

Name: Fuad Choudhury

Course: CIS4400

Semester: Spring 2025

Professor: Jefferson Bien-Aime

Homework: 1

Movie Rating Analysis based on IMDB

Business Requirements 1pts

Goal: Based on IMDb ratings, analyze and compare three to four movies to determine which one has the highest rating.

- This will help users quickly identify highly-rated movies.
- Additionally, the analysis will consider the release year of each movie.

Functional Requirements 1pts

A simple program that allows users to compare movie ratings.

- The interface should be user-friendly and provide relevant details.
- Display the release year along with the rating score for better comparison.
- Users can filter movies based on year to find top-rated films.
- Provide the option to compare multiple movies.
- Include graphical analysis for visual comparison.

Data Requirements 1pts

- IMDb Movie Ratings
- Release Year
- Number of Votes

Data Sourcing

- Web Scraping using Python for scraping movie details.
- Using two CSV files from IMDb websites.

Information Architecture 2pts

This system enables users to evaluate movie ratings by entering specific movie titles. It will extract information from IMDb, including details such as the primary title, release year, number of votes and IMDb rating score.

- Users enter a movie title/name to retrieve its IMDb rating and release year along with the number of votes.
- The system will get necessary details from IMDb's database.
- In addition, users can analyze and compare movies based on their ratings and release years.

Data Architecture 1pts

- When a user inputs a movie title, the system extracts the release year, rating score, and number of votes from IMDb.
- After that, the data will be processed and analyzed to identify the highest-rated movie.

Dimensional Modeling 3pts

- The data warehouse is composed of a fact table that includes ratings, votes, and the year, along with dimension tables that classify movies based on their release year.
- Also, a surrogate key is employed for both the fact and dimension tables.

Deliverables 1pts

- Links to data sources (<https://datasets.imdbws.com/>)
- Azure Storage Link:

Blob-SAS-Token:

sp=r&st=2025-05-03T20:09:13Z&se=2025-05-31T04:09:13Z&spr=https&sv=2024-11-04&sr=c&sig=yS7zZiabUpvHN5IEa0zoCvtYR4PVXdtqU21tv6vVvQE%3D

Blob-SAS-URL:

<https://cis4400datastorage2025.blob.core.windows.net/cis4400homework?sp=r&st=2025-05-03T20:09:13Z&se=2025-05-31T04:09:13Z&spr=https&sv=2024-11-04&sr=c&sig=yS7zZiabUpvHN5IEa0zoCvtYR4PVXdtqU21tv6vVvQE%3D>

- Data origin: IMDb Non-Commercial Datasets; Customers can access subsets of IMDb data for personal and non-commercial purposes.

[\(https://developer.imdb.com/non-commercial-datasets/\)](https://developer.imdb.com/non-commercial-datasets/)

- Data dictionary (Using MS Excel)
- Git repository Link (https://github.com/fuadchoudhury/CIS4400_Homework)
- Data Model (Using Draw.io)
- Accessible IMDb Data Warehouse:

[\(https://developer.imdb.com/non-commercial-datasets/\)](https://developer.imdb.com/non-commercial-datasets/)