



FlatIron Capstone: Letterboxd Recommendation System

Cristina Townsend

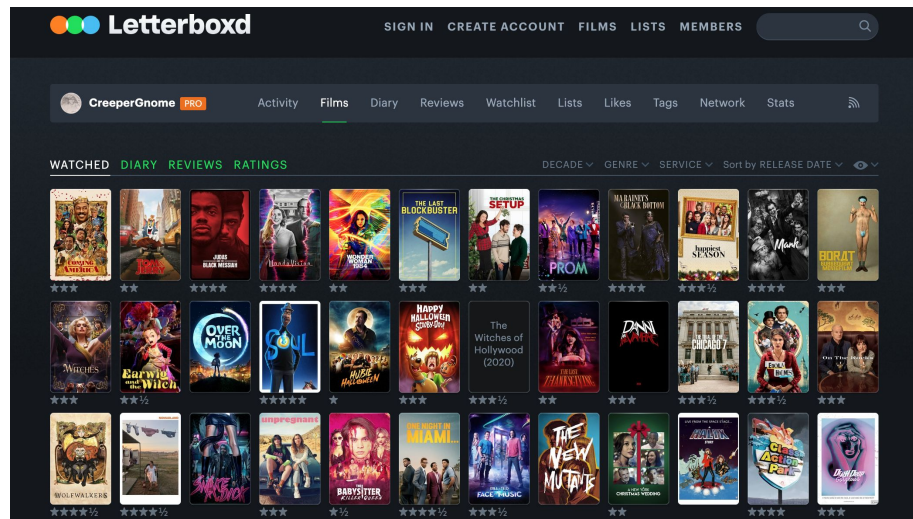


Business Case

LetterBoxd is a website used by film watchers to log, rate, and review movies. It is a website I use frequently and while it has a lot going for it, there isn't a recommended for you section. I decided to create my own recommendation system using data I scraped from the website.

Data

The main data frame has 579,707
rating from 1,199 users rating 77,772
different movies.



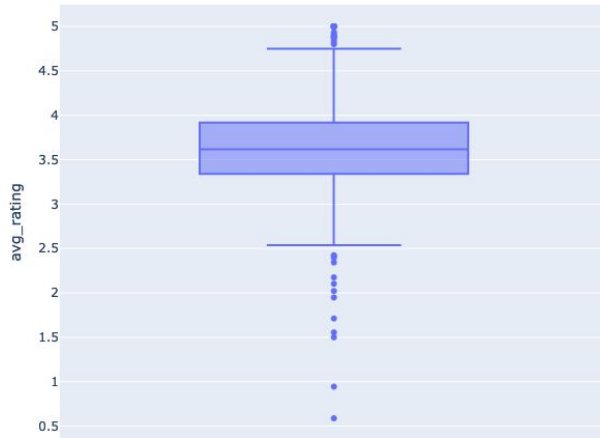
Data

After running and debugging a function that goes through Letterboxd pulling user ratings and film_id information we got a dataframe such as this.

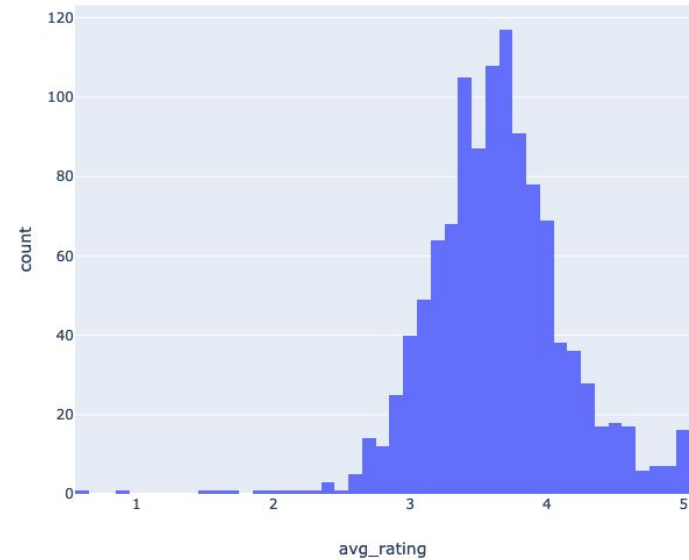
	Username	film_id	Rating
0	xbrookelee	320000	5.0
1	xbrookelee	564784	4.5
2	xbrookelee	579510	3.5
3	xbrookelee	510995	3.5
4	xbrookelee	460155	4.5
5	xbrookelee	441862	3.0
6	xbrookelee	451698	5.0
7	xbrookelee	268380	4.0
8	xbrookelee	318913	5.0
9	xbrookelee	228628	5.0
10	xbrookelee	456712	5.0
11	xbrookelee	421847	3.0
12	xbrookelee	394711	3.0
13	xbrookelee	397942	4.5
14	xbrookelee	433461	3.5

