Project Proposal

1. Basic Information

Project title : YT Chronicles

Member details

1) Kalyan Shankar Ragam u1471323@utah.edu
2) Vijay Surya Vempati u1472324@utah.edu

Github Link - https://github.com/dataviscourse2024/group-project-yt-chronicles

2. Background and Motivation

This project is driven by my deep interest in exploring the dynamics and patterns of digital media, with a particular focus on YouTube. As a major platform for global content creation and consumption, YouTube has evolved into a hub for influencers, brands, and creators to connect with audiences worldwide. Our motivation stems from a desire to understand the underlying factors that contribute to a channel's success. By visualizing how these factors interplay to influence a channel's performance, we aim to gain insights into user engagement and content virality all over the world and in a specific country.

Given YouTube's global reach, this project provides a unique opportunity to visualize data from channels across various countries and categories. By examining these, we hope to uncover broader trends in global content consumption and identify potential correlations between channel performance and regional factors. This insight can be valuable for aspiring content creators and digital marketers seeking to understand and leverage the nuances of digital media engagement.

3. Project Objectives

The primary objectives of this project are to explore and visualize various aspects of YouTube channel metrics to gain insights into content creation and consumption patterns across different countries and categories. Specifically, the project aims to:

- **Analyze Country Variations**: Identify the top 10 YouTubers in each country based on their view counts and subscriber numbers.
- **Identify Recent Trends**: Examine recent changes in subscriber counts and video views over the past 30 days for various channels to uncover current content consumption trends.

- **Determine Popular Content Categories**: Visualize the most popular YouTube channel categories (e.g., gaming, education, lifestyle) in specific countries to gain insights into regional content preferences.

4. Data

We have obtained our data from Kaggle. The youtube dataset contains information about the data of top 1000 youtubers and contains fields like rank, id, sub count, view count, category, number of uploads, country, monthly and yearly earnings, date of creation which we believe should be good for this project. The link to the dataset is https://www.kaggle.com/datasets/nelgiriyewithana/global-youtube-statistics-2023/data

5. Data Processing

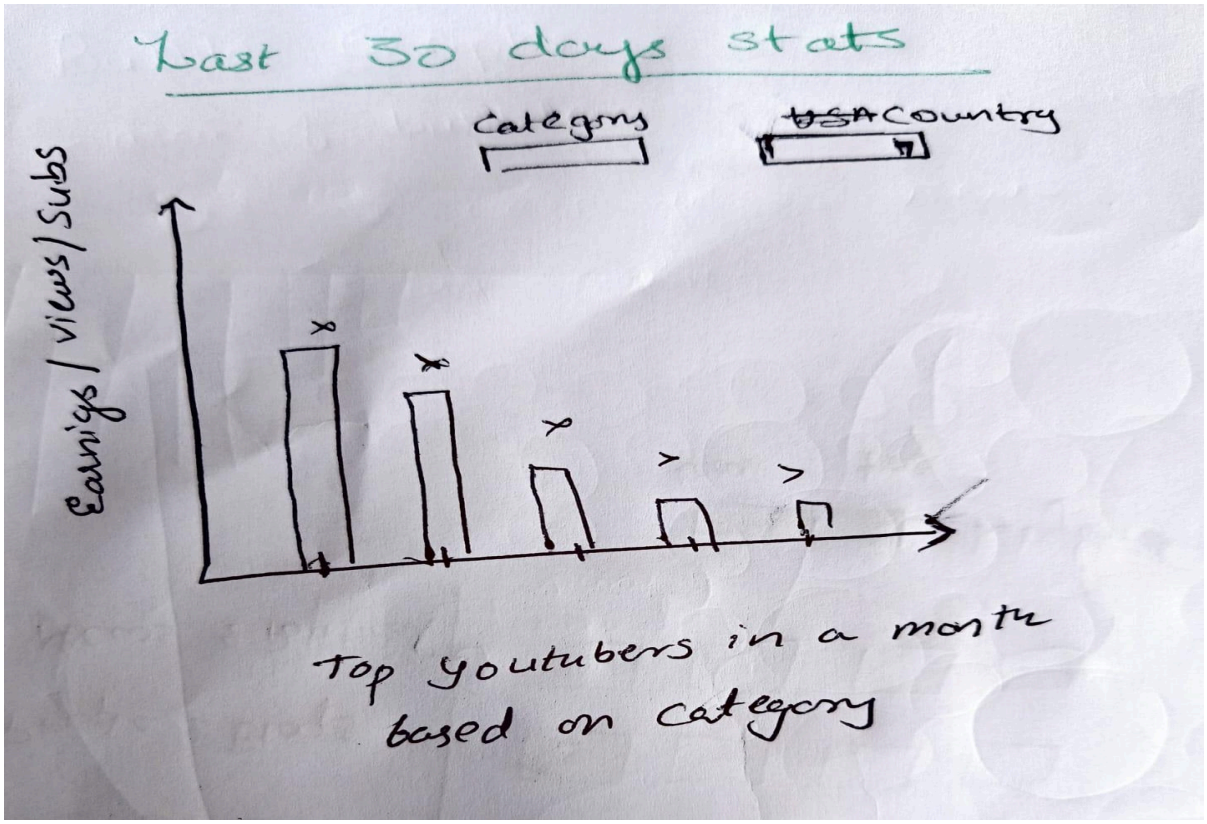
We do not anticipate extensive data processing since our initial analysis indicates that the dataset has been well-curated. While there is some irrelevant data that we plan to omit, the majority of the dataset appears to be in good shape for use. However, to ensure accuracy and thoroughness, we plan to conduct a more in-depth review before drawing final conclusions. As a precaution, we have allocated a one-week buffer for this activity.

