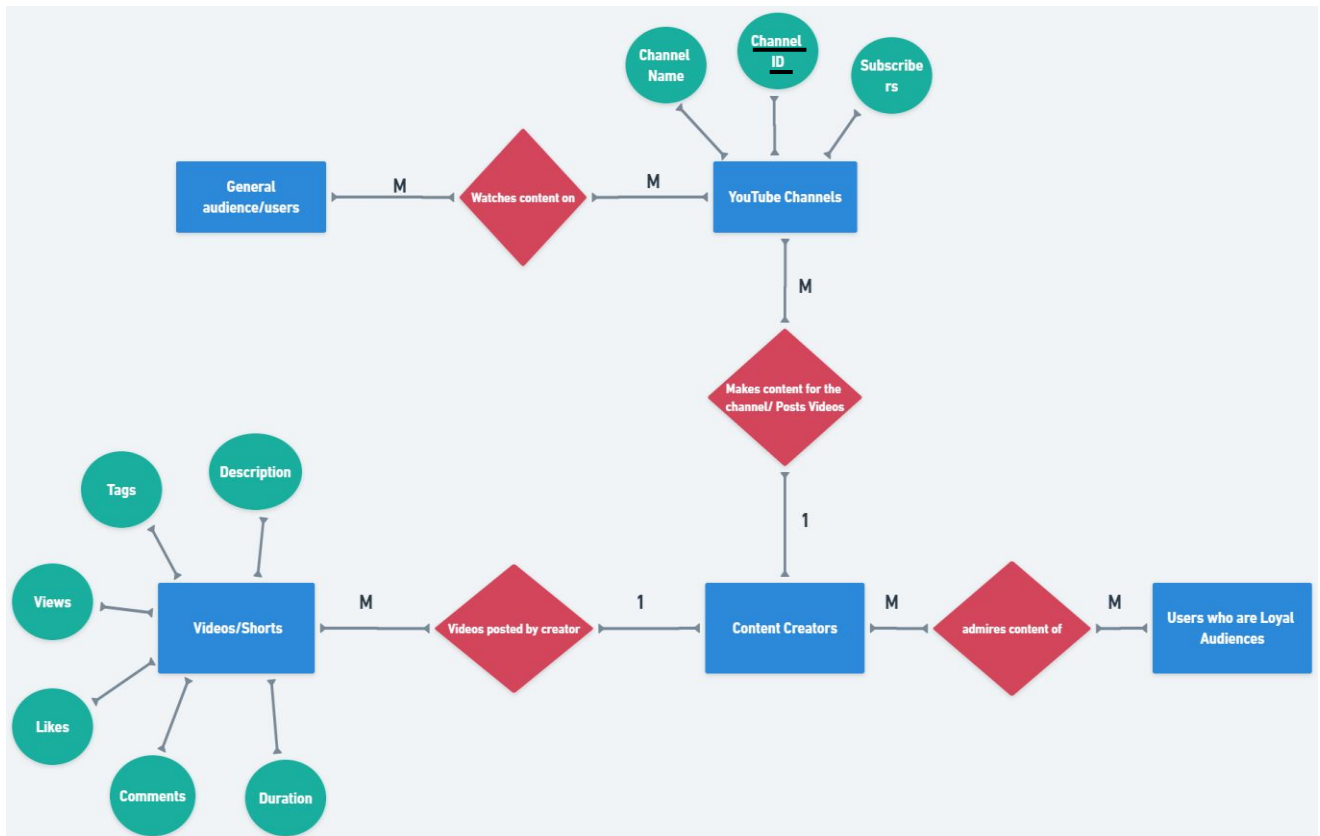


Figure 5.1: ER Diagram

**Entities of the system are:**

- General users
- YouTube Channels
- Videos/Shorts
- Content Creators
- Users who are loyal audience

**Primary Key :**

- ChannelId

**Attributes of the Entities:**

- YouTube Channels : Channel Name, Subscribers , Channel Ids
- Video/Shorts : Likes, Views Comments , Description ,Duration, Tags

**5.6 Cardinality:**

ENTITIES	RELATIONSHIP/CARDINALITIES
Content Creator & YouTube Channel	One To Many
General Audience/Users & YouTube Channel	Many To Many
Content Creators & videos/Shorts	One To Many
Content Creators & Loyal Audiences	Many To Many

*Table 5.1: ER Diagram Cardinalities*

# CHAPTER 6

## 6. Development of the system

The web app SearchTube Analytica integrates a search engine with comparative analysis of channels. The search engine fetches out the data and displays it in a form of table with various attributes such as ChannelName ,SubscriberCount ,ViewCount, ChannelID, AverageDuration on the basis of a query inputted by the user. The results are being fetched from the YouTube database with the help of a YouTube API key.

### What is an API?

API stands for Application Programming interface, which are a set of rules and protocols that allow different Software Applications to communicate and interact with each other .They helps the users/developers to access specific functionalities or data provided by a service or platform. In order to access the data or specific functionalities , the developer needs an API key. An API key is a unique identifier that allows access to an API. It is typically used to authenticate and authorize access to the services of an API. [3]

