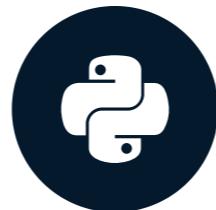


Why learn how to build recommendation engines?

BUILDING RECOMMENDATION ENGINES WITH PYSPARK

Jamen Long

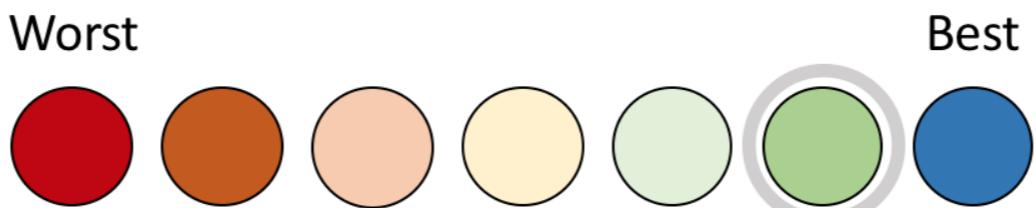
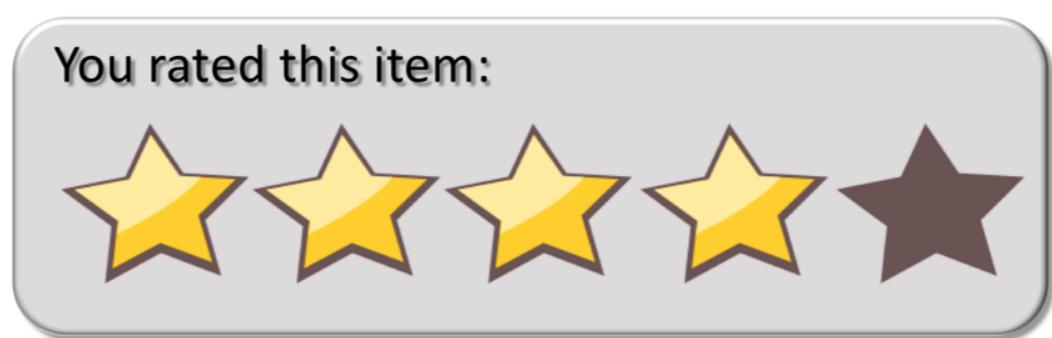
Data Scientist at Nike



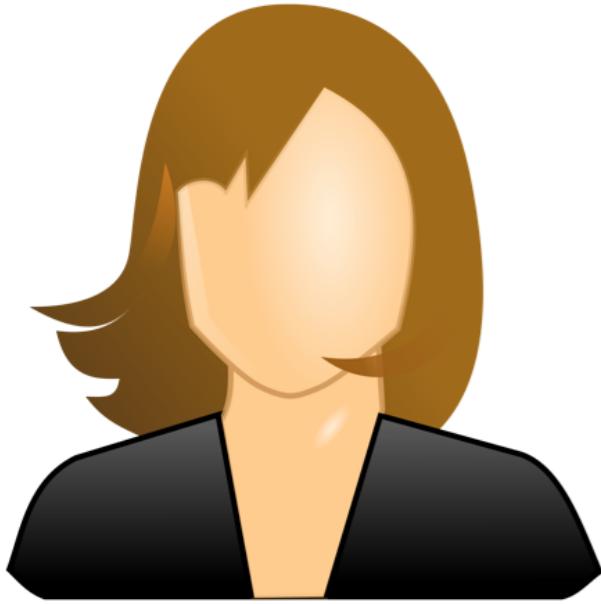
What recommendations look like



Learning about you



How recommendation engines work

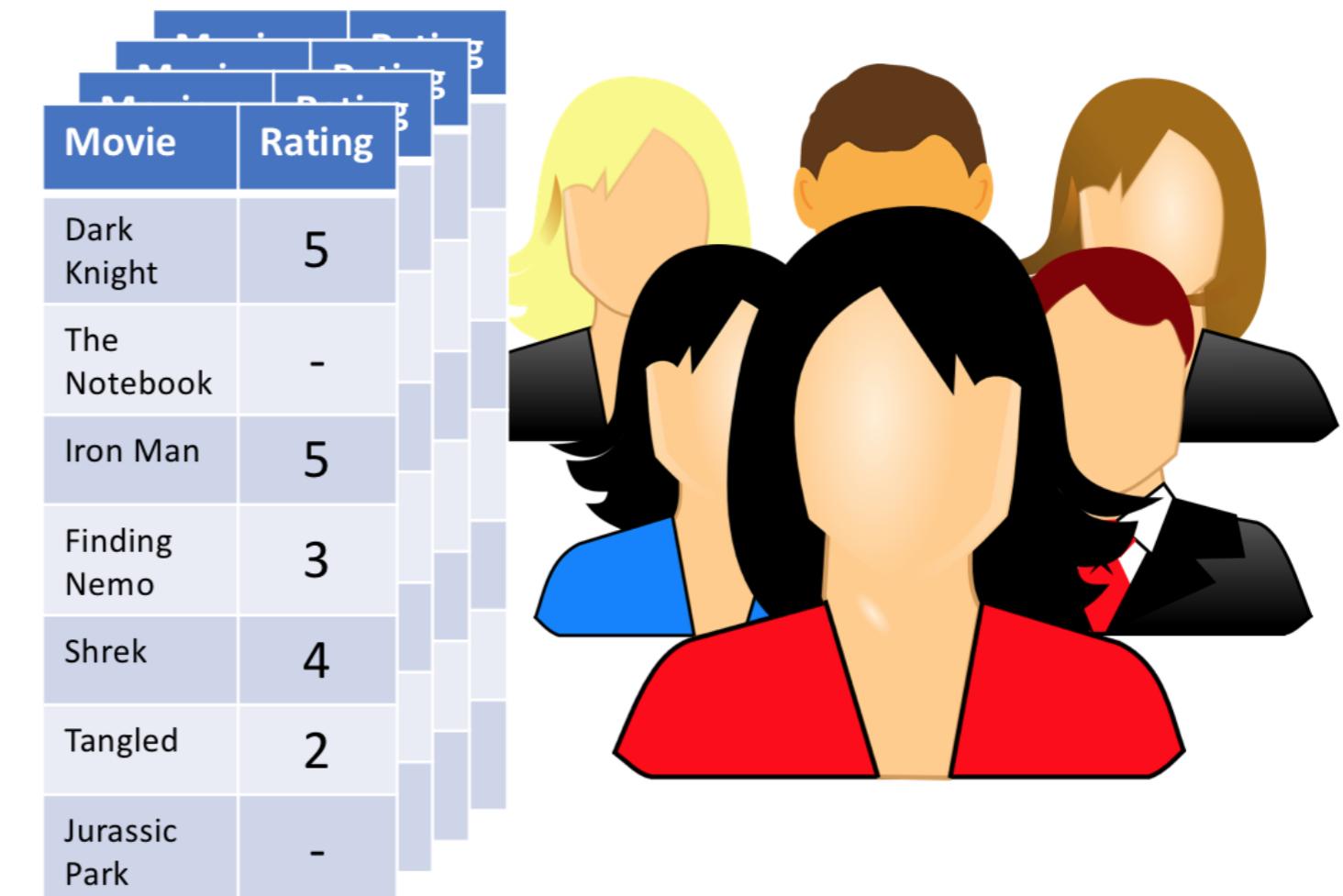


Movie	Rating
Dark Knight	5
The Notebook	-
Iron Man	4
Finding Nemo	3
Shrek	-
Tangled	1
Jurassic Park	4

