

Abstract geometric lines in the top-left corner of the page, consisting of several thin black lines forming overlapping, irregular polygons and triangles.

SQL-BASED ANALYSIS OF THE IMDB MOVIE DATABASE

By: Axton Cahyadi

CONTENTS

- Profile
- Problem Statement
- Data Description
- Analysis Approach
- Insights
- Recommendations
- Conclusion

AXTON CAHYADI



- Bachelor of Advanced Computing student at the University of Sydney
- Majoring in Computer Science, minoring in Software Development
- Passionate about Big Data, Cloud native web applications and Agile Workflows.

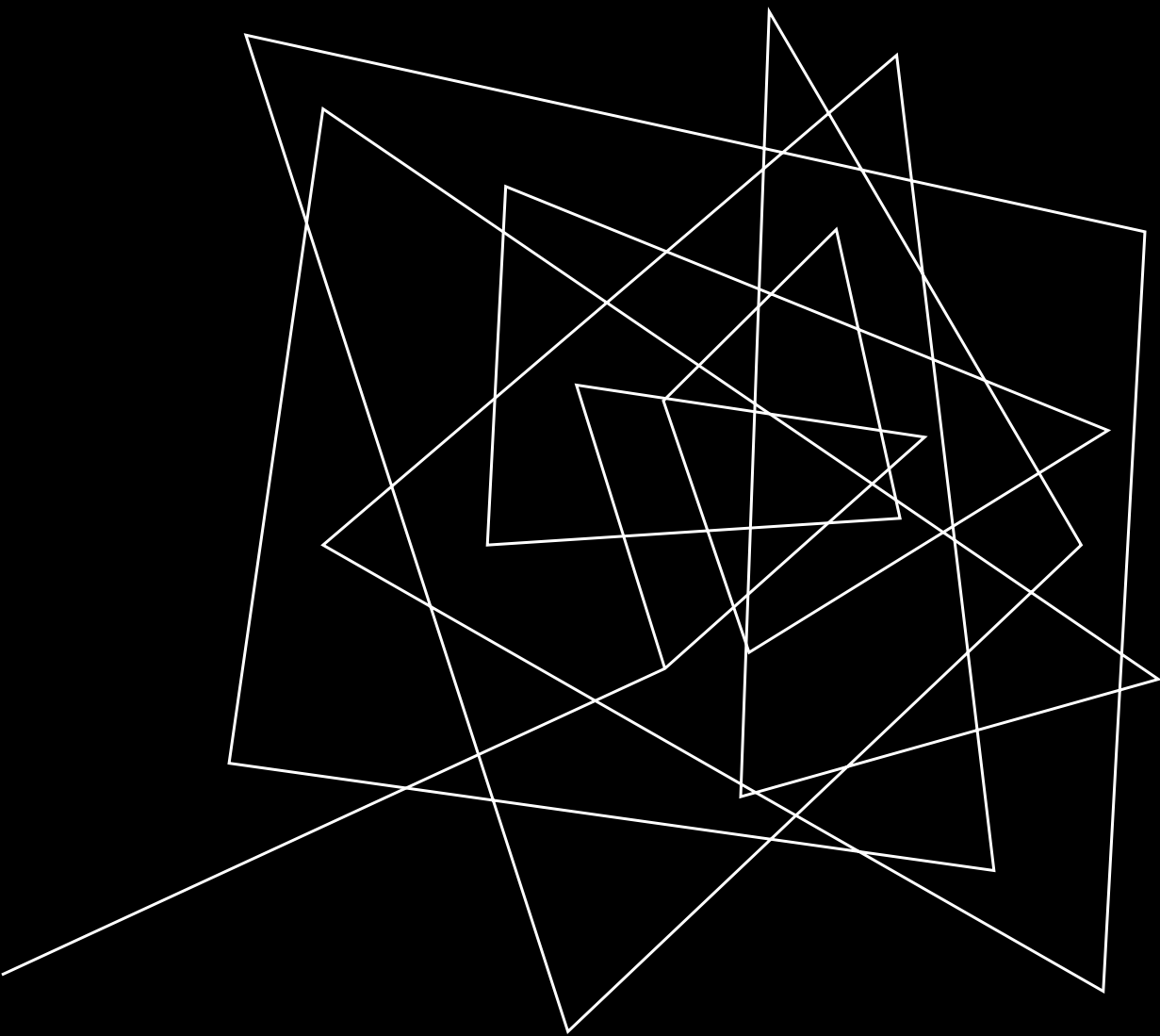
The image features a minimalist design on a light gray background. Two thin, dark gray lines intersect diagonally. One line runs from the top-left towards the bottom-right, while the other runs from the top-right towards the bottom-left. To the right of the intersection, the words "PROBLEM STATEMENT" are written in a bold, black, sans-serif font, stacked in two lines.

