

Tmdb Movie Dataset Analysis Using Tableau

¹Hana Yadullah, ²Ayesha Asna, ³Izma Fatima, ⁴Hiba Anjum

^{1,2,3}UG Scholar, Department of AI & DS, Methodist College of Engineering and Technology, Hyderabad, India

⁴Assistant Professor, Department of CSE (AI&ML), Methodist College of Engineering and Technology, Hyderabad, India

Abstract - This project involves conducting an exhaustive analysis and visual representation of the Tmdb movie data. In connection with the rapidly growing movie industry, it becomes evident that there exist many data sets which can be useful for analysis and prediction purposes based on the extracted information. The present paper relies upon the structured data set of movies with various features such as genre, year of release, budget, profitability, length of the film, average rating of users, vote counts, and countries involved in the movie production. One can conduct the study of the dependencies related to the success of the movie by using these features. It is important to understand how financial features affect the success of the movie, the preferences of the audience in regard to certain movie genres, and trends related to the movie itself. Besides, it is possible to create models predicting future revenue and popularity of the movie using particular input attributes. Significant dependencies have been revealed in respect of the budget and profitability, preferences of audience concerning movie genres, and movie rating.

Keywords: Movie data set, Tableau, Data visualization, Exploratory Data Analysis (EDA), Dashboard design, Film analytics, Data storytelling, Box office analysis.

I. INTRODUCTION

The movie Industry has been expanding really fast, and with that growth comes a ton of data, such as budgets, revenues, ratings, genres, etc. The problem is, just looking at all that data doesn't tell you much. A big list of movies isn't helpful unless you actually check it out. That's where visualization tools such as Tableau come in. Instead of getting lost in numbers, you can actually see what's happening, like which genre makes the most money, whether ratings really matter, and so on. It makes everything a lot clearer and less emotional.

When you begin asking the right questions, things get really interesting. For example, does spending more on a film actually lead to more profit? It turns out there are real patterns in the data, but one needs to know where to look.

Using the Tmdb movie dataset and Tableau to explore those questions, such as like revenue by genre, audience

engagement, and what drives popularity. Less theory, more just looking at what the data actually shows.

II. LITERATURE SURVEY

The film industry continues on its rapid growth trajectory creating huge volumes of data. Box office statistics, production budgets, viewer ratings, genre performance and viewing behavior are just a few examples of the vast quantities of available film data. The issue arises from attempting to comprehend everything the data is telling us. Staring at piles of raw data accomplishes nothing; you need useful analysis tools to interpret the information.

Researchers have been discovering these useful tools for some time now. According to Shukla et al., machine learning has the capability to predict box office performance. Further than that, utilizing machine learning, Shukla and team found that budgets, timing of film releases and cast popularity are 3 significant predictors of box office success. This intuitively seems reasonable. [1]

On the visualization side Afzal et al. provide a strong case for the importance of visual analytics when working with large datasets [2]. When data is vast, it is essential to see the relationships in your data rather than merely reading about them. Roy builds upon this idea with his use of Tableau, illustrating how to transform data into graphs and dashboards can provide more effective decision making than guessing [6]. Bhoomika *et al.* build upon Roy's ideas and technology to show that interactive dashboards are very effective for analyzing trends in genre, ratings and audience engagement. Sentiment analysis has proven to be a valuable approach for examining audience opinions [7].

III. MATERIALS AND METHODS

Dataset: The data set being utilized in this research paper is called the Tmdb movie dataset that consists of structured data related to movies from various genres, nations, and periods of time. This dataset consists of categorical and numeric attributes including movie name, genre, release date, budget, revenue, vote average, number of votes, runtime, and production nation.

The main software tool used in this study for visualization and analysis is Tableau Desktop. This is an

interactive data visualization tool that makes it easy to explore the data using dashboards, charts, story and maps without necessarily knowing how to program.