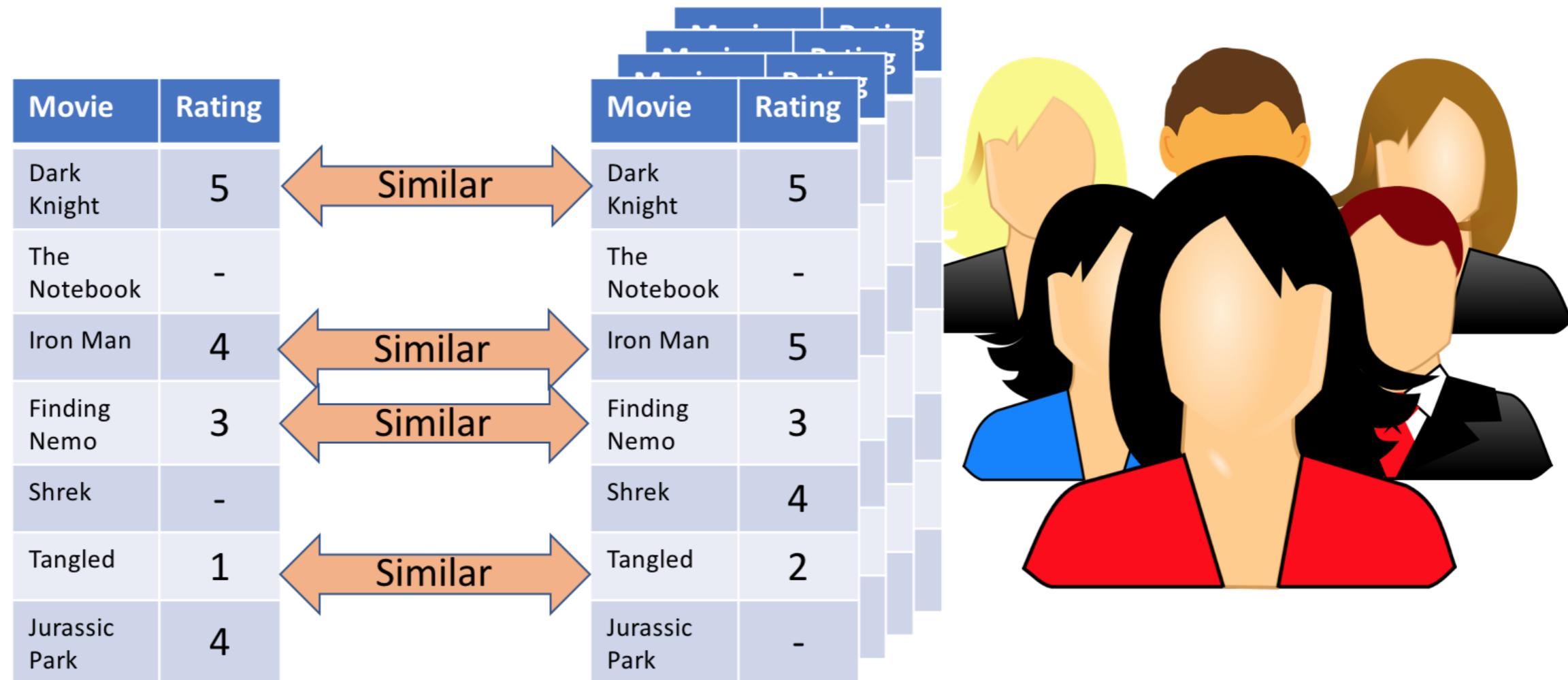
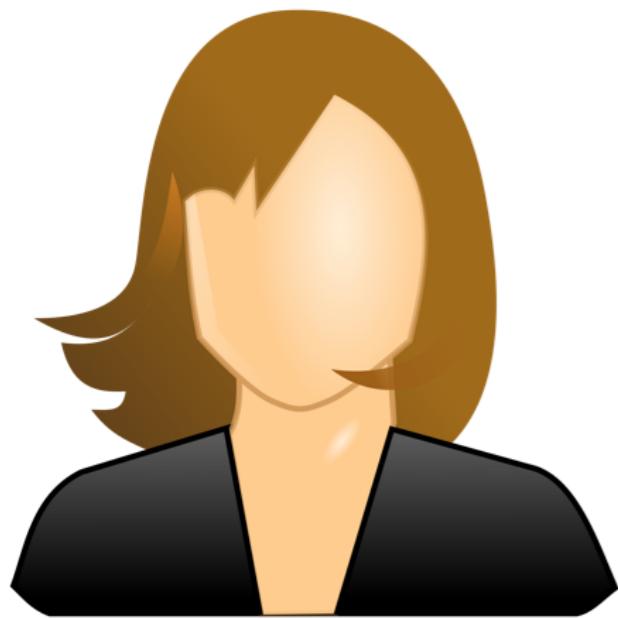
How recommendation engines work



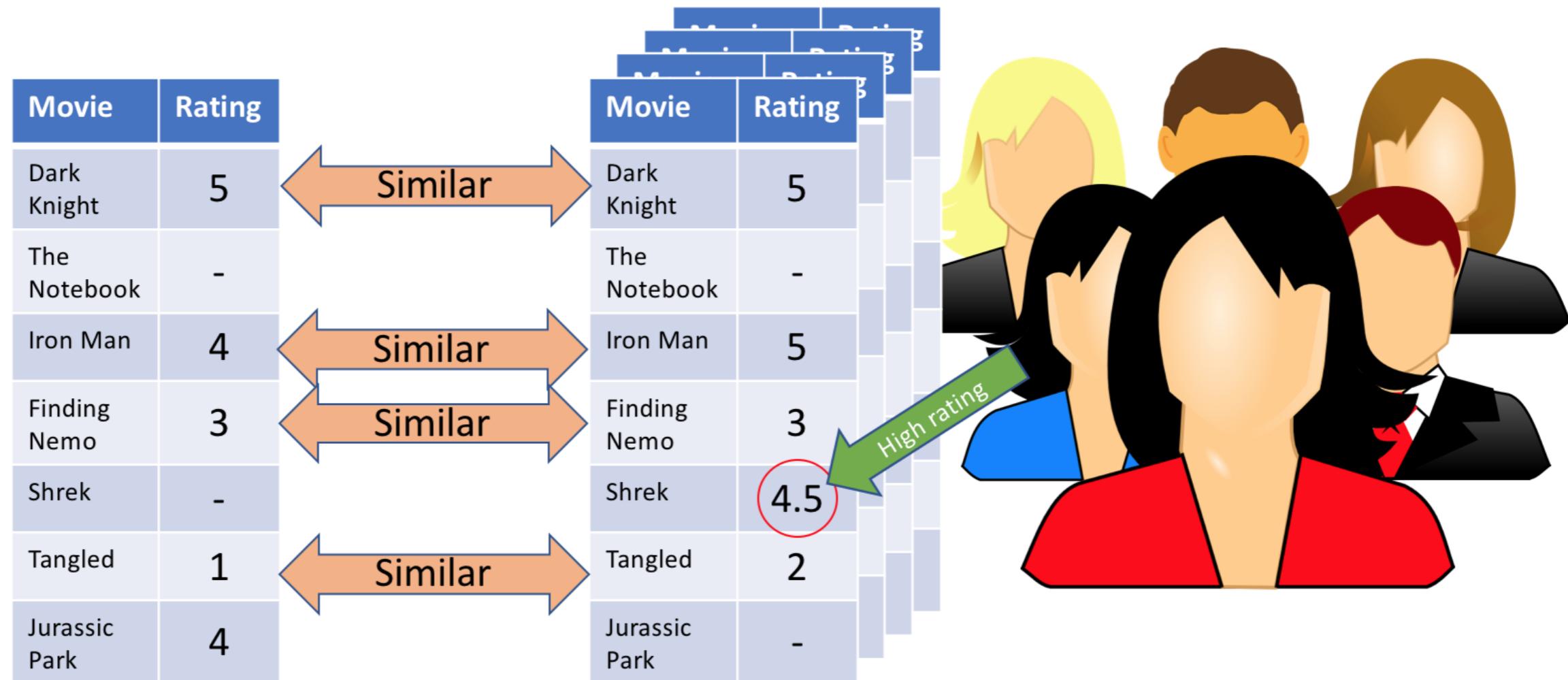
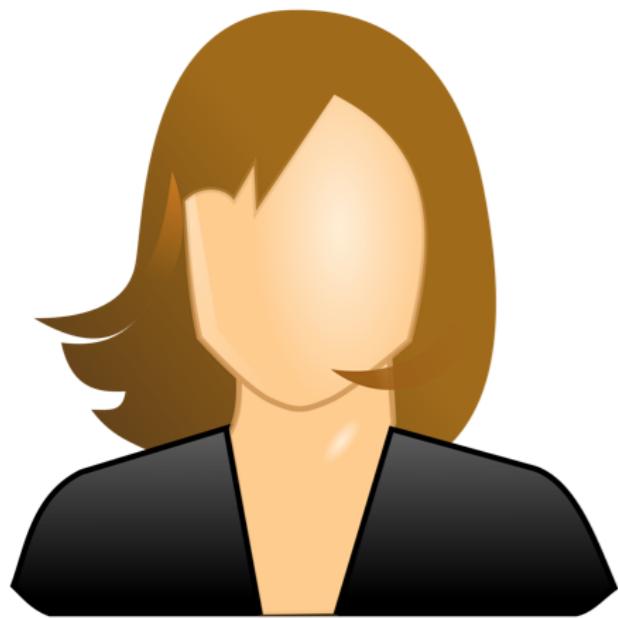
Movie	Rating
Dark Knight	5
The Notebook	-
Iron Man	4
Finding Nemo	3
Shrek	-
Tangled	1
Jurassic Park	4



