Youtube Data Analysis

**Introduction**

In the era of digital transformation, platforms like YouTube have emerged not just as entertainment hubs, but as powerful engines of content-driven income. With millions of content creators contributing daily, YouTube has grown into a global economy of its own. This project focuses on analyzing YouTube-related data, specifically examining individual YouTubers — their names, earnings, and countries of origin.

The analysis doesn’t stop there. To explore the broader socio-economic context, this project also incorporates country-level indicators such as unemployment rates and educational attainment levels. By combining creator-specific data with macroeconomic variables, the goal is to uncover patterns and insights that link digital content creation with national economic and educational landscapes.

### **Abstract**

This project explores the dynamic landscape of YouTube content creation by analyzing data on individual YouTubers — including their names, earnings, and countries of origin. To gain deeper insight into the socio-economic factors that might influence or reflect YouTube success, the dataset is further enriched with country-level indicators such as unemployment rates and educational attainment levels.

Using data analysis techniques and visualization tools, the project uncovers potential correlations and trends between digital earnings and national economic conditions. The study aims to provide a unique perspective on how socio-economic contexts might impact or align with the success of online creators, highlighting the intersection between digital platforms and real-world development indicators.

### **Tools Used**

To carry out the analysis and visualization of the YouTube dataset, the following tools and technologies were utilized:

* **Python**: Used extensively for data preprocessing and exploratory data analysis (EDA). Key libraries included:  
  + **NumPy**: For numerical operations and efficient data handling.
  + **Pandas**: For data manipulation, cleaning, and structuring.
  + **Seaborn**: For generating insightful and aesthetically appealing statistical plots.
* **Excel**: Assisted in refining the dataset by handling missing values, organizing fields, and making minor adjustments for better readability and consistency.
* **Tableau**: Used to create interactive and visually compelling dashboards that presented key insights from the dataset, making the analysis more intuitive and accessible.

### **Steps Involved in Building the Project**

The project was carried out in three major stages:

1. **Excel – Data Cleaning:** The raw dataset was first imported into Excel for basic cleaning, such as removing unnecessary columns, correcting formats, and handling missing values. This step helped in structuring the data for further analysis.
2. **Python – Exploratory Data Analysis (EDA):** Cleaned data was then analyzed in Python using libraries like Pandas, NumPy, and Seaborn. This stage focused on identifying patterns, visualizing distributions, and exploring relationships between YouTuber metrics and country-specific indicators like education and unemployment.
3. **Tableau – Dashboard Creation:** Finally, the processed data was visualized in Tableau through interactive dashboards. These visualizations highlighted key insights such as top-earning creators, popular content categories, and regional trends in viewership and digital income.

### **Conclusion**

This YouTube data analysis revealed how content creation varies across countries based on factors like income, education, and unemployment. Countries with high literacy and urban salaries—such as the U.S., India, and Mexico—showed a strong presence of YouTubers and viewers. Interestingly, countries with high unemployment also had many content creators, indicating YouTube's role as an alternative income source.

The study highlighted that entertainment, education, and kids’ content dominate in popularity and earnings. Overall, it shows how YouTube is not just a platform but a reflection of global digital trends, economic conditions, and evolving employment opportunities.