Project Report

Extract

The data sets used in this project were collected from Kaggle.com. These sets originated from the YouTube API and include data from 10 countries.

For each of the countries there were two files:

A CSV file which houses the the data for daily trending videos from November 2018 to June 2018.

A JSON file containing the data on specific categories for each country.

Transform

An ERD was completed to enable visualization of what the final tables would be for future analysis.

Data from the JSON files had to be converted to CSV.

The CSV files were converted to Panda data frames, which matched the ERD tables

This involved cleaning and removing unnecessary data

The data was then joined to create the final dataframe

Tables were created in the project-2 database in Postgres using PgAdmin4

Load

The data was loaded into the tables in postgres. This was chosen to allow for future updates and analysis