Project -03 Group-04

Global YouTube Statistics 2023

Project Write-Up

Gunel Garayeva

Fatina Hamadi

Lina Blanco

Herbert Dennis

Professor

Alexander Booth

TA

Kourt Baily

Data Analysis Bootcamp

https://hdenni3.pythonanywhere.com/

August 13, 2024

Why we choose this dataset:

   As Data-driven Students We intend to utilize the information including global statistics of the most popular Youtubers. The platform showcases prominent international Youtubers and the corresponding statistics they generate, including their number of subscribers, video views, video categories, and earnings. YouTube is widely used by many folks today, not just for gaming or entertainment purposes.

   The primary purpose of the data set is to highlight the achievements of top performers and identify the channels that are most likely to attract more views and subscriptions. This dataset is exclusively intended for the use of YouTubers to monitor the performance of their content creators on a global level, and it has not been adjusted for any other uses. As the team shared their experiences, it became evident that YouTube is also utilized globally for purposes such as teaching, research, and disseminating business information.

Focus Questions:

1. Top Channels by Subscribers and Views
2. What are the top 25 YouTube channels by the number of subscribers?
3. What are the top 25 YouTube channels by the number of views?
4. What categories will get the most views
5. Which channels have the highest total video views?
6. What are the top 25 YouTube channels by upload

Growth and Trends

Which channels gained the most subscribers? What types of content are most popular when looking on categories with the most views and highest engagement
Which channels had the highest video views?

Earnings Estimates

What are the estimated monthly and yearly earnings for the top 25 YouTube channels?
Which channels have the highest estimated earnings?

Country Comparisons

How do YouTube channels from different countries compare in terms of average subscribers and views?

What is the correlation between a country's population and the number of top-ranked YouTube channels it has?

Impact of Education and Unemployment

Is there a correlation between a country's tertiary education enrollment and the success of YouTube channels from that country?
How does the unemployment rate in a country affect the number of YouTube channels and their success?

Individual Channel Performance

What is the video upload frequency of a channel, and how does it correlate with its subscriber growth?

Channel Categories

Which category has the most channels in the top 100 by subscribers?

How do channels in different categories compare in terms of average views and earnings, by country?

Geographic Insights

What is the distribution of top YouTube channels by geographic coordinates (latitude and longitude)?

How does urban population percentage relate to the number of successful YouTube channels in a country? (TBA)

Historical Trends

What are the oldest YouTube channels in the top 100 by subscribers?

How does the creation year of a channel relate to its current subscriber count and viewership?

Country-Specific Rankings

Which country has the most YouTube channels in the top 100 globally?

Global YouTube Statistics 2023 Project Write-Up

Introduction

The Global YouTube Statistics 2023 project is an in-depth analysis aimed at understanding the trends, growth, and dynamics of YouTube channels across different countries. By leveraging data from YouTube, our project investigates various metrics such as subscriber growth, video views, content categories, and country-specific performance. The insights drawn from this project provide a comprehensive look at the global landscape of YouTube in 2023, highlighting the factors that contribute to a channel's success on this ever-evolving platform.

Project Goals

- To analyze the top-performing YouTube channels globally in 2023.
- To identify trends in content categories and their popularity.
- To examine the correlation between a country's demographics and the success of YouTube channels.
- To create interactive visualizations and dashboards that allow users to explore the data in an engaging way.

Methodology

Our project is built on a full-stack data application using Flask for the backend, SQLite for the database, and various front-end technologies such as Plotly, Leaflet.js, and D3.js for interactive visualizations. We utilized data from a Kaggle dataset, which provided detailed information on YouTube channels, including subscriber counts, video views, and more. The data was cleaned, processed, and analyzed using Python libraries such as Pandas, before being integrated into our web application.

Data Processing

- **Data Cleaning:** We removed any irrelevant or missing data points to ensure the accuracy of our analysis.
- **Data Transformation:** We performed various transformations, such as calculating averages, ranking channels, and categorizing content types.

- **Database Setup:** The processed data was stored in an SQLite database, facilitating efficient queries and data retrieval for our application.

Visualization and Analysis

- **Subscriber Growth Analysis:** We created visualizations to track which channels experienced the most significant growth in subscribers over the past year.
- **Content Category Popularity:** We analyzed and visualized which content categories received the most views and engagement.
- **Country-Specific Comparisons:** We compared YouTube performance metrics across different countries, revealing interesting correlations with demographics.

Challenges and Solutions

During this project, we encountered several challenges, including data discrepancies, managing large datasets, and integrating various technologies into a cohesive application. Through collaboration and continuous learning, our team overcame these obstacles by implementing effective data processing techniques, optimizing our database queries, and ensuring that our application remained user-friendly and responsive.

Contributor Experiences

Fatina Hamadi

Personal Experience on the Global YouTube Statistics 2023 Project

Undertaking the Global YouTube Statistics 2023 project proved to be a stimulating and demanding project. I primarily focused on analyzing specific aspects of YouTube channel performance. This included understanding which channels gained the most subscribers, identifying the most popular content types, and comparing YouTube channel metrics across different countries.

Challenges and limitations

During the project, I encountered a notable obstacle in working with the dataset's constraints. While the dataset from Kaggle contained a wealth of information on YouTube channels, there were certain gaps that posed challenges in drawing conclusive findings for some of my research inquiries. For example, the dataset did not include specific details about external factors such as collaborations or shout-outs that may have influenced sudden increases in subscriber counts. Due to this gap, I had to rely on external research and make assumptions in order to hypothesize the reasons behind the success of channels with rapid growth.

