Report for Viz Lab 1

Overview

In today's world of trends and social interactions, individuals rely on their daily infusion of content from various online platforms. Working with this dataset featuring trendy videos was enjoyable. The amalgamation of these attributes guarantees a comprehensive exploration of YouTube channels, covering aspects such as popularity, financial dynamics, and growth patterns. This contributes to a holistic analysis of the digital content landscape in 2023.

The assignment is based on the youtube statistics of 2023 globally on the ranks of the videos and the views of the same, it depends mainly on the popularity factor, the dataset i have chosen here is the "Global YouTube Statistics 2023", this dataset gives a snapshot of YouTube channels worldwide. It covers essential details like rankings, subscribers, video views, etc. This data helps understand where channels are popular, how much money they might earn, and how they've evolved over time. It's a valuable resource for insights into the global YouTube landscape.

Dataset

Name of the dataset - Global YouTube Statistics 2023
Source of the dataset - Kaggle.com
URL – https://www.kaggle.com/datasets/nelgiriyewithana/global-youtube-statistics-2023

Number of attributes in the dataset - 16
Number of data points - 588

Attributes (16):

- **rank**: Position of the YouTube channel based on the number of subscribers
- **Youtuber**: Name of the YouTube channel
- **subscribers**: Number of subscribers to the channel
- **video views**: Total views across all videos on the channel
- **Title**: Title of the YouTube channel
- **uploads**: Total number of videos uploaded on the channel
- **Country**: Country where the YouTube channel originates
- **channel_type**: Type of the YouTube channel (e.g., individual, brand)
- **country_rank**: Ranking of the channel based on the number of subscribers within its country
- **channel_type_rank**: Ranking of the channel based on its type (individual or brand)
- **video_views_for_the_last_30_days**: Total video views in the last 30 days
- **lowest_yearly_earnings**: Lowest estimated yearly earnings from the channel
- **highest_yearly_earnings**: Highest estimated yearly earnings from the channel
- **subscribers_for_last_30_days**: Number of new subscribers gained in the last 30 days
- **created_year**: Year when the YouTube channel was created

