CS411 PROJECT PROPOSAL

WETUBE

Lumi Xu + Bruce Yin+ David Jiang + John Randolph

09.02.2023

DATA DETAILS

Our group plans to use the following datasets to build a platform to show the trending video of the day, allow users to analyze their video browsing habits, and help them form a friends buble from it by recommendation.

First, we choose a dataset from the course recommendation. The dataset we chose is *YouTube Trending Video Dataset* from "https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset?select=US_youtube_trending_data.csv". It contains details (title, channel, views, etc.) of daily trending videos in 11 different countries, including US, Korean, and Canada since August 2020. The data was retrieved in JSON format using YouTube API and organized into several CSV files.

The dataset is in high quality, which benefits the future development of the project. Different regions are categorized into separate files, and there is no null entry present in the entry.

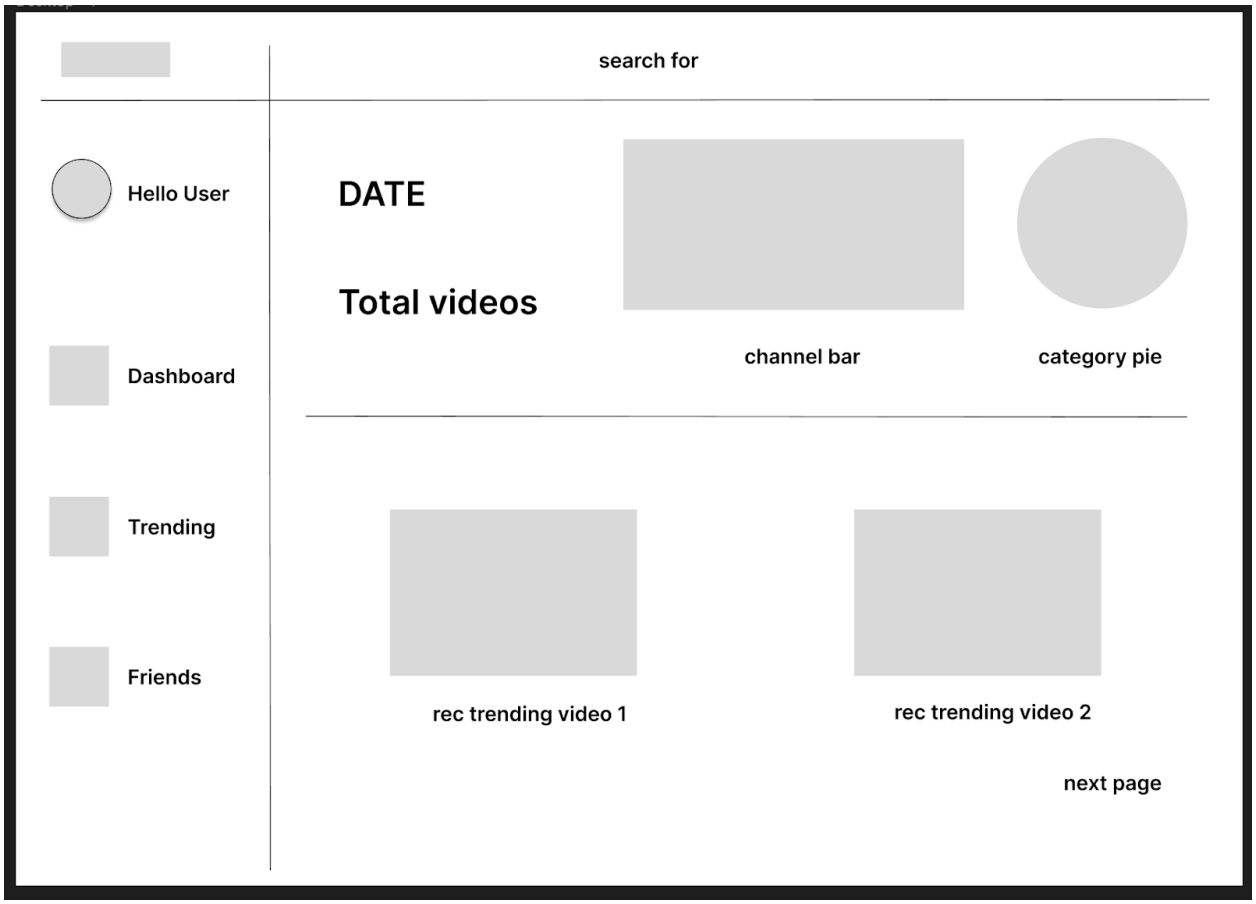
Then, we also will need several other datasets like "users", "playlists". We will these datasets dynamically build as users use the app.