6. Visualization Design

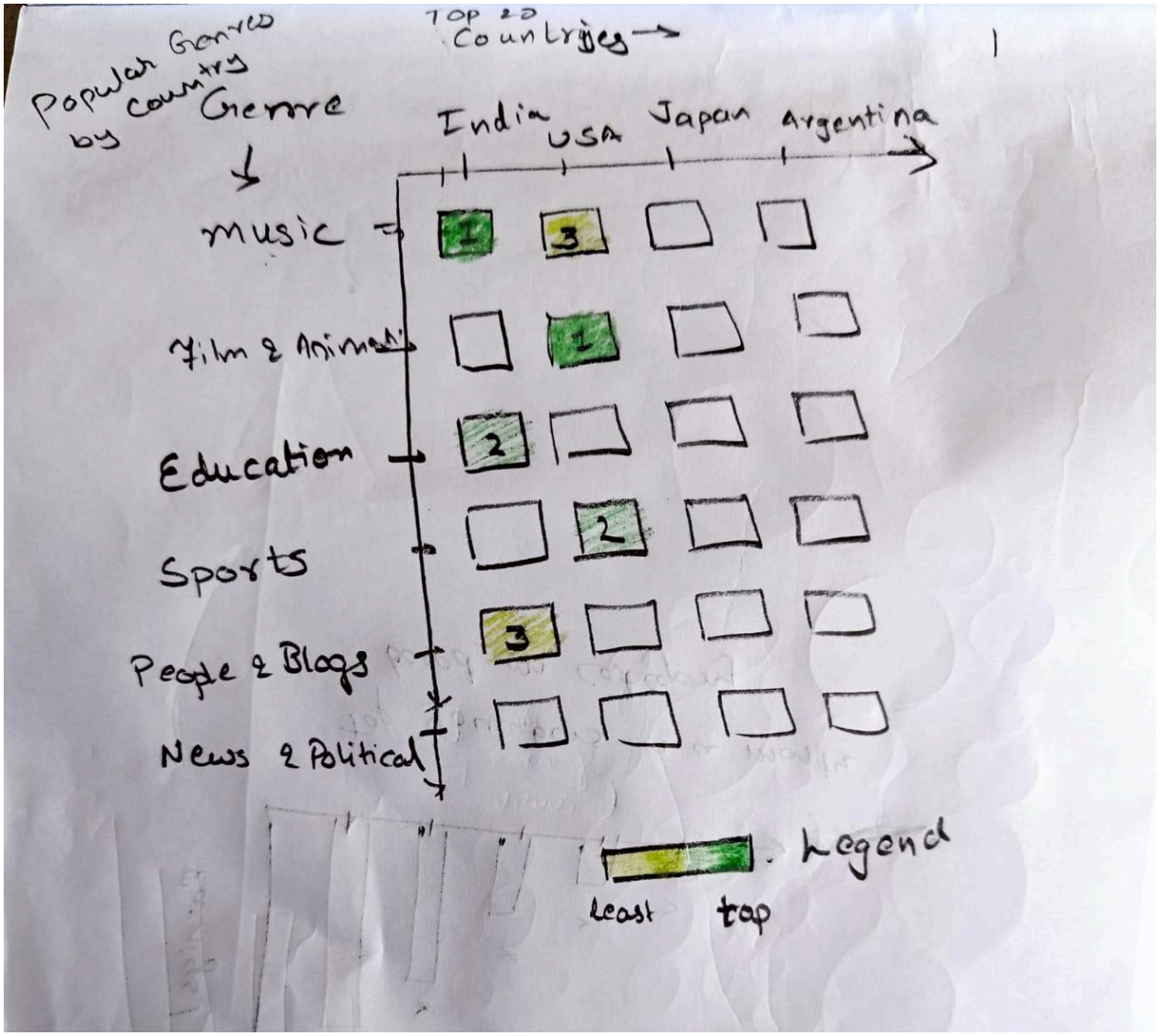
We aim to keep the page minimalistic to ensure a clean, user-friendly experience. Instead of overwhelming the user with multiple pages, our design will focus on a single, dynamic page that adapts based on user input.

