**690: Special Topics in Data Science (SP2020)**

Deliverable 1: Project Proposal

- Harsh Bhanderi

YouTube (the world-famous video sharing website) maintains a list of the [top trending videos](https://www.youtube.com/feed/trending) on the platform. [According to Variety magazine](http://variety.com/2017/digital/news/youtube-2017-top-trending-videos-music-videos-1202631416/), “To determine the year’s top-trending videos, YouTube uses a combination of factors including measuring users interactions (number of views, shares, comments and likes). Note that they’re not the most-viewed videos overall for the calendar year”. Top performers on the YouTube trending list are music videos (such as the famously virile “Gangam Style”), celebrity and/or reality TV performances, and the random dude-with-a-camera viral videos that YouTube is well-known for.

The dataset I choose is This is a daily record of the top trending YouTube videos. This dataset includes several months (and counting) of data on daily trending YouTube videos. Data is included for the US, GB, DE, CA, RU, MX, KR, JP, IN and FR regions (USA, Great Britain, Germany, Canada, Russia, Mexico, South Korea, Japan, India and France, respectively), with up to 200 listed trending videos per day.

Each region’s data is in a separate file. Data includes the video title, channel title, publish time, tags, views, likes and dislikes, description, and comment count. The data also includes a category\_id field, which varies between regions.

This dataset was collected using the YouTube API. <https://github.com/mitchelljy/Trending-YouTube-Scraper>

Possible uses for this dataset could include:

* Sentiment analysis in a variety of forms
* Categorizing YouTube videos based on their comments and statistics.
* Training ML algorithms like RNNs to generate their own YouTube comments.
* Analyzing what factors affect how popular a YouTube video will be.
* Statistical analysis over time

I want to answer questions like:

* How many views do our trending videos have? Do most of them have a large number of views? Is having a large number of views required for a video to become trending?
* The same questions above but applied to likes and comment count instead of views.
* Which video remained the most on the trending-videos list?
* How many trending videos contain a fully capitalized word in their titles?
* What are the lengths of trending video titles? Is this length related to the video becoming trendy?
* How are views, likes, dislikes, comment count, title length, and other attributes correlate with (relate to) each other? How are they connected?
* What are the most common words in trending video titles?
* Which YouTube channels have the largest number of trending videos?
* Which video category (e.g. Entertainment, Gaming, Comedy, etc.) has the largest number of trending videos?
* When were trending videos published? On which days of the week? at which times of the day?