Break up of Ratings

Average Rating Per User



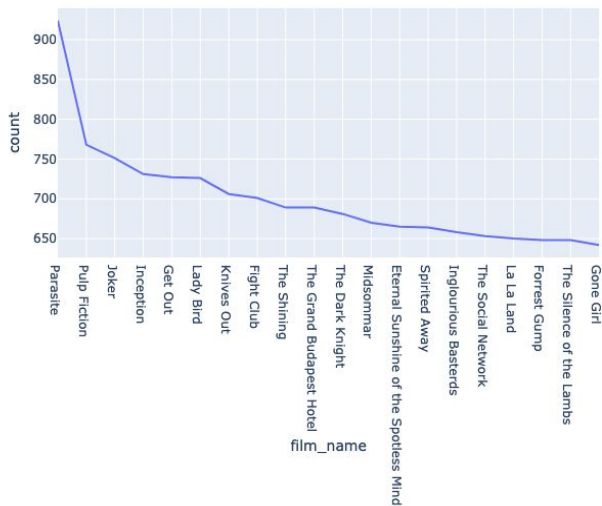
Average Rating Per User



Average Rating per User

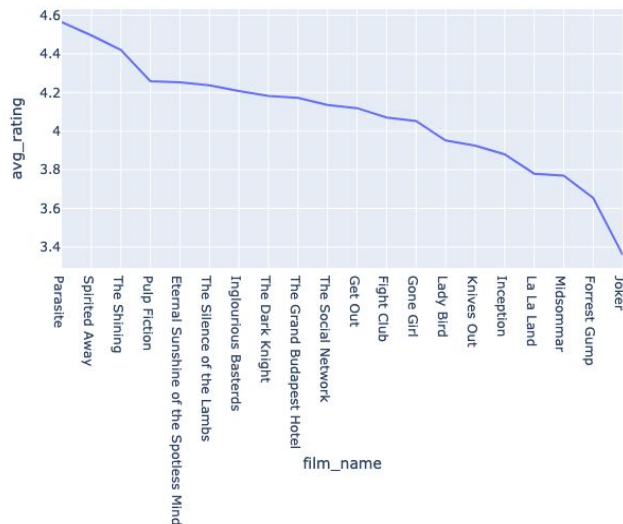
Non-Personalized Recommendation System

20 Most Rated Films



This recommends the 20 most rated films.

20 Most Rated Films Average Rating



This recommends the 20 most rated films by rating

Non-Personalized Recommendation System (cont.)

	film_name	avg_rating	count
0	Parasite	4.564453	924
13	Spirited Away	4.495781	664
8	The Shining	4.420570	689
1	Pulp Fiction	4.258126	768
12	Eternal Sunshine of the Spotless Mind	4.252809	665
18	The Silence of the Lambs	4.237589	648
14	Inglourious Basterds	4.207343	658
10	The Dark Knight	4.182403	681
9	The Grand Budapest Hotel	4.172584	689
15	The Social Network	4.135135	653
4	Get Out	4.119091	727
7	Fight Club	4.070513	701
19	Gone Girl	4.052350	642
5	Lady Bird	3.951585	726
6	Knives Out	3.925386	706
3	Inception	3.879798	731
16	La La Land	3.779959	650
11	Midsommar	3.770147	670
17	Forrest Gump	3.654255	648
2	Joker	3.359143	751

Collaborative Recommendation System

Test user: dougbones with 201 films logged and 22 ratings

- SVD system RMSE: 0.7041

Recommendations for dougbones :

- # 1 : Come and See
- # 2 : Parasite
- # 3 : 12 Angry Men
- # 4 : The Godfather
- # 5 : Portrait of a Lady on Fire

- ALS with Pyspark RMSE: 0.7057

Recommendations for dougbones:

- # 1: Parasite
- # 2: Come and See
- # 3: 12 Angry Men
- # 4: La Haine
- # 5: Portrait of a Lady on Fire

Collaborative Recommendation System

Test user2: creepergnome with 2,901 films logged and 2,234 ratings

- SVD system RMSE: 0.7041

Recommendations for creepergnome :

- # 1 : Come and See
- # 2 : Stalker
- # 3 : The Godfather: Part II
- # 4 : The Godfather
- # 5 : Persona

- ALS with Pyspark RMSE: 0.7057

Recommendations for creepergnome:

- # 1: Come and See
- # 2: Modern Times
- # 3: Sunset Boulevard
- # 4: Do the Right Thing
- # 5: The Godfather

Conclusion/Recommendations

- For new users or user that choose not to log or rate movies, the Non-Personal Recommendation system can be used
- For users that have some movies rated, either the SVD or ALS model can be used. I would lean more to the SVD model due to being more time efficient.

Future Work:

- Expand the number of users used in the models
- Use tmdb api to add genre, film length, and other information to film_info table
- Create recommendations on additional information in film_info table
- Create a front end so anyone can input a letterboxd username and receive movie recommendations.

Any Questions?

Thank you

Please check out the github page for more information.

https://github.com/cristinamtown/flatiron_capstone