Dataset

This report offers a thorough exploration of a dataset featuring the top 1000 YouTube channels, focusing on subscriber count as a key metric. Beyond subscribers, it delves into crucial aspects such as video views, upload frequency, country of origin, and earnings data.

Tools and Libraries

We will be using Python for this project, along with several libraries for data analysis, machine learning, and data visualization. Here are the main libraries we'll be using:

- Pandas: For data manipulation and analysis.
- Matplotlib and Seaborn: For data visualization.

Objective

The goal is to extract actionable insights to guide strategic decisions in the dynamic YouTube landscape. The report covers content strategy, regional influence, earnings analysis, and geospatial trends, aiming to decipher the success factors that distinguish top YouTube channels.

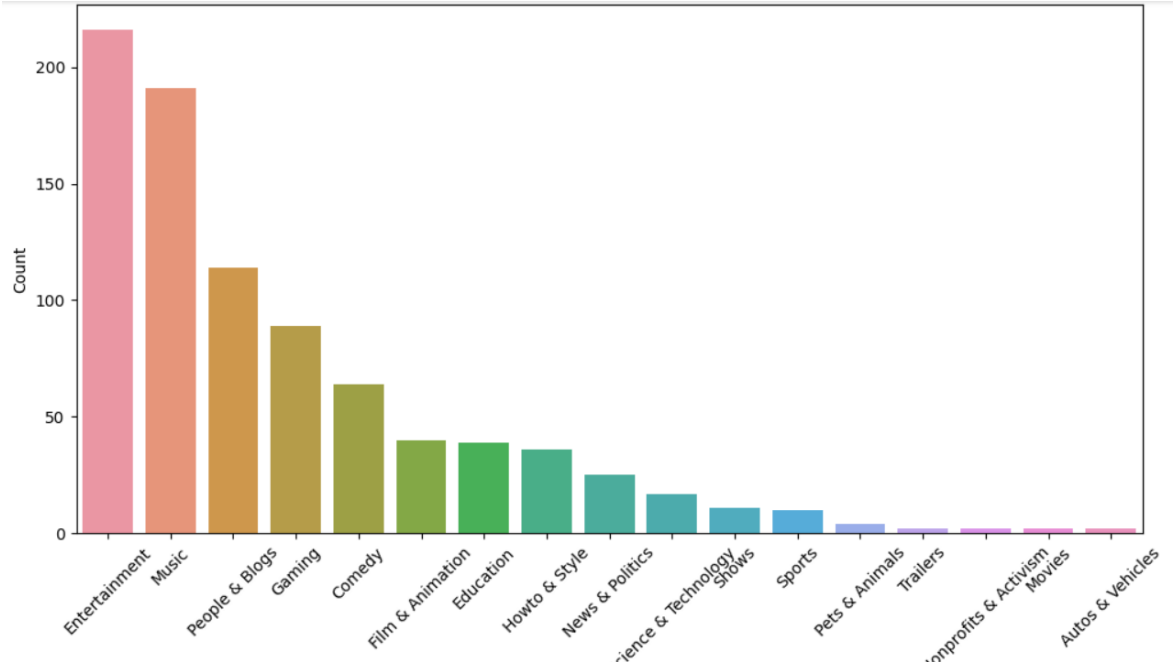
Domain Knowledge

- Youtuber: This represents the unique name of the YouTube channel for which the data is presented.
- Subscribers: Subscribers denotes the count of users or viewers who have chosen to follow or subscribe to the particular YouTube channel.
- Video Views: Video Views refers to the total count of views that the channel has amassed across all its videos.
- Category: Category signifies the specific genre, theme, or subject matter to which the YouTube channel is dedicated or caters to.
- Uploads: Uploads represent the total count of videos that the channel has published or uploaded.

