IMDB MOVIES ANALYSIS

USING SQL

Done by - Nileaysh Jadhav

Phone – 9422173500

Email – [nileshjadhav991@gmail.com](mailto:nileshjadhav991@gmail.com)

INDEX

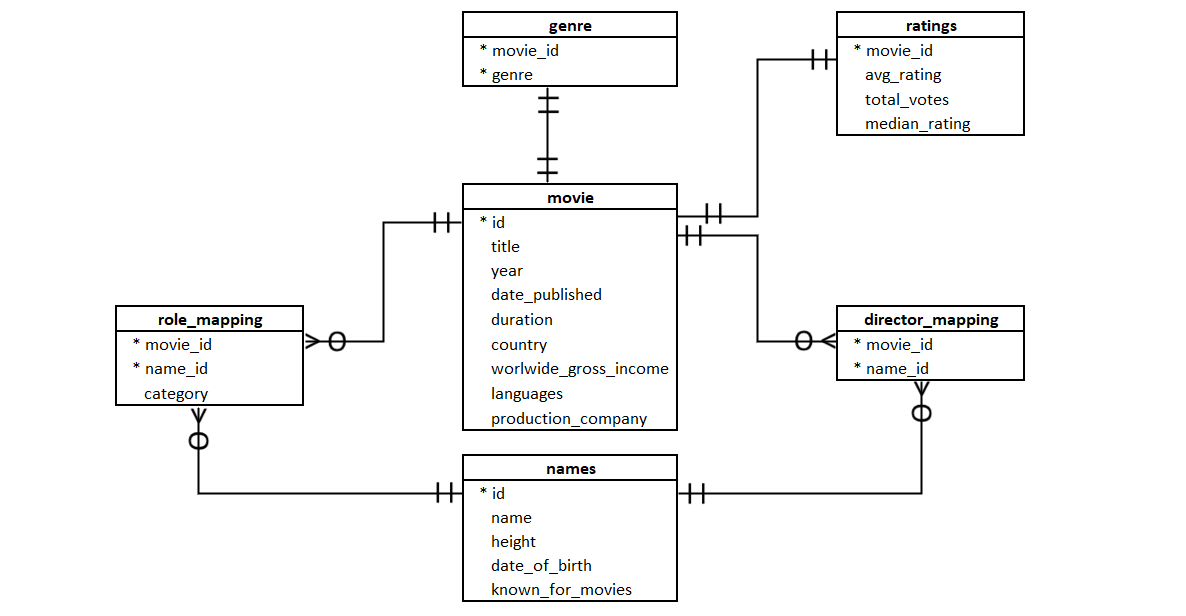
* Objective 03
* Segment 1: Database - Tables, Columns, Relationships 04
* Segment 2: Movie Release Trends 07
* Segment 3: Production Statistics and Genre Analysis 09
* Segment 4: Ratings Analysis and Crew Members 12
* Segment 5: Crew Analysis 16
* Segment 6: Broader Understanding of Data 22
* Segment 7: Recommendations 29

Objective :

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

**Q) What are the different tables in the database and how are they connected to each other in the database?**

****

**Result:**

There are total 6 tables in database namely Movies, Genre, Ratings, Role\_mapping, Names, Director\_Mapping

1. movie: This table stores information about movies. It has a primary key `id` and columns such as `title`, `year`, `date\_published`, `duration`, `country`, `worlwide\_gross\_income`, `languages`, and `production\_company`.

2. genre: This table represents the genres of movies. It has a composite primary key `(movie\_id, genre)’ and contains the movie ID and genre for each movie.

3. director\_mapping: This table maps movies to directors. It has a composite primary key `(movie\_id, name\_id)’ and contains the movie ID and director name for each movie.

4. role\_mapping: This table maps movies to actors/actresses and their roles. It has a composite primary key `(movie\_id, name\_id)’ and contains the movie ID, actor/actress name, and role category for each movie.

5. names: This table stores information about people involved in movies, such as actors, actresses, and directors. It has a primary key `id’ and columns like `name’, `height’, `date\_of\_birth’, and `known\_for\_movies’.

6. ratings: This table contains ratings information for movies. It has a primary key `movie\_id’ and columns such as `avg\_rating’, `total\_votes’, and `median\_rating’.

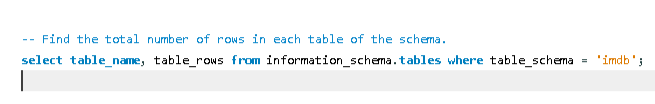
These tables are connected to each other using foreign keys.

The `movie\_id’ column in the genre, director\_mapping, and role\_mapping tables references the `id’ column in the movie table.

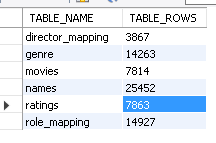
The `name\_id’ column in the director\_mapping and role\_mapping tables references the `id’ column in the names table.

These relationships allow for the association of movies with genres, directors, and actors/actresses.

