Content-Based Movie Recommendation System

Zena Lisa Karari

Alex Mwera

Noel Seda

Business Understanding

Objective:

Build a content-based engine to suggest movies by genre similarity.

Goal:

Improve user discovery and engagement in a large movie catalog.

Problem Statement

- Users overwhelmed by thousands of titles lack personalized guidance.
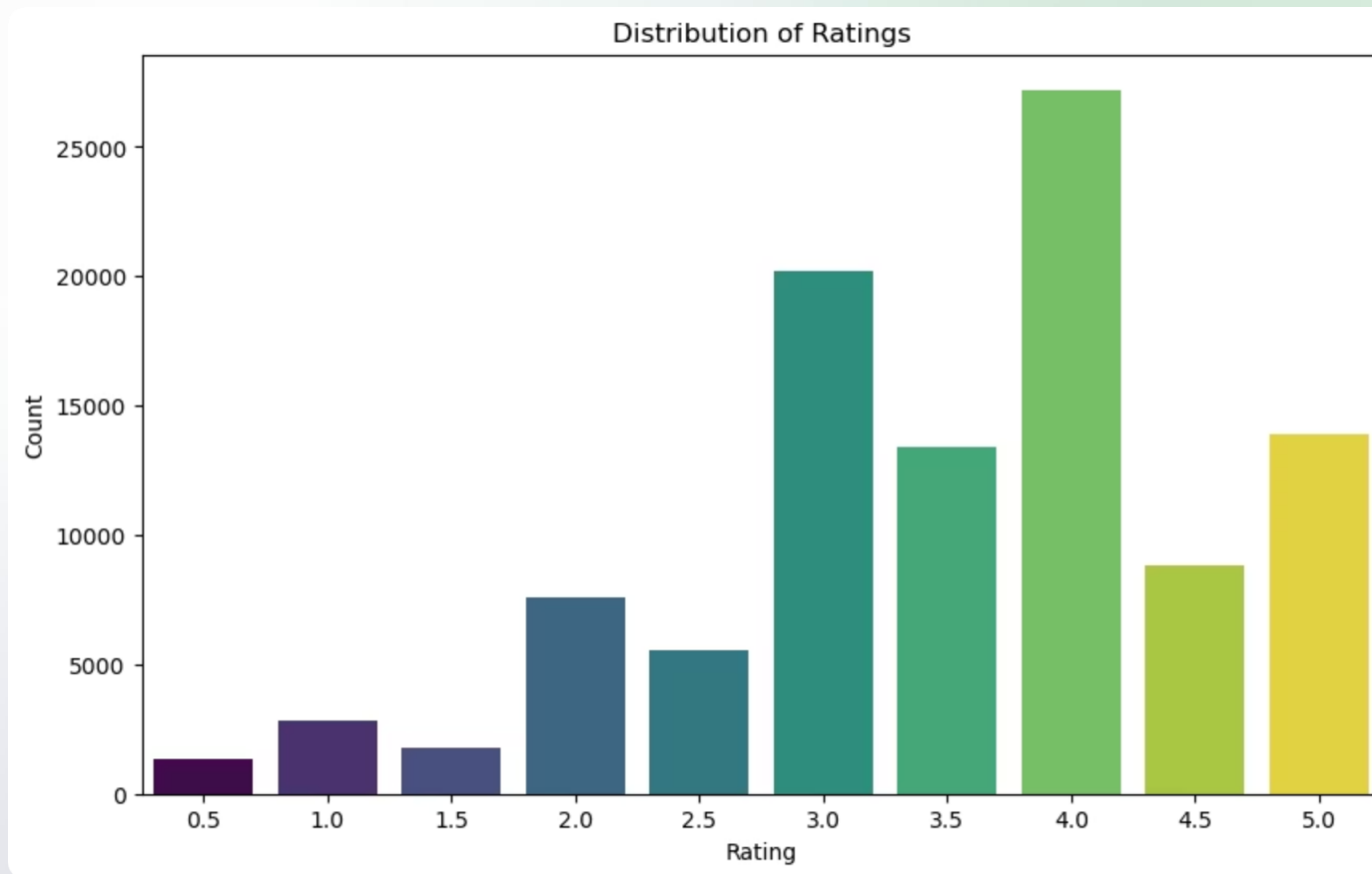- Need an automated system to filter and surface relevant movies.

