

















 Haender-Michael / **Movie-recommendation**

🔍 Type  to search

  |      

 **Code**  Issues  Pull requests  Actions  Projects  Wiki  Security  Insights  Settings



🔗 main ▾

[Movie-recommendation](#) / README.md 

🔍 Go to file

t

**Haender-Michael** Update README with user recommendations f98e704 · now  History

Preview

Code

Blame



Raw



Movie recommendations.



[source](#)

Introduction:

For this project, we will use the "small" dataset from [GroupLens](#). This dataset includes four main files:

- **movies.csv**: Contains information about each movie, including its ID, title, and genres.
- **ratings.csv**: Records user ratings, with fields for user ID, movie ID, rating, and timestamp.
- **tags.csv**: Aggregates user-generated keywords associated with movies.
- **links.csv**: Provides mappings between MovieLens IDs and external databases such as IMDb and TMDb.

Since good recommendations play an important role in client satisfaction we will make of Our objective to develop a recommendation system that can automatically suggest the top 5 movies to a given user. These recommendations will be based on the user's past ratings and those of similar users.

Methodology

Exploratory data Analysis

- dataset overview & cleaning
- dataset description
- data/ business understanding

Major questions

this part will sharpen our understanding of the data and the business task. Major questions will be answered. they will help us grasp the data.

Modelisation

- data preparation for modelisation
- training/ optimisation of SVD model
- recommendations

Contact information

- name
- email
- phone number
- LinkedIn

Exploratory data Analysis

Dataset overview & cleaning

the datasets movies_csv and ratings.csv that we will use are already cleaned which simplifies the process

Dataset Description

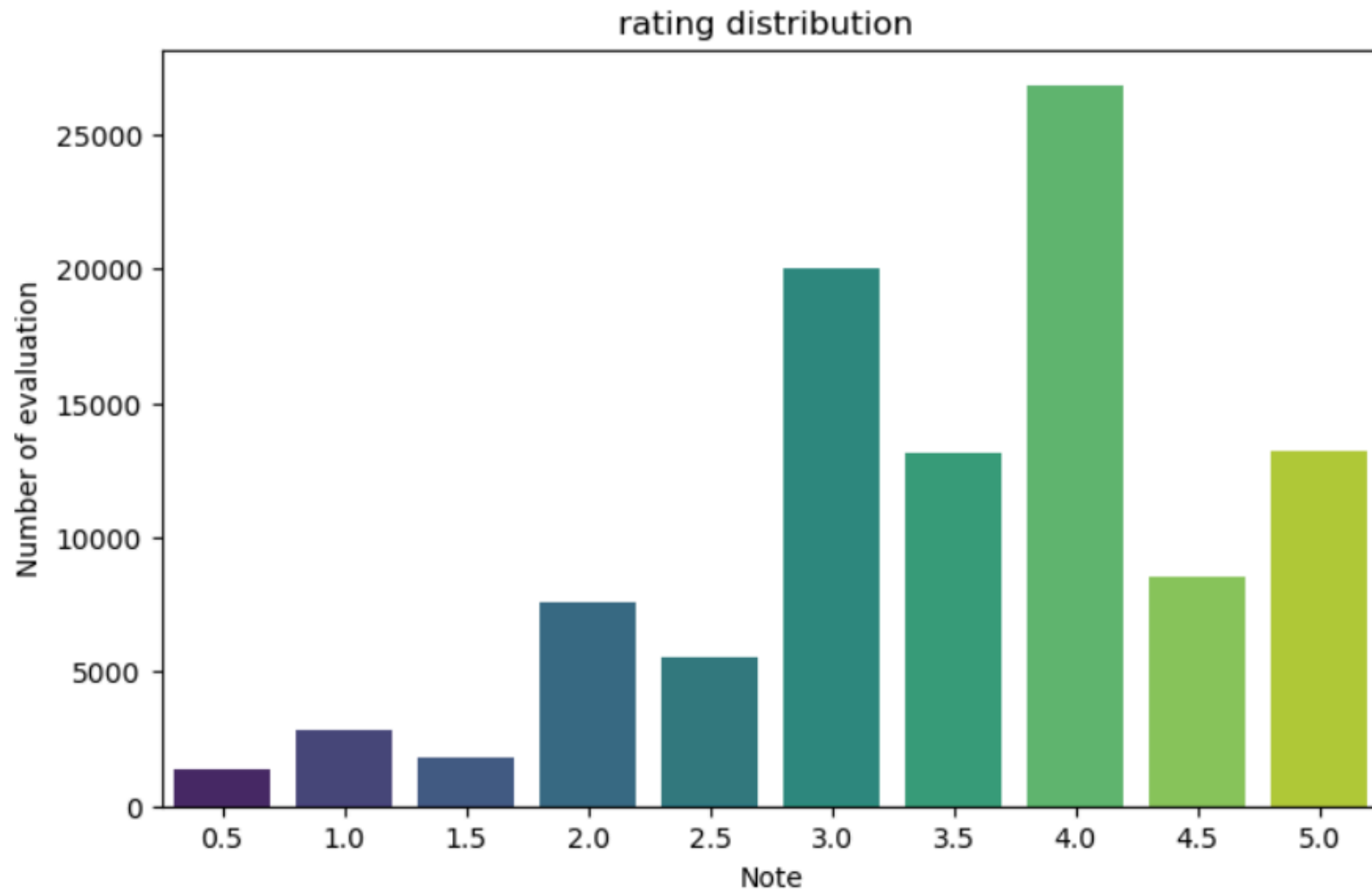
Some description have been provided in the introduction of this project. For more description of the data you may wanna read this: The dataset is composed of four main files: movies.csv: contains information about the movies (ID, title, genres). ratings.csv: contains user ratings (userId, movieId, rating, timestamp). tags.csv: gathers keywords associated with movies by users. links.csv: provides mappings between MovieLens IDs, IMDb, and TMDb. For this project, we will focus on movies.csv and ratings.csv, as they form the foundation of collaborative filtering: ratings.csv allows us to build the user–movie matrix based on the ratings given.

