Names: **Aung Phyo, Chen-Kai Liu, Gebriel Gidey**

Course: **CSC 805 - Data Visualization**

Instructor: **Shahrukh Humayoun**

Date: **11/24/2023**

Phase 3 - Project Implementation

1. System architecture

The system architecture was designed to efficiently handle various functionalities and components using specific technologies and frameworks.

The backend was built using Node.js and Express.js. This makes it well-suited for handling high-traffic web applications. Express.js is a minimal and flexible Node.js framework that makes it easy to develop RESTful APIs and handle HTTP requests.

The front-end was built using React and Vite. React is a popular JavaScript library for building dynamic and interactive web interfaces. Vite is a next-generation front-end build tool that offers rapid development and optimized bundling for enhanced performance.

Data visualization was achieved with the use of Chart.js and D3, two JavaScript libraries renowned for their ability to create a variety of charts, graphs, and other visual representations. Both libraries offer a wide range of chart types, including bar charts, pie charts, and more, providing flexibility and ease of integration for visualizing processed backend data.

The dataset source was a CSV file. The backend parsed and processed the data from the CSV file, transforming it into usable JSON objects for handling and visualization on the front-end.

This architectural setup utilized the strengths of each technology to create a cohesive system that seamlessly managed data retrieval, processing, visualization, and user interaction. The system ensures a smooth and effective pipeline from backend data handling to front-end user interface presentation.

Finally, we used GCP as our cloud server. We created a Compute Engine instance and used the external IP it provided to deploy our website.

2. Dataset description

The data used in this system was obtained from Kaggle, a platform known for storing a variety of datasets. This dataset specifically focused on the top 1,000 YouTube channels, providing a vast range of information about them. The dataset included several important features for understanding and analyzing YouTube channel performance. These features included core attributes like the channel's rank among the top 1,000, the country it originated from, the number of views it received, the number of subscribers, and the monthly and annual income it generated. Additionally, it contained other relevant information, providing a comprehensive overview of these prominent YouTube channels. The dataset's wide range and depth of information which included key metrics and contextual details about the channels, made it a great resource for our analysis and to visualize the insights into the performance and trends of the YouTube channels.

3. System description

The system architecture incorporates a streamlined data retrieval pipeline between the server and front end, enabling the smooth extraction of data from a CSV file. The file serves as the dataset, containing the information to be processed. To efficiently access the desired data, we utilized the capabilities of the "*csvparser*" tool.

After retrieval, the data is parsed into JSON objects, making it easier to manipulate and utilize within the system. The majority of data processing, including computations like sorting and color assignments for visualization, is performed within the backend system. This approach ensures efficient data handling and preparation for visualization purposes. It also alleviates the computational load on the front-end system, leaving it to just the visualization task and improving load times for the client. Our backend system functions as a server, providing endpoints and APIs to facilitate data transmission to the front end, which is built with React.

