

## ETL Project Proposal

Team Name: Data Miners (Seobin, Vaughn, Johnny, William, Habiba)

Project Title: Netflix Trends by Google

### Project Description:

This project will be comparing the popularity of movie titles, rated by google trends, with their duration, the country they're shown in, category, and rating. We will do this using two datasets, "Netflix Movies and TV Shows" by Shivam Bansal and "Google Trends Dataset" by Dhruvil Dave. The former dataset is updated monthly while the latter contains data from 2001 to 2020. Both of the datasets selected are sourced from Kaggle and have a usability rating of 10.0. These datasets source global data from their respective companies, we will be cleaning the data to only reflect trends in the United States. We will be analyzing the data to determine whether individuals in the United States are searching for shows outside of the United States.

### Datasets/Sources:

<https://www.kaggle.com/dhruvildave/google-trends-dataset/code>

<https://www.kaggle.com/shivamb/netflix-shows>

### Extract:

- Downloaded CSV's for google trends and Netflix shows from Kaggle.

### Transform:

- Since the 'trends' dataset contains the top searches from 2001 by each country, we are going to clean the unnecessary data
  - Clean the 'trends' dataset to keep the necessary data related to Netflix such as 'TV Shows' and 'Movies'
  - We are going to keep the specific locational data in the 'trends' dataset such as 'Global' or 'United States'
- Join two datasets that we are going to utilize to find out the popular movies and shows

### Load:

- The final database
- Possible Charts
  - Trend by Duration
  - Trend by Country
  - Trend by Category
  - Trend by Rating