Data Management CSCI 320
Movies Domain - Team 22

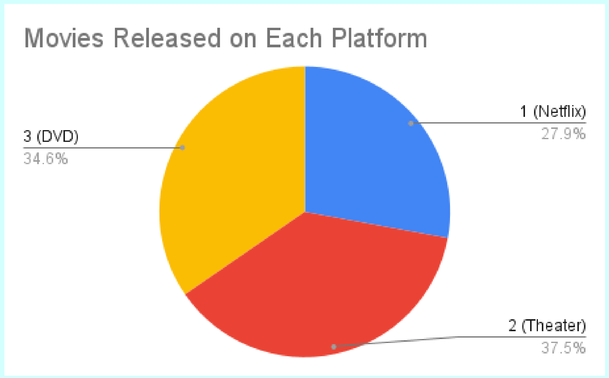
By: Danish Naveed, Graham Zorich, Saadhan Pittala, Jason Yeung, Michael Ambrose

Introduction

Our project introduces a dynamic database system tailored for the modern movie enthusiast. It orchestrates user data, film details, and viewing habits to deliver a personalized movie recommendation experience. This poster outlines our application's design, showcases the analytical approach behind our recommendation engine, and highlights the novel interaction possibilities it offers to users. Below is a rough sketch on the functionalities our API provides to users.

Commands:
- Collections ✅
  - Create Collection ✅
    - {name} ✅
    - List Collections ✅
    - Select {Collection} ✅
      - Play All ✅
        - List Movies ✅
        - Change Name {name} ✅
        - Add Movie {title} ✅
        - Select Movie {title} ✅
          - Play ✅
            - Rate {1-5} ✅
            - Remove Movie ✅
    - Delete {Collection} ✅
- Search Movie
  - By Movie Name ✅
  - By Release Date ✅
  - By Cast Members ✅
  - By Studio ✅
  - By Genre ✅
  - Sort By(Movie Name, Studio, Genre, Released Year) (Ascending, Descending
  - Select Movie {title} ✅
    - Play ✅
      - Rate {1-5} ✅
      - Add to Collection {name} ✅
- Movie Recommendations ✅
- Search User ✅
  - {email} ✅
  - Follow ✅
    - Unfollow ✅
- View Profile ✅
  - Sort Top Ten movies by {Rating, Most Viewed, or Both} ✅

Figure 1

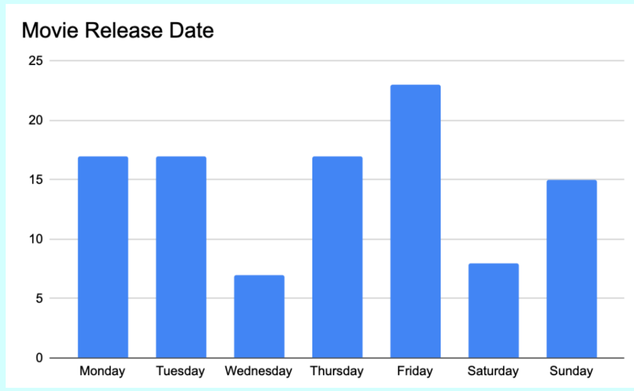


Analysis

The dataset and corresponding pie chart provide information about the number of movies released across three different platforms: Netflix, Theaters, and DVD.

1. **Platform Distribution**: The pie chart illustrates the distribution of movie releases across the three platforms in percentages:
   - Theatrical releases constitute the largest share with 37.5%.
   - DVD releases account for 34.6%.
   - Netflix releases make up 27.9%.
2. **Numerical Data**: According to the dataset:
   - There are a total of 104 movies in the current database
   - 36 movies is released on DVD.
   - 39 movies is released in theaters.
   - 29 movies is released on Netflix.
3. **Majority Platform**: Theaters are the dominant platform for movie releases within this dataset, DVD coming in close second by around 2.9% difference.
4. **Comparative Analysis**: When comparing release platforms, theaters has the most movie releases of 39 movies while DVD comes in a close second of being a 3 movie difference.

Figure 2



Analysis

The bar graph titled "Movie Release Date" represents the distribution of movie releases across different days of the week. The data suggests the following insights:

- **Popularity of Release Days**: Friday leads as the most popular day for movie releases with 23 movies, which is notable as weekends are traditionally considered favorable for movie-goers.
  - To optimize weekend box office earnings, Hollywood studios deliberately selected Fridays as the best day for new movies. This makes sense because Friday is considered the last working day of the week.
- **Weekend Releases**: Contrary to what might be expected, Saturday has significantly fewer releases, with only 8 movies, indicating that this dataset does not follow the typical pattern of heavy weekend releases. And for the data to be more accurate we must have a significantly larger data set.
- **Midweek Trends**: Wednesday shows the lowest number of releases, just 7 movies, which may indicate a less favorable reception for midweek releases or a strategic avoidance of this day by distributors.

Complex Graph Analysis

Creating the Graph

To find the average genre rating for every genre of 2023 and 2024, we had to use the tables "Rating", "MovieHasGenre", "Genre", and "Releasing."