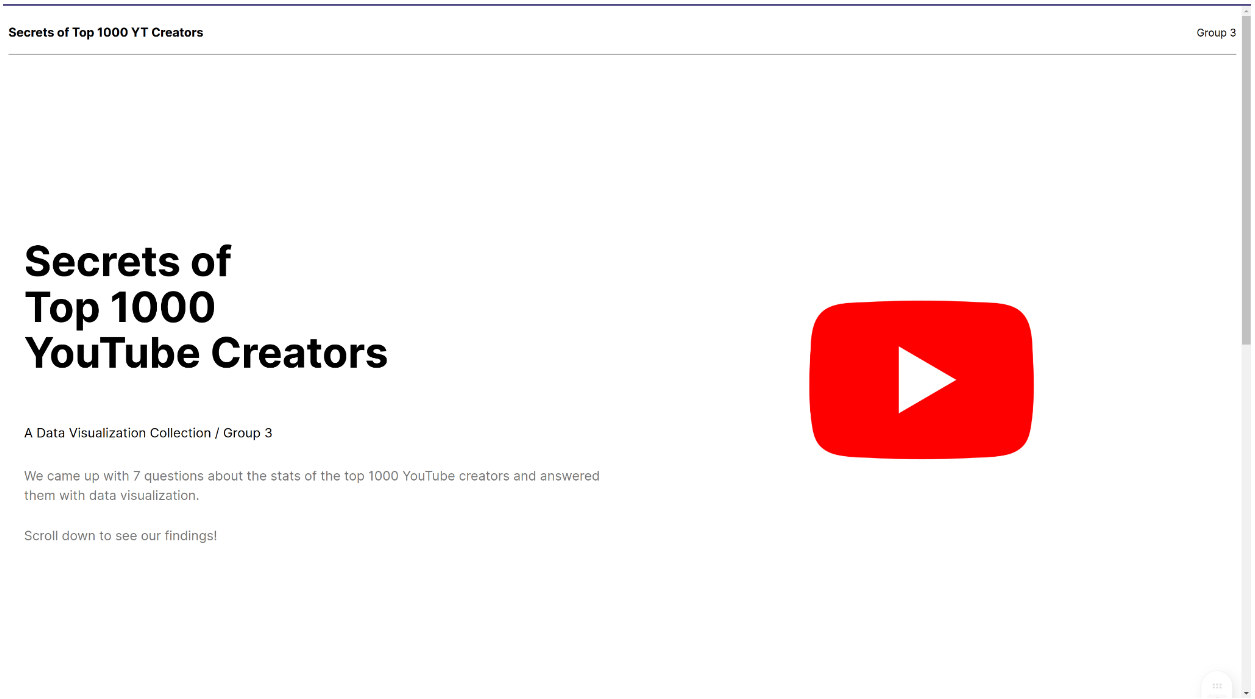
The utilization of Vite as a bundler for the frontend files significantly improved the development experience by optimizing the bundling process and hastening the development workflow to further optimize the user experience. To establish communication between the frontend and backend systems, we used "*CORS*," which enables secure cross-origin resource sharing and ensures seamless data exchange between the client-side React application and the server.

For data visualization, we used both Chart.js and D3 libraries. Chart.js facilitated the creation of the bar charts, pie charts, and maps. Similarly, D3 contributed to providing our system's visual elements with an array of visually appealing charts.

After we merged our branches to our main working branch, we cloned the repository on our GCP instance. From there we configured the firewall rules to give external IPs access to the two ports we were utilizing for our website. We were then able to successfully access our web app from any device, and we have provided the link to it further down below. Given the costly nature of running cloud instances, we had to leverage the smallest available instance and make use of resources like Vite to optimize our site.

4. Screenshots of each view and 2-3 sentences description of each screenshot

1.



Secrets of Top 1000 YT Creators

Group 3

Secrets of
Top 1000
YouTube Creators

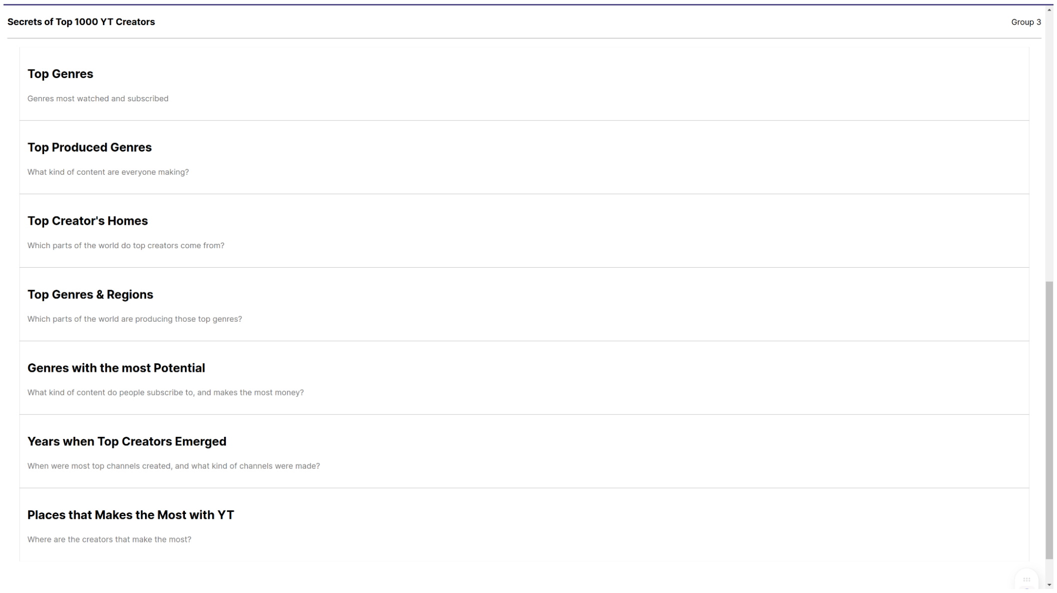
A Data Visualization Collection / Group 3

We came up with 7 questions about the stats of the top 1000 YouTube creators and answered them with data visualization.

