IMDB Movie Analysis Report

Submitted by

Nalina Lingasamy

DTM10

Introduction:

The IMDB Movie Analysis Project aims to explore the factors influencing the success of movies on IMDB, focusing on their ratings. By analysing a dataset containing information about various movies, including genres, directors, budgets, durations, and languages, the project seeks to provide insights that can guide stakeholders in the film industry, such as producers, directors, and investors, in making informed decisions for future projects.

Problem statement:

The primary problem addressed in this project is understanding what factors contribute to a movie's success on IMDB, particularly in terms of high ratings. The project seeks to uncover the relationships between different variables such as movie genres, durations, languages, directors, and budgets, and their impact on IMDB scores.

The analysis conducted in this project holds significant implications for stakeholders in the film industry. Understanding the factors influencing a movie's success can help producers and directors make strategic decisions in various aspects of filmmaking, including genre selection, budget allocation, and directorial choices. Investors can also use these insights to assess the potential profitability of investing in specific movie projects.

Data cleaning and Pre-processing:

- Dataset was loaded in Jupyter notebook.
- Null values and datatypes where checked for all columns in the dataset.
- For various numerical columns in the data, distribution plot was made by importing seaborn library.
- If the distribution was normal, null values were replaced with mean of the particular column. If it is not normal, null values were replaced with median of the particular column.
- For columns like Actor's name, Director's name, null values were filled with "Not specified" string.
- Some of the column's null values were filled with mode of the particular column.

- The column "genres" contains many categories separated by "|". It was splitted using string split method and splitted categories were exploded into separate rows.
- The genres list was further grouped based on IMBD score and its mean, median, mode, max, min, standard deviation, variance was calculated.
- The treated data frame was then saved as an excel file and loaded into PowerBI.

Data Analysis:

A. Movie Genre Analysis:
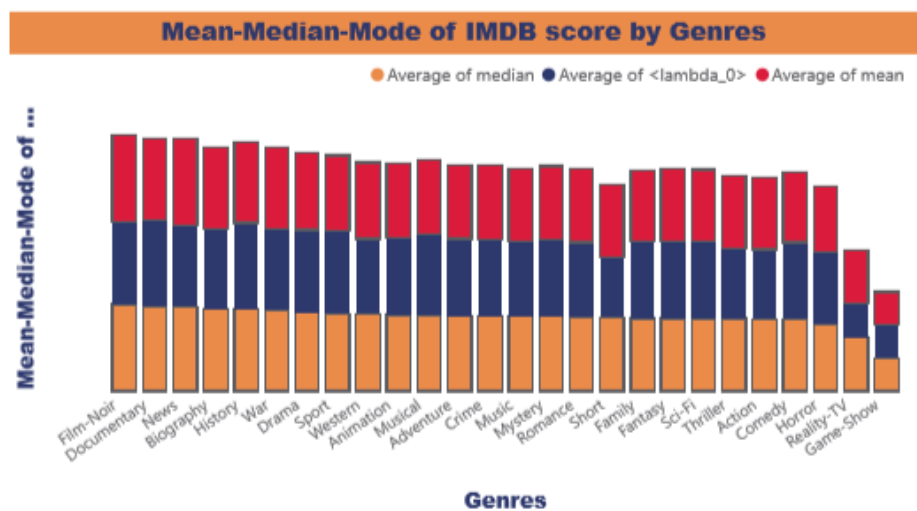
- At 7.65, Film-Noir had the highest Average of median and was 163.79% higher than Game-Show, which had the lowest Average of median at 2.90.
- Average of median and total Average of mode are positively correlated with each other.
- Across all 26 genres list, Average of median ranged from 2.90 to 7.65, Average of mode ranged from 2.90 to 7.50, and Average of mean ranged from 2.90 to 7.63.
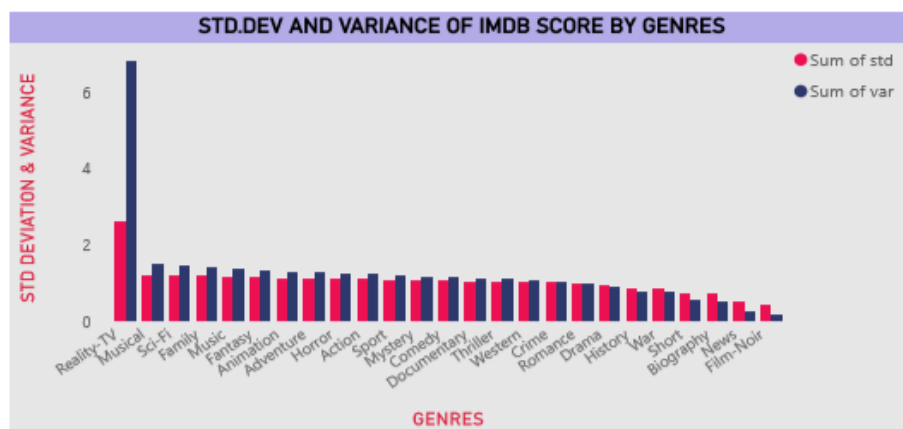
- At 2.62, Reality-TV had the highest Sum of std and was 505.55% higher than Film-Noir, which had the lowest Sum of std at 0.43.
- Sum of std and total Sum of var are positively correlated with each other.
- Reality-TV accounted for 9.80% of Sum of std.
- Sum of var and Sum of std diverged the most when the genres_list was Reality-TV, when Sum of var were 4.23 higher than Sum of std.