Data/business understanding

Video streaming platforms face a major challenge: delivering relevant content to each user from an immense catalog. Without personalized recommendations, users may feel lost and leave the platform, leading to lower retention. Our goal is to build a recommendation system capable of automatically suggesting the top 5 movies to a user, based on their past ratings and those of similar users. This collaborative filtering model will analyze shared behaviors among users to predict future preferences. By integrating such a solution, a streaming company can enhance the user experience, increase engagement, and reduce churn. The key idea behind collaborative filtering is that similar users share similar interests, and users tend to like items that others with similar tastes also enjoy.

Major Questions

What does the distribution of ratings reveal?



Answer : most ratings fall between 3 and 5, suggesting a generally positive trend.

What does the distribution of ratings per user reveal about user activity levels?

Answer: some users are highly active while others contribute very few ratings, highlighting the imbalance.

What does the number of ratings per movie reveal about the dataset's structure?

Answer: some movies have far more ratings than others.

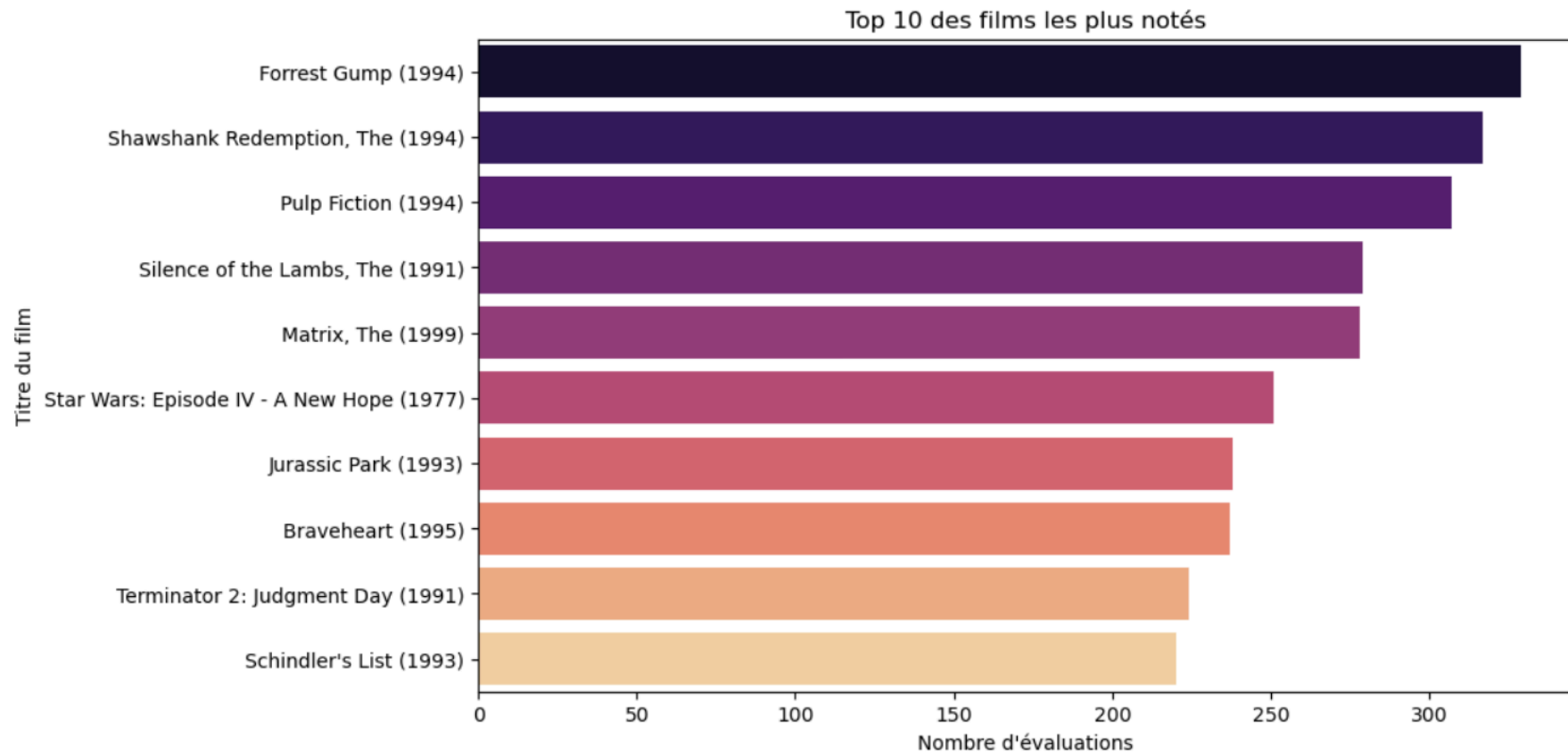
What does the average rating per movie suggest about overall film reception?

Answer: It reflects the observation that most movies are rated between 3 and 4.5, with few receiving very low scores.

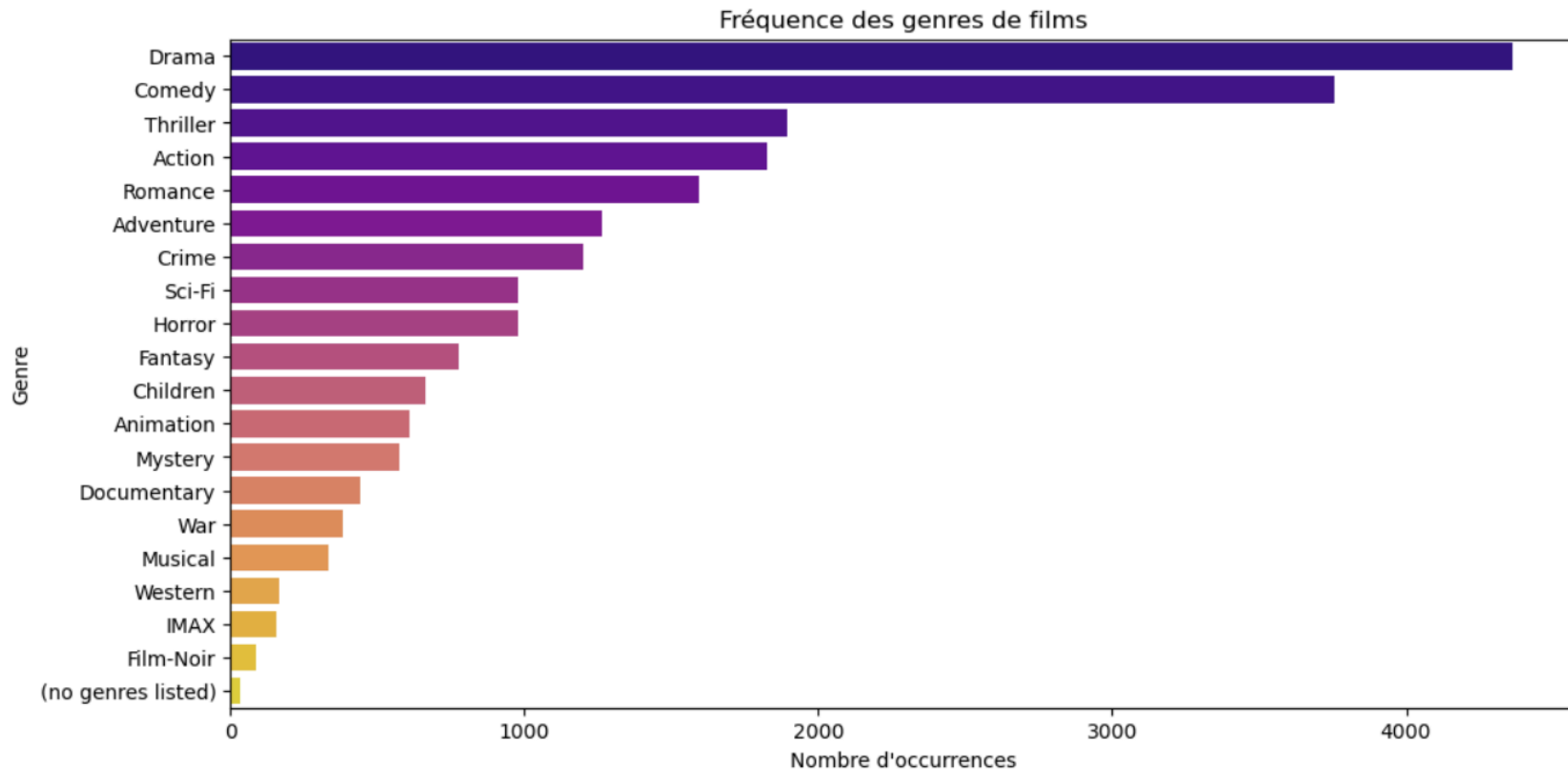
How are average ratings affected by the number of votes?

