MovieLens Recommendation System

What is Recommendation System?

- Makes recommendations to users based on their user information
- Benefits company and users
- Two types: Content Filtering & Collaborative Filtering

Movielens Dataset

- Public dataset used for education purposes
- Subset of the full dataset

movies.csv
9,000 movies

	movieId	title	genres
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
1	2	Jumanji (1995)	Adventure|Children|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
4	5	Father of the Bride Part II (1995)	Comedy

ratings.csv
100,000 ratings, 600 users

	userId	movieId	rating	timestamp
0	1	1	4.0	964982703
1	1	3	4.0	964981247
2	1	6	4.0	964982224
3	1	47	5.0	964983815
4	1	50	5.0	964982931

Analysis of Data

Scatter Plot of Rating by Number of Ratings
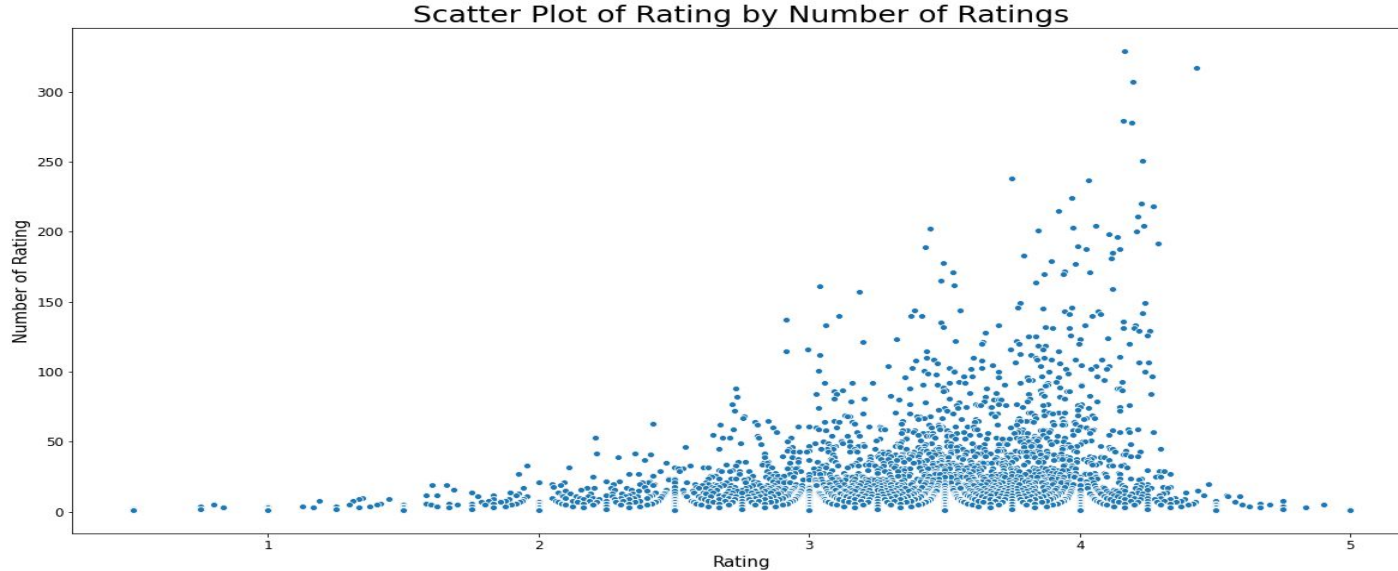
- Ratings of 1 and 5 have low number of ratings
- Average rating = 3.5, Std = 1.0

