

Interactive Visual Analytics for Movie Ratings:

Exploring Audience Preferences and Trends in Movie Data

Aitor Jauregui Jimenez
Sapienza University

A decorative light blue triangle is located in the bottom right corner of the slide, pointing towards the top right.

Título de la sección

Goal of the Project

To design and develop an interactive visual analytics system for exploring movie ratings data.

Empower users to uncover trends, identify patterns, and gain insights into audience preferences.

Focus on:

- Enabling exploratory data analysis.
- Combining multiple visualizations for coordinated interactions.
- Supporting filtering and search functionality for targeted exploration.

Dataset

Dataset Overview

- **Title:** Dataset: MovieLens Small Dataset
- **Content:**
 - Source: MovieLens datasets (ml-latest-small).
 - **Structure:**
 - **movies.csv:** Contains movie metadata (movieId, title, genres).
 - **ratings.csv:** User ratings for movies (userId, movieId, rating, timestamp).
 - **tags.csv:** User-generated tags for movies (userId, movieId, tag, timestamp).
 - **Size:** 100,836 ratings, 9,742 movies, and 610 users.
 - **Preprocessing:**
 - Data cleaning, genre extraction, and rating aggregation.
 - PCA applied for dimensionality reduction.

movieId,title,genres

1,Toy Story (1995),Adventure|Animation|Children|Comedy|Fantasy
2,Jumanji (1995),Adventure|Children|Fantasy
3,Grumpier Old Men (1995),Comedy|Romance
4,Waiting to Exhale (1995),Comedy|Drama|Romance
5,Father of the Bride Part II (1995),Comedy
6,Heat (1995),Action|Crime|Thriller
7,Sabrina (1995),Comedy|Romance
8,Tom and Huck (1995),Adventure|Children
9,Sudden Death (1995),Action
10,GoldenEye (1995),Action|Adventure|Thriller

userId,movieId,rating,timestamp

1,1,4.0,964982703
1,3,4.0,964981247
1,6,4.0,964982224
1,47,5.0,964983815
1,50,5.0,964982931
1,70,3.0,964982400
1,101,5.0,964980868
1,110,4.0,964982176
1,151,5.0,964984041

Data Structure

Data Structure

- **Title:** Structure of Preprocessed Data
- **Content:**
 - **Movies:**
 - Fields: movieId, title, genres (as a list).
 - **Ratings:**
 - Fields: userId, movieId, rating (rounded and original), timestamp.
 - **Aggregated Statistics:**
 - Fields: movieId, average_rating, total_ratings.
 - **After PCA:**
 - Fields: PC1, PC2 (principal components from dimensionality reduction).

Data Structure

```
1  userId,movieId,rating,timestamp,title,genres,year,tag,average_rating,total_ratings
2  1,1,4.0,2000-07-30 18:45:03,Tot Story (1995),["Adventure', 'Animation', 'Children', 'Comedy', 'Fantasy']",1995.0,["pixar', 'pixar', 'fun']",3.9209302325581397,215
3  1,3,4.0,2000-07-30 18:20:47,Grumpier Old Men (1995),["Comedy', 'Romance']",1995.0,["moldy', 'old']",3.2596153846153846,52
4  1,6,4.0,2000-07-30 18:37:04,Heat (1995),["Action', 'Crime', 'Thriller']",1995.0,,3.946078431372549,102
5  1,47,5.0,2000-07-30 19:03:35,Seven (a.k.a. Se7en) (1995),["Mystery', 'Thriller']",1995.0,["mystery', 'twist ending', 'serial killer']",3.9753694581280787,203
6  1,50,5.0,2000-07-30 18:48:51,"Usual Suspects, The (1995)",["Crime', 'Mystery', 'Thriller']",1995.0,["mindfuck', 'suspense', 'thriller', 'tricky', 'twist ending', 'heist']",4.237745098039215,204
7  1,70,3.0,2000-07-30 18:40:00,From Dusk Till Dawn (1996),["Action', 'Comedy', 'Horror', 'Thriller']",1996.0,,3.5090909090909093,55
8  1,101,5.0,2000-07-30 18:14:28,Bottle Rocket (1996),["Adventure', 'Comedy', 'Crime', 'Romance']",1996.0,["crime', 'off-beat comedy', 'quirky']",3.782608695652174,23
9  1,110,4.0,2000-07-30 18:36:16,Braveheart (1995),["Action', 'Drama', 'War']",1995.0,["beautiful scenery', 'epic', 'historical', 'inspirational', 'Medieval', 'mel gibson', 'Oscar (Best Cinematography)', 'revenge', 'sword fight']",
10  1,151,5.0,2000-07-30 19:07:21,Rob Roy (1995),["Action', 'Drama', 'Romance', 'War']",1995.0,,3.5454545454545454,44
```