This choice may cause the dataset to be in unstable quality. To minimize the offset, we will restrict user input to only meet our requirements. Such as when adding new videos to the playlist, users have to fill in all the input entry to successfully add it.

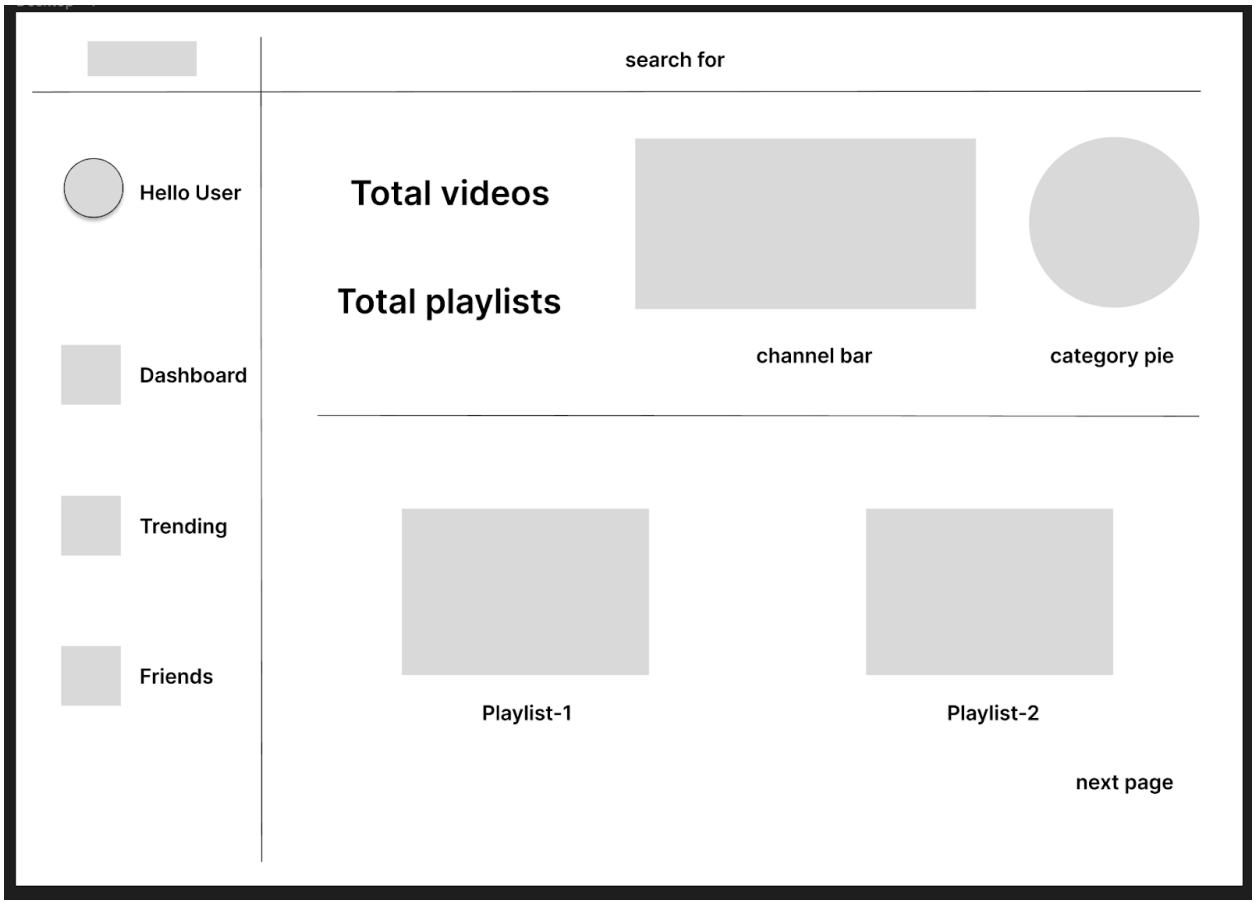
WEB FUNCTIONS

We intend to build four major functions. One is "Trending Now", one is "User DashBoard", one is "Friends Buble"

