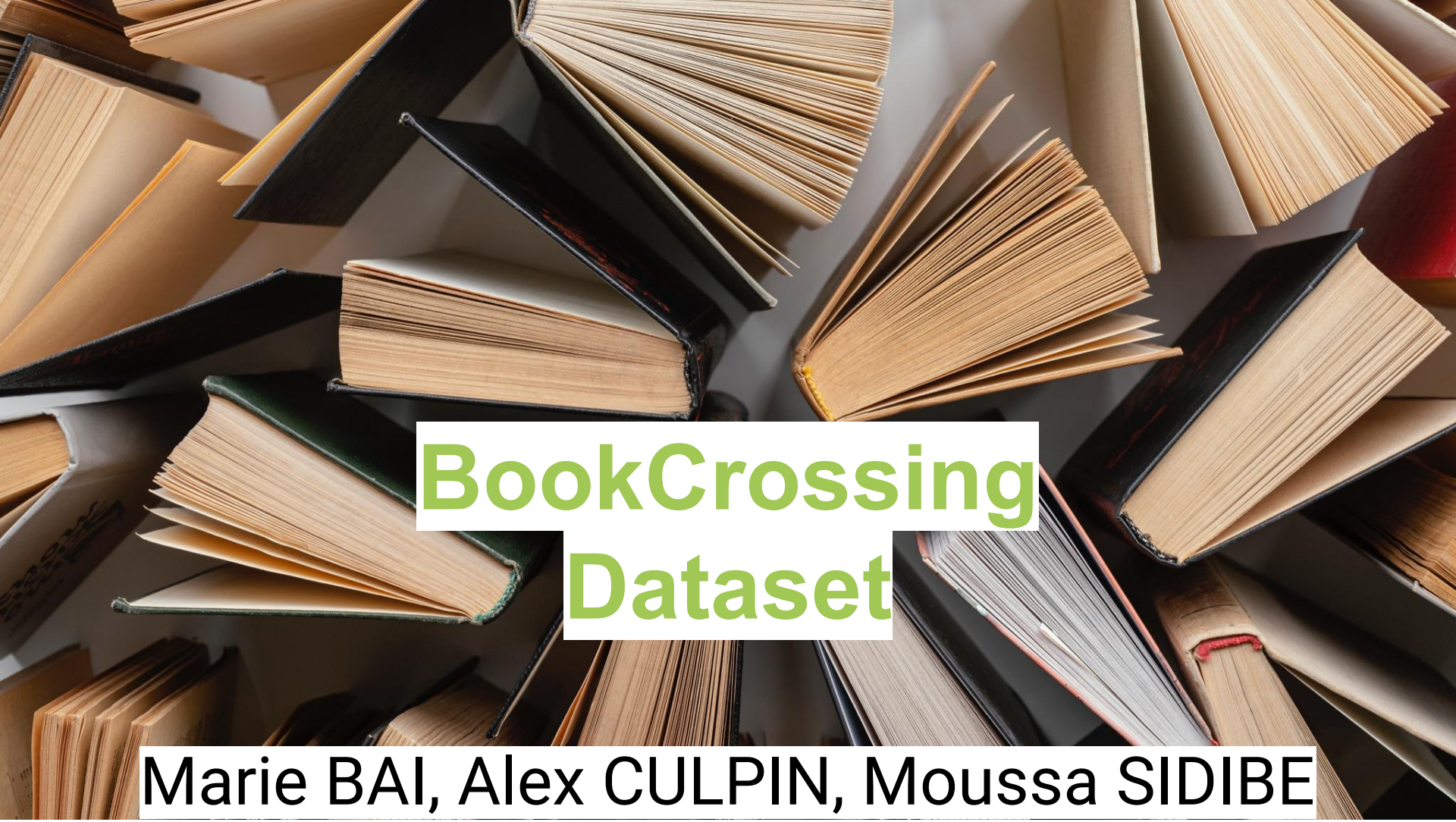


BookCrossing Dataset

Marie BAI, Alex CULPIN, Moussa SIDIBE

9 project

LEARNING PLANET INSTITUTE

AGENDA

1 EDA,
Feature Engineering

2 Recommendation task

3 Recommender system

EDA

```
1 df_users.shape, df_ratings.shape, df_books.shape
```

```
((278858, 3), (1149780, 3), (271360, 8))
```

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...