In addition, the dataset did not consistently offer up-to-date information, which introduced a level of intricacy when attempting to examine patterns or make comparisons with current metrics. These constraints frequently resulted in stress, as I had to devise strategies to account for incomplete data while maintaining the precision and significance of my analysis.

Examining the Expansion of Channels and the Appeal of Content

Although there were obstacles, researching the channels that gained the highest number of subscribers was a gratifying experience. I have identified the top 10 channels that have witnessed substantial growth in subscribers. Additionally, I have thoroughly examined the content strategies that have likely played a crucial role in their achievements. Through a thorough analysis of posting frequency, viewer engagement, and popular topics, I have obtained valuable insights into the factors that contribute to rapid subscriber growth on YouTube. In addition to analyzing the data, it was important to acknowledge the potential impact of external factors such as collaborations and viral moments on a channel's success.

One important aspect of my analysis involved determining the most popular types of content based on views and engagement. I conducted an analysis to determine the content categories that had the highest total views and engagement rates. I utilized visualizations, specifically categorized bar charts, to analyze and contrast the performance of various content

types. This allowed me to delve into the reasons behind the consistent high levels of engagement observed in certain categories, such as entertainment or gaming. It was intriguing to observe trends in content preferences among various demographics and regions. However, the analysis was limited in depth due to the absence of detailed demographic data in the dataset.

Examining Variations Across Regions

The project also included fascinating regional comparisons. I examined the comparative performance of YouTube channels from various countries in terms of average subscriber count and views. Analyzing this aspect posed a significant challenge because the dataset had limitations in capturing all the factors that could potentially impact these metrics, such as internet penetration or cultural variations in content consumption. Using visual aids such as boxplots and bar charts, I conducted a comparative analysis that effectively showcased significant disparities and anomalies across different countries.

I investigated the relationship between a nation's population and its quantity of highly ranked YouTube channels. In this analysis, scatter plots were used to visually represent any potential correlation between population size and channel success. The findings were intriguing yet inconclusive, indicating that population size is just one factor among others, such as content creation culture and technological infrastructure, that are equally crucial.

In summary

In summary, this project provided me with a valuable learning opportunity to enhance and utilize various data analysis skills. Although the dataset had its limitations, it compelled me to engage in critical thinking to effectively interpret and present the data. Working through these challenges was occasionally stressful, but witnessing the final analysis come together was highly rewarding. I look forward to further exploring data-driven insights in future projects and expanding upon the skills I have acquired through this experience.

Analysis:

Based on our analysis we created the following graph, which reveals top channels by country.

Total Number of YouTube Channels by Country
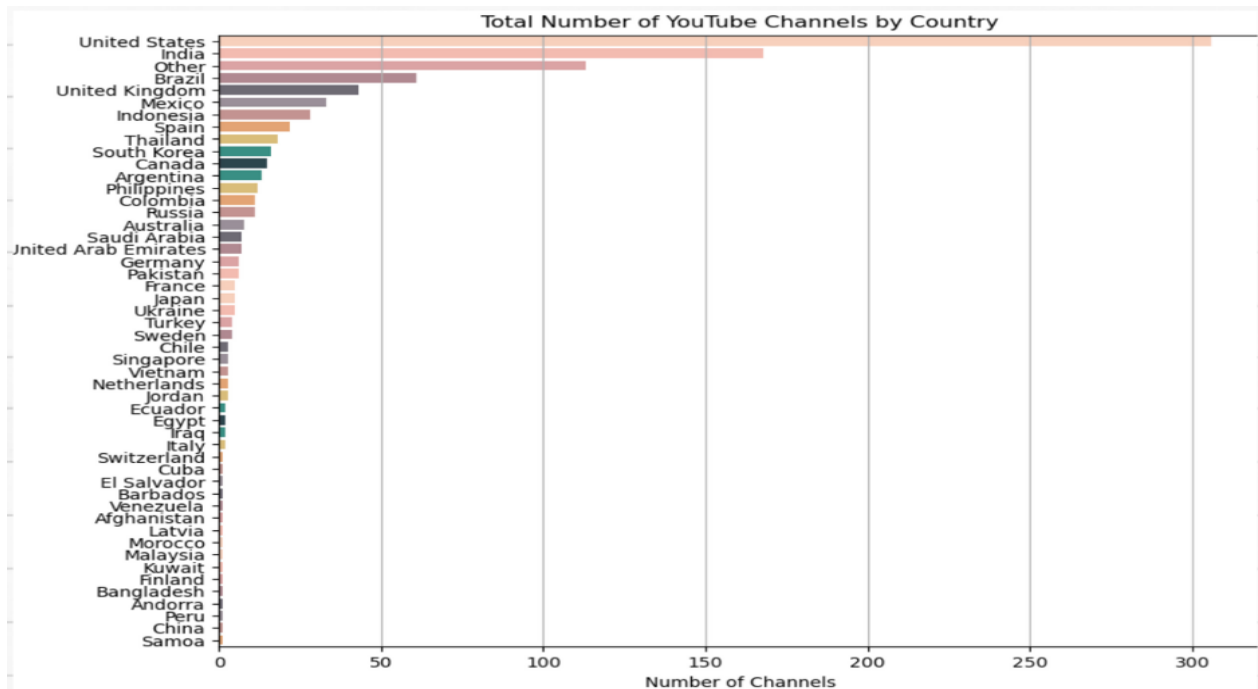
It is clear that the United States has the highest count of YouTubers (306), which is significantly higher than any other country. The reason is the fact that the U.S. is a leading country in content creation on YouTube.

India is in second place with 168 YouTubers. This indicates that India has a strong and growing presence on YouTube, likely driven by its large population and increasing internet penetration.