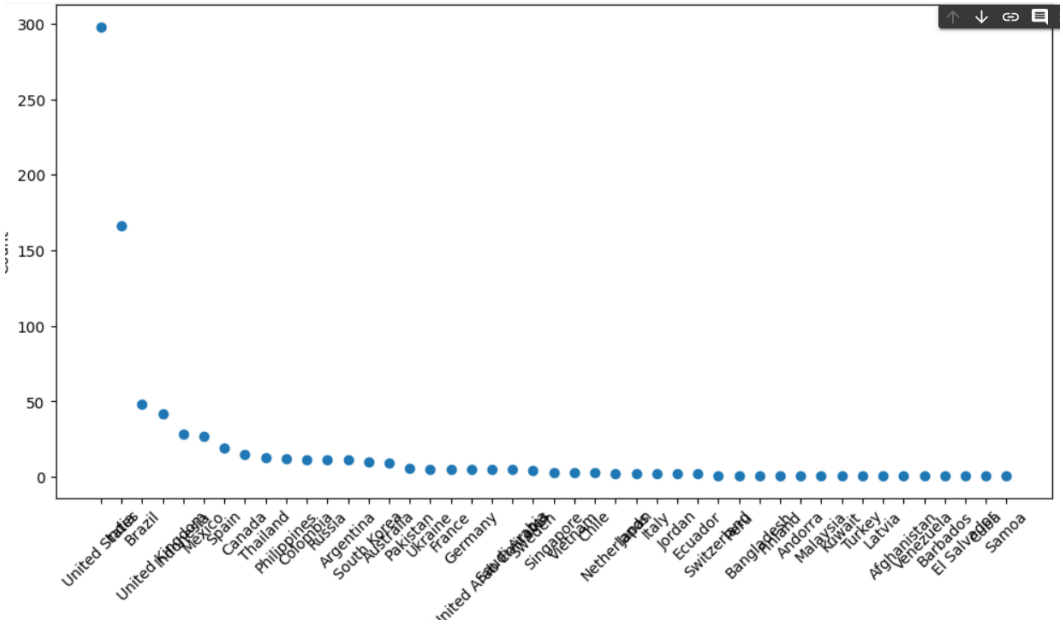
- Country: Country indicates the nation or geographical location from where the YouTube channel primarily operates or originates.
- Video Views for Last 30 Days: This metric shows the total video views that the channel has received within the last 30 days from the present date.
- Subscribers for Last 30 Days: It quantifies the number of new subscribers gained by the channel in the most recent 30-day period.
- Created Year: Created Year specifies the year in which the YouTube channel was established or created.
- Earning Range: Earning Range represents the lower and upper financial limits of the earnings generated by the channel, providing insight into the potential revenue the channel might accrue.

