Analyzing Film Data to Guide RSVP Movies

Introduction:
RSVP Movies, a leading Indian film production house celebrated for its blockbuster films, is launching an exciting new project in 2022 aimed at a worldwide audience. To maximize the potential of this initiative, they are leveraging data analytics. In my role as a data analyst and SQL specialist, I have been assigned the responsibility of examining movie data from the last three years to deliver critical insights that will assist RSVP Movies in making informed strategic choices.

Segmented Analysis:
To achieve this goal, our analytical journey is divided into four segments, each targeting specific aspects of the dataset to derive meaningful insights. These insights will inform RSVP Movies' decisions and ensure a data-driven approach to their new global project.

1. ***Understanding the Data:***

We begin by comprehending the dataset's structure and content. Familiarity with the data is crucial before embarking on any analysis. This segment helps us get acquainted with what each column represents.

2. Box Office Performance Analysis:

To gauge the past success of movies, we delve into box office performance, focusing on revenue, budget, and profit. Through SQL queries, we determine the average budget, identify the top revenue-generating movies, and calculate the total profit for movies released in the past three years.

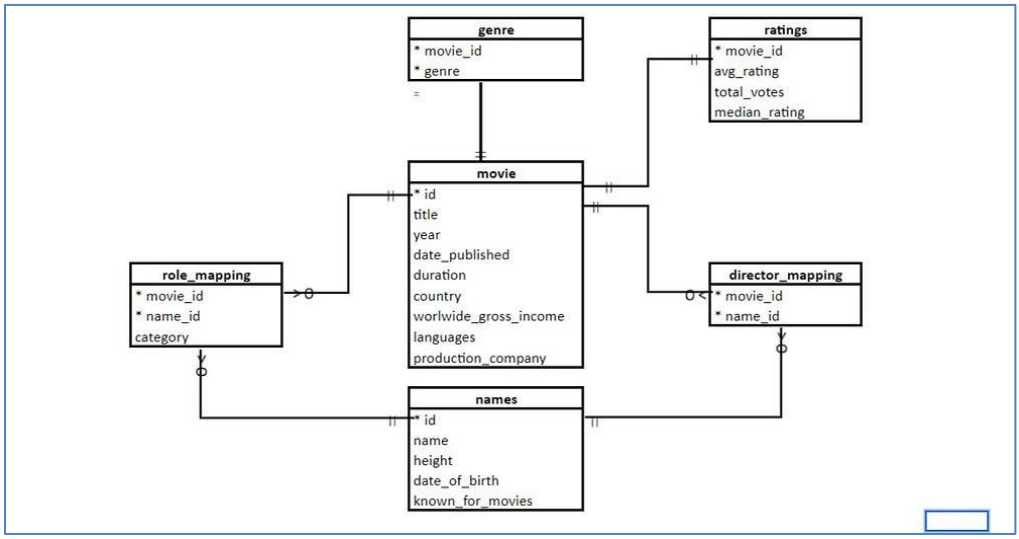
3. Genre Insights:

Understanding the popularity of movie genres is essential. We analyze the dataset to find the most popular movie genre in the past three years based on the number of movies in each genre.

4. Audience Rating Assessment:

To cater effectively to a global audience, we scrutinize audience preferences. We calculate the average audience rating from 2017 -2019 movies, identify the top-rated movies in between this period, and explore the relationship between a movie's collection and its audience rating.

ERD Diagram for RSVP:



We will be using MYSQL workbench for the analysis, Kindly follow the link to download file that contains data to create database.

In this segment, we explore the movie data tables and gather useful information:

Q1. Total Number of Rows in Each Table:

To find the total number of rows in each table within the 'imdb' schema, we execute the following SQL query:

```
14    -- Q1. Find the total number of rows in each table of the schema?
15    -- Type your code below:
16    SELECT 'director_mapping' AS TableName, COUNT(*) AS RowCount FROM director_mapping
17    UNION ALL
18    SELECT 'genre' AS TableName, COUNT(*) AS RowCount FROM genre
19    UNION ALL
20    SELECT 'movie' AS TableName, COUNT(*) AS RowCount FROM movie
21    UNION ALL
22    SELECT 'names' AS TableName, COUNT(*) AS RowCount FROM names
23    UNION ALL
24    SELECT 'ratings' AS TableName, COUNT(*) AS RowCount FROM ratings
25    UNION ALL
26    SELECT 'role_mapping' AS TableName, COUNT(*) AS RowCount FROM role_mapping;
27
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

TableName	RowCount
director_mapping	3867
genre	14662
movie	7997
names	25735
ratings	7997
role_mapping	15615

This query provides the number of rows for each table in the schema, helping us understand the dataset's size.

Q2. Identifying Columns with Null Values in the 'movie' Table:

We inspect the 'movie' table to identify which columns contain null values using the following query:

```
-- Q2. Which columns in the movie table have null values?
SELECT
    SUM(CASE WHEN id IS NULL THEN 1 ELSE 0 END) AS id_null_count,
    SUM(CASE WHEN title IS NULL THEN 1 ELSE 0 END) AS title_null_count,
    SUM(CASE WHEN year IS NULL THEN 1 ELSE 0 END) AS year_null_count,
    SUM(CASE WHEN date_published IS NULL THEN 1 ELSE 0 END) AS date_published_null_count,
    SUM(CASE WHEN duration IS NULL THEN 1 ELSE 0 END) AS duration_null_count,
    SUM(CASE WHEN country IS NULL THEN 1 ELSE 0 END) AS country_null_count,
    SUM(CASE WHEN worlwide_gross_income IS NULL THEN 1 ELSE 0 END) AS worlwide_gross_income_null_count,
    SUM(CASE WHEN languages IS NULL THEN 1 ELSE 0 END) AS languages_null_count,
    SUM(CASE WHEN production_company IS NULL THEN 1 ELSE 0 END) AS production_company_null_count
FROM movie;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

ate_published_null_count	duration_null_count	country_null_count	worlwide_gross_income_null_count	languages_null_count	production_company_null_count
	0	20	3724	194	528

This query calculates the count of null values for various columns in the 'movie' table, providing insights into data quality.

Q3. Number of Movies Released Each Year and Month:

To determine the number of movies released each year, we utilize the following queries:

- For the yearly count:

```
SELECT Year, count(id) as number_of_movies
FROM movie
GROUP BY year
ORDER BY year;
```

Result Grid | Filter Rows: | Export: | Wr

Year	number_of_movies
2017	3052
2018	2944
2019	2001

- For the monthly count:

```
94 •  SELECT Month(date_published) as month_num, count(id) as num_of_movies
95    FROM movie
96    GROUP BY month_num
97    ORDER BY num_of_movies;
98
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

month_num	num_of_movies
12	438
7	493
6	580
11	625
5	625
2	640
8	678
4	680
10	801
1	804
9	809
3	824

The highest number of movies is produced in the month of March.

Q4. Number of Movies Produced in the USA or India in 2019:

To find the number of movies produced in the USA or India in 2019, we execute the following query:

```
113 •  SELECT
114        COUNT(id) AS movie_count
115    FROM
116        movie
117    WHERE
118        (country LIKE '%INDIA%' OR country LIKE '%USA%')
119        AND year = 2019;
120
121
```

Result Grid | Filter Rows: | Export: | Wrap Cell Conten

movie_count
1059

This query provides the count of movies produced in either the USA or India during 2019.

Q5: Discovering Unique Movie Genres

Let's begin our quest by finding out the unique list of genres present in the dataset. This will set the stage for our exploration.

```
137 •   SELECT DISTINCT genre
138     FROM genre;
139
```

Result Grid | Filter Rows: | Export

genre
Drama
Fantasy
Thriller
Comedy
Horror
Family
Romance
Adventure
Action
Sci-Fi
Crime
Mystery
Others

genre 6

Our investigation reveals that there are a total of 13 distinct movie genres. The cinematic universe is more diverse than you might have imagined!

Q6: Finding the Dominant Genre

One key question for a budding filmmaker is, "Which genre has the highest number of movies produced overall?" We need to know where the competition is most fierce.

```
154 •   SELECT GENRE,
155         COUNT(ID) AS TOTAL_MOVIES
156     FROM MOVIE M
157     INNER JOIN GENRE G ON M.ID = G.MOVIE_ID
158     GROUP BY GENRE
159     ORDER BY TOTAL_MOVIES DESC
160     LIMIT 5;
161
```

Result Grid | Filter Rows: | Export:

GENRE	TOTAL_MOVIES
Drama	4285
Comedy	2412
Thriller	1484
Action	1289
Horror	1208

Drama takes the spotlight with the highest number of movies produced. Perhaps, it's time to contemplate the world of dramatic storytelling.

Q7: Single-Genre Wonders

But wait, there's more to the story. How many movies belong to only one genre? These movies, known as "single-genre wonders," have a distinct charm of their own

```
193  WITH movies_with_one_genre AS (
194      SELECT movie_id
195      FROM genre
196      GROUP BY movie_id
197      HAVING COUNT(DISTINCT genre) = 1
198  )
199  SELECT COUNT(*) AS movies_with_one_genre
200  FROM movies_with_one_genre;
201
202
```

Result Grid | Filter Rows: | Export: | Wrap

movies_with_one_genre
3289

Surprisingly, there are over 3,000 movies associated with a single genre. These films have carved a niche for themselves in the vast world of cinema.

Q8: Average Duration of Movies in Each Genre

The duration of a movie can greatly impact its audience. Let's find out the average duration of movies in each genre, keeping in mind that a movie can belong to multiple genres.