- The rating table holds all the user ratings for every movie We then found the average rating for every movie which we made into a separate table
- We joined this with our "MovieHasGenre" table, which contains movie IDs and Genre Ids. We used this to find a list of movies for every genre
- Our Genre table, which contains GenreID by Genre, we used to connect our GenreIDs from the previous step
- Lastly, we filtered in only movies released in 2023 and 2024.

Graph Analysis

Highly Rated Genres: Adventure, Fantasy, Spy, Supernatural

These 4 genres all rated above 4 stars. They often involve elements of escapism, high stake conflicts, and imaginative words, which can be exciting for both children and adults.

Poorly Rated Genres: Art, Parody, Melodrama, Alternate History

These genres rated averaged between 1-2 stars, with the Art and Parody genres in particular being closer to 1 star. Art, Melodrama, and Alternate History are much more grounded genres than those that rated highly, often with less exciting themes and emotions. While these genres can be deeply meaningful and thought provoking, they often require additional context to fully appreciate or simply don't appeal to a mainstream audience.

Exploratory Analysis Using R

In our exploratory data analysis using R, we utilized various statistical measures to understand the distribution and characteristics of our data. focusing on key metrics such as the mean, median, and standard deviation allowed us to analyze user ratings, age ratings, movies rated per user, and movies directed per director. By employing built-in functions in R, such as **mean()**, **median()**, and **sd()** (standard deviation), we efficiently computed these statistics for our data.

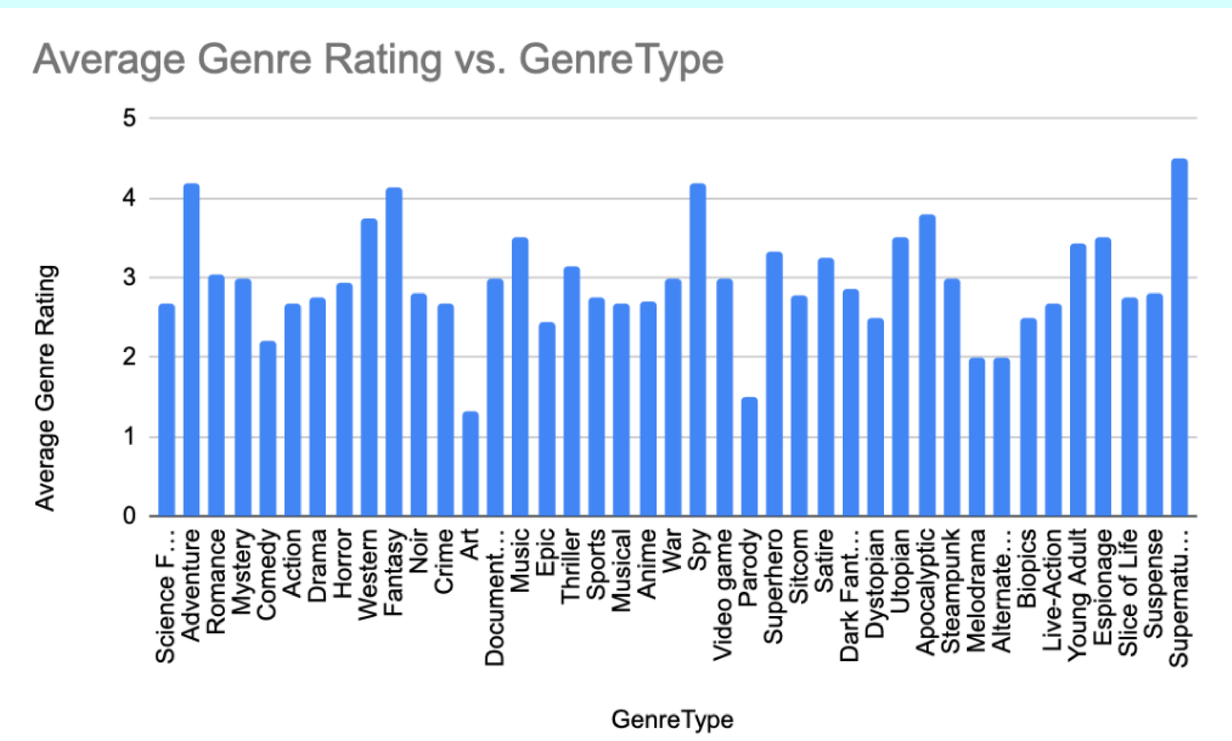
User Ratings:
mean: 2.941667
median: 3.000000
standard deviation: 1.404051

Age Ratings:
Note: For age ratings, we had to approximate "G" as 0, and "PG" as 5 to allow for numerical analysis
mean: 9.900000
median: 13.000000
standard deviation: 6.551274

Movies Rated Per User
mean: 1.329114
median: 1.000000
standard deviation: 0.593156

Movies Directed by Each Director:
mean: 1.316456
median: 1.000000
standard deviation: 0.610681

Complex Graph



Technologies Used

- Excel
- R
- Datagrip
- Python