Data Overview

Dataset:

MovieLens "ml-latest-small" – 9,700 movies, 100 K ratings, tags & links.

Key fields: movieId, title, genres, userId, rating, timestamp.

Data Preprocessing & Merging

- Loaded and merged movies.csv, ratings.csv, tags.csv, links.csv.

- Cleaned duplicates, handled missing values.

- Created a "description" field by combining genres.

Distribution of Ratings
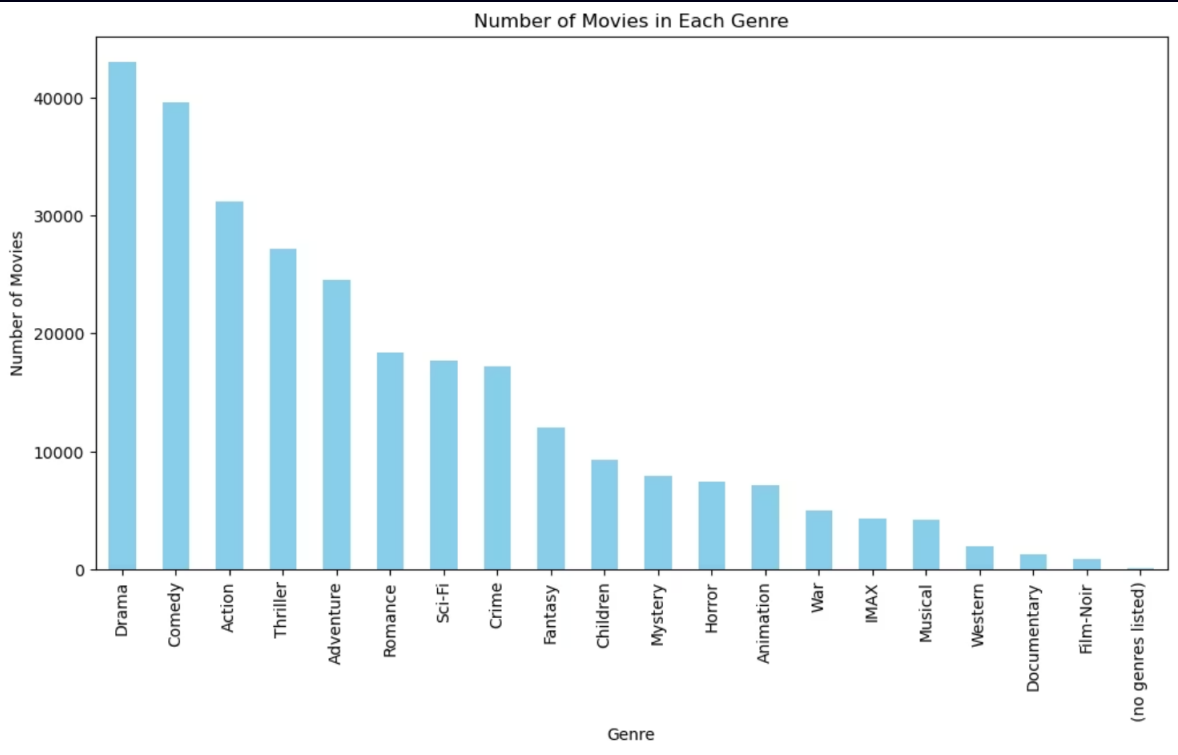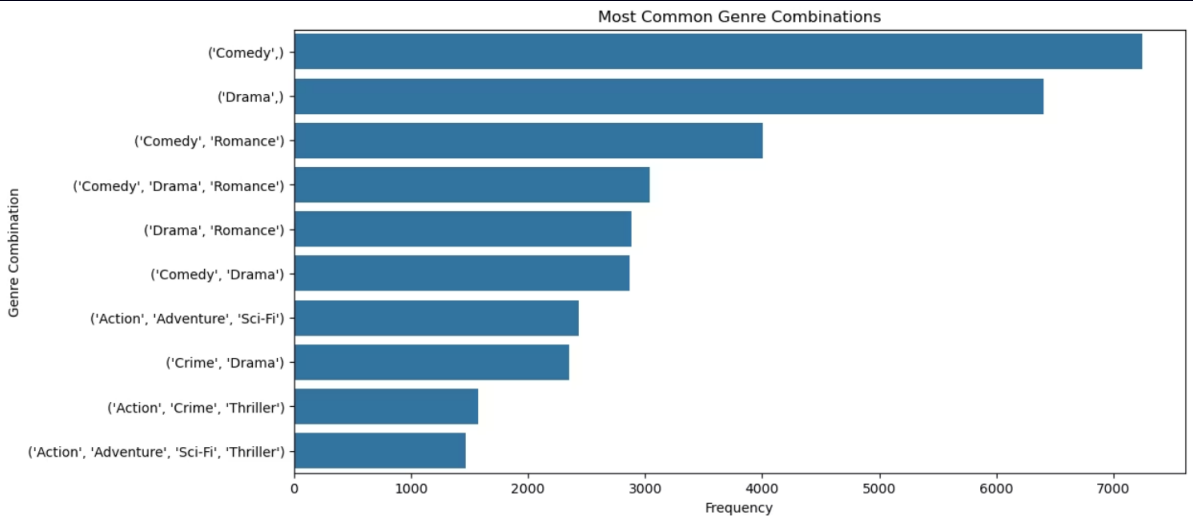
EDA – Ratings Distribution
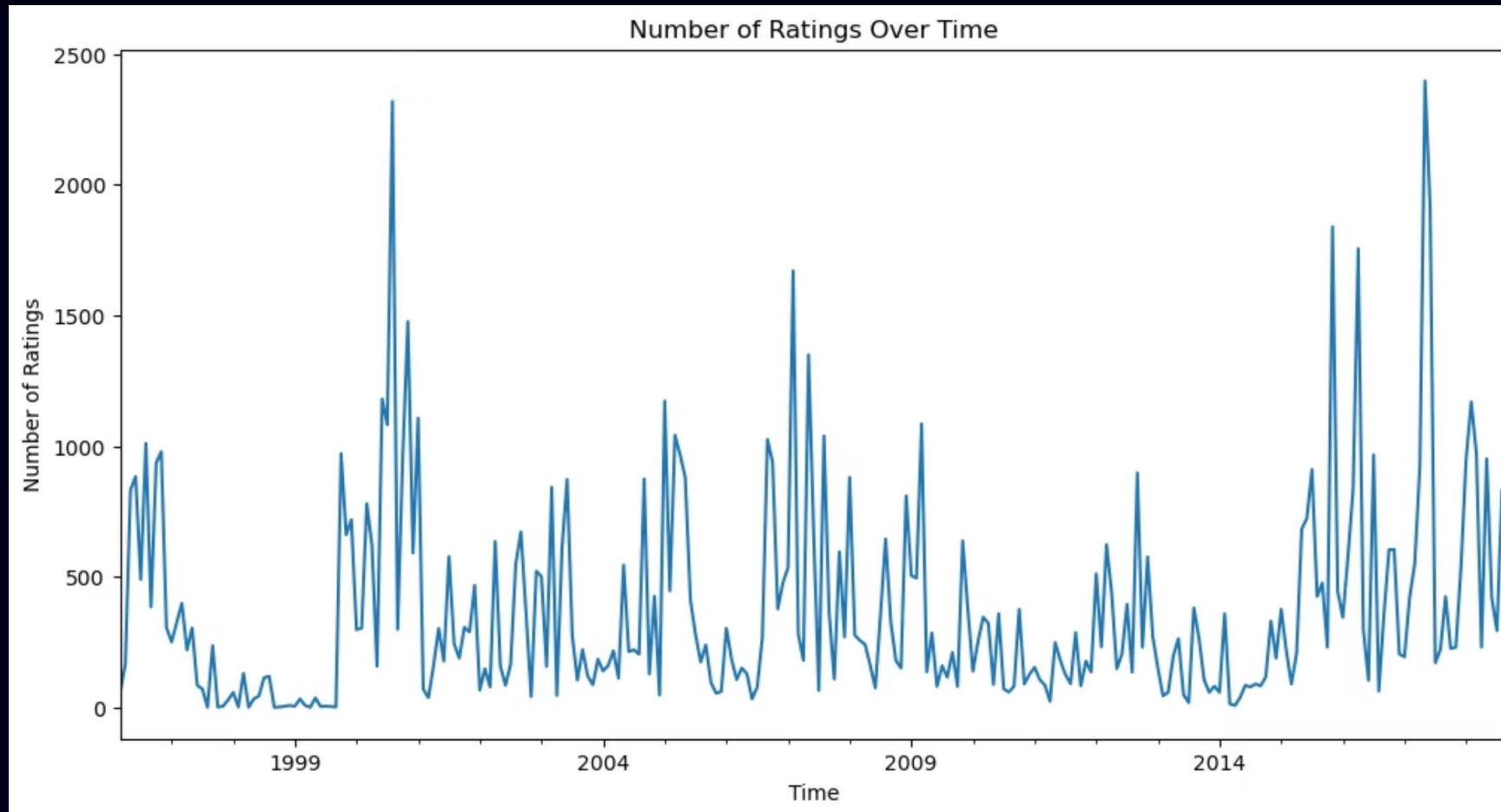
Visualized how users rate movies to detect skew.