**PROBLEM
STATEMENT**

Project aims to explore and extract meaningful insights from the IMDb dataset using SQL, focusing on production trends, genres, ratings, and talent performance.

Going into this we have some key questions:

- Which countries and genres dominate global cinema?
- What time of year sees the most releases?
- Who are the top directors and actors based on ratings?
- Which genres perform best in terms of quality and quantity?



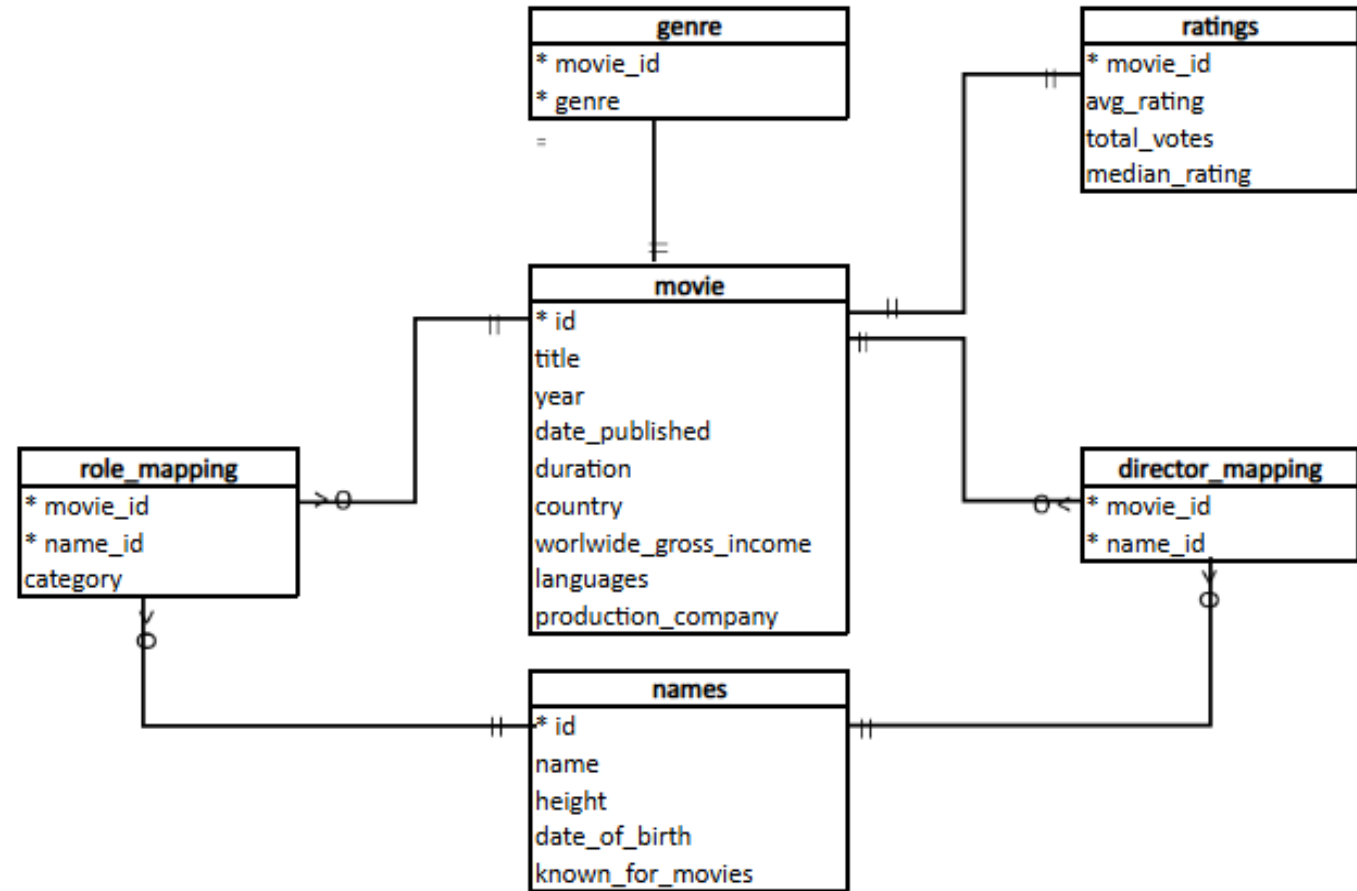
DATA DESCRIPTION

Tables Used: movie, genre, ratings, names, role_mapping, director_mapping

Time Span: Various years up to 2019

Source: IMDb SQL schema with approximately:

- 25k+ names
- 8k+ movies
- 14k+ genre entries
- 15k+ role mappings



An abstract graphic consisting of two thin, dark gray lines intersecting on a light gray background. One line is oriented diagonally from the top-left towards the bottom-right, while the other is oriented more steeply from the top-left towards the bottom-right, crossing the first line.

ANALYSIS APPROACH

By making use of MySQL I was able to:

- Generate queries using aggregations, joins, CTEs, ranking functions, and window functions.
- Query outputs are segmented per business-relevant insights.

Query coverage includes:

- Table data analysis for missing values and structure.
- Release and Productions trends.
- Genre analysis.
- Actor, Director, Movies insights and their performance.

An abstract graphic on a light gray background. Two thin, dark gray lines intersect. One line is nearly vertical, starting from the top center and extending towards the bottom right. The other line is nearly horizontal, starting from the top left and extending towards the middle right. The word "INSIGHTS" is written in a bold, black, sans-serif font, positioned to the right of the intersection point.

INSIGHTS

RELEASE TRENDS

Based on Q3 and Q4

1. Most movies released in **2019**, peak month: **March**
2. Production surges in Q1 suggest **targeting early-year releases**
