

Final Project Part 1

- a. I want to create a database based on movies and the movie industry. I can use a resource like IMDB or any streaming service's databases if they're open to public use. I imagine the different tables could be focused on cast, crew, individual movies, genres, and reviews. Queries could then be focused around these tables, trying to figure out ideas such as how many movies an actor has been in, what genres make the most money, the average rating per actor, and more.
- b. Questions:
 - i. How many movies has each actor starred in, and which actor has the most movie appearances?
 - ii. What is the average rating of movies in each genre?
 - iii. Who are the top 3 directors with the most movies released in the last decade?
 - iv. What is the highest rating for a movie an actor who has starred in more than 10 movies has received?
 - v. What is the total number of reviews and the average review score for each movie in the database?
 - vi. How many movies did a specific actor star in within a particular year or decade?
 - vii. What is the correlation between movie ratings and the number of reviews for each movie?
 - viii. Which genres are most common among movies with a runtime over 2 hours?
 - ix. Who are the top 5 actors who have appeared in the most number of different genres?
 - x. What is the average rating of movies for each production studio?
 - xi. What is the longest movie in each genre and how does it compare to the genre's average movie length?

- xii. Which director has the highest average rating across their movies, and how many movies have they directed?
 - xiii. What are the most common co-actors in movies that received a rating above 8?
 - xiv. Which movies have a runtime longer than 2 hours and received a rating above 7?
 - xv. Which movies with the same genre were released in the same year, and how did their ratings compare?
 - xvi. Which cast member has been in the most movies with a score above 7?
 - xvii. What was the distribution of movie ratings for all movies released in the last 5 years?
 - xviii. Which genre had the most movies released in a specific year, and which movie was the highest-rated in that genre?
 - xix. How many actors have received an average rating above 8 across all of their movies?
 - xx. Which movie has the longest runtime within a given genre, and how does it compare to the average runtime for that genre?
- c. IMDB has a readily available database I can download and use. I've looked at them, and the files appear to be tsvs, which must be reformatted into CSVs. Additionally, I may have to reduce their size if they're too big to be imported into the database. These should be easily remedied by either creating my own database so I can set the size restrictions myself and doing some pre-processing with Panda or manually.
- d. https://github.com/dodsonta/CS586_Final_Project