ETL Project – Proposal

Jen Mattingly

Tyler Motter

Richard Nwokoye

For our Extract, Transform, Load (ETL) project we will work with IMDB data to compare actors and movies and ratings from different databases. We will refine the data, extracting only relevant information such as names, movies, ratings, and discard the columns we don’t think are relevant to our customer (you), such as divorce statistics, number of children, dates of death, etc..

Once we have completed the cleaning process, and transformed the data into a more sensible format, we will load it into Postgres for storage.

Research Question to answer:

Which movie stars and what movies had the highest ratings?

Datasets:

<https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset>

Breakdown of Tasks:

1. Pick a topic
2. Download the data
3. Clean the data
4. Extract relevant columns
5. Load into database.