**Team 6:**

Saeger Godson

Ishin A Yavuz

Jenny Yang

Netflix Movies, IMDB and Rotten Tomatoes Rating, Good Books ETL

A lot of people check IMDB or Rotten Tomatoes score for a movie before watching it on Netflix. Although Netflix has a good recommender system, it still can be hard to use for people to determine how good a movie is. The exercise collects Netflix movies data along with IMDB and Rotten Tomatoes rating data, cleans them and load them into a SQL database, so further analysis can be performed to identify top rated movies on Netflix based on IMDB and Rotten Tomatoes ratings. The database also includes a 10k good books list to identifies which movies were produced from popular books.

**Example queries:**

1. Top 10 IMDB & Rotten Tomatoes rated movies on Netflix
2. Netflix movies produced based on popular books

**Data:**

1. <https://www.kaggle.com/shivamb/netflix-shows?select=netflix_titles.csv>
2. <https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset>
3. <https://www.kaggle.com/stefanoleone992/rotten-tomatoes-movies-and-critic-reviews-dataset?select=rotten_tomatoes_movies.csv>
4. <https://www.kaggle.com/zygmunt/goodbooks-10k?select=books.csv>

**Breakdown of tasks:**

1. Load csv data files
2. Transform the data including cleaning, filtering, aggregating, etc
3. Design the tables structure in SQL- use QuickDBD to generate tables and relationships and create the schema file
4. Load data from Jupyter notebook to SQL
5. Create queries