CINEMATIC SUCCESS

404 NOT FOUND

COM-480
DATA VISUALIZATION
@EPFL 2025

BURKI BENJAMIN
CORNAZ ANTOINE
HUANGFU CHARLINE

PROJECT GOAL

The film industry stands as one of the most powerful and profitable sectors in global entertainment. Over the decades, it has undergone significant transformations, shaped by shifts in popular genres, production budgets, audience expectations, and technological innovations. For filmmakers, production companies, and marketers, gaining a clear understanding of movie trends, viewer preferences, and financial outcomes is essential.

This project explores a dataset covering over six decades of cinematic history (1960–2025), aiming to identify patterns behind successful films and to offer insights into essential movie attributes, box office performance, and audience reception. The analysis is driven by a central question:

---

*With success in cinema being measured through multiple lenses—**AUDIENCE RECEPTION**, **BOX-OFFICE REVENUE**, and **INDUSTRY RECOGNITION**—what are the key factors influencing these metrics, and how have they evolved over the decades?*

DATASET

The dataset we selected from Kaggle, IMDB Movies From 1960 to 2025, provides annual data on budget, worldwide gross, duration, IMDb rating, and more for the most popular 500–600 movies per year from 1960 to 2025, extracted from IMDb. It includes over 30000 movies spanning more than 60 years, offering valuable insights into long-term trends in the film industry.

To enhance our visual storytelling, we also used the OMDb API to automatically retrieve movie posters.

1. EXPLORATORY DATA ANALYSIS

The exploratory data we conducted, revealed key insights : **success in cinema cannot be captured by a single metric.** Commercial performance, critical reception, and industry recognition each tell different stories.

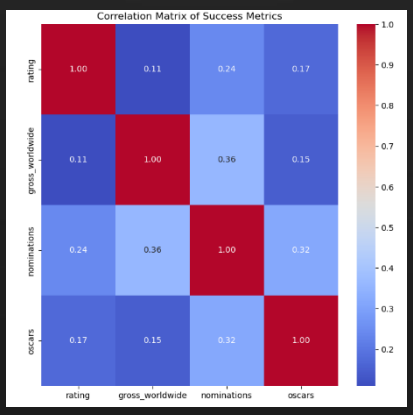

Figure 1- Correlation Matrix of Success Metrics

Indeed, the correlation (*Figure 1*) between IMDb rating and worldwide gross is quite low (0.11), indicating that audience ratings have little influence on a movie's financial success. Similarly, the correlation between ratings and nominations is slightly stronger (0.24) but still weak.

2. ORIGINALITY OF OUR APPROACH

Guided by the EDA, we focused on the interplay between different success dimensions across variables like genre, budget category, and language. Unlike many studies that focus solely on IMDb ratings, our approach integrates several metrics , including worldwide gross, ratings and Oscars, to better capture the multifaceted nature of cinematic success.

The full cleaning process, exploratory analysis, and enriched dataset (including posters) can be found in our Jupyter Notebook.

PROJECT STRUCTURE

1. ARCHITECTURE

To ensure reusability and clarity in development, our project is organized into a structured hierarchy of HTML pages and ressources. Below is a breakdown of the architecture and the purpose of each major element:

…/404NotFound/tree/master

- index.html : Serves as the homepage, introducing the project and inviting users to dive into the visualizations.
- explore.html : First exploratory page featuring a Sankey diagram, allowing users to compare key success metrics (Worldwide Gross, IMDb Rating, Awards) and understand their relationships.
- analysis.html : Offers a bubble chart for deeper analysis of key characteristics (budget, genre, language) that influence movie success across different metrics.
- detail.html : Provides detailed information for each movie

…/ assets :

… /components : Contains reusable UI components used across different pages.

- navbar.html : Defines the top navigation bar, included on all pages for consistent and intuitive navigation throughout the site
- footer.html : Provides a consistent footer layout containing credits
- timeline.html : Contains the interactive timeline filter used to dynamically adjust visualizations based on the selected year range

… / css : Contains stylesheets that define the visual identity of the site
…/ data : Stores the cleaned dataset
…/ img : Includes visual assets
…/ js : JavaScript codes responsible for dynamic content, interactivity and pages logic

2. TOOLS

Our project was developed focusing on interactivity, responsiveness and ease of use. The frontend stack combines standard web technologies with powerful visualization libraries to create an engaging user experience :

- HTML & CSS : to structure the web pages and apply consistent styling across the interface.
- JavaScript : to power all dynamic behavior on the site. Key libraries :
  - D3.js : Sankey diagram, Bubble chart
  - Tom Select.js : dropdown filters with multi-select capability and search functionality

VISUALIZATIONS & STORYTELLING

1. GRAPHIC CHARTER

Our visual identity draws inspiration from IMDb's iconic brand colors, reflecting the world of cinema. The primary color is **yellow (#f5c518)**, a bold and energetic tone that draws attention to key interface elements such as titles, buttons and important insights.

