

TOP Movie Recommend



Here is where your
journal begins



Jiayi Sun
Jiantian Lei
Zhentao Liu





Introduction



Since the 1960s and 1970s, cinema has evolved into one of the most popular entertainment genres worldwide, becoming an integral part of business and daily life. The flourishing success of Hollywood films during this time laid the foundation for its global reach. In other words, **movies are becoming an important part of people's entertainment life**. With the exponential growth of movie websites and the rise of film reviewers providing in-depth analyses, **choosing which movie to watch** has become increasingly challenging.

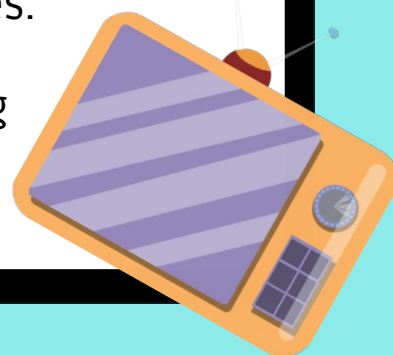
Our goal

Interesting Visualization: Several kinds of interesting visualization are given to offer target readers good and direct reading experiences.

Recommend Quality Films: Movie lovers can easily locate a good movie through a website made by pure love rather than commercial websites with spam and advertise.

Films Analysis Tools: Filtering and useful interactive analysis charts functions are to be implemented and offer the targets good tools to select and understand movies.

In the end, we don't want to make a boring recommendation website. Instead we want to visualize the recommended process.



Website Design



Dataset: We hope to recommend the quality movies and thus **IMDb TOP250 movies** dataset can provide the most welcomed movies. However, the dataset only contains basic info like years, rating, and so on. It lacks movie_id and more additional information. To make movies information more completed, this dataset and **IMDb Movies in different categories** are merged and duly treated. It not only contains the necessary information for later visualization but also benefits the later potential web scraper applications.



Braining storm: Filter should be based on certain criteria.

1) There are lots of **genres** of movies, and people like their specific movie hobbies. Thus we want to develop one selection system on the type. 2) **Country** is also one important information to filter the films. For instance, some films fans have extreme passion in the film produced by british directors. 3) Some **bonus analysis tools** would be the icing on the cake.

Chosen Visualization:

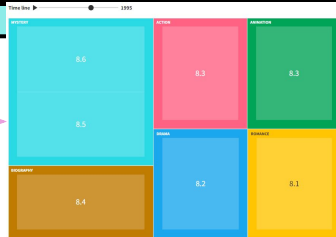
Interactive Graph: Relationship graph of movies and interactive filtering for movie recommendation.

Interactive Map: Word Map to Link geography and movies information.

bonus analysis tools: Some analysis tools like bubble chart is useful for analysis on rating or box office.

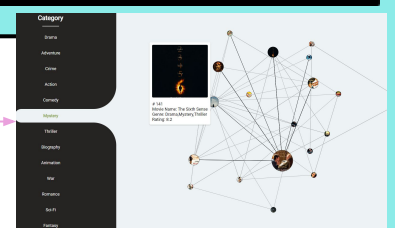
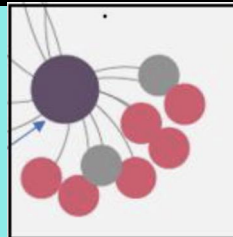
What we change? From Sketch to Final

Mystery
Sci Fi History



Instead of simply making an interactive word cloud, we decide to make a hierarchical chart that you can see more informations like ratings or published time. Users can easily check the information about movies they interest in.

WordCloud



We make the relationship network able to move around, users can click on the movie they are interested in and the neuron will zoomed in with the pop up of details of that movie.

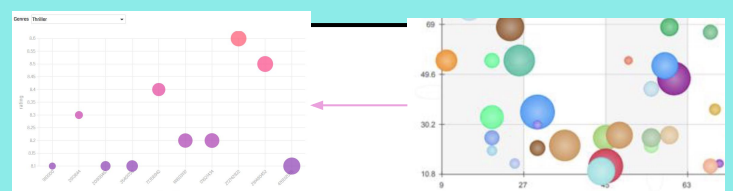
Relationship Graph



Click on U.S.A to see the list of American actors and directors.

Initially our idea was to visualize the nationality of actors and directors by using a flat world map. During the implementation, we found it will be interesting to make a 3D globe that you can rotate the map and click on the country to see further info.

Interactive Map



We basically use the similar design from the milestone II. However, the final chart looks very messy, so we add a filter of genre of movies. The users can easily to see relationships between rating and box-office for different types of movie.

Bubble chart

Challenges and Solutions



Challenge 1: Data Collection and Data Integration

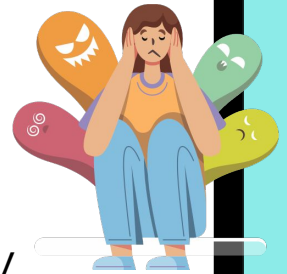
We aim to explore the movie features in depth, which single data source does not have all feature columns.

—> We combine several existing dataset and web scraper techniques, and integrate movie data from multiple sources

Challenge 2: Interactive Visualization

Tabular Data visualization is static, while movie recommendation requires user interaction

—> In our interactive Movie Relationship Graph and Actor/Director Maps, users are more easy to find movies/actor that they are interested in.



Challenge 3: Understand the movie data in depth

Movie Recommendation is not enough, we need further exploration.

—> Exploratory Data Analysis, Bubble Chart, Word cloud, etc.

Challenge 4: Full stack Development

This projects requires data visualization, as well as other data science tech-stack from data collection to data processing. Different techniques including ML, web scraper, EDA are applied.