Brazil (61) and the United Kingdom (43) also have notable numbers of YouTubers, indicating a strong digital presence in these countries.

Countries like Mexico, Indonesia, Spain, Thailand, South Korea, and Canada have fewer YouTubers but still represent important markets for YouTube content.

Argentina (13) and the Philippines (12) have the smallest numbers on this list, suggesting that these countries might have smaller YouTube communities or that their creators are less prominent on a global scale.

The following pie chart gives the clear visualization of YouTube channel distribution by Country

Top YouTube Channels Subscribers Distribution by Country
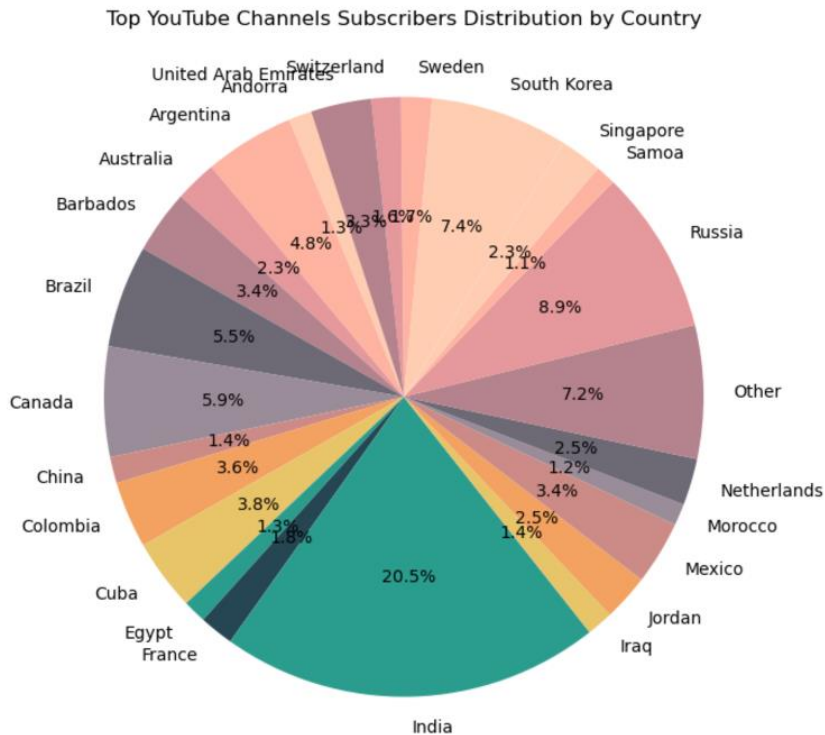
Total Views by Category and Country

The graph below lists various content categories on YouTube and their total view counts

The U.S. leads with 31.42%, making it the largest contributor to Entertainment content on YouTube. This reflects the country's dominance in producing a wide range of entertainment content, including movies, TV shows, and digital series.

India is the largest contributor of Music with 18.32%. The strong presence of Bollywood, regional films, and other entertainment industries in India drives this significant share
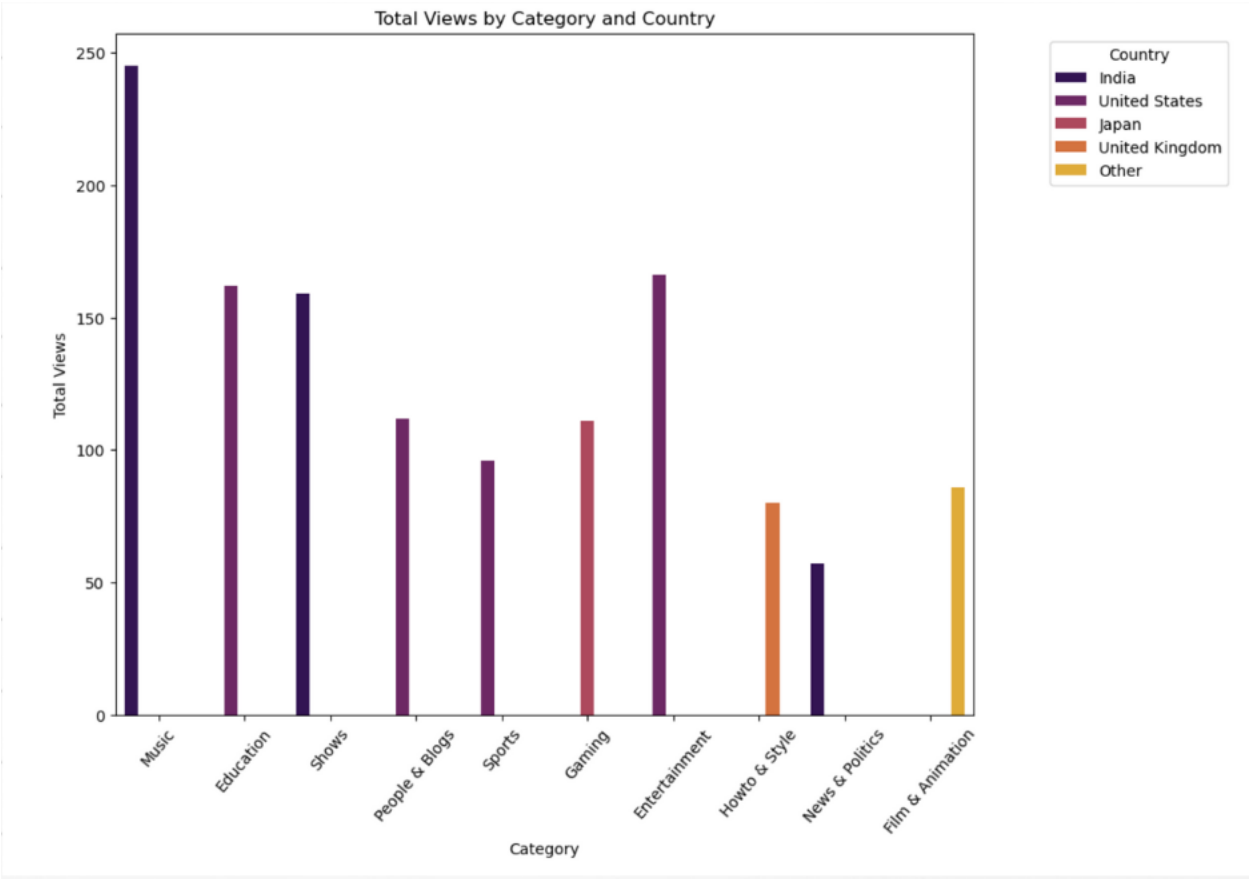
Based on this table:

North America leads with 8.12 billion subscribers, showing the region's dominance in YouTube popularity. This can be attributed to the high internet penetration and strong content creation culture.

Asia follows closely with 6.89 billion subscribers, reflecting the vast population and the rapid growth of internet users and content consumption in countries like India, Japan, and Southeast Asia.

Europe has 2.21 billion subscribers, indicating a significant presence on the platform, although smaller than Asia and North America.

South America has 1.96 billion subscribers, showing a growing but smaller YouTube community compared to the leading regions.

Australia and Oceania account for 167.5 million subscribers, reflecting a smaller population base but a vibrant digital culture.

Africa has 45.1 million subscribers, the smallest share, but this is likely to grow as internet access expands across the continent.

Top Category Channels by Year

The following bar chart reveals how the count of YouTubers in different categories has changed over the years (since 2005). It is clear that entertainment and music have significant increase first 10 years (2005-2015). Nowadays, People and Blogs category getting more YouTube audience.
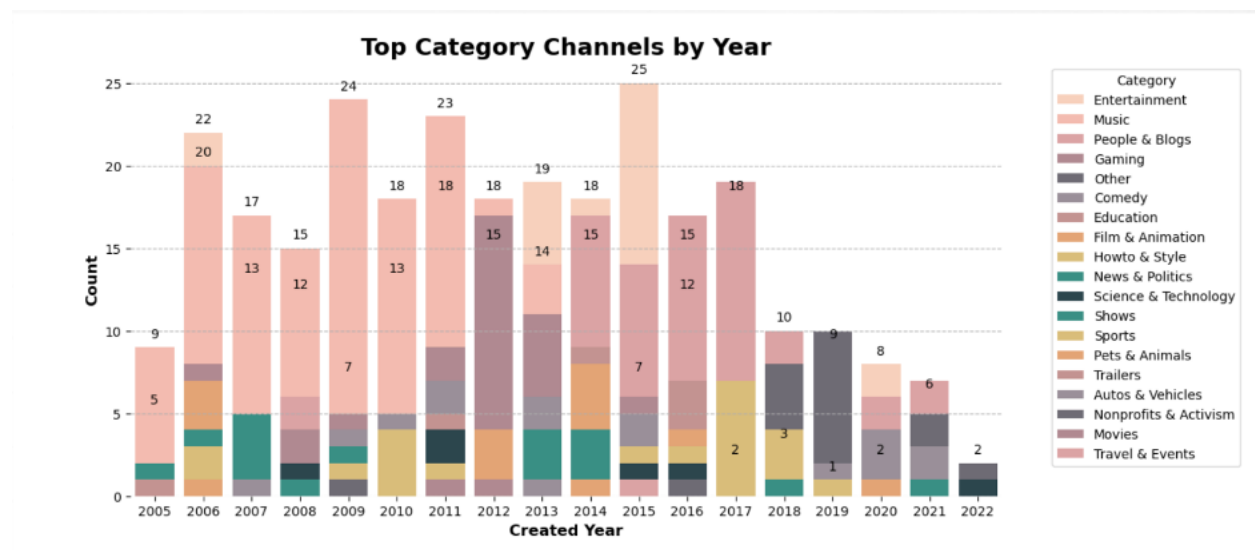
Lina Blanco:

YOUTUBE CHANNEL CATEGORIES BY SUBSCRIBERS, VIEWS, AND EARNINGS

	rank	youtuber	subscribers (million)	channel_type	category	country	total_views
0	1	T-Series	245	Music	Music	India	3103516219048
1	4	Cocomelon - Nursery Rhymes	162	Education	Education	United States	691690418531
2	5	SET India	159	Entertainment	Shows	India	432356612140
3	7	Kids Diana Show	112	Entertainment	People & Blogs	United States	1251773521276
4	12	WWE	96	Sports	Sports	United States	147986290620
5	8	PewDiePie	111	Entertainment	Gaming	Japan	706166809255
6	3	MrBeast	166	Entertainment	Entertainment	United States	2500154780534
7	17	5-Minute Crafts	80	Entertainment	Howto & Style	United Kingdom	194799530478
8	34	Aaj Tak	57	News	News & Politics	India	270391919347
9	15	Goldmines	86	Music	Film & Animation	Other	507963782337
10	92	Vlad v	37	Entertainment	Other	United States	278191674947
11	276	That Little Puff	23	Animals	Pets & Animals	United States	44866745899
12	71	Ishtar Music	41	Music	Trailers	India	33262717607
13	183	Aditya Movies	28	Film	Movies	India	15383255129
14	334	DUDU e CAROL	21	Entertainment	Autos & Vehicles	Brazil	15003458141
15	85	TEDx Talks	38	Nonprofit	Nonprofits & Activism	United States	10862911785
16	144	MR. INDIAN HACKER	31	Tech	Science & Technology	India	66976747110
17	60	whinderssonnunes	44	Comedy	Comedy	Brazil	550112774315
18	957	M	12	Entertainment	Travel & Events	Mexico	3140883140

