DATA ILLUMINATORS

Data Analytics Bootcamp 2023

UNIVERSITY OF TORONTO SCHOOL OF CONTINUING STUDIES

Data Illuminators Project 3

Daniel Marquez
Diana Cao
Kevin Mosweu
Mert Oztop

CineMetrics: Discovering the Best Movies through Data Analytics and Visualizations

Introduction

In recent years, the field of data analytics and visualization has emerged as a powerful tool for gaining insights into various aspects of the world around us. With the vast amounts of data available today, it has become increasingly important to have effective ways of processing, analyzing, and presenting that data in a way that is both informative and visually compelling.

> *The primary objective of this visualization is to provide good movie recommendations based on thorough analysis of movie data. The user of this visualization can be anyone who is interested in finding good movies based on reliable data and analysis. We believe that this project has the potential to provide a valuable service to movie enthusiasts and casual moviegoers alike.*

For our project, we have chosen to focus on the topic of movies and their reception among audiences and critics. Movies are a fascinating subject for data analytics and visualization, as they provide a rich and diverse set of data that can be analyzed in a number of ways. From box office numbers and critical reviews to audience reactions and social media buzz, there is a wealth of information available for those interested in studying the world of cinema.

To ensure that our data is as accurate and unbiased as possible, we have selected three sources of information: Metacritic, IMDB, and TMDB. These sources are well-known in the movie industry for providing comprehensive and reliable data on movies, including reviews, ratings, and other key metrics.

By using these sources, we can be confident that our analysis will be based on data that is both trustworthy and comprehensive.

By using the latest data analytics and visualization techniques, we hope to create a compelling and informative narrative that will engage and inform audiences.

Requirements and steps taken:

Blue font for Rubric requirements:

Data and Data Delivery

1. Collect Data from at least two sources of information.
    - We collected data from three sources coming from two different APIs: OMDB[1] and TMDB[2]

2. The dataset contains at least 100 unique records.
    - We wanted to have complete data from all three sources.
    - We started by collecting movies sorted by the highest TMDB rating. After analyzing the data, we realized that many of those ratings were skewed because there needed to be more votes. Meaning a movie with 9.5 Starts and three votes cannot be fairly comparable with a movie with 9.3 starts and 250K votes.
    - A large percentage of the Top-Rated Movies were obscure and had very little information besides title, rating, and number of votes (e. g. no data for Box office, language, etc.)
    - Based on the time given for this project, it was not time efficient to analyze and get information from more than two APIs/sources of information. We decided to pivot the focus and recollect the data now based on the count of votes (number of unique people that rated the movie), selecting close to 800 movies with votes higher than five thousand.
3. A database is used to house the data:
    - The API's JSON's calls responses were saved in Mongo.
    - For easy manipulation, we migrate the information to Postgres.
4. The project is powered by a Python Flask API and includes professional-level use of HTML/CSS, JavaScript, and the chosen database.

[1] https://www.omdbapi.com/
[2] https://www.themoviedb.org/

- We made our API using Flask in Python. It provides the data for the Webpage Visualizations.
- There is the use of HTML and CSS as frameworks for the Webpage.
- JavaScript was to build the visualizations.

Back End

5. JavaScript library previously unintroduced during class is included and functioning correctly.
    - We used "Taucharts3", which was NOT previously introduced in class.

6. The page created to showcase data visualizations runs without error.
    - Visualizations are presented on the page without errors.

7. A dashboard page was updated using the same data.
    - Indeed, the visualization page uses the data from the Postgres Database that was introduced from the two APIs mentioned in point 1.

Visualizations

8. A minimum of three unique views to present the data.
    - In a sense, we have 20 different views because every time the user selects a different genre, the visualizations are updated.
    - Information will be provided:
        - View 1: Genre Vs Rating
        - View 2: Genre Vs Runtime
        - View 3: IMDB vote counts Vs Box-office**. (Tauchart – JS not in class)**
        - View 4: Listed movies in a table for the selected Genre.

9. Multiple user-driven interactions are included on the final page (such as dropdowns, filters, or a zoom feature):
    - Users can use a drop-down to select the genre.
    - Users can zoom in/out on the bar chart Genre Vs Rating.

10. The final page displays visualizations in a clear, professional-level manner. The data story is easy to interpret for users of all levels.

[3] https://taucharts.com/

- Visualizations were using techniques used in the market and learned in class. Professionally presented and easy to understand and interpret.
- Data charts are very intuitive and straightforward, meaning everyone would like to see this kind of comparison.

