**Categorizing YouTube videos**

**Capstone 1 Project Proposal**

**Springboard Data Science Career Track**

Problem Statement: 300 hours of videos are uploaded to YouTube every minute and 5 billion videos are watched every day. In this huge amount of content, it is very tough to know what makes the video stand out and be the trending or most watched video. This project is intended to do a data analysis of the existing videos and come up with answers to questions such as:

* How many views do trending videos have?
* How many comments, likes do the trending videos have?
* How long do these videos remain trending?
* Is the length of video relevant and have relationship to trending videos?
* What are the most common words in title of trending videos?
* Do the day the week the video is published matter?
* How long in each country it remains trending?
* Analysis based on the category of video.

Dataset: the dataset used for this project is from Kaggle: <https://www.kaggle.com/datasnaek/youtube-new>

Though in future, the solution can be extended to read continuous data from API provided by YouTube:

<https://developers.google.com/youtube/reporting/>

Client: The client for this would be companies trying to advertise their products, individuals looking to maximize their monetization for videos.

Approach: Data wrangling will be done to drop any irrelevant features, adding new features and merging the different datasets per country. Since the dataset for each country have different columns there will be some data cleansing to be done and I have to come up with a logic to fill blank values.

This will be followed by exploratory data analysis to answer the questions stated above using visualization tools. Several classification and regression machine learning algorithms will be used to investigate and build prediction models.s

Deliverables: Project code, PowerPoint presentation, report with findings