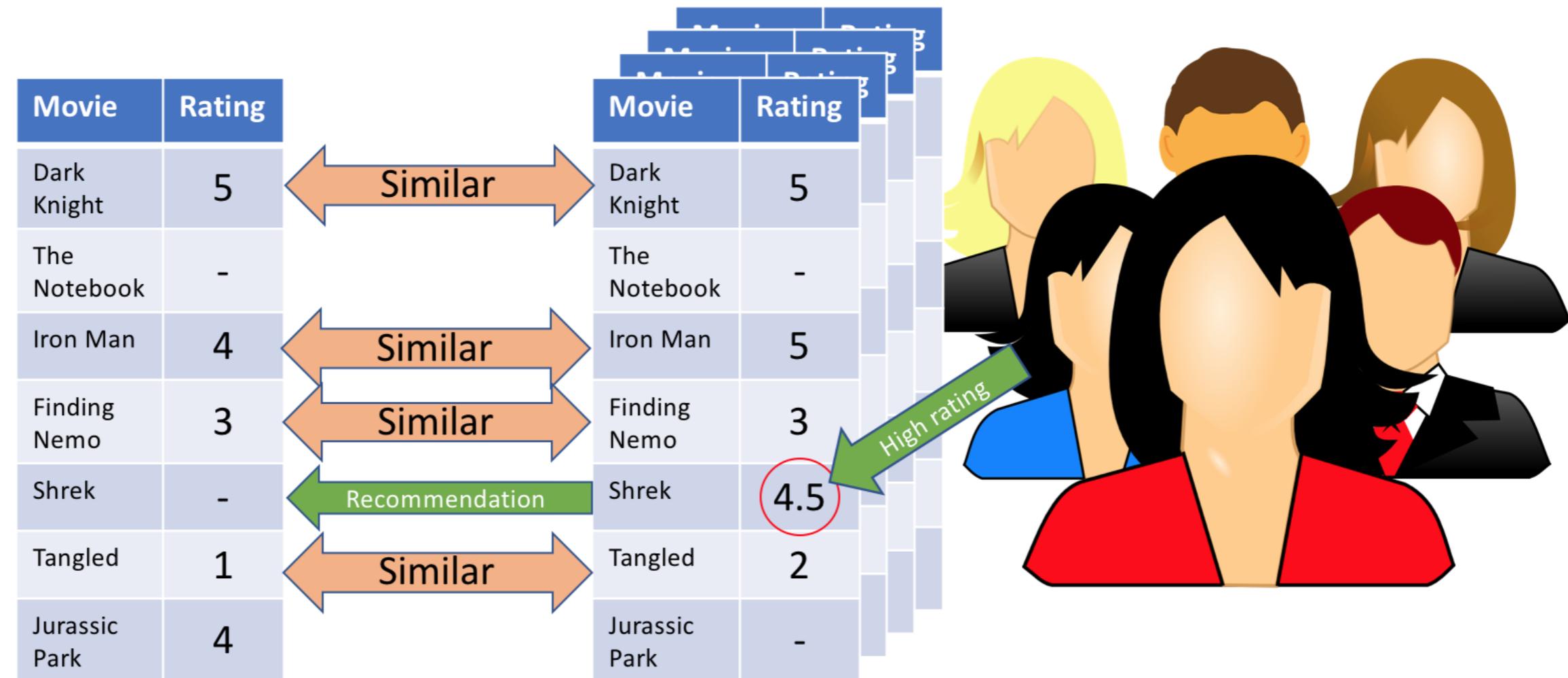
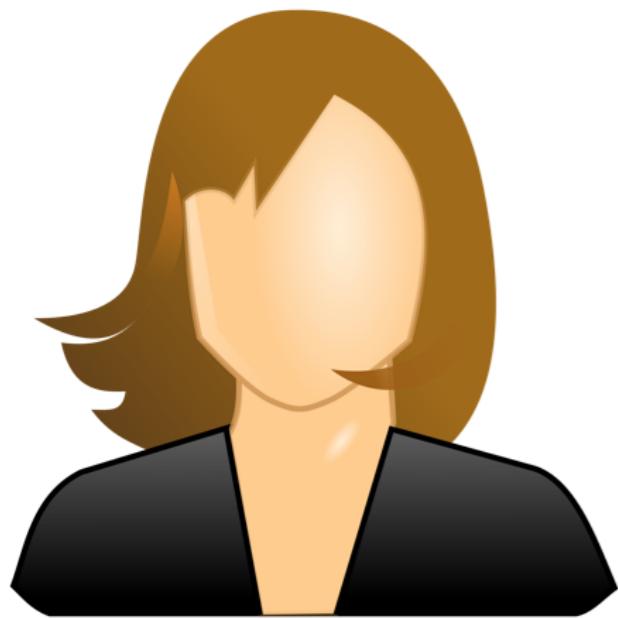
How recommendation engines work



How recommendation engines work



How recommendation engines work



The Power of Recommendation Engines

Powered by increasingly sophisticated models that analyze transaction data and digital signals (for example, what topics are hot on social networks). Already, 35 percent of what consumers purchase on Amazon and 75 percent of what they watch on Netflix come from product recommendations based on such algorithms. Company-directed marketing is also competing for attention through social networks, user-generated content, and other channels. In fact, companies are competing for the average consumer's attention.

Ian Mackenzie, Chris Meyer, and Steve Noble
McKinsey & Company, October 2013

Prerequisites

- [Introduction to PySpark](#)
- [Intermediate Python](#)
- [Supervised Learning with scikit-learn](#)

Let's practice!

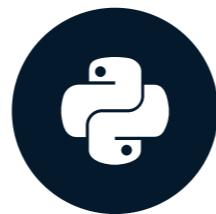
BUILDING RECOMMENDATION ENGINES WITH PYSPARK

Recommendation engine types and data types

BUILDING RECOMMENDATION ENGINES WITH PYSPARK

Jamen Long

Data Scientist at Nike



Two types of recommendation engines:

CONTENT-BASED FILTERING

Based on features of items

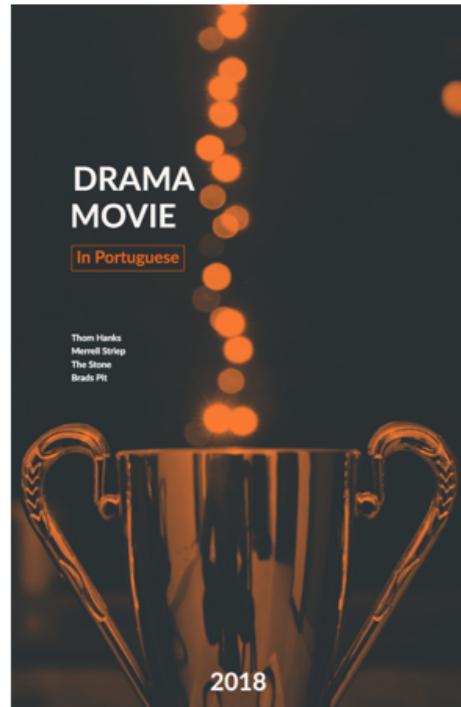
- Genre: Comedy, Action, Drama
- Animation: Animated, Not animated
- Language: English, Spanish, Korean
- Decade Produced: 1950's, 1980's
- Actors: Meryl Streep, Tom Hanks

