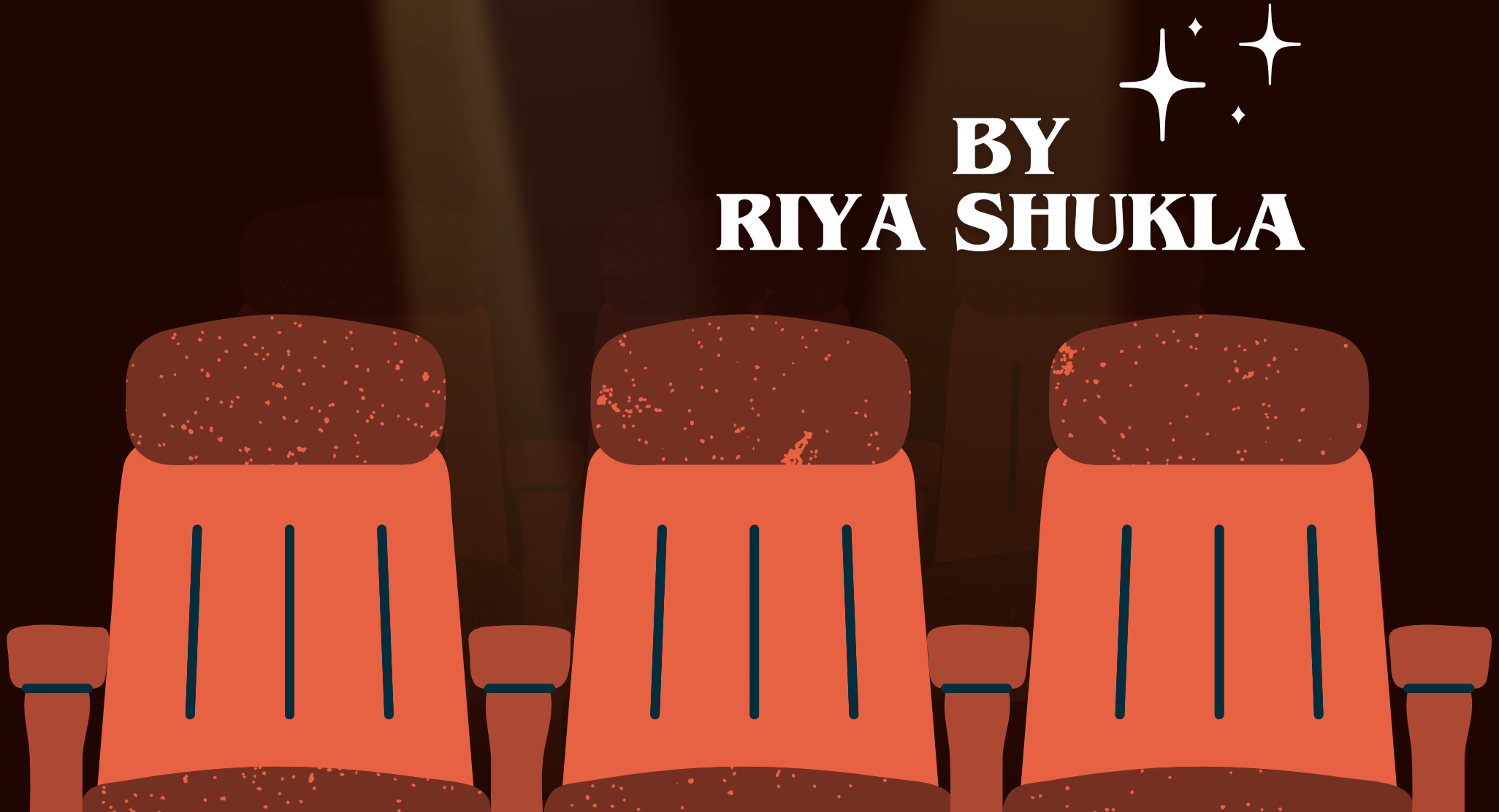
RSVP MOVIES

USING SQL

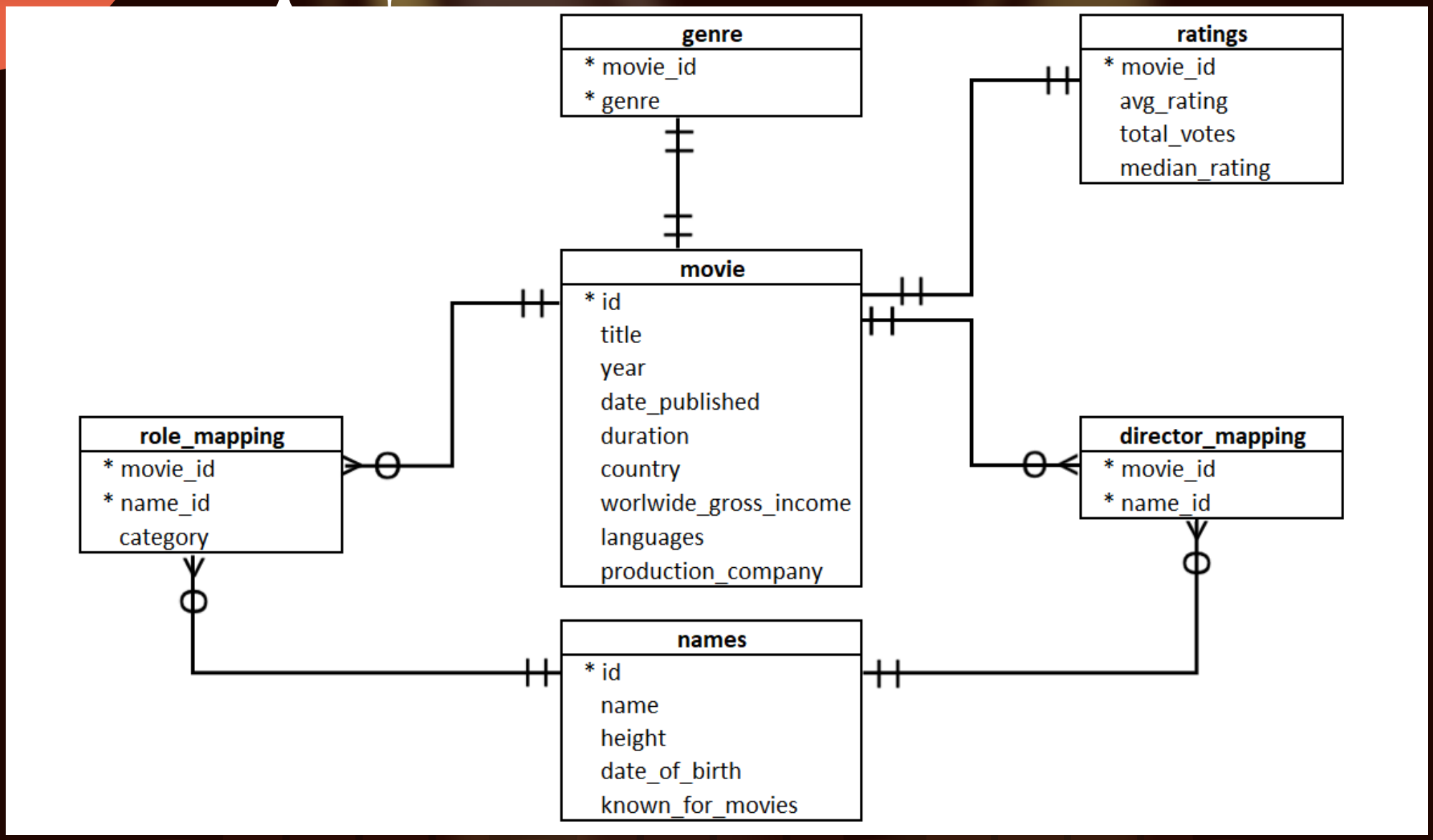
BY
RIYA SHUKLA



OBJECTIVE

To provide RSVP Movies with data-driven insights and recommendations using SQL analysis of the past three years' movie data. The analysis aims to identify key trends, audience preferences, and performance metrics to guide the company's strategic decisions in planning their first movie release targeted at a global audience in 2022. The project will involve segment-wise analysis, covering various aspects of the data to ensure comprehensive insights that align with the company's business objectives.