Table 1: Dataset Parameters

Feature	Data Type	Description	Feature
Title	String	Name of the movie	Title
Genres	Categorical	Type of movie (Action, Comedy, Drama, etc.)	Genres
Release Date	Date	Movie release date	Release Date
Popularity	Float	Popularity score based on user activity	Popularity
Vote Average	Float	Average user rating (0–10)	Vote Average
Vote Count	Integer	Number of users who rated the movie	Vote Count
Runtime (min)	Integer	Duration of the movie	Runtime (min)
Budget (USD)	Integer	Production budget	Budget (USD)
Revenue (USD)	Integer	Earnings generated by the movie	Revenue (USD)

Software: Tableau

The main software tool used in this project is Tableau Desktop. Tableau is a powerful data visualization and business intelligence tool. The platform makes it easy to analyze complex data and visualize it in the form of tables, charts, maps, etc.

Tableau has been employed for data import, preprocessing, and data visualization tasks. Some of the key features in Tableau that made our work convenient include data filtering, data aggregation, calculations, and interactive dashboards. These allowed us to analyze the relationships among variables including revenues, budgets, genres, and rating by audience.

Apart from Tableau, the Microsoft Excel program has also been used for preliminary analysis of data and performing some initial preprocessing steps. Overall, the above-mentioned tools were helpful in conducting our data analysis work effectively.

The key features of Tableau include Data Connectivity, Interactive Dashboards, User-friendliness (no coding required), Powerful Analytics, Sharing & Collaboration, and Data Preparation.

IV. DATA VISUALIZATION

4.1 Revenue Analysis with Genre Filtering

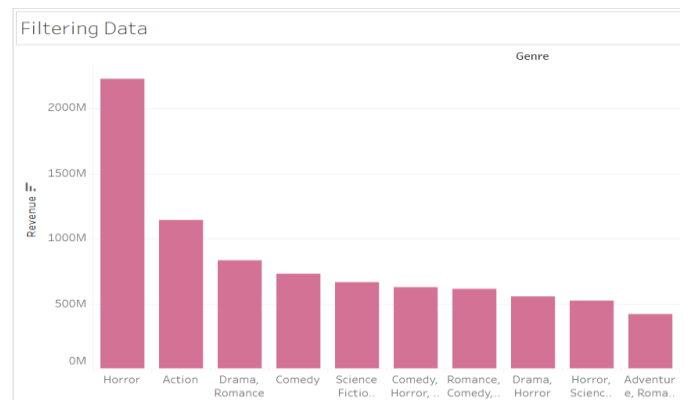


Figure 4.1: Filtering Data to refine dataset

The bar chart illustrates how much money each film category brings in. The x-axis represents different types of movies, while the y-axis indicates how much money was made from each individual movie type, or combination of movie types.

The bar chart shows that the highest grossing movie genre is Horror. It’s bar towers over any other bar. Second to that is Action. Other areas on the bar chart are moderate and include Drama - Romance, Comedy, and Science Fiction.

In comparison to Horror and Action, other types of movies such as Romance – Horror, Action – Comedy and Science Fiction have lower than average earnings, and thus their bars are shorter than the other types of movies. Using filters assists in comparing genres, and permits for a more in-depth view of certain types of movies. As a result, there is significantly less money made for specific genres compared to other types of movies.

