Final Project Memo

Ying Zhou

10/2/2022

The goal of my final project is to predict the potential YouTube views of the title track music video for Straykids' future third album. Straykids is a South Korean boy group under JYP Entertainment.

An overview of my dataset

I will scrape my dataset basically from YouTube API, kaggle. Here's the source:

https://www.kaggle.com/datasnaek/youtube-new,

https://kworb.net/youtube/artist/straykids.html,

https://developers.google.com/youtube/v3/docs.

Observations are all the video they published before. Since they started the channel at May 11, 2017, they published videos 638 in total. I don't think there is missing data.

Variables might contain in this project are:

- Video Title
- Channel title
- Publish time
- Tags
- Views
- Likes
- Dislikes
- Description
- Comment count

An overview of my research question(s)

I am interested in predicting the amount of YouTube views of the future title music video to analysis the popularity of the group.

My outcome variable is the amount of views. This will be answered with a regression approach.

I think publish time, likes, and description will be especially useful. My goal model is going to be a combination of descriptive and predictive statistic. Because descriptive statistic describes the actual characteristics of the dataset and predictive statistic predicts what is most likely to happen in the future. The combination of these two statistic will contribute to my goal.

My proposed project timeline

I plan to load and clean the data as we learn more knowledge from the lecture in week 3. Probably I will start the exploratory data analysis in week 5 and 6, start model building in week 7 and 8, and finish up in week 9 and 10.

Questions

The question I have is that I am not very sure of the selection of the data. Should I just focus on just the music video or also include the video of other types since my prediction is on aspect of music? If only focus on aspect of music, should I train the model with the data of non-title music videos as well? I am a little bit confused about this.