ERD


genre
* movie_id
* genre
ratings
* movie_id
avg_rating
total_votes
median_rating
movie
* id
title
year
date_published
duration
country
worlwide_gross_income
languages
production_company
role_mapping
* movie_id
* name_id
category
director_mapping
* movie_id
* name_id
names
* id
name
height
date_of_birth
known_for_movies

QUESTIONS

Q1. Find the total number of rows in each table of the schema?
Q2. Which columns in the movie table have null values?
Q3. Find the total number of movies released each year? How does the trend look month wise?
Q4. How many movies were produced in the USA or India in the year 2019?
Q5. Find the unique list of the genres present in the data set?
Q6.Which genre had the highest number of movies produced overall?
Q7. How many movies belong to only one genre?
Q8.What is the average duration of movies in each genre?
Q9.What is the rank of the ‘thriller’ genre of movies among all the genres in terms of number of movies produced?
Q10. Find the minimum and maximum values in each column of the ratings table except the movie_id column?
Q11. Which are the top 10 movies based on average rating?
Q12. Summarise the ratings table based on the movie counts by median ratings.
Q13. Which production house has produced the most number of hit movies (average rating > 8)?
Q14. How many movies released in each genre during March 2017 in the USA had more than 1,000 votes?
Q15. Find movies of each genre that start with the word ‘The’ and which have an average rating > 8?

QUESTIONS

Q16. Of the movies released between 1 April 2018 and 1 April 2019, how many were given a median rating of 8?
Q17. Do German movies get more votes than Italian movies?
Q18. Which columns in the names table have null values?
Q19. Who are the top three directors in the top three genres whose movies have an average rating > 8?
Q20. Who are the top two actors whose movies have a median rating >= 8?
Q21. Which are the top three production houses based on the number of votes received by their movies?
Q22. Rank actors with movies released in India based on their average ratings. Which actor is at the top of the list?
Q23.Find out the top five actresses in Hindi movies released in India based on their average ratings?
Q24. Select thriller movies as per avg rating and classify them in the following category:
Rating > 8: Superhit movies
Rating between 7 and 8: Hit movies
Rating between 5 and 7: One-time-watch movies
Rating < 5: Flop movies
Q25. What is the genre-wise running total and moving average of the average movie duration?
Q26. Which are the five highest-grossing movies of each year that belong to the top three genres?
Q27. Which are the top two production houses that have produced the highest number of hits (median rating >= 8) among multilingual movies?
Q28. Who are the top 3 actresses based on number of Super Hit movies (average rating >8) in drama genre?
Q29. Get the following details for top 9 directors (based on number of movies)
Director id
Name
Number of movies
Average inter movie duration in days
Average movie ratings
Total votes
Min rating
Max rating
total movie durations

Q1. Find the total number of rows in each table of the schema?

```sql
SELECT COUNT(*) as number_of_rows_director_mapping
FROM director_mapping;
SELECT COUNT(*) as number_of_rows_genre
FROM genre;
SELECT COUNT(*) as number_of_rows_movie
FROM movie;
SELECT COUNT(*) as number_of_rows_names
FROM names;
SELECT COUNT(*) as number_of_rows_ratings
FROM ratings;
SELECT COUNT(*) as number_of_role_mapping
FROM role_mapping;
```

number_of_rows_director_mapping
3867

number_of_rows_names
25735

number_of_rows_genre
14662

number_of_rows_ratings
7997

number_of_rows_movie
7997

number_of_role_mapping
15615

Q2. Which columns in the movie table have null values?

```sql
SELECT
    COUNT(CASE WHEN id IS NULL THEN 1 END) AS id_null,
    COUNT(CASE WHEN title IS NULL THEN 1 END) AS title_null,
    COUNT(CASE WHEN year IS NULL THEN 1 END) AS year_null,
    COUNT(CASE WHEN date_published IS NULL THEN 1 END) AS date_published_null,
    COUNT(CASE WHEN duration IS NULL THEN 1 END) AS duration_null,
    COUNT(CASE WHEN country IS NULL THEN 1 END) AS country_null,
    COUNT(CASE WHEN worlwide_gross_income IS NULL THEN 1 END) AS worlwide_gross_income_null,
    COUNT(CASE WHEN languages IS NULL THEN 1 END) AS languages_null,
    COUNT(CASE WHEN production_company IS NULL THEN 1 END) AS production_company_null
FROM
    movie;
```