3. USA & India together produced **887** movies in 2019 alone

Movies produced in
USA/india 2019

	movie_count
▶	887

Movies made per year

	year	number_of_movies
▶	2017	3052
	2018	2944
	2019	2001

Movies made per month of
all years

	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

GENRE PREFERENCES

Based on Q5, Q6, Q7, Q8, Q9

1. **Drama** leads in production count
2. **Thriller** ranked top 3 in frequency
3. Most movies belong to only **one genre**
4. Average duration by genre helps set editorial guidelines (e.g., Drama ~106.8 mins)

Genres

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

Rank of genres movie count

	genre	movie_count	genre_rank
►	Drama	4285	1
	Comedy	2412	2
	Thriller	1484	3
	Action	1289	4
	Horror	1208	5
	Romance	906	6
	Crime	813	7
	Adventure	591	8
	Mystery	555	9
	Sci-Fi	375	10
	Fantasy	342	11
	Family	302	12
	Others	100	13

Average genre duration

	genre	avg_duration
►	Drama	106.77
	Fantasy	105.14
	Thriller	101.58
	Comedy	102.62
	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

Single movie genre count

	single_genre_movies
►	3289

Genre with most movies

	genre	movie_count
►	Drama	4285

MOVIE QUALITY

Based on Q10, Q11, Q12

1. Top-rated movie: *Fan* (9.6)
2. Highest number of movies have **median rating = 7**
3. Genre-wise rankings help studios / RSVP choose genres with consistent audience approval

Highest ranked movies based on average ratings

	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

Movie counts per 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

Ratings statistics

	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

TALENT & CREW INSIGHTS

Production house ranking
with most avg_rating of 8+

Based on Q13, Q15, Q19, Q20

1. **Top Directors** with avg rating > 8 in leading genres:
2. James Mangold, etc.
3. **Top Actors** with median rating ≥ 8 :
4. Christian Bale, Mohanlal
5. **Top Production Houses (by votes)**: Marvel Studios, Dream Warrior Pictures

