



Lights, Camera, Earnings: What Drives the Show Biz?

Team 7: Divyam Rana, Achinthya Sreedhar, Neha Shastri, Wenlin Zhao, Xingzhu Chen, Atishay Jain

Table of Contents



- 01** | **Introduction**
Executive Summary
- 02** | **Data Cleanup & Mapping**
- 03** | **EDA & Dashboards**
- 04** | **Conclusion**
Challenges

01

Introduction





WHAT are we going to talk about?

- ❑ **Fierce Competition:** Requires innovative strategies to stand out
- ❑ **Audience Understanding:** Essential for creating impactful content
- ❑ **Data-Driven Insights:** Needed to understand movie success metrics (ratings, revenue, genre trends)





WHY is this relevant?

01

Maximize Appeal and Profitability:
Critical for survival in a competitive market

02

Understand Seasonal Trends:
Optimize release timing for high-grossing movies

03

Identify Success Drivers: Focus on strategies with the best ROI

04

Tailored Strategies: Tap into diverse audience preferences and boost engagement

WHO are the stakeholder?



**CONTENT
CREATORS**



MARKETERS



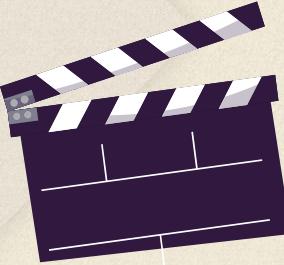
**STUDIO
EXECUTIVES**



**DATA
ANALYSTS**



AUDIENCE



Executive Summary

Summary

- Analyzed the IMDB dataset
- Explored the factors of successful **movies and entertainment media**
- **Identified the key trends in the show biz**

Headwinds

- **JSON** type objects in dataset
- Type Inaccuracies
- Missing Values

Insights

- **Comedy, Drama, and Documentary** are most produced
- **Action, Adventure, and Animation** tend to perform better
- High-grossing films often release in peak seasons: **June and December**.
- Telugu films and specific titles like **Pirates of the Caribbean** show regional popularity (like in Caribbean)



