



CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE • INDIA

Comprehensive Movie Industry Insights: A Multidimensional Overview

By

Alaka P (2348006)

Ann Mathew (2348010)

Navya Maria Jomy (2348044)

Philsy V Martin (2348046)

4 MSc Data Science

(CAC-2)

1.DATASET DESCRIPTION

Dataset Name: TMDB 5000 Movie Dataset

Source: Kaggle, originally from The Movie Database (TMDb)

This dataset contains comprehensive information about movies, including details about their genres, production, cast, crew, and financial performance.

Key Features:

1. **Comprehensive Credits:** The dataset includes full credits for both cast and crew, providing a more complete picture of each movie's production team.
2. **Ordered Cast Listing:** Actors and actresses are listed in the order they appear in the movie credits, offering insight into their prominence in each film.
3. **Updated Financial Data:** The revenue figures appear to be more current than in previous versions, potentially providing a more accurate representation of each movie's financial performance.
4. **Genre Information:** The dataset includes binary indicators for various genres, allowing for easy categorization and analysis of movies by type.
5. **Production Details:** Information about production companies and countries is included, offering insights into the global nature of film production.
6. **Language Information:** The dataset contains data about the languages spoken in each film.
7. **Keywords and Themes:** There's a column for keywords, which likely describes themes, plot elements, or other notable features of each movie.

Dataset Structure:

- **Rows:** 4,803 (each row likely represents a unique movie)

- **Columns:** Multiple columns covering various aspects of each movie:

Budget: The movie's production budget, **Genres:** List of genres associated with the movie, **Homepage:** URL of the movie's homepage, **ID:** Unique identifier for the movie, **Keywords:** List of keywords relevant to the movie's plot, **Original Language:** Language in which the movie was originally released, **Original Title:** The original title of the movie, **Overview:** Brief description of the movie's plot, **Popularity:** Popularity score of the movie, **Production Companies:** List of companies involved in the production of the movie, **Production Countries:** Countries where the movie was produced, **Release Date:** Release date of the movie, **Revenue:** Total revenue generated by the movie, **Runtime:** Duration of the movie in minutes, **Spoken Languages:** Languages spoken in the movie, **Status:** Current status of the movie (e.g., Released), **Tagline:** Movie tagline, **Title:** Title of the movie, **Vote Average:** Average vote rating given by viewers, **Vote Count:** Total number of votes.

2.OBJECTIVES

Financial Performance Analysis:To explore the relationship between movie budgets and revenues, analyzing how financial investments affect profitability and identifying key financial trends and successful films across different decades.

Temporal Analysis: To analyze the evolution of the movie industry over time, focusing on genre-specific financial performance, popularity trends, and overall industry growth from 1916 to 2016.

Production Company Performance: To evaluate the performance of production companies in terms of revenue, profitability, and popularity, identifying key industry players and their impact on the movie landscape from the 1990s to the 2010s.

Audience Reception: To assess how different genres and production companies influence audience engagement and popularity, examining trends in audience interaction, vote averages, and genre performance over the decades.

3. DATA PREPROCESSING

Normalization and Parsing:

Genres and Keywords: Converted from JSON format to a readable list of strings. Each genre and keyword was extracted and normalized for consistent analysis.

Production Companies and Countries: Parsed from JSON format into individual fields for each movie, ensuring consistent naming conventions.

Data Cleaning:

Handling Missing Values: Rows with missing or null values in critical fields like Revenue, Budget, or Vote Average were either filled with appropriate placeholders or removed, depending on their impact on the analysis.

Consistency Checks: Verified that all fields adhered to the expected data types and formats (e.g., dates were standardized, numeric fields were checked for correctness).

