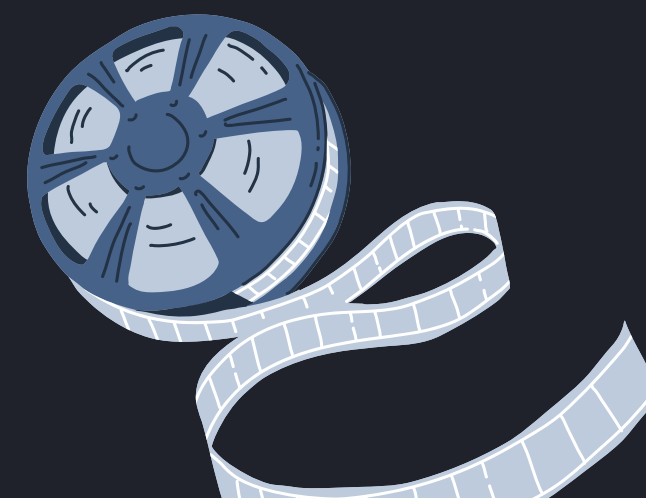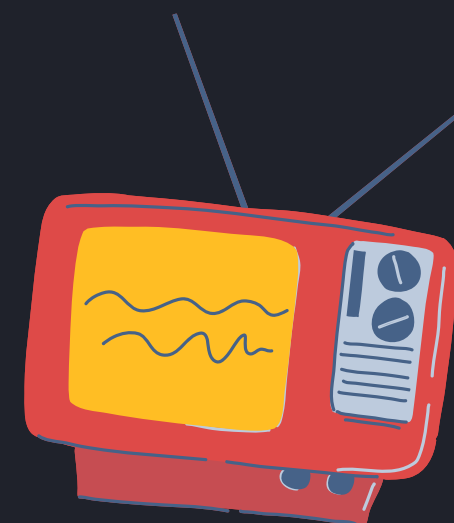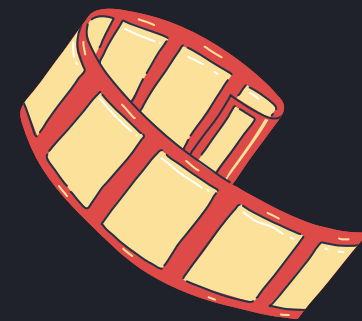# RSVP Movies Case Study (Segment-1)

PROBLEM BACKGROUND

RSVP Movies is an Indian film production company which has produced many super-hit movies. They have usually released movies for the Indian audience but for their next project, they are planning to release a movie for the global audience in 2022.
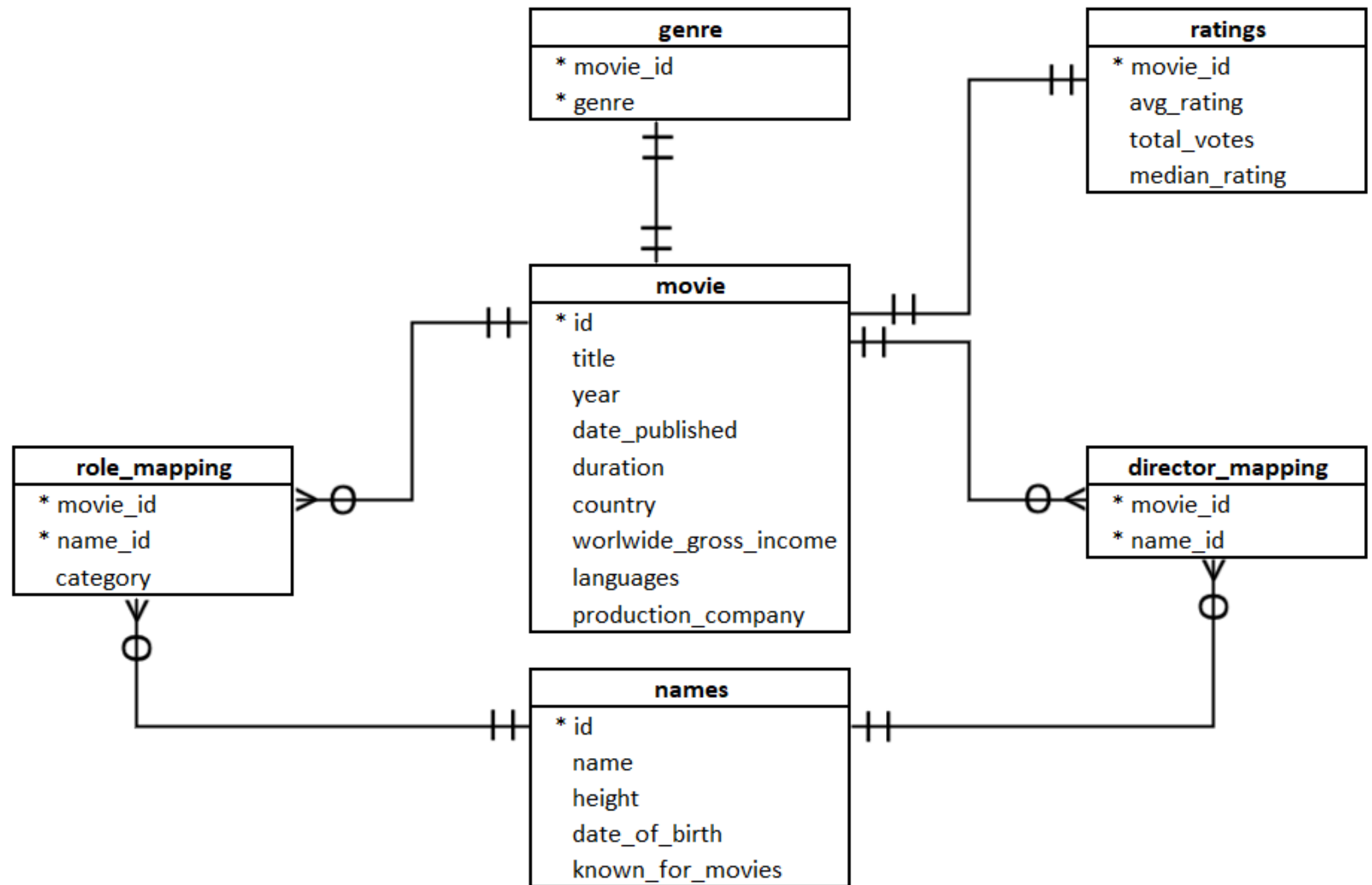
The production company wants to plan their every move analytically based on data and have approached us for help with this new project. We have been provided with the data of the movies that have been released in the past three years. We have to analyze the data set and draw meaningful insights that can help them start their new project.

