RSVP Movies is an Indian film production company which has produced many super hit movies.

-- They have usually released movies for Indian audience but for their next project, they are planning to release a movie for the global audience in 2022.

-- The production company wants to plan their every move analytically based on data. We analysed the dataset and

-- draw meaningful insights that can help them to start their new project

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

SELECT table\_name, table\_rows

FROM INFORMATION\_SCHEMA.TABLES

WHERE TABLE\_SCHEMA = 'rsvp';

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

WITH null\_info

AS (SELECT 'id' AS 'Column\_Name', Count(\*) AS Null\_Values

FROM movies

WHERE id IS NULL

UNION ALL

SELECT 'title' AS 'Column\_Name', Count(\*) AS Null\_Values

FROM movies

WHERE title IS NULL

UNION ALL

SELECT 'year' AS 'Column\_Name', Count(\*) AS Null\_Values

FROM movies

WHERE year IS NULL

UNION ALL

SELECT 'date\_published' AS 'Column\_Name', Count(\*) AS Null\_Values

FROM movies

WHERE date\_published IS NULL

UNION ALL

SELECT 'duration' AS 'Column\_Name', Count(\*) AS Null\_Values

FROM movies

WHERE duration IS NULL

UNION ALL

SELECT 'country' AS 'Column\_Name', Count(\*) AS Null\_Values

FROM movies

WHERE country IS NULL

UNION ALL

SELECT 'worlwide\_gross\_income' AS 'Column\_Name', Count(\*) AS Null\_Values

FROM movies

WHERE worlwide\_gross\_income IS NULL

UNION ALL

SELECT 'languages' AS 'Column\_Name', Count(\*) AS Null\_Values

FROM movies

WHERE languages IS NULL

UNION ALL

SELECT 'production\_company' AS 'Column\_Name', Count(\*) AS Null\_Values

FROM movies

WHERE production\_company IS NULL)

SELECT column\_name

FROM null\_info

WHERE Null\_Values > 0

ORDER BY null\_values DESC;

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

select Year, count(title) as number\_of\_movies from movies group by Year;

SELECT MONTH(date\_published) AS month\_num, COUNT(id) AS number\_of\_movies

FROM movies

GROUP BY MONTH(date\_published)

ORDER By MONTH(date\_published);

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

select count(title) as Movie\_count from movies

where country = 'USA' or country = 'India' and Year = 2019;

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

SELECT distinct genre FROM genre order by genre;

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

SELECT g.genre, COUNT(m.id) AS num\_of\_movie

FROM genre g

INNER JOIN movies m ON g.movie\_id = m.id

GROUP BY genre

ORDER BY COUNT(id) DESC

LIMIT 1;

Q7. How many movies belong to only one genre?

WITH one\_genre AS

(

SELECT movie\_id, COUNT(distinct genre) AS number\_of\_genre FROM genre

GROUP BY movie\_id

HAVING number\_of\_genre=1

)

SELECT COUNT(\*) AS number\_of\_movies

FROM one\_genre;

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

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

select g.genre , avg(duration) as avg\_duration from movies m

join genre g on g.movie\_id=m.id

group by genre

order by avg\_duration DESC;

Q9.What is the rank of the â€˜thrillerâ€™ genre of movies among all the genres in terms of number of movies produced?

-- (Hint: Use the Rank function)

With genredetail

As

(select g.genre, count(title) as movie\_count,rank() over (order by count(title) desc ) genre\_rank from movies m

join genre g on g.movie\_id=m.id

-- where genre = 'Thriller'

group by g.genre)

SELECT \* from genredetail

where genre = 'Thriller';

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

select min(avg\_rating) as min\_avg\_ratings,max(avg\_rating)as max\_avg\_ratings,

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;

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

WITH top\_avg\_rating\_rank

AS

(

SELECT

m.title AS title, r.avg\_rating as avg\_rating,

DENSE\_RANK() OVER(ORDER BY r.avg\_rating DESC) movie\_rank

FROM movies m

INNER JOIN ratings r

ON m.id = r.movie\_id

)

SELECT \* FROM top\_avg\_rating\_rank WHERE movie\_rank<=10;

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

select median\_rating, count(movie\_id) as movie\_count from ratings

group by median\_rating

order by median\_rating;

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

select movies.production\_company,count(movies.id) as movie\_count,rank() over (order by count(movies.id)) as Prod\_comp\_rank

from movies

inner join ratings on movies.id = ratings.movie\_id

where ratings.avg\_rating > 8

group by production\_company;

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

SELECT g.genre,

COUNT(m.id) AS movie\_count

FROM genre g

INNER JOIN movies m ON g.movie\_id = m.id

INNER JOIN ratings r ON m.id = r.movie\_id

WHERE MONTH(m.date\_published) = 3

AND m.year = 2017

AND m.country = 'USA'

AND r.total\_votes > 1000

GROUP BY g.genre

ORDER BY movie\_count DESC;

Q15. Find movies of each genre that start with the word â€˜Theâ€™ and which have an average rating > 8?

select movies.title,ratings.avg\_rating, genre.genre from ((movies

inner join ratings on movies.id=ratings.movie\_id)

inner join genre on movies.id=genre.movie\_id)

where movies.title like 'the%' and avg\_rating > 8

order by avg\_rating desc;

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

SELECT COUNT(id) as MovieReleased from movies m

INNER JOIN ratings r ON m.id = r.movie\_id

WHERE r.median\_rating = 8 AND date\_published BETWEEN '01-04-2018' AND '01-04-2019'

GROUP BY median\_rating;

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

select movies.country,sum(ratings.total\_votes) as Votes from movies

inner join ratings on movies.id=ratings.movie\_id

where movies.country='Germany' or movies.country='Italy'

group by country;