Scroll down to see our findings!

This is the landing page of our website. We chose to settle with a clear, simple user interface to provide a smooth user experience and focus on data visualization.

2.



Secrets of Top 1000 YT Creators

Group 3

Top Genres

Genres most watched and subscribed

Top Produced Genres

What kind of content are everyone making?

Top Creator's Homes

Which parts of the world do top creators come from?

Top Genres & Regions

Which parts of the world are producing those top genres?

Genres with the most Potential

What kind of content do people subscribe to, and makes the most money?

Years when Top Creators Emerged

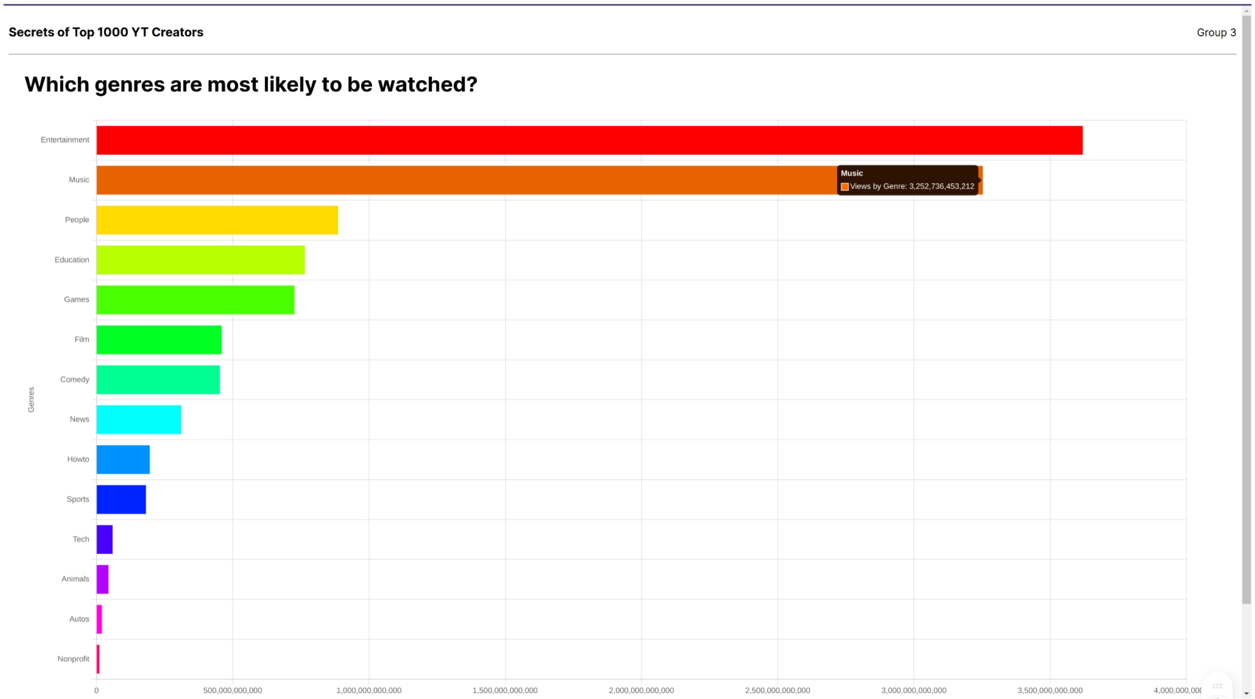
When were most top channels created, and what kind of channels were made?

Places that Makes the Most with YT

Where are the creators that make the most?

As you scroll down, you will see a list of questions that we asked and answered with the help of visualizations. We kept the amount of text to a minimum to maintain our simplistic user interface, allowing the visuals to speak for themselves.

3.