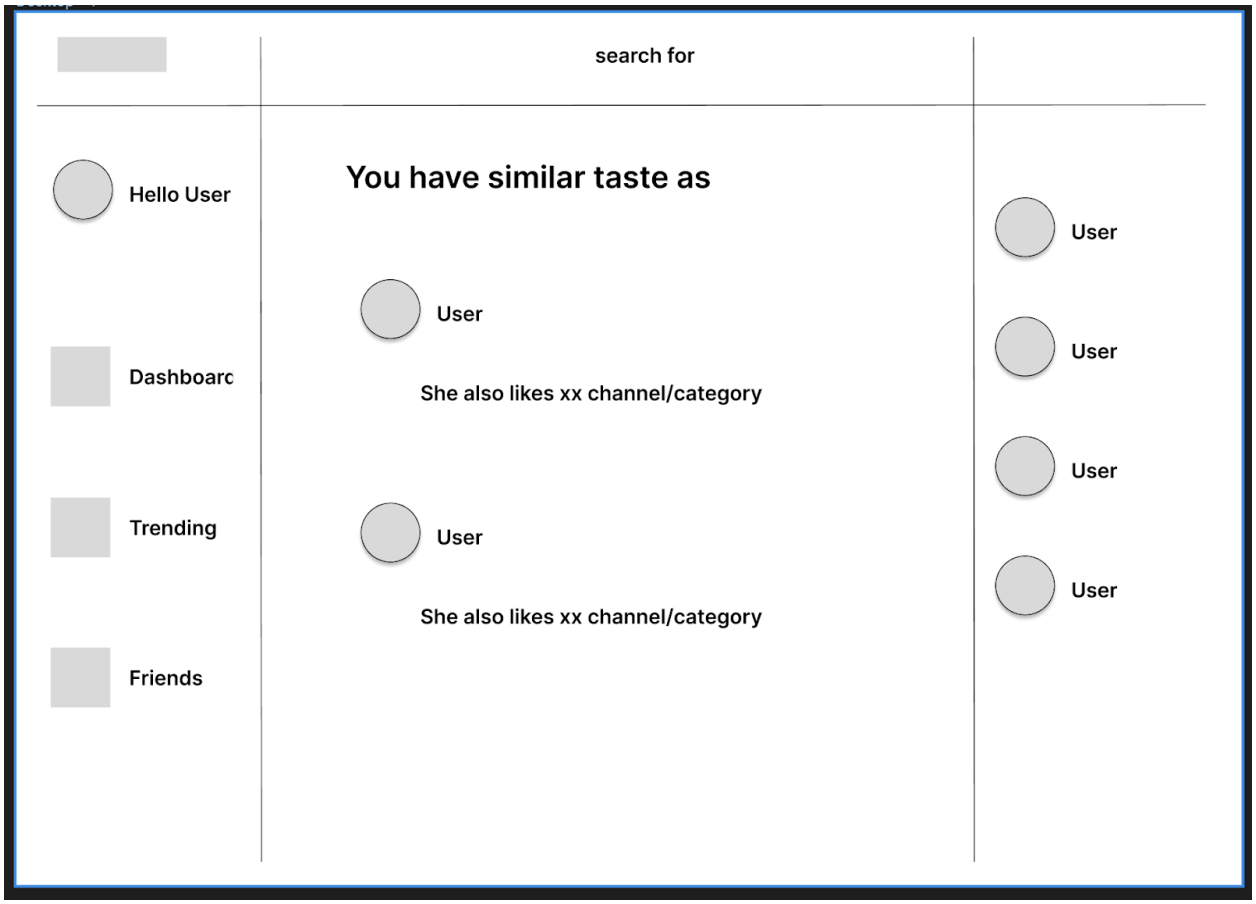
In "Trending Now", once logged in, users are recommended the top 200 trending videos in the selected region. Users can click each video to view detailed information. Inside video details, users can add it to playlists.



In “User DashBoard”, once logged in, users can view their playlists and also charts generated from overall data. Users can click each playlist to view detailed playlist information. Inside playlist details, users can also edit the current playlist.



In “Friends”, once logged in, users can view their friends login information on the sidebar, or more detailed information by clicking into their friends icon and going inside their dashboard. Users will also be recommended new friends based on their common browsing habits(category, region, channels).



CREATIVITY

1. Our project renovates the way users view their playlist by visualizing it:
    - categories pie chart
    - favorite channels bar chart

    We will use ReCharts (https://recharts.org/en-US) to build the graph

2. Our project renovates the way users shares their playlist:
    - private/public playlist
    - recommend friends based on view habits.

    We will recommend friends based on public playlists, by randomly selecting public users with the same favorite category.

3. Our project renovates the way users exposed to new videos
    - Recommendations based on most viewed category
    - Recommendations based on favorite channels

- Recommendations based on friends browsing history

TITLE

Title: WeTube

Meaning: users can form a friend buble in the app.