02 | Data Cleanup & Mapping

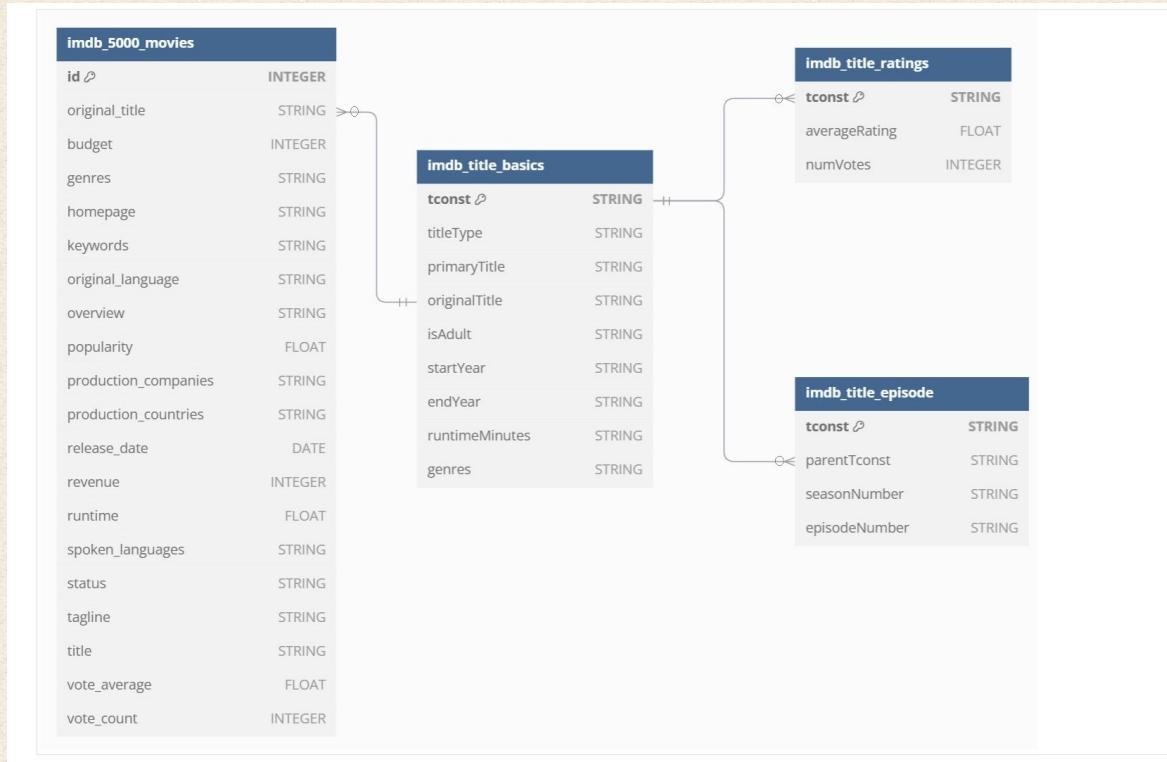
Datasource

IMDb Non-Commercial Datasets

Movie Box Office Revenue

Data Types: String , Integer, Float, DateTime, JSON

Entity Relationship Diagram



Data Cleaning and Preparation

- Created Views to obtain a clean interface for analysis and reporting by organizing, filtering, and combining datasets.
- Converted numeric fields in ideal formats for consistency.
- Mapped original_language codes to readable names.
- Applied thresholds for relevance wherever required.



03 | EDA & Dashboards

Genre

Which Genre combinations have the highest ratings?

- Top Genre: "Comedy, Music, Short" – Avg. Rating: 9.15, Bayesian: 8.4
- Highest Avg. Rating: "Documentary, Mystery, Sci-Fi" – 9.5
- Balanced Genre: "Animation, Documentary, Family" – Avg. Rating: 8.7, Bayesian: 8.1

	genres	Num_Titles	Average_Rating	Bayesian_Avg_Rating
0	Comedy,Music,Short	2	9.150000	8.40
1	Adventure,Family,Reality-TV	4	8.575000	8.24
2	Documentary,Mystery,Sci-Fi	1	9.500000	8.20
3	Animation,Comedy,Talk-Show	2	8.800000	8.16
4	Documentary,Family,Reality-TV	4	8.425000	8.12
5	Animation,Documentary,Family	2	8.700000	8.10
6	Adventure,Sport	2	8.650000	8.06
7	Crime,Drama,Reality-TV	2	8.600000	8.03
8	Documentary,Family	37	8.043243	8.01
9	Animation,Documentary	9	8.122222	8.00



TV Shows

What are the most popular genres in TV Series, TV Mini Series, and TV Episode?

- **Top Genre:** War leads with an average rating of 8.03.
- **Second Place:** Western follows at 7.63.
- **Other Strong Performers:** Documentary (7.59) and History (7.59).



	Genre	AvgRating
0	War	8.033333
1	Western	7.626415
2	Documentary	7.593682
3	History	7.588462
4	Sport	7.516842
5	Biography	7.463345
6	Music	7.383688
7	Crime	7.328591
8	Drama	7.325025
9	Romance	7.303456

Movie

What are the most popular movie genres and what is their average popularity and average rating?

- Drama (7.12), Thriller (7.08), and Action (6.99) are top-rated genres by average rating.
- High ratings don't always align with average popularity, showing a gap between acclaim and audience appeal.



	genre_names	genre_count	avg_popularity	avg_rating
0	Drama	8892	28.17	7.12
1	Thriller	7258	30.82	7.08
2	Action	5623	41.11	6.99
3	Comedy	5562	28.90	6.94
4	Adventure	3877	50.34	6.95
5	Crime	3657	28.46	7.14
6	Romance	3445	25.19	7.01
7	Science Fiction	2835	50.74	7.03
8	Horror	2302	27.32	6.95
9	Fantasy	2144	49.38	6.95
10	Mystery	2118	31.20	7.08
11	Family	2069	44.81	6.85
12	Animation	1088	57.12	6.85
13	History	669	23.99	7.28
14	Music	635	20.63	7.14
15	War	587	34.30	7.16
16	Western	230	24.76	7.08
17	Documentary	121	7.23	7.16
18	Foreign	71	1.87	7.18

