

---

# COMPSCI 571

## Project Proposal

---

**Soniya Gaikwad**  
32927398  
sgaikwad@umass.edu

**Manan Parikh**  
34968148  
mananrajeshb@umass.edu

**Shiva Sravani Mudiyanur**  
34965596  
shivasravani@umass.edu

**Project Title: Exploring Movie Trends Over Time (Montage)**

**GitHub Repository: <https://github.com/soniyagaikwad/movie-trends-data-vis>**

### 1 Background & Motivation

Our team decided to pursue “Exploring Movie Trends Over Time (Montage)” because of our love and interest in entertainment, especially movies. Also, some of us have projects revolving around media, so we wanted to continue looking into this industry and its crossover with data visualization.

### 2 Project Objectives

The primary questions we are aiming to answer with our visualization include:

- While looking at a specific (movie/genre/MPAA rating/budget), how much revenue was generated?
- What are the relationships between the MPAA ratings, budget, gross revenue, release dates, genres, runtimes, and ratings the movies have received?
- What are the highest and lowest revenue-generating movies?
- What are the highest and lowest-budget movies?
- What are the ratings for the movies depending on their genres?
- Which genre has the most interest over the years?

We would like to learn how these pieces of data related to movies correlate with one another and theorize how they potentially play out in the entertainment industry. Some benefits of learning and accomplishing these aspects include helping us and others understand what factors could be critical to the movie industry for potential box office success and how the public interacts with these movies based on these factors.

### 3 Data

We are using a dataset from Kaggle by Yashwanth Sharaff called “Movies Performance and Feature Statistics: Analyzing Box Office Performance, Rating and Audience Reactions” [1] to visualize our project.

<https://www.kaggle.com/datasets/thedevastator/movies-performance-and-feature-statistics>

## 4 Data Processing

In the dataset “Movies Performance and Feature Statistics: Analyzing Box Office Performance, Rating and Audience Reactions,” we will gain access to a large amount of data related to movies, such as their titles, MPAA ratings, budget, gross revenue, release dates, genres, runtimes, ratings they have received, and summaries. Based on our initial designs, we plan to focus on each movie’s title, MPAA ratings, budget, gross revenue, genres, and summaries. However, we may potentially look into adding the other additional pieces of data if we can find a way to fit it with our visions. Otherwise, we will clean up the data accordingly, but not significantly.

We perform data cleaning and processing by removing duplicate and null values. Additionally, we analyze the dataset to identify missing values and determine whether they can be filled based on relationships and patterns. If no suitable replacement is found, we remove those rows to prevent them from affecting the outcome.

Moreover, there is a column named 'release date' in our dataset, which we don't intend to use in its entirety. Based on our brainstorming, we will only extract the release year from it.

## 5 Visualization Design: Five Design Sheet Methodology

Sheet 1, Brainstorm:

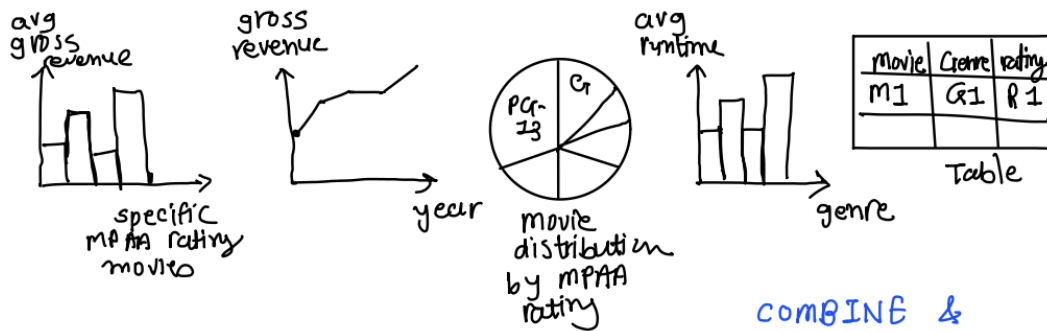
### Brainstorm

Movie trends over time (Montage)

dataset

movie\_id, title, MPAA rating, budget, gross, release, genre,  
runtime, rating

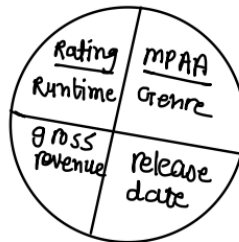
most of the columns are useful for generating visualizations



