



# Visualizing Movie Recommendation System

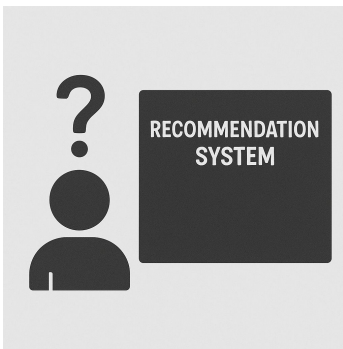


A Data Visualization Approach to Personalized Movie Recommendations

Sanjana J D  
Ashutosh Rane

# Challenges in Movie Recommendation Systems

- **Lack of Transparency:** Users frequently question why specific movies are recommended, resulting in skepticism and reduced trust in the system.
- **Cold-Start Problem:** New users with no prior rating history receive generic or irrelevant suggestions, hindering their initial experience.
- **User Engagement:** Conventional systems often feel static and impersonal, leading to lower interaction and satisfaction levels.





# Motivation and Goals

**Enhance Transparency:** Use visualizations to explain recommendation logic (e.g., similarity charts, latent space plots).

**Solve Cold-Start Problem:** Develop an intuitive interface for new users to input preferences via genre and movie selection.

**Improve User Experience:** Create an interactive and engaging system with visual insights into recommendations.

**Integrate Future Movies:** Include upcoming 2025 movies to make recommendations more relevant and forward-looking.

## Recommend for New User

Let's find movies you'll love! First, select the genres you enjoy.

Drama

Comedy

Action

Thriller

Romance

Sci-Fi

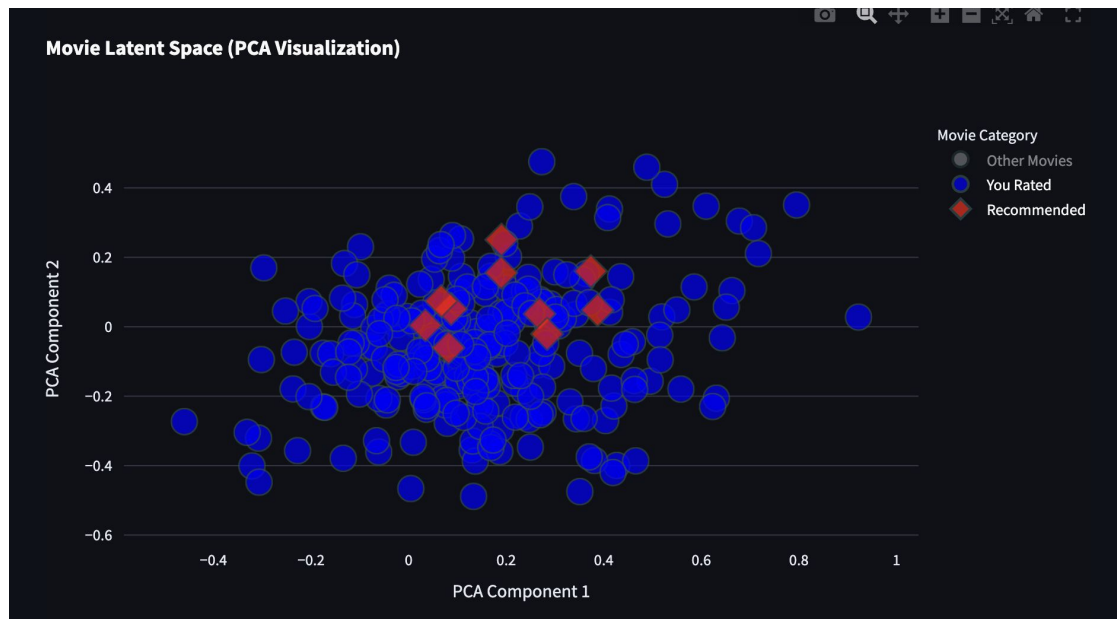
Adventure

Fantasy

Crime

Animation

# Motivation and Goals





# Motivation and Goals

**Enhance Transparency:** Use visualizations to explain recommendation logic (e.g., similarity charts, latent space plots).