id_null	title_null	year_null	date_published_null	duration_null	country_null	worlwide_gross_income_null	languages_null	production_company_null
0	0	0	0	0	20	3724	194	528

Q3. Find the total number of movies released each year? How does the trend look month wise?

```sql
SELECT year,
       COUNT(title) as Number_of_movies
FROM movie
GROUP BY year;
```

year	Number_of_movies
2017	3052
2018	2944
2019	2001

```sql
SELECT month(date_published) as month_num,
       COUNT(title) as number_of_movies
FROM movie
GROUP BY month_num
ORDER BY month_num;
```

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

Q4. How many movies were produced in the USA or India in the year 2019?

```sql
SELECT COUNT(id) as number_of_movies
FROM movie
WHERE (country like '%USA%' or
       country like '%India%')
       and year = 2019;
```

number_of_movies
1059

Q5. Find the unique list of the genres present in the data set?

```sql
SELECT distinct(genre) as unique_genre
FROM genre;
```

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

Q6.Which genre had the highest number of movies produced overall?

```sql
SELECT genre,
       count(m.id) as number_of_movies
FROM movie as m
INNER JOIN genre as g
ON m.id = g.movie_id
GROUP BY genre
ORDER BY number_of_movies DESC
limit 1;
```

	genre	number_of_movies
▶	Drama	4285

Q7. How many movies belong to only one genre?

```sql
SELECT COUNT(*) AS num_movies
FROM(
     SELECT movie_id
     FROM genre
     GROUP BY movie_id
     HAVING COUNT(DISTINCT genre)=1)
     as single_genre_movies;
```

	num_movies
▶	3289

Q8.What is the average duration of movies in each genre?

```sql
SELECT genre,
       Round(avg(m.duration),2) as avg_duration
FROM movie as m
INNER JOIN genre as g
ON m.id = g.movie_id
GROUP BY genre
ORDER BY avg_duration DESC;
```

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

Q9.What is the rank of the ‘thriller’ genre of movies among all the genres in terms of number of movies produced?

```sql
WITH genre_summary AS
(
        SELECT     genre,
                   Count(movie_id) as movie_count ,
                   Rank() OVER(ORDER BY Count(movie_id) DESC) as genre_rank
        FROM       genre
        GROUP BY   genre )
SELECT *
FROM   genre_summary
WHERE  genre = "THRILLER" ;
```

genre	movie_count	genre_rank
Thriller	1484	3

Q10. Find the minimum and maximum values in each column of the ratings table except the movie_id column?

```sql
SELECT min(avg_rating) as min_avg_rating,
       max(avg_rating) as max_avg_rating,
       min(total_votes) as min_total_votes,
       max(total_votes) as max_total_votes,
       min(median_rating) as min_median_rating,
       max(median_rating) as max_median_rating
FROM ratings;
```

min_avg_rating	max_avg_rating
1.0	10.0

min_total_votes	max_total_votes
100	725138

min_median_rating	max_median_rating
1	10

Q11. Which are the top 10 movies based on average rating?

```sql
SELECT title,
       avg_rating,
       RANK() OVER( ORDER BY avg_rating DESC) as movie_rank
FROM movie as m
INNER JOIN ratings as r
ON m.id = r.movie_id
GROUP BY title, avg_rating
ORDER BY movie_rank
LIMIT 10;
```

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

Q12. Summarise the ratings table based on the movie counts by median ratings.

```sql
SELECT median_rating,
       count(movie_id) as movie_count
FROM ratings
GROUP BY median_rating
ORDER BY median_rating;
```