```
214  SELECT genre,
215      ROUND(AVG(duration), 2) as avg_duration
216  FROM movie m
217  INNER JOIN genre g
218  ON m.id = g.movie_id
219  GROUP BY genre;
220
221
```

Result Grid | Filter Rows: | Export: | Wrap Ce

genre	avg_duration
Drama	106.77
Fantasy	105.14
Thriller	101.58
Comedy	102.62
Horror	92.72
Family	100.97
Romance	109.53
Adventure	101.87
Action	112.88
Sci-Fi	97.94
Crime	107.05
Mystery	101.80
Others	100.16

For example, movies of the ‘Drama’ genre, which were produced in the highest numbers in 2019, have an average duration of approximately 106.77 minutes. Filmmakers can use this insight to plan their next project.

Q9: Ranking the ‘Thriller’ Genre

Speaking of genres, where does the ‘Thriller’ genre stand in terms of the number of movies produced? Let’s find out by using the Rank function.

```
250  WITH genre_rank AS (
251      SELECT genre,
252          COUNT(ID) AS movie_count,
253          ROW_NUMBER() OVER(ORDER BY COUNT(ID) DESC) AS genre_rank
254      FROM MOVIE M
255      INNER JOIN GENRE G
256      ON M.ID = G.MOVIE_ID
257      GROUP BY GENRE )
258  SELECT * FROM genre_rank
259  WHERE genre = 'Thriller';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

genre	movie_count	genre_rank
Thriller	1484	3

It turns out that the ‘Thriller’ genre ranks among the top 3 genres with the most movies. Filmmakers seeking an exciting genre should keep their eyes on ‘Thriller.’

Q10: Exploring Ratings — Min and Max Values

To ensure the data is clean and free of outliers, we start by finding the minimum and maximum values in each column of the ratings table. These statistics help us understand the data’s range.

```
293  SELECT
294      MIN(avg_rating) AS min_avg_rating, MAX(avg_rating) AS max_avg_rating,
295      MIN(total_votes) AS min_total_votes, MAX(total_votes) AS max_total_votes,
296      MIN(median_rating) AS min_median_rating, MAX(median_rating) AS max_median_rating
297  FROM ratings;
298
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

min_avg_rating	max_avg_rating	min_total_votes	max_total_votes	min_median_rating	max_median_rating
1.0	10.0	100	725138	1	10

As we can see, the data falls within the expected range, indicating the absence of significant outliers.

Q11: The Top 10 Movies by Average Rating

Now, it's time to explore the most highly rated movies. We identify the top 10 movies based on their average rating.

```sql
-- Q11. Which are the top 10 movies based on average rating?
WITH RankedMovies AS (
    SELECT m.title, r.avg_rating,
           ROW_NUMBER() OVER (ORDER BY r.avg_rating DESC) AS movie_rank
    FROM ratings r
    INNER JOIN movie m ON r.movie_id = m.id
)
SELECT title, avg_rating, movie_rank
FROM RankedMovies
WHERE movie_rank <= 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

title	avg_rating	movie_rank
Kirket	10.0	1
Love in Kilnerry	10.0	2
Gini Helida Kathe	9.8	3
Runam	9.7	4
Fan	9.6	5
Android Kunjappan Version 5.25	9.6	6
Yeh Suhaagraat Impossible	9.5	7
Safe	9.5	8
The Brighton Miracle	9.5	9
Shibu	9.4	10

Your favorite movie 'Fan' might just make it to the list, with an impressive average rating of 9.6!

Q12: Summarizing Ratings by Median Rating

To understand the distribution of ratings and their impact on movie counts, we summarize the ratings table based on movie counts by median ratings.

```
363 •  SELECT
364        median_rating,
365        COUNT(m.id) AS movie_count
366    FROM ratings r
367    INNER JOIN movie m ON r.movie_id = m.id
368    GROUP BY median_rating
369    ORDER BY movie_count DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

median_rating	movie_count
7	2257
6	1975
8	1030
5	985
4	479
9	429
10	346
3	283
2	119
1	94

Movies with a median rating of 7 seem to be the most numerous. This information can be invaluable for movie production planning.

Q13: The Most Successful Production House

Next, we delve into the production companies behind the movies. We aim to find the production house that has produced the most hit movies (with an average rating > 8).