	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
	The Archers	1	11
	Blaze Film Enterprises	1	11
	Bradeway Pictures	1	11
	Bert Marcus Productions	1	11
	A Studios	1	11
	Ronk Film	1	11
	Benaras Mediaworks	1	11
	Bioscope Film Framers	1	11
	Bestwin Production	1	11
	Studio Green	1	11
	AKS Film Studio	1	11
	Kaargo Cinemas	1	11
	Animonsta Studios	1	11
	O3 Turkey Medya	1	11
	StarVision	1	11
	Synergy Films	1	11
	PVP Cinema	1	11

Top 3 Directors with median
rating of 8+

	director_name	movie_count
▶	James Mangold	4
	Joe Russo	3
	Anthony Russo	3

Top 2 actors in median
rating of 8+ movies

	actor_name	movie_count
▶	Mammootty	8
	Mohanlal	5

Movies starting with 'The'
with average 8+

	title	avg_rating	genre
▶	The Blue Elephant 2	8.8	Drama
	The Blue Elephant 2	8.8	Horror
	The Blue Elephant 2	8.8	Mystery
	The Brighton Miracle	9.5	Drama
	The Irishman	8.7	Crime
	The Irishman	8.7	Drama
	The Colour of Darkness	9.1	Drama
	The Mystery of Godliness: The Sequel	8.5	Drama
	The Gambinos	8.4	Crime
	The Gambinos	8.4	Drama
	The King and I	8.2	Drama
	The King and I	8.2	Romance

REGIONAL INSIGHTS

Based on Q22, Q23

1. Indian actor **Vijay Sethupathi** ranked #1 by weighted avg rating in ≥ 5 movies
2. **Taapsee Pannu** leads among Hindi actresses (≥ 3 movies)
3. Targeting regional talent boosts authenticity and audience reach

Actress ranked based on avg_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

	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
	Naseeruddin Shah	12604	5	6.54	17
	Anandraj	2750	6	6.54	18
	Mohanlal	17244	6	6.51	19
	Aju Varghese	2237	5	6.43	20
	Siddique	5953	7	6.43	21
	Prakash Raj	8548	6	6.37	22
	Jimmy Sheirgill	3826	6	6.29	23
	Mahesh Achanta	2716	6	6.21	24
	Biju Menon	1916	5	6.21	25
	Suraj Venjaramo...	4284	6	6.19	26
	Abir Chatterjee	1413	5	5.80	27
	Sunny Deol	4594	5	5.71	28
	Radha Ravi	1483	5	5.70	29
	Prabhu Deva	2044	5	5.68	30

RECOMMENDATIONS

- Focus on **Drama and Thriller genres** as they dominate in quantity and ranking, indicating high audience demand and easier market penetration.
- Target **March release** windows for maximum exposure as it has the highest number of movie releases as it has better box office positioning as it aligns the release with peak audience activity.
- Cast **top-rated regional actors** for credibility and loyalty.
- Partner with **high-performing production houses** with strong rating history as trusted brands increase viewer expectation and loyalty.
- Certain directors such as **James Mangold thrive in Drama & Thiller** and **Russo Brothers thrive in action movies**. Hire top ranked directors in leading genre's to reduce risk and elevate project profitability.
- Focus on **USA and India for production and distribution** as they are the highest-output markets which means **increased ROI** from higher-consumption markets.



CONCLUSION

- This project made use of SQL analytics to get actionable insights from the IMDb database.
- We discovered data-backed trends in all aspects in order to improve business and marketing strategies that's based on evidence and not intuition.
- Key findings suggest focusing on **Drama and Thriller genres**, releasing content in **peak seasons like March**, and collaborating with **top-rated directors and production houses** to maximize audience engagement and revenue.
- Ultimately, this analysis proves that **carefully structured queries** can generate **better strategic directions** for companies based on historic data.



THANK YOU

Axton

0406823507

www.linkedin.com/in/axton-cahyadi-378115233

axtoncahyadi@gmail.com