- **created_month**: Month when the YouTube channel was created

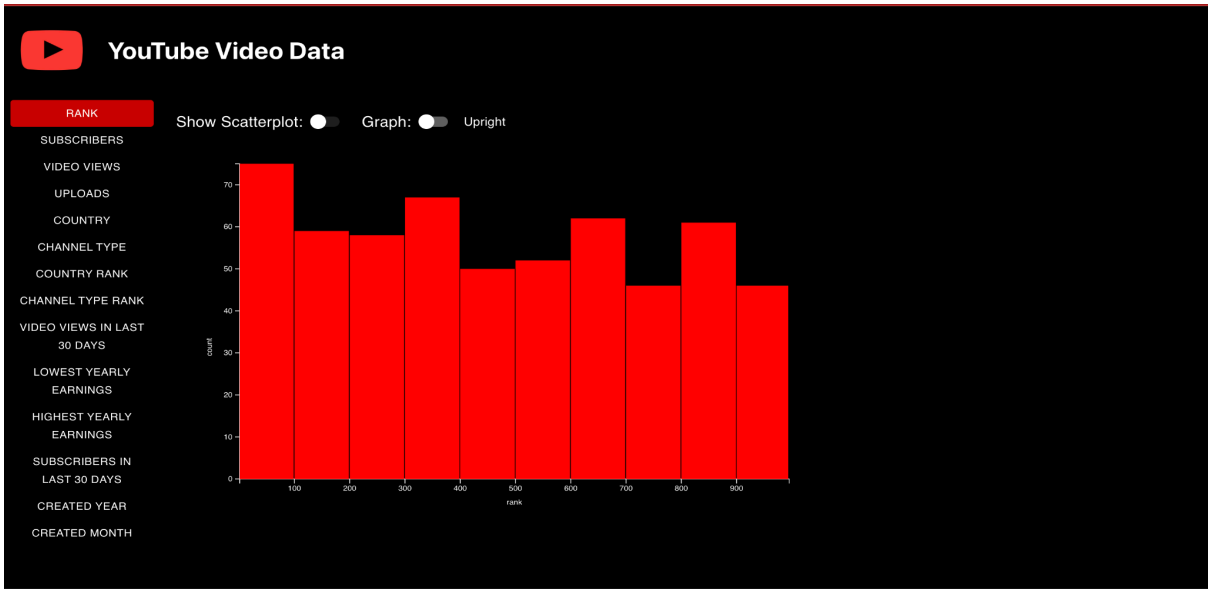
Data Cleaning

Firstly after selecting the dataset from the website, I initiated the data cleaning process, which involved handling null values represented by '0', '0.0', and NaN. I replaced these instances with NaN and subsequently removed them, effectively refining the dataset. This meticulous approach significantly reduced the data from approximately 900 data points to around 600, as the removal of missing data cells enhanced the dataset's clarity and facilitated seamless interpretation and visualization.

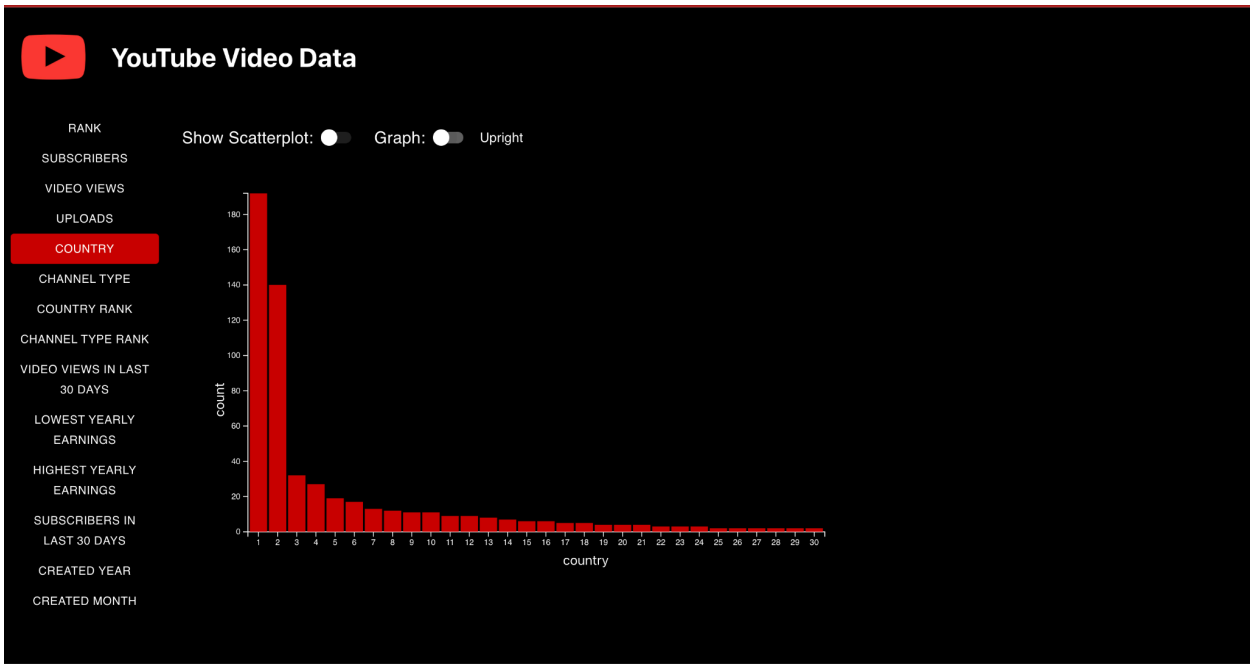
Features / Work done :

The webpage is designed with a dark theme, featuring a black background. The page includes toggles above the charts to provide users with interactive options. The first toggle allows users to change the graph structure, with the default being an upright orientation. Clicking on the toggle switches the graph structure to a sideways orientation. The second toggle is designed for displaying a scatter plot, but only activates when two attributes are selected from the menu. Upon selecting an attribute, the other attribute can be chosen by double-clicking on the desired one.