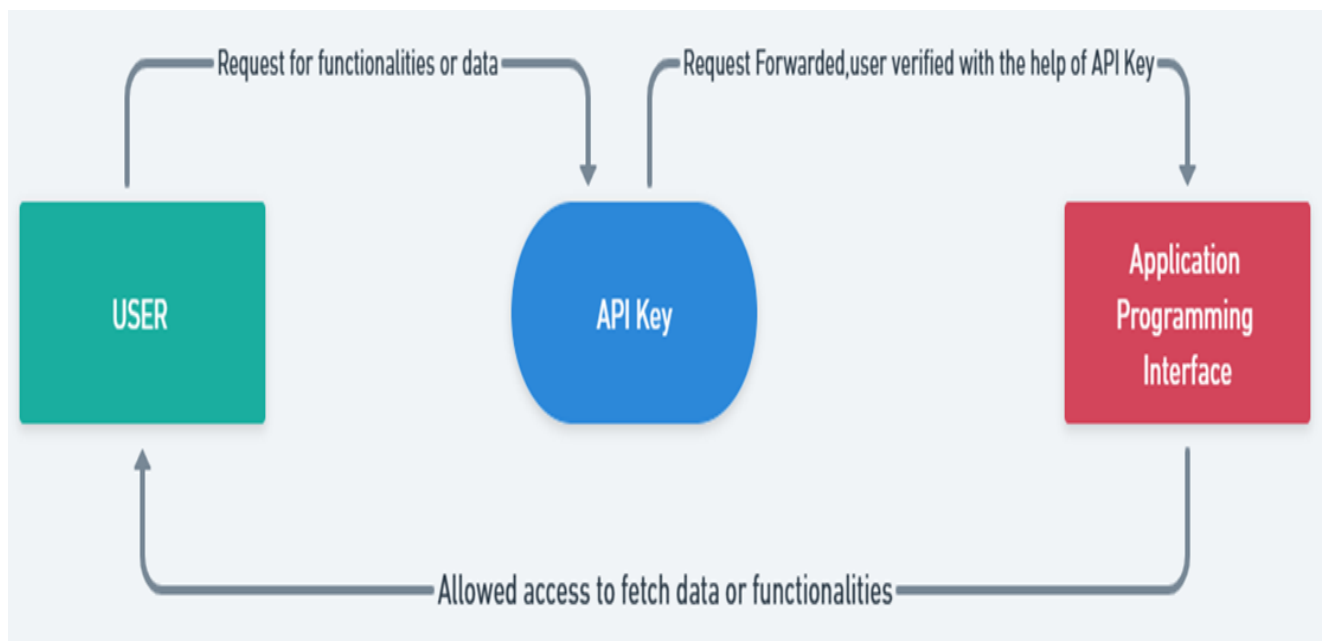


Figure 6.1: General Demonstration of Working of API key

## **Back-End Side:**

### **1. Fetching data from YouTube Database with the help of API Key:**

To fetch data from YouTube using the YouTube API key, one need to follow these general steps:

- Obtain a YouTube API Key:
- Set up the API Client
- Authenticate your API Requests
- Choose the API Endpoint
- Make API Requests
- Process the Response along with preprocessing (if required)

These all above steps were executed for the implementation of the Discover Section(Search engine) so as to display the data fetched out on the webapp.

### **2. Creating temporary Dataframes with the resources fetched out as response:**

The channelId is the unique identifier for a specific channel. It can act as a primary key for the temporary database created. Hence, for the analysis section the user needs to input the number of channels he/she want to compare along with the unique identifier of the respective channels i.e., the ChannelID. On the basis of the ChannelId , the resources are fetched from the YouTube Database for that particular channel. These resources are as follows:

- For channel\_data :
  - channelTitle
  - subscriberCount
  - videoCount
  - viewCount
- For video\_df (for each video uploaded on the Channel) :
  - ChannelTitle
  - title
  - description
  - tags
  - published at

The resources fetched with the help of API key are stored as attributes in their respective dataframes which are further utilized in the implementation of the overall analysis i.e., graphs for channels comparison

### **3. Pre-Processing and Data Enrichment:**

Pre-Process and enriching the data fetched for analysis. Preprocessing and Data Enrichment steps implemented are as follows:

- Converting specific columns to numeric datatype.
- Extracting the days on which the videos were uploaded.
- Converting the duration of the videos to seconds.
- Quantifying the number of tags used.
- Computing the length of Description.
- Computing the length of Title.

### **4. Creating functions for visualizations:**

The Analysis section of SearchTube Analytica displays the comparative analysis of channels according to the user with the help of visualizations. Functions for the following visualizations were created so as to display them on webapp ,deployed with the help of Streamlit.

- Bar Chart
- Pie Chart
- Box Plot
- Channel Engagement Charts (line Chart)
- Histograms
- Scatter Plot
- Word Cloud

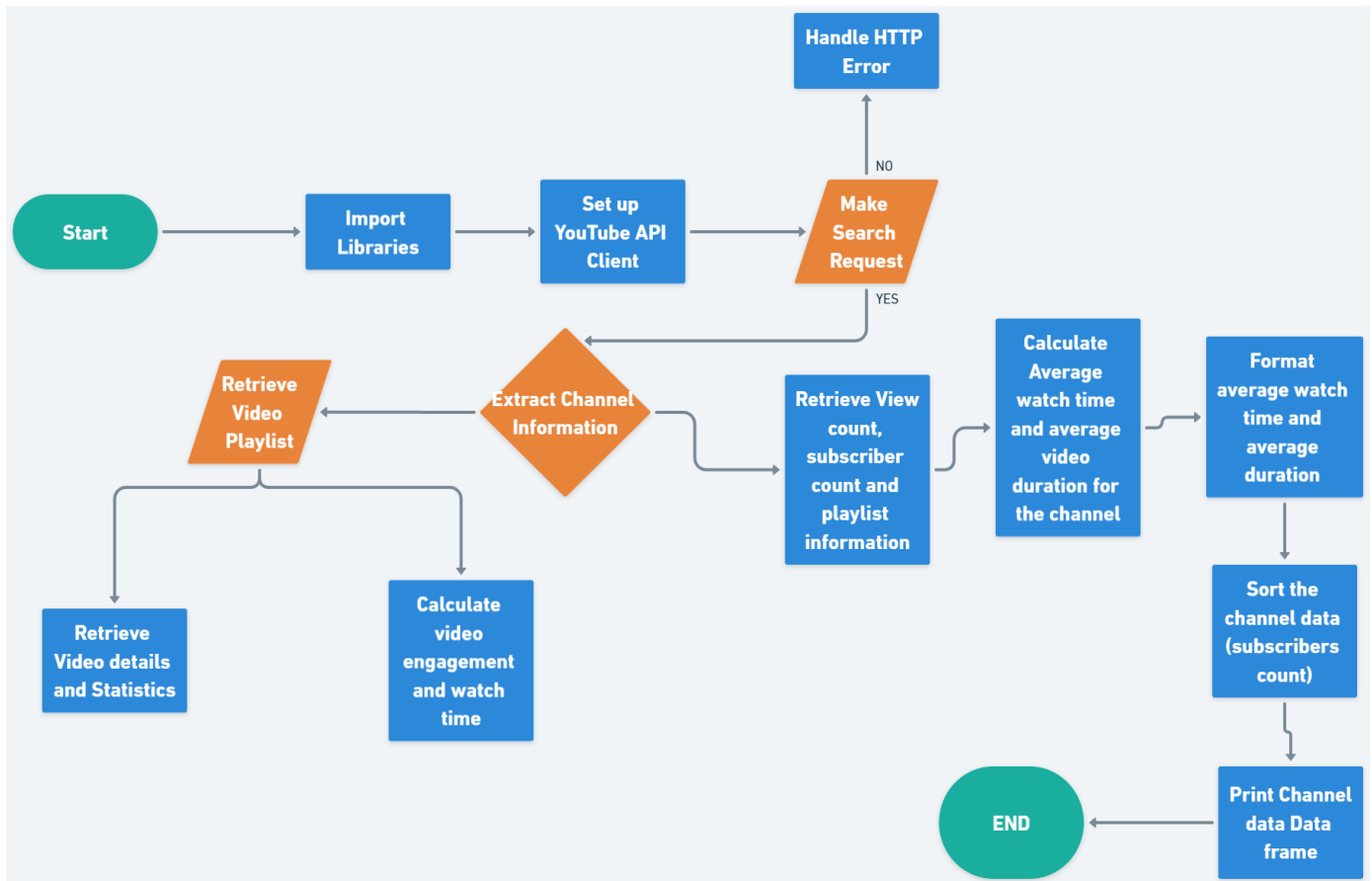


Figure 6.2: Backend Overview of SearchTube Analytica

## Front -End Side

### 1. User Interface Development:

The backend side of the project is deployed on a web app with the help of Streamlit and its compatible libraries. The name proposed for the web app is SearchTube Analytica, which consists of the following sections:

- Home
- Instructions : General Instructions for using SearchTube Analytica
- Discover : Search Engine
- Analysis : Comparative Analysis of Channels

# SearchTube Analytica

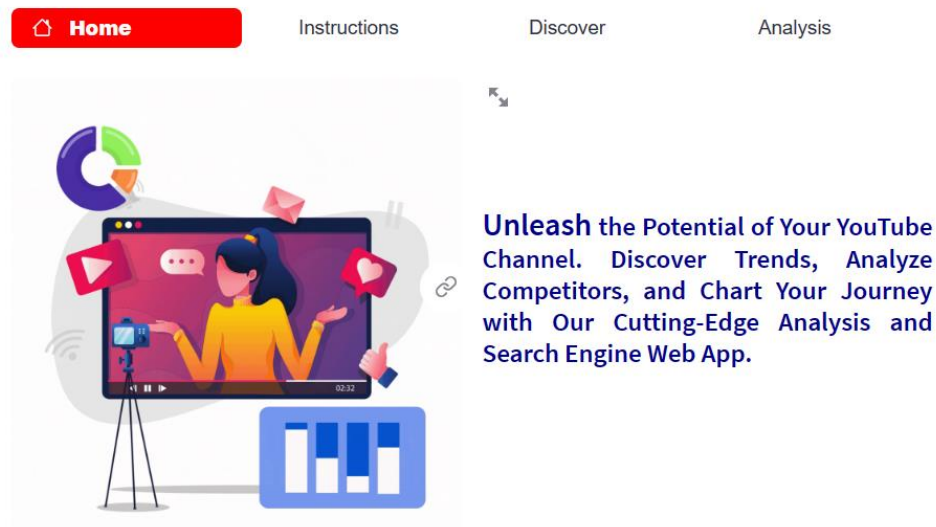


Figure 6.3: Home Page of SearchTube Analytica

# SearchTube Analytica

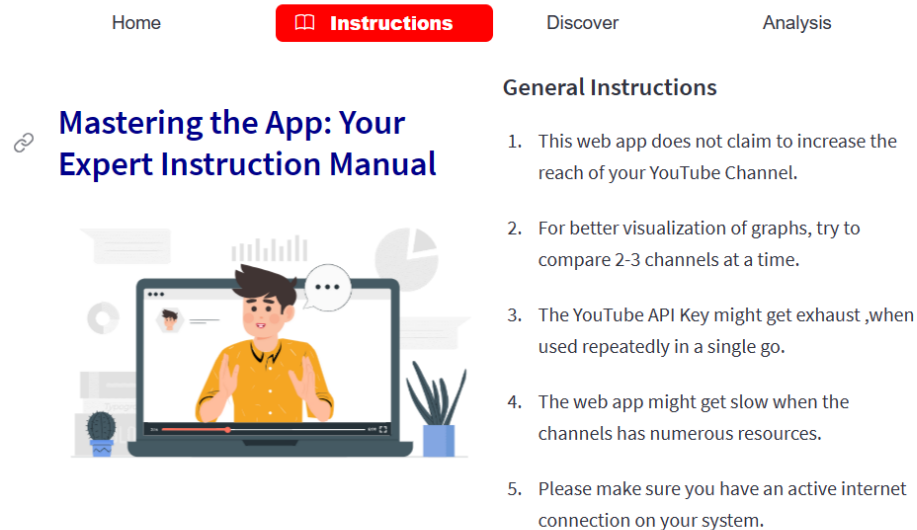


Figure 6.4: Instructions page of SearchTube Analytica

# SearchTube Analytica

[Home](#)[Instructions](#)[Discover](#)[Analysis](#)

Enter the query:

## YouTube Channel Analysis

	Channel Name	Channel ID	View Count	Subscriber Count	Average D
0	Gate Smashers	UCJihyK0A38SZ6SdJirEdIOw	305,578,798	1,440,000	
1	codebasics	UCh9nVJoWXmFb7sLApWGcLPQ	56,212,696	797,000	
2	Great Learning	UCObs0kLlrDjX2LLSybqNaEA	94,241,385	774,000	
3	Krish Naik	UCNU_lfiWBdtULKow6X0Dig	76,386,784	762,000	
4	Learn With Lokesh Lalwani	UCFar3FQxRhS2VNOD3vq1gMQ	22,178,568	563,000	
5	Alex The Analyst	UC7cs8q-gJRIgwj4A8OmCmXg	19,673,712	496,000	
6	5 Minutes Engineering	UCyHta2dyCTkf29AB67AYn7A	74,217,822	479,000	
7	Chandoo	UC8uU_wruBMHeeRma49dtZKA	36,257,679	467,000	

Figure 6.5: Discover Section of SearchTube Analytica

# SearchTube Analytica

[Home](#)[Instructions](#)[Discover](#)[Analysis](#)

## Chart Your YouTube Journey

Enter number of channels you want to compare

Enter the Channel id 1

Enter the Channel id 2

Press Enter to apply

[Analyze](#)

Figure 6.6: Analysis section of SearchTube Analytica



## 2. Integration:

Integrated the backend and frontend side so as to enhance user experience.