```
404 •  SELECT production_company, COUNT(m.id) AS movie_count,
405 ⊗      RANK() OVER (ORDER BY COUNT(m.id) DESC) AS prod_company_rank
406    FROM movie m
407    INNER JOIN ratings r
408    ON m.id = r.movie_id
409    WHERE avg_rating > 8 AND production_company IS NOT NULL
410    GROUP BY production_company;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

production_company	movie_count	prod_company_rank
Dream Warrior Pictures	3	1
National Theatre Live	3	1
Lietuvos Kinostudija	2	3
Swadharm Entertainment	2	3
Panorama Studios	2	3
Marvel Studios	2	3
Central Base Productions	2	3
Painted Creek Productions	2	3
National Theatre	2	3
Colour Yellow Productions	2	3

This analysis can guide RSVP Movies in finding potential partners for their next project.

Q14: Genre Analysis — March 2017, USA, and Votes

For a unique perspective, we analyze movies released in each genre during March 2017 in the USA. We focus on movies with more than 1,000 votes.

```
438 •  SELECT genre,
439        COUNT(id) AS movie_count
440    FROM genre g INNER JOIN movie m
441    ON m.id = g.movie_id
442    INNER JOIN ratings r
443    ON m.id = r.movie_id
444    WHERE MONTH(date_published) = 3
445        AND YEAR(date_published) = 2017
446        AND total_votes > 1000
447        AND country LIKE '%USA%'
448    GROUP BY genre
449    ORDER BY movie_count DESC;
```

Result Grid | Filter Rows: | Export: | Wra

genre	movie_count
Drama	24
Comedy	9
Action	8
Thriller	8
Sci-Fi	7
Crime	6
Horror	6
Mystery	4
Romance	4
Fantasy	3
Adventure	3
Family	1

Result 25

This information can help filmmakers identify the genres that captivate audiences in the USA.

Q15: Movies with ‘The’ in Their Title and High Average Rating

To add a touch of excitement, we search for movies of each genre that start with the word ‘The’ and have an average rating > 8.

```
478 •  SELECT m.title, r.avg_rating, g.genre
479    FROM movie m
480    INNER JOIN ratings r ON m.id = r.movie_id
481    INNER JOIN genre g ON m.id = g.movie_id
482    WHERE m.title LIKE 'The%' AND r.avg_rating > 8
483    ORDER BY r.avg_rating DESC;
```

Result Grid | Filter Rows: | Export: | Wrap C

title	avg_rating	genre
The Brighton Miracle	9.5	Drama
The Colour of Darkness	9.1	Drama
The Blue Elephant 2	8.8	Drama
The Blue Elephant 2	8.8	Horror
The Blue Elephant 2	8.8	Mystery
The Irishman	8.7	Crime
The Irishman	8.7	Drama
The Mystery of Godliness: The Sequel	8.5	Drama
The Gambinos	8.4	Crime
The Gambinos	8.4	Drama
Theeran Adhigaaram Ondru	8.3	Action
Theeran Adhigaaram Ondru	8.3	Crime
Theeran Adhigaaram Ondru	8.3	Thriller
The King and I	8.2	Drama
The King and I	8.2	Romance

Result 26

This query unveils a collection of movies that blend 'The' with high ratings and diverse genres.

Q16: Of the movies released between 1 April 2018 and 1 April 2019, how many were given a median rating of 8?

```
516 •  SELECT
517        COUNT(id) AS Movie_released_April2018_April2019
518    FROM movie m INNER JOIN ratings r
519    ON m.id = r.movie_id
520    WHERE (date_published BETWEEN '2018-04-01' AND '2019-04-01')
521    AND (median_rating = 8);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Movie_released_April2018_April2019
361

The answer is that there was a total of 361 movies released between 1 April 2018 and 1 April 2019 with a median rating of 8.

Q17: Do German movies get more votes than Italian movies?

```
533 •   SELECT country,
534         SUM(total_votes) AS total_votes
535     FROM movie AS m
536     INNER JOIN ratings AS r
537     ON m.id = r.movie_id
538     WHERE country IN ('Germany', 'Italy')
539     GROUP BY country;
540
```

Result Grid | Filter Rows: | Export: | Wrap Cell

country	total_votes
Germany	106710
Italy	77965

The answer is that German movies received the highest number of votes, indicating that they received more votes than Italian movies.

Q18: Searching for Null Values

We begin by looking for columns in the “names” table that have null values. This information is crucial for data integrity and completeness.

