

Midterm Project Proposal

Chun Gao

Personal statement:

I will land a job as a data analyst at SM entertainment (a leading entertainment company in Asia) within the next seven years. For an entertainment group which owns a complete system of producing and managing artists, a strong sense of music and cultural trends is critical. As the largest online video-sharing platform, the data provided by YouTube could show cultural trends to some extent. Under this condition, my project is to obtain useful information from YouTube dataset, and this information contributes to set position for artists' style and increase the publicity.

Question:

I want to investigate whether there is relationship between the view counts, likes counts, time spent from videos published to listed on trending list, video tags and category of videos. More specifically, which factors might affect videos' view counts.

The data source :

<https://www.kaggle.com/datasnaek/youtube-new>

I would combine the "KR_youtube_trending_data" and "US_youtube_trending_data" to a new dataset.

Proposed Timeline of work :

- ☐ EDA: 11/19/2020
- ☐ Data Processing: 11/12/2020
- ☐ Modeling and Validation: 11/26/2020
- ☐ Write up: 12/01/2020