DATABASE SCHEMA

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

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

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

```sql
SELECT table_name, table_rows
FROM INFORMATION_SCHEMA.TABLES
WHERE TABLE_SCHEMA = 'imdb';
```

TABLE_NAME	TABLE_ROWS
director_mapping	3867
genre	14662
movie	7438
names	27679
ratings	7927
role_mapping	15714

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

```sql
select
sum( case when id is null then 1 else 0 end) as id_nulls,
sum( case when title is null then 1 else 0 end) as title_nulls,
sum( case when year is null then 1 else 0 end) as year_nulls,
sum( case when date_published is null then 1 else 0 end) as date_published_nulls,
sum( case when duration is null then 1 else 0 end) as duration_nulls,
sum( case when  country is null then 1 else 0 end) as country_nulls,
sum( case when worlwide_gross_income is null then 1 else 0 end) as worlwide_gross_income_nulls,
sum( case when languages is null then 1 else 0 end) as languages_nulls,
sum( case when production_company  is null then 1 else 0 end) as production_company_nulls
from movie;
```

id_nulls	title_nulls	year_nulls	date_published_nulls	duration_nulls	country_nulls	worlwide_gross_income_nulls	languages_nulls	production_company_nulls
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(*) AS number_of_movies
FROM movie
GROUP BY year
ORDER BY year;
```

year	number_of_movies
2017	3052
2018	2944
2019	2001

```sql
SELECT MONTH(date_published) AS month_num,
       COUNT(*) 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

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

```sql
SELECT COUNT(*) AS total_Movies_released FROM movie
WHERE year = 2019 AND country IN ('USA' , 'India');
```

total_Movies_released
887

Q5.Find the unique list of the genres present in the data set.Which genre had the highest number of movies produced overall?

```sql
SELECT DISTINCT genre FROM genre;
```

genre
Drama
Fantasy
Thriller
Comedy
Horror
Family
Romance
Adventure
Action

```sql
SELECT genre, COUNT(movie_id) AS total_movies
FROM genre g INNER JOIN movie m ON g.movie_id=m.id
GROUP BY genre
ORDER BY total_movies DESC
LIMIT 1;
```

genre	total_movies
Drama	4285

Q6. How many movies belong to only one genre?

```sql
WITH movie_count AS
(
SELECT movie_id, COUNT(movie_id) AS number_of_movies
FROM genre
GROUP BY movie_id HAVING number_of_movies = 1
)
SELECT COUNT(movie_id) AS number_of_movies FROM movie_count;
```

number_of_movies
3289

Q7.What is the average duration of movies in each genre?

