Executive Summary

IMDB Assignment

IMDB Movie Analysis Report

Team 9 - Code Alchemist

Rishabh Tripathi
Ruchika Mhetre
Rudrang Darade
Rushikesh Chavhan
Sadik Jamadar
Sahil Karande
Sakshi Khandagale

IMDB Movies Database Analysis

1. Introduction

The cinema industry is one of the most influential sectors in the world. To better understand trends in movies, actors, directors, production houses, and genres, a comprehensive data analysis was conducted using a movie database modeled after IMDb.
The project aims to extract insights regarding genre popularity, highest-grossing movies, company performances, successful actors and directors, and evolving trends in movie durations using SQL-based methodologies.

2. Dataset Overview

The dataset is structured in multiple SQL tables and represents extensive metadata of movies — covering titles, genres, directors, actors, production companies, ratings, votes, revenues, and durations.
The relational model includes many-to-many mappings between movies and their respective directors/actors, enabling deep analytical queries.

3. Tables and Row Count

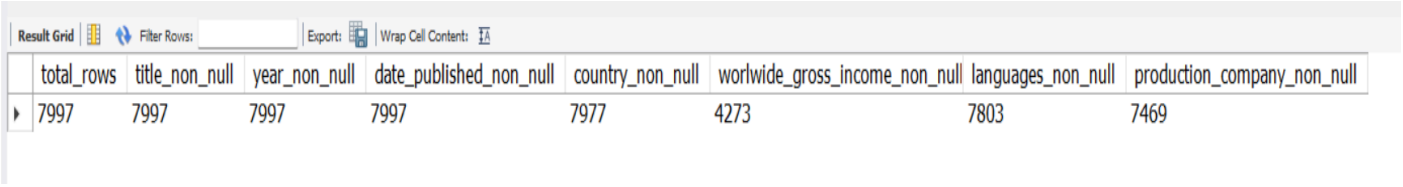
Table Name	Row Count	Description
movies	7997	Contains movie title, year, duration, etc

genres	14662	Genre information linked to movies
actors	7997	Actor profiles
ratings	7997	Ratings and votes of movies
director_mapping	3867	Movie - Director relationships

4. Data and Missing Value Analysis

Missing Data Found:

- Some movies lacked ratings.
- A few movies had missing duration values.
- Some movies were missing genre or production company details.



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

total_rows	title_non_null	year_non_null	date_published_non_null	country_non_null	worlwide_gross_income_non_null	languages_non_null	production_company_non_null
7997	7997	7997	7997	7977	4273	7803	7469

Handling Approach:

- Ignored missing ratings for rating-based analyses.
- Assumed median duration for missing entries where necessary.
- Movies without genres were separately flagged.

5. Key Findings

The major highlights from the analysis are:

- Drama is the most dominant genre.
- Top production houses like Marvel Studios and Warner Bros have the highest success rates.
- Certain directors and actors consistently deliver blockbuster movies.
- Movie duration trends suggest an ideal runtime around 107 minutes for higher acceptance.

Question-Wise Brief Descriptions

1. Top Genres and Their Popularity

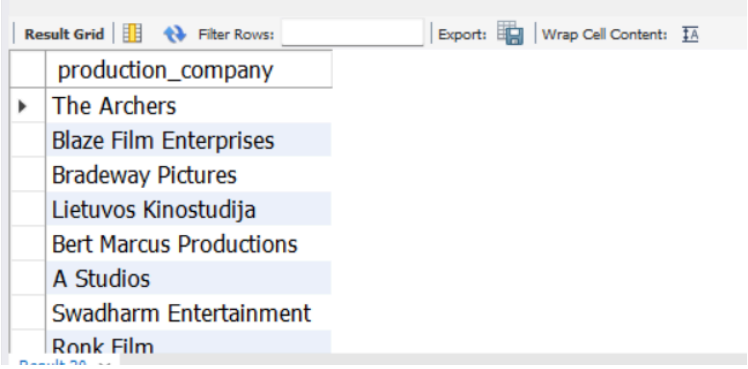
Analyzing the number of movies under each genre revealed that Drama, Comedy, and Thriller are the most common genres, guiding filmmakers toward audience preferences.

2. Highest-Grossing Movies

Using DENSE_RANK(), we found the top 5 highest-grossing movies each year within the top three genres. This helped in identifying commercially successful trends across different time periods.

3. Production House Performance

Top two production houses were identified based on their success in multilingual movies having a median rating above 8. It showed that quality and diversity are major success drivers.