This primary yellow is paired with **a dark base palette**, including **black (#000000)** and **dark gray (#121212)**, to create strong contrast and ensure that content and visualizations stand out clearly. These dark tones provide a cinematic, immersive background reminiscent of a theater experience, while maintaining high readability.

For typography we use **BEBAS NEUE** for titles and headers to covey a bold, modern, and cinematic tone, while in contrast, **Arial** is used for body text and UI components to maintain a professional and neutral appearance.

This colors and typograph scheme ensures consistency, readability, and emotional resonance with our theme, while reinforcing the data storytelling through visually engaging and intuitive design.

2. VISUAL STORYTELLING

Our website is structured as a **narrative-driven experience**, guiding users through a visual exploration of a rich movie dataset. By structuring the content from general context to specific analytical insights, we help users form meaningful connections with the data and understand complex relationships in an accessible way. Each visualization is crafted not only for aesthetic appeal, but also for purposeful interaction and data insight. Here's how the story unfolds:

A. INTRODUCTION - SETTING THE SCENE

The journey begins with a concise overview of the project's purpose **: What makes a movie successful? How success is defined in the world of cinema?**

This section emphasizes that success in film is multifaceted, encompassing not just box office performance but also critical reception and award. By presenting these success dimensions early, we guide the user to consider the many possible interpretations of "success."

To support engagement, we highlight key initial statistics immediately demonstrating the richness of the dataset. This context grounds the experience, catches the cinephile attention and helps the audience know what to look for.

B. TIMELINE FILTER – NAVIGATING THE ERAS

The timeline filter serves as the user's first point of interaction with the dataset. Implemented as a horizontal slider styled like a filmstrip, it not only ties visually to the cinematic theme but also helps users intuitively filter movies by release year or time range.

This feature enables users to isolate films from a particular decade or year (e.g., 1990s blockbusters or post-2010 trends), offering a temporal dimension to the exploration. The ability to compare eras is crucial for understanding how success factors shift over time.

The timeline thus acts as both a thematic and functional introduction to data filtering, giving users agency to explore based on personal curiosity or analytical intent.

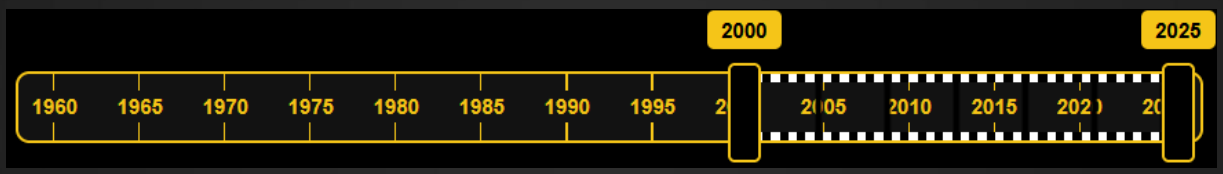

Figure 2 – Timeline filter

C. SANKEY DIAGRAM – UNRAVELING THE PATHWAYS OF SUCESS

The Sankey diagram serves as a narrative centerpiece in our visual storytelling approach. Rather than focusing on any single measure of success, it offers a dynamic overview of how different success metrics interconnect across the movie dataset. The goal is not to determine what makes a film successful, but rather to reveal that success itself is multi-dimensional and sometimes contradictory.

We enhance readability and engagement through:

- Interactive highlights on hover, to help isolate and read individual flows
- Tooltips that precise film counts in each link, helping users grasp both proportions and exact figures.
- Clickable nodes that transforms the diagram into a filtering interface, updating the rest of the interface to reflect only the selected pathway (Note that the categories used in the Sankey nodes are rounded grouping while the statistics displayed on top are based on exact value.)
- Draggable elements, giving users freedom to rearrange nodes for clarity or just playful interaction