### FILTER

- search for a movie name, and start getting data as you start typing
- get the details of popular genres among viewers over time
- generate a graph
- ~~for all MPAA ratings vs avg gross revenues~~
- ↓
- better to generate
- a boxplot using which we can see outliers too

### CATEGORIZE



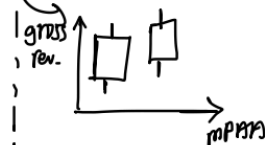
most of the visualizations can be generated using these four categories  
for example,  
genre → gross revenue  
genre → rating

### COMBINE & REFINER

Table

movie	genre	rating
M1	G1	R1

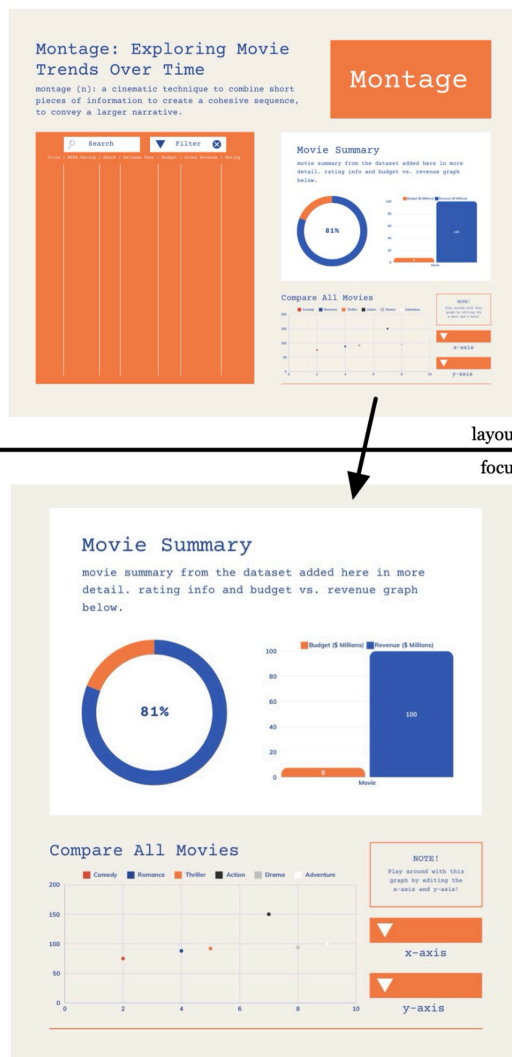
genre rating/popularity  
Action 63%  
Sci-Fi 35%



### QUESTION

- designs are effective and useful - they don't mislead
- pros: interactive & dynamic designs
- cons: boxplot could be difficult
- We have more than 3 types of distinct designs

## Sheet 2, Initial Design by Soniya:



title: movie trends over time - montage

group: Soniya Gaikwad, Manan Parikh, Shiva Sravani Mudiyanur

date: 2/28/2025

task: represent information about movies effectively

sheet: 2, initial design

### operations

- the user can use the search engine to find a specific movie title, which they can click on to get more information and data about the movie, such as the MPAA rating, genre, release year, budget, gross revenue, rating, and summary.
- the user can use the filter button to filter the dataset in the table by their MPAA rating, genre, release year, budget range, revenue range, and rating range.
- after clicking on a movie, the right side, which shows the movie data summary, will give the movie's summary, rating, and the comparison between the movie's budget and revenue.
- on the bottom right of the website, the user can compare all the movies that are provided in the dataset. however, with this graph, the user can interact with it by changing the x-axis and y-axis labels, such as the ones provided in the table.

### discussion

#### advantages

- get individual information and data about a movie, especially through a simple visual representation of the data
- have the ability to compare all movies in a graphical representation
- more user freedom by allowing the user to change the x-axis and y-axis labels of the graph related to all movies
- allows the user to interact with all the movies by allowing searches and filters
- the color palette most likely shouldn't cause issues for any groups of people