Total Views by Category
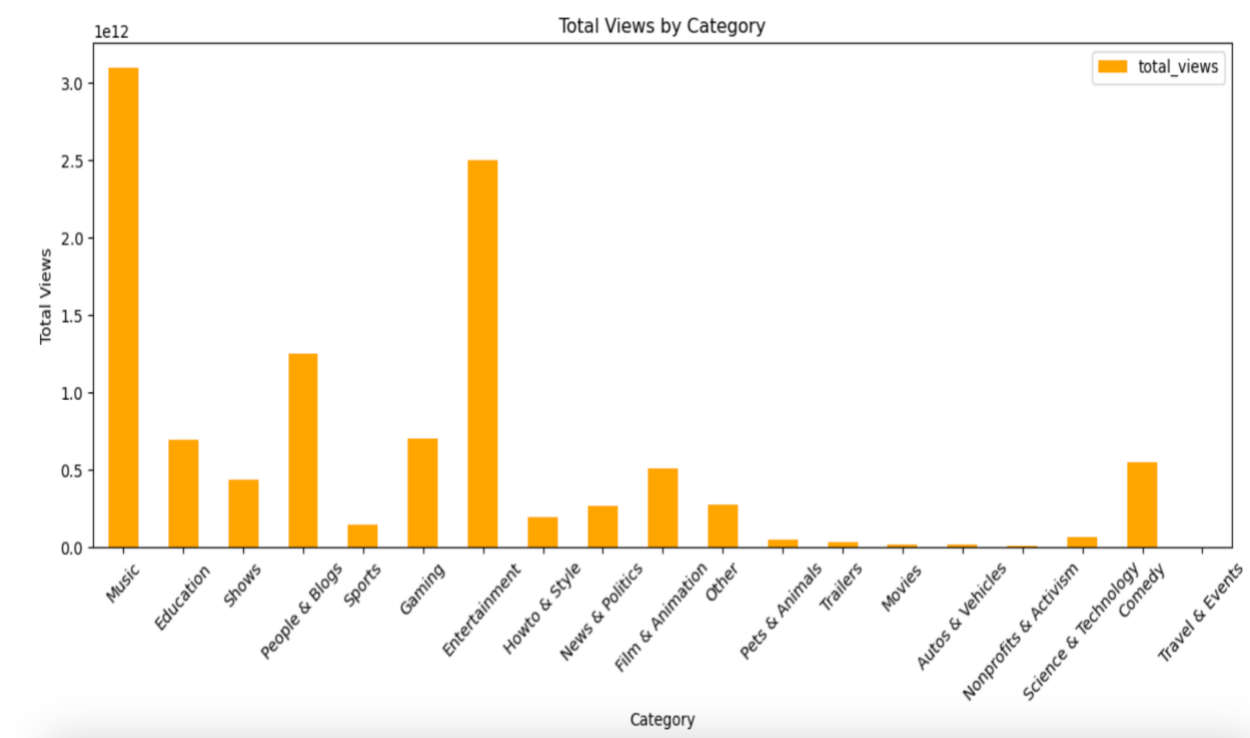
Categories vs video views:

1. Music Category:

Music channels tend to have very high views, as music videos are frequently replayed and shared. Was and is how YouTube is most known for music videos. However, the original idea has evolved a lot more for the music category, a place where artists and Youtubers can create specialized channels for all music flavors old and new.

Country: India for example (e.g., T-Series, Ishtar Music) and the United States have some of the most viewed music channels.

Earnings: This one rate High, driven by consistent viewership and strong advertising revenue.

2. Entertainment Category:

Channels in this category also boast high average views, with creators like MrBeast and PewDiePie leading in views.

Countries: Predominantly the United States, but also other countries like Japan (PewDiePie) and India (Goldmines).

Earnings: Among the highest, especially for channels that produce viral content and engage in brand partnerships.

3. People & Blogs, Education and Gaming Categories:

These categories rate a medium high rank in views.

People & Blogs and Education usually have the tendency to compete for similar content – targeting children content – for example channels like "Cocomelon - Nursery Rhymes" have significant views

Countries: The United States dominates this category. However, Gaming has majority of views in Japan.

Earnings: Moderate to high, depending on the niche and audience demographic.

4. Sports:

Sports channels like WWE maintain a strong viewership, particularly for live events and highlights. The US is a major player in this category. However, Channels containing FIFA (world cup and other soccer cups) have higher presence in countries in Europe and Latin America. This category slips by demography.

Earnings: are Moderate high, driven by advertising and sponsorships associated with major sports brands.

Curios note this category has the particularly individual youtubers with a smaller percentage but rating high on this because they provide a more provocative and instructional content compared to paid-cable sport channels.

5. Howto & Style:

Average Views rating. Channels like "5-Minute Crafts" attract substantial views with DIY content.

Countries: The United Kingdom and other English-speaking countries.

Earnings: Moderate, with earnings boosted by affiliate marketing and sponsorships.

6. News & Politics:

Low to average views. News channels, such as Aaj Tak from India, have lower average views compared to entertainment or music, but they maintain a steady, loyal audience.

Countries: India has a strong presence in this category.

Earnings: Low high, with revenue primarily from ads and subscriptions.

