Building a Recommendation System with the MovieLens Dataset
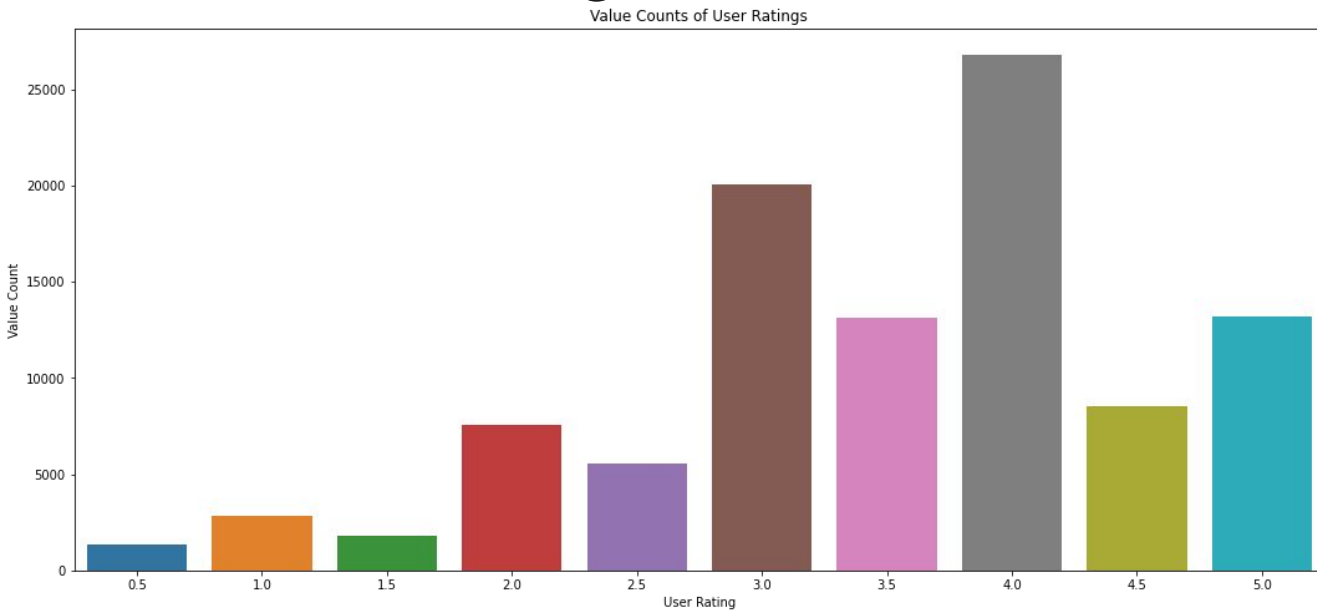
Jonah Flateman
August 12, 2021
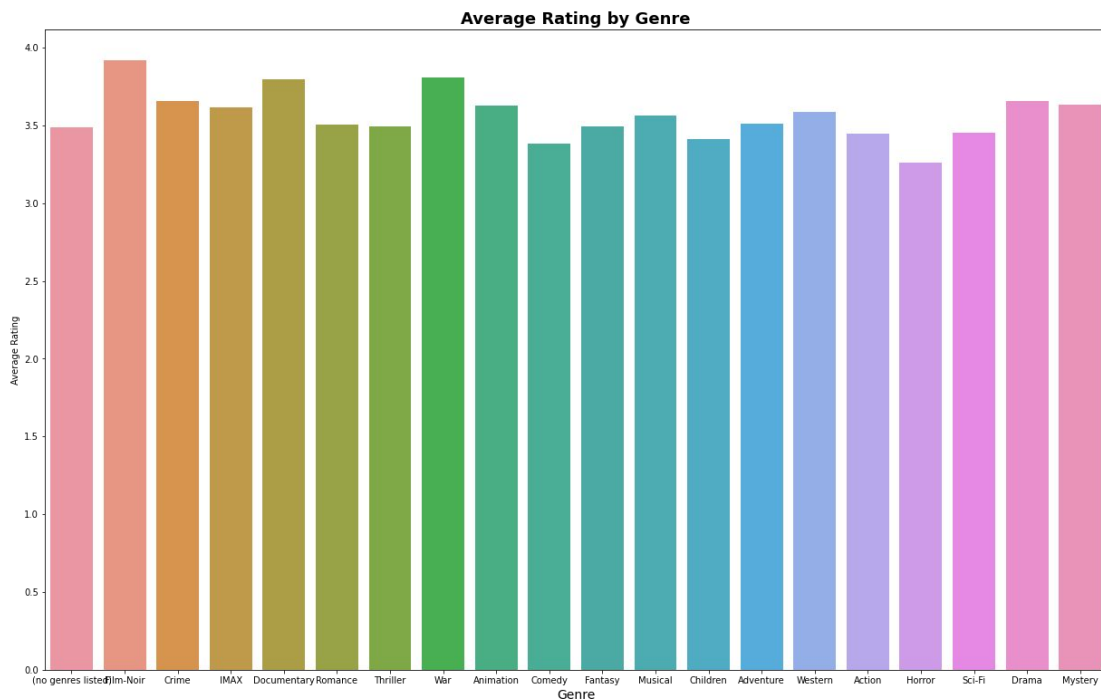
Understanding the data