COLLABORATIVE FILTERING

Two types of recommendation engines

CONTENT-BASED FILTERING

Based on features of items

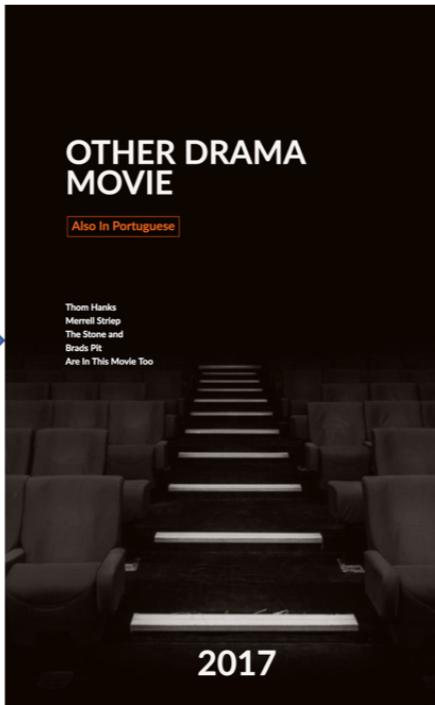
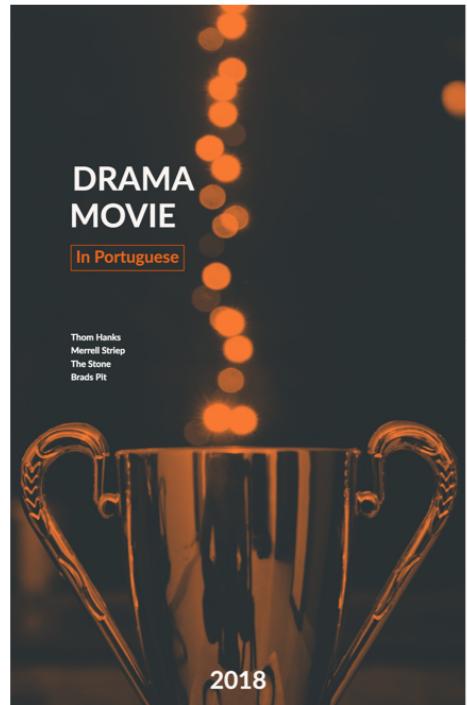


COLLABORATIVE FILTERING

Two types of recommendation engines

CONTENT-BASED FILTERING

Based on features of items

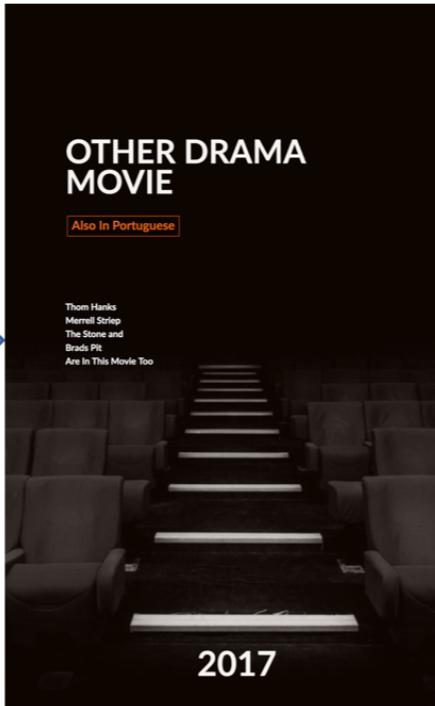
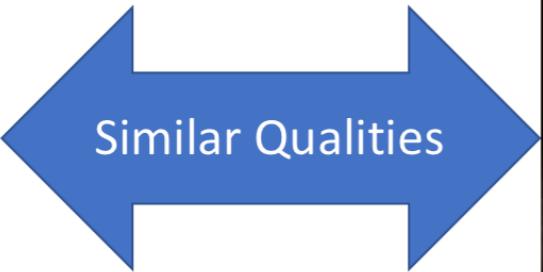
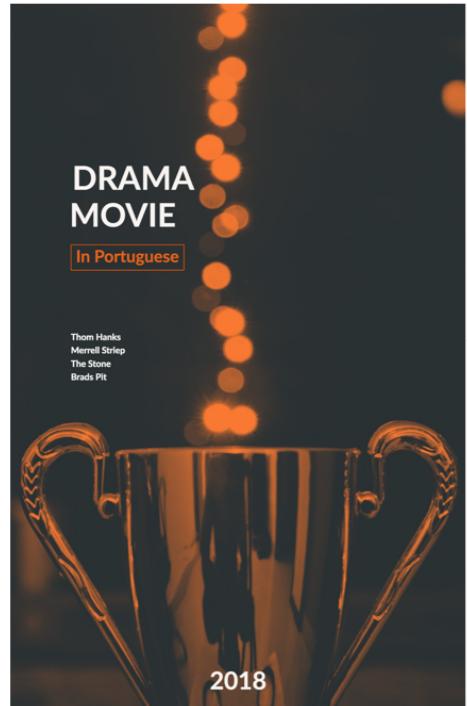


COLLABORATIVE FILTERING

Two types of recommendation engines

CONTENT-BASED FILTERING

Based on features of items



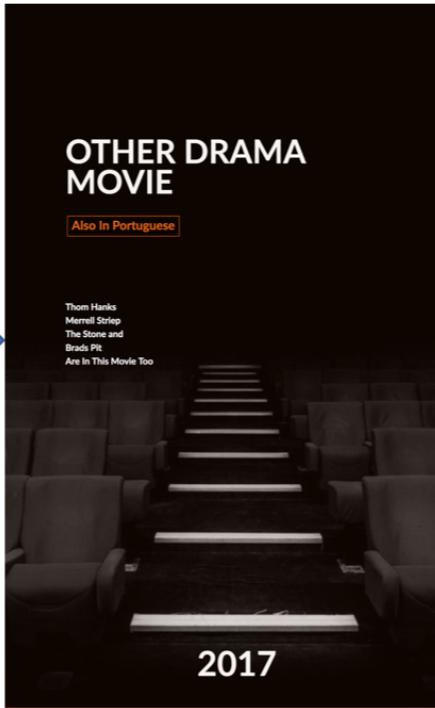
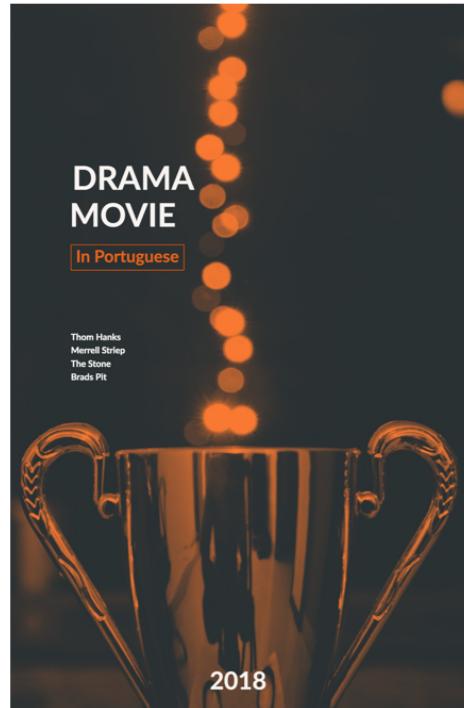
Recommended for you!

COLLABORATIVE FILTERING

Two types of recommendation engines

CONTENT-BASED FILTERING

Based on features of items



Recommended for you!