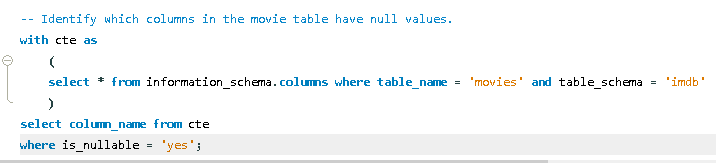
**Q) Find the total number of rows in each table of the schema.**



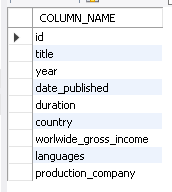
**Result:**

****

**Q) Identify which columns in the movie table have null values.**

****

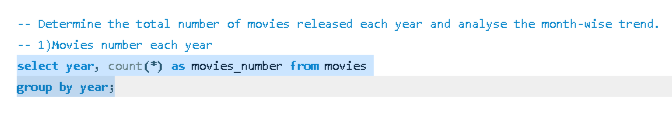
**Result:**

****

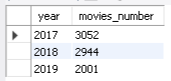
Segment 2: Movie Release Trends

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

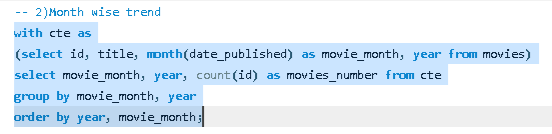
**1) Movies number each year**

****

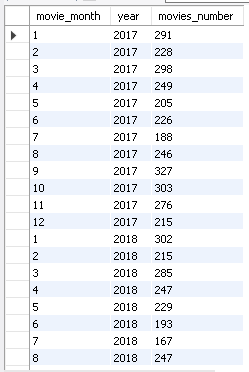
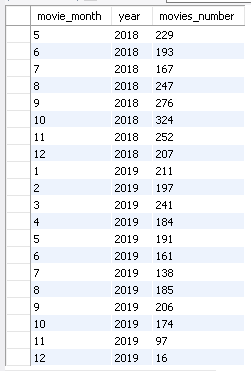
**Result:**

****

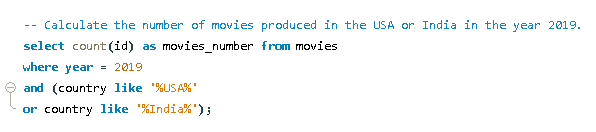
**2) Month wise trend**

****

**Result:**

****

**Q) Calculate the number of movies produced in the USA or India in the year 2019.**

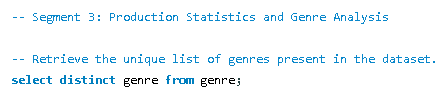
****

**Result:**

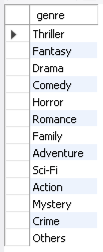
****

Segment 3: Production Statistics and Genre Analysis

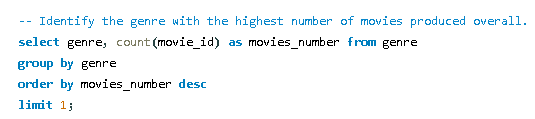
**Q) Retrieve the unique list of genres present in the dataset.**

****

**Result:**

****

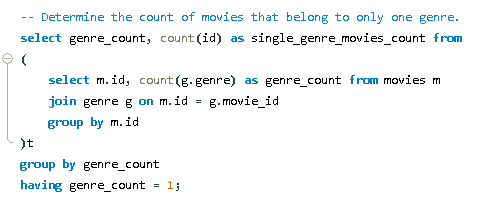
**Q) Identify the genre with the highest number of movies produced overall.**

****

**Result:**

****

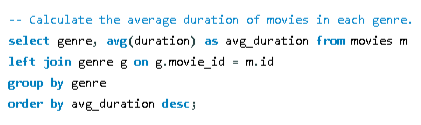
**Q) Determine the count of movies that belong to only one genre.**

****

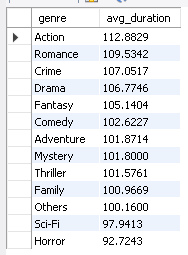
**Result:**

****

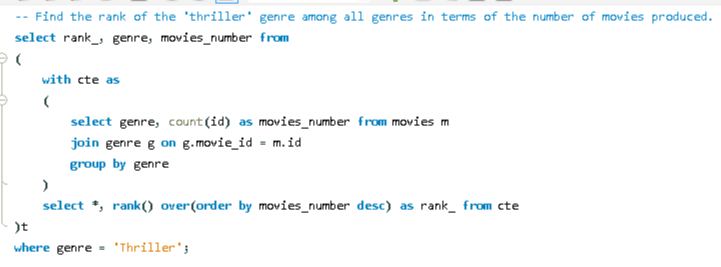
**Q) Calculate the average duration of movies in each genre.**

****

**Result:**

****

**Q) Find the rank of the 'thriller' genre among all genres in terms of the number of movies produced.**

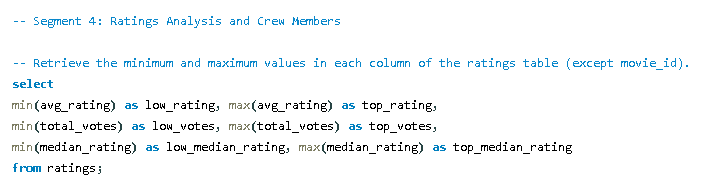
****

**Result:**

****

Segment 4: Ratings Analysis and Crew Members

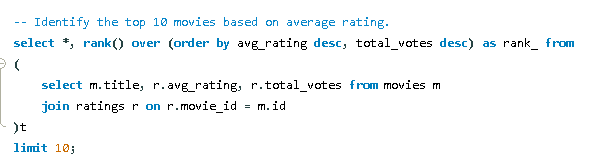
**Q) Retrieve the minimum and maximum values in each column of the ratings table (except movie\_id).**

****

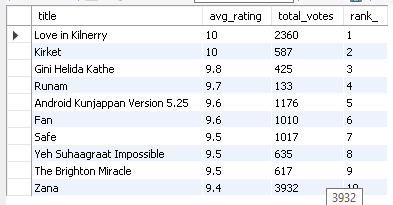
**Result:**

****

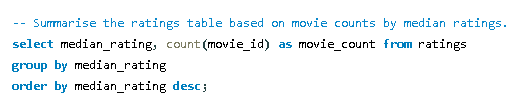
**Q) Identify the top 10 movies based on average rating.**