Univariate Analysis

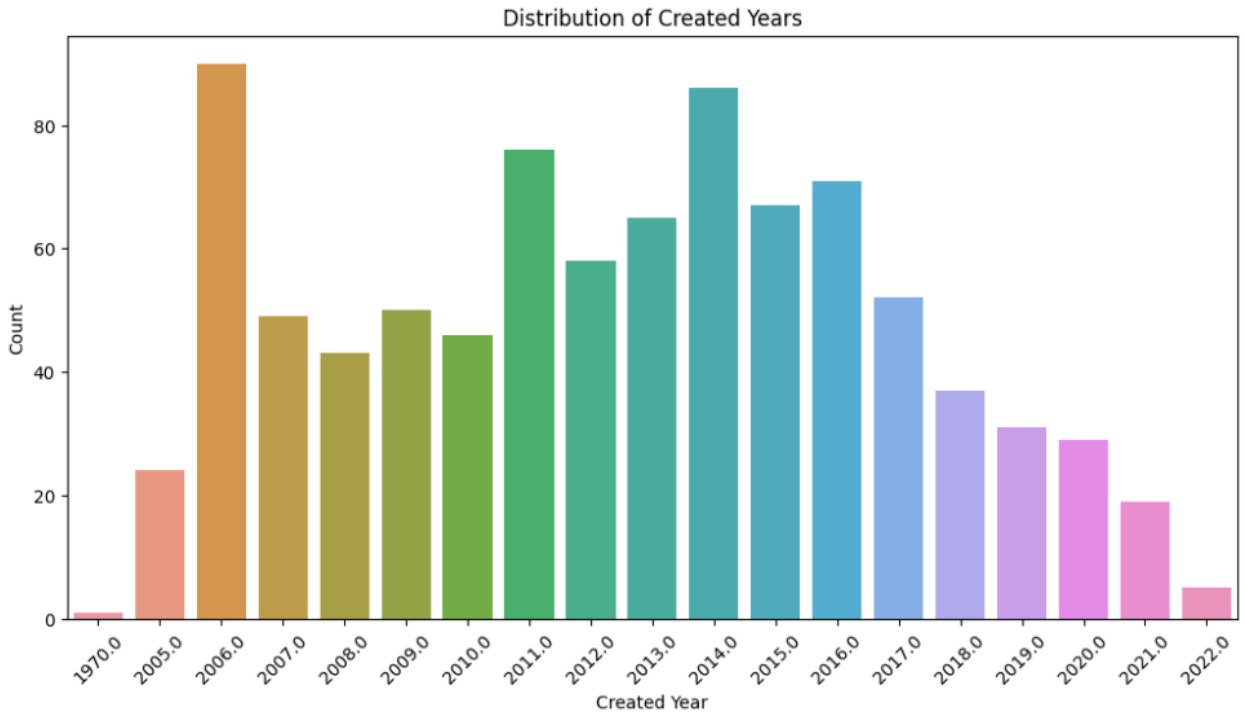
Distribution of amount of top 1000 most subscribed channels by categories :