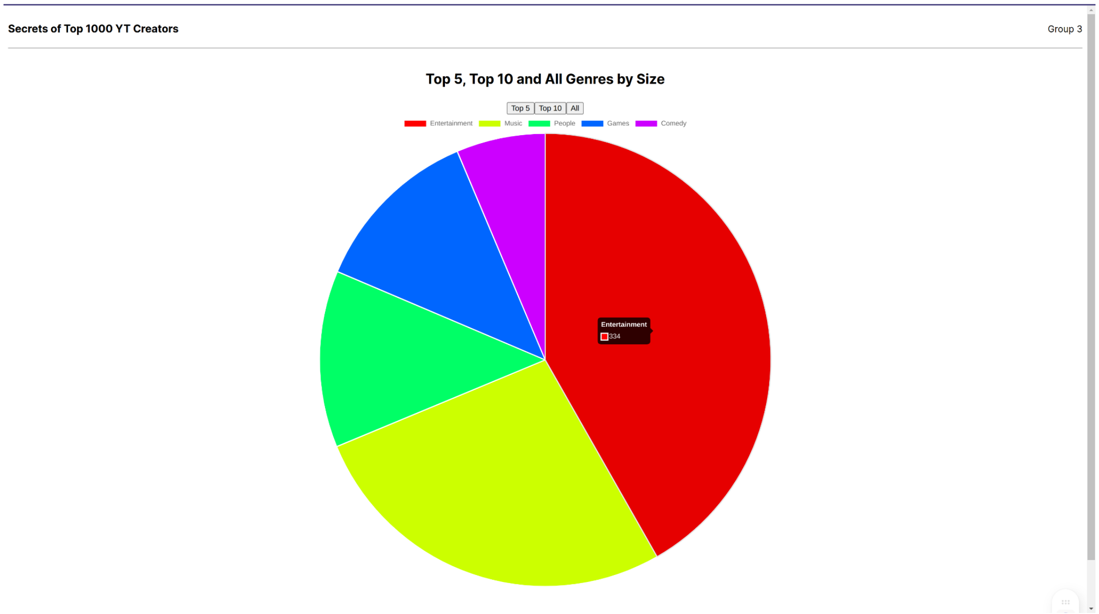
Secrets of Top 1000 YT Creators

Group 3

Which genres are most likely to be watched?

The answer to our first question is here. The number of views per genre from the 1000 largest YouTube channels is presented. Entertainment, Music, and People-oriented channels are the most viewed genres, with the first two bringing in multiple trillions of views each.

4.



