



### Capstone Projects

## YouTube Trending Video Analysis

#### Prompt:

For this project, the goal is to determine the features or attributes that make a video successful enough to be placed on YouTube's trending list. To do so, you will need to conduct multiple forms of exploratory analysis on the data provided. The analysis may include but is not limited to, sentiment analysis of video comments or implementing a form of natural language processing to determine similarity of comments or talking points from the video. This prompt is intentionally vague as you and your group are to decide on an approach which will be able to identify the qualities of a trending video. Feel free to get as creative as you would like with this project, there is a lot of value that can be extracted from the dataset and we are eager to see what you and your team are able to find.

#### Deliverables:

The final deliverable will be via presentation at our last meeting of the semester on December 8th, 2019. You and your group will put together a powerpoint presentation that outlines the work done throughout the project and any findings that were notable from your analysis.

Presentations should be between 8 to 15 minutes and should consist of data preprocessing methods, project scope, approach taken, findings, potential areas of error, challenges, and takeaways. Each member should speak during the presentation about his or her contribution.

#### Dataset:

The data for this project has been sourced through Kaggle. The dataset includes several months of data pertaining to daily trending videos on YouTube for multiple countries. Attributes such as video title, channel title, publish time, tags, views, likes and dislikes, description, and comment count are present in the dataset. Video specific categorical labels can be found in the JSON files that are also provided. The dataset can be found here:

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