Total Count of each Rating

Movie Count per Genre
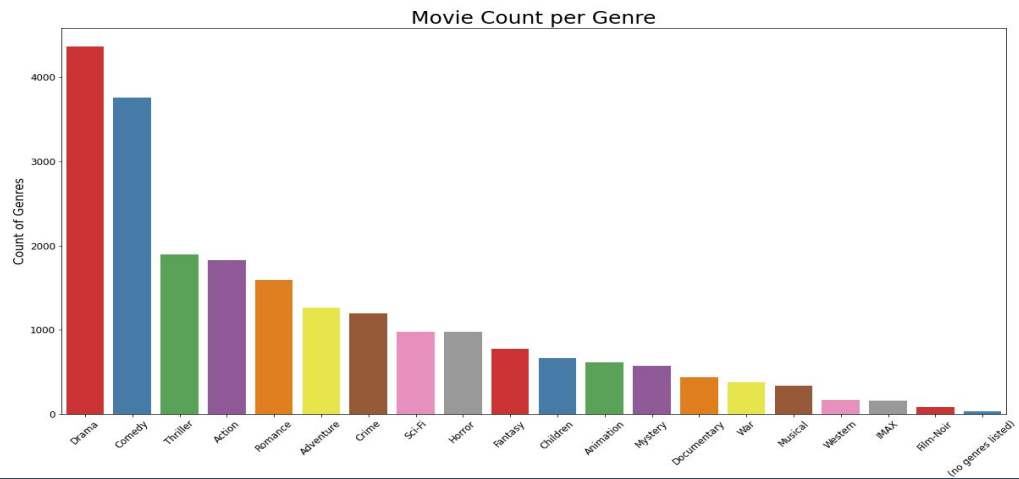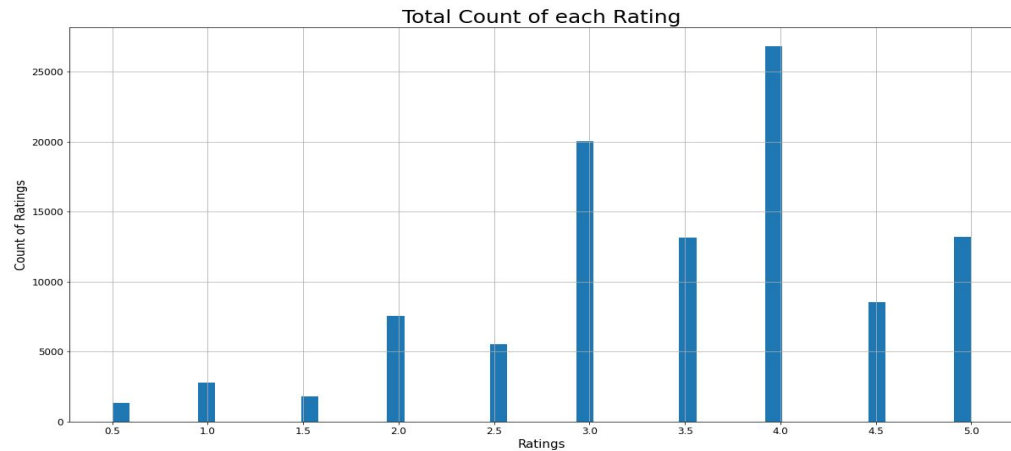
Content Based Recommender System

- New user added to ratings data
- Find similar movies with the same movie genres
- Output movies with the highest weighted ratings for user

New User

	title	rating
0	Jumanji (1995)	3.0
1	No Game No Life: Zero (2017)	4.5
2	Grumpier Old Men (1995)	4.5

Recommended Movies

	movieId	title	weighted_ratings
0	84637	Gnomeo & Juliet (2011)	1.000000
1	56152	Enchanted (2007)	1.000000
2	4306	Shrek (2001)	1.000000
3	92348	Puss in Boots (Nagagutsu o haita neko) (1969)	1.000000
4	1064	Aladdin and the King of Thieves (1996)	0.904762

Collaborative Filtering (Best Model)

Algorithm	test_rmse	fit_time	test_time
SVDpp	0.870703	431.107839	16.553896
BaselineOnly	0.876901	0.213474	0.267595
SVD	0.879918	5.074416	0.378455
KNNBaseline	0.882909	0.336863	3.794513
KNNWithZScore	0.905582	0.187470	3.363470
KNNWithMeans	0.905808	0.132291	3.089450
SlopeOne	0.912458	4.904305	12.645922
NMF	0.935876	5.739594	0.343495
CoClustering	0.955678	2.377111	0.335019
KNNBasic	0.956891	0.110424	2.879292
NormalPredictor	1.424418	0.175244	0.493909

Best Model (RMSE): BaselineOnly

Collaborative Based Recommender System (BaselineOnly)

- New user added to ratings data
- Neighborhood-based filtering
- Output movies with the highest similarities

New User

	userId	movieId	rating
0	1000	133712	5
1	1000	3773	4
2	1000	60904	5
3	1000	2385	2
4	1000	114818	1

Recommended Movies

```
Recommendation #  1 :    277      Shawshank Redemption, The (1994)
Name: title, dtype: object

Recommendation #  2 :    906      Lawrence of Arabia (1962)
Name: title, dtype: object

Recommendation #  3 :    602      Dr. Strangelove or: How I Learned to Stop Worr...
Name: title, dtype: object

Recommendation #  4 :   9618       Three Billboards Outside Ebbing, Missouri (2017)
Name: title, dtype: object

Recommendation #  5 :    841      Streetcar Named Desire, A (1951)
Name: title, dtype: object
```

Future Work

Full Dataset
- Full dataset contains more movies and ratings
- Improve model results

Include Recent Data
- Include recent movies and ratings
- MovieLens last updated 9/2018

Thank you!