	User-ID	Location	Age
0	1	nyc, new york, usa	NaN
1	2	stockton, california, usa	18.0
2	3	moscow, yukon territory, russia	NaN
3	4	porto, v.n.gaia, portugal	17.0
4	5	farnborough, hants, united kingdom	NaN

	User-ID	ISBN	Book-Rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6

LEARNING PLANET INSTITUTE

EDA

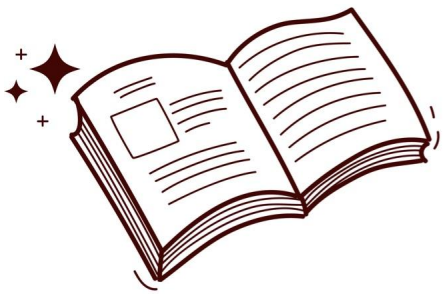
	User-ID	Location	Age
0	1	nyc, new york, usa	NaN
1	2	stockton, california, usa	18.0
2	3	moscow, yukon territory, russia	NaN
3	4	porto, v.n.gaia, portugal	17.0
4	5	farnborough, hants, united kingdom	NaN

Missing values

User-ID	0
Location	0
Age	110762

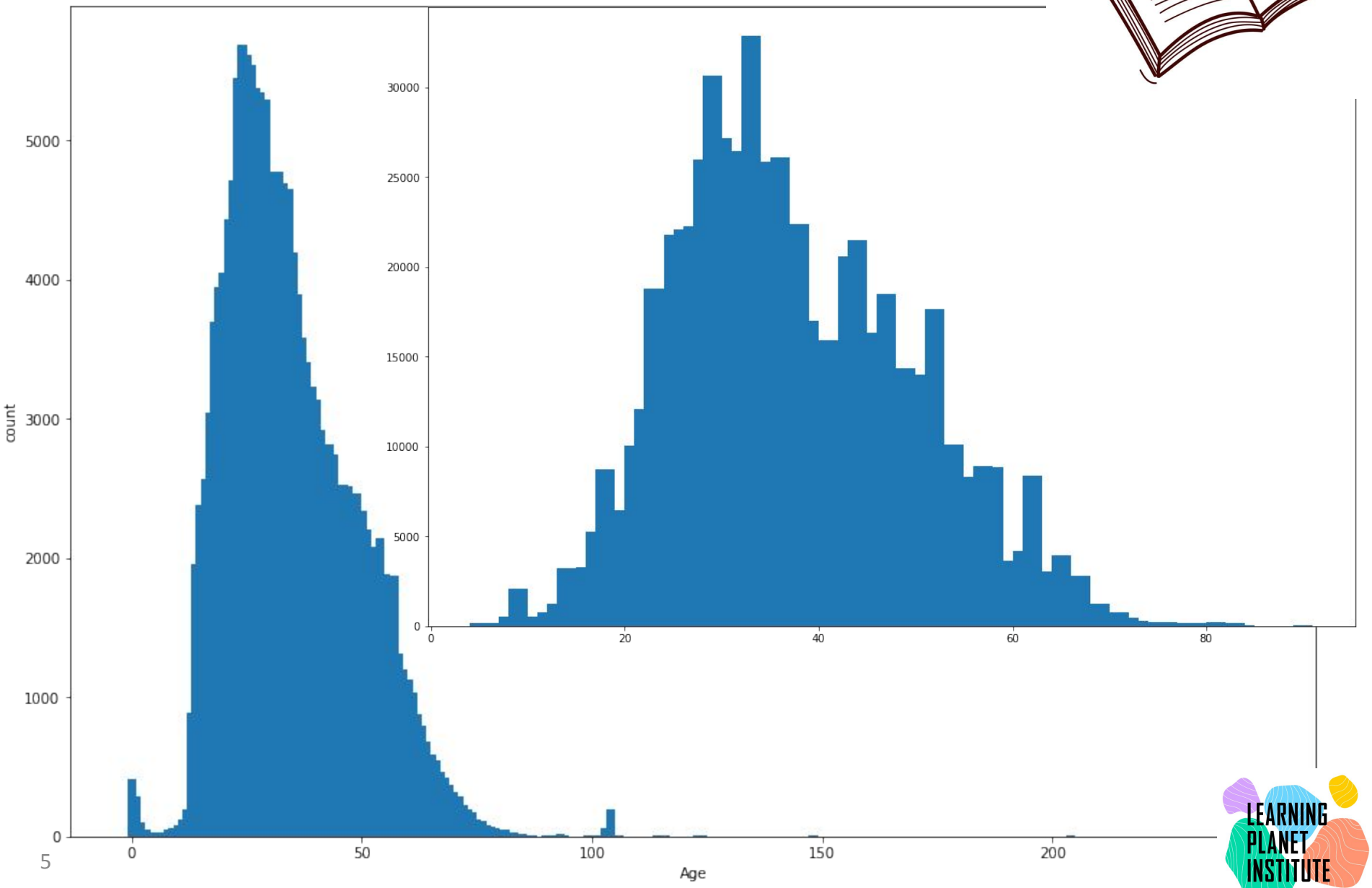
	User-ID	Age
count	168096.000000	168096.000000
mean	139394.611865	34.751434
std	80561.659564	14.428097
min	2.000000	0.000000
25%	69914.750000	24.000000
50%	139363.500000	32.000000
75%	209162.500000	44.000000
max	278855.000000	244.000000