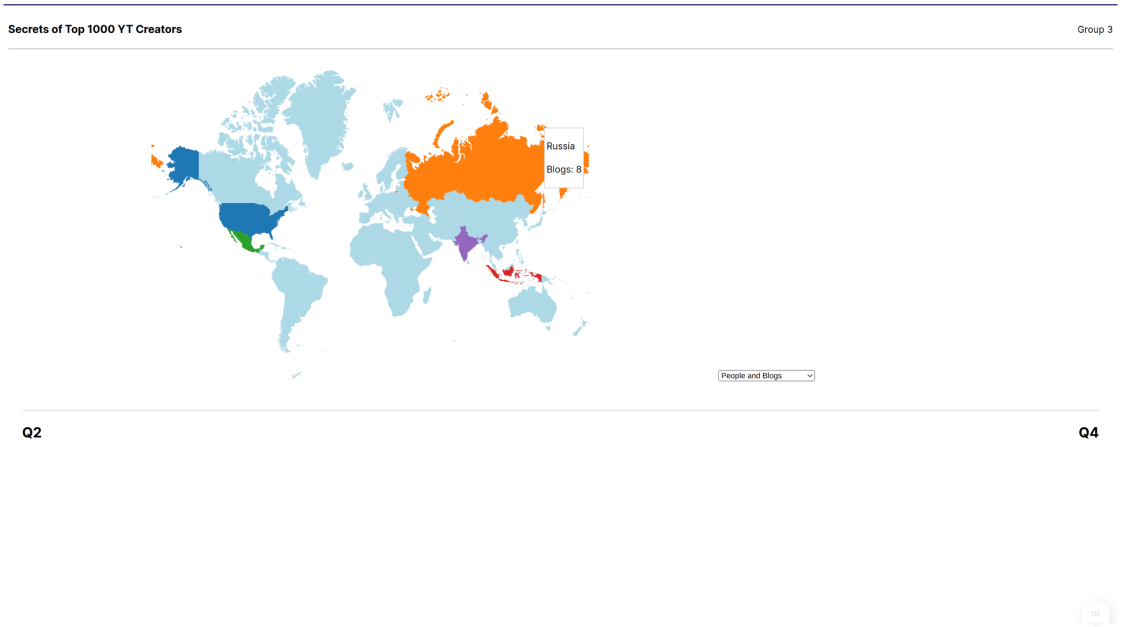
Secrets of Top 1000 YT Creators

Group 3

Top 5, Top 10 and All Genres by Size

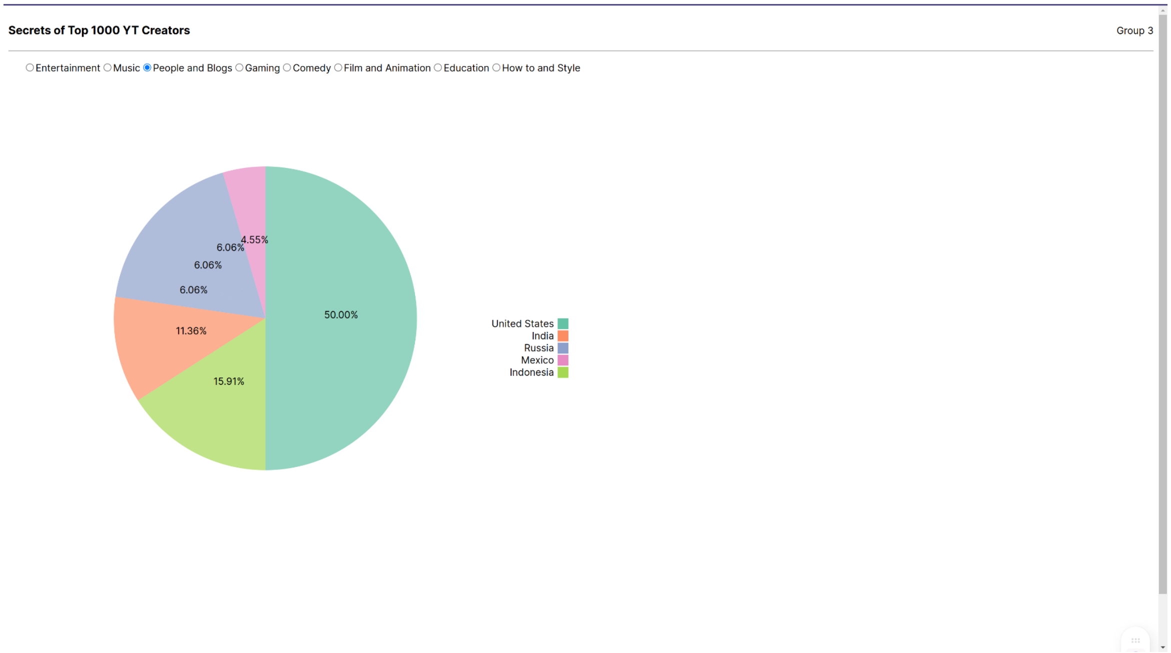
Here is the response to our second query. We can see the most popular genres in terms of how much of the largest 1,000 YouTube channels they make up. The user can choose to see the top 5, 10, or all genres.

5.



The third question is answered here. The countries of origin for the different genres can be seen. Users can choose a genre from the dropdown menu, and the graph will automatically render the countries of origin with different colors. Users can also hover over the countries to see the name of the country and the number of channels that originate there.

6.