COLLABORATIVE FILTERING

Based on similar user preferences



Movie	Rating
Dark Knight	5
The Notebook	-
Iron Man	4
Finding Nemo	3
Shrek	-
Tangled	1
Jurassic Park	4

Movie	Rating
Dark Knight	5
The Notebook	-
Iron Man	5
Finding Nemo	3
Shrek	4.5
Tangled	2
Jurassic Park	-



Two types of ratings

Two types of ratings

EXPLICIT RATINGS

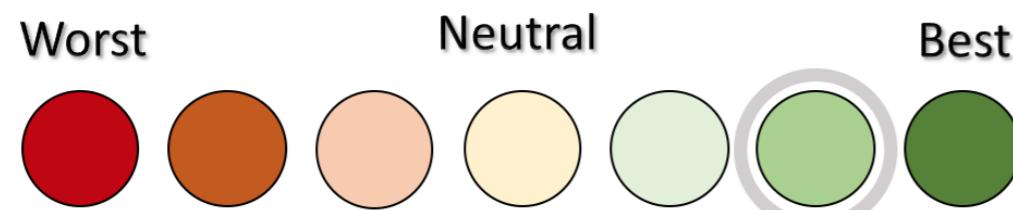
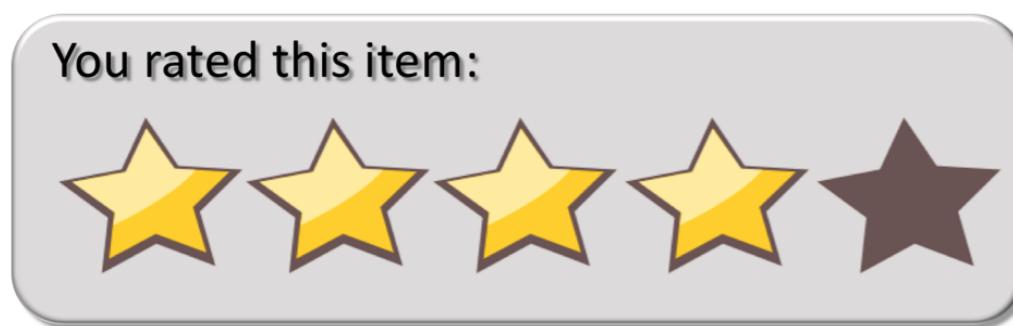
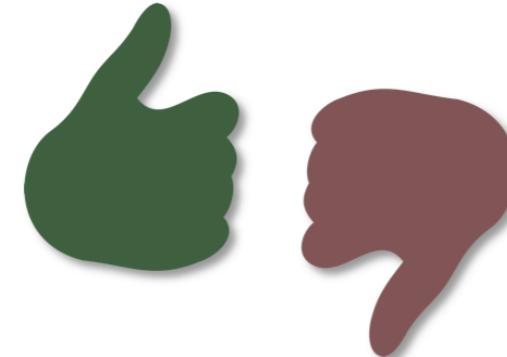
Two types of ratings

EXPLICIT RATINGS

IMPLICIT RATINGS

Two types of ratings

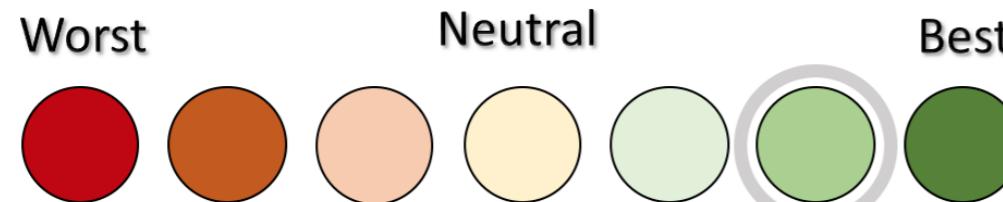
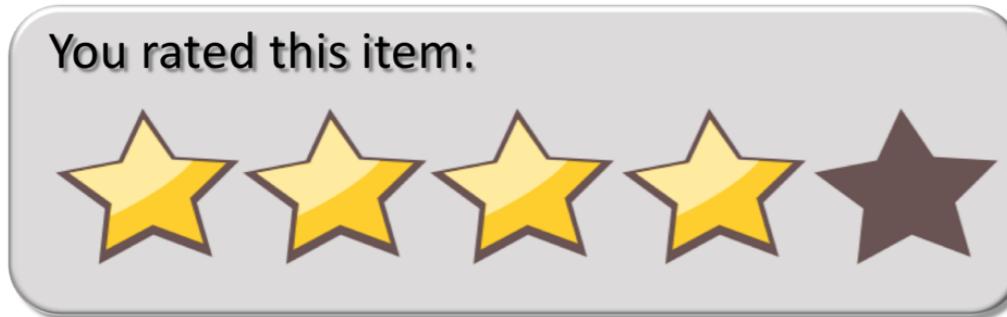
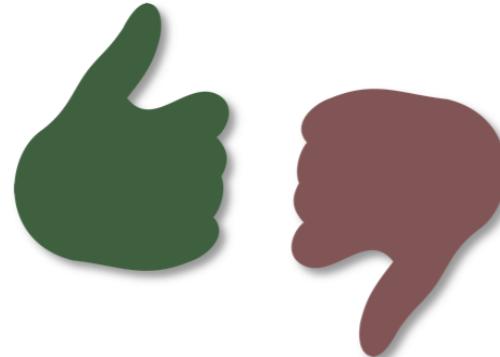
EXPLICIT RATINGS



IMPLICIT RATINGS

Two types of ratings

EXPLICIT RATINGS

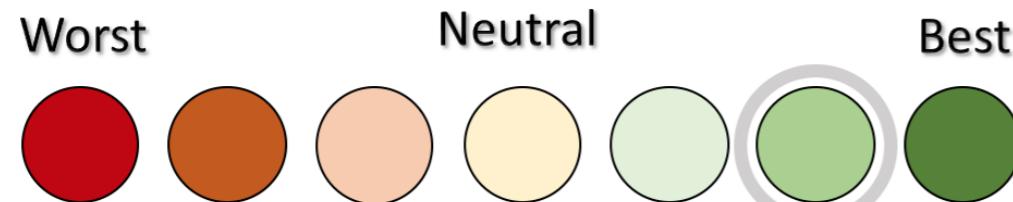
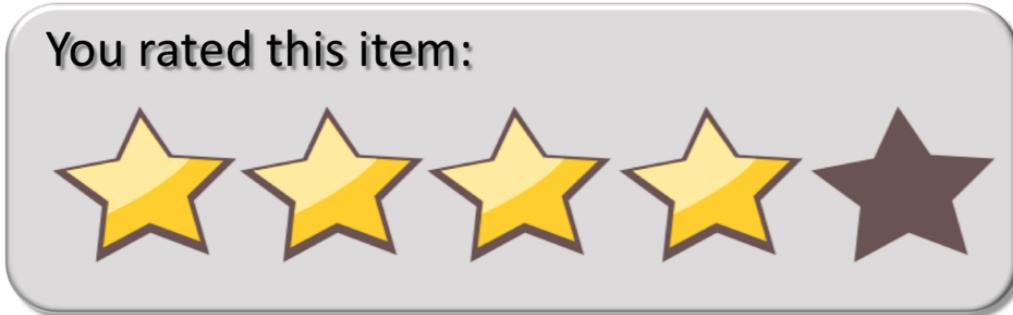


IMPLICIT RATINGS



Two types of ratings

EXPLICIT RATINGS



IMPLICIT RATINGS

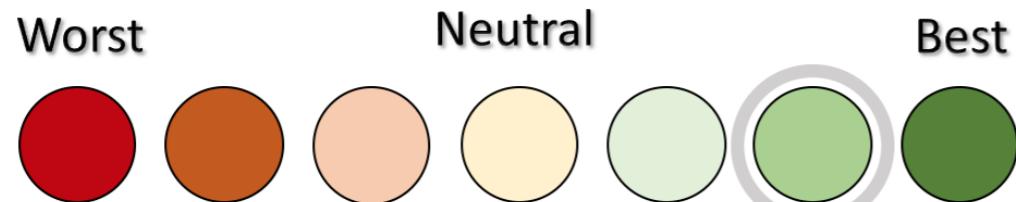
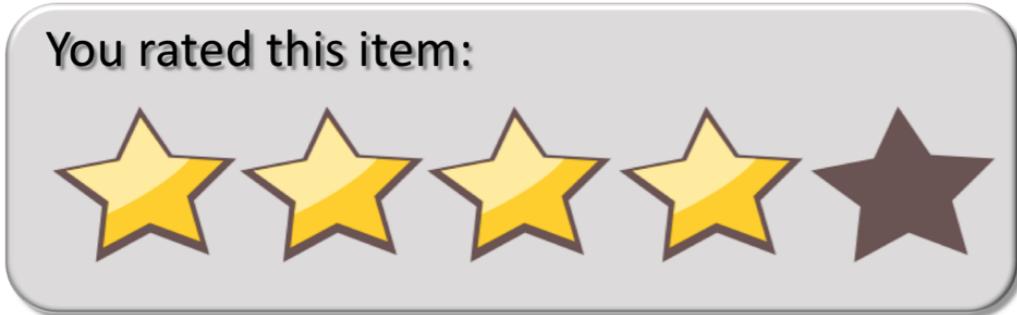
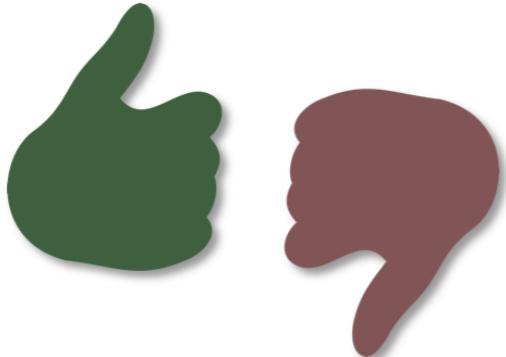


≈ Low Confidence Rating



Two types of ratings

EXPLICIT RATINGS



IMPLICIT RATINGS



≈ Low Confidence Rating



≈ High Confidence Rating

Let's practice!

