Midterm Project Proposal

The dataset I chose is YouTube's daily list of trending videos from 2017 to 2018. The authors of the data provided data from different countries, and I selected data from the United States for the following analysis. YouTube has a special algorithm to determine popular videos, including the number of views, shares, comments and likes, and many other variables as references, instead of just taking the most watched videos as trending videos. There are about 16 variables in this dataset, such as publish time, tags, number of views, like or dislike, etc., which provide me with enough predictors for using multilevel model and prediction. My career goal has always been to become a data scientist and working in an outstanding technology company. I really want to go to the parent company of TikTok of China, ByteDance to take up a data-related position, because I think this company is promising and they get high quality data, which is worth analyzing. I think the core competitiveness of TikTok is the popular video recommendation algorithms, which can capture users' preferences and make users addicted for a long time, thus bringing more web traffic and money to the company. As far as I know, the data analysts of ByteDance are basically committed to improving this recommendation algorithm by analyzing different user interaction data. Therefore, as a world-famous video website, YouTube's trending video selecting system is suitable for me to study. Because I believe that using YouTube data to analysis for this midterm project will definitely help me get closer to my career goals and lay a foundation for me to experience similar projects or problems in the future. The primary question I am going to figure out is that what are the factors, and what extent that affect how popular YouTube videos would be.

Proposed Timeline of work:

1. EDA: Nov 13th
2. Data Processing: Nov 18th
3. Modeling and Validation: Nov 24th
4. Write up: Dec 1st

The data source: https://www.kaggle.com/datasnaek/youtube-new