4.2 Budget vs Revenue by Year

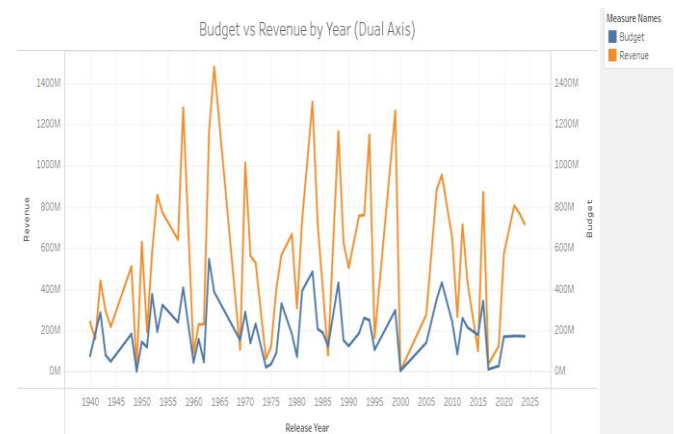


Figure 4.2: Budget vs Revenue by Year

The table also enables multi-dimensional analysis through the combination of categorical fields such as genre and production country with numerical measures such as vote average and vote total. This creates a way for the user to look at how movies from different production countries and genres accumulate audience ratings and popularity through ballots (from multiple production countries and multiple genres), thus illustrating the variance of production output and audience response across Germany, China, Japan, India, and the United Kingdom. The genres present for each movie further shows the amount of diversity that is available within each movie's content and gives an idea of how that diversity may correlate (or possibly correlate) with audience ratings.

4.5 Tree map (Revenue by Genre)

Treemap – Revenue by Genre

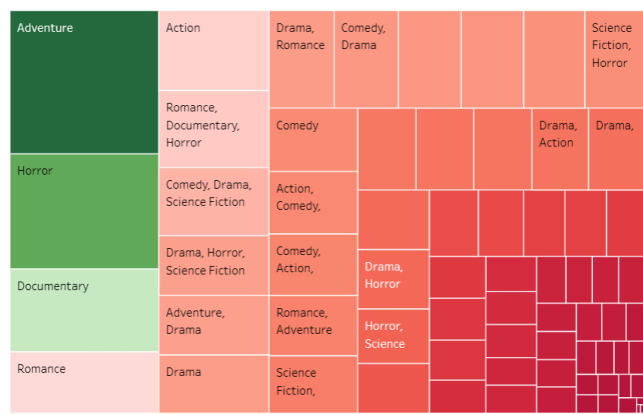


Figure 4.5: Tree map (Revenue by Genre)

The tree map graph gives an excellent insight into the distribution of total revenue generated by different types of movies according to the genre. Each box or rectangle represents either one genre or several combinations of genres. The size of the box represents the total revenue generated, while the color represents the amount of revenue generated, where green depicts high revenues and red depicts low revenues. It can be seen from the visualization that Adventure has the largest box area represented in green and thus has the highest revenue of all other genres. Horror and Documentary also generate relatively higher revenues. On the other hand, genres like Romance are depicted in red showing low revenue generation.

Few boxes represent various genre combinations like Drama-Romance, Comedy-Action and Science Fiction-Horror. The boxes are comparatively smaller and are in shades of red. This can suggest that multi-genre movies have revenue contributions ranging from medium to lower levels.

There are many smaller boxes representing uncommon genre combinations that have even lower revenues. It can be

clearly stated that from the tree map visualization, some major genres contribute substantially to the total revenue whereas other genres and combinations contribute relatively minor revenues.

4.6 Audience Engagement Efficiency by Genre

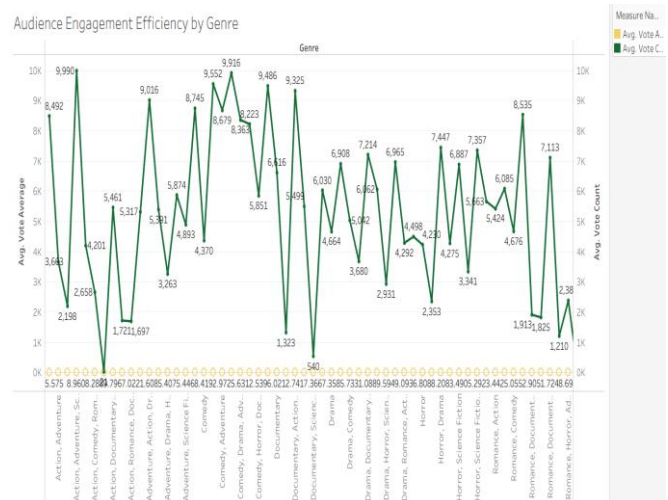


Figure 4.6: Audience Engagement Efficiency by Genre

This visual presents a comparison of movie genres in accordance with two indicators related to the audience behavior, namely the average rating and average vote count. On the horizontal axis, various genres and combinations of them are shown. The vertical axes show the average vote rating or average vote count depending on the genre.

