**YouTube Client Brief**



**Context**

YouTube (the world-famous video sharing website) maintains a list of the top trending videos on the platform. According to Variety magazine - to determine the year’s top-trending videos, YouTube uses a combination of factors, including, measuring users’ interactions:

* Number of views
* Shares
* Comments
* Likes

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.

### Content

This dataset is a daily record of the top trending YouTube videos.

The data includes several months (and counting) of data on daily trending YouTube videos. Data is included for the GB regions (Great Britain), with up to 200 listed trending videos per day over the same period. 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. To retrieve the categories for a specific video, find it in the associated JSON. More information on specific columns in the dataset refer to the [column metadata](https://www.kaggle.com/datasnaek/youtube-new/data).

This dataset was collected using the YouTube API.

**Task**

The CEO of YouTube has reached out to Stronger Rhondda for an important piece of work. YouTube are keen to utilise the data they have available to answer a handful of key business questions.

**Part 1 – Data Analysis**

YouTube’s key stakeholders are keen to understand the below questions.

1. 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?
2. The same questions above but applied to likes and comment count instead of views.
3. Which video remained the most on the trending-videos list?
4. How many trending videos contain a fully capitalized word in their titles?
5. What are the lengths of trending video titles? Is this length related to the video becoming trendy?
6. How are views, likes, dislikes, comment count, title length, and other attributes correlate with (relate to) each other? How are they connected?
7. What are the most common words in trending video titles?
8. Which YouTube channels have the largest number of trending videos?
9. Which video category (e.g. Entertainment, Gaming, Comedy, etc.) has the largest number of trending videos?
10. When we’re trending videos published? On which days of the week? at which times of the day?

**Part 2 – Machine Learning & Modelling**

YouTube CEO requests a model which predicts the number of views a music video (category 10) has based on many factors, including likes, dislikes and comment count. She is also keen to dig deep into the data available, to enhance the model to its full potential, by feature engendering dimensions such as channel title, title, description and publish time and adding these to the inputs of the model.

**Part 3 - Design**

**DESIGN – WEB PAGE?**

**Part 4 – YouTube Recommendations Engine**

YouTube are keen to build a new recommendations engine. Find out what videos your peers like, or dislike and create a machine learning model to make predictions and recommendations of what they should watch next.