```
573 •   SELECT SUM(CASE WHEN NAME IS NULL THEN 1 ELSE 0 END) as name_nulls,
574             SUM(CASE WHEN HEIGHT IS NULL THEN 1 ELSE 0 END) as HEIGHT_nulls,
575             SUM(CASE WHEN DATE_OF_BIRTH IS NULL THEN 1 ELSE 0 END) as DATE_OF_BIRTH_nulls,
576             SUM(CASE WHEN KNOWN_FOR_MOVIES IS NULL THEN 1 ELSE 0 END) as KNOWN_FOR_MOVIES_nulls
577     FROM NAMES;
578
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

name_nulls	HEIGHT_nulls	DATE_OF_BIRTH_nulls	KNOWN_FOR_MOVIES_nulls
0	17335	13431	15226

We find that there are no null values in the “name” column, ensuring data completeness and accuracy.

Q19: Top Directors in Top Genres

RSVP Movies aims to collaborate with top directors in the top three genres with movies that have an average rating greater than 8. This analysis helps identify potential directors for their upcoming projects.

```
609  WITH Top_Three_Genre AS (
610  SELECT genre, COUNT(m.id) AS Movie_count
611  FROM movie m INNER JOIN genre g ON m.id = g.movie_id
612  INNER JOIN ratings r ON r.movie_id = m.id
613  WHERE avg_rating > 8
614  GROUP BY genre
615  ORDER BY Movie_count DESC LIMIT 3)
616  SELECT n.name AS director_name, COUNT(m.id) AS Movie_count
617  FROM movie m INNER JOIN director_mapping d ON m.id = d.movie_id
618  INNER JOIN names n ON n.id = d.name_id INNER JOIN genre g ON g.movie_id = m.id
619  INNER JOIN ratings r ON m.id = r.movie_id
620  WHERE g.genre IN (SELECT genre FROM Top_Three_Genre)
621  AND avg_rating > 8
622  GROUP BY director_name
623  ORDER BY Movie_count DESC LIMIT 3;
624
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

director_name	Movie_count
James Mangold	4
Joe Russo	3
Anthony Russo	3

The analysis reveals potential directors for RSVP Movies' next projects, based on the highest-rated genres.

Q20: Top Actors with High Median Ratings

Next, we explore the top two actors whose movies have a median rating greater than or equal to 8. This information can help RSVP Movies choose actors who have a strong track record.

```
650  SELECT n.name AS Actor_name,
651      COUNT(m.id) AS Movie_count
652  FROM movie m
653  INNER JOIN ratings r ON m.id = r.movie_id
654  INNER JOIN role_mapping rm ON m.id = rm.movie_id
655  INNER JOIN names n ON n.id = rm.name_id
656  WHERE median_Rating >= 8
657  GROUP BY Actor_name
658  ORDER BY Movie_count DESC LIMIT 2;
659
```

Result Grid | Filter Rows: | Export: | Wrap Cell Con

Actor_name	Movie_count
Mammootty	8
Mohanlal	5

The results may help identify acclaimed actors for upcoming regional projects.

Q21: Top Three Global Production Houses

RSVP Movies plans to partner with global production houses. To find potential partners, we explore the top three production houses based on the number of votes received by their movies.

```
695 •  SELECT production_company,
696        SUM(total_votes) AS Vote_count,
697        ROW_NUMBER() OVER (ORDER BY SUM(total_votes) DESC) AS prod_comp_rank
698    FROM movie m
699    INNER JOIN ratings r ON m.id = r.movie_id
700    GROUP BY production_company
701    ORDER BY Vote_count DESC
702    LIMIT 3;
703
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

production_company	Vote_count	prod_comp_rank
Marvel Studios	2656967	1
Twentieth Century Fox	2411163	2
Warner Bros.	2396057	3

Marvel Studios stands tall as one of the top global production houses, making it a potential partner for RSVP Movies.

Since RSVP Movies is based out of Mumbai, India also wants to wow its local audience. RSVP Movies also wants to hire a few Indian actors for its upcoming project to give a regional feel.
Let's find who these actors could be

Q22: Rank actors with movies released in India based on their average ratings. Which actor is at the top of the list?

```
738 •  SELECT name AS actor_name,
739        SUM(total_votes) AS total_votes,
740        COUNT(m.id) AS movie_count,
741        ROUND(SUM(avg_rating * total_votes) / SUM(total_votes), 2) AS actor_avg_rating,
742        ROW_NUMBER() OVER (ORDER BY  ROUND(SUM(avg_rating * total_votes) / SUM(total_votes), 2) DESC) AS actor_rank
743    FROM names n
744    INNER JOIN role_mapping rm ON n.id = rm.name_id
745    INNER JOIN ratings r ON rm.movie_id = r.movie_id
746    INNER JOIN movie m ON m.id = rm.movie_id
747    WHERE category = "actor" AND country LIKE "%india%"
748    GROUP BY actor_name
749    HAVING movie_count >= 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

actor_name	total_votes	movie_count	actor_avg_rating	actor_rank
Vijay Sethupathi	23114	5	8.42	1
Fahadh Faasil	13557	5	7.99	2
Yogi Babu	8500	11	7.83	3
Joju George	3926	5	7.58	4
Ammy Virk	2504	6	7.55	5
Dileesh Pothan	6235	5	7.52	6
Kunchacko Boban	5628	6	7.48	7
Pankaj Tripathi	40728	5	7.44	8
Rajkummar Rao	42560	6	7.37	9
Dulquer Salmaan	17666	5	7.30	10

Result 35

The top actor is “Vijay Sethupathi.”

Q23: Find out the top five actresses in Hindi movies released in India based on their average ratings.

