PROCESS BOOK

Nandini Sharma   *Lekhana Chowdary*   *Sravanth Thatavarthi*

nandins@clemson.edu   lsingam@clemson.edu   sravant@clemson.edu

Group: GR8

Dataset: https://docs.google.com/spreadsheets/d/12bnGB7w5T03f7Bq1PG6gqFIdsLccdP6E-86xdVjbCOc/edit#gid=423546026

GitHub Repository: https://github.com/nandini1799/toWatch-orNot

GitHub Link: https://nandini1799.github.io/toWatch-orNot/

Project Screencast: https://youtu.be/Qh9Bw1lhmCA

REEL INSIGHTS

Overview and Motivation:

Enrico Bertini argues, "We do not visualize data, we visualize ideas based on data." Our primary motivation was to highlight how studying past trends in movies, will help in creating films that will be better appreciated by the audience. The world of cinema has always fascinated and highly influenced our society. Movies not only entertain us, but they also function as cultural mirrors, reflecting our dreams, how our society thinks, and how it functions, including everything the good and the bad. The film industry has changed dramatically over the years, it has now become a complex industry where creativity meets intricate commercial methods. Movies are no exception in this digital age, a vast amount of data is produced every day in the movie business. For thousands of films, millions of movie reviews, box office collections, production budgets, and other associated data points are available. "Reel Insights," our project, seeks to bridge this gap by bringing the intriguing world of movies to life through data visualization.

The dashboard provides a comprehensive exploration of the film industry through various visualizations. The Simple Bar Chart offers a straightforward view of IMDb ratings over the years, while the dynamic Heat Map illustrates the density of films with similar ratings annually. The Pie Chart dynamically showcases script distribution, and the Stacked Bar Chart presents financial perspectives based on IMDb ratings. The Grouped Bar Chart delves into the financial performance of different film genres globally and domestically. Additionally, Line Charts track genre and runtime shifts across years.

Related Work:

A. Visual inspirations:

We examined various dashboards to gain visual inspiration for conceptualizing our dashboard layout and color scheme. Among the dashboards reviewed, two particularly resonated with the theme of our project, aligning well with the data we were handling and complementing

our conceptual vision. These two dashboards are identified as Fig 1: Dashboard based on IMDB movies dataset and Fig 2: Dashboard based on Disney movies dataset.

Fig 1: Dashboard based on IMDB movies dataset
Source: https://www.thedataschool.co.uk/shreya-arya/imdb-movies-dashboard-week-day-2/

Fig 2: Dashboard based on Disney movies dataset

Source: https://insights-on.com/disney-movies-analysis-power-bi/

B. Pragmatic Inspiration:

We explored various blog posts and research papers to get ideas for our project. One blog post that stood out was "Exploration of The Movie Database (TMDB) Data" by Alvira Swalin and Kishan Panchal[1]. It gave us good advice on which columns in the dataset to focus on so that our visualizations are clear and efficient. The insights derived from this blog post significantly influenced our decision-making process, particularly in the selection of appropriate graph types for our visualizations.

Questions:

The initial questions that guided our project in the beginning were aimed at understanding various aspects of movie performance. These included:
- Which genres are thriving at the box office, considering factors like budget, revenue, and budget recovery?
- How does the performance of different genres compare in the domestic market versus the international market?
- What patterns exist in the distribution of movie scripts throughout the year?

However, as the project progressed, these questions evolved into more refined and specific inquiries:
- Does the Rotten Tomato rating of movies have an influence on their opening weekend box office collections?
- How do the domestic and international gross earnings of various genres change from year to year?
- How many movies were produced in each genre across different years?
- How has the runtime of movies evolved over the years for different genres?
- How many movies released in a particular year share a similar IMDB rating, and what is the average IMDB rating for these movies?

Data:

The dataset was sourced from Kaggle [2] using search terms like "movie," "genre performances," "IMDb ratings," and "box office collections." It consists of 26 columns and 1695 rows, providing a comprehensive set of information related to movies. The dataset's dimensions include 26 columns and 1695 rows, offering a diverse range of information on movie details, rotten tomato ratings, IMDb ratings, box office collections and more.