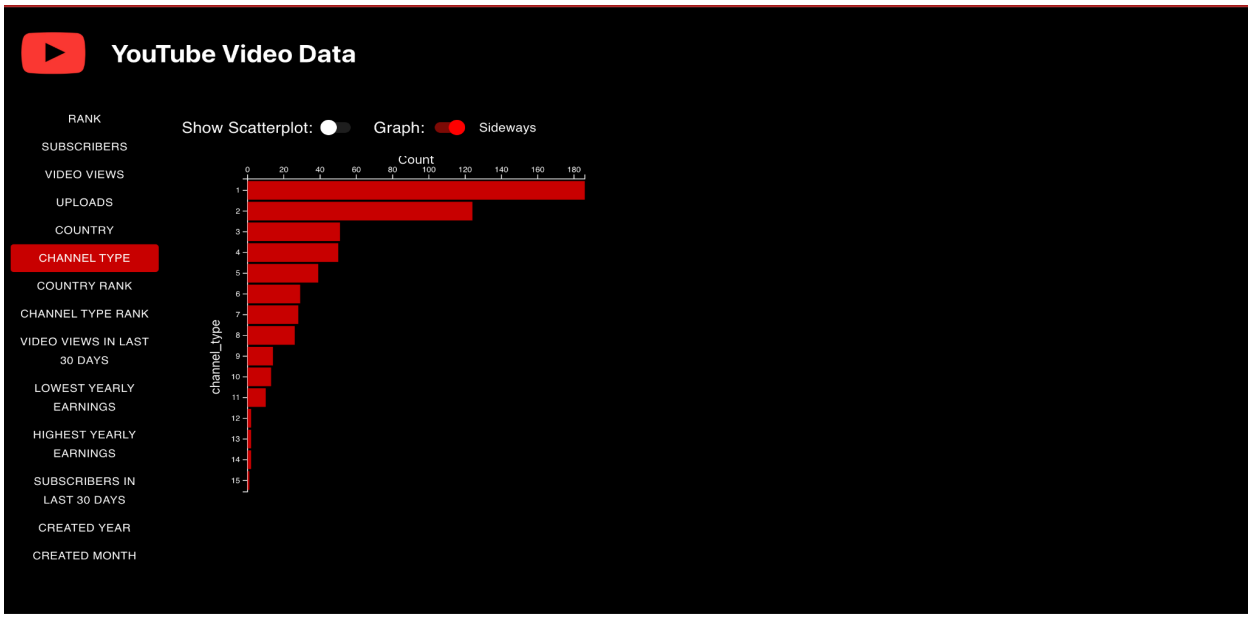
Additionally, when the scatter plot toggle is clicked, another toggle appears, offering options for selecting the x and y axes of the scatter plot. The visualizations on the webpage use bar charts for categorical data and histograms for numerical data. This design ensures an intuitive representation of data, providing a dynamic and user-friendly experience. There are two kinds of data here: categorical which is in the form of bar charts and numerical which is in the form of histograms