Average of gross vs genres_list by budget

imdb_score by genres_list
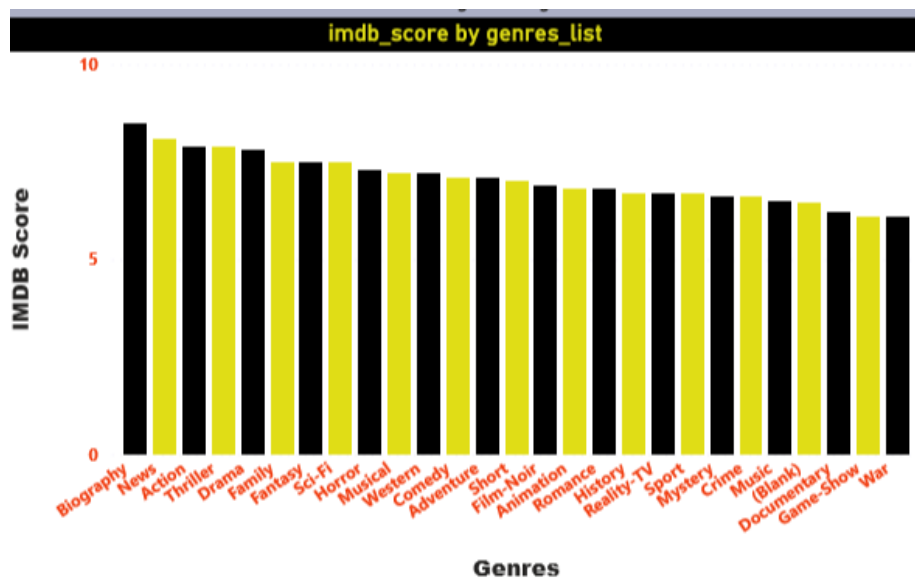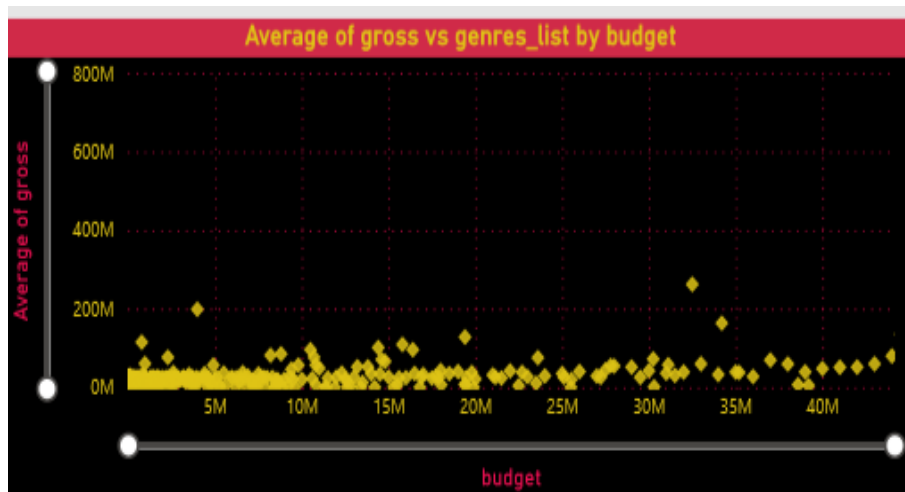
- Average of imdb_score was highest for Biography at 8.50, followed by News and Action.
- Across all 27 genres_list, Average of imdb_score ranged from 6.10 to 8.50.