In order to gather additional movie data, we utilized the OMDB API [3] to extract runtime and movie director information from IMDb. The API calls retrieved this data, and we stored the results in a JSON file on the backend. Recognizing the large volume of data handled by the dashboard, we opted to host the backend on RENDER [4]. This approach helps ensure that the dashboard remains responsive and does not become burdensome, allowing for quicker loading times and an overall smoother user experience.

Exploratory Data Analysis:

In our initial exploration of the dataset, we utilized the following visualizations:

1. Grouped Bar Chart for Domestic vs. International Genre Performances: This chart illustrates how different genres perform in terms of box office revenue, distinguishing between domestic and international markets. It helps capture genre-specific audience preferences.
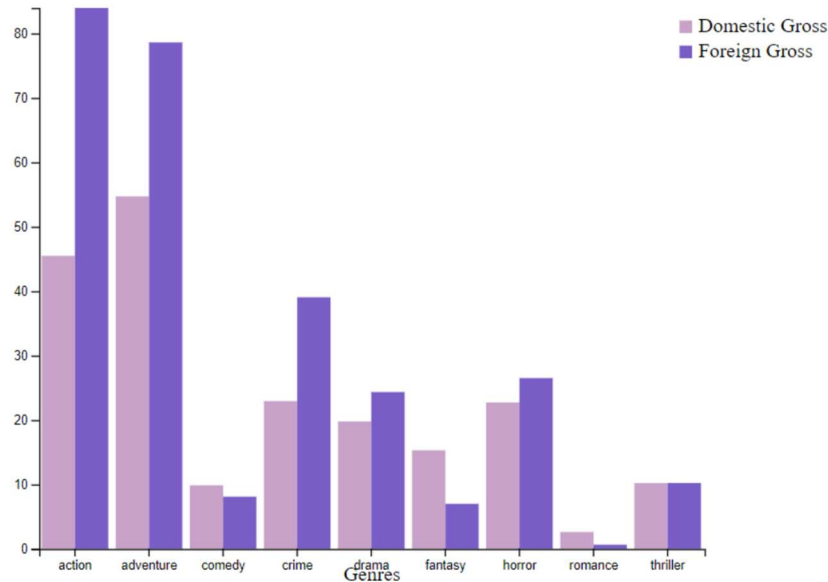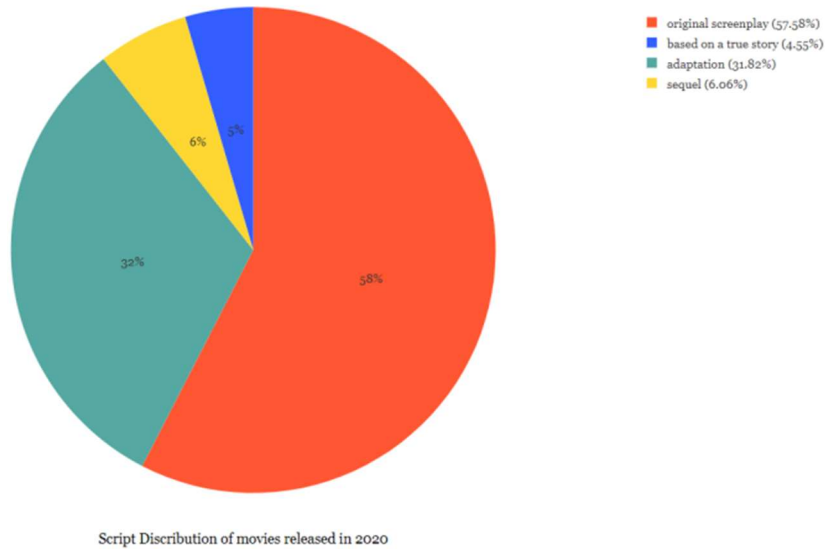