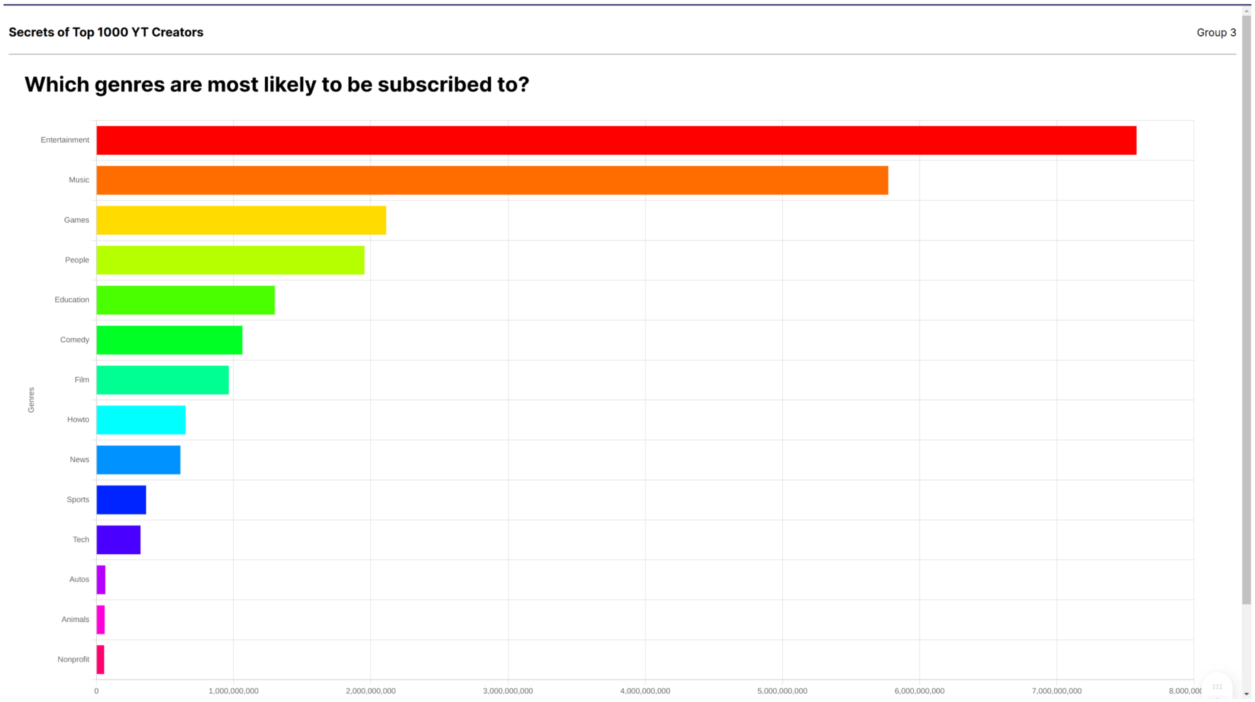
Secrets of Top 1000 YT Creators

Group 3

○ Entertainment ○ Music ● People and Blogs ○ Gaming ○ Comedy ○ Film and Animation ○ Education ○ How to and Style

The visualization below shows the countries where the largest genres originate from. We can select the genres of our choice, and the pie chart will automatically render the visuals along with how much of the chart they comprise.

7.