#### disadvantages

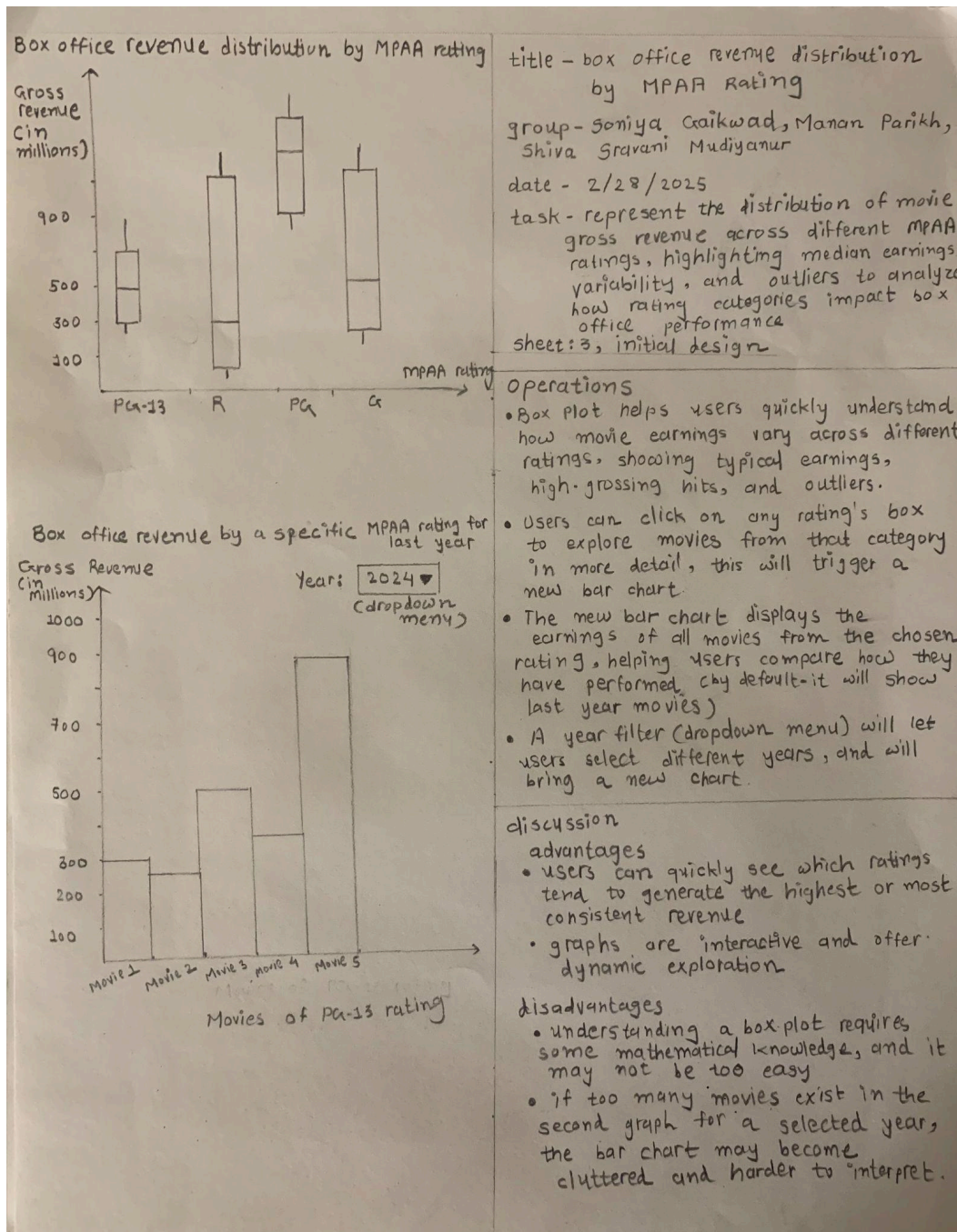
- the table format can get involved and overwhelming at first glance, though it allows the user to get more of an idea of the dataset

the website seems to be suitable for our tasks, especially in terms of how we want to present information and trends about movies.