BUILDING RECOMMENDATION ENGINES WITH PYSPARK

Uses for recommendation engines

BUILDING RECOMMENDATION ENGINES WITH PYSPARK

Jamen Long

Data Scientist at Nike



Original Ratings Matrix

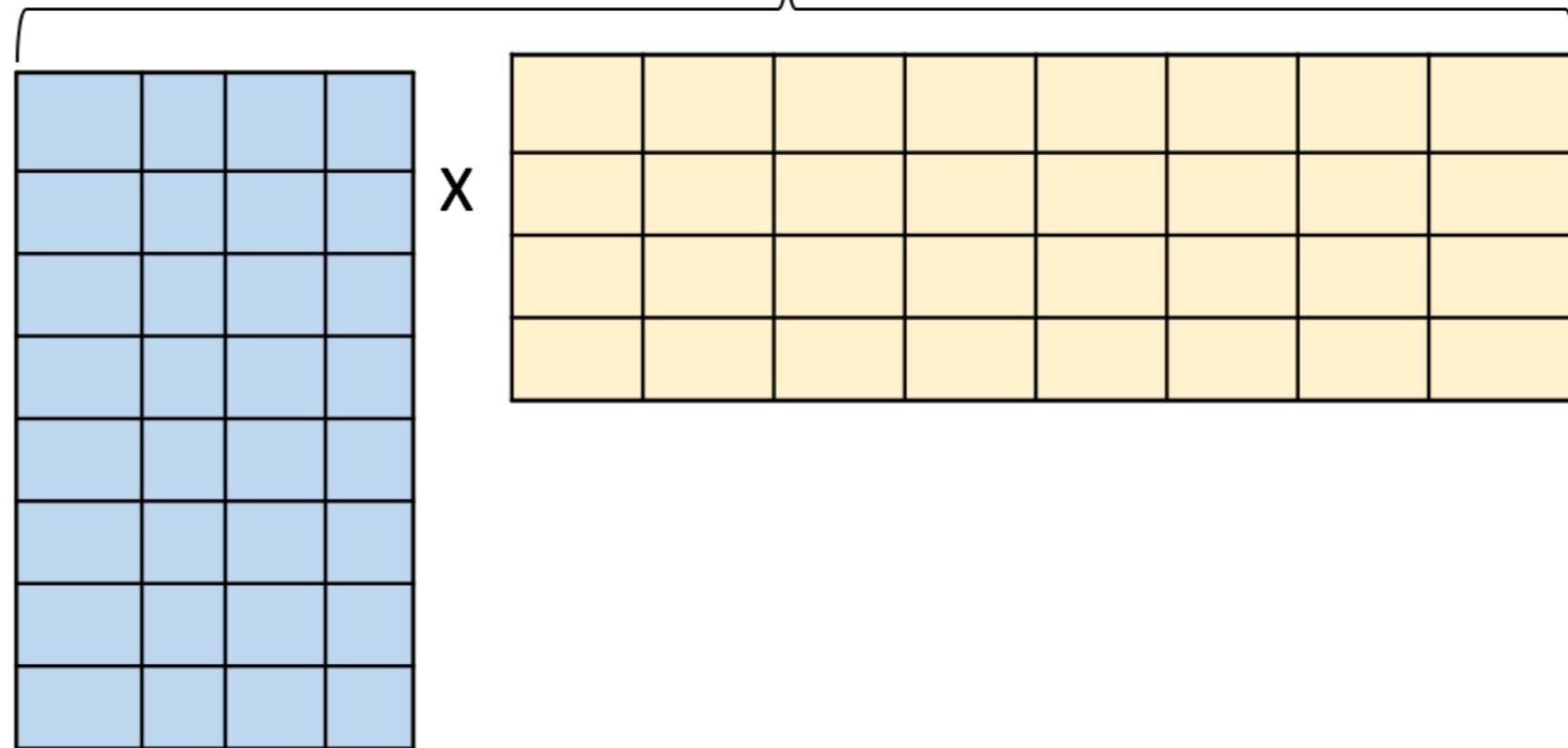
	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie ...	Movie <i>n</i>
User 1							
User 2							
User 3							
User 4							
User 5							
User...							
User <i>m</i>							

Original Ratings Matrix

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie ...	Movie <i>n</i>
User 1							
User 2							
User 3							
User 4							
User 5							
User ...							
User <i>m</i>							

ALS

Factor Matrices



Original Ratings Matrix

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie ...	Movie <i>n</i>
User 1							
User 2							
User 3							
User 4							
User 5							
User...							
User <i>m</i>							

ALS

Factor Matrices

	LF1	LF...	LF <i>k</i>
User 1			
User 2			
User 3			
User 4			
User 5			
User...			
User <i>m</i>			

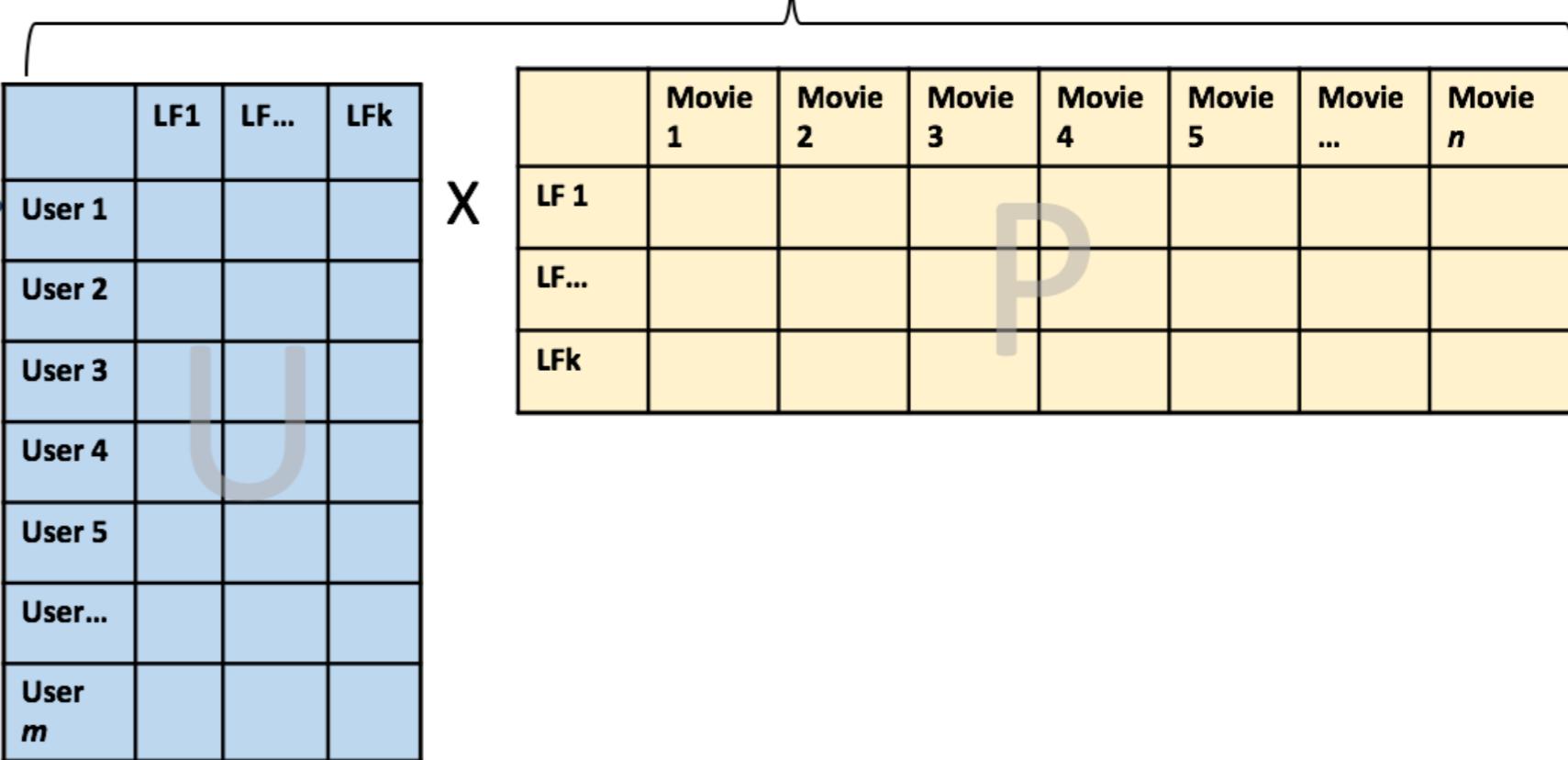
X

Original Ratings Matrix

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie ...	Movie n
User 1							
User 2							
User 3							
User 4							
User 5							
User...							
User m							

