YouTube Best Collection Project
Team 083

1. **Summary**

As the world's leading video-sharing platform, YouTube generates vast amounts of data everyday. These trend data, such as trending date, likes, view count, are valuable resources for understanding user preferences and developing website functionalities. However, with millions of videos uploading to YouTube everyday, it becomes harder and harder for users to select videos that they would like to watch. To solve this problem, YouTube creates a functionality called "Trending" that collects some of the most popular videos across different categories with special formula. However, this functionality has its own drawbacks. For example, users are not able to find out the trending videos for a specific category or region. To enhance the experience of watching YouTube videos, our group decided to create a project called "YouTube Best Collection" that makes further categorization based on the selected trending videos.

 In our project, we will implement functionalities like sorting, filtering, and basic searching to our webpage. Our primary goal is to improve users' watching experience, convenience, and spread high-quality videos. Also, through these implements, we will be able to create weekly recommendations based on preferences of users.

2. **Description**
    a. **Usefulness**

    Based on our research, applications and websites with similar functionality are rare. Websites like YouTube have a "trending" functionality but it cannot be filtered or sorted based on user preferences. Most video websites have the functionality of searching, but most of them cannot filter by categories or rating. Therefore, our web application would be much more targeted for users who would like to find anything specific that they would like to watch.

    b. **Functionality**

    Our front-end functionalities include:
    a. Select top 10 videos/channels/tags with highest views and likes for each day, listed by category and region.
    b. Summarize the must watch videos per week globally
    c. Display all trending videos where users are able to sort and filter based on their preferences.
    d. Search for relevant videos based on user input and able to sort and filter.

    Our back-end functionalities include:
   a. Add new videos regularly from the original dataset
    b. Delete videos that have been removed from the dataset
    c. Updating video attributes that has been updated in the dataset
    d. Retrieving information from the database based on various requirements

c. Creativity

A creative functionality we are planning to embed into our project is searching. This functionality would be a significant improvement for our website because it offers users the freedom to search for contents they are interested in. We are planning to implement the search bar by listing all the videos that partially match the prompt provided by the user. Another functionality that we will implement is a list of best watch videos. These videos are collected from all trending videos across the eight countries the database includes and select approximately 10 best videos based on a particular formula. These creative functionality would definite help users better select video types of their interests and benefit their watching experiences.

d. UI design

Our design of the front-end UI is shown below:
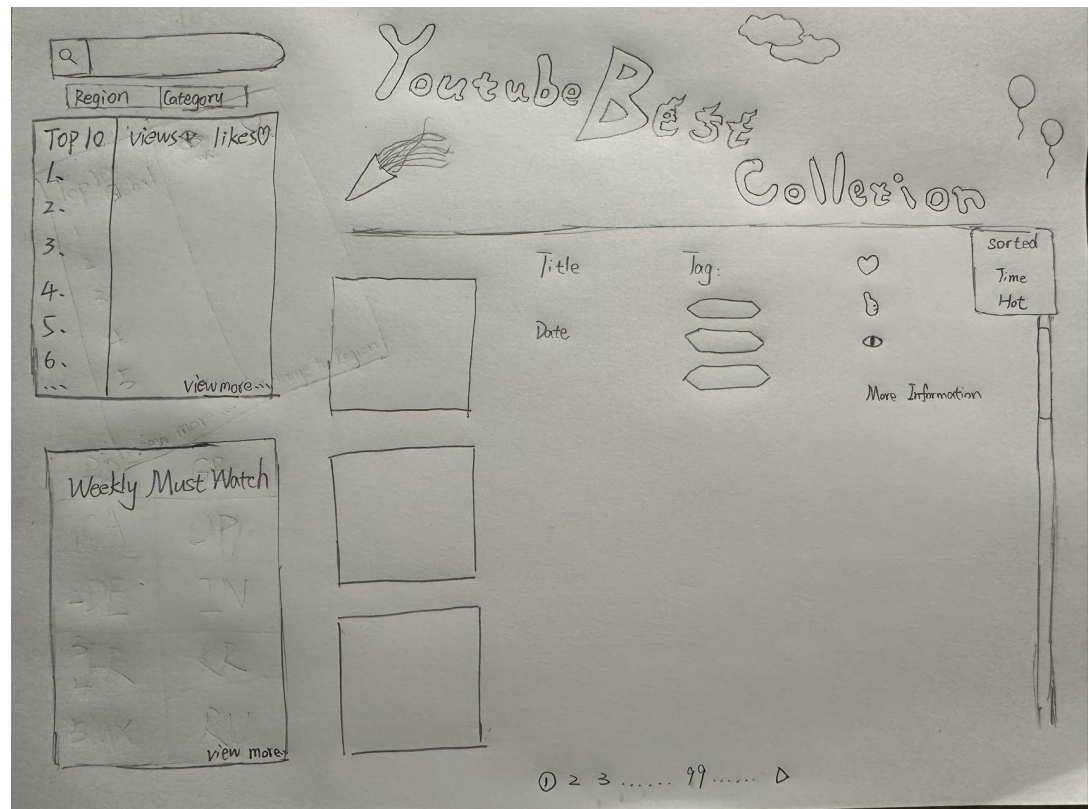
3. Source

The data stored is the real-world dataset about YouTube trending videos acquired from Kaggle. The original dataset consists of trending video data from eleven different countries, including US, UK, and Japan, each with 16 columns describing the attributes of the video. In our dataset, we are going to columns of store video ID, title, published date, channel ID, channel title, category ID, trending date, tags, view count, likes, dislikes, comment count, thumbnail link, and description. We also need to obtain tables

about the meaning of each category ID from the database so that we can group each video into corresponding categories. We might be generating some random data for testing purposes but would not use them in production environments. The dataset link is at https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset?select=US_youtube_trending_data.csv.

4. **Work Distribution**

The following is an outline of distribution of work:

a. Front-end development
   - Zihuang Li: Select top 10 videos/channels/tags with highest views and likes for each day, listed by category and region
   - Tianyi Liu: Summarize the must watch videos per week globally
   - Jiayi Zhang: Display all trending videos where users are able to sort and filter based on their preferences
   - David Fu: Search for relevant videos based on user input and able to sort and filter

b. Back-end development
   - Jiayi Zhang: Link construction of front-end and back-end servers
   - Tianyi Liu: Cloud server deployment
   - David Fu: database operation and maintenance,
   - Zihuang Li: database construction