Filtering and Aggregation:

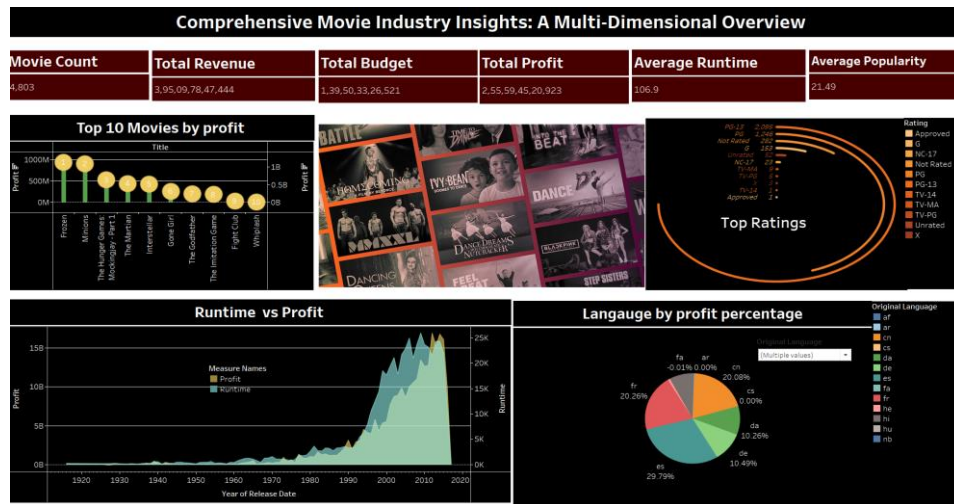
Date Ranges: Filtered data to include specific decades (1990s, 2000s, 2010s) for focused analysis.

4. STATISTICAL OVERVIEW

Movie Count	– 4803
Total Revenue	– 3,95,09,78,47,444
Total Budget	– 1,39,50,33,26,521
Total Profit	– 2,55,59,45,20,923
Maximum Revenue	-2,78,79,65,087
Maximum Budget	-380,00,000
Maximum Profit	-2,55,59,5,20,923
Average Runtime	– 106.9
Average Popularity	– 21.49
Average Budget	– 29,045,040
Average Profit	-53,215,599
Average Rvenue	-82,260,639
Average Age	-21
Average Vote Count	-6
Top Movie	– Avatar
Top Movie Revenue	-2,787,965,087
Total Companies	-1313

5.DASHBOARDS(Story)

5.1. OVERVIEW



The "Comprehensive Movie Industry Insights" dashboard presents an overview of the film industry, focusing on key metrics such as Movie Count, Total Revenue, Budget, Profit, Average Runtime, and Popularity. It features visualizations like the "Top 10 Movies by Profit," with "Frozen" leading, a chart on movie ratings dominated by PG-13 and PG, and a "Runtime vs Profit" graph that tracks profitability trends over time. Additionally, a pie chart shows profit distribution by movie language, highlighting English and French. This dashboard offers a concise analysis of industry trends, profitability, and production characteristics.

5.2. FINANCIAL ANALYSIS

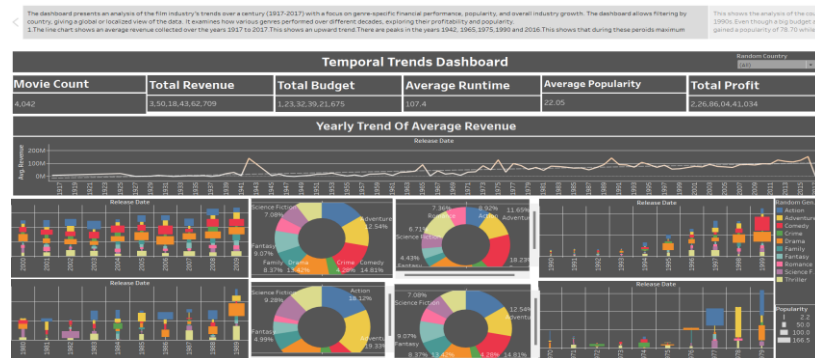
Movie Industry Financial Analysis by Decade



Description: **This dashboard** provides a deep dive into the financial landscape of the movie industry, offering insights into key metrics like revenue, profit, budget, genre distribution, company performance, and popularity trends. The top section of the dashboard highlights essential figures such as the average budget, total revenue, top-grossing movies, and the total number of films analyzed, offering a clear financial overview. The dashboard also breaks down the genre distribution of movies over the years, revealing trends and shifts in audience preferences, which can help identify potential investment opportunities. Movie Profit Margin Breakdown categorizes films into different profitability groups, offering a snapshot of the industry's financial health and potential for high returns. Additionally, the dashboard explores