1) Upon visiting the site, users will be directed to the main page, where they will be greeted with a world map and a bar graph displaying the top 10 YouTubers in the world(by default). This graph can be dynamically sorted by either views or subscriber count and by countries as well. When a user selects a specific country on the map, the graph will update to show the top 10 YouTubers from that country, with the ability to sort them by views or subscriber count, as desired.
   The world map can be used as a heatmap where the country's color intensity is proportional to the category(views/subscribers/earnings).

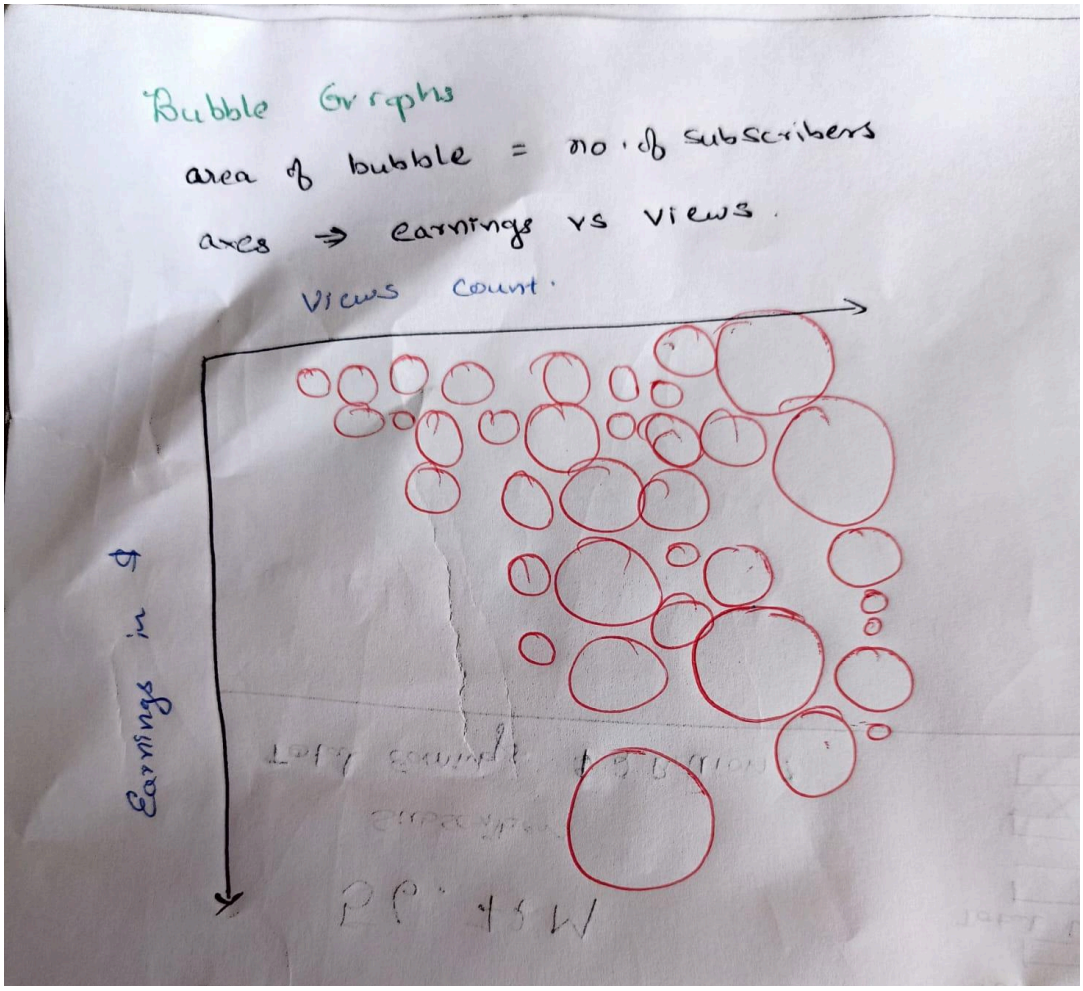
2) Below the main graph, we will display a second graph focused on recent trends, highlighting changes in subscriber counts and video views over the past 30 days. This will help uncover current content consumption trends either globally(by default) or for the selected country on the world map.



3) At the bottom of the page, we will feature a heat map that shows the popularity of different YouTube channel categories across countries. This visualization will reveal which types of content creators are most popular in specific regions.



Optional : We aim to implement a bubble graph where the size of the bubble indicates the subscriber count, axes represent Earnings vs View counts.



7. Must Have Features

- Interactive heatmaps and charts for visualizing country-specific YouTube stats.
  The world map should function as a heatmap where the color intensity reflects a selected metric (views, subscribers, or earnings).
- Filter options to select categories, countries
  By default, the bar graph displays the top 10 YouTubers globally.
  The graph should allow sorting by views or subscriber count.