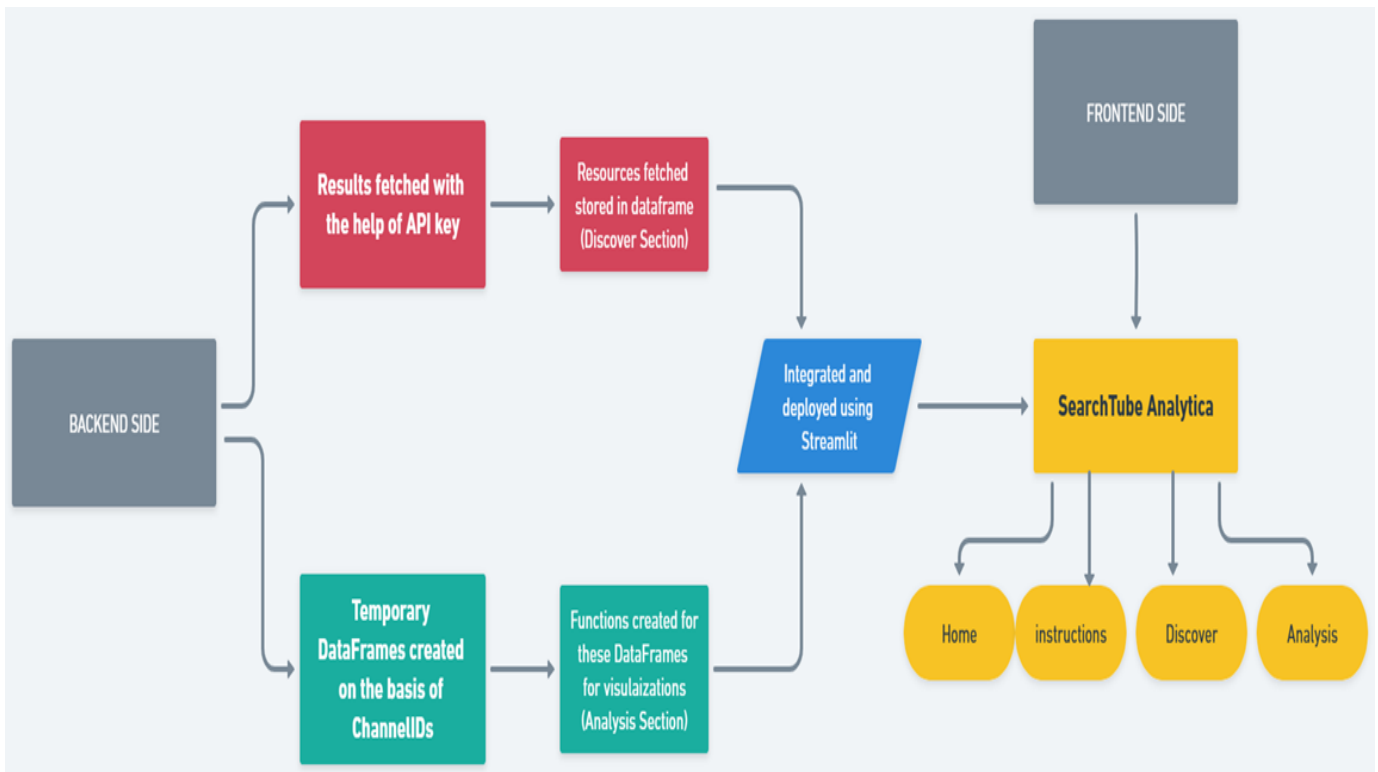


Figure 6.7: Combining the back end and front-end sides.

# CHAPTER 7

## 7. OUTCOMES:

Visualizing data is one of the most valuable skills every Data Scientist and Analyst should have. There are a lot of data visualizations to learn in data science Analysis for a particular domain becomes effective with the implementation of visualizations. [5]Below are the few visualizations which are deployed in the Analysis section of SearchTube Analytica:

### Channel Wise Subscriber Count:

A bar chart is a visual representation of data in the form of rectangular bars, generally used to display and compare categorical data or discrete values.

For the comparison of the subscriber count of each channel, a bar chart is plotted in order to provide a general comparison of how many people have subscribed to a particular channel?

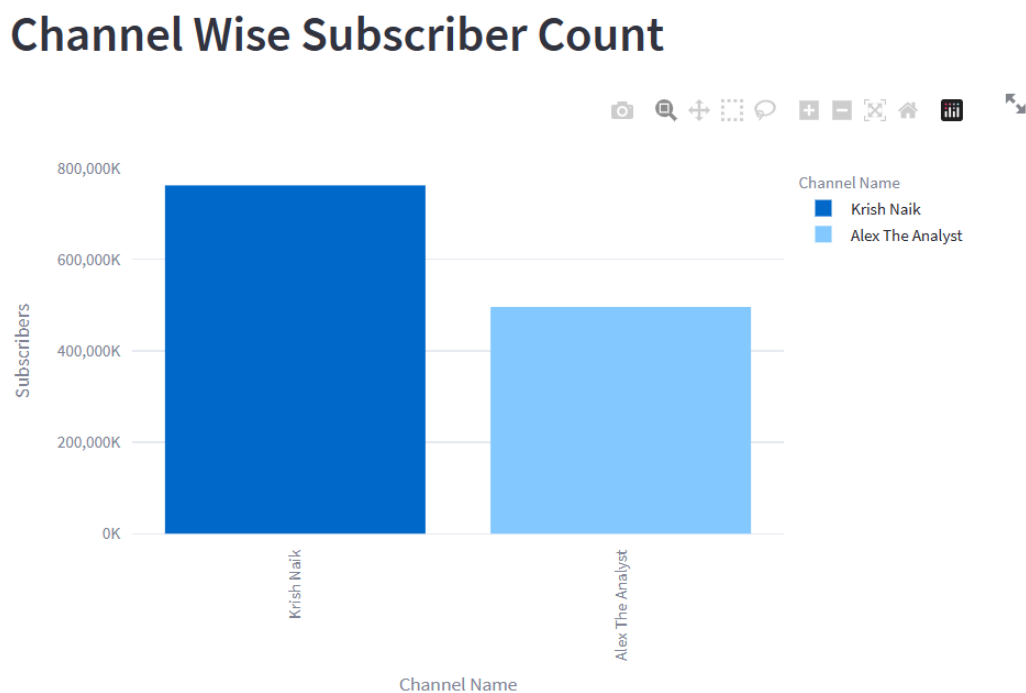


Figure 7.1: Bar Chart for Subscriber Count

### Proportion of Views Each Channel:

A pie chart is a visual representation of data in a form of a circle, divided into sectors to represent the proportion of different categories of a whole.

To visualize the proportion of views of each channel from of a whole, a pie chart is plotted to represent which channel is getting more views?

## Proportion Of Views For Each Channel

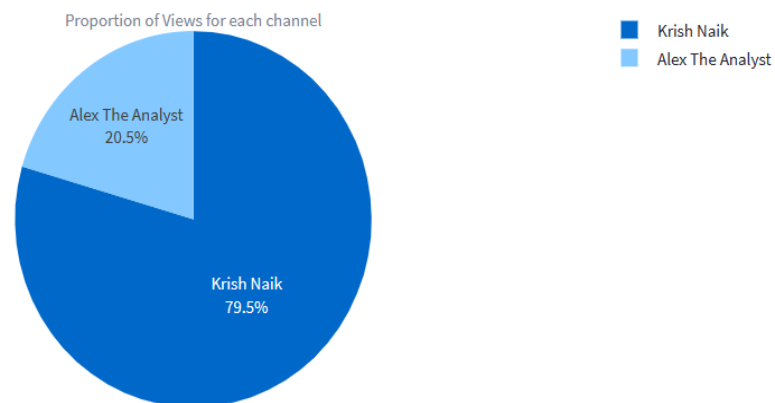


Figure 7.2: Pie Chart for View Count Proportion

### Box Plot analysis for View Count:

A box plot is a statistical visualization tool that displays the distribution of a dataset with key summary. It basically provides a concise summary of dataset's statistical measures such as central tendencies, variance, outliers (if present) etc.

In order to provide the concise summary of the viewCount attribute of the dataset, a box plot is plotted so as to have a general analysis of the viewCount (of whole channel) the channel is getting. Outliers here display the extreme value of viewCount for a particular channel.

