RSVP MOVIES' GLOBAL EXPANSION: SQL CASE STUDY

SEGMENT 1:

1) Find the total number of rows in each table of the schema?

```
SELECT table_name, table_rows
FROM information_schema.tables
WHERE table_schema = 'imdb'
ORDER BY table_rows DESC;
```

TABLE_NAME	TABLE_ROWS
names	25735
genre	14662
role_mapping	14397
movie	7997
ratings	7997
director_mapping	3867

2) Which columns in the movie table have null values?

```
SELECT
SUM(Case when id is null then 1 else 0 end) as id_null,
SUM(Case when title is null then 1 else 0 end) as title_null,
SUM(Case when (year) is null then 1 else 0 end) as year_null,
SUM(Case when date_published is null then 1 else 0 end) as date_published_null,
SUM(Case when duration is null then 1 else 0 end) as duration_null,
SUM(Case when country is null then 1 else 0 end) as country_null,
SUM(Case when worlwide_gross_income is null then 1 else 0 end) as ww_null,
SUM(Case when languages is null then 1 else 0 end) as lang_null,
SUM(Case when production_company is null then 1 else 0 end) as production_null
FROM imdb.movie;
```

id_null	title_null	year_null	date_published_null	duration_null	country_null	ww_null	lang_null	production_null
0	0	0	0	0	20	3724	194	528

-- **country, worldwide_gross_income, language, production columns have null values in movie table**

-- Now as you can see four columns of the movie table has null values. Let's look at the at the movies released each year.

3) Find the total number of movies released each year? How does the trend look month wise? (Output expected)

Output format for the first part:

```
+---------------+-------------------+
| Year          |   number_of_movies|
+-------------------+----------------
|   2017        |   2134            |
|   2018        |       .           |
|   2019        |       .           |
+---------------+-------------------+
```

Output format for the second part of the question:

```
+---------------+-------------------+
|   month_num   |   number_of_movies|
+---------------+----------------
|   1           |    134            |
|   2           |    231            |
|   .           |       .           |
+---------------+-------------------+ */
```

-- Number of movies released each year

SELECT YEAR AS Year,

COUNT(*) AS number_of_movies

FROM movie

GROUP BY YEAR

ORDER BY Year;

Year	Number_of_movies
2017	3052
2018	2944
2019	2001

The number of movies released shows a steady decline from 2017 to 2019, with a significant drop in 2019.

-- Number of movies released each month

SELECT Month(date_published) AS Month_Num,

Count(*) AS Number_of_movies

FROM movie

GROUP BY Month_Num

ORDER BY Number_of_movies DESC;

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

The highest number of movies is released in March and the lowest number of movies is released in December.

So, now that you have understood the month-wise trend of movies, let’s take a look at the other details in the movies table.

We know the USA and India produce a huge number of movies each year. Let's find the number of movies produced by the USA or India for the last year.

4) How many movies were produced in the USA or India in the year 2019??

SELECT COUNT(id) AS Total_Movies

FROM movie

WHERE (country LIKE '%INDIA%' OR country LIKE '%USA%')

AND year = 2019;

Total_Movies
1059

The USA and India produced 1059 movies in the year 2019.

5) Find the unique list of the genres present in the data set?

SELECT DISTINCT genre

FROM imdb.genre

ORDER BY genre ASC;

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

6) Which genre had the highest number of movies produced overall?

SELECT genre, COUNT(movie_id) AS Movie_Count

FROM genre

GROUP BY genre

ORDER BY movie_count DESC

LIMIT 1;

genre	Movie_Count
Drama	4285

A total of 4,265 Drama movies were produced, making it the highest among all genres.

7) How many movies belong to only one genre?

WITH unique_movie AS (

SELECT movie_id,

COUNT(genre) AS Total_genre

FROM genre

GROUP BY movie_id

)

SELECT COUNT(*) as Movie_With_One_Genre

FROM unique_movie

WHERE Total_genre = 1;

Movie_With_One_Genre
3289

3289 movies belong to only one genre.

There are more than three thousand movies which has only one genre associated with them.

So, this figure appears significant. Now, let's find out the possible duration of RSVP Movies' next project.

8) What is the average duration of movies in each genre?

-- (Note: The same movie can belong to multiple genres.)