**Solve Cold-Start Problem:** Develop an intuitive interface for new users to input preferences via genre and movie selection.

**Improve User Experience:** Create an interactive and engaging system with visual insights into recommendations.

**Integrate Future Movies:** Include upcoming 2025 movies to make recommendations more relevant and forward-looking.

## Recommend for New User

Let's find movies you'll love! First, select the genres you enjoy.

Drama

Comedy

Action

Thriller

Romance

Sci-Fi

Adventure

Fantasy

Crime

Animation



# Data Sources and Insights

## MovieLens Small Dataset:

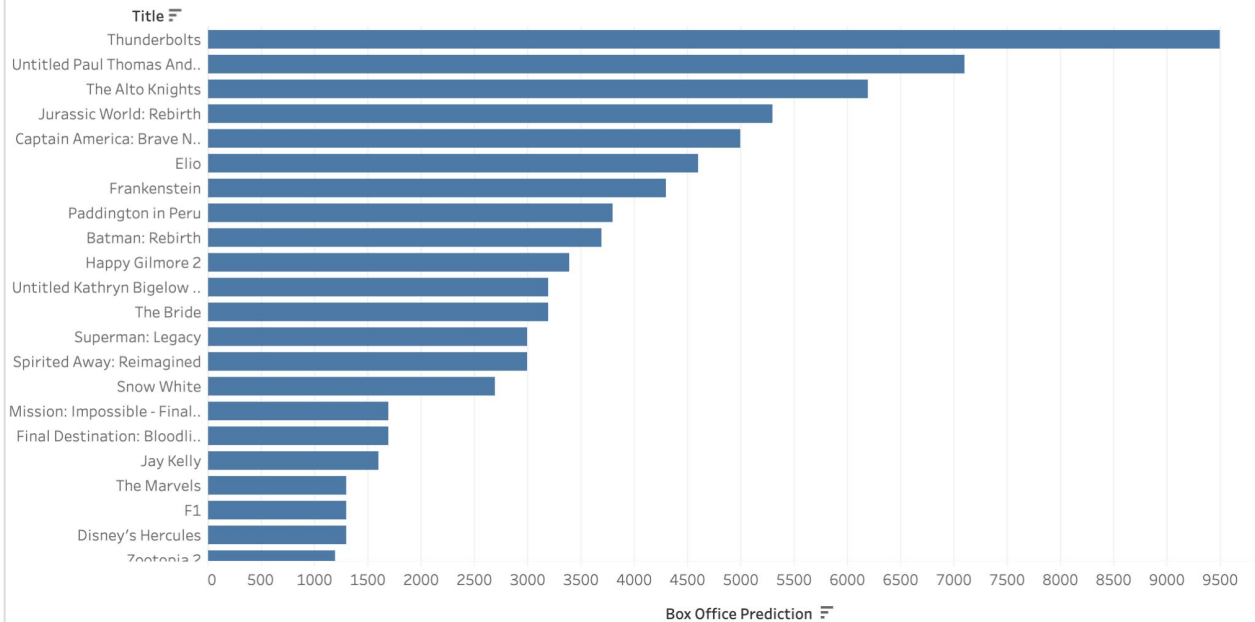
- Source: movies.csv, ratings.csv
- Details: 100,000 ratings, 9,000+ movies, 600+ users.
- Purpose: Historical ratings for collaborative filtering.

## Top 100 Trending Movies of 2025:

- Source: Kaggle dataset by taimoor888.
- Details: Upcoming movies with box office predictions.
- Purpose: Recommend future movies and estimate ratings.