**B. Movie Duration Analysis:**
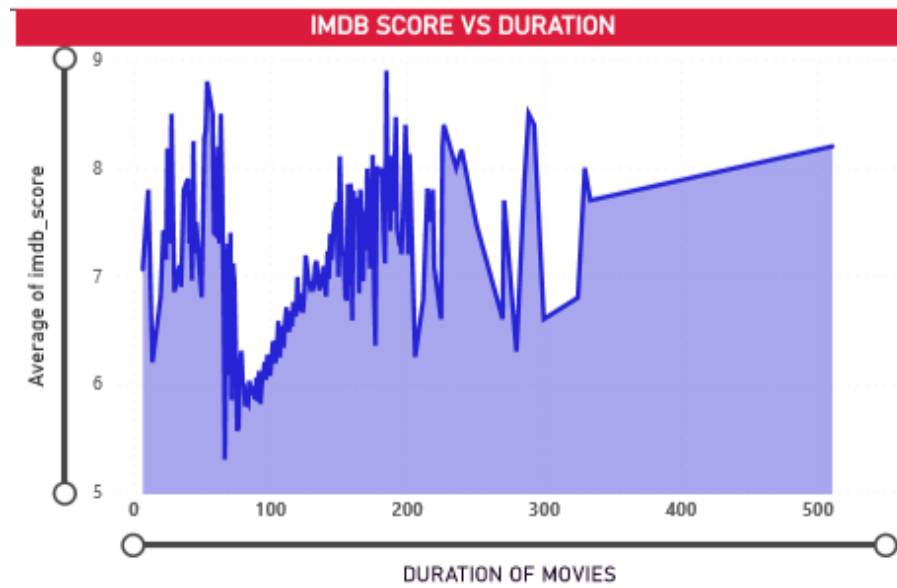
- At 8.90, 185 had the highest Average of imdb_score and was 67.92% higher than 67, which had the lowest Average of imdb_score at 5.30.
- 185 had the highest Average of imdb_score at 8.90, followed by 54 and 55. 67 had the lowest Average of imdb_score at 5.30.
- Across all 192 duration, Average of imdb_score ranged from 5.30 to 8.90.

C. Language Analysis:
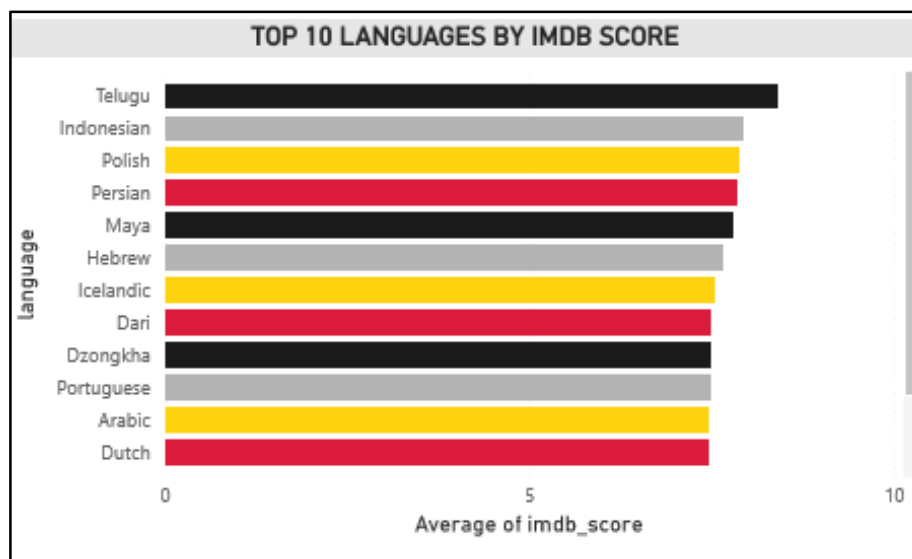
- Average of imdb_score was highest for Telugu at 8.40, followed by Indonesian and Polish.
- Across all 15 language, Average of imdb_score ranged from 7.40 to 8.40.