```
780  SELECT name AS actress_name,
781      SUM(total_votes) AS total_votes,
782      COUNT(m.id) AS movie_count,
783      ROUND(SUM(avg_rating * total_votes) / SUM(total_votes), 2) AS actress_avg_rating,
784      ROW_NUMBER() OVER ( ORDER BY ROUND(SUM(avg_rating * total_votes) / SUM(total_votes), 2) DESC,
785      SUM(total_votes) DESC
786      ) AS actress_rank
787  FROM names n
788  INNER JOIN role_mapping rm ON n.id = rm.name_id
789  INNER JOIN ratings r ON rm.movie_id = r.movie_id
790  INNER JOIN movie m ON m.id = rm.movie_id
791  WHERE category = "actress"
792      AND country LIKE "%india%"
793      AND languages LIKE "%hindi%"
794  GROUP BY actress_name
795  HAVING movie_count >= 3;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

actress_name	total_votes	movie_count	actress_avg_rating	actress_rank
Taapsee Pannu	18061	3	7.74	1
Kriti Sanon	21967	3	7.05	2
Divya Dutta	8579	3	6.88	3
Shraddha Kapoor	26779	3	6.63	4
Kriti Kharbanda	2549	3	4.80	5
Sonakshi Sinha	4025	4	4.18	6

Result 36

TaapSee Pannu?? Really? I guess there’s something wrong with this world :P

Still according to dataset answer is correct.

Q24: Select thriller movies based on average rating and classify them into categories:

```
836  SELECT title, genre, avg_rating,
837      CASE
838          WHEN avg_rating > 8 THEN 'Superhit movies'
839          WHEN avg_rating BETWEEN 7 AND 8 THEN 'Hit movies'
840          WHEN avg_rating BETWEEN 5 AND 7 THEN 'One-time-watch movies'
841          WHEN avg_rating < 5 THEN 'Flop movies' END AS rating_category
842  FROM movie AS m
843  INNER JOIN genre AS g ON m.id = g.movie_id
844  INNER JOIN ratings AS r ON r.movie_id = m.id
845  WHERE genre = 'Thriller';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

title	genre	avg_rating	rating_category
Der müde Tod	Thriller	7.7	Hit movies
Fahrenheit 451	Thriller	4.9	Flop movies
Pet Sematary	Thriller	5.8	One-time-watch movies
Dukun	Thriller	6.9	One-time-watch movies
Back Roads	Thriller	7.0	Hit movies
Countdown	Thriller	5.4	One-time-watch movies
Staged Killer	Thriller	3.3	Flop movies
Vellaipookal	Thriller	7.3	Hit movies
Uriyadi 2	Thriller	7.3	Hit movies
Incitement	Thriller	7.5	Hit movies
Rakshasudu	Thriller	8.4	Superhit movies

Result 37

Q25: What is the genre-wise running total and moving average of the average movie duration?

```
SELECT genre,
    ROUND(AVG(duration), 2) AS avg_duration,
    SUM(ROUND(AVG(duration), 2)) OVER (ORDER BY genre) AS running_total_duration,
    ROUND(AVG(ROUND(AVG(duration), 2))
    OVER (ORDER BY genre ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW), 2) AS moving_avg_duration
FROM movie AS m
INNER JOIN genre AS g ON m.id = g.movie_id
GROUP BY genre
ORDER BY genre;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

genre	avg_duration	running_total_duration	moving_avg_duration
Action	112.88	112.88	112.88
Adventure	101.87	214.75	107.38
Comedy	102.62	317.37	105.79
Crime	107.05	424.42	106.11
Drama	106.77	531.19	106.24
Family	100.97	632.16	105.36
Fantasy	105.14	737.30	105.33
Horror	92.72	830.02	103.75
Mystery	101.80	931.82	103.54
Others	100.16	1031.98	103.20
Romance	109.53	1141.51	103.77
Sci-Fi	97.94	1239.45	103.29
Thriller	101.58	1341.03	103.16

Q26: Which are the five highest-grossing movies of each year that belong to the top three genres?