Fig 3: Grouped Bar chart

2. Stacked Bar Chart for Budget Recovery by Genres: The stacked bar chart visualizes the relationship between box office collection, budget, and revenue for each genre. This allows us to identify patterns in the performance of movie genres.
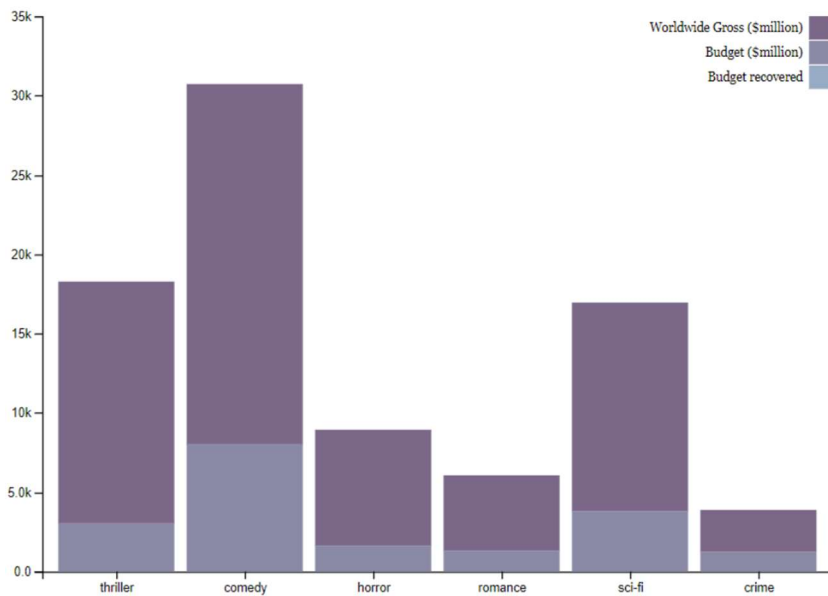
Fig 4: Stacked Bar chart

3. Pie Chart for Movie type Distribution: This visualization describes how many scripts of movies that are made in a year are original, a remake, an adaptation, or a sequel. To give an overview a general in movie's scripts.

Script Discribution of movies released in 2020

- original screenplay (57.58%)
- based on a true story (4.55%)
- adaptation (31.82%)
- sequel (6.06%)

Fig 5: Pie Chart

4. Stacked Area Chart for Shift of Genres Over the Years: The stacked area chart displays the number of movies released in different genres over the years, with each genre represented by a distinct color. This visualization helps highlight genre popularity trends and shifts over time.

We opted against using the stacked area chart due to potential for misinterpretation. The visual complexity with multiple categories and color reliance further prompted us to choose alternative visualization method like the line chart better aligned with our dataset and analytical objectives.

Expanding on the insights gained from our visualizations, we recognized an opportunity to enhance the stacked bar chart by introducing an additional category: Rotten Tomato ratings. This modification aims to deepen our understanding of the relationship between movie ratings and box office collections, offering a more comprehensive view of audience preferences and potentially revealing correlations between critical acclaim and financial success.

Design Evolution:
We deviated from our initial proposal in terms of the chosen visualizations to address more insightful questions, leading us to modify the number of visualizations on our dashboard. This adjustment was made to better align with our evolving goals and the pursuit of deeper insights in our analysis. Initially, we employed D3 and JavaScript (JS) for the development of our visualizations. However, as the project progressed, we transitioned to using ReactJS in conjunction with D3. This shift allowed us to leverage the component-based structure of React for more efficient and modular development, enhancing the overall scalability and maintainability of our visualizations.

The following are the descriptions of final visualizations in our dashboard.

- **Simple Bar Chart:** The primary bar chart on our dashboard has an x-axis for release years and a y-axis for IMDb ratings. Each bar represents a film released in a certain year, with the height representing the IMDb rating.

- **Heat Map:** Our primary heatmap visualization is a dynamic representation, featuring IMDb ratings ranging from 2 to 9 on the y-axis and years on the x-axis. The color saturation in each cell serves as a unique indicator: the greater the saturation, the higher the number of films sharing the same IMDb rating in a particular year.

- **Pie Chart:** The pie chart dynamically reflects the distribution of film scripts, adjusting with the selected year. The chart categorizes script types into five sectors: Original Screenplay, Remake/Reboot, Sequel, Adaptation, and Others.