ALS

Factor Matrices



Original Ratings Matrix

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie ...	Movie n
User 1							
User 2							
User 3							
User 4							
User 5							
User ...							
User m							

ALS

	LF1	LF...	LFk
User 1			
User 2			
User 3			
User 4			
User 5			
User ...			
User m			

Factor Matrices

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie ...	Movie n
LF 1							
LF...							
LFk							

X

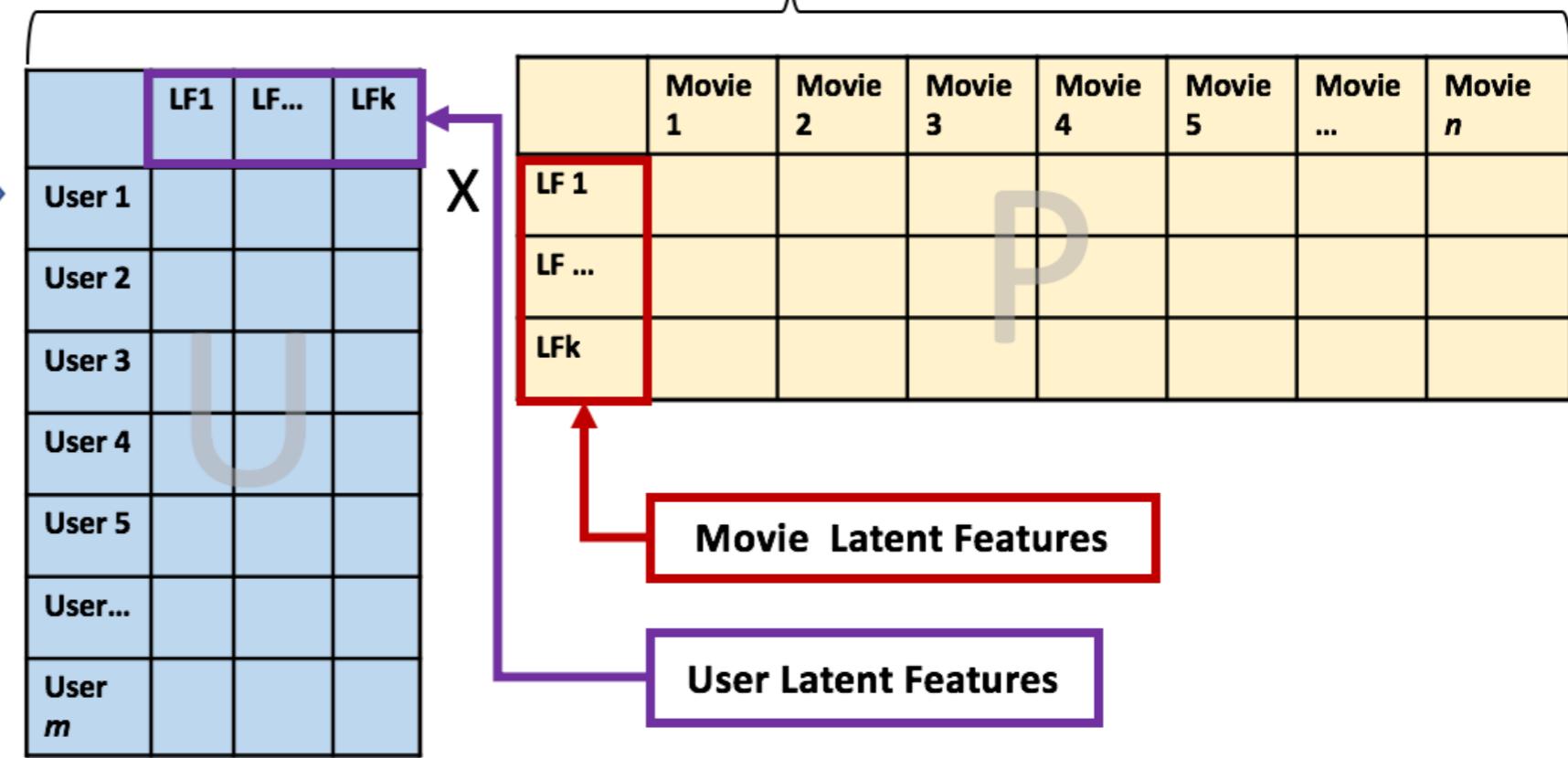
P

Original Ratings Matrix

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie ...	Movie n
User 1							
User 2							
User 3							
User 4							
User 5							
User ...							
User m							

ALS

Factor Matrices



Original Ratings Matrix

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Drama MovN
User 1	5	5	1	1	1	...	1
User 2	4	4	1	1	1	...	1
User 3	4	5	1	1	1	...	1
User 4	2	2	1	1	1	...	1
User 5	1	1	1	5	5	...	5
User...
User <i>m</i>	1	1	1	4	5	...	5

ALS

Factor Matrices

	LF1	LF...	LFk		Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Movie <i>n</i>
User 1	2.200	LF 1	2.1	2.2	.01	.2	.041
User 2	1.900	LF
User 3	2.101	LFk	.1	.01	.04	2.0	2.1	...	2.2
User 4	.801								
User 5	.01	...	2.3								
User...								
User <i>m</i>	.00	...	2.2								

Original Ratings Matrix

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Drama MovN
User 1	5	5	1	1	1	...	1
User 2	4	4	1	1	1	...	1
User 3	4	5	1	1	1	...	1
User 4	2	2	1	1	1	...	1
User 5	1	1	1	5	5	...	5
User...
User <i>m</i>	1	1	1	4	5	...	5

ALS

Factor Matrices

	LF1	LF...	LFk		Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Movie <i>n</i>
User 1	2.200		2.1	2.2	.01	.2	.041
User 2	1.900	
User 3	2.101	
User 4	.801	
User 5	.01	...	2.3	
User...
User <i>m</i>	.00	...	2.2	

Original Ratings Matrix

[No Title]	Horror	Horror	Horror	Drama	Drama	Movie	Drama
	Mov1	Mov2	Mov3	Mov1	Mov2	...	MovN
User 1	5	5	1	1	1	...	1
User 2	4	4	1	1	1	...	1
User 3	4	5	1	1	1	...	1
User 4	2	2	1	1	1	...	1
User 5	1	1	1	5	5	...	5
User...
User <i>m</i>	1	1	1	4	5	...	5

ALS

Factor Matrices

	LF1	LF...	LFk		Horror	Horror	Horror	Drama	Drama	Movie	Movie
	Mov1	Mov2	Mov3	Mov1	Mov2	...	MovN	Mov1	Mov2	...	n
User 1	2.200		2.1	2.2	.01	.2	.041
User 2	1.900	
User 3	2.101	
User 4	.801	
User 5	.01	...	2.3	
User...
User <i>m</i>	.00	...	2.2	

Original Ratings Matrix

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Drama MovN
User 1	5	5	1	1	1	...	1
User 2	4	4	1	1	1	...	1
User 3	4	5	1	1	1	...	1
User 4	2	2	1	1	1	...	1
User 5	1	1	1	5	5	...	5
User...
User <i>m</i>	1	1	1	4	5	...	5