There were 3 movies with INR value. When analyzed the sheet the corresponding movie worlwide_gross_income were in Dollar. So in the dataset I made changes with respect to those three movies.

-- Select the top three genres with the most number of movies.

```
WITH TOP_3_GENRE
AS
  (SELECT   GENRE
   FROM     GENRE
   GROUP BY GENRE
   ORDER BY COUNT(GENRE) DESC
   LIMIT    3 ),
  TOP_MOVIES
```

```
  TOP_MOVIES
AS(SELECT genre, year, TITLE AS movie_name,
   CAST(REPLACE(IFNULL(WORLWIDE_GROSS_INCOME,0),'$ ','') AS DECIMAL(10)) AS worldwide_gross_income_$,
   ROW_NUMBER() OVER (PARTITION BY YEAR ORDER BY CAST(REPLACE(IFNULL(WORLWIDE_GROSS_INCOME,0),'$ ','')
   AS DECIMAL(10)) DESC) AS movie_rank
   FROM  MOVIE M
   INNER JOIN GENRE G
```

-- Select the top five movies from each of the top three genres for each year based on worldwide gross

```
920      ON M.ID = G.MOVIE_ID
921      WHERE  GENRE IN
922       (SELECT * FROM   TOP_3_GENRE) )
923     SELECT * FROM   TOP_MOVIES
924     WHERE  MOVIE_RANK<=5;
925
```

-- Retrieve the results of the top movies in the top genres.

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

genre	year	movie_name	worldwide_gross_income_$	movie_rank
Thriller	2017	The Fate of the Furious	1236005118	1
Comedy	2017	Despicable Me 3	1034799409	2
Comedy	2017	Jumanji: Welcome to the Jungle	962102237	3
Drama	2017	Zhan lang II	870325439	4
Thriller	2017	Zhan lang II	870325439	5
Drama	2018	Bohemian Rhapsody	903655259	1
Thriller	2018	Venom	856085151	2
Thriller	2018	Mission: Impossible - Fallout	791115104	3
Comedy	2018	Deadpool 2	785046920	4
Comedy	2018	Ant-Man and the Wasp	622674139	5
Drama	2019	Avengers: Endgame	2797800564	1
Drama	2019	The Lion King	1655156910	2
Comedy	2019	Toy Story 4	1073168585	3
Drama	2019	Joker	995064593	4
Thriller	2019	Joker	995064593	5

Q27: Which are the top two production houses that have produced the highest number of hits (median rating >= 8) among multilingual movies?

```
1141  WITH top_production_houses AS (
1142  SELECT production_company, COUNT(*) AS movie_count,
1143  ROW_NUMBER() OVER (ORDER BY COUNT(*) DESC) AS prod_comp_rank
1144  FROM movie m INNER JOIN ratings r ON m.id = r.movie_id
1145  WHERE median_rating >= 8
1146          AND POSITION(',' in languages) > 0 -- Movies with multiple languages
1147          AND production_company IS NOT NULL
1148      GROUP BY production_company)
1149  SELECT production_company, movie_count,prod_comp_rank
1150  FROM top_production_houses
1151  WHERE prod_comp_rank <= 2;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

production_company	movie_count	prod_comp_rank
Star Cinema	7	1
Twentieth Century Fox	4	2

Q28: Who are the top 3 actresses based on the number of Super Hit movies (average rating >8) in the drama genre?

```
1183    SELECT name AS actress_name,
1184        SUM(total_votes) AS total_votes,
1185        COUNT(m.id) AS movie_count,
1186        AVG(avg_rating) AS actress_Avg_rating,
1187        ROW_NUMBER() OVER (ORDER BY count(m.id) DESC) AS actress_rank
1188    FROM names n
1189    INNER JOIN role_mapping rm ON n.id = rm.name_id
1190    INNER JOIN movie m ON m.id = rm.movie_id
1191    INNER JOIN ratings r ON r.movie_id = m.id
1192    INNER JOIN genre g ON g.movie_id = m.id
1193    WHERE avg_rating > 8 AND category = "actress" AND genre = "drama"
1194    GROUP BY actress_name
1195    ORDER BY movie_count desc LIMIT 3;
1196
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

actress_name	total_votes	movie_count	actress_Avg_rating	actress_rank
Parvathy Thiruvothu	4974	2	8.20000	1
Susan Brown	656	2	8.95000	2
Amanda Lawrence	656	2	8.95000	3

Q29: Get details for the top 9 directors (based on the number of movies).

