SQL – RSVP MOVIE CASE STUDY

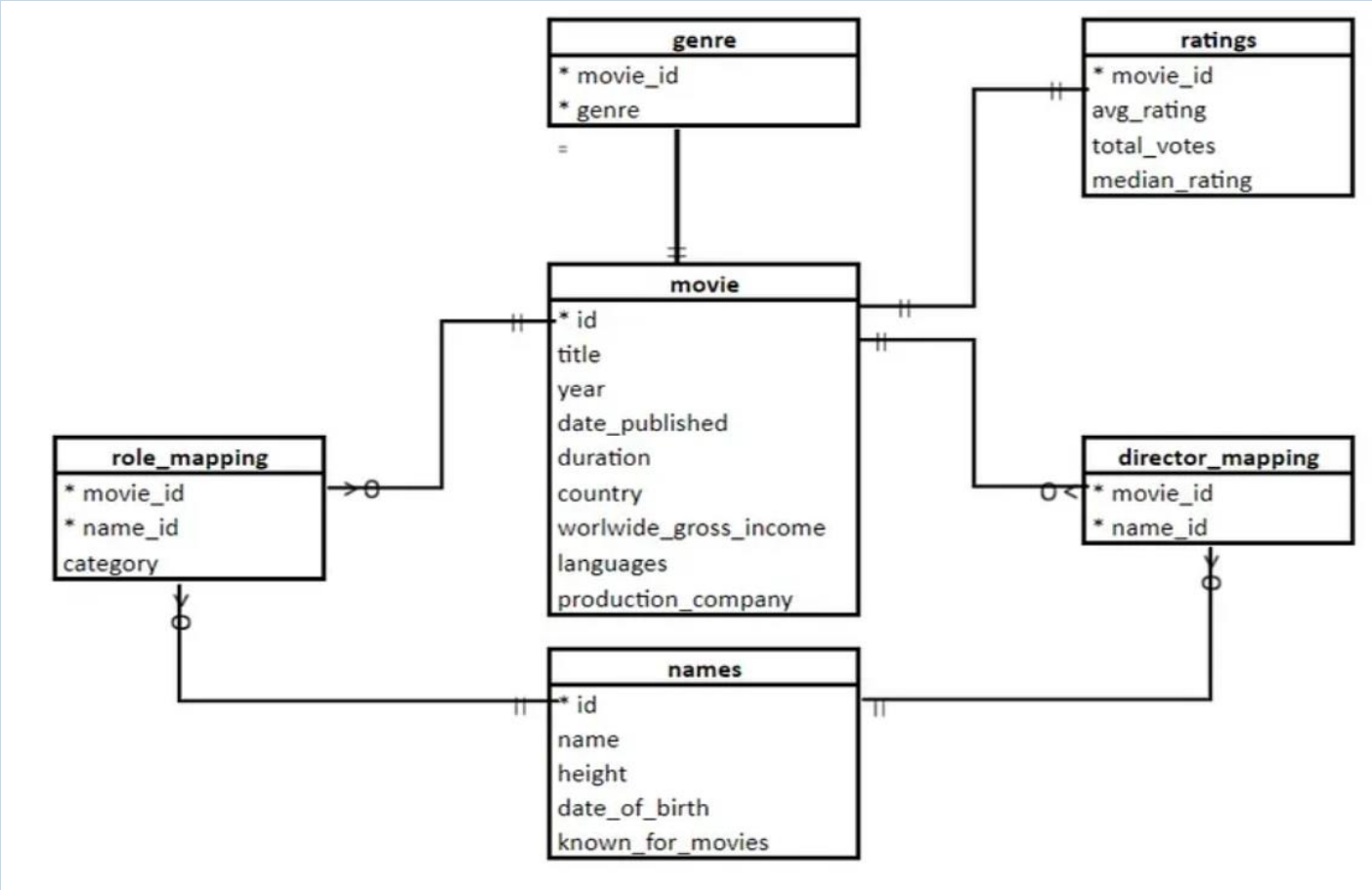# RSVP MOVIE CASE STUDY|| PROBLEM STATEMENT

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 you for help with this new project. You have been provided with the data of the movies that have been released in the past three years. You have to analyse the data set and draw meaningful insights that can help them start their new project.