Age from 5 to 90



LEARNING PLANET INSTITUTE

EDA

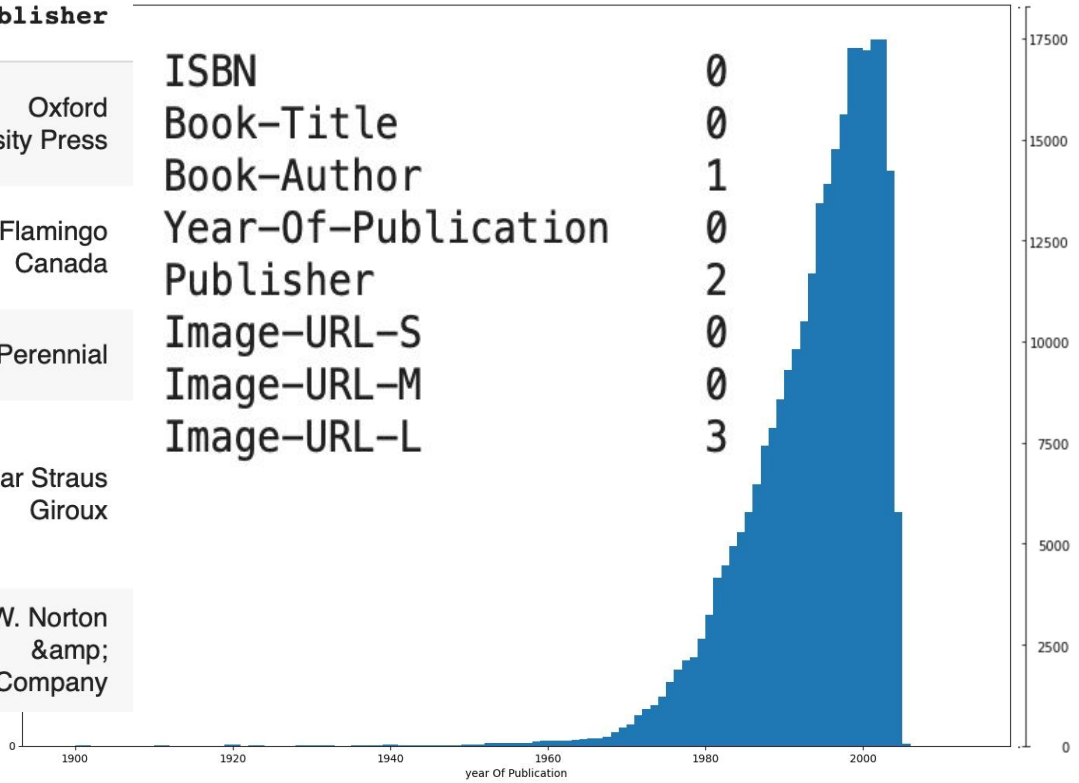


LEARNING PLANET INSTITUTE

EDA

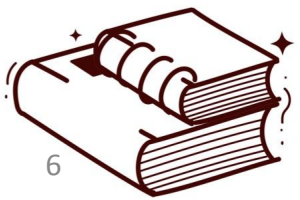
	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company

```
ISBN                  0
Book-Title            0
Book-Author           1
Year-Of-Publication   0
Publisher             2
Image-URL-S           0
Image-URL-M           0
Image-URL-L           3
```




```
Agatha Christie          632
William Shakespeare      567
Stephen King             524
Ann M. Martin            423
Carolyn Keene            373
```