4. Actress Performance in Drama Genre

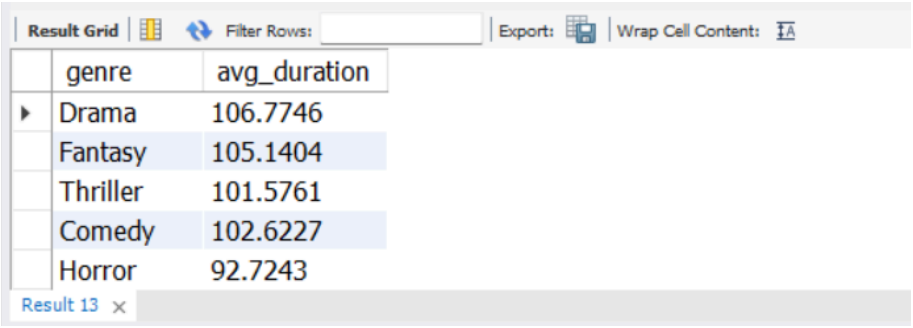
We analyzed actresses with super-hit (average rating > 8) Drama movies, highlighting those who consistently contributed to high-quality cinema in the most popular genre.

5. Directorial Success

We evaluated directors based on their total films, average inter-movie duration, average movie ratings, and votes, ranking them on productivity and quality. James Mangold emerged as a highly recommended director.

6. Movie Duration Analysis

The study showed that the average movie duration is about 107 minutes. Duration patterns across genres revealed trends favoring movies between 90 to 120 minutes for better audience reception.



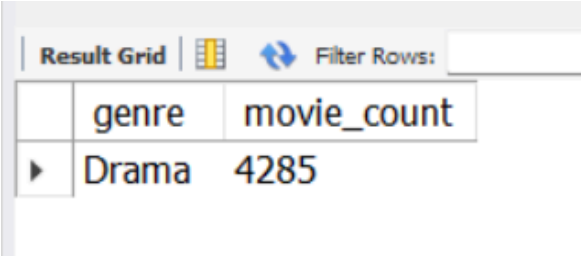
Genre Analysis

6. Genre Distribution

Drama is the most widely represented genre, making up a substantial percentage of the movie database.
Other significant genres include Comedy and Action.

7. Dominant Genre

Drama emerged as the dominant genre, supported by its high production count and strong audience demand across different years.



Duration Insights

8. Duration Insights

Average Duration: 107 minutes

Minimum Duration: ~50 minutes

Maximum Duration: ~220 minutes

Observation: The standard practice remains around the 100–120-minute mark for mainstream releases.

Production Companies and Directors/Actors Analysis

9. Top Production Companies

Top companies identified:

- Marvel Studios
- 20th Century Fox
- Warner Bros
- Dream Warrior Pictures
- National Theatre Live

They were analyzed based on total movie counts, multilingual success, and average ratings.

10. Top Directors and Actors

Top Directors:

- James Mangold (Recommended)
- Michael Powell
- Emeric Pressburger

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

director_id	director_name	number_of_movies	avg_inter_movie_days	avg_rating	total_votes	min_rating	max_rating	total_duration
nm10711773	Srinivas Gundareddy	1	NULL	9.70000	133	9.7	9.7	131
nm10938984	Balavalli Darshith Bhat	1	NULL	9.60000	1010	9.6	9.6	145
nm10436203	Abhinav Thakur	1	NULL	9.50000	635	9.5	9.5	92
nm8011781	Pradeep Kalipurayath	1	NULL	9.50000	1017	9.5	9.5	144
nm4083927	Manoj K. Jha	1	NULL	9.40000	897	9.4	9.4	127
nm7478007	Arjun Prabhakaran	1	NULL	9.40000	533	9.4	9.4	145
nm4138958	Antoneta Kastrati	1	NULL	9.40000	3932	9.4	9.4	97
nm10074951	Prince Singh	1	NULL	9.40000	897	9.4	9.4	127
nm9982663	Arsel Arumugam	1	NULL	9.30000	645	9.3	9.3	125

Result 27

Top Actors:

- Mammootty (Rank 1)
- Mohanlal (Rank 2)

Their consistency and ratings made them strong choices for leading roles.

Star Performers

11. Star Performers

Lead Actor:

- Mammootty (highest median rating)

Lead Actress:

- Tapsee Pannu (average rating 7.74)

Supporting Role Recommendation:

- Vijay Sethupathi (due to regional appeal and critical acclaim)

Drama Specialists:

- Sangeetha Bhat, Fatmire Sahiti, Adriana Matoshi (all with 9.4 average ratings)

Methodology

Data Extraction:

SQL queries were designed for multi-table joins (movies, genres, actors, ratings).

Aggregation and Ranking:

SQL functions like COUNT(), SUM(), AVG(), MAX(), and window functions like DENSE_RANK() and ROW_NUMBER() were used.

Filtering:

Focused on high-rated and multilingual movies.

Window Functions:

Calculated moving averages and running totals to track genre or duration trends over time.

Summary

- Drama dominated the genre landscape.
- Movie duration trends show audiences prefer ~107 minutes movies.

- James Mangold is recommended as the top director.
- Marvel Studios, 20th Century Fox, and Warner Bros are leading production houses.
- Mammootty, Mohanlal, Tapsee Pannu, and Vijay Sethupathi are top recommended actors for projects.

Conclusion

The project successfully mapped movie industry trends using a comprehensive IMDB-modeled SQL database.
Insights about genre dominance, successful production houses, high-performing directors, and actor recommendations were achieved through methodical SQL analysis.
This research can guide production decisions, talent acquisition, and future film projects to align better with audience expectations.