ETL Project
Members:
Crisharon Beale
Julie Pyle
Ben Stork
Robert Payne

ETL Project Proposal:

Purpose:
Create a movie information data frame in SQL that will allow users to quickly and conveniently do the following when looking for a movie:

1) Compare movie ratings by release year. What were the top 5 rated movies based on the combined data sets (ratings on Rotten Tomatoes, IMDb, and possibly user reviews)?
2) Compare movie ratings by genre. What were the top 5 rated movies based on the combined data sets (ratings on Rotten Tomatoes, IMDb, and possibly user reviews)?
3) Target age group of movies (Top 5). This will allow users to get movies that target their specific age group and interest users of the same age.
4) Inform the consumer where to stream a movie.

Work Breakdown:

- Schema - Crisharon will take lead for the shema
- GitHub - Robert will take lead for the GitHub repository with assistance from Ben if required
- HTML - Julie
- ER Diagram - Julie
- Cleaning data - Ben and Robert will take lead on cleaning

Current Datasources:
https://www.kaggle.com/rounakbanik/the-movies-dataset

https://www.kaggle.com/ruchi798/movies-on-netflix-prime-video-hulu-and-disney

Possible Issues:

- How to combine data sources. If completed off of movie titles these are not unique. Can we do it off of Movie Title or another attribute?
- Genre column in one of the datasets has multiple entries. WIll need to create a separate column for all Genres and have a boolean response if that movie is part of said genre.
- Is there any other information that is grouped and we are unaware?
- Are there any abbreviations from the dataset?