Insight: Most ratings cluster between 3–4 stars, indicating generally positive feedback.

EDA – Genre Analysis

Counted movies per genre and most frequent multi-genre combos.

Insight: Drama, Comedy, Thriller dominate. Drama/Comedy is top combo.

Number of Ratings Over Time

EDA – Rating Activity Over Time

Trended rating volume by month to gauge engagement growth.

Insight: Steady increase in ratings activity.

Modelling Approach

Step 1

Applied TF-IDF vectorization on the combined genres text (top 50 features).

Step 2

Computed cosine similarity matrix to measure pairwise movie similarity.

Recommendation Function

get_recommendations(title)– Finds a movie's index, sorts by descending cosine similarity, returns top N.– Outputs title, genres, average rating

Example: Run get_recommendations('Toy Story (1995)')

Recommendations align closely with the input movie's genres (e.g., Animation, Adventure).

```
                                               title  \
0                               Toy Story 2 (1999)
1                             Monsters, Inc. (2001)
2                                       Antz (1998)
3           Adventures of Rocky and Bullwinkle, The (2000)
4                      Emperor's New Groove, The (2000)
5                              Shrek the Third (2007)
6                            The Good Dinosaur (2015)
7   Asterix and the Vikings (Astérix et les Viking...
8                         Tale of Despereaux, The (2008)
9                                      Moana (2016)

                                        genres  rating
0  Adventure, Animation, Children, Comedy, Fantasy     4.5
1  Adventure, Animation, Children, Comedy, Fantasy     4.0
2  Adventure, Animation, Children, Comedy, Fantasy     3.0
3  Adventure, Animation, Children, Comedy, Fantasy     4.0
4  Adventure, Animation, Children, Comedy, Fantasy     4.5
5  Adventure, Animation, Children, Comedy, Fantasy     4.5
6  Adventure, Animation, Children, Comedy, Fantasy     3.0
7  Adventure, Animation, Children, Comedy, Fantasy     5.0
8  Adventure, Animation, Children, Comedy, Fantasy     3.0
9  Adventure, Animation, Children, Comedy, Fantasy     4.5
```

Conclusion & Future Work

Conclusion:

- Successfully implemented a genre-based recommender.

- Generated sensible, content-driven movie suggestions.

Future Work:

- Incorporate collaborative filtering (user profiles & ratings).

- Expand features (movie descriptions, tags, metadata).

- Deploy as a web service with real-time scoring.