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

Q13. Which production house has produced the most number of hit movies (average rating > 8)?

```sql
SELECT production_company,
       movie_count,
       prod_company_rank
FROM (
SELECT production_company,
       count(id) as movie_count,
       RANK() OVER (ORDER BY COUNT(id) DESC) as prod_company_rank
FROM movie as m
INNER JOIN ratings as r
ON m.id = r.movie_id
WHERE r.avg_rating > 8 AND production_company IS NOT NULL
GROUP BY production_company
) as most_hit_movies
WHERE prod_company_rank=1;
```

production_company	movie_count	prod_company_rank
Dream Warrior Pictures	3	1
National Theatre Live	3	1

Q14. How many movies released in each genre during March 2017 in the USA had more than 1,000 votes?.

```sql
SELECT genre,
       count(id) as movie_count
FROM movie as m
INNER JOIN ratings as r
ON m.id=r.movie_id
INNER JOIN genre as g
ON g.movie_id=r.movie_id
WHERE
    country LIKE '%USA%' AND
    year=2017 AND
    monthname(date_published)='March'AND
    total_votes>1000
GROUP BY genre
ORDER BY movie_count DESC;
```

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

Q15. Find movies of each genre that start with the word ‘The’ and which have an average rating > 8?

```sql
SELECT title,
       avg_rating,
       genre
FROM movie as m
INNER JOIN ratings as r
ON m.id=r.movie_id
INNER JOIN genre as g
ON r.movie_id=g.movie_id
WHERE title like 'The%' AND
      avg_rating > 8
GROUP BY title,
         avg_rating,
         genre
ORDER BY avg_rating DESC;
```

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

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

```sql
SELECT count(id) as movies_released,
       median_rating
FROM movie as m
INNER JOIN ratings as r
ON m.id=r.movie_id
WHERE median_rating=8 AND
      date_published BETWEEN '2018-04-01' AND '2019-04-01'
GROUP BY median_rating;
```

movies_released	median_rating
361	8

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

```sql
SELECT country, sum(total_votes) as total_votes
FROM movie AS m
    INNER JOIN ratings as r ON m.id=r.movie_id
WHERE country = 'Germany' or country = 'Italy'
GROUP BY country;
```

country	total_votes
Germany	106710
Italy	77965

Q18. Which columns in the names table have null values?

```sql
SELECT
    SUM(CASE WHEN name is null THEN 1 ELSE 0 END) AS name_nulls,
    SUM(CASE WHEN height is null THEN 1 ELSE 0 END) AS height_nulls,
    SUM(CASE WHEN date_of_birth is null THEN 1 ELSE 0 END) AS date_of_birth_nulls,
    SUM(CASE WHEN known_for_movies is null THEN 1 ELSE 0 END) AS known_for_movies_nulls
FROM names;
```

name_nulls	height_nulls	date_of_birth_nulls	known_for_movies_nulls
0	17335	13431	15226

Q19. Who are the top three directors in the top three genres whose movies have an average rating > 8??

```sql
WITH top_3_genres AS (SELECT genre, count(m.id) as movie_count
FROM movie as m
INNER JOIN genre as g
ON m.id = g.movie_id
INNER JOIN ratings as r
ON m.id = r.movie_id
WHERE avg_rating > 8
GROUP BY genre
ORDER BY movie_count DESC LIMIT 3)
SELECT n.name as director_name, count(id) as movie_count
    FROM director_mapping as d
    INNER JOIN names as n
    ON d.name_id = n.id
    INNER JOIN genre as g
    USING (movie_id)
    INNER JOIN ratings as r
    USING (movie_id)
    INNER JOIN top_3_genres
    USING (genre)
WHERE avg_rating > 8
GROUP BY director_name
ORDER BY movie_count DESC
LIMIT 3
```