In Summary Music channels generally have the highest number of views, especially in countries like India and the United States. In terms of earnings, Entertainment and Music channels are at the forefront, profiting from high viewership and strong money opportunities. The Entertainment category dominates in the number of top channels by subscribers, showcasing a broad appeal across different content types.

IMPACT OF EDUCATION AND UNEMPLOYMENT

How education and unemployment rate in a country affect the number of YouTube channels and their success?

The dataset provided unemployment rate, population, urban population, and channel count by country data. Interesting connection due to the increasing number

of "influencers" performance and the higher tertiary education enrollment meaning education outside college or university.

The reasoning in the correlation with more successful YouTube channels is centered on increased digital literacy, creativity, and access to resources.

Higher unemployment rates could lead to more people starting YouTube channels as an alternative income source, but this might not directly correlate with channel success.

The data showed the United States, with a high number of channels (179) and a higher unemployment rate (14.7%). In contrast, the data showed India, despite a lower unemployment rate (5.36%), has a significant number of channels (139), potentially influenced by its large population and growing digital economy.   Lots of IT and data analysis helpful YouTube channels originate in India.

Brazil for example has a high unemployment rate (12.08%) and moderate channel count (33) might reflect the economic challenges and opportunities driving content creation in the country. Lack of resources and traffic.

Country YouTubers Performance based on Population, Unemployment Rate, and Urban Population
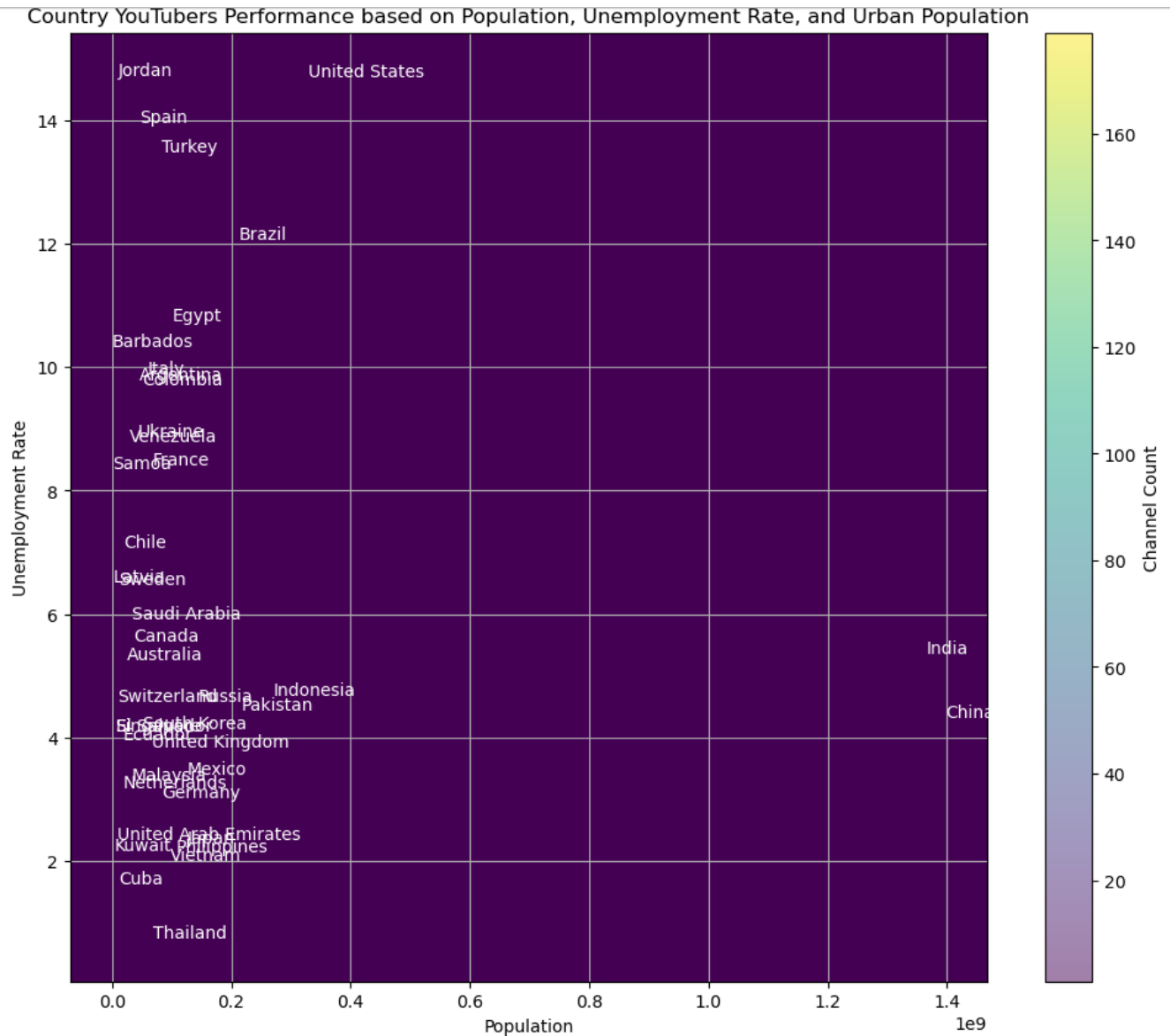
Our team was surprised by the high number of unemployment for the US and compared the 14.7% on the dataset with the percentage published by the U.S. Bureau of Labor Statistics. Which indicated a 4.1% significant difference. The presence of outliers challenges the results.

