ETL Report

Tammy Ura-Aguilar

Michael Park

Extract

We looked in Kaggle and found data sets for Anime Recommendations Database. The data frames we got were two csv files. We wanted to find out what shows (TV) and movies anime to watch and what not to watch. The data sets rated Anime shows and movies on a scale from 1-10, 10 being the best. We will load the data frames into a database.

Transform

We had to merge the two data frames first then rename a couple of columns. Then we grouped the data frames by type. We removed -1 from both ratings because -1 meant that the users did not rate the Anime. Then we combined the two ratings to create one data frame. We filtered it out by TV first then sorted it by top 10 TV Anime and then bottom 10 TV Anime. We did the same for Anime movies, top 10 Anime movies and bottom 10 Anime movies.

Loading

We then connected to the local database and then used pandas to load the csv converted data frame into the database. We double checked to see if everything was in place by confirming the data has been added by querying the TV top Anime table, TV last Anime table, movie top Anime table, and movie last Anime table.