# INF 551 Project Proposal report

**Cao Yixiang SID: 4246043356**
**Tang Qianqian SID: 1289405231**

## I. Dataset Introduction

We chose dataset for Trending Youtube Video Statistics. It contains attributes like video_id, trending_date title, channel_title, category_id, publish_time, tags views, likes, dislikes, comment_count, thumbnail_link, comments_disabled, ratings_disabled, video_error_or_removed, description. The link of data set is : https://www.kaggle.com/datasnaek/youtube-new

Example data:

| publish_time | tags | views | likes | dislikes | comment_c | thumbnail_li | comments_c | ratings_disal | video_error_ | description |
|---|---|---|---|---|---|---|---|---|---|---|
| 2017-11-10 | God\|"Sheldo | 505161 | 4135 | 976 | 1484 | https://i.ytim | FALSE | FALSE | FALSE | Sheldon is roasting pastor of the church\nyoung Sheldon season 01 episode 3 |

## II. App Design

### a. Database:

We choose firebase to store our data. And we will try to use Firebase cloud function in our app.

### b. Attributes showing:

According to the requirement, our App's default interface will display all the attributes in the data set.

### c. Sorting functions:

According to the second requirement of project, we will implement sorting functions on four attributes. The numerical attributes are: "*likes*", "*views*" and "*publish time*". And the non-numerical attribute is the "*title*" whose type is string.

### d. Filter:

According to the third requirement, we will implement four filters, which include two numerical filters — "*publish time*" and "*likes*", and two non-numerical filters — "*tag*" and "*video error*". The data type of "*tag*" is string and the "*video error*" 's data type is bool.

The attributes showing, sorting function and filters will be realized by JavaScript and python. We will process data in the back-end with python and

represent it on the front-end with JS.

**e. Data Visualization:**

In order to help users more intuitive to observe data trends, we decide to visualize some of the attributes in the dataset.

For "*views*" and "*comment_count*" attributes , we plan to use statistical histogram to visualized it. And we will use pie chart to present the proportion of "*category_id*". For "*tag*" we will use word cloud to crawl keywords.

The Data visualization will be realized by using D3 library of JavaScript.

## III. Group Formation.

Team member 1:

Name: Cao Yixiang:

Responsibility: filter non-numerical attributes, visualization

Team member 2:

Name: Qianqian Tang

Responsibility: web frame, sorting function, filter numerical attribute,

Both of us have experience in developing web and analyzing data, and are proficient in JavaScript and python language. Based on the past experience and skills of the two team members of our team, we are confident that we can complete this project on time and with high quality.

## IV: Milestones:

Below is the Gantt Chart of our project. We plan to finish realizing all sorting functions by week 7 of this semester. And by week 9, all filter should be built. The visualization will start in week 8 and end in week 10. We also reserve a few weeks at the end of the semester to debug or add new features if new ideas are raised.

| tasks | week 5 | week 6 | week 7 | week 8 | week 9 | week 10 | week 11 |
|---|---|---|---|---|---|---|---|
| frame and dataset | web frame | | | | | | |
| | attributes | | | | | | |
| sorting function | | sort numerical attribute | | | | | |
| | | | sort nonnumerical attribute | | | | |
| filter function | | | | filter numerical attribute | | | |
| | | | | | filter nonnumerical attribute | | |
| visualization | | | | stastical histogram | | | |
| | | | | | pie chart | | |
| | | | | | | word cloud | |
| other features | | | | | | | other features |