D. Director Analysis:
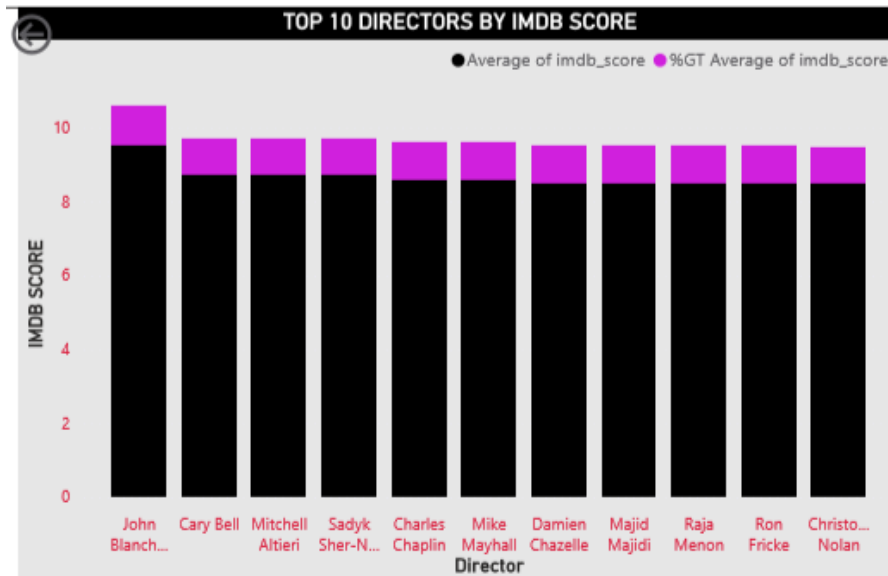
- The top 10 directors were shown based on IMDB score.
- The top most director is John Blanch with an average score of 9.50 followed by Cary Bell, Mitchell Alteri and Sadyk Sher-N with an average score of 8.70.
- Charles Chaplin and Mike Mayhall holds the third position with an average score of 8.60.