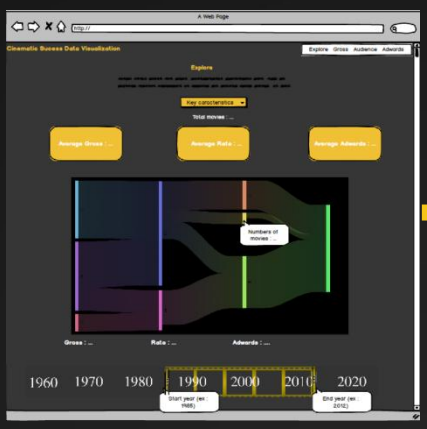

Figure 3 – Sketch of the Sankey diagram

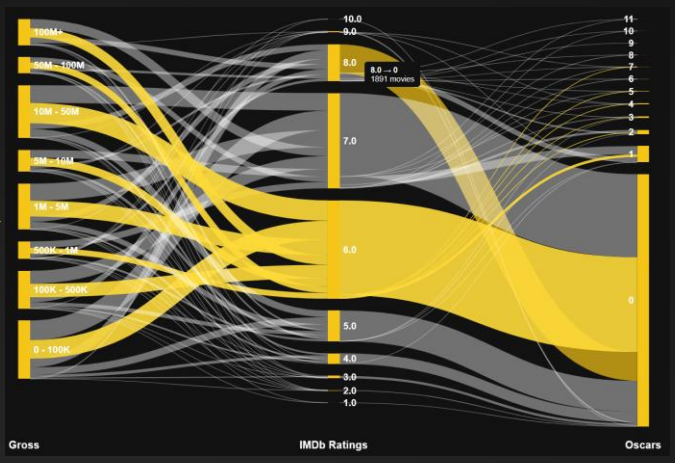

Figure 4 – Sankey diagram of success metrics

The narrative purpose of the Sankey diagram is to provoke exploration by encouraging users to reconsider simplified assumptions. It visually suggests that film might succeed in one area while failing in another

D. BUBBLE CHART – ZOOMING IN ON BLOCKBUSTERS

The bubble chart brings a more focused lens to the analysis by allowing users to compare standout films based on layered attributes. This chart is structured by:

- Three layers, chosen by the user : representing key characteristics such as budget level, language, or main genre;
- A focus metric which determines what "success" means for the chart : box office revenue, critic score, or number of awards.

This hierarchical layered approach empowers users to isolate specific trends such as "low-budget thrillers in English tend to perform well in ratings" and compare apples to apples. As users progress through the layers the bubbles become more granular, ultimately displaying individual movie posters on which the users can click to get detailed information about that movie.

This visualization not only supports analytical discovery but also invites a personal layer of engagement. A user might come across a high-performing film he hadn't seen, understand why it stood out and even be inspired to watch it.

Thus this page is a blend of data exploration and cinematic recommendation.

Figure 6 –Sketch of the Bubble chart

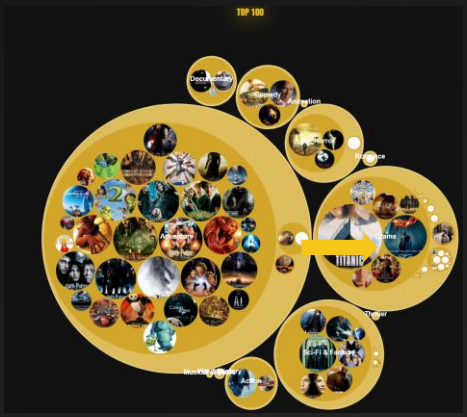

Figure 7 – Bubble chart zoom

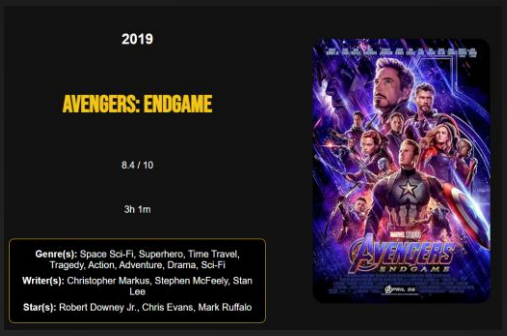

Figure 8 Movie details

CHALLENGES

1. CHALLENGES ENCOUNTERED

One major challenge we encountered was the **limitation of GitHub Pages**, which only supports frontend technologies and does not allow backend operations. As a result, we couldn't dynamically fetch movie posters via the OMDb API when the user changes filters. To address this, we pre-downloaded a large number of posters in advance by selecting the top-rated films and then the top-grossing ones. However, since we couldn't fetch everything, a significant number are still missing and won't appear on the site.

Another challenge involved the **Sankey diagram**: due to its fixed column structure, it is not possible to compare some metrics like gross revenue and Oscars directly, since they don't appear side by side. To address this limitation, we enabled filtering for example by gross categories, allowing users to see related statistics such as the average number of Oscars in the summary panel.