****

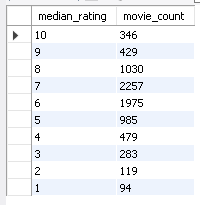
**Result:**

****

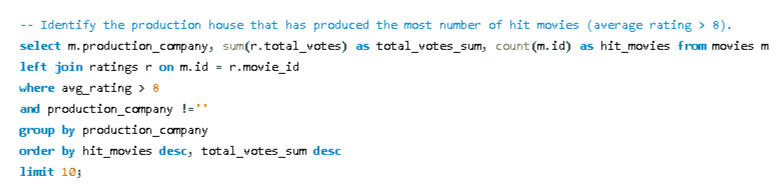
**Q) Summarise the ratings table based on movie counts by median ratings.**

****

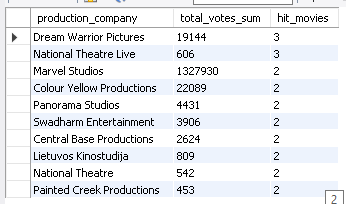
**Result:**

****

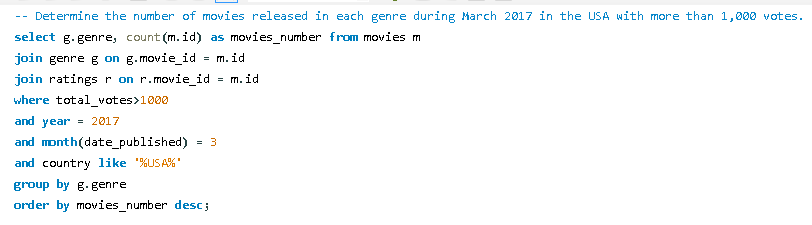
**Q) Identify the production house that has produced the most number of hit movies (average rating > 8).**

****

**Result:**

****

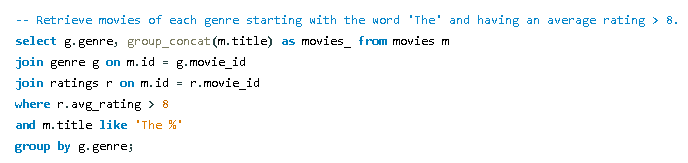
**Q) Determine the number of movies released in each genre during March 2017 in the USA with more than 1,000 votes.**

****

**Result:**

****

**Q) Retrieve movies of each genre starting with the word 'The' and having an average rating > 8.**

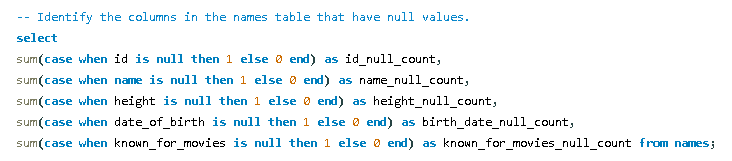
****

**Result:**

****

Segment 5: Crew Analysis

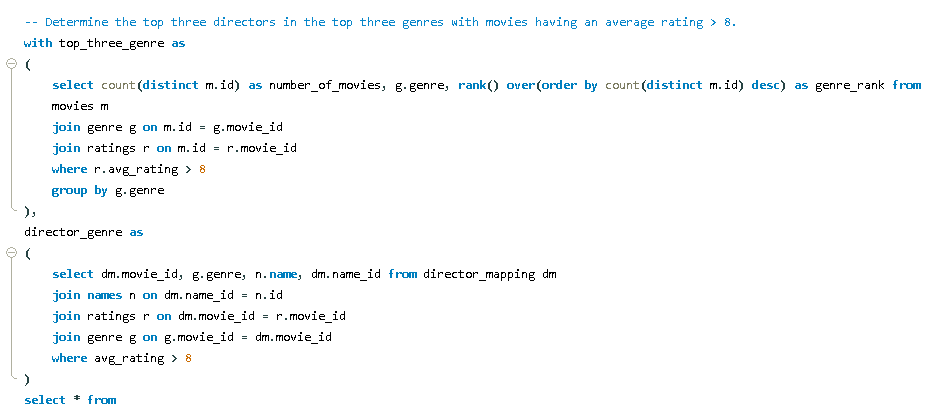
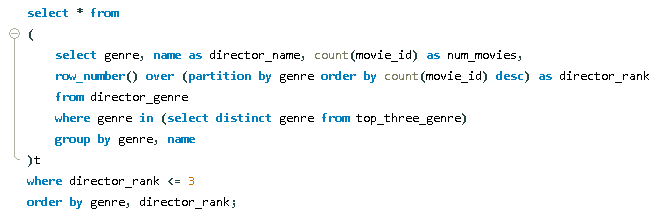
**Q) Identify the columns in the names table that have null values.**