Visualizations

The selection of the visualizations

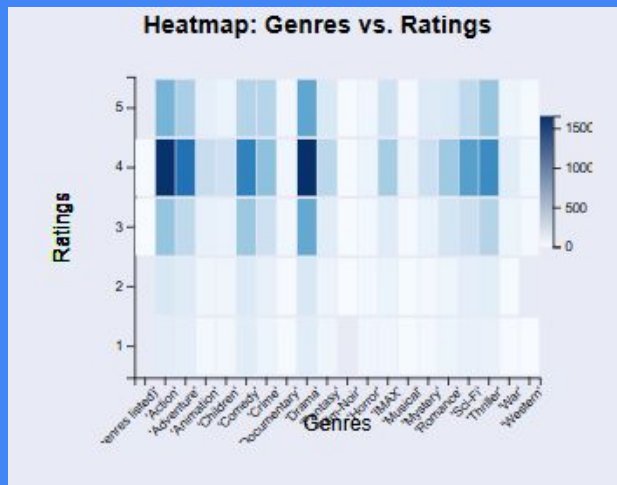
Chosen Visualizations - Overview

- **Title:** Chosen Visualizations
- **Content:**
 - The system uses four key visualizations to present data:
 1. **Heatmap**
 2. **Bar Chart**
 3. **Scatter Plot (years)**
 4. **Scatter Plot (Original Data)**
 5. **Scatter Plot (PCA)**

Visualizations were chosen to emphasize trends in ratings, genre preferences, and movie popularity.

Heatmap

Genre vs. Rating Correlation.



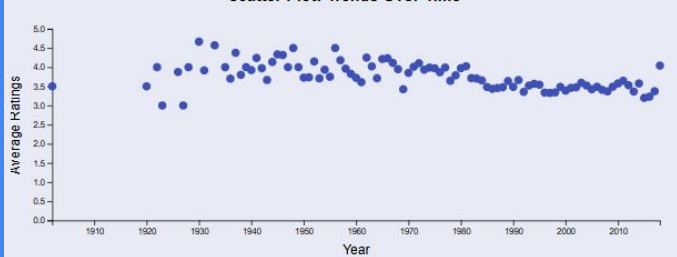
Content:

- **Purpose:** Explore how ratings are distributed across genres.
- **Axes:**
 - X-axis: Genres (sorted alphabetically).
 - Y-axis: Ratings (rounded to integers from 1 to 5).
- **Color Encoding:**
 - Represents the count of ratings for each genre-rating pair.
 - Darker colors = higher counts.
- **Example Insight:** Most ratings cluster around 3.5 to 4 across all genres

Scatter Plot (years)

Showing years with
the average rating
of each year

Scatter Plot: Trends Over Time

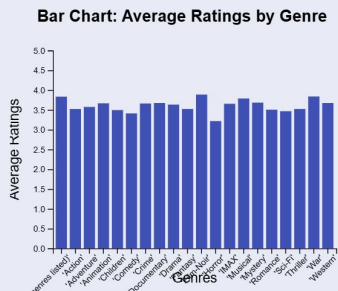


Content:

- **Purpose:** Explore how ratings are distributed across years.
- **Axes:**
 - X-axis: Years (from 1900 to 2018)
 - Y-axis: Ratings (from 0 to 5).
- **Example Insight:**
- We can see that during the late 50's the scatter plot suffers a fall in the ratings.
- We can appreciate that the older the films are the more spread out they are

Bar Chart

Average Rating by Genre



Content:

- **Purpose:** Provide an overview of genre performance.
- **Axes:**
 - X-axis: Genres.
 - Y-axis: Average rating.
- **Features:**
 - Bars sorted by average rating.
 - Dynamic updates based on applied filters.
- **Example Insight:** Film noir is the worst rated genre of all genres

Scatter Plots

RAW vs PCA

We see the relation between the avg rating and total rating displayed normally and after applying PCA

Content:

- **Purpose:** Visualize the relationship between a movie's popularity and its average rating AND simplify high-dimensional data for easier interpretation.
- **Axes:**
 - X-axis: Total ratings (popularity) vs Principal Component 1
 - Y-axis: Average rating (quality) vs Principal Component 2
- **Features:**
 - Each point represents a movie, clustered based on similar characteristics.
 - Hovering displays movie details.
- **Example Insights:**
 - PCA reveals distinct clusters of genres and user preferences.
 - Older movies with fewer ratings often have higher average ratings, indicating a potential nostalgia bias.

Thanks