- **Stacked Bar Chart:** The stacked bar chart is another robust visual depiction, with profits in millions of dollars on the y-axis and IMDb ratings on the x-axis. Each bar is divided into two sections: the blue area shows opening week earnings, while the orange area shows overall earnings in millions.

- **Grouped Bar Chart:** The grouped bar chart illustrates the financial performance for different film genres at the domestic and worldwide box office. The x-axis shows earnings in millions of dollars, while the y-axis shows different genres. There are two bars for each genre, one for domestic gross and one for foreign gross.

- **Line Chart:** We included two different line charts in our dashboard. The first line chart reflects genre shift across years with the years on x-axis and number of movies of the specific genre on y-axis. The second line chart represents runtime shift across years with years on x-axis and average runtime on y axis.

We selected the above visualizations by applying the following perceptual and design principles learned in the course:

- **Clarity and Simplicity:** To make the visualisations easy to understand, we tried to keep the design simple. Simple and uncomplicated charts improve the viewer's comprehension.
- **Use of Colors:** The selection of colours was done with great care to improve visual clarity and prevent possible problems with colour perception. Different categories or genres were represented by distinct colour choices. We decided to use a dark mode background in our dashboard serves to enhance user experience by reducing eye strain, providing a modern aesthetic, and emphasizing content vibrancy.
- **Accessibility:** To ensure that the charts are understandable to a wide range of viewers, accessibility criteria were considered during the visualisation selection process.

Implementation:

Interactive Features: The primary bar chart serves as a control element for each of the three visualizations on our dashboard. It allows users to interactively navigate and filter data by selecting a year of their interest. The selection of a year on this control chart dynamically changes and updates the information displayed in the accompanying visualizations like the pie chart, stacked bar chart and grouped bar chart, and line charts resulting in a cohesive user experience (Fig 6).

Fig 6: Interactivity with the simple bar chart at the top of the dashboard.

Upon selecting a specific year within the heatmap, the dashboard undergoes a synchronized update. This includes the pie chart, stacked bar, and the grouped bar chart, line charts, all adjusting dynamically to showcase relevant data for the chosen year (Fig 7).