Group Presentation

(To be graded on presentation day)

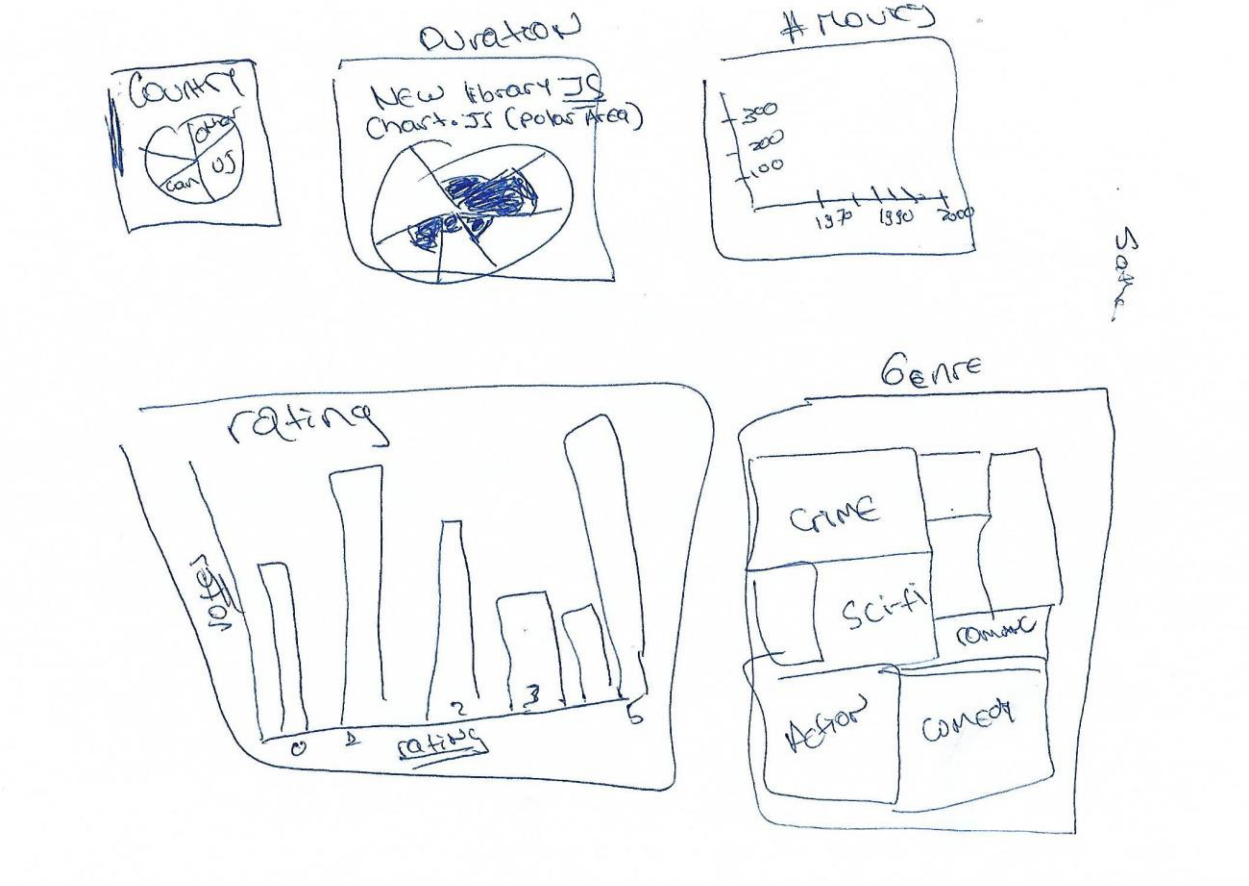
11. All group members spoke during the presentation
12. Group was well prepared
13. Presentation was relevant to material
14. Presentation maintains audience interest

Slide Deck

Click Here to see the Slide Deck

15. Slides are visually clean and professional
16. Slides are relevant to material
17. Slides effectively demonstrate the project
18. Slides are clear and maintain audience interest

Napkin Drawings


Country
Other
US
Can
Duration
New library JS
Chart.JS (polar Area)
Movies
300
200
100
1970
1990
2000
Date
rating
Votes
0
1
rating
2
3
5
Genre
Crime
Sci-fi
Comic
Action
Comedy

Napkin Drawings to requirements