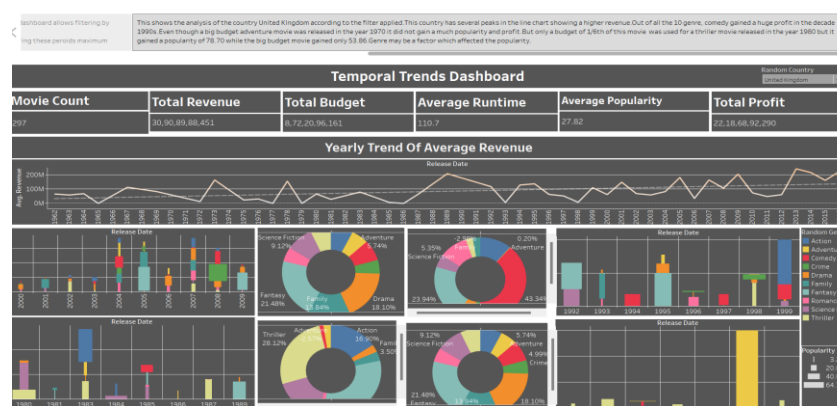
the correlation between movie popularity and profit, informing marketing strategies. The Top Production Companies section reveals the market share of leading production companies, helping to identify dominant players and emerging competitors.

5.3. TEMPORAL TRENDS



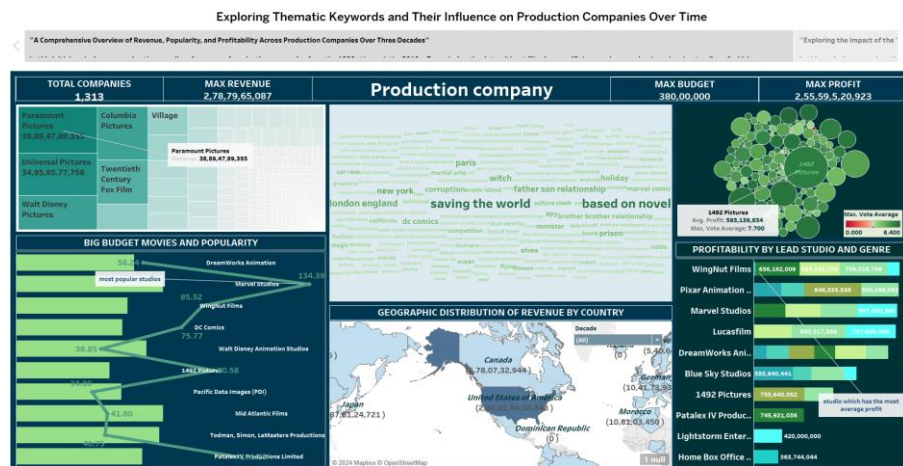
Description: The dashboard presents an analysis of the film industry's trends over a century (1917-2017) with a focus on genre-specific financial performance, popularity, and overall industry growth. The dashboard allows filtering by country, giving a global or localized view of the data. It examines how various genres performed over different decades, exploring their profitability and popularity.

The line chart shows an average revenue collected over the years 1917 to 2017. This shows an upward trend. There are peaks in the years 1942, 1965, 1975, 1990 and 2016. This shows that during these periods maximum revenue was collected globally. The second chart is the stacked bar chart for the years 2000s genre wise budget and popularity is given as the width size. The second is the doughnut chart which shows the percentage of the profit for each genre at 2000s. The third set of chart which is the same stacked bar chart and the doughnut chart is for the years 1990s. Likewise for 4 decades the same analysis is done.