****

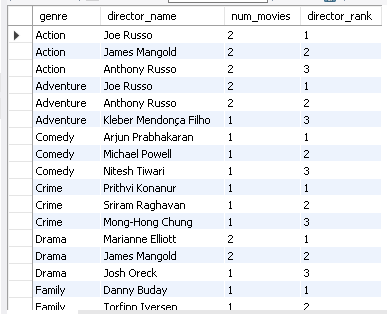
**Result:**

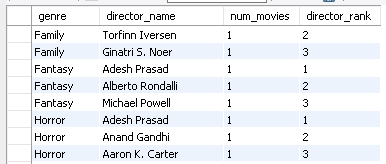
****

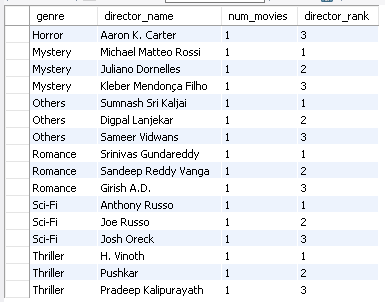
**Q) Determine the top three directors in the top three genres with movies having an average rating > 8.**

****

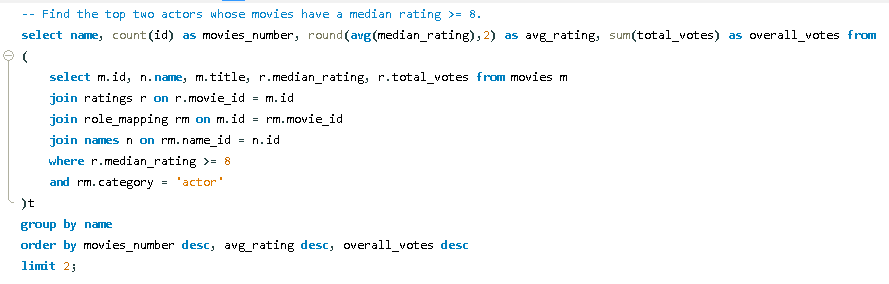
**Result:**

****

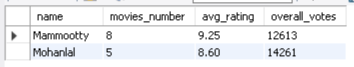
****

****

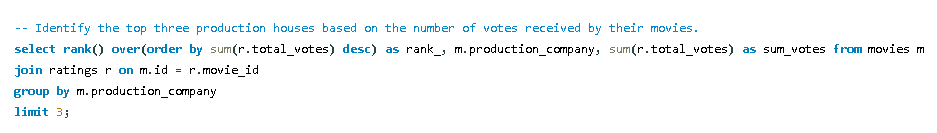
**Q) Find the top two actors whose movies have a median rating >= 8.**

****

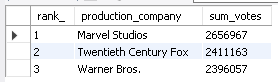
**Result:**

****

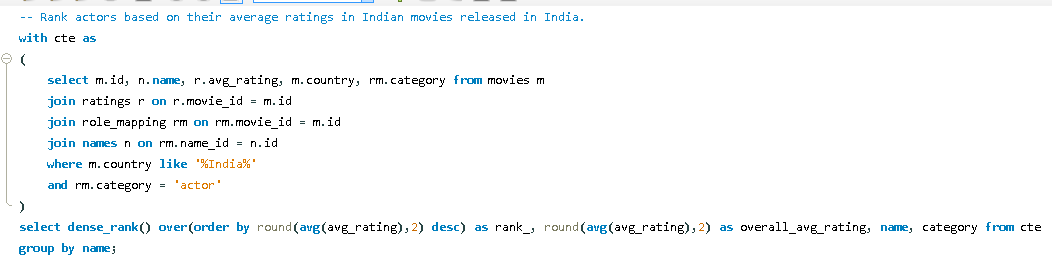
**Q) Identify the top three production houses based on the number of votes received by their movies.**

****

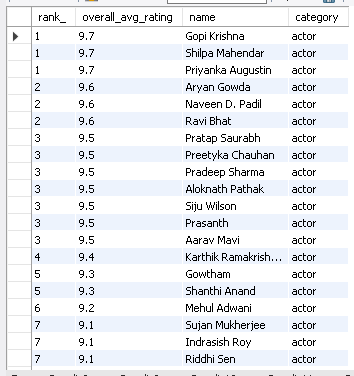
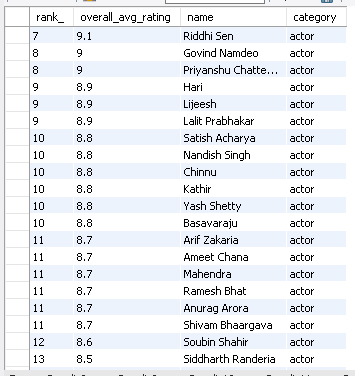
**Result:**

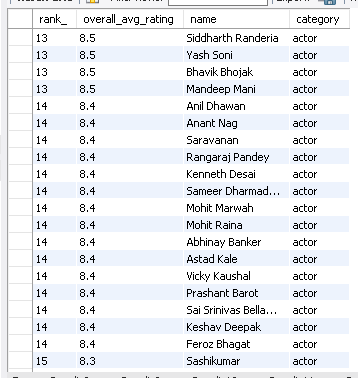
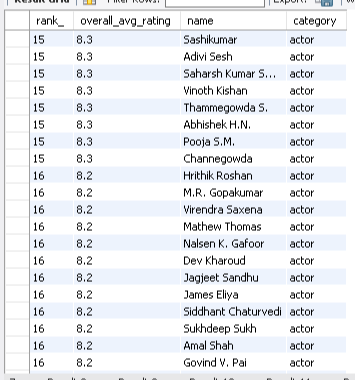
****