director_name	movie_count
James Mangold	4
Anthony Russo	3
Soubin Shahir	3

Q20. Who are the top two actors whose movies have a median rating >= 8?

```sql
SELECT name as actor_name,
       count(movie_id) as movie_count
FROM role_mapping as rm
INNER JOIN names as n
ON n.id = rm.name_id
INNER JOIN ratings as r
USING (movie_id)
WHERE category= 'actor' AND
      median_rating >= 8
GROUP BY actor_name
ORDER BY movie_count DESC
limit 2;
```

actor_name	movie_count
Mammootty	8
Mohanlal	5

Q21. Which are the top three production houses based on the number of votes received by their movies?

```sql
SELECT production_company,
       sum(total_votes) as vote_count,
       RANK() OVER (ORDER BY sum(total_votes) DESC) as prod_comp_rank
FROM movie as m
INNER JOIN ratings as r
ON m.id = r.movie_id
WHERE production_company is not null
GROUP BY production_company
ORDER BY vote_count DESC
LIMIT 3;
```

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

Q22. Rank actors with movies released in India based on their average ratings. Which actor is at the top of the list?
Note: The actor should have acted in at least five Indian movies.

```sql
SELECT
    name as actor_name,
    SUM(total_votes) as total_votes,
    COUNT(m.id) as movie_count,
    ROUND(SUM(avg_rating * total_votes) / SUM(total_votes), 2) as actor_avg_rating,
    ROW_NUMBER() OVER (ORDER BY  ROUND(SUM(avg_rating * total_votes) / SUM(total_votes), 2) DESC) as actor_rank
FROM names n
INNER JOIN role_mapping rm
ON n.id = rm.name_id
INNER JOIN ratings r
ON rm.movie_id = r.movie_id
INNER JOIN movie m
ON m.id = rm.movie_id
WHERE category = "actor" AND
      country LIKE "%india%"
GROUP BY actor_name
HAVING movie_count >= 5;
```

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
Amit Sadh	13355	5	7.21	11
Tovino Thomas	11596	8	7.15	12
Mammootty	12613	8	7.04	13
Nassar	4016	5	7.03	14
Karamjit Anmol	1970	6	6.91	15
Hareesh Kanaran	3196	5	6.58	16

Q23.Find out the top five actresses in Hindi movies released in India based on their average ratings?
Note: The actresses should have acted in at least three Indian movies.

```sql
WITH rank_actress as
(SELECT name as actress_name,
        sum(total_votes) as total_votes,
        count(distinct(movie_id)) as movie_count,
        ROUND(sum(avg_rating * total_votes)/sum(total_votes),2) as actress_avg_rating
FROM ratings as r
INNER JOIN role_mapping as rm
USING (movie_id)
INNER JOIN names as n
ON n.id = rm.name_id
INNER JOIN movie as m
ON m.id = r.movie_id
WHERE category= 'Actress' AND
      country like '%India%' AND
      languages like '%Hindi%'
GROUP BY name_id,name
HAVING count(distinct(movie_id))>=3)
SELECT * ,
   RANK() OVER (ORDER BY actress_avg_rating DESC)  as actress_rank
FROM rank_actress
limit 5;
```

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

Q24. Select thriller movies as per avg rating and classify them in the following category:

Rating > 8: Superhit movies

Rating between 7 and 8: Hit movies

Rating between 5 and 7: One-time-watch movies

Rating < 5: Flop movies

```sql
SELECT title, avg_rating,
       case
       WHEN avg_rating > 8 THEN 'Superhit movies'
       WHEN avg_rating BETWEEN 7 AND 8 THEN 'Hit movies'
       WHEN avg_rating BETWEEN 5 AND 7 THEN ' One-time-watch movies'
       ELSE 'Flop movies'
       END as Category
FROM genre AS g
INNER JOIN ratings AS r
USING (movie_id)
INNER JOIN movie AS m
ON m.id = r.movie_id
WHERE genre like '%Thriller%';
```

title	avg_rating	Category
Der müde Tod	7.7	Hit movies
Fahrenheit 451	4.9	Flop movies
Pet Sematary	5.8	One-time-watch movies
Dukun	6.9	One-time-watch movies
Back Roads	7.0	Hit movies
Countdown	5.4	One-time-watch movies
Staged Killer	3.3	Flop movies
Vellaipookal	7.3	Hit movies
Uriyadi 2	7.3	Hit movies
Incitement	7.5	Hit movies
Rakshasudu	8.4	Superhit movies
Trois jours et ...	6.6	One-time-watch movies
Killer in Law	5.1	One-time-watch movies