"The U.S. labor market continued to expand in 2023. Although there was an uptick in unemployment in the second half of the year, the national unemployment rate

remained below 4 percent throughout 2023. The labor force participation rate, at 62.6 percent in the fourth quarter, increased over the year. Notable labor market improvements occurred among people of prime working age (those ages 25 to 54); these improvements were most pronounced among women." [*Unemployment rate inches up during 2023, labor force participation rises : Monthly Labor Review: U.S. Bureau of Labor Statistics (bls.gov)*](https://www.bls.gov/opub/mlr/2024/article/unemployment-rate-inches-up-during-2023-labor-force-participation-rises.htm) [*https://www.bls.gov/opub/mlr/2024/article/unemployment-rate-inches-up-during-2023-labor-force-participation-rises.htm*](https://www.bls.gov/opub/mlr/2024/article/unemployment-rate-inches-up-during-2023-labor-force-participation-rises.htm)

Herbert Dennis:

**Correlation of Video Uploads and Subscribers**

The graph below displays the correlation between the number of videos a channel uploads and their subscriber count. Most of the channels in this dataset have uploaded less than 25,000 videos and many of those channels have over 25 million subscribers. As channels get into a higher upload range, such as much news based channels, the return in subscribers does not seem to be affected much.

The channel with the highest number of uploads on the dataset, with slightly over 300,000 uploads only has about 45 million subscribers compared to the top channel with 245 million subscribers and less than 45,000 uploads. As mentioned before, channels with high upload frequency are typically news, sports, and weather channels. These types of channels typically have content that constantly needs to be updated throughout the day and a lot of times are being sourced from a tv channel and are "drag and dropped" to YouTube. That can lead to a split viewer demographic. For example, if a TV news station has a YouTube channel and uploads its nightly broadcast to YouTube after airing. It is most likely only the people who did not view the live broadcast on TV would be interested in watching it on YouTube. Other channel types do not have this problem when YouTube is the only platform their content can be sourced from leading to more possible subscribers.
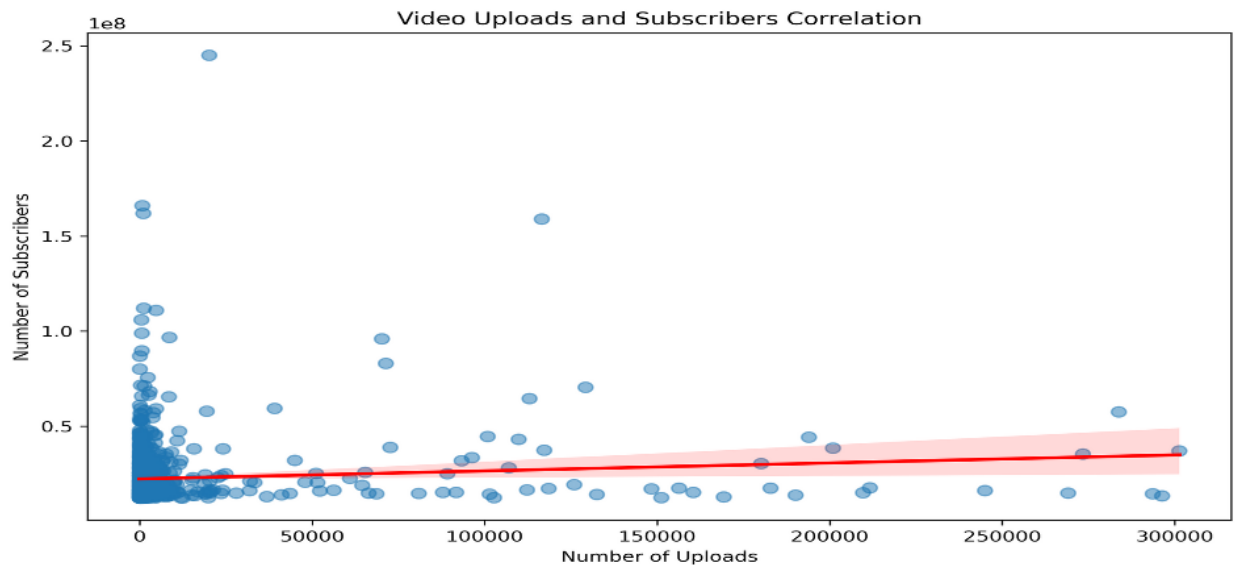
Earnings of Top YouTubers

The graph below shows how lucrative successful YouTube channels can be. Some channels generate millions of dollars per month and 10s of millions of dollars per year. The top earning YouTube channel, T-Series, is an Indian music-based channel. Music channels are some of the most popular and lucrative channels on YouTube with a combination of replay value and loyal subscribers. In 20223, T-Series made close to 60 million dollars. This should come as no surprise since the country the channel is based out of has one of the largest populations in the world. If you take a look at other channels on the graph, you can also see many other music-based channels as top revenue earners.