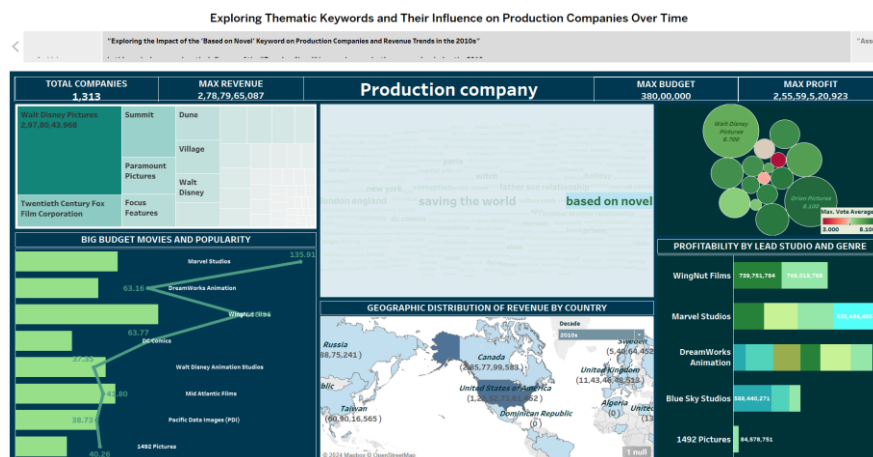
Description: This shows the analysis of the country United Kingdom according to the filter applied. This country has several peaks in the line chart showing a higher revenue. Out of all the 10 genre, comedy gained a huge profit in the decade 1990s. Even though a big budget adventure movie was released in the year 1970 it did not gain a much popularity and profit. But only a budget of 1/6th of this movie was used for a thriller movie released in the year 1980 but it gained a popularity of 78.70 while the big budget movie gained only 53.86. Genre may be a factor which affected the popularity. Like wise, each country's revenue profit and buget can be analyzed for each decade from 1970 to 2010.

5.4. PRODUCTION COMPANY



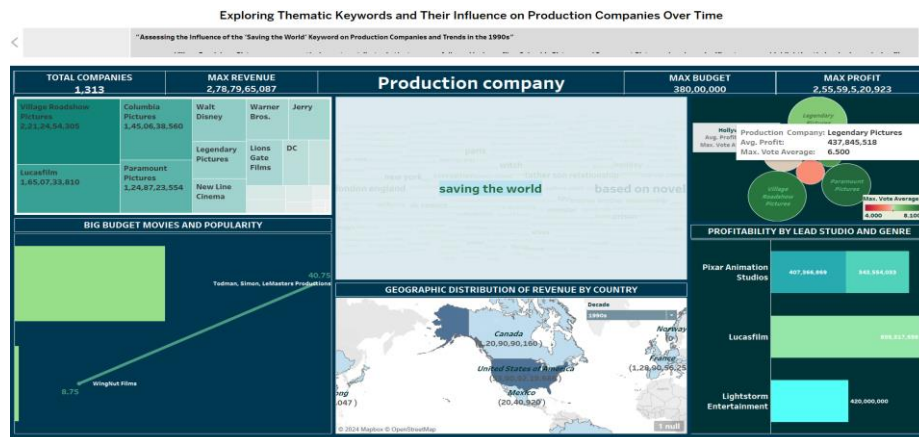
Description: In this initial analysis, we examine the overall performance of production companies from the 1990s through the 2010s. By exploring the data without filtering specific keywords, we gain a broad understanding of which production companies consistently generate high revenue, receive strong vote averages, and achieve significant profitability. This view allows us to identify the key players in the movie industry and observe the trends that have shaped the landscape over three decades, providing a foundational overview of revenue distribution, regional success, and the impact of high budgets on popularity.

Paramount Pictures emerged as the production company with the highest revenue. 1492 Pictures was identified as the most profitable studio when considering only profits without factoring in losses. When losses were also considered, WingNut Films stood out as the most profitable production company over the last three decades. Marvel Studios was found to be the most popular production company during this period