Distribution by the country of origin



Distribution by the year created



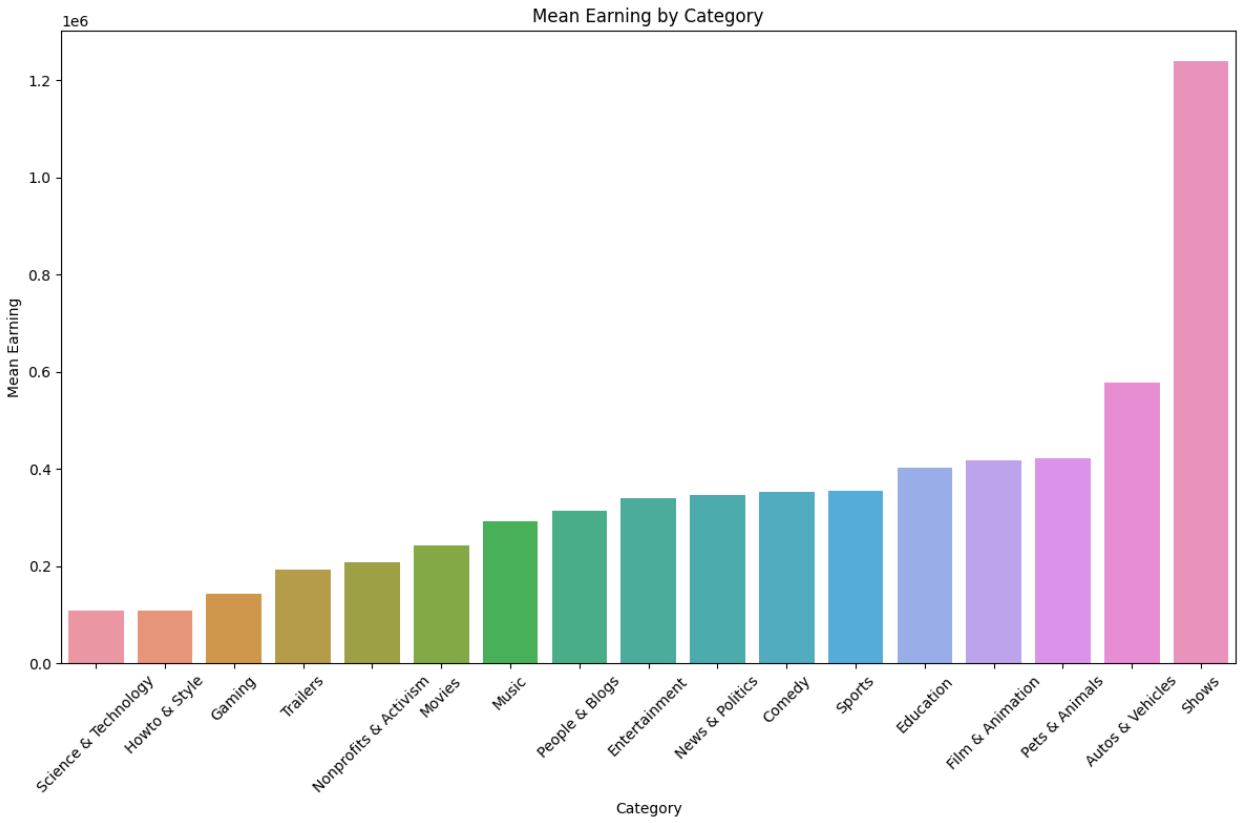
Interpretation

- Over 200 Channels in the Entertainment Genre: With a count surpassing 200, this dataset encapsulates a significant segment of YouTube channels belonging to the Entertainment category.
- Nearly 200 Channels in the Music Genre: Additionally, there are close to 200 YouTube channels categorized specifically under the enthralling Music genre.
- Majority of Channels Belong to Entertainment or Music Category: The dominant proportion of channels within this dataset aligns with either the lively Entertainment or melodious Music category.
- Channels Belonging to Different Countries:Among the top 1000 channels in the dataset, a notable distribution is observed based on their country of origin.
- The dataset showcases a diverse representation of channels, with a significant presence from various countries.
- Concentration of Top Channels: Remarkably, more than half of the top 1000 channels, a significant more than 500, are dedicated to content from the USA, India, or Brazil

- Around 300 channels hail from the United States (USA),
- Exceeding 150 channels are attributed to India, and
- Close to 50 channels represent Brazil.
- Channel Distribution Across Creation Years: The distribution of channels concerning their creation years follows a pattern closely resembling a normal distribution, indicating a typical trend in channel creation over time.
- Anomaly in 2006: A noteworthy anomaly surfaces in the distribution for the year 2006, merely a year after YouTube's inception,This anomaly in 2006 can be attributed to a surge in channel creation by corporations and entities, aligning with YouTube's burgeoning popularity and recognition during that period.
- Surge in 2013-2016: Another interesting trend unfolds in the years 2013 to 2016, revealing a significant upswing in channel creation during this period. This surge can be associated with the expanding availability of the internet in third-world countries, leading to an increase in internet users and subsequently, a higher presence of users establishing their profiles on YouTube.

Relationship Between the Mean earning and the category of the channel

Result:



Interpretation:

A comprehensive analysis of mean earnings across different genres unveiled a remarkable finding: the genre of 'Shows' exhibited the highest mean earnings, showcasing a substantial lead over all other categories.

Following 'Shows', the 'Autos and Vehicles' genre secured the second position. However, the earnings in this category were strikingly lower, amounting to less than half of the earnings witnessed in the 'Shows' genre.

The Relation of Subscribers and Views

To identify the relationship of subscribers gained in accordance to the amount of views a channel has received, we calculated the average ratio of the amount of views on average a channel has received in the past 30 days to the amount of subscribers it has gained in the past 30 days.

Result:

Average Ratio of Video Views to Subscribers for the Last 30 Days: 916.880533715767

Interpretation:

The average ratio of the amount of views a channel has received in last 30 days to the subscribers gained is 916.880533715767 i.e. roughly 917, which means roughly for every 917 views a channel receives it gains 1 subscriber.