# Data Sources and Insights

Top Upcoming Movies





# Project Toolkit: Powering the Recommendation System

## Core & Data:



**Python:** Core programming language for development.



**Pandas:** Essential for data loading, manipulation, cleaning, and preparation (EDA summaries).

## Machine Learning & Analysis



**Surprise:** Building and evaluating the SVD collaborative filtering recommendation model.



**Scikit-learn:** Performing PCA (Principal Component Analysis) on SVD factors for dimensionality reduction.





# Project Toolkit: Powering the Recommendation System

## Visualization & Interaction



**Tableau:** Used for initial Exploratory Data Analysis (visualizing distributions, trends)



**Plotly :** Creating interactive visualizations (genre comparison, similarity scores, latent space plot) embedded in the web app.



**Streamlit:** Building the interactive web application/dashboard for user interaction, recommendation display, and model explanation visuals.



# Novel Aspects of the Project

- **Interactive User Input:** Genre and movie selection blocks for new users, simplifying the input process.
- **Latent Space Visualization:** Used PCA to visualize movie relationships in a 2D scatter plot.
- **Transparent Recommendations:** Provided visual explanations of recommendation logic through similarity charts.



# Novel Aspects of the Project

**User:** User 1

**User 1's Ratings:**

- Movie 1 (Drama, Thriller): 4.0
- Movie 2 (Drama, Comedy): 5.0
- Movie 3 (Comedy): 3.0

**Step 1: Compute Genre Averages**  
(`get_user_genre_averages`):

- Drama:  $(4.0 + 5.0) / 2 = 4.5$
- Thriller:  $4.0 / 1 = 4.0$
- Comedy:  $(5.0 + 3.0) / 2 = 4.0$
- Overall average:  $(4.0 + 5.0 + 3.0) / 3 = 4.0$
- Result: `genre_averages = {'Drama': 4.5, 'Thriller': 4.0, 'Comedy': 4.0}`,  
`overall_avg = 4.0`

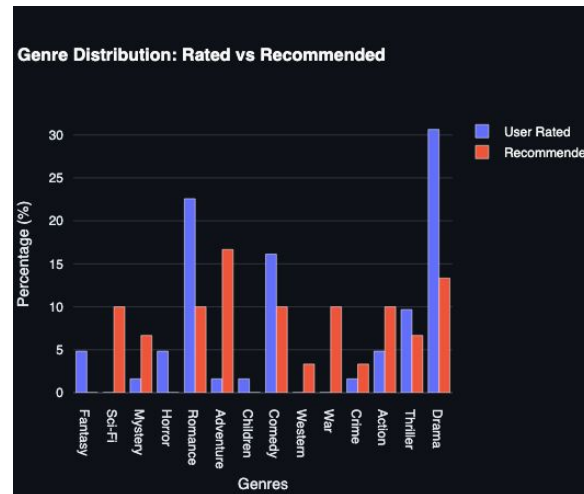
**Step 2: Estimate Rating for a New Movie**  
(`estimate_rating_for_new_movie`):

- New Movie: Thunderbolts (movieid 193610, Genres: Action, Adventure)
- User 1 hasn't rated Action or Adventure movies, so both genres use the `overall_avg`:
  - Action: 4.0
  - Adventure: 4.0
- Estimated rating:  $(4.0 + 4.0) / 2 = 4.0$

# Visualization Techniques Used

## Genre Comparison Bar Chart:

- Purpose: Compares user preferences with recommended movies.
- Example: Shows percentage of genres in user-rated vs. recommended movies.



## Latent Space Scatter Plot:

- Purpose: Visualizes movie relationships in 2D using PCA.
- Example: Blue dots for user-rated movies, red for recommended.

