# Team 1 Project Proposal

**Team Members**
- Kyden Pham
- Bryan Corona
- Daniel Palma

**Project description**
In this data mining project, we want to utilize YouTube Data API to scrape and analyze to collect valuable insights from some of the most famous YouTube channels. Our selected channels include well-known creators such as MrBeast, Pewdiepie, and Cocomelon, etc.

**Obtaining and scraping data**
We will research and combine existing methods to create a different data frame that we can work on. Some data frames that we can scrape for example:
- Channel name, subscribers, views, total video
- Channel name, video title, description, tags, published at, view count, like count, comment count, duration, definition, caption, title length.

For us to use the data for analysis, we can do some preprocessing to format some of the attributes if needed, split attributes, or remove them if too much data is missing.

**Objectives**
The primary objective of this project is to extract, analyze, and derive meaningful insights from the content and metadata associated with these influential YouTube channels. By doing so, we aim to gain a comprehensive understanding of their content strategies, audience engagement, and various other key performance indicators.

- Data collection: leverage YouTube Data API to scrape a wide range of data from specified channels. This could include video details, comments, likes, dislikes, views, etc.
- Data preprocessing: cleaning, structuring, and formatting the collected data for further analysis.
- Data analysis: identifying trends, patterns, key metrics for each channel. The main goal is uncovering what makes these channels successful and how they resonate with their audiences.
- Comparative analysis: highlight similarities and differences in their content strategies, audience demographics, and engagement levels.
- Visualization: generate relative graphs and charts to convey insights effectively.
- Regression: create and train model that can predict video KPI (views, likes).