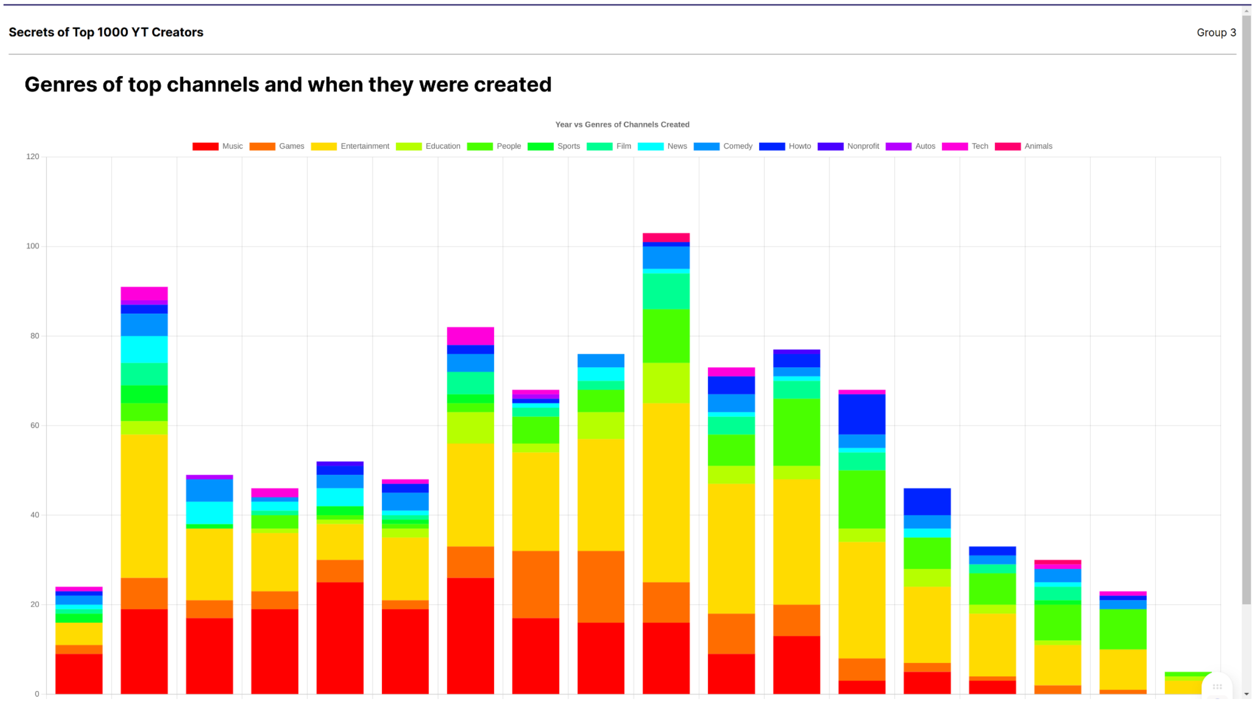
Secrets of Top 1000 YT Creators

Group 3

Which genres are most likely to be subscribed to?

This visualization answers the question of which genres are statistically the most likely to attract subscribers. Given that these are the largest YouTube channels, this chart illustrates which genres are the ones that guarantee growth.

8.



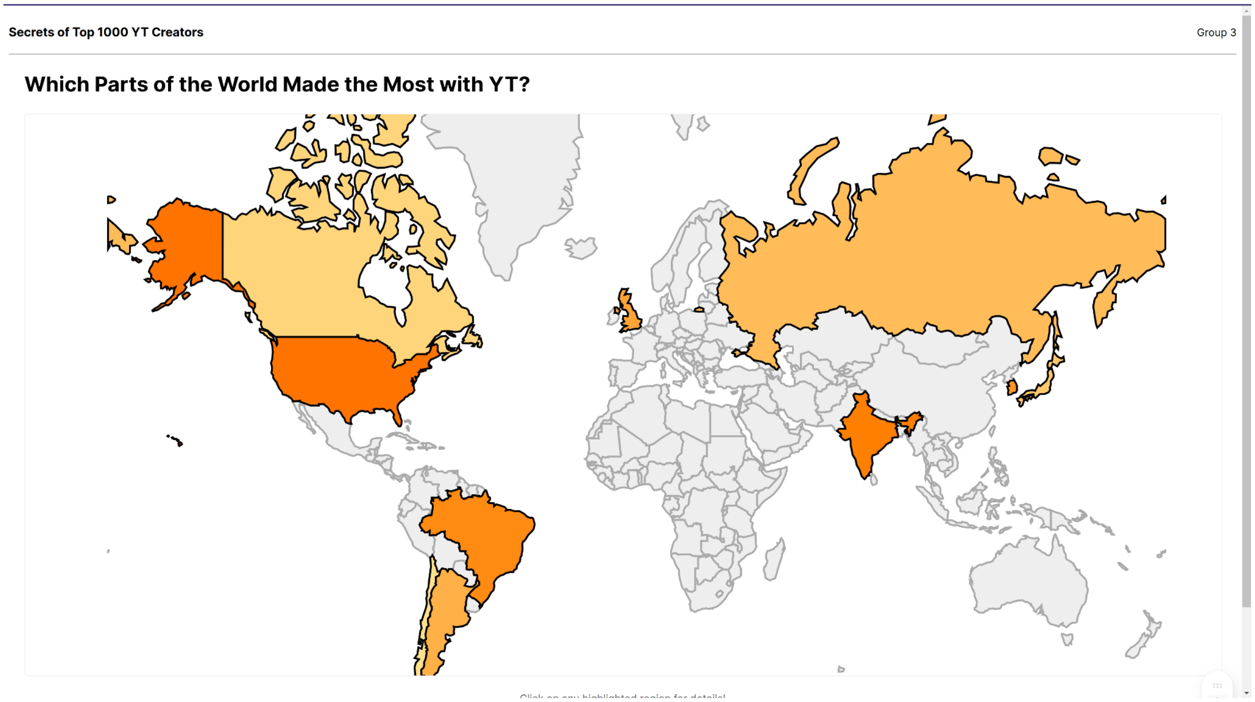
Secrets of Top 1000 YT Creators

Group 3

Genres of top channels and when they were created

This visualization provides an answer to the question of when the most successful channels were created. Given that YouTube was created in 2005/2006, we can infer that the year immediately following the platform’s creation and the years in which smartphone adoption exploded into the mass market were key periods for channel creation.