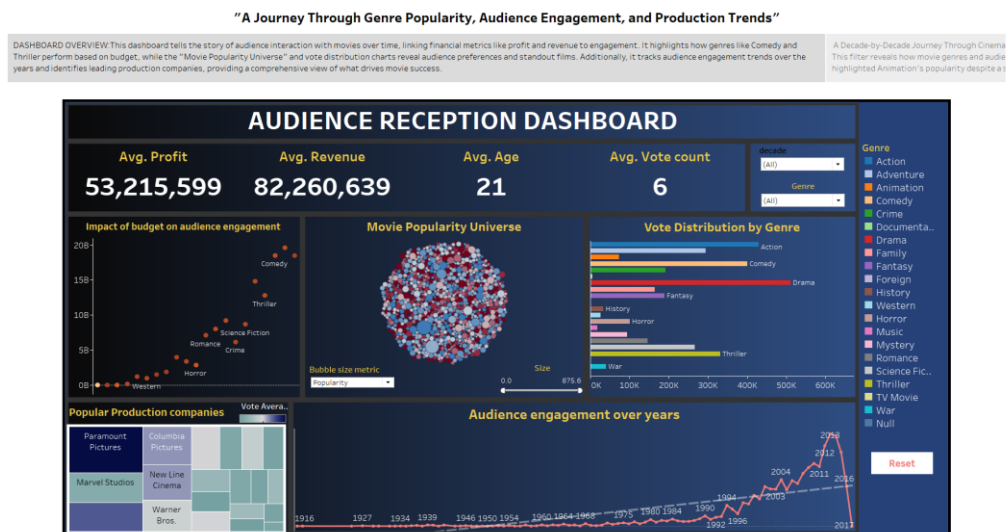
Description : In this analysis, we explore the influence of the "Based on Novel" keyword on production companies during the 2010s.

Walt Disney Pictures stands out in the treemap and bubble chart indicating its significant role in producing novel-based films. The larger size in the treemap reflects high total revenue, while the darker color emphasizes this revenue magnitude. Orion Pictures also appears prominently in the bubble chart, where the darker shade signifies a higher vote average, despite its slightly smaller size than Walt Disney. The filter applied to the 2010s reveals that WingNut Films achieved the highest total revenue while Marvel Studios showed greater popularity, as indicated by the higher peak in the line chart. In the map the United States dominates revenue, with American production companies generating the most from novel-based movies. Marvel Studios led in average profit, particularly in the Science Fiction genre followed by DreamWorks Animation and WingNut Films.

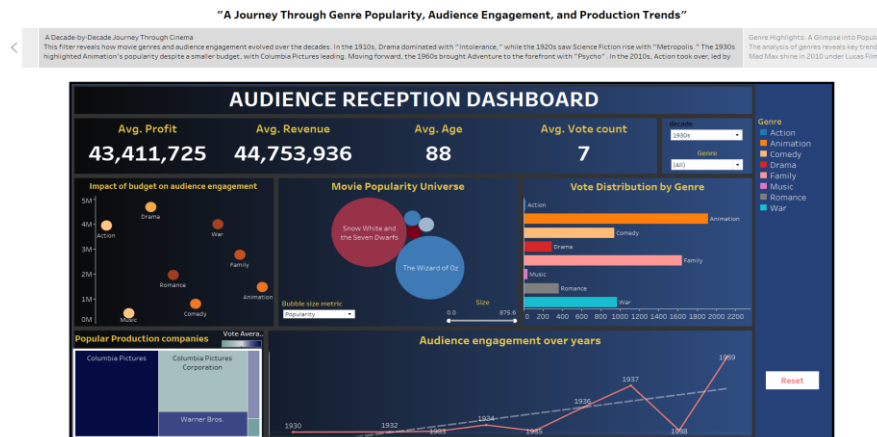


Village Roadshow Pictures emerges as the largest contributor in the treemap, followed by Lucasfilm. Columbia Pictures and Paramount Pictures also show significant revenue, highlighting their roles in producing films with this theme. In the bubble chart, Legendary Pictures and Village Roadshow stand out in both large size and darker shades, indicating high maximum vote averages. Hollywood Pictures, however, appears much smaller and in a redder shade, signaling a lower vote average. The filter applied to the 1990s reveals in the dual axis bar chart and line chart that Todman, Simon, Lemasters Production, and WingNut Films were notable, with Todman achieving the highest popularity during these decades. In the bar chart Lucasfilm led in average profit, driven by the Adventure genre, while Pixar Animation Studios followed closely with contributions from the Comedy and Animation genres.

