

The background is a dark blue gradient with various decorative elements. On the left, there is a blue arrow pointing right with five white chevrons inside. On the right, there is a blue arrow pointing left with five white chevrons inside. Scattered throughout are green circuit-like lines with small circles at the ends, and horizontal rows of small white dots. The main title is centered in large, bold, light blue and white text.

IMDB DATASET

AN SQL-BASED ANALYSIS

Key Insights from the World of Cinema

Presented By: MAYINI CHARAN TEJ




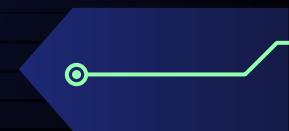
Project Overview & Objectives



Introduction to the Dataset:

- This project analyzes a rich dataset from IMDB, containing information across 6 key tables: Movie, Genre, Ratings, Names, Director_Mapping, and Role_Mapping.
- It covers movie details, genres, audience ratings, and information about the actors and directors.

Project Objectives:

- To reinforce advanced SQL skills, including joins, aggregations, and filtering.
 - To analyze the data to uncover trends and extract meaningful insights about the film industry.
 - To present these findings in a clear and concise manner.
- 
- 

Dataset at a Glance (Data Exploration)

Scale of the Data (Query 1):

- Movies: **[Count from movie table]** records
- Ratings: **[Count from ratings table]** records
- Names (Actors/Directors): **[Count from names table]** records

Data Quality Check (Queries 2 & 16):

- Initial analysis revealed that some columns contain a significant number of null values.
- **Key columns with missing data:** worldwide_gross_income in the movie table and height in the names table.
- **Insight:** This highlights the importance of data cleaning and handling missing data in any real-world analysis.

Key Finding 1 - Movie Production Trends

Yearly Production (Query 3):

- The number of movies produced each year shows a distinct trend. (*Describe the trend you observed, e.g., "a steady increase over the decades"*).

Geographic Focus (Query 4):

- In 2019, the USA and India were major production hubs, accounting for a combined total of **[Result from Query 4]** movies.

Insight: Movie production is not static; it evolves over time and is concentrated in key global markets.



Key Finding 2 - The World of Genres

Most Popular Genre (Query 6):

- The most dominant genre in the dataset is **[Result from Query 6]**, with a total of **[Count from Query 6]** movies.

Genre and Duration (Query 7):

- Average movie duration varies significantly by genre.
- Longest Average Duration: **[Genre with highest avg_duration]**
- Shortest Average Duration: **[Genre with lowest avg_duration]**

Insight: Genre is a defining characteristic of a film, influencing not just its story but also its runtime.



Key Finding 3 - What Makes a "Top" Movie?

Critical Acclaim (Query 10):

- The top 10 movies by average rating include titles like [Example Title 1] and [Example Title 2].

Audience Popularity (Query 22):

- However, the movies with the most votes are not always the highest-rated. The top 5 by total votes include [Example Title from Query 22].

A Unique Trait (Query 13):

- Interestingly, we found [Count] movies with an average rating above 8 that start with the word "The".

Insight: A movie's success can be measured by both critical rating and audience engagement (votes), and these two metrics don't always align.

Key Finding 4 - The People Behind the Camera

Top Production Companies (Query 18):

- The production companies whose movies have received the most total votes are

Prolific Directors (Query 19):

- There are **[Result from Query 19]** directors in the dataset who have directed more than three movies.

Insight: A small number of influential production companies and experienced directors have a major impact on the film landscape.

Key Finding 5 - Actor & Actress Analysis

Top-Tier Talent (Query 17):

- The top two actors who consistently appear in films with a median rating of 8 or higher are [Actor 1] and [Actor 2].

A Surprising Correlation (Query 8):

- Conversely, we identified actors who have appeared in more than 3 movies that, on average, are rated below 5.0. This shows that a long filmography doesn't always guarantee high ratings.

A Fun Fact (Query 20):

- When analyzing physical attributes, the average height for actors is [Avg Actor Height], while for actresses it is [Avg Actress Height].

Insight: Data can reveal which actors are associated with critically acclaimed films and uncover interesting, non-obvious patterns.

Interesting Outliers & Comparisons

The Longest Film (Query 23):

- The movie with the longest duration is **[Movie Title from Query 23]**, produced by **[Production Company from Query 23]**.

A Look Back in Time (Query 21):

- The oldest films in our dataset provide a glimpse into early cinema, including **[Oldest Movie Title]** from the year **[Year]**.

International Vote Comparison (Query 15):

- When comparing two European film industries, we found that **[German/Italian]** movies receive more votes on average than **[Italian/German]** movies.

Insight: Exploring the extremes and making direct comparisons within the data can reveal fascinating, specific facts about film history and audience behavior.

Conclusion & Key Takeaways

Summary of Findings:

- This analysis of the IMDB dataset demonstrates the power of SQL to extract valuable insights, from high-level production trends to specific details about individual movies and artists.

Key Takeaways:

1. **Data Quality is Paramount:** The presence of null values underscores the need for careful data assessment.
2. **Genre Defines More Than Story:** It correlates strongly with other film attributes like duration.
3. **Success is Multi-faceted:** Critical acclaim (rating) and audience popularity (votes) are two different, important metrics.

Final Thought

By querying a structured database, we can transform raw data into a compelling narrative about the art and business of filmmaking.

Thank You
Any Questions.....?