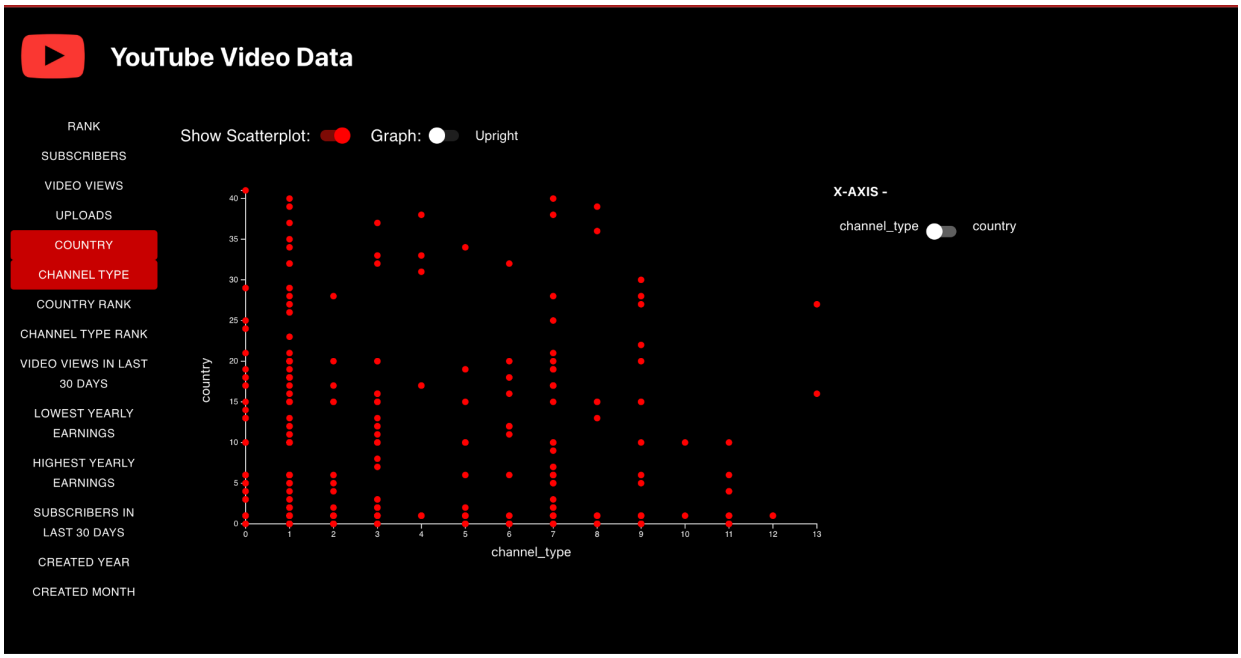
The 'rank' attribute graph visually depicts the competitive landscape of YouTube channels based on subscriber count. Lower values on the y-axis represent higher-ranked channels, showcasing their popularity and influence. This graph serves as a quick reference for understanding channel hierarchy and tracking trends in subscriber dynamics over time.



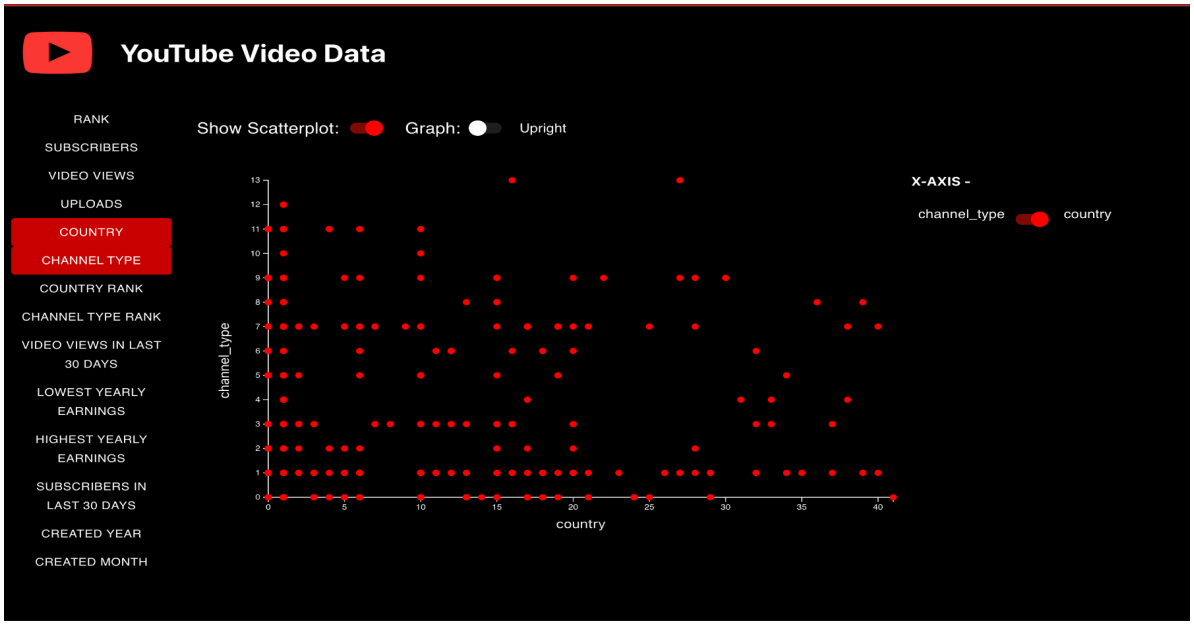
The 'Country' attribute graph visually illustrates the global distribution of YouTube channels, offering insights into regional content creation trends. It serves as a valuable tool for understanding the diversity of YouTube content and making informed decisions based on geographical insights.