Answer: Movies with only a handful of ratings often show extreme values—either very high or very low—whereas those with many ratings tend to stabilize around an average of 3.5. This highlights the importance of having enough ratings to reliably gauge a film's actual popularity.

Top 10 Most Rated Movies



What are the most common genres ?



Drama, comedy and thriller are the most common genre in this dataset.

Modelisation

Data Preparation for modelisation

Remark

Some users have rated very few movies, and some films have received very few ratings. To improve the quality of collaborative filtering, we will we filter out users or movies with fewer than 5 ratings.

training an SVD Model

RMSE: 0.8504 RMSE SVD : 0.8504026440686286 the model's predictions are fairly close to the actual values but not perfect. we will optimize the model.

Optimisation of the SVD Model with GridSearchCV

Remark :

n_factors=150 → the model learns 150 latent dimensions (more complexity, greater ability to capture subtle preferences).
n_epochs=40 → more training iterations. lr_all=0.01 → a faster learning rate. reg_all=0.1 → stronger regularization to prevent overfitting.

Recommendations based on the model

Top 5 recommendations for user 1 : Recommendation #1: Shawshank Redemption, The (1994) (predicted rating : 5.00)

Recommendation #2: Vanya on 42nd Street (1994) (predicted rating : 5.00)

Recommendation #3: Wallace & Gromit: The Best of Aardman Animation (1996) (predicted rating : 5.00)

Recommendation #4: Godfather, The (1972) (predicted rating : 5.00)

Recommendation #5: Philadelphia Story, The (1940) (predicted rating : 5.00)

Top 5 recommendations for user 2 : Recommendation #1: Three Billboards Outside Ebbing, Missouri (2017) (predicted rating : 4.62)

Recommendation #2: Day of the Doctor, The (2013) (predicted rating : 4.56)

Recommendation #3: Swept Away (Travolti da un insolito destino nell'azzurro mare d'Agosto) (1975) (predicted rating : 4.54)

Recommendation #4: Last Tango in Paris (Ultimo tango a Parigi) (1972) (predicted rating : 4.52)

Recommendation #5: Captain Fantastic (2016) (predicted rating : 4.51)

contact information

- First Name: Haender Michael
 - Last Name: Jean Louis
 - Email: michaelhaenderjeanlouis@gmail.com
 - Phone Number: +509 41 75 0264
 - LinkedIn: https://www.linkedin.com/in/michael-haender-jean-louis-4b7320316?utm_source=share&utm_campaign=share_via&utm_content=profile&utm_medium=ios_app
- For further inquiries, feedback, or collaboration on this analysis, feel free to reach out. I welcome discussions and any contract to work.