What are the Top 10 Most Popular Movies(by vote count) Released in the 2000s

	movie_title	release_year	vote_count
0	The Dark Knight	2008	2939644
1	The Lord of the Rings: The Fellowship of the Ring	2001	2054876
2	The Lord of the Rings: The Return of the King	2003	2025045
3	The Lord of the Rings: The Two Towers	2002	1825670
4	Gladiator	2000	1668967
5	Inglourious Basterds	2009	1626167
6	Batman Begins	2005	1610495
7	The Prestige	2006	1473959
8	The Departed	2006	1450779
9	Avatar	2009	1407149

- ❖ **Top Voted Movies:** *The Dark Knight* (2.94M votes) leads the list, showcasing its immense popularity and impact among audiences.
- *The Lord of the Rings* trilogy dominates with all three movies
- ❖ **Consistently High Performers:** Movies like *Gladiator* (1.67M votes), and *Batman Begins* (1.61M votes) also rank highly, highlighting their strong audience appeal.



Revenue

What is the categorization of the movies according to the revenue?

	revenue_range	movie_count
0	Medium Revenue	1688
1	Low Revenue	2271
2	High Revenue	844

Movies are categorized into three revenue ranges using statistical thresholds:

- **High Revenue:** Defined by the **75th percentile** and above --- 844.
- **Medium Revenue:** Falls between the **median** and **75th percentile**---1688 .
- **Low Revenue:** Defined by the **25th percentile** and below --- 2271.
- Most movies perform moderately — opportunity to boost



What are the highest revenue generating movie for each country (TOP 10)

- ❖ **Global Blockbusters:** Avatar dominates in both US and UK with the highest revenue
- ❖ **Regional Preferences:** Pirates of the Caribbean 2 leads in Jamaica, Bahamas and Dominica ---- strong interest in adventure
- ❖ **Localized Success:** Iron Man 3 leads in China (\$1.2 billion), while Harry Potter and the Chamber of Secrets is the favorite in Germany (\$876 million).

	country	title	revenue
0	United States of America	Avatar	2787965087
1	United Kingdom	Avatar	2787965087
2	Japan	Furious 7	1506249360
3	China	Iron Man 3	1215439994
4	New Zealand	The Lord of the Rings: The Return of the King	1118888979
5	Bahamas	Pirates of the Caribbean: Dead Man's Chest	1065659812
6	Jamaica	Pirates of the Caribbean: Dead Man's Chest	1065659812
7	Dominica	Pirates of the Caribbean: Dead Man's Chest	1065659812
8	Germany	Harry Potter and the Chamber of Secrets	876688482
9	Canada		2015595





Dashboards



04 | Conclusion & Challenges

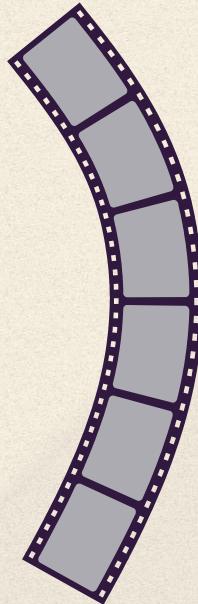
Recommendations

- ★ Focus on High-Performing Genres and Themes
- ★ Optimize Runtime
- ★ Strategic Release Timing
- ★ Regional Tailoring
- ★ Brand Building
- ★ Expanding Global Reach



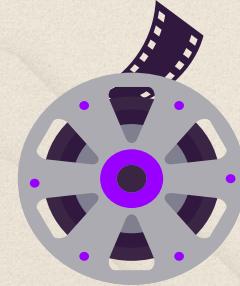
Challenges

- **Missing and Inconsistent Data**
- **Data Format Issues (JSON)**
- **Data Cleaning in SQL**



Future Steps

- Impact of Streaming Platforms
- Star Power and Directorial Influence
- Analyse Relationship between Budget Size and Box Office
- Evolving Audience Behavior



References

- ❑ IMDb Non-Commercial Datasets
(<https://datasets.imdbws.com/>)
- ❑ Movie Box Office Revenue Prediction
(<https://data.mendeley.com/datasets/xv9wtc9gdk/2>)



Thanks