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

```sql
SELECT genre,
       ROUND(AVG(duration),2) AS avg_duration,
       SUM(ROUND(AVG(duration),2)) OVER(ORDER BY genre) AS running_total_duration,
       AVG(AVG(duration)) OVER(ORDER BY genre) AS moving_avg_duration
       FROM movie m
       INNER JOIN genre g
       ON m.id = g.movie_id
       GROUP BY genre;
```

genre	avg_duration	running_total_duration	moving_avg_duration
Action	112.88	112.88	112.88290000
Adventure	101.87	214.75	107.37715000
Comedy	102.62	317.37	105.79233333
Crime	107.05	424.42	106.10717500
Drama	106.77	531.19	106.24066000
Family	100.97	632.16	105.36170000
Fantasy	105.14	737.30	105.33008571
Horror	92.72	830.02	103.75436250
Mystery	101.80	931.82	103.53721111
Others	100.16	1031.98	103.19949000
Romance	109.53	1141.51	103.77537273
Sci-Fi	97.94	1239.45	103.28920000
Thriller	101.58	1341.03	103.15742308

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

```sql
WITH top_three_genre AS (
    SELECT genre, COUNT(m.id) AS movie_count
    FROM movie m
    INNER JOIN genre g
    ON g.movie_id = m.id
    GROUP BY genre
    ORDER BY movie_count DESC
    LIMIT 3
),
final_tab AS (
    SELECT
        g.genre, m.year, m.title AS movie_name, worlwide_gross_income,
        ROW_NUMBER() OVER (PARTITION BY m.year ORDER BY worlwide_gross_income DESC) AS movie_rank
    FROM movie m
    INNER JOIN genre g
    ON g.movie_id = m.id
    WHERE g.genre IN (SELECT genre FROM top_three_genre)
)
SELECT *
FROM final_tab
WHERE movie_rank <= 5;
```

genre	year	movie_name	worlwide_gross_income	movie_rank
Drama	2017	Shatamanam Bhavati	INR 530500000	1
Drama	2017	Winner	INR 250000000	2
Drama	2017	Thank You for Your Service	$ 9995692	3
Comedy	2017	The Healer	$ 9979800	4
Drama	2017	The Healer	$ 9979800	5
Thriller	2018	The Villain	INR 1300000000	1
Drama	2018	Antony & Cleopatra	$ 998079	2
Comedy	2018	La fuitina sbagliata	$ 992070	3
Drama	2018	Zaba	$ 991	4
Comedy	2018	Gung-hab	$ 9899017	5
Thriller	2019	Prescience	$ 9956	1
Thriller	2019	Joker	$ 995064593	2
Drama	2019	Joker	$ 995064593	3
Comedy	2019	Eaten by Lions	$ 99276	4
Comedy	2019	Friend Zone	$ 9894885	5

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

```sql
SELECT production_company,
       count(id) as movie_count,
       RANK() OVER ( ORDER BY count(id) DESC) as prod_comp_rank
FROM movie as m
INNER JOIN ratings as r
ON m.id = r.movie_id
WHERE median_rating >=8 AND
      production_company is not null AND
      POSITION(',' IN languages)>0
GROUP BY production_company
limit 2;
```

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 number of Super Hit movies (average rating >8) in drama genre?

```sql
SELECT name as actress_name,
       sum(total_votes) as total_votes,
       count(movie_id) as movie_count,
       Round(Sum(avg_rating*total_votes)/Sum(total_votes),2) AS actress_avg_rating,
       RANK() OVER (ORDER BY count(movie_id) DESC) as actress_rank
FROM ratings as r
INNER JOIN role_mapping as rm
USING (movie_id)
INNER JOIN genre as g
USING(movie_id)
INNER JOIN names as n
ON rm.name_id=n.id
WHERE genre like '%Drama%' AND
avg_rating >8 AND
category = 'Actress'
GROUP BY name
LIMIT 3;
```

actress_name	total_votes	movie_count	actress_avg_rating	actress_rank
Parvathy Thiruvothu	4974	2	8.25	1
Susan Brown	656	2	8.94	1
Amanda Lawrence	656	2	8.94	1