Fig 7: Interactivity with heatmap

Additionally, upon selection of bar in a particular release year on simple the bar chart, a movie detail card appears, providing a curated summary of key information. This includes information such as the film's director, genres, IMDb rating, and a visual touch with the movie poster. Users can quickly access a variety of information about a selected film, enriching their exploration (Fig 8,9).
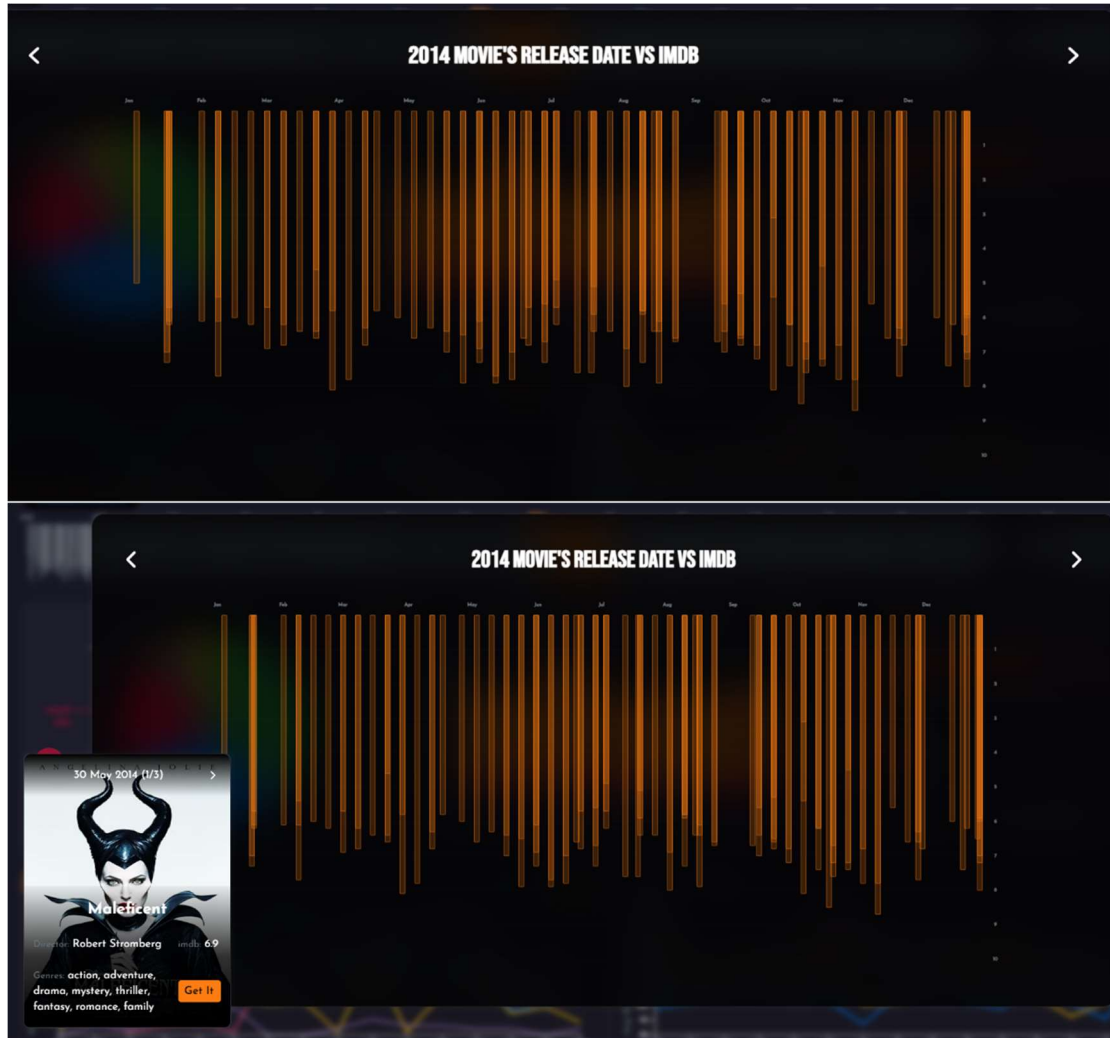
Fig 8,9: Movie title card interaction

To enhance clarity and minimize visual clutter, we implemented an interaction feature that enables users to select their desired genre, facilitating a focused understanding of the two line charts positioned at the end of the dashboard. These charts specifically illustrate shifts in genre popularity and movie runtimes over time, allowing users to explore the data in a more targeted and comprehensible manner(Fig 10).

Fig 10: Feature to select a specific genre

Evaluation:

We believe doing a pilot study in which participants interact with the dashboard would help to further improve the visualisation. Evaluating their insights and using this feedback to enhance the visualisation and make it more user-friendly.

Learnings from visualisations:

- Each sector in the pie chart emphasizes distinct components of storytelling by examining writing preferences across cinematic time periods, indicating shifts in preferred scripts with each chosen year.

- The stacked bar chart combines critical and financial metrics, allowing users to understand the relationship between IMDb ratings and movie financial success. The stacked structure allows for a clear comparison of the initial commercial impact during the opening week and the cumulative earnings, providing useful insights into the relationship between critical acclaim and box office performance.

- The grouped bar chart visualization allows viewers to easily browse between years, providing a thorough overview of genre-specific performances on both local and worldwide parameters. Users can quickly determine whether genre prevails in domestic or international gross earnings for a given year by analysing the chart. Furthermore, it allows an examination of genres with the greatest discrepancies between domestic and international box office profits, providing useful information into audience preferences and box office market dynamics.

- The line chart about movie genres helps us understand which genres have been more popular during different times, giving insights into how many movies where made in a genre across years. For example, the number of adventure movies made as steadily declined from year 2020 and after in comparison to the years 2018, 2019
- The line chart about runtime helps in understanding the change in movie runtimes across years. For example, there is a significant increase in the average runtime of action movies from the year 2020.

References:

[1] Source: https://sites.google.com/dons.usfca.edu/movieviz/home

[2] Kaggle: https://www.kaggle.com/

[3] OMDB API https://www.omdbapi.com/

[4] Render: Cloud Application Hosting for Developers | Render