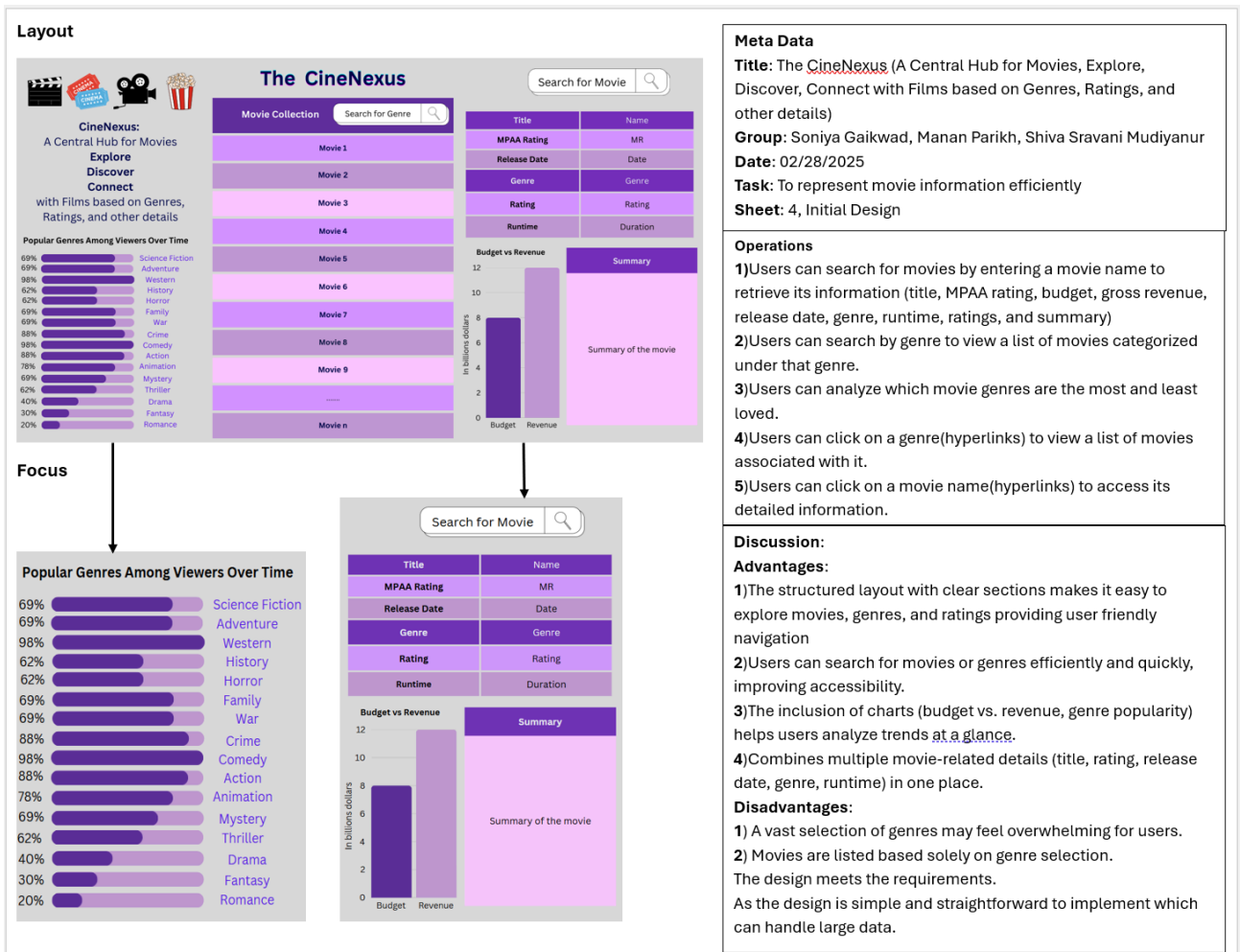
the website should be feasible to implement and scalable with more data.

the design also seems clear and easy to interact with.

Sheet 3, Initial Design by Manan:

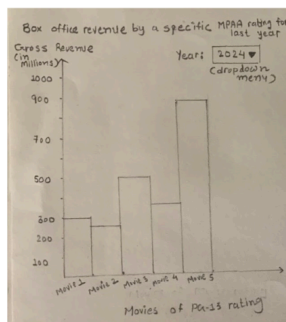
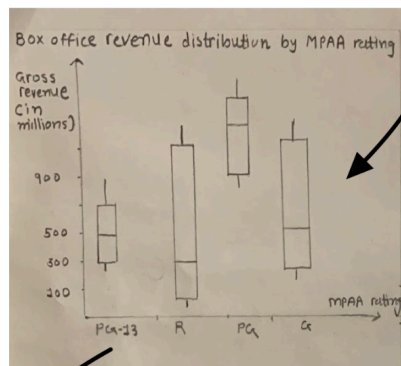
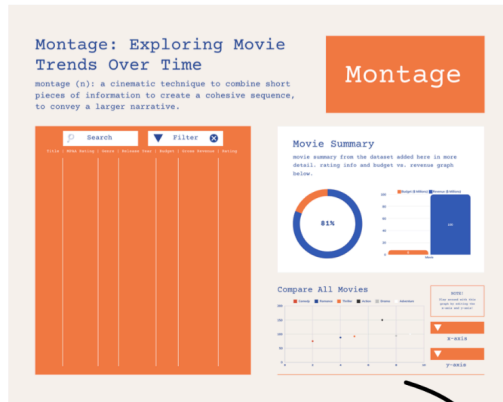