## Box Plot Analysis For View Count



Figure 7.3: Box Plot for View Count

### Relationship Between View And Like Count:

For exploring the correlation/relationship between variables , generally scatter plots are plotted. Scatter plots uses a collection of points ,where each point represents the values of the two variables for a specific data point. [6]

Therefore, to find the correlation between the viewCount and LikeCount of a channel a scatter plot is plotted so as to find whether the channel is getting views and likes synchronously.

## Relationship Between View And Like Count

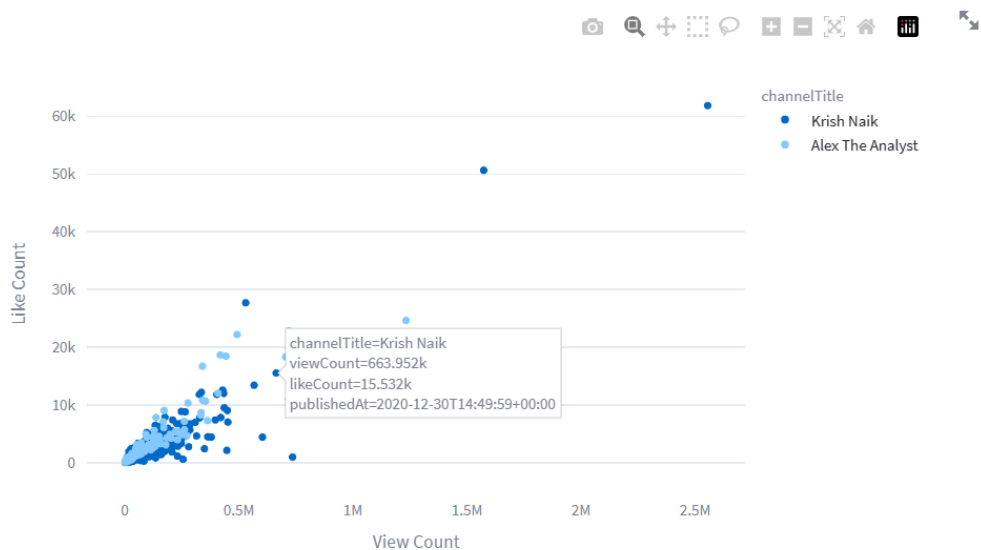


Figure7.4 :Scatter Plot for View and Like Count

### Channel Engagement Chart-ViewCount, LikeCount, CommentCount:

For a YouTube channel ,engagement is important as it estimates the level of interaction and participation from individuals /specific audience .It also shows how many people are viewing your videos, how many people are connecting with the channel ,liking and commenting on the videos. All such metrics defines the channel's loyal audience, eventually leading to higher reach of the channel.

To visualize these metrics, line charts are plotted. Line Charts are visual representation of data that displays the data points connected by line segments. They are usually used to demonstrate the trend the data points are following over a period of time .Channel's overall viewCount ,likeCount and commentCount are plotted against the year from which the YouTube channel is active so as to estimate the channel's engagement.

## Channel Engagement : View Count

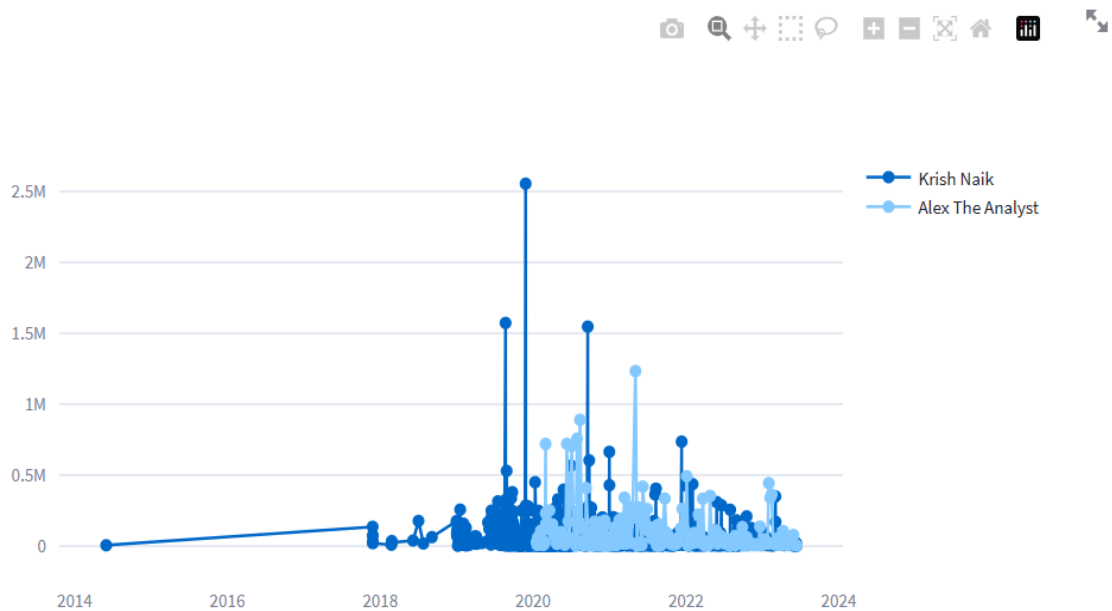


Figure 7.5: Line Chart for Engagement(views)

## Channel Engagement : Like Count

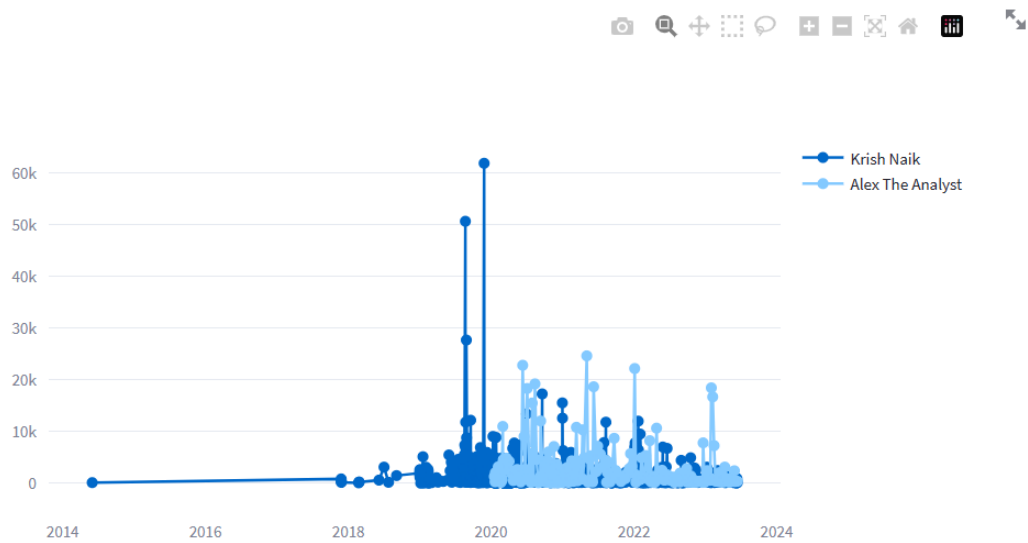


Figure7.6: Line Chart for Engagement(Likes)

