

# CS 411 Detailed Project Description

*Marvin Zou, Sambuddha Biswas, Sohyun Park, Leo Lu*

## 1. Describe what data is stored in the database:

The data that is stored in the database is daily trending YouTube videos. The data includes the video titles, channel titles, publish times, tags, views, likes and dislikes, description, and comment counts for each video. It also categorizes the videos using a category ID field that varies between different regions.

The dataset was collected using the Youtube API and includes several months of videos in 11 different regions with up to 200 listed trending videos per day.

### Database Interaction:

In addition to reading the data, we plan to interact with the database in the following ways:

- Updating data: We will ensure that new data is inserted into the database by using automated scripts to pull the data from the YouTube API at regular intervals.
- Data cleaning and parsing: We will clean and preprocess the data to ensure that it is consistent and accurate.
- User input: We will provide a form for users to input their own data, such as information about their YouTube channel or video performance. This user input will be stored in the database and used for analysis.

### Data Analysis:

We plan to use SQL queries to analyze the data, including user input data. Some of the queries we will use include:

- Joining tables to compare different metrics such as views and likes.
- Filtering data by demographic variables such as age and gender to identify patterns in viewership.
- Aggregating data to identify trends and patterns over time.

### User Input:

The user input data will be used to supplement the existing dataset and provide a more comprehensive analysis. Users will be able to input data such as their video titles, descriptions, tags, and engagement metrics. This data will be stored in the database and used for analysis along with the existing dataset. The user input data will be used in the following ways:

- Comparing user input data with the existing dataset to identify patterns and trends in user performance.
- Creating personalized recommendations for users based on their input data.
- Identifying content trends and preferences specific to certain users or groups of users.

## 2. What are the basic functions of web applications?

Some basic functionalities that we would like to implement:

- View trending videos: This will provide us with trending videos and their related data so the creator can see what's trending.
- Video metrics dashboard: This would be a good graphical representation of the data using charts and graphs, showing the number of likes, dislikes, video views, etc.
- Trending video insights: This would be a function to provide in-depth insights on videos that do well, and the factors and metrics that contribute to their popularity.
- User input: This function will be used to compare the user input data with the existing dataset so that the user can identify patterns and trends.
- Live database: We plan to have a live database so that it gets updated live and the user can have access to the most up-to-date metrics.

## 3. What would be a good creative component that can improve the functionality of the application?

A good creative component would be data visualization analyzing trends of different viewership of videos over time and identifying factors that drive popularity.

We can use a variety of graphs and charts to display this information such as:

**Scatter Plots:** Scatter plots can be used to visualize the relationship between different variables, such as the number of views and the length of a video, to identify the factors that contribute to a video's popularity.

**Pie Charts:** Pie charts can be used to display the distribution of views for different video categories, providing a visual representation of the relative popularity of each category.

**Line Charts:** Line charts can be used to display the number of views for each video category over time, allowing users to see how the popularity of each category changes over time.

Other creative components besides graphs and charts would be:

**Video Embedding:** Allow users to watch the trending videos directly in the web application, instead of having to leave the page and go to YouTube.

**Trending Video Comparison:** Provide a feature that allows users to compare the popularity and engagement of two or more trending videos. This could include information such as views, likes, and dislikes over time, as well as other relevant data.

Real-time Updates: Periodically retrieve updated data on the trending videos and display it in real time.

**Video Categorization:** Organize the trending videos into categories, such as music, gaming, or sports, to make it easier for users to find the type of content they are interested in.

**4. Project Title:**

YTrend: Analysing the popularity of youtube videos

**5. Project Summary:**

Content creators are constantly looking to identify ways to increase engagement and expand their audiences. By using a dataset that stores several months of daily trending videos on YouTube, we are hoping to identify patterns and trends that exist within the trending videos on YouTube. By the end of this project, we will ideally have a website that will guide content creators in general with their decisions regarding how they are publishing and marketing the content that they produce. Creators may find recommendations for when the ideal times are to publish a video or formatting techniques to improve engagement by using the aforementioned functions described. Our goal by the end of the semester is to have a system that seamlessly recommends content creators the optimal choices when producing content.

**6. Description of an application:**

The problem that most creators have is ideas to create new content. By implementing this application, we plan to inspire up-and-coming content creators with ideas that flourish in the community through analyzing data.

Our goal is to make it easier to see trends and patterns on youtube, and make data more visual and appealing so we can burst the growth of content creators.

**7. Usefulness:**

The chosen application would be useful to future content creators looking to increase engagement with their content and acquire new viewers. By being able to recognize patterns within the trending YouTube videos that cause a certain video to be popular, creators could attempt to implement these successful patterns in their own unique manner to drive higher engagement. For instance, if our application were to discover that a certain time to publish videos often resulted in increased engagement with comment count, views, tags, and likes, this could be significantly useful to content creators. With the success and revenue of content creators being so heavily dependent on customer engagement, identifying the optimal ways to produce their videos and market them to the public is crucial to maximizing their success.

Moreover, analyzing these trends would also allow us to identify possible negative trends content creators should avoid when producing and publishing videos. Perhaps there are certain publish times that are found to be unideal or certain ways of formatting titles that

seem to minimize engagement. Avoiding these bad practices could also assist content creators in maximizing the viewership and revenue they receive.