Top - 5 authors



LEARNING PLANET INSTITUTE

Feature Engineering

	User-ID	ISBN	Book-Rating	Location	Age	Book-Title	Book-Author	Year-Of-Publication	Publisher
0	276727	0446520802	0	h, new south wales, australia	16.0	The Notebook	Nicholas Sparks	1996	Warner Books
1	638	0446520802	0	san diego, california, usa	20.0	The Notebook	Nicholas Sparks	1996	Warner Books
2	3363	0446520802	0	knoxville, tennessee, usa	29.0	The Notebook	Nicholas Sparks	1996	Warner Books
3	7158	0446520802	10	omaha, nebraska, usa	30.0	The Notebook	Nicholas Sparks	1996	Warner Books
4	8253	0446520802	10	tulsa, oklahoma, usa	26.0	The Notebook	Nicholas Sparks	1996	Warner Books

	User-ID	ISBN	Book-Rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6

	User-ID	Location	Age
0	1	nyc, new york, usa	NaN
1	2	stockton, california, usa	18.0
2	3	moscow, yukon territory, russia	NaN
3	4	porto, v.n.gaia, portugal	17.0
4	5	farnborough, hants, united kingdom	NaN

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...

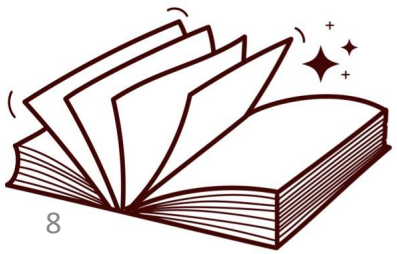
LEARNING PLANET INSTITUTE

Recommendation task

- Simple recommendation based on the popularity of the product

- Item based collaborative filtering recommendation system by using rating user matrix

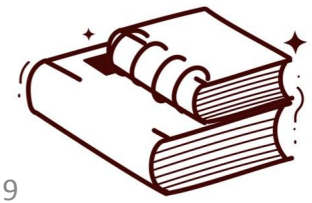
- User based collaborative filtering recommendation system



LEARNING PLANET INSTITUTE

Recommender system
Build a simple rating base recommendation model

- Extract the data
- Defind a new scoring variable
- $(\mathrm{nb}_v ote/(nb_v ote + m) * nb_v ote.mean()) + (m/(m + nb_v ote) * meanRating))sortthedatatogetthebestreco$
- Adapt to new and old user



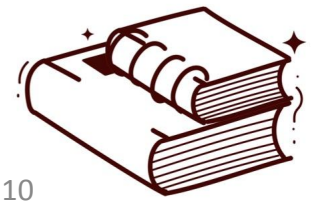
LEARNING PLANET INSTITUTE

Recommender system

Collaborative filtering model Using SVD model

- Using the surprise package
- Merge the data
- Matrix reduction base algorithme
- Prediction formula:

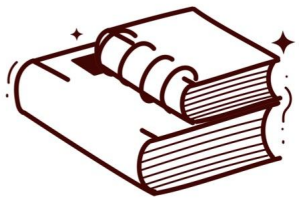
$$_{ui} = \mu + b_u + b_i + q_i^T p_u$$



LEARNING PLANET INSTITUTE

Recommender system
Collaborative filtering model Using KNNWithMeans model

- Using the surprise package
- Merge the data
- Use cos distance
- Use mean square difference (msd)



LEARNING PLANET INSTITUTE

Example

userID = 276768

	title	meanRating	score
671	The Lovely Bones: A Novel	8.185290	11.313177
149	The Da Vinci Code	8.435318	11.307284
1328	Harry Potter and the Sorcerer's Stone (Harry P...	8.939297	11.300696
507	The Red Tent (Bestselling Backlist)	8.182768	11.298335
979	The Secret Life of Bees	8.452769	11.293856

userID = 44842

	title	meanRating	score
149	The Da Vinci Code	8.435318	11.307284
1328	Harry Potter and the Sorcerer's Stone (Harry P...	8.939297	11.300696
507	The Red Tent (Bestselling Backlist)	8.182768	11.298335
979	The Secret Life of Bees	8.452769	11.293856
2636	Divine Secrets of the Ya-Ya Sisterhood: A Novel	7.887500	11.288363

LEARNING PLANET INSTITUTE

Thank you for your attention