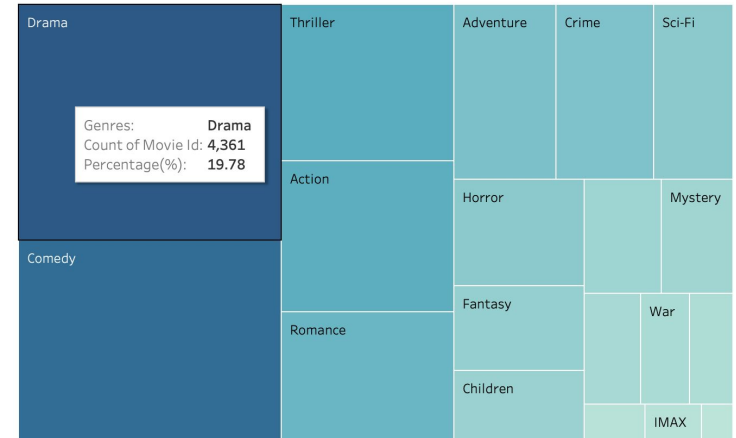


# Visualization Techniques Used

## Tableau Dashboard:

- Purpose: Displays Genre Distribution and top movies.  
To show distribution of ratings across each genre.
- Example: Drama has the highest number of user rating.

Genre Distribution





# Visualization Techniques Used

[Dashboard](#)



# Visualization Techniques Used

## Interactive Blocks on Streamlit:

- Purpose: Allows new users to select genres and movies.
- Example: Genre buttons (e.g., Drama, Comedy) and movie "Like" buttons.

Let's find movies you'll love! First, select the genres you enjoy.

Drama	Comedy	Action	Thriller	Romance
Sci-Fi	Adventure	Fantasy	Crime	Animation

---

Great! Now, select the movies you like from these genres.

### Random Drama Movies

Hard-Boiled (Lat sau san taam) (1992) <i>Action Crime Drama Thriller</i> <div>Like</div>	Facing Windows (Finestra di fronte, La) (2003) <i>Drama Romance</i> <div>Like</div>	Lady Vanishes, The (1938) <i>Drama Mystery Thriller</i> <div>Like</div>
McFarland, USA (2015) <i>Drama</i> <div>Like</div>	Long Walk Home, The (1990) <i>Drama</i> <div>Like</div>	Oxygen (1999) <i>Crime Drama Thriller</i> <div>Like</div>



# Individual Contribution

Contribution Area	Sanjana	Ashutosh
Data Exploration & Curation	Performed <b>Exploratory Data Analysis (EDA)</b> using <b>Tableau</b> ; Researched & added latest/upcoming movies	
Machine Learning	Integrated and implemented <b>SVD</b> for new movies dataset	Implemented <b>SVD Recommendation Model</b> Developed recommendation logic
Application Development	Embedded <b>Plotly</b> visualizations. Added on to the Streamlit app for new users	Built the interactive <b>Streamlit Web Application</b> ;Embedded <b>Plotly</b> visualizations



## About

This app uses SVD collaborative filtering for existing users and similarity-based recommendations for new users, with genre-based estimation for new movies, similarity analysis, and latent space visualization.

- Existing User: SVD + genre estimation.
- New User: Similarity-based with SVD factors and genre overlap.  
Data: MovieLens (movies\_ratings.csv) + Future Movies (updated\_new\_movies.csv)  
Model: Surprise SVD + Scikit-learn PCA Visualization:  
Streamlit + Plotly





# Visualising Movie Recommendation System

Movies Ratings data loaded successfully.

New Movies data loaded successfully.

Data loaded and preprocessed successfully!

Model trained and PCA computed!

 [Recommend for Existing User](#)  [Recommend for New User](#)

## Select Existing User

Choose a User ID:

578





# Future Improvements

## Advanced Recommendation Models:

- **Hybrid Models (CF + Content):** Improve cold-start performance & diversity.
- **Deep Learning Models:** Explore for higher accuracy & complex patterns.

## Enhanced Explainability & User Experience:

- **Enhanced Explainability:** Deeper insights into "Why?".
- **Diversity/Serendipity Controls:** User choice for recommendation novelty.
- **User Accounts:** Profiles, watchlists, persistent history.

## Deployment & Evaluation:

- **Containerization (Docker):** Simplify deployment.
- **Robust Evaluation:** Implement more comprehensive offline evaluation metrics like Precision/Recall, NDCG, and Mean Average Precision (MAP). Consider strategies for online A/B testing if the app were deployed.



# Demo

Link to our application: [Demo](#)