  When a country is selected on the map, the graph should update to show the top 10 YouTubers from that country.
- Recent Trends Graph
  Display changes in subscriber counts and video views over the past 30 days.
  By default, show global trends, but when a country is selected, it should update to reflect trends for that specific country.
- Category Popularity Heatmap

A heatmap at the bottom of the page that visualizes the popularity of different YouTube categories (e.g., Gaming, Education, Music) across countries.

8. Optional Features

Here are some additional features we believe would enhance our project, though we recognize they may be challenging to implement within our given timeframe. Nevertheless, we will make an effort to incorporate them to further enrich the project

- We aim to create individual profiles for the top YouTubers, including their profile picture and key information displayed in a dashboard format, allowing users to quickly learn more about them.
- We want to add a feature that randomly recommends a YouTuber to watch, based on the country and category the user selects.
- We plan to leverage unused data, such as population, unemployment rate, and gross tertiary education enrollment, to find correlations with YouTube metrics and create insightful visualizations that reveal deeper patterns.

9. Project Schedule

Our tentative project schedule is as follows:

- **Week 1-2**: Data collection and cleaning to prepare the dataset for the project.
- **Week 3-4**: Design and prototype development.
- **Week 5-8**: Webpage development and implementation.
- **Week 9-10**: Testing and debugging the web page.
- **Weeks 11-12**: Final adjustments, polishing, and completing the documentation.

Throughout each phase, we will update the process logs and push changes to GitHub to track modifications and progress for reference.