First of all, one may notice that average rating values are rather constant for different genres, with minor variations observed. At the same time, the values of the average vote count vary greatly among various genres, with some spikes and drops observed. This means that there are some movies which gain more interest from the audience than other ones.

In particular, action, adventure, and mixed genres seem to attract high audience interest since their votes per review ratio is rather high. Some rare genres, such as documentaries or combined genres, receive less attention from the audience. Moreover, it appears that popularity does not coincide with a better average rating of a movie; that is, higher vote counts do not correlate with the rating.

V. CONCLUSION

This research involved analyzing the Tmdb movies dataset through data visualization tools provided by Tableau in order to establish trends within the aspects of profit generation, audience involvement, and international performance of films. Various data visualization tools such as

tree maps, bar graphs, line graphs, and maps have been used to identify trends based on genre, year, and origin of production.

Findings from this project include the realization that there is some contribution to revenue from different genres while ratings have not changed significantly among genres regardless of their popularity levels. Analysis of budget and revenue shows trends in profitability while geographically, profits have been generated only in a few nations.

The use of Tableau was quite successful as far as making visual sense out of data has been concerned. As seen above, insights from data have been drawn and the findings are useful in understanding the movie industry.

VI. FUTURE SCOPE

In light of the Tmdb dataset used, the research has been focused on exploratory analysis and visualization in Tableau. Future researches may continue working in this direction by adding machine learning models for prediction of the success of movies taking into consideration their budget, genre, cast members, release date, among other factors.

More precise conclusions could be reached by including current datasets and big data from different sources, for example, from media streaming services and social networking websites, thus obtaining a better knowledge of the preferences of audiences. In addition, a sentiment analysis of online comments about films and reviews would contribute significantly to understanding the preferences of viewers.

Also, it is possible to further develop technology solutions such as dynamic visualizations and analytics applications. In particular, an individual approach can be used, thus offering new opportunities to analyze and predict future developments in a variety of industries. Comparing several industries or domains at the same time would provide deeper insights into film studies.

Thus, it is important to combine visualization with predictive analytics and current datasets to improve movie data analysis.

REFERENCES

- [1] A. Shukla *et al.*, "Predictive Modeling of Box Office Success Using Machine Learning and Historical Movie Data," *ShodhKosh: Journal of Visual and Performing Arts*, 2025.
- [2] S. Afzal *et al.*, "Visualization and Visual Analytics Approaches for Image and Video Datasets: A Survey," *ACM Transactions on Interactive Intelligent Systems*, 2023.
- [3] D. S. Kwaghzan, "A Machine Learning-Based Platform for Sentiment Analysis of Movie Reviews," *International Journal of Computer Science and Mathematical Theory*, 2025.
- [4] Y. S. You *et al.*, "Design and Development of Visualization Tool for Movie Review and Sentiment Analysis," *Proceedings of EDB Conference*, 2016.
- [5] S. Bohra *et al.*, "Hybrid Machine Learning Based Recommendation Algorithm for Multiple Movie Dataset," *Indian Journal of Science and Technology*, 2023.
- [6] S. Roy, "Data Visualization on Movies Dataset using Tableau," *International Journal of Engineering Trends and Technology*, 2019.
- [7] R. Bhoomika *et al.*, "Data Analytics and Visualization Using Tableau for IMDb Movies & Shows," *IJPREAMS*, 2024.
- [8] R. Roopa and H. Khanam, "A Comprehensive Study on Advanced Video Data Preprocessing Techniques for Enhanced Object Detection," *REST Journal on Data Analytics and Artificial Intelligence*, 2024.
- [9] S. Few, *Information Dashboard Design: The Effective Visual Communication of Data*, O'Reilly Media, 2013.
- [10] The Movie Database (TMDb), "TMDB 5000 Movie Dataset," *Kaggle*, 2017.

Citation of this Article:

Hana Yadullah, Ayesha Asna, Izma Fatima, & Hiba Anjum. (2026). Tmdb Movie Dataset Analysis Using Tableau. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(5), 484-488. Article DOI <https://doi.org/10.47001/IRJIET/2026.105067>