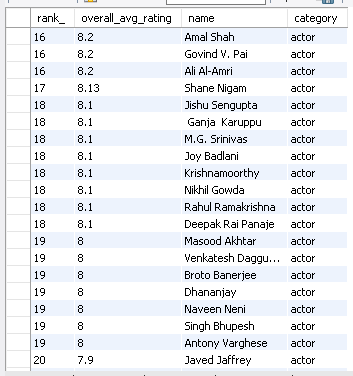
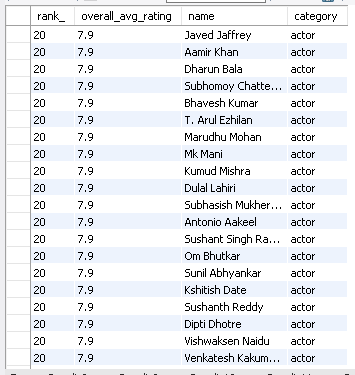
**Q) Rank actors based on their average ratings in Indian movies released in India.**

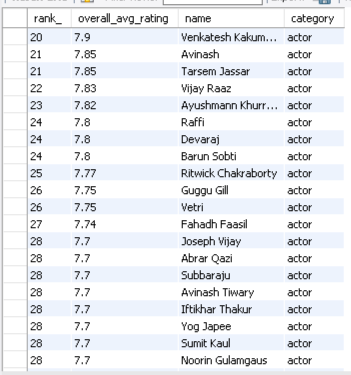
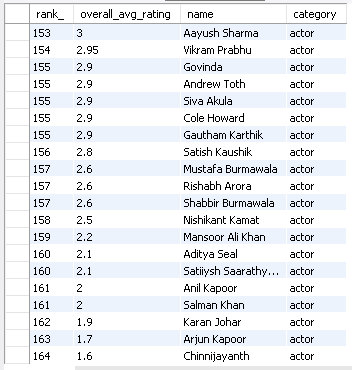
****

**Result:**

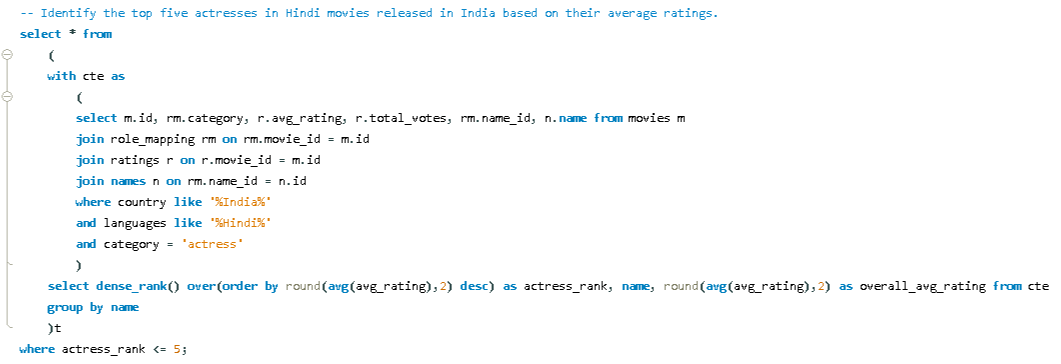
****

****

****

****

**Q) Identify the top five actresses in Hindi movies released in India based on their average ratings.**

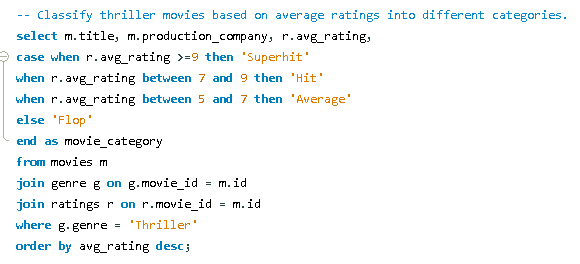
****

**Result:**

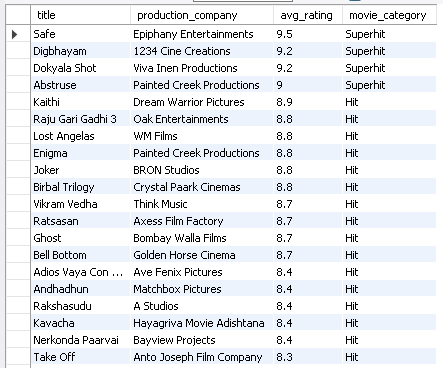
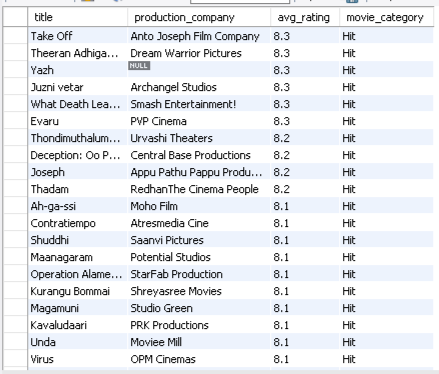
****

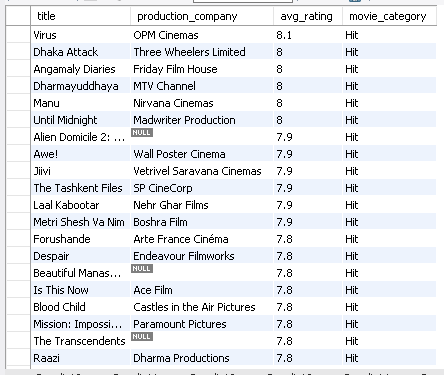
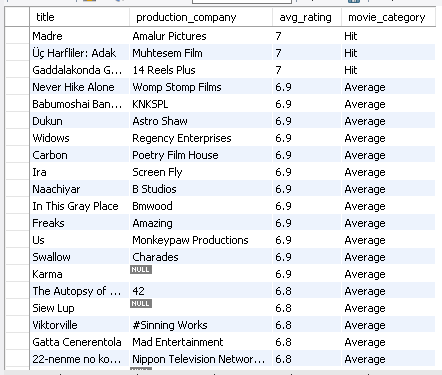
Segment 6: Broader Understanding of Data