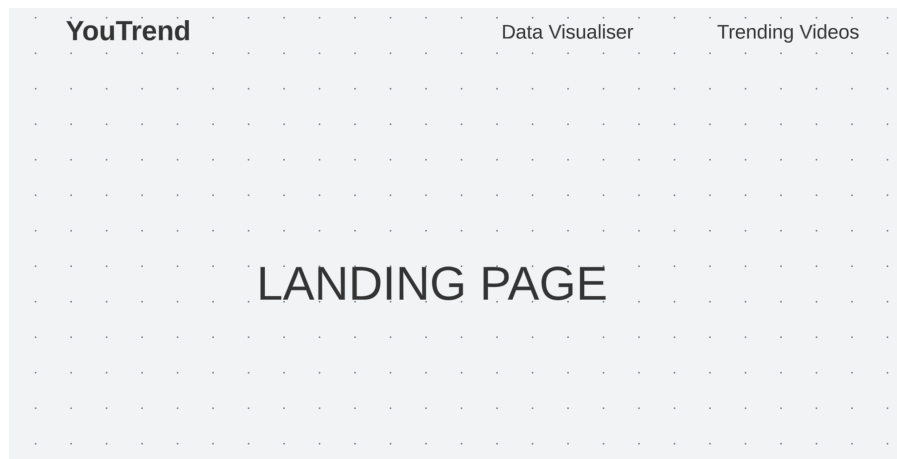
8. **Realness:**

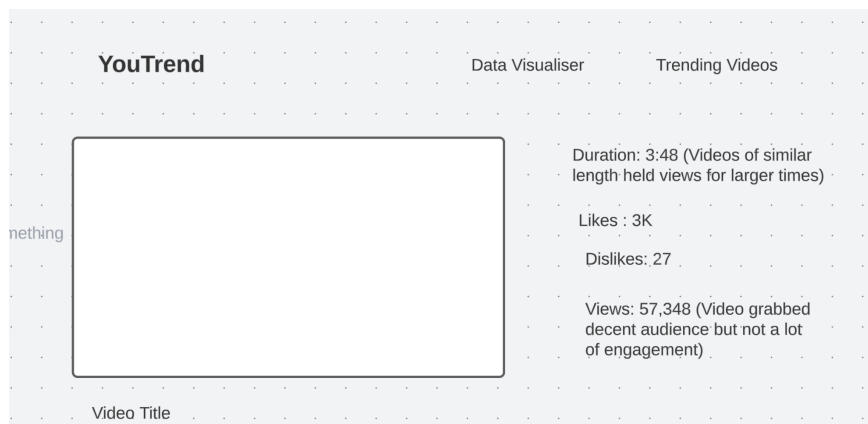
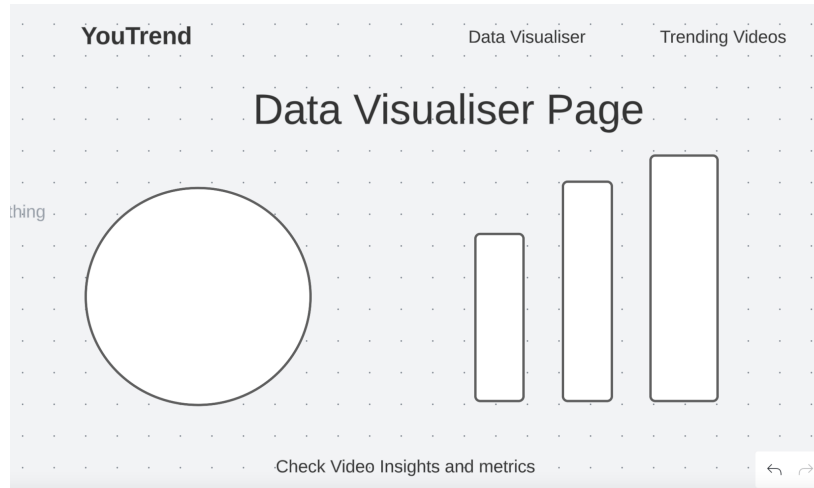
The YouTube Trending Video dataset contains information about the most popular videos on YouTube, as determined by a combination of factors such as views, likes and dislikes, channel titles, video titles, and more. Most of the data would be based on the YouTube Trending Video Dataset, and we will extract the specific data we need from the dataset.

Other data including a description of the video, URL, and thumbnail can be obtained from the YouTube API, which provides access to various data points related to YouTube videos, including trending videos. To use API, we need to create a project in the Google Developers Console and obtain an API key. We will then use the API key to make requests to the API and retrieve the desired data.

9. **Description of the functionality that the website offers:**

a. **Low-Fidelity UI Mockup**





## b. Project Work Distribution

We would try to distribute the work as evenly as possible with all members of the team having the opportunity to work on the front end as well as the back end. As for the specific functions each member will be responsible for, Marvin will be in charge of the view trending videos function, Sambuddha will be in charge of the metric video dashboard, and Leo and Sohyun will be in charge of the trending video insights function. Although these team members will be responsible for overseeing each of these functions, all group members will still be contributing to each of the functions.