

Proposal 13 : Data Analysis for Trending Youtube Videos

Problem:

Analyze the trending videos information from the Youtube

Dataset:

Trending YouTube Video Statistics (<https://www.kaggle.com/datasnaek/youtube-new>)

The entire dataset contains 5 csv files and 5 json files(for 5 different countries), including various kind of information like video titles, channels, video categories, publish time, number of views, number of likes and dislikes, etc.

Motivation:

Nowadays, an increasing number of people post their videos on Youtube, and it is interesting to know whether a video is popular is dependent on its category and the cultural background of viewers. We plan to use the acquired dataset to analyze the composition and popularity associated with different factors of online videos on YouTube. We'd like to dig in deeper to elaborate the relationship between them. To be more specific, some videos are highly controversial because of its content or types. Also, we want to show how the cultural divergence affects people's likes and the overall most popular video types to shed light on how YouTubers are supposed to refine their videos to get more subscribers, and recommend popular channels of particular video genres.

Plan :

Firstly, we count the number of trending videos of different types at different times, with which we can see the fluctuation. We also plan to show the popularity of particular video types by examining the total views of different video types. Then, we evaluate how these attributes are associated with cultures in different countries. Finally, we calculate the ratio of likes to dislikes for every channel to show the viewers' attitudes towards them.

Project steps

Step	Estimated completion time
1. Collecting, extracting and analyzing data	1 week

2. Data visualization	1.5 weeks
3. Preparing for the report and presentation	1.5 week