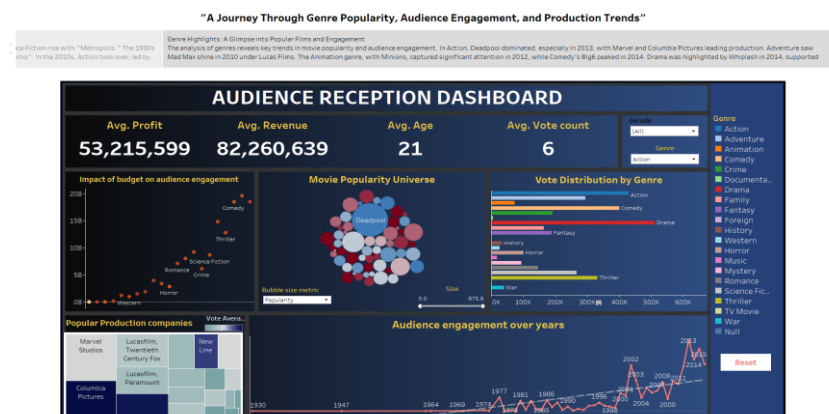
5.5 AUDIENCE RECEPTION



Description : This dashboard narrates how audiences have interacted with and received movies over time, revealing the intricate relationships between financial aspects and audience engagement. We gain a high-level view of overall movie performance by examining average profit, revenue, and vote counts. The dashboard delves deeper into genre-specific analysis, showing how budget influences audience engagement, with Comedy and Thriller genres standing out. The "Movie Popularity Universe" visualizes the widespread appeal of films, highlighting which ones capture the most attention. Additionally, the vote distribution by genre uncovers audience preferences, illustrating the competitive landscape among different genres. Over time, the "Audience Engagement over Years" chart reveals evolving trends, while insights into leading production companies show who commands the most audience favor. This dashboard tells a comprehensive story of how financial and creative elements converge to shape audience reception in the movie industry.



: Description This interactive dashboard unravels the evolution of the movie industry across decades, spotlighting the shifting dynamics of genres, popular films, and audience engagement. Starting from the 1910s, where Drama dominated and "Intolerance" stood out, the story progresses to the 1920s, revealing the rise of Science Fiction with "Metropolis" as the most popular film. The journey through the 1930s highlights the growing presence of genres, with Animation emerging as a surprise leader despite a lower budget, while Columbia Pictures shined in 1939 with the highest audience engagement. Adventure takes center stage in the 1960s with "Psycho," while United Artists tops the charts. The 1970s revisit Drama, with "Alien" and Lucasfilm capturing the spotlight. Action explodes in the 1980s, with Paramount Pictures leading in audience engagement. Moving into the 2000s, Drama remains prominent, with New Line Cinema rising, while the 2010s shift towards Action, led by Marvel Studios, illustrating the changing tides of the film industry over the years.



Description: Exploring various genres reveals intriguing trends in movie popularity and audience engagement over the years. In the Action genre, Deadpool emerged as the most popular film, with top production companies like Marvel, Columbia Pictures, and Lucas Films leading the charge, especially in 2013, a peak year for audience engagement. Adventure films, led by Mad Max, saw Lucas Films dominating, with 2010 marking a significant year. In Animation, Minions captured hearts in 2012, while in Comedy, Big 6 reached new heights in 2014 under Paramount Pictures' banner. With Whiplash at the forefront, the Drama genre experienced a surge in 2014, bolstered by Columbia and Paramount Pictures. Other notable mentions include the Crime genre's Dark Knight in 2008, Family's Despicable Me 2 in 2004, and Fantasy's Batman vs Superman in 2012. These findings underscore the vital role of production houses in shaping audience reception across different genres over time.

CONCLUSION

Based on the comprehensive analysis of the TMDB 5000 Movie Dataset, several key insights emerge about the film industry's evolution from 1916 to 2016. The data reveals significant trends in financial performance, audience reception, and genre popularity across different decades, with notable shifts in dominant production companies and emerging markets. This analysis provides valuable information for industry professionals, highlighting the importance of adapting to changing audience preferences, the impact of high-budget productions on popularity, and the growing influence of specific genres and production companies in shaping the global cinematic landscape.