E. Budget Analysis:
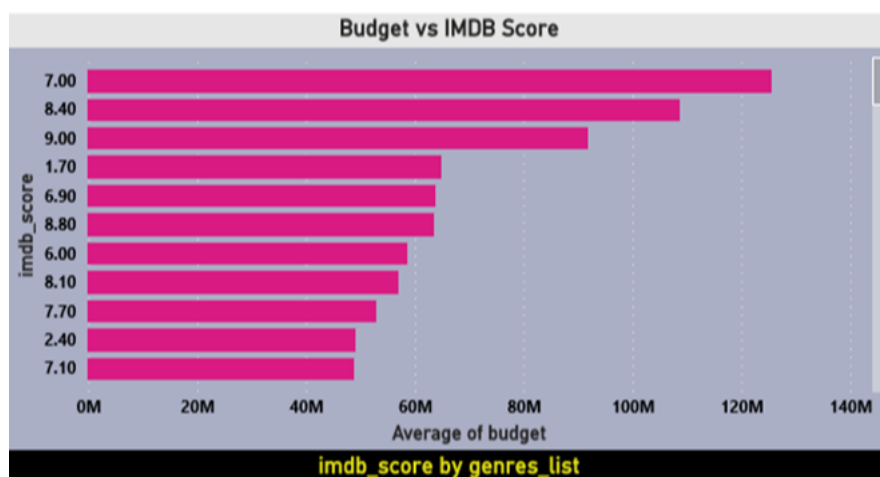
Top profit earned movies

movie_title	budget	gross	profit.1	profit
12 Rounds	22000000	12232937	9767063	TRUE
A Man Apart	36000000	26183197	9816803	TRUE
Alone with Her	1000000	10018	989982	TRUE
An Alan Smithee Film: Burn Hollywood Burn	10000000	15447	9984553	TRUE
Blonde Ambition	10000000	5561	9994439	TRUE
Cinco de Mayo, La Batalla	10000000	169379	9830621	TRUE
Edmond	10000000	131617	9868383	TRUE
Godzilla 2000	1000000000	10037390	989962610	TRUE
High Heels and Low Lifes	10000000	226792	9773208	TRUE
In Her Line of Fire	1000000	721	999279	TRUE
Jersey Girl	35000000	25266129	9733871	TRUE
Killer Elite	35000000	25093607	9906393	TRUE
Of Horses and Men	10000000	11835	9988165	TRUE
Poseidon	160000000	60655503	99344497	TRUE
Ravenous	12000000	2060953	9939047	TRUE
Resurrecting the Champ	13000000	3169424	9830576	TRUE
Safe Men	1000000	21210	978790	TRUE
September Dawn	11000000	1066555	9933445	TRUE
Special	1000000	6387	993613	TRUE
Swimming Pool	20000000	10105505	9894495	TRUE
The Adventures of Pinocchio	25000000	15091542	9908458	TRUE
The Art of War	40000000	30199105	9800895	TRUE
The Crew	23000000	13019253	9980747	TRUE
The Good, the Bad, the Weird	10000000	128486	9871514	TRUE
The Hebrew Hammer	1000000	19539	980461	TRUE
The In Crowd	15000000	5217498	9782502	TRUE
The Open Road	10000000	19348	9980652	TRUE
The Oxford Murders	10000000	3607	9996393	TRUE

- At 12,56,80,222.98, 7 had the highest Average of budget and was 6,030.74% higher than 2.60, which had the lowest Average of budget at 20,50,000.00.
- Across all 78 imdb_score, Average of budget ranged from 20,50,000.00 to 12,56,80,222.98.

F. Rating analysis:

- Count of content_rating was highest for 2009 at 740, followed by 2006 and 2014. 2009 accounted for 5.10% of Count of content_rating.
- Across all 91 title_year, Count of content_rating ranged from 2 to 740.

- At 8.30, TV-MA had the highest Average of imdb_score and was 32.50% higher than PG-13, which had the lowest Average of imdb_score at 6.27.
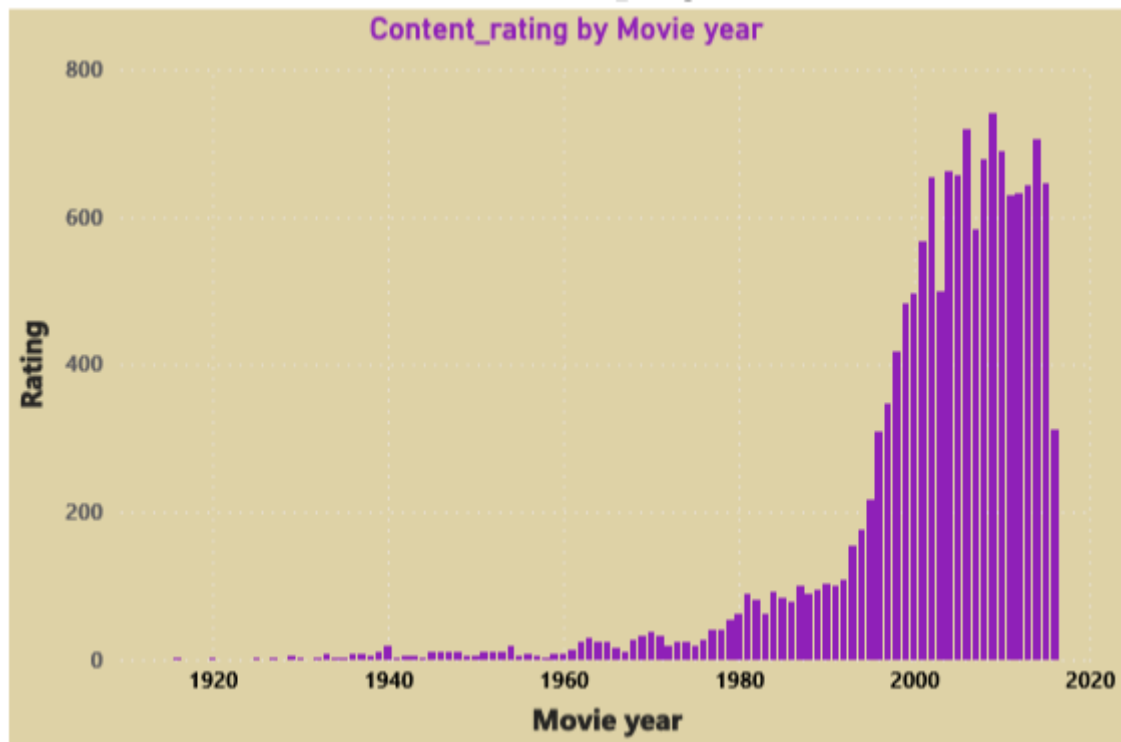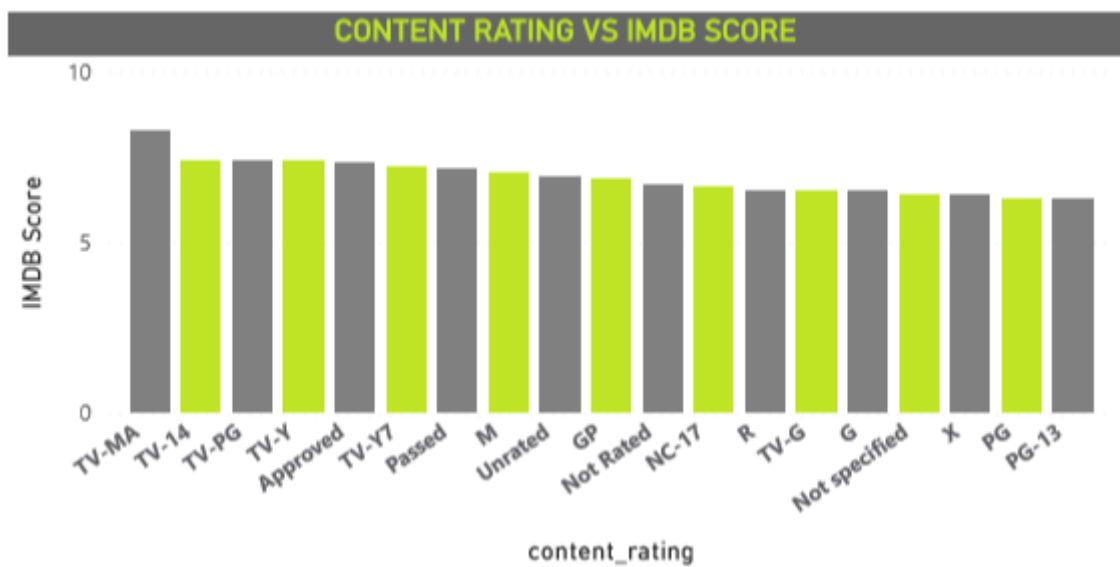- Across all 19 content_rating, Average of imdb_score ranged from 6.27 to 8.30.