## Channel Engagement : Comment Count

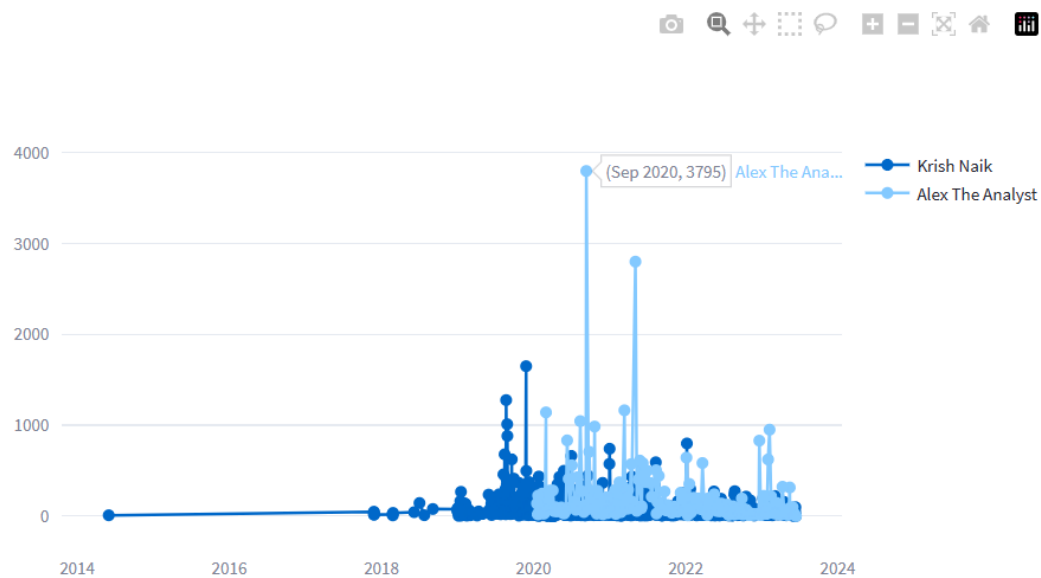


Figure7.7: Line Chart for Engagement(Comments)

**Does description length matters for viewCount**

The description in a YouTube video refers to the text content that accompanies the video, providing additional information to the viewers. The description is an essential component of a YouTube video as it provides valuable context, information, and opportunities for engagement.

A histogram is a graphical representation of the distribution of a dataset. It displays the frequencies or counts of data values within specific intervals, known as bins or classes, along a continuous axis. They provide a visual representation of the data's frequency distribution and help identify patterns, skewness, outliers, and other characteristics of the dataset. To show the distribution of viewCount of the channel for a particular range of description length, histogram is plotted.

## Does Description Length Matters For View Count

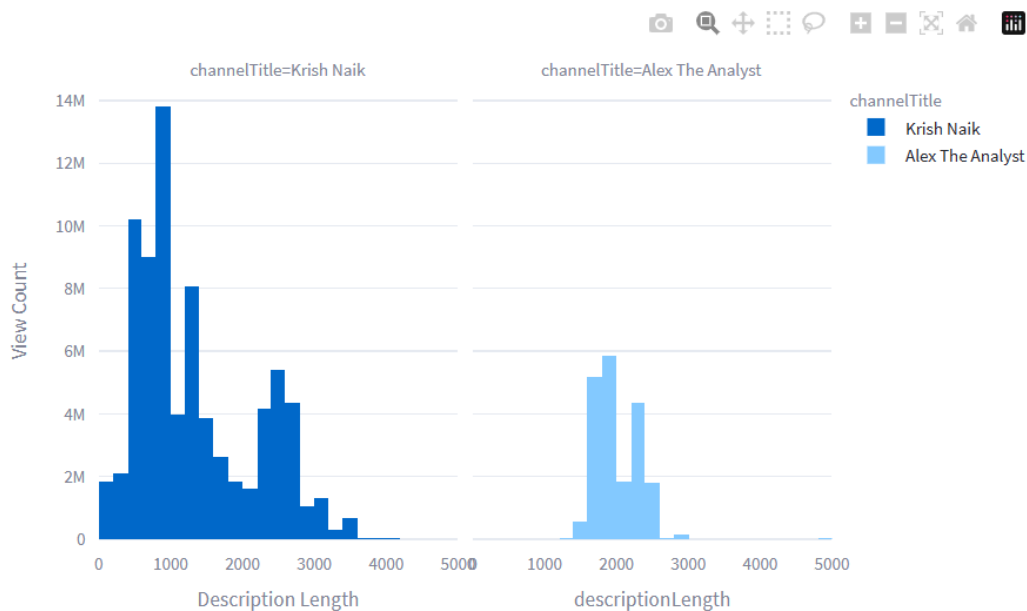


Figure7.8: Histogram for Description length and View Count

### When others are posting Content?

A stacked bar chart, is a data visualization tool that represents multiple variables or categories using horizontal bars or vertical columns. It displays the composition of a whole category while also showing the relative proportions of its subcategories. For finding the number of videos a channel is posting on a particular day of a week stacked bar chart is plotted. It will help the user to know when a channel's competitors is posting content on the YouTube platform?



Word Clouds are the visual representations of textual data where the maximum occurring words are prominent in sizes and are of high importance. Word Cloud are visualized for channel's description and tags in order to show the type of content the channel is posting, the tags used, the collaborations the channels is doing ,the unique topics the channel is aware about which is enhancing their channel's reach and engagement.

[illegible]

