**Project 1: IMDb Movies Analysis using SQL**

Bolly Movies, an Indian film production company, has a successful track record of producing numerous blockbuster films. While primarily catering to the Indian audience, they have decided to venture into the global market with their upcoming project scheduled for release in 2022.

Objective:

Recognizing the value of data-driven decision-making, Bolly Movies has enlisted your expertise as a data analyst and SQL specialist. The objective of this case study is to analyse the movie dataset using SQL queries and extract valuable insights to guide Bolly Movies in planning their new project. The analysis will cover various aspects such as table exploration, movie release trends, production statistics, genre popularity, ratings analysis, crew members, and more.

Segment 1: Database - Tables, Columns, Relationships

1. What are the different tables in the database and how are they connected to each other in the database?

Answer: The different tables are:

(i)Genre Table: Columns: movie\_id (foreign key referencing movie.id), genre. Relationships: Connected to the Movie table through movie\_id.

(ii) Ratings Table: Columns: movie\_id (foreign key referencing movie.id), avg\_rating, total\_votes, median\_rating

Relationships: Connected to the Movie table through movie\_id.

(iii) Movie Table: Columns: id, title, year, date\_published, duration, country, worldwide\_gross\_income, languages, production\_company Relationships: No explicit foreign key relationships mentioned, but it seems to be the central table.

(iv) Role Mapping Table: Columns: movie\_id (foreign key referencing movie.id), name\_id (foreign key referencing names.id), category

Relationships: Connected to the Movie and Names tables through movie\_id and name\_id, respectively.

(v) Director Mapping Table: Columns: movie\_id (foreign key referencing movie.id), name\_id (foreign key referencing names.id)

Relationships: Connected to the Movie and Names tables through movie\_id and name\_id, respectively.

(vi) Names Table: Columns: id, name, height, date\_of\_birth, known\_for\_movies Relationships: No explicit foreign key relationships mentioned, but it seems to be related to Role Mapping and Director Mapping tables through id.

It appears that the central table is the Movie Table, and other tables like Genre, Ratings, Role Mapping, Director Mapping, and Names are connected to it through foreign keys.

1. Find the total number of rows in each table of the schema.

Answer: Total no. of rows in each table of the schema is as follows:

(i) Genre Table: 14662

(ii) Ratings Table: 7976

(iii) Movie Table: 7997

(iv) Role Mapping Table: 15615

(v) Director Mapping Table: 3867

(vi) Names Table: 25735

1. Identify which columns in the movie table have null values.

Answer: worldwide, production\_company

Segment 2: Movie Release Trends

1. Determine the total number of movies released each year and analyse the month-wise trend.

ANSWER:

SELECT

year,

COUNT(\*) AS total\_movies

FROM

amitdb.movies

GROUP BY

year

ORDER BY

year;

-- to analyze month wise trend

SELECT

SUBSTRING\_INDEX(date\_published, '-', -1) AS year,

SUBSTRING\_INDEX(SUBSTRING\_INDEX(date\_published, '-', 2), '-', -1) AS month,

COUNT(\*) AS total\_movies

FROM

amitdb.movies

WHERE

date\_published IS NOT NULL

GROUP BY

SUBSTRING\_INDEX(date\_published, '-', -1), SUBSTRING\_INDEX(SUBSTRING\_INDEX(date\_published, '-', 2), '-', -1)

ORDER BY

year, month;

1. Calculate the number of movies produced in the USA or India in the year 2019.

Answer:

SELECT

COUNT(\*) AS total\_movies

FROM

amitdb.movies

WHERE

(country = 'USA' OR country = 'India')

AND year = 2019;

Segment 3: Production Statistics and Genre Analysis

-- (a) Retrieve the unique list of genres present in the dataset.

SELECT DISTINCT genre

FROM amitdb.genre;

-- (b) Identify the genre with the highest number of movies produced overall.

SELECT

genre,

COUNT(movie\_id) AS movie\_count

FROM

amitdb.genre

GROUP BY

genre

ORDER BY

movie\_count DESC

LIMIT 1;

-- (c) Determine the count of movies that belong to only one genre.

SELECT COUNT(\*) AS single\_genre\_count

FROM (

SELECT movie\_id

FROM amitdb.genre

GROUP BY movie\_id

HAVING COUNT(DISTINCT genre) = 1

) AS single\_genre\_movies;

-- (d) Calculate the average duration of movies in each genre.

SELECT

genre,

AVG(duration) AS average\_duration

FROM

amitdb.genre

JOIN amitdb.movies ON amitdb.genre.movie\_id = amitdb.movies.id

GROUP BY

genre;

-- (e) Find the rank of the 'thriller' genre among all genres in terms of the number of movies produced.

SELECT

genre,

RANK() OVER (ORDER BY movie\_count DESC) AS genre\_rank

FROM (

SELECT

genre,

COUNT(movie\_id) AS movie\_count

FROM

amitdb.genre

GROUP BY

genre

) AS genre\_counts

WHERE

genre = 'Thriller';

Segment 4: Ratings Analysis and Crew Members

* Retrieve the minimum and maximum values in each column of the ratings table (except movie\_id).
* Identify the top 10 movies based on average rating.
* Summarise the ratings table based on movie counts by median ratings.
* Identify the production house that has produced the most number of hit movies (average rating > 8).
* Determine the number of movies released in each genre during March 2017 in the USA with more than 1,000 votes.
* Retrieve movies of each genre starting with the word 'The' and having an average rating > 8.

Segment 5: Crew Analysis

* Identify the columns in the names table that have null values.
* Determine the top three directors in the top three genres with movies having an average rating > 8.
* Find the top two actors whose movies have a median rating >= 8.
* Identify the top three production houses based on the number of votes received by their movies.
* Rank actors based on their average ratings in Indian movies released in India.
* Identify the top five actresses in Hindi movies released in India based on their average ratings.

Segment 6: Broader Understanding of Data

1. Classify thriller movies based on average ratings into different categories.
2. analyse the genre-wise running total and moving average of the average movie duration.
3. Identify the five highest-grossing movies of each year that belong to the top three genres.
4. Determine the top two production houses that have produced the highest number of hits among multilingual movies.
5. Identify the top three actresses based on the number of Super Hit movies (average rating > 8) in the drama genre.
6. Retrieve details for the top nine directors based on the number of movies, including average inter-movie duration, ratings, and more.

Segment 7: Recommendations

* Based on the analysis, provide recommendations for the types of content Bolly movies should focus on producing.