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

Q8.What is the rank of the 'thriller' genre of movies among all the genres in terms of number of movies produced?
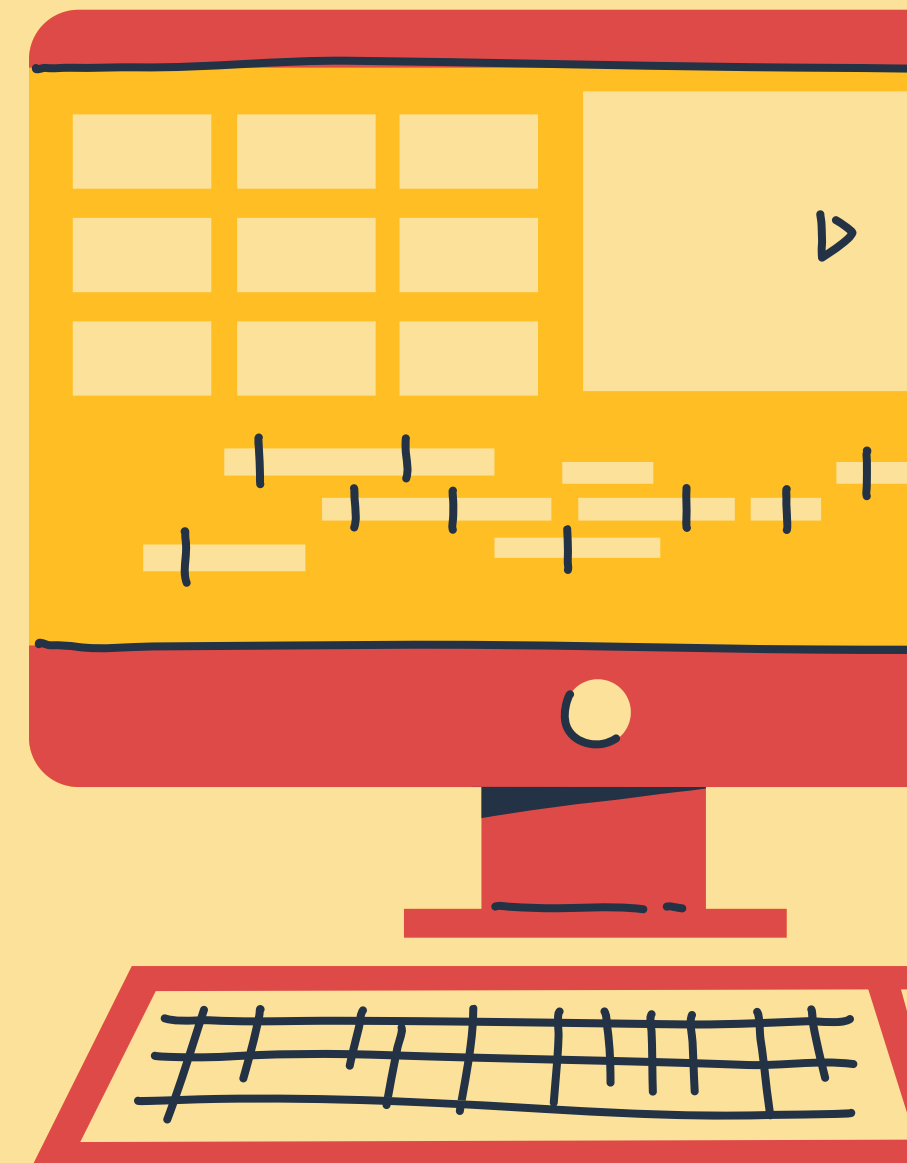
```sql
WITH genre_rank AS
(
    SELECT genre, COUNT(movie_id) AS movie_count,
            RANK() OVER(ORDER BY COUNT(movie_id) DESC) AS genre_rank
    FROM genre
    GROUP BY genre
)

SELECT *
FROM genre_rank
WHERE genre='thriller';
```

genre	movie_count	genre_rank
Thriller	1484	3

SUMMARY

- We can see 4 columns of the movie table has null values - country,worlwide_gross_income,languages, production_company.
- The highest number of movies is produced in the month of March.
- USA and India produced more than 800 movies(887) in the year 2019.
- 'Drama' genre had the highest number of movies produced overall.
- There are more than three thousand movies which has only one genre associated with them.
- Movies of genre 'Drama' (produced highest in number) has the average duration of 106.77 mins.
- 'Thriller' movies is in top 3 among all genres in terms of number of movies.

THANK YOU!