The 'Channel Type' attribute graph visually categorizes YouTube channels into individual creators and brand channels. This allows for a quick comparison of their prevalence and popularity, offering valuable insights into the diverse dynamics of content creation on the platform. Once the toggle is selected, the graph is displayed in a sideways orientation.



The scatter plot, combining 'Country' and 'Channel Type' attributes, visually explores the global distribution of YouTube channels and the prevalence of individual and brand types. This dynamic representation allows for quick insights into regional preferences and categorical trends, aiding creators and marketers in making informed decisions for content strategies.



If you toggle the scatter plot axes, it means you can switch between looking at different aspects of YouTube channels. This feature allows you to easily explore and compare global distribution and channel types from various angles, providing a more customized and detailed view of the data.Once the toggle is selected, the axes are interchanged and then a graph is displayed.

There are few features which I wanted to highlight and I have mentioned the same below, which justify my dataset and attributes.

Insights into Online Influence: The dataset includes attributes like "subscribers," "video views," and "rank," providing insights into the influence and popularity of YouTube channels. Analyzing these metrics can help understand the impact of content creators on a global scale.

Global Content Creation Trends: With the "Country" attribute, users can explore the geographical distribution of YouTube channels. This allows for an examination of global content creation trends, identifying regions with a high concentration of popular channels.

Economics of Content Creation: The inclusion of earnings-related attributes ("lowest_yearly_earnings" and "highest_yearly_earnings") enables users to delve into the economic aspects of content creation. Understanding earnings ranges provides insights into the financial viability of YouTube channels.

Competitive Landscape: Rankings such as "rank," "country_rank," and "channel_type_rank" allow users to analyze the competitive landscape. Identifying top channels globally, within specific countries, and based on channel types contributes to a comprehensive understanding of the YouTube ecosystem.

Temporal Trends and Evolution: Attributes like "created_year" and "created_month" facilitate the exploration of temporal trends. Analyzing the dataset over time can reveal the evolution of YouTube, including periods of rapid growth and changes in content creation dynamics.

Comparison Across Channel Types: The "channel_type" attribute enables users to compare individual creators with brand channels. This comparison sheds light on the different strategies and performance metrics associated with various channel types.

Real-time Performance Metrics: Metrics like "video_views_for_the_last_30_days" and "subscribers_for_last_30_days" provide a snapshot of recent channel performance. These real-time metrics are crucial for understanding current trends and the responsiveness of audiences.

Subscriber and View Dynamics: Analyzing metrics related to recent performance allows for insights into subscriber and view dynamics. This understanding is valuable for creators and stakeholders looking to adapt strategies to current audience behavior.

In summary, the chosen attributes in the dataset offer a holistic view of the YouTube landscape, covering aspects of popularity, geographical distribution, economic aspects, competition,

evolution over time, and real-time performance. Researchers, analysts, and content creators can leverage this dataset to derive valuable insights into one of the largest online content platforms.

Conclusion

The analysis of the 'Global YouTube Statistics 2023' dataset offers rich insights into global YouTube channels. The interactive web page, featuring dark-themed visualizations and toggles, provides a user-friendly exploration of subscriber dynamics, content trends, and economic aspects. The dynamic toggles enhance the user experience, offering flexibility in graph orientation and scatter plot customization. This dataset serves as a valuable resource for stakeholders, providing actionable insights and a foundation for strategic decision-making in the ever-evolving landscape of online content creation.