9/10.

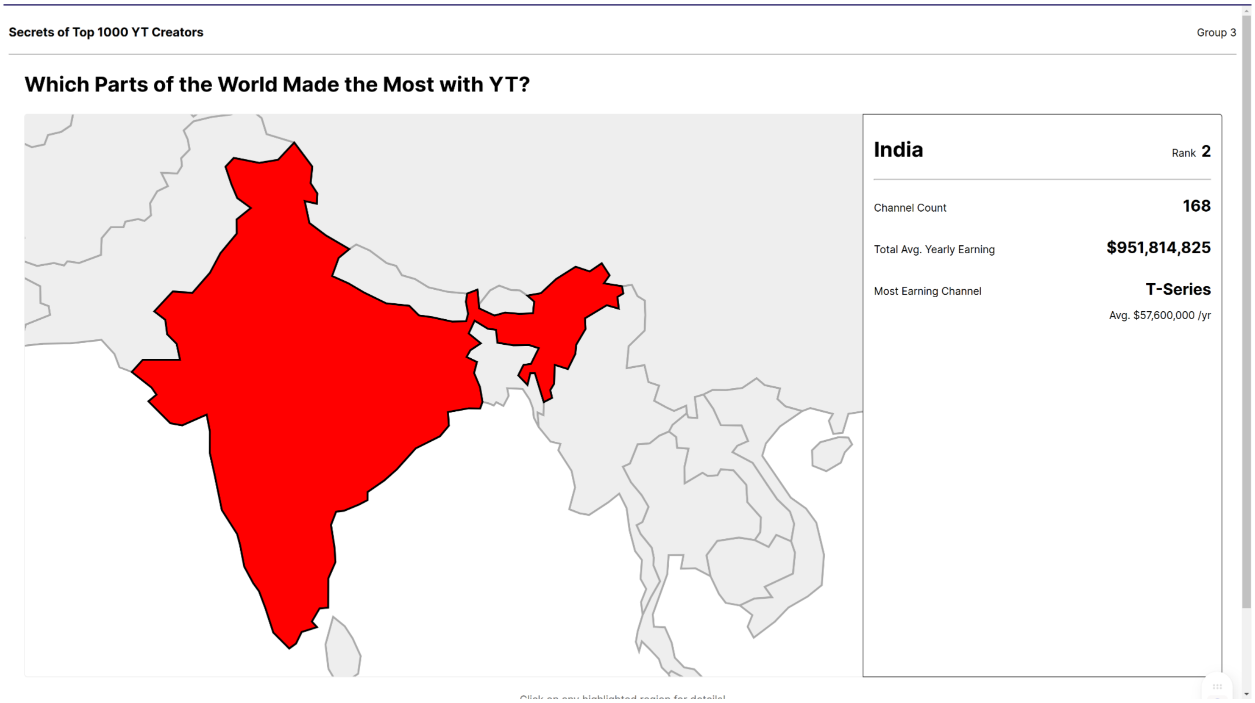


Secrets of Top 1000 YT Creators

Group 3

Which Parts of the World Made the Most with YT?

Click on any highlighted region for details!



Secrets of Top 1000 YT Creators

Group 3

Which Parts of the World Made the Most with YT?

India — Rank 2

Channel Count	168
Total Avg. Yearly Earning	$951,814,825
Most Earning Channel	T-Series Avg. $57,600,000 /yr

Click on any highlighted region for details!

To conclude, we address the number of YouTube channels in a country, the average income per channel, the top earning channel, and the country's income ranking. In this instance, India is the second highest earning country, bringing in close to one billion dollars. Additionally, T-Series is the highest-earning individual channel, bringing in an average of almost 60 million dollars, which is understandable given that T-Series has the most subscribers of any channel.

5. 2-3 minute Demo link

- **Link:** **https://www.youtube.com/watch?v=d7KD22q6oaY**

6. Link to the system source

- **GitHub Link:** **https://github.com/Gebriell/data_visualization**

- **Website Link:** **http://34.105.44.80:5173/**
- *In order to reduce costs, we had to choose the smallest server instance. Data may not render as expected. The address is bound to change should instance reboot in which case the most up-to-date version will be on GitHub.*