Output format:

```
+---------------+-------------------+
| genre         |   avg_duration    |
+-------------------+---------------
|   thriller    |       105         |
|   .           |       .           |
|   .           |       .           |
+---------------+-------------------+ */
```

SELECT genre, ROUND(AVG(duration), 2) as avg_duration

FROM movie m

INNER JOIN genre 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

Action movies have the longest average duration at 112.88 minutes, while horror is the shortest at 92.72 minutes, with most genres ranging between 100 to 110 minutes.

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

Output format:

```
+---------------+-------------------+---------------------+
| genre         |       movie_count |       genre_rank    |
+---------------+-------------------+---------------------+
|drama          |   2312            |           2         |
+---------------+-------------------+---------------------+*/
```

```
WITH Genre_Rank AS (
    SELECT genre, COUNT(ID) AS Movie_Count,
    ROW_NUMBER() OVER(ORDER BY COUNT(ID) DESC) AS Genre_Rank
    FROM MOVIE M INNER JOIN GENRE G
    ON M.ID = G.MOVIE_ID
    GROUP BY GENRE
)
SELECT *
FROM genre_rank
WHERE genre = 'Thriller';
```

genre	Movie_Count	Genre_Rank
Thriller	1484	3

The Thriller genre has 1,484 movies, ranking 3rd in popularity.

SEGMENT 2:

10) Find the minimum and maximum values in each column of the ratings table except the movie_id column?

/* Output format:

```
+--------------+------------------+--------------------+---------------------+----------------+----------------+
| min_avg_rating|   max_avg_rating  |   min_total_votes   | max_total_votes      |min_median_rating|min_median_rating|
+--------------+------------------+--------------------+---------------------+----------------+----------------+
|       0      |          5       |         177        |    2000             |      0         |    8           |
+--------------+------------------+--------------------+---------------------+----------------+----------------+*/
```

```
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 imdb.ratings;

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

This data shows that movie ratings range from a minimum of 1.0 to a maximum of 10.0, with total votes varying between 100 and 725,138, and median ratings also ranging from 1 to 10.

So, the minimum and maximum values in each column of the ratings table are in the expected range. This implies there are no outliers in the table.

Now, let's find out the top 10 movies based on average ratings.

11) Which are the top 10 movies based on average rating?

/* Output format:

```
+---------------+-------------------+---------------------+
| title         |       avg_rating  |       movie_rank    |
+---------------+-------------------+---------------------+
| Fan           |       9.6         |           5         |
|   .           |       .           |           .         |
|   .           |       .           |           .         |
|   .           |       .           |           .         |
+---------------+-------------------+---------------------+*/
```

SELECT title, avg_rating,

DENSE_RANK() OVER(ORDER BY avg_rating DESC) AS movie_rank

FROM movie AS m INNER JOIN ratings AS r ON r.movie_id = m.id

LIMIT 10;

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

Multiple movies share the same average rating, with "Kirket" and "Love in Kilnerry" tied at rank 1, and several films tied at rank 5 with a 9.5 rating.

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

/* Output format:

```
+---------------+-------------------+
| median_rating |   movie_count     |
+-------------------+----------------
|   1           |       105         |
|   .           |       .           |
|   .           |       .           |
+---------------+-------------------+ */
```

SELECT median_rating, Count(movie_id) AS Movie_Count

FROM ratings

GROUP BY median_rating

ORDER BY movie_count DESC;

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

Most movies have a median rating of 7, with very few rated 1 or 2, suggesting that the majority of movies tend to receive mid-range to high ratings.

Now, let's find out the production house with which RSVP Movies can partner for its next project.

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

Output format:

```
+------------------+------------------+---------------------+
|production_company|movie_count       |    prod_company_rank|
+------------------+------------------+---------------------+
| The Archers      |        1         |            1        |
+------------------+------------------+---------------------+*/
```

```
WITH prod_company_ranks as
(
SELECT m.production_company,
   COUNT(m.id) AS movie_count,
   RANK() OVER (ORDER BY COUNT(m.id) DESC) AS prod_company_rank
FROM movie AS m
JOIN ratings AS r
ON m.id = r.movie_id
WHERE r.avg_rating > 8 and m.production_company is not NULL
GROUP BY m.production_company
)
SELECT production_company, movie_count, prod_company_rank
FROM prod_company_ranks
WHERE prod_company_rank = 1;
```

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

Both Dream Warrior Pictures and National Theatre Live are tied for the top rank, each producing 3 movies.

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

Output format:

```
+---------------+-------------------+
| genre         |   movie_count     |
+-------------------+----------------
|   thriller    |       105         |
|   .           |       .           |
|   .           |       .           |
+---------------+-------------------+ */
```

SELECT genre, COUNT(id) AS Movie_Count

FROM genre g

INNER JOIN movie m

ON m.id = g.movie_id

INNER JOIN ratings r

ON m.id = r.movie_id

WHERE MONTH(date_published) = 3

AND YEAR(date_published) = 2017

AND total_votes > 1000

AND country LIKE '%USA%'

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

In March 2017 in the USA, Drama had the most releases with 24 movies having more than 1,000 votes, followed by Comedy with 9 movies, while Family had the fewest with only 1 release.

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

Output format:

```
+---------------+-------------------+---------------------+
| title         |       avg_rating  |       genre         |
+---------------+-------------------+---------------------+
| Theeran       |       8.3         |       Thriller      |
|   .           |       .           |           .         |
|   .           |       .           |           .         |
|   .           |       .           |           .         |
+---------------+-------------------+---------------------+*/
```

SELECT m.title, r.avg_rating, g.genre

FROM movie m

INNER JOIN ratings r ON m.id = r.movie_id

INNER JOIN genre g ON m.id = g.movie_id

WHERE m.title LIKE 'The%'

AND r.avg_rating > 8

ORDER BY r.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

The list features high-rated movies starting with "The" and has an average rating greater than 8 across various genres, with Drama being the most common, including "The Brighton Miracle" (9.5) and "The Colour of Darkness" (9.1).

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

```
SELECT  r.median_rating,
COUNT(m.id) AS Movie_released_april2018_april2019
FROM movie AS m
INNER JOIN ratings AS r
ON m.id = r.movie_id
WHERE r.median_rating = 8
AND date_published BETWEEN '2018-04-01' AND '2019-04-01'
GROUP BY r.median_rating;
```

median_rating	Movie_released_april2018_april2019
8	361

A total of 361 movies with a median rating of 8 were released between April 2018 and April 2019.

17) Do German movies get more votes than Italian movies?

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 IN ('Germany', 'Italy')

GROUP BY country;

country	Total_Votes
Germany	106710
Italy	77965

Yes, German movies received more votes (106,710) compared to Italian movies, which received 77,965 votes.

SEGMENT 3:

18) Which columns in the names table have null values??

```
+---------------+-------------------+---------------------+----------------------+
| name_nulls    |   height_nulls    |date_of_birth_nulls  |known_for_movies_nulls|
+---------------+-------------------+---------------------+----------------------+
|       0       |           123     |          1234       |    12345             |
+---------------+-------------------+---------------------+----------------------+*/
```

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

The data indicates many missing entries, with 17,335 nulls for height, 13,431 nulls for date of birth, and 15,226 nulls for known movies, but no missing values for names.

The director is the most important person in a movie crew.

Let’s find out the top three directors in the top three genres who can be hired by RSVP Movies.

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

Output format:

```
+---------------+-------------------+
| director_name |   movie_count     |
+---------------+-------------------|
|James Mangold  |       4           |
|   .           |       .           |
|   .           |       .           |
+---------------+-------------------+ */
```

-- First, we find the top three genres with high-rated movies.

```
WITH Top_Three_Genre AS (
    SELECT genre, COUNT(m.id) AS Movie_Count
    FROM movie m
    INNER JOIN genre g ON m.id = g.movie_id
    INNER JOIN ratings r ON r.movie_id = m.id
    WHERE avg_rating > 8
    GROUP BY genre
    ORDER BY Movie_Count DESC
    LIMIT 3
)
SELECT
    n.name AS Director_Name,
    COUNT(m.id) AS Movie_Count
FROM movie m
```

```
INNER JOIN
    director_mapping d ON m.id = d.movie_id
INNER JOIN
    names n ON n.id = d.name_id
INNER JOIN
    genre g ON g.movie_id = m.id
INNER JOIN
    ratings r ON m.id = r.movie_id
WHERE g.genre IN (SELECT genre FROM Top_Three_Genre)
    AND avg_rating > 8
GROUP BY director_name
ORDER BY Movie_Count DESC
LIMIT 3;
```

Director_Name	Movie_Count
James Mangold	4
Joe Russo	3
Anthony Russo	3

The top directors with movies averaging a rating above 8 are James Mangold (4 movies), Joe Russo (3 movies), and Anthony Russo (3 movies).

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

Output format:

```
+---------------+-------------------+
| actor_name    |   movie_count     |
+-------------------+----------------
|Christain Bale |       10          |
|   .           |       .           |
+---------------+-------------------+ */
```

```
SELECT
    n.name AS Actor_Name,
    COUNT(m.id) AS Movie_Count
```

```
FROM movie m
INNER JOIN ratings r ON m.id = r.movie_id
INNER JOIN role_mapping rm ON m.id = rm.movie_id
INNER JOIN names n ON n.id = rm.name_id
WHERE median_Rating >= 8
GROUP BY Actor_name
ORDER BY Movie_count DESC
LIMIT 2;
```

Actor_Name	Movie_Count
Mammootty	8
Mohanlal	5

The top two actors whose movies have a median rating of 8 or higher are Mammootty with 8 movies and Mohanlal with 5 movies.

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

Output format:

```
+------------------+--------------------+---------------------+
|production_company|vote_count          |       prod_comp_rank|
+------------------+--------------------+---------------------+
| The Archers      |       830          |       1             |
|   .              |       .            |           .         |
|   .              |       .            |           .         |
+------------------+--------------------+---------------------+*/
```

```
SELECT
    production_company,
    SUM(total_votes) AS Vote_count,
    ROW_NUMBER() OVER (ORDER BY SUM(total_votes) DESC) AS
prod_comp_rank
FROM movie m
INNER JOIN ratings r
```

ON m.id = r.movie_id

GROUP BY production_company

LIMIT 3;

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

The top three production houses based on the number of votes received by their movies are Marvel Studios (2,659,667 votes), Twentieth Century Fox (2,411,163 votes), and Warner Bros. (2,396,057 votes).

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

Output format:

```
+--------------+------------------+--------------------+----------------------+-----------------+
| actor_name   |   total_votes    |   movie_count      | actor_avg_rating     |actor_rank       |
+--------------+------------------+--------------------+----------------------+-----------------+
|   Yogi Babu  |          3455    |         11         |    8.42              |      1          |
|      .       |          .       |         .          |    .                 |      .          |
|      .       |          .       |         .          |    .                 |      .          |
|      .       |          .       |         .          |    .                 |      .          |
+--------------+------------------+--------------------+----------------------+-----------------+*/
```

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

The actor at the top of the list based on average ratings is Vijay Sethupathi, with an average rating of 8.42.

23) Find out the top five actresses in Hindi movies released in India based on their average ratings?

Output format:

```
+---------------+-------------------+---------------------+----------------------+-----------------+
| actress_name  |   total_votes     |   movie_count       | actress_avg_rating   |actress_rank     |
+---------------+-------------------+---------------------+----------------------+-----------------+
|   Tabu        |          3455     |          11         |    8.42              |      1          |
|       .       |          .        |          .          |    .                 |      .          |
|       .       |          .        |          .          |    .                 |      .          |
|       .       |          .        |          .          |    .                 |      .          |
+---------------+-------------------+---------------------+----------------------+-----------------+*/
```

SELECT

name AS actress_name,

SUM(total_votes) AS total_votes,

```
   COUNT(m.id) AS movie_count,
  ROUND(SUM(avg_rating * total_votes) / SUM(total_votes), 2) AS
actress_avg_rating,
    ROW_NUMBER() OVER (
   ORDER BY ROUND(SUM(avg_rating * total_votes) / SUM(total_votes), 2)
DESC, SUM(total_votes) DESC
    ) AS actress_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 = "actress"
    AND country LIKE "%india%"
    AND languages LIKE "%hindi%"
GROUP BY actress_name
HAVING movie_count >= 3;
```

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

Taapsee Pannu tops with an average rating of 7.74.

24) 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

```
SELECT m.title, r.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'
WHEN avg_rating <5 THEN 'Flop movie'
ELSE 'Invalid'
END AS 'avg_rating_category'
FROM movie m
INNER JOIN ratings r
ON m.id = r.movie_id
INNER JOIN genre g
ON m.id = g.movie_id
WHERE g.genre = 'thriller';
```

title	avg_rating	movie_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

SEGMENT 4:

25) What is the genre-wise running total and moving average of the average movie duration?

Output format:

```
+---------------+-------------------+---------------------+----------------------+
| genre         |   avg_duration    |running_total_duration|moving_avg_duration  |
+---------------+-------------------+---------------------+----------------------+
|   comdy       |           145     |          106.2      |    128.42            |
|       .       |           .       |          .          |    .                 |
|       .       |           .       |          .          |    .                 |
|       .       |           .       |          .          |    .                 |
+---------------+-------------------+---------------------+----------------------+*/
```

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

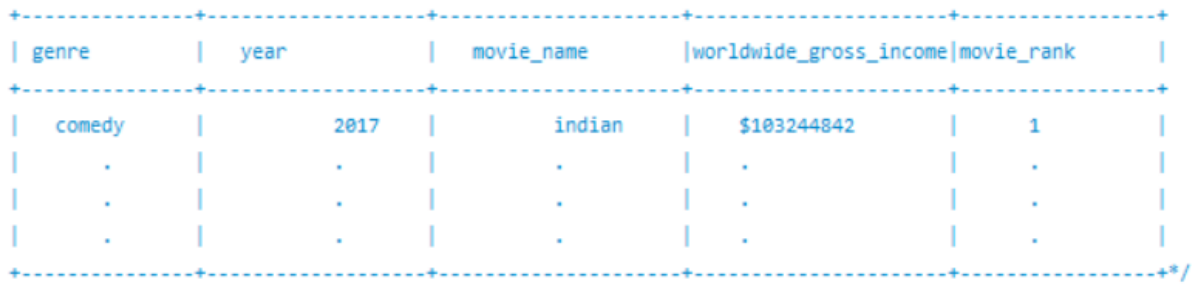
ORDER BY genre;

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

The table shows average movie durations by genre, with Action having the longest average duration (112.88 minutes) and a cumulative running total and moving average that gradually increase as genres are listed.

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

Output format:



genre	year	movie_name	worldwide_gross_income	movie_rank
comedy	2017	indian	$103244842	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

*/

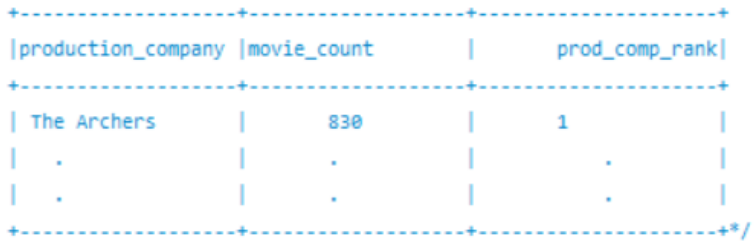
```
WITH top3genre AS (
    SELECT genre
    FROM genre
    GROUP BY genre
    ORDER BY COUNT(movie_id) DESC
    LIMIT 3
), currency_converted AS (
    SELECT
        id,
        CASE
            WHEN worlwide_gross_income LIKE 'INR%'
            THEN Cast(Replace(worlwide_gross_income, 'INR', '') AS DECIMAL(12)) / 84
            WHEN worlwide_gross_income LIKE '$%'
                THEN Cast(Replace(worlwide_gross_income, '$', '') AS DECIMAL(12))
            ELSE Cast(worlwide_gross_income AS DECIMAL(12))
        END AS worlwide_gross_income
```

```
    FROM
        movie), ranked_movies AS (
    SELECT
        genre,
        year,
        title AS movie_name,
        CONCAT('$', ROUND(cc.worlwide_gross_income, 0)) AS
'worldwide_gross_income ($)',
        DENSE_RANK() OVER(PARTITION BY year ORDER BY
cc.worlwide_gross_income DESC) movie_rank
    FROM
        movie m
        INNER JOIN genre g ON m.id = g.movie_id
        INNER JOIN currency_converted cc ON cc.id = m.id
    WHERE genre IN (SELECT * FROM top3genre)
)
SELECT * FROM ranked_movies
WHERE movie_rank <= 5;
```

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	4
Comedy	2017	Guardians of the Galaxy Vol. 2	$863756051	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	4
Thriller	2019	Ne Zha zhi mo tong jiang shi	$700547754	5

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

Output format:



```
+------------------+------------------+--------------------+
|production_company |movie_count       |      prod_comp_rank|
+------------------+------------------+--------------------+
| The Archers      |      830         |      1             |
|   .              |      .           |          .         |
|   .              |      .           |          .         |
+------------------+------------------+--------------------+*/
```

-- Top 2 production houses for high-rated multilingual movies

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

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

Star Cinema is the top production company with 7 movies, followed by Twentieth Century Fox with 4 hits.

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

Output format:

```
+---------------+-------------------+---------------------+----------------------+-----------------+
| actress_name  |   total_votes     |   movie_count       |actress_avg_rating    |actress_rank     |
+---------------+-------------------+---------------------+----------------------+-----------------+
|   Laura Dern  |           1016    |          1          |    9.60              |      1          |
|       .       |           .       |          .          |    .                 |      .          |
|       .       |           .       |          .          |    .                 |      .          |
+---------------+-------------------+---------------------+----------------------+-----------------+*/
```

-- Top 3 actresses in the drama genre with Super Hit movies

```
SELECT
    name AS actress_name,
    SUM(total_votes) AS total_votes,
    COUNT(m.id) AS movie_count,
    ROUND(AVG(avg_rating),2) AS actress_Avg_rating,
    ROW_NUMBER() OVER (ORDER BY count(m.id) DESC) AS actress_rank
FROM
    names n
INNER JOIN role_mapping rm ON n.id = rm.name_id
INNER JOIN movie m ON m.id = rm.movie_id
INNER JOIN ratings r ON r.movie_id = m.id
INNER JOIN genre g ON g.movie_id = m.id
WHERE
    avg_rating > 8
    AND category = "actress"
    AND genre = "drama"
GROUP BY actress_name
```

ORDER BY movie_count desc

LIMIT 3;

actress_name	total_votes	movie_count	actress_Avg_rating	actress_rank
Parvathy Thiruvothu	4974	2	8.20	1
Susan Brown	656	2	8.95	2
Amanda Lawrence	656	2	8.95	3

Parvathy Thiruvothu ranks first among actresses with an average rating of 8.20 across 2 movies, followed by Susan Brown and Amanda Lawrence, both with an 8.95 rating but fewer total votes.

29) Get the following details for the top 9 directors (based on the 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

Format:

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	177	5.65	1754	3.7	6.9	613
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

WITH director_info AS

(

SELECT

dm.name_id AS director_id,

```
    name AS director_name,
    dm.movie_id AS movie_id,
    date_published,
    LEAD(date_published, 1) OVER(PARTITION BY dm.name_id ORDER BY
date_published) AS next_date_published,
    total_votes,
    avg_rating,
    duration
FROM
 names n
  INNER JOIN
 director_mapping dm ON dm.name_id = n.id
  INNER JOIN
 movie m ON m.id = dm.movie_id
  INNER JOIN
 ratings r ON r.movie_id = m.id
), top_directors AS
(
SELECT
 director_id,
    director_name,
 COUNT(movie_id) number_of_movies,
 ROUND(AVG(datediff(next_date_published, date_published)), 2) as
avg_inter_movie_days,
    ROUND(SUM(avg_rating * total_votes)/ SUM(total_votes), 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,
    RANK() OVER(ORDER BY COUNT(movie_id) DESC) as movie_rank
FROM
 director_info
GROUP BY director_id
)
SELECT director_id,
       director_name,
       number_of_movies,
       avg_inter_movie_days,
       avg_rating,
       total_votes,
       min_rating,
       max_rating,
       total_duration
FROM   top_directors
WHERE  movie_rank <= 9;
```

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	176.75	5.65	1754	3.7	6.9	613
nm2096009	Andrew Jones	5	190.75	3.04	1989	2.7	3.2	432
nm0001752	Steven Soderbergh	4	254.33	6.77	171684	6.2	7.0	401
nm0425364	Jesse V. Johnson	4	299.00	6.10	14778	4.2	6.5	383
nm0515005	Sam Liu	4	260.33	6.32	28557	5.8	6.7	312
nm0814469	Sion Sono	4	331.00	6.31	2972	5.4	6.4	502
nm0831321	Chris Stokes	4	198.33	4.32	3664	4.0	4.6	352
nm2691863	Justin Price	4	315.00	4.93	5343	3.0	5.8	346
nm6356309	Özgür Bakar	4	112.00	3.96	1092	3.1	4.9	374

The top 9 directors, led by A.L. Vijay and Andrew Jones with 5 movies each, vary in average ratings and total votes, with Steven Soderbergh having the highest average rating (6.77) and most votes (171,684).