Summary of IMDB Movie Analysis:

The analysis of IMDB movie data reveals several key insights regarding the factors influencing movie success, including content ratings, budgets, directors, languages, durations, and genres.

Content Rating Analysis:

- TV-MA content rating had the highest average IMDB score, significantly higher than PG-13.

- Content ratings varied widely in terms of IMDB score, with 2009 having the highest count of content ratings.

Budget Analysis:

- Movies with a budget of 7 had the highest average budget, followed by a substantial margin compared to the lowest average budget of 2.60.

- Budgets ranged from 20,50,000.00 to 12,56,80,222.98 across all movies.

Director Analysis:

- Top directors were identified based on their average IMDB scores, with John Blanch leading the list.

- Directors play a significant role in determining movie success, as reflected in the IMDB scores of their movies.

Language Analysis:

- Telugu emerged as the language with the highest average IMDB score, indicating its popularity among viewers.

- Languages vary in terms of their impact on movie ratings, suggesting potential opportunities for language-specific film production.

Duration and Genre Analysis:

- Movies with a duration of 185 minutes had the highest average IMDB score, indicating potential viewer preference for longer films.

- Biography genre had the highest average IMDB score, followed by News and Action, suggesting audience interest in these genres.

Implications for Stakeholders:

- Producers, directors, and investors can leverage insights from content ratings, budgets, directors, languages, durations, and genres to make informed decisions about future movie projects.

- Understanding the preferences of target audiences based on content ratings, languages, and genres can help in developing more appealing and successful movies.

Recommendations for Future Analysis:

- Further exploration of the relationship between specific genres, directors, and audience demographics could provide deeper insights into movie success factors.

- Conducting sentiment analysis on viewer reviews and social media mentions could supplement IMDB scores to provide a more comprehensive understanding of movie reception.

- Investigating the impact of marketing strategies, release timing, and competition within the film industry could offer additional insights into movie success determinants.

By leveraging these insights and recommendations, stakeholders in the film industry can enhance their decision-making processes and increase the likelihood of producing successful movies that resonate with audiences.