Indian Movie Box Office Analysis using SQL

# 1. Objective

To analyze Indian movie data and uncover insights regarding box office collections, performance by language, genre, and director, and identify top-performing actors and OTT platforms using SQL queries.

# 2. Tools & Technologies Used

- Microsoft SQL Server Management Studio (SSMS)  
- Flat File Import  
- SQL (DDL, DML, Joins, Subqueries, Aggregates, CTEs, Window Functions)

# 3. Dataset Details

Table Name: Movies\_2025

Source: Custom dataset (imported via flat file)

Columns in Dataset:  
- FilmID (INT)  
- Title (VARCHAR)  
- Release Date (DATE)  
- DirectorID (INT)  
- Lead Actor/Actress (VARCHAR)  
- LanguageID (INT)  
- Industry (VARCHAR)  
- GenreID (INT)  
- Budget in Crores (DECIMAL)  
- First Day Collection Worldwide in Crores (DECIMAL)  
- Worldwide Collection in Crores (DECIMAL)  
- Overseas Collection in Crores (DECIMAL)  
- India Gross Collection in Crores (DECIMAL)  
- Verdict (VARCHAR)  
- IMDb Rating (DECIMAL)  
- Runtime (INT)  
- OTT Platform (VARCHAR)

Also created lookup tables:  
- Director (DirectorID, DirectorName)  
- Language (LanguageID, LanguageName)  
- Genre (GenreID, GenreName)

# 4. SQL Concepts Used

Easy Level Queries:  
- SELECT  
- ORDER BY  
- WHERE filters  
- TOP N queries

Moderate Level Queries:  
- INNER JOIN  
- GROUP BY with aggregates (AVG, MAX)  
- Subqueries  
- Nested queries

Advanced Level Queries:  
- Common Table Expressions (CTEs)  
- RANK() OVER() PARTITION BY (Window Functions)  
- Complex nested queries with multiple joins  
- Conditional filtering in subqueries

# 5. Sample Questions Solved

## Easy:

Q-1: List all movies with their title and IMDb rating

Q-2: Find movies released after 2020

Q-3: Get total number of movies in the database

## Moderate:

Q-1: Top 5 movies with highest worldwide collections

Q-2: Average IMDb rating by Language

Q-3: Top 3 Directors Based on Average IMDb Rating of Their Movies

## Advanced:

Q-1: Find the top-grossing movie for each genre released after 2015 with IMDb rating above 6

Q-2: For each director, find their most profitable movie, along with its genre, language, and OTT platform

Q-3: Actor with Highest Grossing Movie in Each Language

# 6. Key Takeaways

- Learned and applied real-world SQL concepts.  
- Understood how to clean, analyze and derive insights from structured movie data.  
- Practiced solving business-type analytical questions in a project format.

# 7. Future Scope

- Build a Power BI dashboard using the same dataset.  
- Automate query reporting using SQL jobs or scripts.  
- Add more data like critics' reviews, awards, and social media buzz for deeper analysis.

# 8. Conclusion

This project helped reinforce practical SQL knowledge and simulate how data analysts handle entertainment industry datasets. The insights from the queries can assist production houses, OTT platforms, and investors in making data-driven decisions.

Prepared by: Kavyanshu Pawar