IMDb Data Analysis

🔍 **Project Overview :** This project explores trends and performance metrics in the film industry using IMDb's public dataset. **SQL** is used to query and transform the data, and insights are visualized through interactive **BI** dashboards. The goal is to identify patterns in movie success, genre evolution, and contributor influence over time.

📦 Dataset Description :

Series_Title	**Released_Year**	**Certificate**	**Runtime**	**Genre**	**IMDB_Rating**	**Meta_score**
Director	**Star1**	**Star2**	**Star3**	**Star4**	**No_of_Votes**	**Gross**

🧰 Tools we used in an IMDb Data Analytics Project :

1. **Data Storage & Preparation**

   - Microsoft Excel
   - Google Sheet
   - Python (Data preprocessing, cleaning if required)

2. **Data Transformation & Querying**

   - SQL (Structured Query Language) Filtering, grouping, bucketing, aggregation, joins)
   - Views / CTEs (Create reusable logical tables in SQL)
   - Window Functions (Ranking, running averages, percentiles)

3. **Data Visualization & Reporting**

   - Power BI Desktop (Primary dashboard & visualization tool)
   - Power Query (inside Power BI) Clean, reshape, and transform data
   - DAX (Data Analysis Expressions)
   - PowerPoint (For final presentation with visual insights from Power BI)

🚀 Execution Blueprint :

1. **Import** the CSV into a SQL database (e.g., PostgreSQL).
2. **Run SQL queries** to validate metrics and exports for reference.
3. **Load cleaned data into Power BI**.
4. **Define relationships**, create dimension tables, and establish modeling.
5. **Build dashboard pages**
6. **Add interactivity** via slicers (Year, Genre, Director).
7. **Annotate insights**: highlight top performers, anomalies, and recommendations.

SQL Business Queries

1. Preview First 10 Rows.
2. Get Table Schema (Column Names & Data Types)
3. Count Total Rows
4. Count of NULLs per COLUMN
5. Summary Statistics for Numeric Columns
6. Distribution of Movies by Certificate and YEAR
7. Top 10 Most Frequent Genres / UNIQUE Genres
8. Count how many movies each director has directed
9. Most Frequent Lead Actor (Star1)
10. Correlation Between IMDB Rating and Meta Score
11. Outliers in Gross Revenue (Top and Bottom)
12. Distribution Buckets of IMDB Rating
13. Get the top 10 movies by IMDB rating
14. Most Common Movie Runtimes
15. Duplicate Titles Check
16. Find the highest Grossing & Lowest Grossing movies.
17. List all movies with a meta score below 50
18. List all movies with certificate 'PG-13'
19. Find top 5 directors by average IMDB rating
20. Find year-wise average gross
21. Calculate the percentage of movies with a Meta_score above 70
22. List all movies with IMDB rating above average
23. Top 3 genres with highest average IMDB rating
24. Find total gross revenue generated by each genre
25. Find directors with more than 5 movies and average rating > 8
26. Create a star-cast frequency leaderboard (count how many times each actor appears across all star columns)
27. Rank Movies by IMDB Rating
28. Dense Rank of Movies Within Each Genre by Gross
29. Running Total of Votes by Year
30. Lag/Lead to Compare Ratings
31. Cumulative Gross, Avg Gross, Avg Votes Using Window Functions
32. Query with PRECEDING Window Frame
33. Cumulative Sum & Avg Over Top 5 Movies by IMDB Rating
34. Creating View for this table