-- First, create a view to calculate the average difference between two movie dates

```
WITH movie_dates AS (
SELECT nm.id AS director_id, nm.name AS director_name, m.id AS movie_id, m.date_published AS movie_date,
LEAD(m.date_published, 1) OVER (PARTITION BY nm.name ORDER BY m.date_published) AS next_movie_date
FROM names nm INNER JOIN director_mapping dm ON nm.id = dm.name_id
INNER JOIN movie m ON dm.movie_id = m.id
)
SELECT director_id, director_name,
AVG(DATEDIFF(next_movie_date, movie_date)) AS avg_inter_movie_days
FROM movie_dates
GROUP BY director_id, director_name;
```

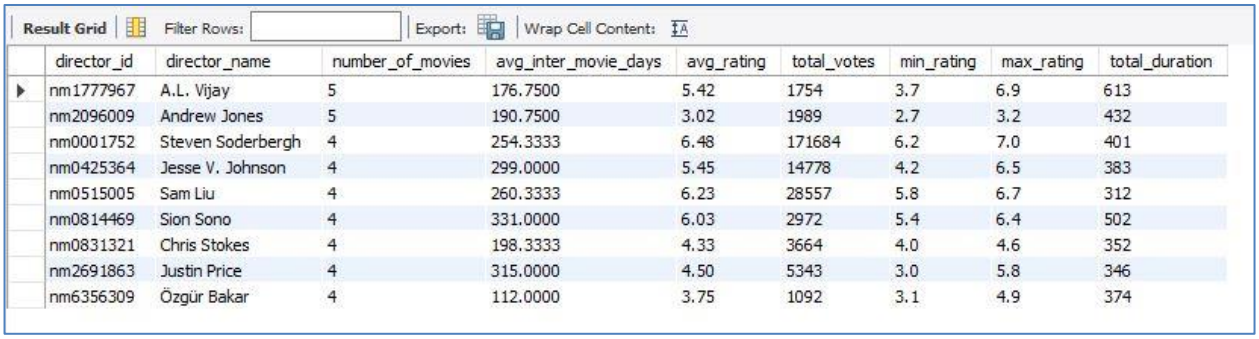
-- Top 9 directors based on the number of movies

```
    SELECT nm.id AS director_id, nm.name AS director_name,
        COUNT(DISTINCT dm.movie_id) AS number_of_movies,
        ROUND(AVG(r.avg_rating), 2) AS avg_rating,
        SUM(r.total_votes) AS total_votes,
        MIN(r.avg_rating) AS min_rating,
        MAX(r.avg_rating) AS max_rating,
        SUM(m.duration) AS total_duration,
        ROW_NUMBER() OVER (ORDER BY COUNT(DISTINCT dm.movie_id) DESC) AS director_rank
FROM names nm
INNER JOIN director_mapping dm ON nm.id = dm.name_id
INNER JOIN movie m ON dm.movie_id = m.id
INNER JOIN ratings r ON m.id = r.movie_id
GROUP BY director_id, director_name
)
```

-- Combine with the average inter-movie days

```
-- Combine with the average inter-movie days
SELECT td.director_id, td.director_name, td.number_of_movies,
    AVGD.avg_inter_movie_days AS avg_inter_movie_days,
    td.avg_rating,td.total_votes, td.min_rating, td.max_rating,
    td.total_duration
FROM top_directors td
LEFT JOIN avg_diff_between_movie_dates AVGD ON td.director_id = AVGD.director_id
WHERE td.director_rank <= 9;
```

--shows the result of the top 9 directors based on their movies



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

director_id	director_name	number_of_movies	avg_inter_movie_days	avg_rating	total_votes	min_rating	max_rating	total_duration
nm1777967	A.L. Vijay	5	176.7500	5.42	1754	3.7	6.9	613
nm2096009	Andrew Jones	5	190.7500	3.02	1989	2.7	3.2	432
nm0001752	Steven Soderbergh	4	254.3333	6.48	171684	6.2	7.0	401
nm0425364	Jesse V. Johnson	4	299.0000	5.45	14778	4.2	6.5	383
nm0515005	Sam Liu	4	260.3333	6.23	28557	5.8	6.7	312
nm0814469	Sion Sono	4	331.0000	6.03	2972	5.4	6.4	502
nm0831321	Chris Stokes	4	198.3333	4.33	3664	4.0	4.6	352
nm2691863	Justin Price	4	315.0000	4.50	5343	3.0	5.8	346
nm6356309	Özgür Bakar	4	112.0000	3.75	1092	3.1	4.9	374

Conclusion:

This analytical exploration using SQL provides RSVP Movies with essential insights. By leveraging data-driven recommendations, they can effectively strategize their global audience-oriented film project. With a solid understanding of their data, RSVP Movies is well-positioned to craft cinematic experiences that connect with viewers around the globe.