RSVP MOVIE CASE STUDY|| DATA OVERVIEW

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

RSVP MOVIE CASE STUDY|| INSIGHTS – SEGMENT 1

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

TableName	RowCount
▶ director_mapping	3867
genre	14661
movie	7997
names	25735
ratings	7997
role_mapping	15615

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

```
-- Columns                    Total Null
-- worlwide_gross_income      3724
-- production_company         528
-- languages                  198
-- country                    20
```

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

	Year	number_
▶	2017	3052
	2018	2944
	2019	2001

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

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

	movie_count
▶	1059

RSVP MOVIE CASE STUDY|| INSIGHTS – SEGEMENT 1 📈

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

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

Insight: We got a diverse cinematic genres which is total of 13

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

GENRE	TOTAL_MOVIES
Drama	4285
Comedy	2412
Thriller	1484

Insight: Genre - Drama takes the front running among its competitors

Q7. How many movies belong to only one genre?

movie_with_one_genre
3289

Insight: There are over 3,000 movies associated with a single genre, which will be quite interesting to explore.

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

genre	avg_duration
Drama	106.77
Thriller	101.58
Comedy	102.62
Fantasy	105.16
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

Insight: This is a good metric to measure the success of the movie. For e.g., the genre – Drama was highest in 2019 with average duration of 106.77 minutes and second genre – Comedy following up closely with 102.62.
This will help in planning the upcoming genre movie productions.

RSVP MOVIE CASE STUDY|| INSIGHTS- SEGMENT 1

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

genre	movie_count	genre_rank
Thriller	1484	3

Insight: "Thriller" genre ranks among the top 3 genres with the most movies and will be good input for producers to plan their next upcoming projects.

RSVP MOVIE CASE STUDY|| INSIGHTS – SEMENT 2

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

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

Insight: Data is well within range and does not have any major outliers for ratings

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

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

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

median_rating	movie_count
7	2257
6	1975
8	1030
5	985
4	479

Insight: Movie rating with 7 is the front runner and makes a good input for movie production. This is one of the industry KPIs to measure the success of a movie and having this input will help plan the movie productions.

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

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

RSVP MOVIE CASE STUDY|| INSIGHTS– SEMENT 2

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

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

Insight: This insight helps in identifying the genres which will be a good sell with audience.

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

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

Insight: This insight provides 'The' coupled with genres and their respective ratings.
This will be helpful to see if movie name is captivating the audience.

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

Movie_released_april2018_april2019
361

Insight: There were 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?

country	total_votes
Germany	106710
Italy	77965

Insight: Yes, German movies received the highest number of votes

RSVP MOVIE CASE STUDY|| INSIGHTS – SEGMENT 3

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

name_nulls	HEIGHT_nulls	DATE_OF_BIRTH_nulls	KNOWN_FOR_MOVIES_nulls
0	17335	13431	15226

Insight: No null values in the "name" column which indicates data completeness and accuracy.

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

director_name	Movie_count
James Mangold	4
Joe Russo	3
Anthony Russo	3

Insight: No null values in the "name" column which indicates data completeness and accuracy.

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

Actor_name	Movie_count
Mammootty	8
Mohanlal	5

Insight: This insight will be helpful for producers to plan their next big picture and budget required.

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

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

Insight: Marvel studio is the front runner for the production houses with number of votes received.

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

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

Insight: Vijay Sethupathi tops the chart and well deserved as well.

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

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

Insight: Taapsee Pannu tops the chart with her selected roles

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

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

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

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	105.16	737.32	105.33
Horror	92.72	830.04	103.76
Mystery	101.80	931.84	103.54

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

genre	year	title	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

Insight: A next level of deep dive with combination of genres, movies and rating will give a good insight to movie production team of what can be planned for next and if it has the chances of making it to "block buster" list

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

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

Insight: The Star_Cinema and Twentieth Century Fox will has the highest number of hits for multilingual movies. Capturing some good pratices in making the movie to marketing will be helpful.

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

actress_name	total_votes	movie_count	actoress_avg_rating	actoress_rank
Sangeetha Bhat	1010	1	9.60	1
Fatmire Sahiti	3932	1	9.40	2
Adriana Matoshi	3932	1	9.40	2

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

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
nm6356309	Özgür Bakar	4	112.00	3.75	1092	3.1	4.9	374
nm2691863	Justin Price	4	315.00	4.50	5343	3.0	5.8	346
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
nm0425364	Jesse V. Johnson	4	299.00	5.45	14778	4.2	6.5	383
nm0001752	Steven Soderbergh	4	254.33	6.48	171684	6.2	7.0	401
nm0515005	Sam Liu	4	260.33	6.23	28557	5.8	6.7	312

Insight: Building a good movie starts with having a good director and this insights provides this data.

Conclusion

In conclusion, the insights provided to movie production – business teams will help plan better on their next project. This is a wealth of information which will help understand how their investments will earn them back the business or even higher profitability.
The insights like having a good director, good actor or actress, production teams are very important for budget planning. Planning to onboard a director shows the kind of impact on genre of the movie being produced – like locations, movie set, colors, customs etc.,

Over and above, having this data provides a strong backing for your business canvas of where investments should be made? Why? What?