Average Monthly and Yearly Earnings of Top YouTubers
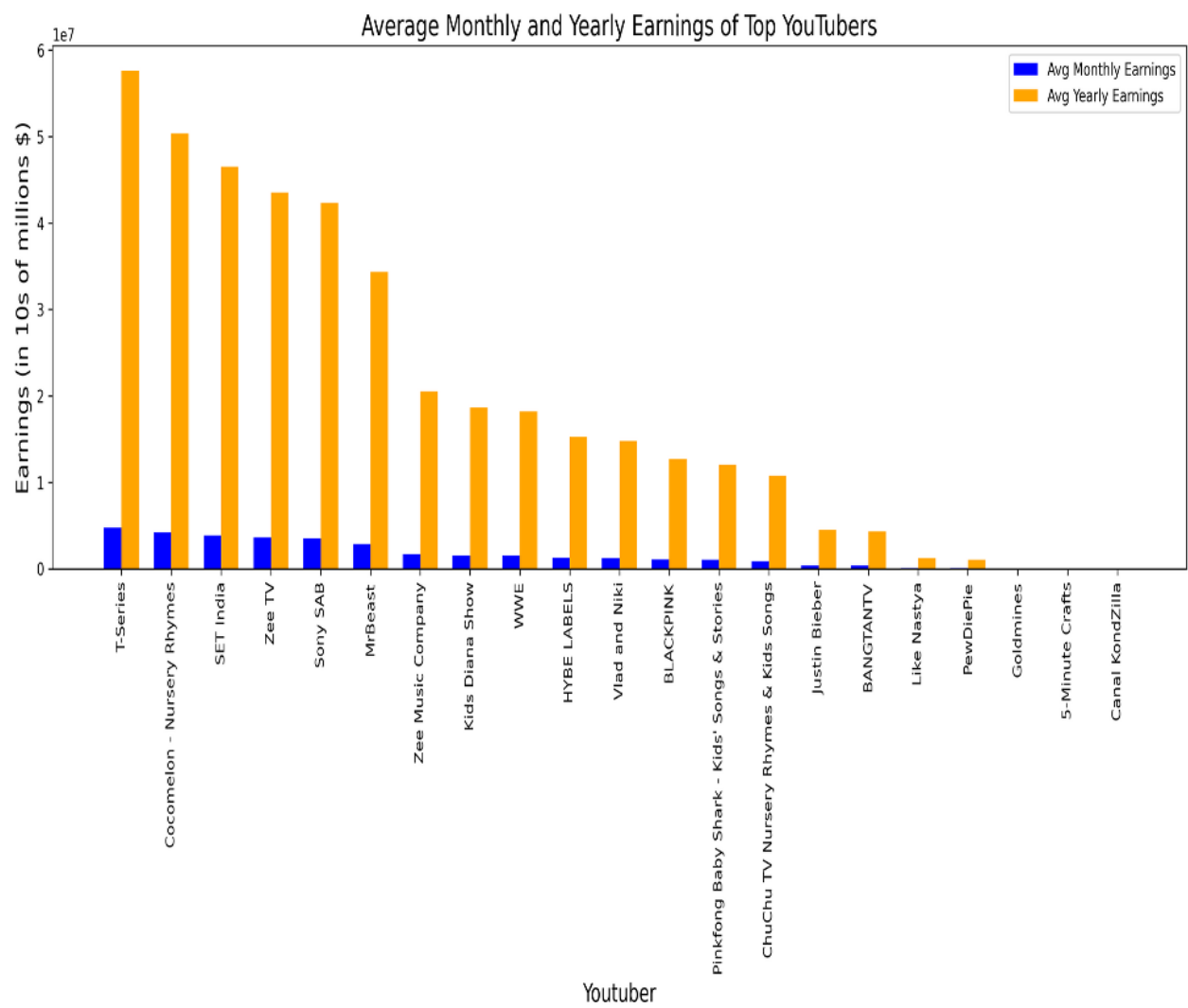
Conclusion

The Global YouTube Statistics 2023 project represents a significant learning experience for our team, combining technical skills with real-world data analysis. Through collaboration and dedication, we were able to deliver a project that not only meets our academic objectives but also provides valuable insights into the current state of YouTube globally. We hope that our work serves as a useful resource for others interested in the dynamics of online content creation and the factors that drive success on platforms like YouTube.

Data Source

YouTube Data: Kaggle Dataset - YouTube
Statistics https://www.kaggle.com/datasets/nelgiriyewithana/global-youtube-statistics-2023

References

Unemployment rate inches up during 2023, labor force participation rises : Monthly Labor Review: U.S. Bureau of Labor Statistics (bls.gov) *https://www.bls.gov/opub/mlr/2024/article/unemployment-rate-inches-up-during-2023-labor-force-participation-rises.htm*

Flask and SQLite Documentation Flask Documentation Flask. (n.d.). Flask Documentation. Retrieved from https://flask.palletsprojects.com/

SQLite Documentation SQLite. (n.d.). SQLite Documentation. Retrieved from https://sqlite.org/docs.html

Frontend Libraries and Tools Leaflet Leaflet. (n.d.). Leaflet JavaScript Library. Retrieved from https://leafletjs.com/

Plotly Plotly. (n.d.). Plotly JavaScript Library. Retrieved from https://plotly.com/javascript/

Bootstrap Grid Bootstrap. (n.d.). Bootstrap Grid System. Retrieved from https://getbootstrap.com/docs/5.0/layout/grid/

Bootswatch Bootswatch. (n.d.). Bootswatch Themes. Retrieved from https://bootswatch.com/

Highcharts.js Highcharts. (n.d.). Highcharts JavaScript Charting Library. Retrieved from https://www.highcharts.com/

Charts.js. (n.d.). Charts.js Documentation. Retrieved from https://www.chartjs.org/

D3.js. (n.d.). D3.js - Data-Driven Documents. Retrieved from https://d3js.org/

Python Tools and Notebooks Jupyter Notebooks Project Jupyter. (n.d.). Jupyter Documentation. Retrieved from https://jupyter.org/

Pandas. (n.d.). Pandas Documentation. Retrieved from https://pandas.pydata.org/

Data Visualization Best Practices Tableau. (n.d.). Data Visualization Best Practices. Retrieved from https://www.tableau.com/learn/articles/data-visualization-best-practices

Bootcamp Exercises Samples and AI: DATA-PT-EAST-APRIL-041524-MTTH-CONS Xpert Learning Assistant EdX platform DATA-PT-EAST-APRIL-041524-MTTH-CONS Lesson Practice Activity Module 15 Alexander Booth.