Landon, Maddie, Ramona

Movie Report

I. Abstract

This project examines Movie Datasets to investigate how insights derived from data can enhance a movie-streaming platform. We focus on three primary areas: recommendation, trend analysis, and budget-revenue correlations. Initially, we create a straightforward content-based recommendation system using features such as genres, directors, and popularity for a visualization that uses bar charts that focus on the shared characteristics of the movie. On the other hand, we analyze long-term trends in cinema by studying how the popularity of genres, average ratings, and audience preferences have evolved over decades. Lastly, we looked at how movie budgets and revenues correlate with the rise in production of certain genres, showing us patterns such as increased investment and success in certain genres. Altogether, these facts show how data can improve user engagement and decision-making for a streaming service, including personalized recommendations to help choose which movies to add and find features that lead to good ratings, and determine what genre is popular and good in pricing while making money from it.

II. Introduction

Streaming services depend significantly on data to maintain user engagement and to determine which films to suggest, promote, or include in their library. Given the huge number of movies released over various decades, genres, and countries, it becomes challenging to find out what the audience preferences are. This idea focuses on movie data to investigate three primary

issues: how to recommend similar films, how cinematic trends have evolved, and how movie budgets and revenues correlate with the rise of specific genres. By examining patterns in genres, directors, decades, popularity, and descriptive keywords, we seek to demonstrate how data can enhance recommendations and assist platforms in comprehending which types of films people like. We also want to see what the most profitable genres are to invest in, providing insights that could help investors maximize returns on production spending. Over the last ten years, research in movie analytics and recommendation systems has really taken off, particularly with datasets like MovieLens. This dataset is often used to analyze user preferences, rating habits, and recommendation methods. Previous studies have indicated that factors like genre combinations, release years, and user viewing history are crucial for enhancing personalized recommendations (Harper and Konstan, 2015). Our project takes these ideas further by using larger, more similar datasets and expanding the analysis to cover genre trends and how they relate to budget, revenue, and genre production.

III.   Datasets

For this project, we gathered various movie-related datasets from Kaggle. These datasets offer a lot of movie information that helps us achieve our three main objectives: creating a recommendation system, analyzing long-term trends, and analyzing how movie attributes relate to budgets and revenues. The combined data features movie attributes like titles, release dates, genres, and languages, along with extra attributes such as popularity, keywords, cast, crew, and directors. Some datasets even provide financial data like budgets and revenues, plus overviews, and production specifics. All these attributes help us with the necessary information to compare

films, spot similarities, monitor how film trends evolve, and determine which traits are linked to ratings.

IV. Visualization Design

We created a scatter plot visualization to look at the correlation of budget and revenue of what we found to be the most popular, horror films, which to other genres. This helps us understand how production and financial success have evolved and points out specific patterns related to popular genres. To show the changes in genre popularity, we made a line chart that shows the number of films produced in each genre by decade. This visualization effectively focuses on which genres are on the rise or decline and aids in showing what long-term changes in audience preferences. Lastly, for our recommendation system, we display bar charts that feature the top recommended films for a specific title, with similarity scores based on genre, ratings, and popularity. This layout is simple and enables users to quickly identify which movies are most similar to the one they chose. In summary, these visualizations are helpful because they are easy to understand, directly related to our research questions, and offer insights into both audience behavior and movie traits, telling us the recommendations and decision-making for a streaming service while helping many people in the movie industry make money.

V. Data Processing/Implementation Details

We used multiple datasets that have movie ratings, pricing, and other types of data, which required a lot of preprocessing steps to make it easier for us to analyze and visualize. We cleaned the datasets by handling missing values, removing duplicate values, and combining columns. We also performed data transformation tasks, such as changing attributes that benefit us more for our project. However, during the processing, we encountered some issues, such as genre labels being

difficult across visualizations and datasets, and incomplete information, which were later resolved through excluding data that really didn't help our goals, which led us to finally do some data merging. Once everything was processed, the data was ready for implementation in our systems. We used the processed data to make clear visualizations showing popular genres, ratings, and trends.

VI.   Results/Findings

We ran several tests trying to find patterns in our visualizations and code. Some of us ran into problems with working with a lot of genres, languages in movies, and trying to efficiently get clean results. We were able to recognize patterns in our movie recommendations when we analyzed ratings, popular genres, and trends across ratings to pricing in movies. We later found that horror movies experienced a huge spike in production (and popularity) in the 80s and 2000s, and when researching more closely, found that many societal factors contributed to this. An article by the University of Puget Sound suggests that the rise in horror movie production is attributed to the cultural and economic uncertainty that loomed over the US in the 1980s following the Cold War (Sheehy, 2024). Other articles suggest that the rise in horror movies reflects growing anxieties about disruptions to social norms and conformity, along with the development of SFX and VFX that continuously make the film industry better (Emanuely, 2025). For example, we noticed in our visualization that action, horror, and thriller movies have seen significant growth in both popularity and box-office performance, likely in part due to enhanced visuals that were made possible due to modern technology. From our scatterplot of combined datasets on the basis of genre and success, we found that horror had a correlation between budget

and revenue compared to other genres, where they are relatively cheap to produce but consistently yield high returns.

VII.   Conclusions

Altogether in this project, we think we all did fairly well. We successfully looked at the data and made our inferences and conclusions, and found patterns that we were happy with and made sense. We found out that horror was the most popular genre, followed by thriller and action, and found out that these genres make most of the money and are the most popular at the moment. We developed a recommendation system for our datasets that took in multiple values and produced a successful result of what patterns in movies would most likely influence why people are watching certain movies and genres. Our visualizations helped us find new insights that we were not expecting through our experiments and testing. From all that work, we did encounter some of the analysis to be difficult due to the problem of limited time and dataset complexity. We feel like if we had more time, and less complexity with the data, we could have produced even more effective visualizations or even better graphs. The result would've been a much easier insight into our visuals and results. Going in future directions for this project would be better visualizations with more information, and updating the dataset yearly or every half-year for more up-to-date results. Overall, we came out of this with new knowledge and understanding of visualizations. This project really helped us strengthen our skills and introduce information about the film industry, which was new to us, and hopefully, helps future people looking for a movie and are curious like us about its history and what factors influence it.

VIII.   References

The movielens datasets: History and context: ACM Transactions on Interactive Intelligent Systems: Vol 5, no 4. (n.d.-l). https://dl.acm.org/doi/10.1145/2827872

Emanuely. (2025, March 22). How film genres have changed over the decades - onebeone.

Onebeone -. https://onebeon.com/film-genres/how-film-genres-have-changed-over-the-decades/

Sheehy, S. (2024, December 4). *Be very afraid: How horror movie trends reflect societal*

fears. The Trail. https://trail.pugetsound.edu/?p=18998

Landon:https://www.kaggle.com/datasets/abdallahwagih/movies

Madison: https://www.kaggle.com/datasets/kutayahin/letterboxd-movies-dataset,

https://www.kaggle.com/datasets/sujaykapadnis/horror-movies-dataset