Figure7.10: Word Cloud for Description



## Word Cloud : Tags

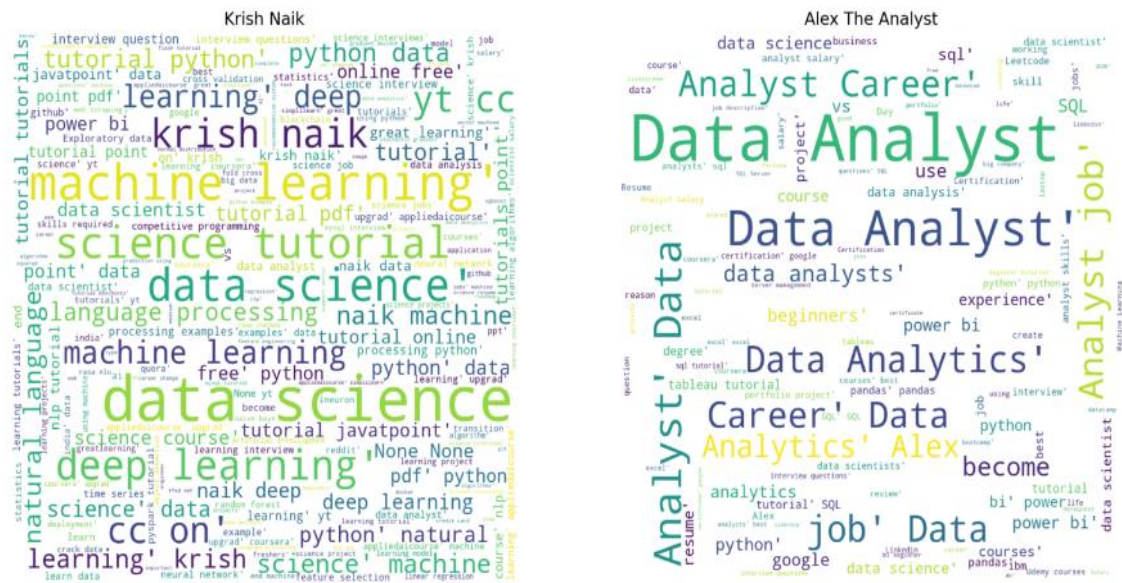


Figure 7.11: Word Cloud for Tags

### How much Proportion of tags the channel is using?

Tags play an important role in optimizing YouTube videos and improving their discoverability. They also help in increasing the reach of the channel. For this purpose pie chart is plotted to find the proportion of tags used by the channel of the whole.

## How much Proportion of Tags the Channel Is Using?

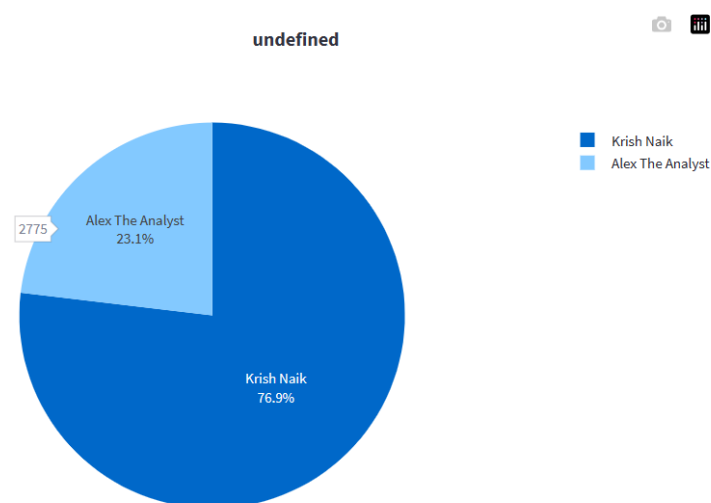


Figure7.12: Pie Chart for tags proportion

## How Much Duration Of Content The Channel Is Posting?

The duration of content might affect the engagement metrics of a channel. Usually people don't opt for videos which have longer duration. Therefore ,a channel must optimize the duration of content to be posted. Histograms are plotted for finding the frequency of videos the channel has posted for a particular range of duration.

## How Much Duration Of Content The Channel Is Posting?

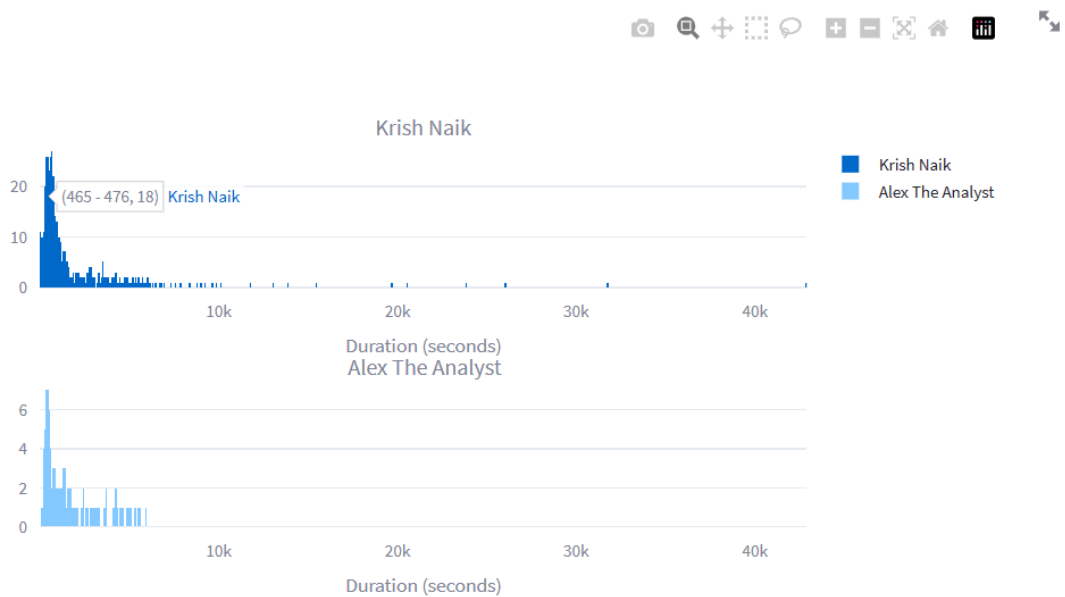


Figure7.13: Histogram for Content Duration

## Relationship between viewCount with Tags Count, Title length and Duration

To explore the relationship of viewCount with other metrics, Scatter plots are plotted .ViewCount might get affected on the basis of duration ,title length and tags used. For e.g.: People might ignore videos having a comparatively longer title or duration, Moreover, it is also possible that the tags used for the videos are not able to rank them at a higher position.

