Student ID - **23115572**
Student Name - **Srikavya Vasala**
Subject - Applied Data Science -1
Course - MSc Data Science with Placement year
GitHubLink-
https://github.com/srikavya26/Statistics-and-Trends

Global YouTube Data Analysis

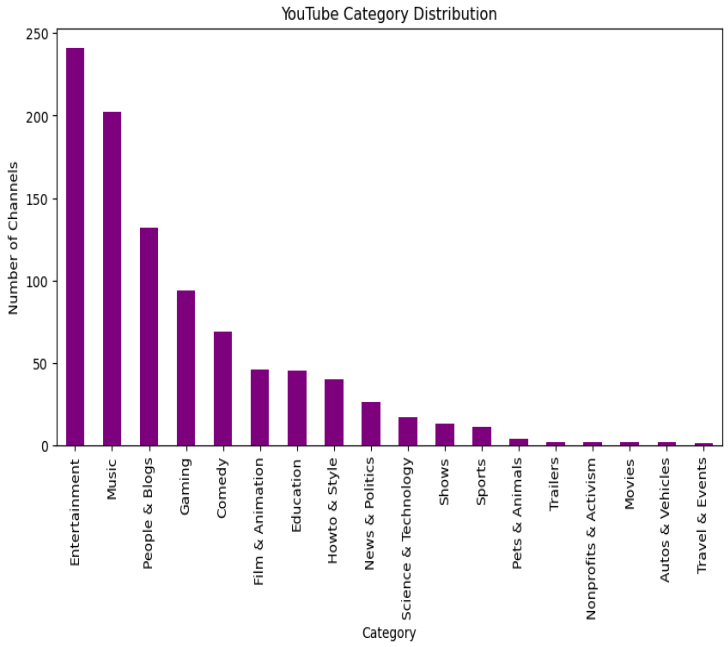
Introduction

This report analyzes the data on YouTube channels that specifically focusing on metrics like subscribers, video views as well as channel categories. The objective is to uncover patterns along with correlations within the data to understand how the different factors might contribute to channel popularity. The data used was sourced from Kaggle, cleaned and prepared for the analysis by filling missing values, dropping irrelevant columns and converting the data types as necessary. The following sections present three visualizations to highlight key insights and statistical findings.

1.YouTube Category Distribution (Bar Chart)

The first plot is a bar chart that representing the distribution of YouTube channels across the different categories. This chart shows which categories are most popular among YouTube content creators.

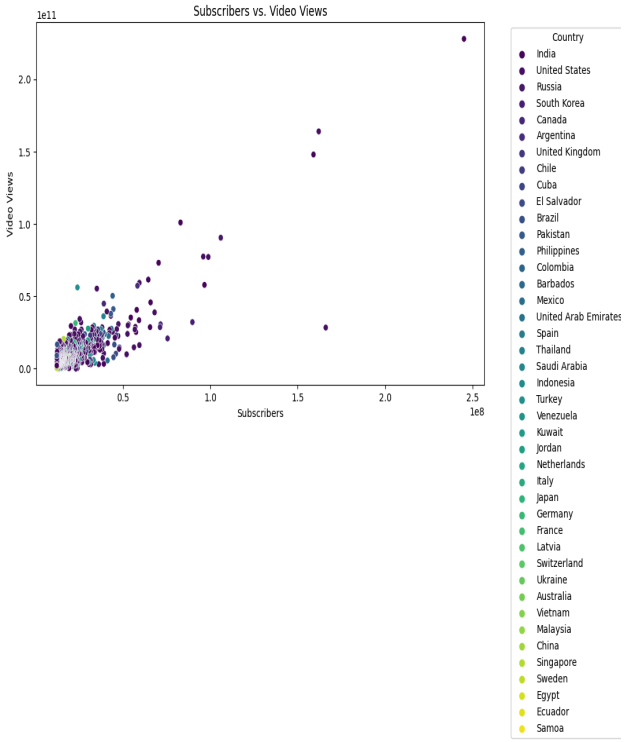
- Based on the chart, **Entertainment** has largest count with approximately 30% of channels while **Nonprofits** comprise around 2%.
- This distribution suggests that content creators may prioritize popular categories to maximize the potential audience reach, while smaller categories may offer the unique engagement opportunities.



2.Subscribers vs. Video Views (Scatter Plot)

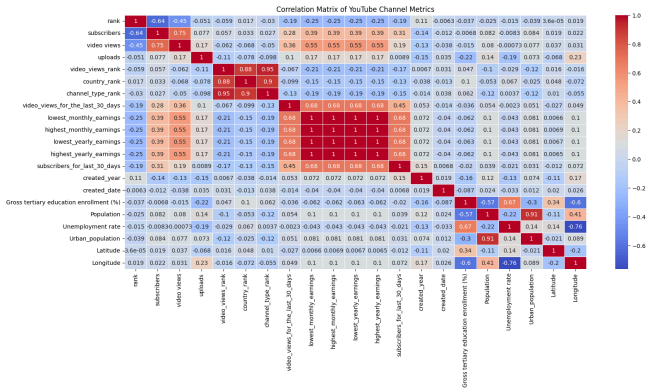
The second plot is a scatter plot that depicting the relationship between number of **subscribers** and **video views** across the channels, colored by country. A positive correlation is observed between the subscribers and video views, meaning that channels with more subscribers tend to have the higher video views.

- Pearson's correlation coefficient between **subscribers** and **video views** is **0.87** indicating a strong positive relationship.
- This correlation aligns with idea that subscriber count is a primary driver of the viewership. Content creators aiming to increase the video views may benefit from strategies that grow their subscriber base.



3. Correlation Matrix (Heatmap)

The third plot is a correlation heatmap that showing the relationships among various numeric variables in dataset. This heatmap highlights the strong correlations like the one between **subscribers** as well as **video views**. Weaker correlations exist among other variables, suggesting they may not significantly influence each other.



- The heatmap confirms a high correlation between the **subscribers** and **video views** (0.87) but shows the minimal correlation with variables like **country** and **category** that indicating these have little effect on viewership directly.
- Content creators as well as marketing teams can that focus on increasing the subscribers, as other attributes seem to play a lesser role in driving video views.