ETL Project

Jeevani & Gaby

Dataset 1: <https://www.kaggle.com/orgesleka/imdbmovies>

Dataset 2: <https://data.world/siyeh/walt-disney-movies/workspace/file?filename=disney_movies_total_gross.csv>

Extract:

Our data was in csv format, so we used the csv files we found through the data websites. All the columns looked like the data we want. One of the files movie titles has the year right next to the title.

Transform:

Since some of the titles has years in them we had to use string split to separate the titles from the year. We also had to separate the year from the release data, we used the same string split function. We also had to get rid of the white spaces after splitting so our merging was done correctly. When we did merger, we choose specific columns from two databases to combine them.

Load:

We loaded the combines data table into the database. This merged table tells you what the cross net, nominations, wins and ratings of all the Disney movies.