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

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;

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

-- (Hint: The top three genres would have the most number of movies with an average rating > 8.)

select names.name as director\_name,count(ratings.movie\_id) as movie\_count from names

inner join ratings on names.known\_for\_movies=ratings.movie\_id

having avg\_rating > 8

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

SELECT n.name AS actor\_name,

COUNT(ro.movie\_id) AS movie\_count FROM role ro

INNER JOIN names n ON n.id = ro.name\_id

INNER JOIN ratings r ON r.movie\_id = ro.movie\_id

WHERE category="actor" AND r.median\_rating >= 8

GROUP BY n.name

ORDER BY movie\_count DESC

LIMIT 2;

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

SELECT production\_company, SUM(total\_votes) AS vote\_count,

DENSE\_RANK() OVER(ORDER BY SUM(total\_votes) DESC) AS prod\_comp\_rank

FROM movies m

INNER JOIN ratings r ON m.id = r.movie\_id

GROUP BY production\_company

ORDER BY vote\_count DESC

LIMIT 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.

-- (Hint: You should use the weighted average based on votes. If the ratings clash, then the total number of votes should act as the tie breaker.)

WITH ind\_actors

AS (SELECT n.NAME,

Sum(total\_votes) AS total\_votes,

Count(r.movie\_id) AS movie\_count,

Round(Sum(total\_votes \* avg\_rating) / Sum(total\_votes), 2) AS actor\_avg\_rating,

Rank() OVER(ORDER BY Round(Sum(total\_votes \* avg\_rating)/Sum(total\_votes), 2) DESC,

Sum(total\_votes) DESC) AS actor\_rank

FROM names n

INNER JOIN role rm ON n.id = rm.name\_id

INNER JOIN movies m ON rm.movie\_id = m.id

INNER JOIN ratings r ON rm.movie\_id = r.movie\_id

WHERE country = 'India'

GROUP BY n.NAME

HAVING movie\_count >= 5)

SELECT \* FROM ind\_actors;

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

WITH ind\_actress

AS (SELECT n.NAME AS actress\_name,

Sum(total\_votes) AS total\_votes,

Count(r.movie\_id) AS movie\_count,

Round(Sum(total\_votes \* avg\_rating) / Sum(total\_votes), 2) AS actor\_avg\_rating,

Rank() OVER(ORDER BY Round(Sum(total\_votes \* avg\_rating)/Sum(total\_votes), 2) DESC,

Sum(total\_votes) DESC) AS actress\_rank

FROM names n

INNER JOIN role rm ON n.id = rm.name\_id

INNER JOIN movies m ON rm.movie\_id = m.id

INNER JOIN ratings r ON rm.movie\_id = r.movie\_id

WHERE country = 'India'

AND category = 'actress'

AND languages = 'Hindi'

GROUP BY n.NAME

HAVING movie\_count >= 3)

SELECT \* FROM ind\_actress

WHERE actress\_rank <= 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

--------------------------------------------------------------------------------------------\*/

SELECT title AS movie\_title,

avg\_rating,

CASE

WHEN avg\_rating > 8 THEN 'Superhit movie'

WHEN avg\_rating BETWEEN 7 AND 8 THEN 'Hit movie'

WHEN avg\_rating BETWEEN 5 AND 7 THEN 'One-time-watch movie'

ELSE 'Flop movie'

END Movie\_type

FROM genre g

INNER JOIN movies m ON g.movie\_id = m.id

INNER JOIN ratings r ON m.id = r.movie\_id

WHERE g.genre = 'Thriller'

ORDER BY movie\_title;

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

SELECT genre,

Round(Avg(duration)) AS avg\_duration,

round(sum(Avg(duration)) OVER w1, 1) AS running\_total\_duration,

round(avg(avg(duration)) OVER w2, 2) AS moving\_avg\_duration

FROM genre g

INNER JOIN movies m

ON g.movie\_id = m.id

GROUP BY genre

WINDOW w1 AS (ORDER BY genre rows UNBOUNDED PRECEDING),

w2 AS (ORDER BY genre rows BETWEEN 2 PRECEDING AND 2 following);

Q26. Which are the top two production houses that have produced the highest number of hits (median rating >= 8) among multilingual movies?

WITH prod\_comp\_info

AS (SELECT production\_company,

Count(movie\_id) AS movie\_count,

Rank() over(ORDER BY Count(movie\_id) DESC) AS prod\_comp\_rank

FROM ratings r

INNER JOIN movies m ON r.movie\_id = m.id

WHERE production\_company IS NOT NULL

AND median\_rating >= 8

AND Position(',' IN languages) > 0

GROUP BY production\_company)

SELECT \* FROM prod\_comp\_info

WHERE prod\_comp\_rank <= 2;

Q27. Who are the top 3 actresses based on number of Super Hit movies (average rating >8) in drama genre?

WITH top\_actress

AS (SELECT n.NAME AS actress\_name,

Sum(total\_votes) AS total\_votes,

Count(r.movie\_id) AS movie\_count,

Round(Sum(total\_votes \* avg\_rating) / Sum(total\_votes), 2) AS actor\_avg\_rating,

Rank() OVER(ORDER BY Count(r.movie\_id) DESC) AS actress\_rank

FROM names n

INNER JOIN role rm ON n.id = rm.name\_id

INNER JOIN genre g ON rm.movie\_id = g.movie\_id

INNER JOIN ratings r ON rm.movie\_id = r.movie\_id

WHERE category = 'actress'

AND genre = 'Drama'

AND avg\_rating > 8

GROUP BY n.NAME

)

SELECT \* FROM top\_actress

WHERE actress\_rank <= 3;