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

Director id

Name

Number of movies

Average inter movie duration in days

Average movie ratings

Total votes

Min rating

Max rating

total movie durations

director_id	director_name	number_of_movies	avg_inter_movie_days	avg_rating	total_votes	min_rating	max_rating	total_duration
nm2096009	Andrew Jones	5	190.75	3.02	1989	2.7	3.2	432
nm1777967	A.L. Vijay	5	176.75	5.42	1754	3.7	6.9	613
nm0814469	Sion Sono	4	331.00	6.03	2972	5.4	6.4	502
nm0831321	Chris Stokes	4	198.33	4.33	3664	4.0	4.6	352
nm0515005	Sam Liu	4	260.33	6.23	28557	5.8	6.7	312
nm0001752	Steven Soderbergh	4	254.33	6.48	171684	6.2	7.0	401
nm0425364	Jesse V. Johnson	4	299.00	5.45	14778	4.2	6.5	383
nm2691863	Justin Price	4	315.00	4.50	5343	3.0	5.8	346
nm6356309	Özgür Bakar	4	112.00	3.75	1092	3.1	4.9	374

```sql
WITH date_summary AS
( SELECT d.name_id,
        NAME, d.movie_id, duration, r.avg_rating, total_votes, m.date_published,
        Lead(date_published,1) OVER(PARTITION BY d.name_id ORDER BY date_published,movie_id ) AS next_date_published
        FROM director_mapping AS d
        INNER JOIN names AS n
        ON n.id = d.name_id
        INNER JOIN movie AS m
        ON m.id = d.movie_id
        INNER JOIN ratings AS r
        ON r.movie_id = m.id ),
top_director_summary AS
( SELECT *, Datediff(next_date_published, date_published) AS date_difference FROM date_summary )
SELECT name_id AS director_id,
       NAME AS director_name,
       COUNT(movie_id) AS number_of_movies,
       ROUND(AVG(date_difference),2) AS avg_inter_movie_days,
       ROUND(AVG(avg_rating),2) AS avg_rating,
       SUM(total_votes) AS total_votes,
       MIN(avg_rating) AS min_rating,
       MAX(avg_rating) AS max_rating,
       SUM(duration) AS total_duration
FROM top_director_summary
GROUP BY director_id
ORDER BY COUNT(movie_id) DESC limit 9;
```

INSIGHTS

Below are the key insights derived from the data and some recommendations to be given to RSVP movies for their project:

1. There is downward trend in the number of movies released over the years. Most of the movies were produced in the month of March.
2. Drama was the most popular genre with 4285 number of movies and an average duration of 106.77. RSVP movies can focus on this genre for its future films. Action and Thriller genres also hold potential.
3. Most of the movies were rated between 6-8 median rating scale. Aiming for 8+ on a median rating will increase chances of a superhit movie.
4. Dream warrior Pictures and National Theatre Live had produced highest rated films. Star Cinema and Twentieth Century Fox are also good contenders due to high number of multilingual movies as the movie will be for Indian audience primarily.
5. Top directors observed from the analysis are James Mangold, Anthony Russo, Joe Russo and Soubin Shahir. With the later 3 tied at 2nd spot. RSVP can have its future projects with them.
6. Mammooty and Mohanlal are the top actors with the highest number of rating.
7. Taapsee Pannu can be chosen as actress as she is on the top of the list with average rating of 7.74.
8. Marvel Studios(1st), Twentieth Century Fox(2nd) and Warner Bros.(3rd) can be chosen as its Global Partners as number of votes of their movies are maximum.
9. In India, Vijay Sethupati can be focused on as he is popular here.
10. Star Cinema and Twentieth Century Fox are the top production houses that have produced the highest number of hits among multilingual movies..com