ETL Project  
compiling a data lake of youtube and billboard trending stats

November 2018

**Team Members:** Erica Rosa Thach Tran

## Project Background and Description

We will prepare a data lake with trending statistics from YouTube and Billboard, which could be used to uncover relationships between most popular songs and related YouTube content over time.

## Data Sets to be Used

Trending YouTube Video Stats (<https://www.kaggle.com/datasnaek/youtube-new>)

Billboard Trending Stats

## Rough Breakdown of Tasks

Identify data sources.

Data cleaning csv files.

Import data sources into Jupyter Notebook.

Import data and structure via data frames.

Deploy as tables to MySQL.

Summarize & document ETL process.