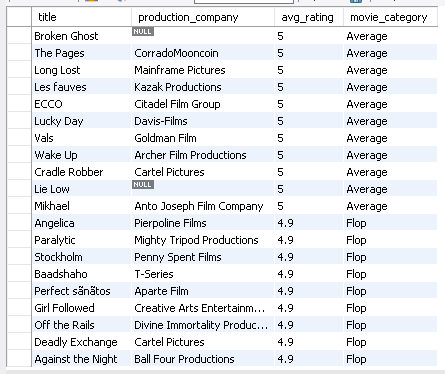
**Q) Classify thriller movies based on average ratings into different categories.**

****

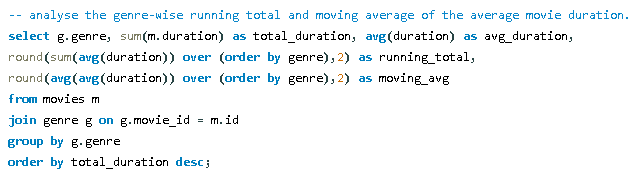
**Result:**

****

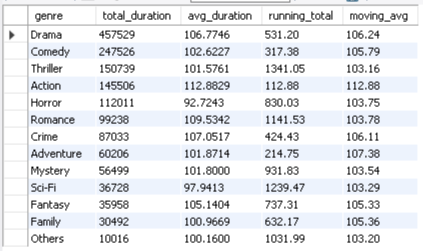
****

****

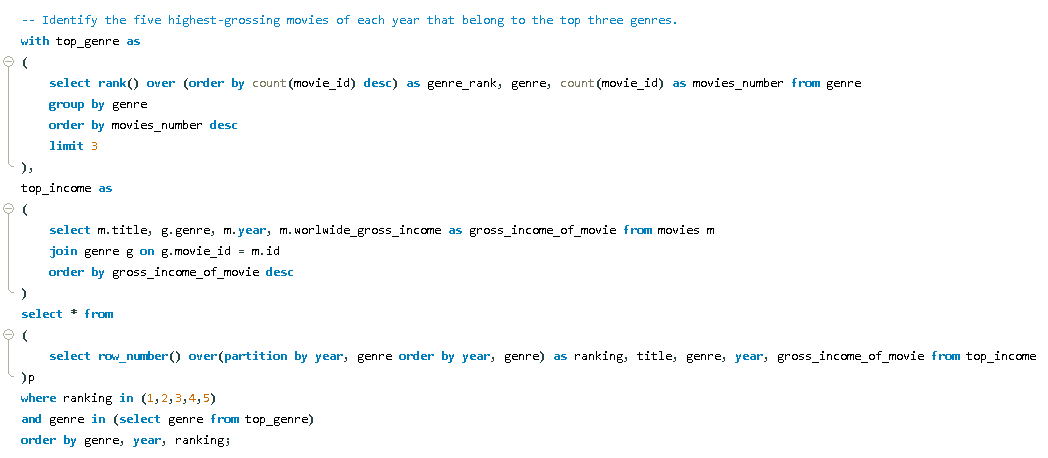
**Q) Analyse the genre-wise running total and moving average of the average movie duration.**