The correlation between uploads and subscribers

Next, we proceeded to examine the correlation between the quantity of videos a channel has uploaded and the number of subscribers they have amassed**.**

Result:

Correlation between 'uploads' and 'subscribers': 0.07630549934768688

Interpretation:

A correlation coefficient of approximately 0.076 between 'uploads' and 'subscribers' suggests a very weak positive linear relationship between the number of uploads on a YouTube channel and the number of subscribers.

Therefore, based on this correlation, the number of uploads a channel has does not strongly influence the number of subscribers it has. Other factors likely play a more significant role in attracting and retaining subscribers.

The correlation between uploads and video views

An analysis was conducted to determine the correlation between the quantity of videos uploaded by a channel and the corresponding number of views the channel has garnered.

Result:

Correlation between video views and uploads: 0.41516791820367904

Interpretation:

The correlation coefficient between 'video views' and 'uploads' is approximately 0.415, indicating a moderate positive correlation between these two variables.

A positive correlation suggests that as the number of 'uploads' increases, there tends to be an increase in the 'video views'.

This suggests that channels with a higher number of uploads often experience a higher number of video views. However, the correlation is not strong enough to imply a direct causation**.**

EDA results and Discussion

1. **Genre Distribution Influence:**

The preeminent observation reveals that a significant portion of the top 100 channels fall under the genres of entertainment or music. Additionally, the subsequent four categories—people and blogs, gaming, comedy, film and animation—also represent content primarily consumer-driven. This strongly underscores the prevalence of consumer-oriented content in the YouTube landscape. Conversely, content centered around personal growth, education, and skill development appears to not resonate on a massive scale.

2. **Global Channel Distribution:**

Notably, a majority of the most prominent YouTube channels are based in the USA, closely followed by India. Despite India's substantial lead in terms of population, this ranking can be attributed to the later adoption of the internet by a significant portion of the Indian population. The USA, having embraced the internet earlier, witnessed a more rapid uptake of YouTube**.**

3. **Earnings Hierarchy by Genre:**

A discernible trend emerged, highlighting that 'shows' genre channels dominate the landscape in terms of average mean earnings. This insight suggests that, on average, the highest-earning channels on YouTube predominantly belong to the 'shows' genre.

4. **Views to Subscribers Ratio:**

The ratio of views received to subscribers gained within the past 30 days approximates 917. This indicates that, on average, for every 917 views, one new subscriber is acquired. However, it's important to note that this ratio can be influenced by a significant portion of the audience already being subscribed. Hence, definitive conclusions regarding the conversion rate from viewer to subscriber cannot be drawn from this dataset.

5. Uploads and Subscriber Correlation:

Analysis revealed a negligible correlation of approximately 0.076 between the number of uploads and the number of subscribers. This suggests that the quantity of uploads does not significantly influence the number of subscribers.

6. Uploads and Views Correlation:

A moderate positive correlation of about 0.415 was observed between the number of uploads and the number of views. This signifies that an increase in a channel's uploads is associated with a rise in its viewership. However, it's important to note that the correlation is not strong enough to indicate direct causation.

These comprehensive descriptions encapsulate the key findings from the exploratory data analysis (EDA) project, shedding light on various aspects of YouTube channel dynamics and user engagement.

Exploratory Data Analysis (EDA) on Top 1000 YouTube Channels

Aspect	Key Findings
Top Categories	**Entertainment and Music dominate the landscape, with over 200 channels in each category.**
Country Distribution	**USA, India, and Brazil account for over half of the channels.**
Channel Creation Trends	**A surge observed in 2013-2016 due to internet expansion in third-world countries.**
Earnings by Genre	**'Shows' has the highest mean earnings, followed by 'Autos and Vehicles.'**
Views to Subscribers Ratio	**Approximately 1 subscriber gained for every 917 views.**
Uploads vs Subscribers	**Weak correlation (0.076), suggesting other factors influence subscriber count.**
Uploads vs Views	**Moderate positive correlation (0.415), indicating more uploads often lead to more views.**