## Is Thier Relationship Between Tags Count And View Count?

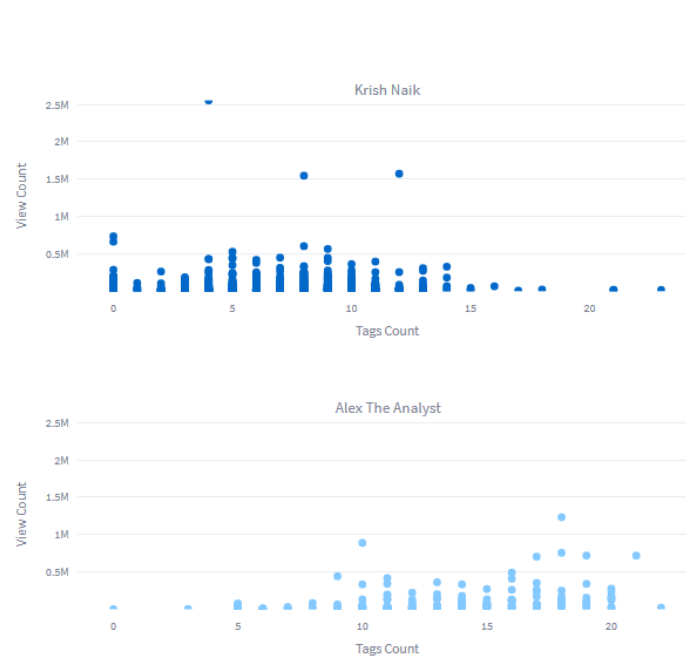


Figure7.14: Scatter plot for Tags and View Count

## Is Thier Relationship Between Title Length And View Count?

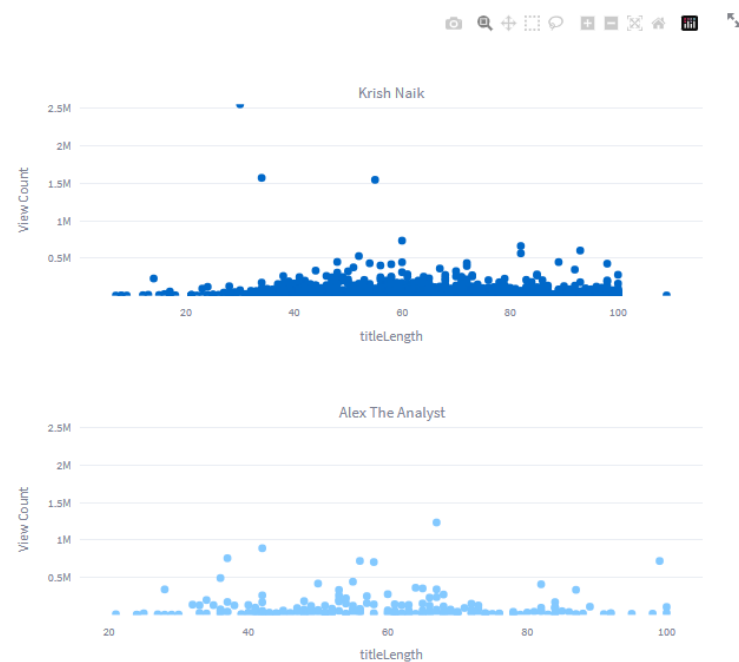


Figure 7.15: Scatter Plot for Title length and View Count

# Is Their Any Realtion Between Duration And View Count?

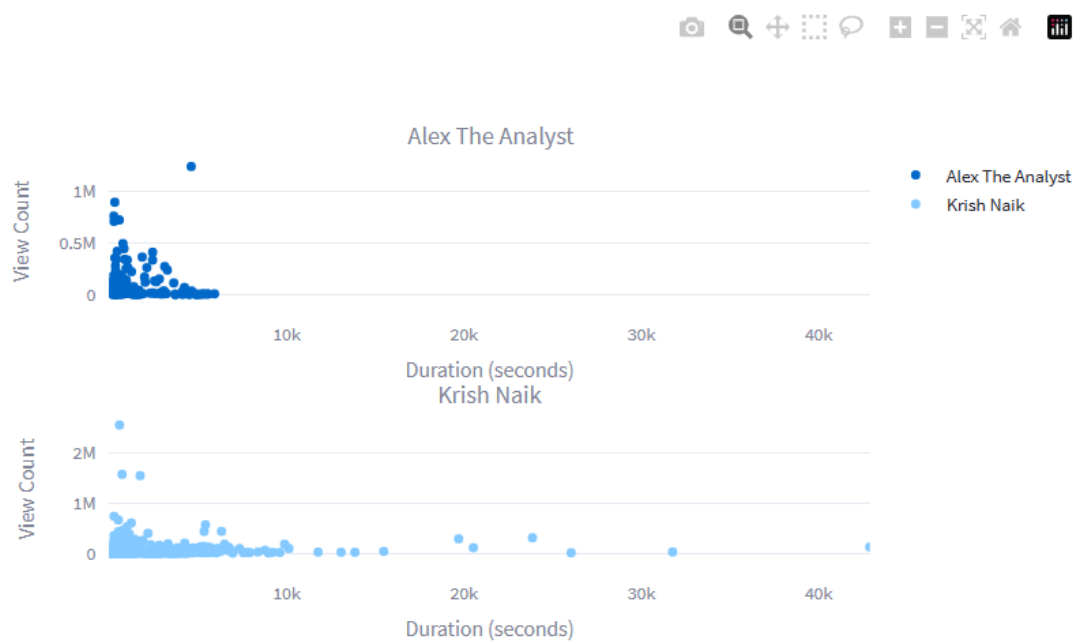


Figure7.16: Scatter Plot for Duration and View Count

# CHAPTER 8

## 8. Future scope :

The future scope of the "SearchTube Analytica" project involves several potential enhancements and expansions. Some of the key areas of future development include:

**1. Enhanced Data Analysis:** The project can incorporate more advanced data analysis techniques and algorithms to provide deeper insights into YouTube channels. This could include sentiment analysis of comments, trend prediction, audience segmentation, and personalized recommendations based on user preferences.

**2. Machine Learning and AI Integration:** By incorporating machine learning and AI techniques, the project can offer more accurate predictions, personalized recommendations, and automated insights generation. This could involve training models to identify trending topics, analyze audience behavior, and suggest optimal content strategies for creators.

**3. Collaboration Features:** Adding collaboration features would enable content creators to connect with each other, share insights, and collaborate on projects. This could involve a platform for creators to find potential collaborators based on their niche, audience, or content type, fostering a community of creators supporting each other.

**4. Monetization Insights:** Providing insights on monetization strategies and revenue generation could be valuable for creators. This could involve analyzing ad performance, identifying opportunities for brand collaborations, and offering guidance on optimizing revenue streams.

**5. Mobile Application:** Developing a mobile application version of the project would allow users to access and analyze their YouTube channels on-the-go. This would enhance convenience and accessibility, catering to the increasing use of mobile devices.

**6. User Feedback and Interaction:** Implementing a feedback system where users can provide their suggestions, report issues, and interact with the development team would help in continuous improvement and customization based on user needs.

# CHAPTER 9

## 9. References:

- [1] "YouTube Data API," [Online]. Available: <https://developers.google.com/youtube/v3>.
- [2] Kharwal, Aman, "All Data Visualizations in Data Science using python," 21 May 2022.
- [3] M. Bartl, "YouTube channels, uploads and views: A statistical analysis of the past 10 years," 2018.
- [4] M. L. K. A. Malik, "Research YouTube: Methods,Tools,Analytics," 2022.
- [5] "Youtube Analytics-Digital Marketing Lesson-DMI," [Online]. Available: <https://www.digitalmarketinginstitute.com/blog/youtube-analytics-digital-marketing-lesson-dmi..>
- [6] "Converting video durations to time function," [Online]. Available: <https://stackoverflow.com/questions/15596753/how-do-i-get-video-durations-with-youtube-api-version-3>.
- [7] P. Covington, J. Adams, E. Sargin, "The youtube video recommendation system," New York, 2016.
- [8] Eliganti Ramalakshmi, A Bindhu Sree Reddy,Sharvani G, "YouTube Data Analysis and Prediction of Views and Categories," 2021.
- [9] Anantharamaiah, Krishna Bhargava, "YouTube Analytics Using Google Data Studio," 2020.
- [10] S. G. Samant Saurabh, "Modelling and Statistical analysis of Youtube's educational videos :a channel owner's perspective," 2018.

