### Decoding YouTube Popularity: Predicting and Analyzing of Trending or Non Trending

### Team 2A - Ruicheng Li, Chun-Hui Tsai, Lalitha Bhupalam, Hilary Nguyen, Abhinav Samavedula

The digital world is evolving with YouTube standing at the forefront as a principal platform for video sharing and content creation. With over 2.51 billion monthly users in 2023, the platform serves well-structured videos for individuals to enjoy. Our team decided to dive into this question: What factors influence the popularity of a YouTube video? We are particularly interested in understanding the factors that contribute to a video’s success on this platform.

**Topic Selection**

We choose to focus on what factors create a trending YouTube video. This exploration includes a deep analysis of various factors that could determine the popularity of a single video, such as the optimal time of day to upload videos, and the criterion like views, likes, dislikes, and comments that determines a video’s popularity.

**Data Collection**

We came across a Trending Youtube videos dataset on Kaggle. But since the data is from 2017-2018, we have two ways to collect the data for our analysis:

1. Getting the recent data from 2022-2023 by using the methods described in the documentation for Youtube dataset available on Kaggle using Youtube API with Python and using web scraping to get the most recent data for non-trending videos.

2. Web scraping data for non trending videos from 2017-2018 to match the data from the Kaggle Youtube Dataset

Through either of these methods we will have plenty of information concerning YouTube video statistics. (Mitchell)

**Justification**

Understanding the mechanics of video popularity on YouTube can offer invaluable insights for content creators, advertisers, and digital marketers in implementing their strategies for the maximization of potential customers and revenue.

The questions that we are trying to answer:

* What category do the trending videos on YouTube platform usually belong to?
* What is the relationship between different metrics in a video, like views, likes, and comment counts?
* What’s the best time in a day or week to upload videos?
* Are there any variations in popularity across different regions and categories?

Our method involves a comprehensive statistical analysis of the data covering views, likes, comment counts, and trending days over time. We will be training a model to predict whether a video is trending or not and analyze the data classified as trending to provide insights.

By using a data-focused approach, we aim to provide a detailed analysis that can serve as a helpful resource for those involved in the YouTube community. This analysis can help them make better decisions in improving creators' content strategies.

Should time allow, we are also aiming to use natural language processing concepts to analyze what words appear the most in trending youtube videos.

**References:**

Mitchell, J. “Trending YouTube Video Statistics”, Kaggle, 2019, <https://www.kaggle.com/datasets/datasnaek/youtube-new>