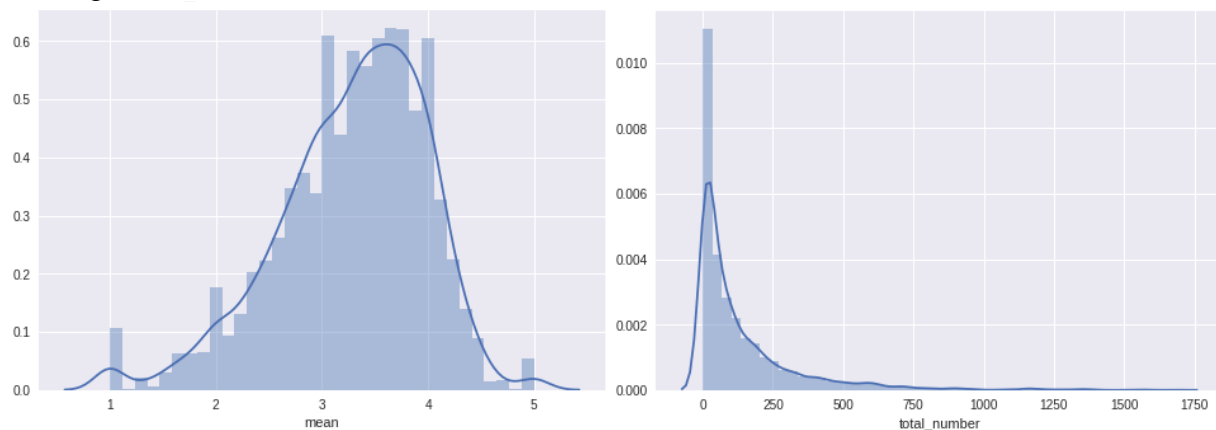Movie Recommendation

Watching movie is one of the most popular things that people will do during their leisure time. There is a lot of different type of movies, including Adventure, Action, Drama, etc. When I need some recommendation about new movies that I can watch during my leisure, others' suggestion is not exactly satisfying my taste. Hence, utilizing personal rating data and other rating data to provide the recommendation is an effective and efficient way. I used the dataset of users' movie ratings in the Kaggle to provide the recommendation for each user. (https://www.kaggle.com/c/movie/data). I utilized the Surprise Framework to create collaborative filtering to do the movie recommendation.

Firstly, I ingested the data into the Colab. Then, I merged the rating data with movie data as a new dataset. Utilizing dropna() to clean the dataset, which will be used in the EDA and Modeling part.

	user	movie	rating	id	names
0	2783	1253	5	2783_1253	1253::Day the Earth Stood Still, The (1951)::D...
1	2788	1253	4	2788_1253	1253::Day the Earth Stood Still, The (1951)::D...
2	2789	1253	4	2789_1253	1253::Day the Earth Stood Still, The (1951)::D...
3	2794	1253	5	2794_1253	1253::Day the Earth Stood Still, The (1951)::D...
4	2795	1253	4	2795_1253	1253::Day the Earth Stood Still, The (1951)::D...

(this is head 5 rows of the final raw dataset)

Secondly, I plotted the distribution of the movie ratings and the total number of ratings that each movie got.

(Left: this is a distribution plot of the movie's rating score. Based on this plot, we can easily figure out that most of movies ratings are around 3 to 4. There are only a few movies to be rated as 1 or 5. Right: this is a distribution plot that the total number of ratings that each movie gained. According to this plot, a large part of movies only gets less than 250 ratings.)

Then, I did the modeling part of movie recommendation. First of all, I evaluated the performance of collaborating filtering, with 100k rows for a faster process, and gain the value of RMSE and

MAE. Then, I only focused on the user 2783 and did the recommendation for him. Below is what user 2783 most like (rating as 5) in the past:

```
       user  movie  rating                                                names
0      2783   1253       5  1253::Day the Earth Stood Still, The (1951)::D...
289    2783    589       5  589::Terminator 2: Judgment Day (1991)::Action...
3232   2783    741       5  741::Ghost in the Shell (Kokaku kidotai) (1995...
3375   2783    750       5  750::Dr. Strangelove or: How I Learned to Stop...
4087   2783    924       5  924::2001: A Space Odyssey (1968)::Drama|Myste...
10481  2783   1301       5               1301::Forbidden Planet (1956)::Sci-Fi
15236  2783   3175       5  3175::Galaxy Quest (1999)::Adventure|Comedy|Sc...
18137  2783   1584       5                  1584::Contact (1997)::Drama|Sci-Fi
18823  2783   1748       5  1748::Dark City (1998)::Film-Noir|Sci-Fi|Thriller
19246  2783   2553       5  2553::Village of the Damned (1960)::Horror|Sci...
19370  2783   2571       5   2571::Matrix, The (1999)::Action|Sci-Fi|Thriller
20743  2783   3527       5     3527::Predator (1987)::Action|Sci-Fi|Thriller
21424  2783   1019       5  1019::20,000 Leagues Under the Sea (1954)::Adv...
22295  2783   1965       5               1965::Repo Man (1984)::Comedy|Sci-Fi
23462  2783   3576       5   3576::Hidden, The (1987)::Action|Horror|Sci-Fi
23602  2783    527       5       527::Schindler's List (1993)::Drama|War
29347  2783   1077       5             1077::Sleeper (1973)::Comedy|Sci-Fi
29768  2783   2968       5  2968::Time Bandits (1981)::Adventure|Fantasy|S...
```

Later, I still used SVD to predict the ratings of the movie that user 2783 will rate. And the following is the top ten movies that user 2783 would like to watch based on recommendation model.

	index	movie	names	Estimate_Score
2836	2836	2905	2905::Sanjuro (1962)::Action|Adventure	4.918621
604	604	608	608::Fargo (1996)::Crime|Drama|Thriller	4.896635
847	847	858	858::Godfather, The (1972)::Action|Crime|Drama	4.782098
740	740	750	750::Dr. Strangelove or: How I Learned to Stop...	4.766366
523	523	527	527::Schindler's List (1993)::Drama|War	4.762625
3065	3065	3134	3134::Grand Illusion (Grande illusion, La) (19...	4.760847
2255	2255	2324	2324::Life Is Beautiful (La Vita è bella) (199...	4.759253
2134	2134	2203	2203::Shadow of a Doubt (1943)::Film-Noir|Thri...	4.744759
662	662	668	668::Pather Panchali (1955)::Drama	4.741984
537	537	541	541::Blade Runner (1982)::Film-Noir|Sci-Fi	4.721818

In conclusion, based on the movie recommendation result, the user 2783 more enjoy the action and drama movie. And when this user wants to enjoy new movies during his leisure time, he can choose one of the movies from the top ten movie list.

My Colab Link:
https://colab.research.google.com/drive/1c1u3W9kCTkf9o_zOL1s8zcaW9SPu_HRn