****

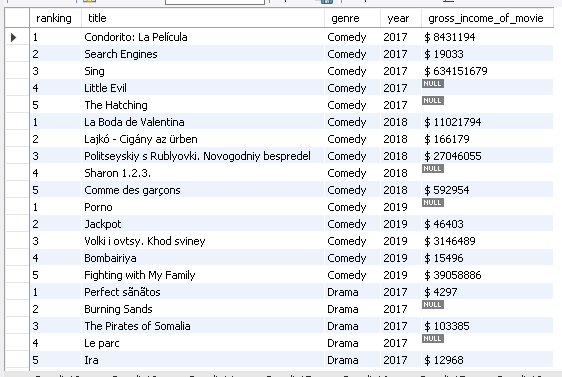
**Result:**

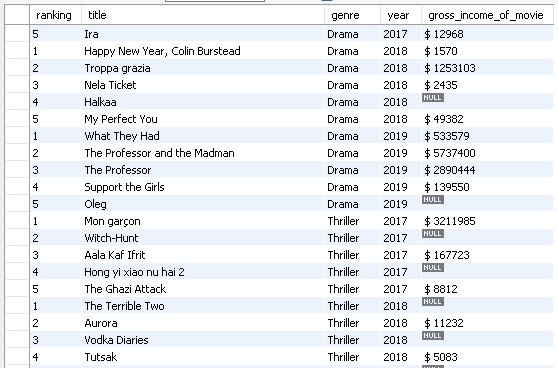
****

**Q) Identify the five highest-grossing movies of each year that belong to the top three genres.**

****

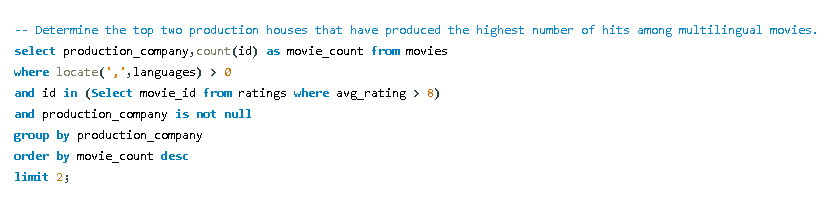
**Result:**

****

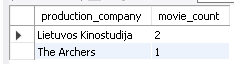
****

****

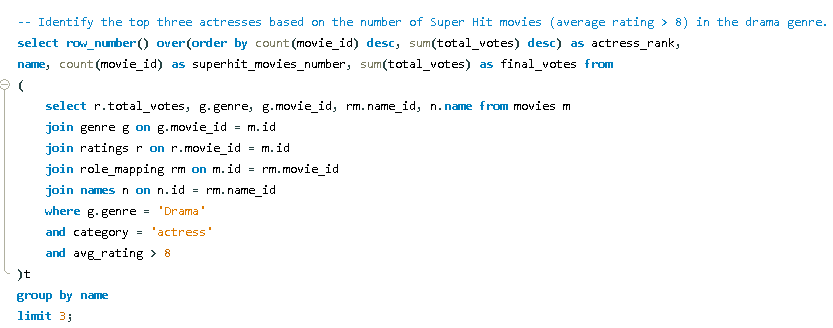
**Q) Determine the top two production houses that have produced the highest number of hits among multilingual movies.**

****

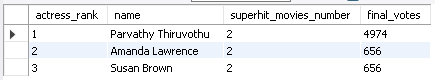
**Result:**

****

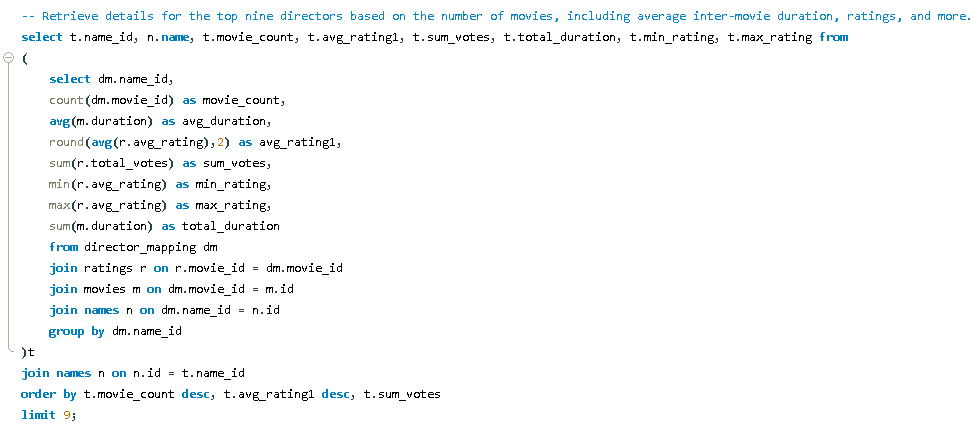
**Q) Identify the top three actresses based on the number of Super Hit movies (average rating > 8) in the drama genre.**

****

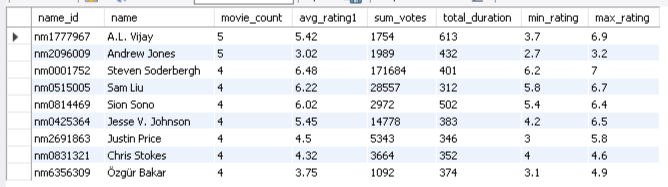
**Result:**

****

**Q) Retrieve details for the top nine directors based on the number of movies, including average inter-movie duration, ratings, and more.**

