

CS 586 Project Description

1. Introduction

The topic that I chose for the final project is Cinema. This is because I have always been passionate about movies and enjoy exploring their impacts on culture and society. Having a movie database will help us gain valuable insights into various aspects of cinema, such as box office performance, reviews, and the people involved in making these films. This database will also be useful for researchers who want to analyze trends and understand how cinema changes over time.

2. Questions

Here is a list of 20 questions that are frequently asked about cinema:

- Question 1: How many movies were released each year?
- Question 2: Which movie genre is the most common?
- Question 3: Which certificate rating (e.g. PG, R, G) earns the most money?
- Question 4: What are the top 10 highest-grossing movies of all time?
- Question 5: Who are the top 5 directors with the most films grossing over \$100 million?
- Question 6: Who are the top 10 actors that appear most frequently in movies with high ratings (e.g. above 7.5)?
- Question 7: Which director has the highest average rating for their movies?
- Question 8: How many movies were released each year in a certain period (e.g. 2000 - 2010)?
- Question 9: What is the average runtime of movies in each genre?
- Question 10: How has the average rating for movies changed in each decade since 1940?
- Question 11: Which movie genre has the highest average rating?
- Question 12: How has the average number of votes for recently released movies (e.g. 2 years or less) compared to old movies (e.g. 15 years or more)?
- Question 13: What is the average box office gross for movies in each genre, considering only movies with a high number of votes (e.g. 100000 or more votes)?
- Question 14: Which actors or actresses have appeared in the most films?
- Question 15: What is the total box office gross for all movies directed by a director (e.g. James Cameron)?
- Question 16: Which year had the highest number of movies with a high rating (e.g. above 7.5)?
- Question 17: What actors appear together in more than one movie?
- Question 18: Which actors have appeared across most movie genres?
- Question 19: Which actor has the highest average rating across all their movies?

- Question 20: What is the average runtime of movies that earn the most money?

3. Data Source

The database will be based on this data source: [IMDB Movies Dataset](#). To prepare the movie dataset for injection, I will utilize Google Sheets for data cleaning and organization. After cleaning the data, I will divide the dataset into multiple CSV files representing different entities, such as movies, directors, and actors. I will also create tables in PostgreSQL that match the structure of each CSV file. Finally, I will import the CSV files into the PostgreSQL database using the import feature.