SUMMARY

We intend to build a web app called WeTube to help users organize and analyze their playlist. Based on their preferences, we also enabled features like “trending” to recommend similar videos and “friends” to allow users to add each other.

To use WeTube, users need to have an account. Once logged in, the app will recommend trending videos based on users' playlist habits. Users can click the recommended video for further details and even add it to the playlist. In the User Dashboard, users can view all its playlists and several charts visualizing its playlist. In the friend page, the app will recommend other accounts sharing the same video preferences.

DESCRIPTION

Our application WeTube would aim to bring a more collaborative and social experience to the usage of YouTube. Over the years, YouTube has deemphasized or deprecated certain features that were oriented towards a community-driven experience in favor of a more widespread yet corporate-powered experience. WeTube aims to utilize this modern YouTube landscape in a way that still maintains focus on individual users and the communities they form.

WeTube would achieve this community-first experience by providing extended functionality focused on building and maintaining an individual profile through statistics about viewing history and personal preferences and on sharing videos with friends maintained through the WeTube site. The use of the trending video database allows for a wide variety of topics and interests to present themselves to users either organically through friends or through a recommendation system.

USEFULNESS

Although YouTube does allow users to create playlists, WeTube puts the emphasis on sharing those playlists with personal friends through the user system. Other external tools allow for enhanced playlist creation tools, such as YouTube Playlist Creator, but they aren't as interested in providing an integrated social experience on a single web application. Data includes the video title, channel title, publish time, tags, views, likes and dislikes, description, and comment count.

REALNESS

The dataset is collected from Kaggle. Kaggle is a prominent online platform and community that serves as a hub for data scientists, machine learning practitioners, researchers, and data enthusiasts. It was founded in 2010 and has since grown into one of the most influential platforms in the field of data science and artificial intelligence. Kaggle offers a wide range of resources and features, including datasets, machine learning competitions, collaborative coding environments, and educational content.

The Kaggle dataset is real and ready to use, contains data of daily trending videos on Youtube of several months. The data is collected from 11 regions including India, USA, Great Britain, Germany, Canada, France, Russia, Brazil, Mexico, South Korea, and Japan respectively, with up to 200 listed trending videos per day since August 2020.

Trending video dataset was collected using Youtube API and stored in separated files according to the category_id varied between regions. To retrieve the category for a specific video, find it in the associated JSON. The datasets of users and playlist is real, dynamically built up from real users. To ensure the quality and reliability of these datasets, we will implement measures such as strictly organize the data collected and restrict the input type.

UI

See here:

https://www.figma.com/file/ogvnDVYkcRMYVqZUa2neqw/CS411_WETUBE?type=design&node-id=0%3A1&mode=design&t=PRUI08C7mFf4CvzG-1

Text explanation:
There will be 6 main pages and one login page for our website. They are “User Dashboard”, “Trending”, “Friends”, “Video Info”, “Playlist”, and “Playlist Info”. The left column of the 6 main pages contains user avatar (click to login), “Dashboard”, “Trending”, and “Friends”. There is a search bar on the top of all these pages for users to look for statistics of a video. Our index page is a dashboard. The total number of videos and playlists are shown at the top, and we’ll use a bar chart to visualize statistics of channels and a pie chart to show categories. There will be two popular playlists at the bottom. The Trending page is almost the same as the dashboard page, except that there is a date on the top, and the playlists at the bottom are replaced by two trending videos. On the “Friends” page, we will compare the user’s playlist to others and show the user’s similar taste with other users. There will be a user list on the right of both “Friends” page and “Video Info” page. The “Video Info” page will show 5 or more statistics of a certain video. The “Playlist” page will show the user’s preferred videos as a list. For the login page, we will use email and password login as default, and add google login as an alternative using Firebase.

WORK DISTRIBUTION

Since all the functions are embedded in the pages. We separate the work based on the front-end part and back-end part.

For the front-end part, since the web page has 6 main pages and 1 login page, our group agreed to divide the work evenly with one person mainly responsible for 1-2 pages.

This is the front-end page work distribution

Dashboard: David, Bruce

Trending: David

Friends: Bruce

Video info: John

Playlist: John

Playlist info: Lumi

Login: Lumi

For the back-end part, we primarily have 4 datasets now, “daily-trending”, “playlist”, “users”, “videos”, so we decided to let one person manage one dataset at the current time. We will separate database work in more detail once we learn more about it. Currently, our group focuses on finishing the UI part.