We also faced difficulties stemming from **the large number of movies in the dataset**. Initially, the bubble visualization attempted to display all films individually, which resulted in overcrowded visuals that were unappealing and unreadable. Moreover, variables like budget and gross were raw numerical values, meaning nearly every film had a unique number and making visual comparison nearly impossible. To solve this, we introduced categorical bins (e.g., "50M - 100M") for budget and gross (more details in the solutions described below). However, this introduced a trade-off by slightly reducing precision and complicating the ordering of non-numerical values.

2. CHANGES TO OVERCOME CHALLENGES

To address these challenges, we implemented several strategic adjustments to improve both readability and interpretability of our visualizations:

- **Bubble Chart Redesign :** The bubble visualization was redesigned into a bubble chart (see *Figure 7*) that only displays the top 100 movies based on user-selected filters. This dramatically reduced visual clutter while still capturing the most relevant data points for analysis.

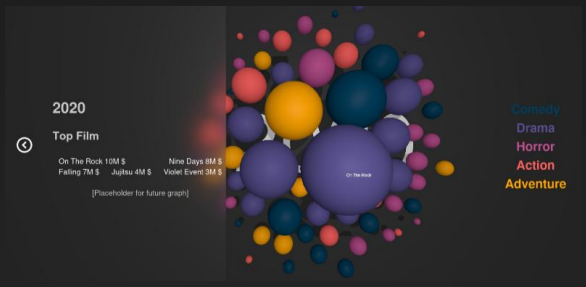

Figure 9 - Initial sketch for the bubble visualization

- **Focused Scope on Key Characteristics :** We streamlined our analysis by focusing on a limited set of influential characteristics (*budget, genre, language*) which were identified during our exploratory data analysis. This allowed us to deliver a clearer narrative and avoid overwhelming the visualization with too many variables.
- **Exclusion of Redundant Success Metrics :** Some secondary success metrics, such as *gross opening weekend* and *number of nominations*, were excluded from the final analysis because they were already largely explained by metrics already included (e.g., *gross worldwide* and *Oscars won*) and offered limited additional insight.
- **Non-Uniform Binning for Budget and Gross :** Continuous variables like budget and gross were converted into categorical bins. We chose non-uniform intervals to better reflect the real-world significance of differences. For example, the gap between a $5M and $10M film represents a major jump in production capabilities, while the difference between $500M and $505M is negligible in the context of high-budget blockbusters. This approach ensured more meaningful and interpretable groupings.

These changes helped us turn a complex and dense dataset into a MORE ENGAGING and INSIGHTFUL EXPERIENCE, while preserving analytical depth and aligning with the overall storytelling goals of the project.

TEAM CONTRIBUTIONS

Benjamin BURKI :

- Participated to the implementation of analysis page
- Contributed to the final screencast
- Integrated additional statistics to enhance the storytelling

Antoine CORNAZ :

- Enriched the dataset by web scraping movie posters
- Designed and implemented the detail view page

- Participated to the implementation of analysis page
- Contributed to the final screencast
- Was in charge of GitHub

Charline HUANGFU :

- Led the exploratory data analysis to identify patterns, cleaned and preprocessed the data
- Designed and built the overall site architecture including pages, navigation, footer and a consistent graphic charter
- Implemented the Sankey diagram and started the bubble chart integration
- Developed and designed the interactive filtering components, especially the timeline filter
- Wrote and designed the process book

CONCLUSION & FUTUR WORK

This project set out to explore what makes a film “successful” through an **ENGAGING** and **NARRATIVE-DRIVEN DATA VISUALIZATION EXPERIENCE**. By leveraging a curated movie dataset and focusing on interactive storytelling, we developed a web interface that guides users from high-level trends to detailed film-level insights. Each visualization was carefully chosen and designed to reveal different dimensions of success, whether commercial, critical, or artistic.

Throughout the process, we prioritized usability, visual clarity, and analytical depth. The end result **challenges assumptions** and **empowers users to form their own conclusions** by interacting directly with the metrics.

While the current version provides a solid foundation, this are some ideas to expand and improve the project:

- Add additional features like cast, director, or studio to enrich the context behind each film’s success
- Incorporate animated or comparative views across time to show how success metrics evolve over decades
- Personalized Recommendations **:** Allow users to save their favorite films in a “My Blockbusters” list. By analyzing the common traits of these selections, we could identify each user’s unique definition of success and suggest similar films they’re likely to enjoy
- Mobile Optimization : Adapt the interface to smaller screens for better accessibility and usability on mobile devices