Sheet 4, Initial Design by Sravani:





## Sheet 5, Realization:



title: movie trends over time - montage

group: Soniya Gaikwad, Manan Parikh, Shiva Sravani Mudiyanur

date: 2/28/2025

task: final overall design of the data visualization components

sheet: 5, realization

### operations

- the user can use the search engine to find a specific movie title, which they can click on to get more information and data about the movie, such as the MPAA rating, genre, release year, budget, gross revenue, rating, and summary.
- the user can use the filter button to filter the dataset in the table by their MPAA rating, genre, release year, budget range, revenue range, and rating range.
- after clicking on a movie, the right side, which shows the movie data summary, will give the movie's summary, rating, and the comparison between the movie's budget and revenue.
- on the bottom right of the website, the user can compare all the movies that are provided in the dataset. however, with this graph, the user can interact with it by changing the x-axis and y-axis labels, such as the ones provided in the table, which will be the main focus with many features.
- for example, the user can explore and visualize the box office revenue distribution by MPAA rating using the x-axis and y-axis label options.
- this part of the graph would work dynamically and change according to the different types of data plots that make sense with the provided axes.
- for example, as we did with the box office revenue distribution by MPAA rating, we could use box plots to help the user quickly understand how movie earnings vary across different ratings, showing typical earnings, high-grossing hits, and outliers.
- the user can click on any rating box to explore movies from that category in more detail, triggering a new bar chart.
- the new bar chart will display the earnings of all the movies from the chosen MPAA rating, helping users compare how they have performed.
- with this bar chart, there is a drop-down menu to represent a year's filter and create the chart accordingly.

### detail

- we estimate that it could possibly take at least 6 weeks to implement the website end-to-end.
- we will be using a dataset from Kaggle, which our data structures and software will heavily rely on.
- we are using an orange and blue color palette for the UI to ensure that all people with different eyesight will be able to use the website easily and to create more contrast. we also use a script font to stay on theme with movies.
- the focus of the data will likely become more prominent in the final result to emphasize its features.
- the dataset may need some cleaning to make sure we are providing the best visualizations and information to the users.

## 6 Must-Have Features

The features we find essential to our project include the following:

- User-Friendly Interface – A clean, intuitive design with easy navigation.
- Movie Search Functionality – Users can search for movies by name to retrieve relevant details.

3. Movie Details Display – Essential information such as title, MPAA rating, budget, revenue, release date, genre, runtime, number of ratings, and summary should be presented.
4. Sorting & Filtering Options – Allow users to sort movies by popularity, revenue, release date, or rating.
5. Clickable Links – Users can click on the movie name, genre or released year to get the information accordingly
6. Trends over Time - Graphs or visualizations showing revenue trends over time.

## 7 Optional Features

The features which we thought would be nice to have, but not critical to our project include the following:

1. User Ratings & Reviews – Users can rate and review movies.
2. Social Sharing – Allow users to share movie details on social media.
3. Streaming Availability – Indicate where the movie is available for streaming (Netflix, Disney+, etc.).
4. Movie Recommendations – Suggest similar movies based on selected titles or genres.

## 8 Project Schedule

We plan to work together and delegate responsibilities based on our interests, experiences, and strengths each step of the way.

Tasks	Deadline
Data Processing + *Project Review	*3/7
Create the UI + *Project Review	*3/14
Work on the Backend Functionality	3/28
Visualization Prototype for 2 Datasets + *Project Milestone	*4/4
Fully Interactive Website w/ Few Bugs	4/11
*Peer Feedback	*4/16 + 4/18
Final Touches	4/25
*Project Screencast	*5/2
*Final Project Submission + Group Feedback	5/9 + *5/12

\* Course Deadlines, Subject to Change

## References

[1] TheDevastator. (2023). Movies performance and feature statistics: Analyzing Box Office Performance, Rating and Audience Reactions. Kaggle.  
<https://www.kaggle.com/datasets/thedevastator/movies-performance-and-feature-statistics>