Implementation Path



Data Collection: To overcome the drawbacks of existing database, we use web scraper to collect more features of movies and people.

Data Integration: We build data processing pipeline to integrate multiple datasource

Data processing: We use ML methods to generate advanced features, e.g: related movies to build relationship graph


Data visualization:

We use React Framework to interactively visualize movies/ actors / director data, including:

Interactive Graph: In the movie relationship graph: Users can filter and search movies, and response with relationship graph to recommend users with similar movies.

Interactive Map: We use 3D interactive Map to visualize Stars / Directors Information. Location Information of Stars/ Directors are shown on the map.

Further Exploration: To further explore movie statistics, We use Bubble Chart to visualize movies sales, and we use Time Chart to show movies in different time periods.



How to use our website



Our main functions

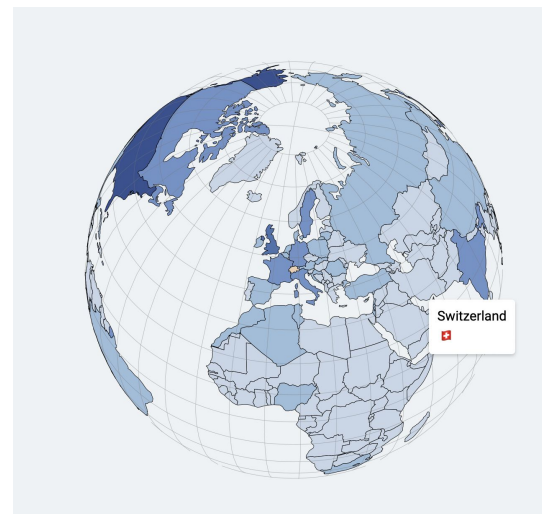
We got three main functions for our website: "Map", "Graph" and "Exploration".

Map Graph Exploration Info



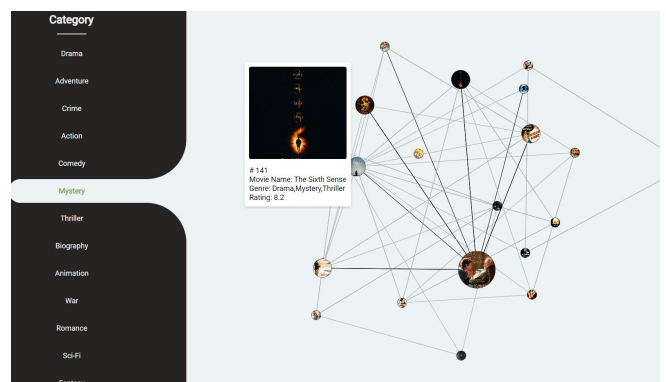
Part I: Map

Our interactive map is designed to showcase the distribution of directors and actors across the globe within the prestigious IMDb top 250 list. Vibrant colors emerge as heat map, indicating the intensity of director and actor presence within each region. Clicking on a specific region unveils the country's name and emoji so that you could know which region has the most famous directors and stars.



Part II: Graph

The central part of our website is the movie recommendation using relationship graph. Users can filter by different genre and search movies they might be interested in, and our module will response with different relationship graphs of related movies. When users click on a specific movie, movie details will be shown and related movies will be highlighted.

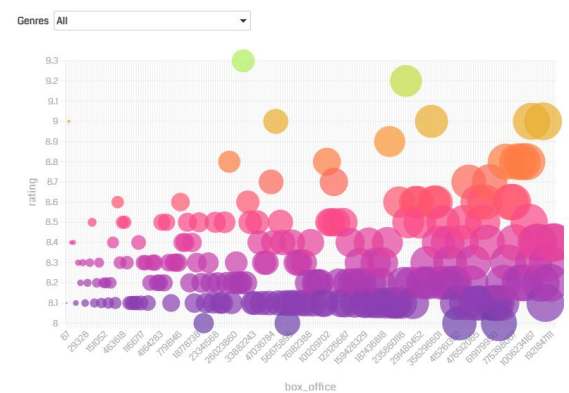


How to use our website

Part III: Exploration

For some film lovers, they are more interested in choosing a film after some research than simply choosing a film based on its actors or title. This section will offer a fun movie discovery feature where users can discover the relationship between the rating and the box office of the IMDB top 250 movies, the user can use the genre filter to explore the specific type of movie they are interested in.

and alternative approach is to use the treemap chart, Click the genre you like, and you will see the movies with different ratings. Click the pop-up window to see the details of the movie you interested in, there is also a timeline filter for you to see the result for each year.



Peer assessment



- Milestone 1 & 3 EDA
- Interactive World Map
- Web Scraper for Directors & Stars data, process the dirty data for use
- Process book and Website
- Deployment to github page

Jiayi Sun

What I learn:

I learned React and D3.js for an interactive frontend, also web scraper and some data processing for display the raw data. It is such a brilliant chance to implement various visualization modules.

Jiantian Lei

- Milestone 1 & 2 text
- "exploration" viz including treemap and bubble chart
- Process book and Website

What I learn:

It is a great opportunity to practice how to tell a story about given data by visualization, and the experience of creating a website with front-end development.



- Interactive Movie Relationship Graph
- Search and Filtering Module for movie recommendation
- Web Scraper for Movie & People Data
- ML & Data processing for related movies
- Milestone 1 EDA and Documentation

Zhentao Liu

What I learns: Apply Frontend (React and D3.js) and Backend(web scraper, ML, etc.) to implement interactive modules. A great opportunity to experience from data collection, data analysis to data visualization.