**You Tube Trending Videos Analysis**

**Machine Learning Model Proposal**

**Dataset:**

You Tube is the most popular and widely used online video sharing platform in the world. The available dataset contains trending videos in the US on YouTube that made it to the trending list over a period of several months during the years 2017 and 2018. For each video, we have data on specific attributes such as trending date, title, channel, publish time, tags, number of views, number of likes and dislikes, comments, thumbnail and description.

**Business Objective:**

The purpose of this proposed Machine Learning Model is to provide useful information to content creators, online media companies and other businesses that use social media platforms, in this case You Tube, to propagate their business by reaching out to a larger audience for a longer period.

On the other hand, this would also present a business opportunity to YouTube to identify videos and channels that are likely to trend and approach them for promotions, thus generating even greater views to the videos and revenue for You Tube through promotions.

**Proposed** **Machine Learning Model:**

According to YouTube help (<https://support.google.com/youtube/answer/7239739?hl=en>),

“Trending helps viewers see what’s happening on YouTube and in the world. Trending aims to surface videos that a wide range of viewers would find interesting.”

The page does not provide sufficient information on how videos make it to the trending list and what determines how long it remains trending.

Since the dataset does not have information on non-trending videos, it is not possible to apply a classification model to determine what factors make it a trending or non-trending video.

With the given dataset, it is proposed that a Regression Model be used to predict how long a video remains trending on YouTube and what features enable it to extend the trending duration.

The trending duration would be a calculated field since we have data for all the dates when each video remained on the trending list. The trending duration would be our y value.

Our X values would be view count, count of likes, count of comments and possibly category and count of dislikes. Since there are multiple features (X values) that may influence the trending duration (y value), the multiple regression model would be applied.

The model will need to be revisited, tested and refined based on accuracy and findings in the upcoming stages.