Value Counts of User Ratings

- 100K total
- Most 3 - 5
- User-focused

Understanding the data

Average Rating by Genre

- Noticing trends (Film-Noir, Documentary, War, Horror)
- Averages between 3.5 - 4.0

Modeling through Collaborative Filtering

- Explicit ratings from user - input process
- Similarity to other users
- Familiar system

Methodologies

- Alternating Least Squares modeling - finding best parameters
- User IDs, movie IDs, movie titles
- Root Mean Square Error (RMSE) as key metric

How did our predictions do?

```
+------+-------+------+----------+
|userId|movieId|rating|prediction|
+------+-------+------+----------+
|   133|    471|   4.0| 2.7365522|
|    91|    471|   1.0| 3.0695984|
|   555|    471|   3.0| 3.5713944|
|   469|    471|   5.0| 3.4271655|
|   104|    471|   4.5| 3.3484874|
|   599|    833|   1.5| 1.3743167|
|   307|    833|   1.0|  1.295717|
|   606|   1088|   3.0| 3.4065895|
|    20|   1088|   4.5|   3.50422|
|   169|   1088|   4.5|  4.082981|
|    41|   1088|   1.5| 2.8915646|
|   286|   1088|   3.5| 3.4070752|
|   583|   1088|   3.5| 3.6121225|
|   391|   1088|   1.0|  2.953916|
|   188|   1088|   4.0| 3.6435459|
|    42|   1088|   3.0| 3.4566991|
|   599|   1342|   2.5| 2.2686872|
|    19|   1342|   2.0| 2.3561733|
|   307|   1342|   2.0|  2.381162|
|   322|   1580|   3.5| 3.1221216|
+------+-------+------+----------+
only showing top 20 rows
```

● RMSE of 0.8649

User Interface In Action

```
Field of Dreams (1989)
How do you rate this movie on a scale of 1-5, press n if you have not seen :
4
Edward Scissorhands (1990)
How do you rate this movie on a scale of 1-5, press n if you have not seen :
5
Atlantis: The Lost Empire (2001)
How do you rate this movie on a scale of 1-5, press n if you have not seen :
4
Prestige, The (2006)
How do you rate this movie on a scale of 1-5, press n if you have not seen :
4
Felon (2008)
How do you rate this movie on a scale of 1-5, press n if you have not seen :
n
Lawnmower Man 2: Beyond Cyberspace (1996)
How do you rate this movie on a scale of 1-5, press n if you have not seen :
n
Hot Shots! Part Deux (1993)
How do you rate this movie on a scale of 1-5, press n if you have not seen :
2
Recommendation 1: On the Beach (1959) | predicted score: 5.213873386383057
Recommendation 2: Saving Face (2004) | predicted score: 5.1715898513793945
Recommendation 3: Come and See (Idi i smotri) (1985) | predicted score: 5.0117950439453125
Recommendation 4: Jetée, La (1962) | predicted score: 5.004154205322266
Recommendation 5: Cosmos | predicted score: 4.967296123504639
Recommendation 6: Lady Jane (1986) | predicted score: 4.946325778961182
Recommendation 7: Crossing Delancey (1988) | predicted score: 4.946325778961182
Recommendation 8: Deathgasm (2015) | predicted score: 4.904405117034912
Recommendation 9: Watermark (2014) | predicted score: 4.888047695159912
Recommendation 10: Woman Under the Influence, A (1974) | predicted score: 4.888047695159912
```

Future Work

- Larger size datasets from MovieLens
- Genre-specific results
- Binary yes/no recommendations

Thank you

jonahflateman@gmail.com