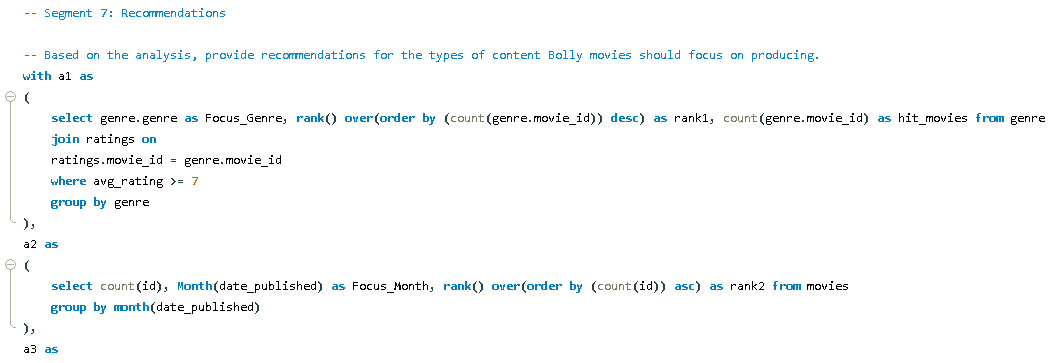
****

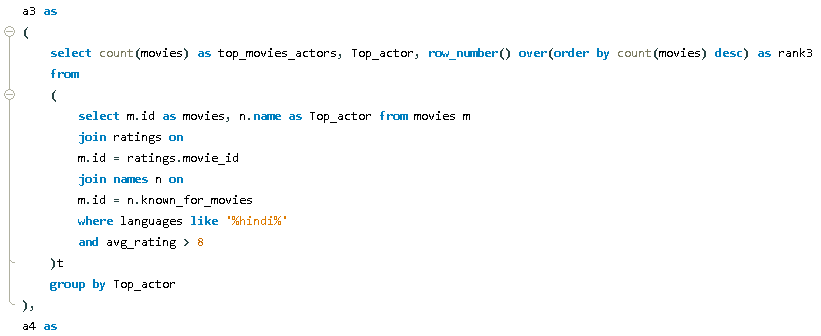
**Result:**

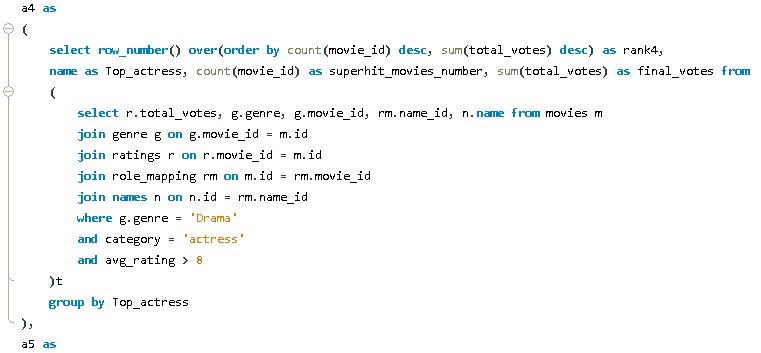
****

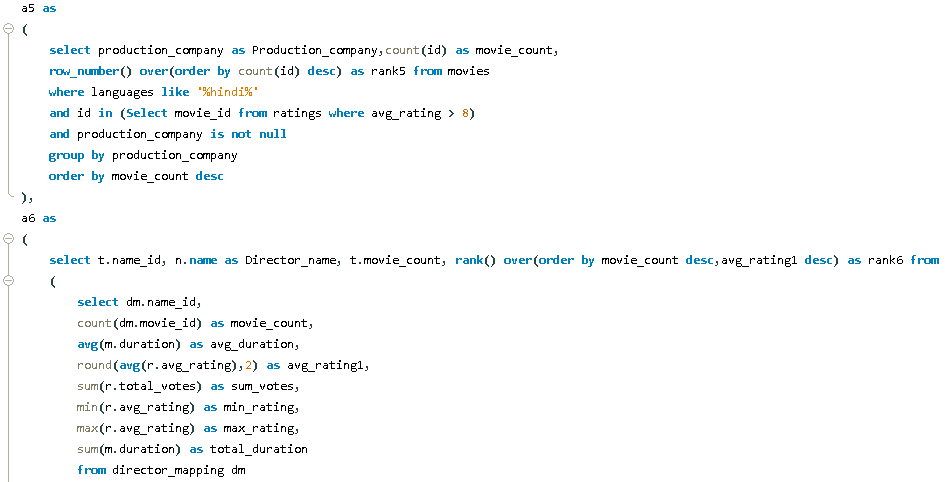
Segment 7: Recommendations

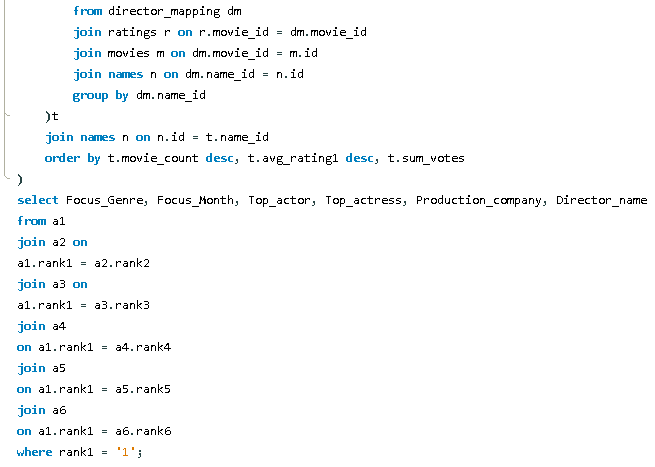
**Q) Based on the analysis, provide recommendations for the types of content Bolly movies should focus on producing.**

****

****

****

****

****

**Result:**

****

* **Diversification of Genre Portfolio:** Given the evident preference for drama within the Bollywood audience, it is advisable for the industry to strategically expand its repertoire of high-quality drama productions. By exploring diverse narrative themes and cinematic styles within this genre, Bollywood can further engage its audience base and foster sustained interest.
* **Strategic Utilization of Talent Resources:** The profound appeal and multifaceted talent of actors like Hrithik Roshan and actress like Parvathy Thiruvothu, coupled with Directors like A.L. Vijay's exemplary track record in filmmaking, present invaluable assets for Bollywood's creative endeavors. It is recommended that industry stakeholders conscientiously leverage their widespread popularity and exceptional acting prowess by offering them a spectrum of roles across various genres. Additionally, collaborating with Director A.L. Vijay on upcoming projects offers an opportunity to infuse Bollywood productions with innovative storytelling and directorial expertise. By strategically tapping into the diverse talents of these individuals, Bollywood can optimize its potential to captivate and resonate with a broad spectrum of audience segments while elevating the caliber of its offerings.
* **Strategic Movie Release Timing:** In light of historical data indicating a notable decrease in movie releases during November and December 2019, attributed to the impact of the COVID-19 pandemic, it is evident that external factors can influence industry dynamics. Moreover, analysis reveals that July typically witnesses a reduced number of competing releases. Therefore, Bollywood stands to benefit from strategically scheduling movie premieres during July, leveraging the comparatively lower competition to maximize audience engagement and box office success. Additionally, a thorough understanding of past trends, including the influence of significant events such as the COVID-19 pandemic, is essential in formulating effective release strategies tailored to current industry conditions.
* **Strategic Partnerships with Esteemed Production Entities:** Collaborative ventures with established production entities such as Colour Yellow Productions hold significant potential for enhancing the promotional efficacy and commercial viability of Bollywood ventures. Leveraging the robust networks and campaign strategies of such esteemed partners can effectively augment the visibility and market penetration of film releases, thus fostering a conducive environment for sustained industry success.