ALS

Factor Matrices

	LF1	LF...	LFk		Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Movie <i>n</i>
User 1	2.200	LF 1	2.1	2.2	.01	.2	.041
User 2	1.900	LF
User 3	2.101	LFk	.1	.01	.04	2.0	2.1	...	2.2
User 4	.801								
User 5	.01	...	2.3								
User...								
User <i>m</i>	.00	...	2.2								

Original Ratings Matrix

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Drama MovN
User 1	5	5	1	1	1	...	1
User 2	4	4	1	1	1	...	1
User 3	4	5	1	1	1	...	1
User 4	2	2	1	1	1	...	1
User 5	1	1	1	5	5	...	5
User...
User <i>m</i>	1	1	1	4	5	...	5

ALS

Factor Matrices

	LF1	LF...	LFk
User 1	2.200
User 2	1.900
User 3	2.101
User 4	.801
User 5	.01	...	2.3
User...
User <i>m</i>	.00	...	2.2

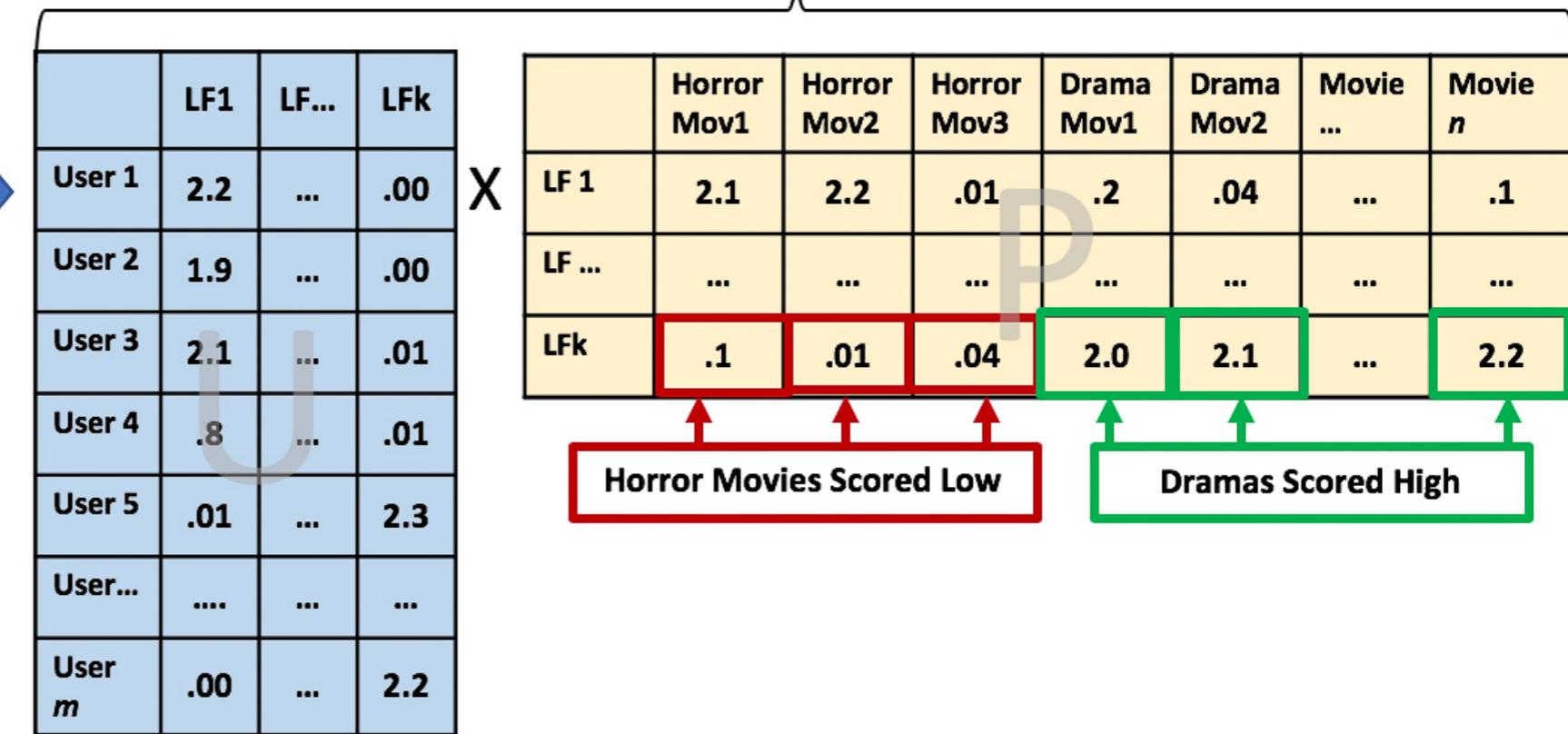
	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Movie <i>n</i>
LF 1	2.1	2.2	.01	.2	.041
LF
LFk	.1	.01	.04	2.0	2.1	...	2.2

Original Ratings Matrix

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Drama MovN
User 1	5	5	1	1	1	...	1
User 2	4	4	1	1	1	...	1
User 3	4	5	1	1	1	...	1
User 4	2	2	1	1	1	...	1
User 5	1	1	1	5	5	...	5
User...
User <i>m</i>	1	1	1	4	5	...	5

ALS

Factor Matrices

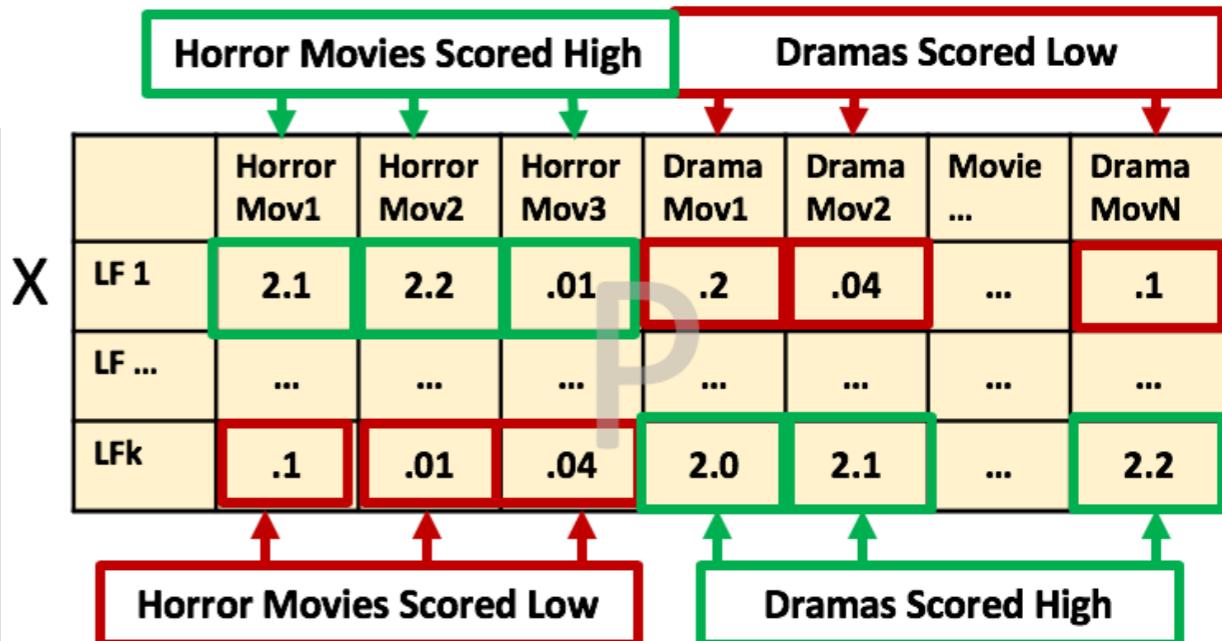


Original Ratings Matrix

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Drama MovN
User 1	5	5	1	1	1	...	1
User 2	4	4	1	1	1	...	1
User 3	4	5	1	1	1	...	1
User 4	2	2	1	1	1	...	1
User 5	1	1	1	5	5	...	5
User...
User <i>m</i>	1	1	1	4	5	...	5

ALS

	LF1	LF...	LFk
User 1	2.200
User 2	1.900
User 3	2.101
User 4	.801
User 5	.01	...	2.3
User...
User <i>m</i>	.00	...	2.2



Factor Matrices

Original Matrix

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
User 1	5	4.8	4.9	...	4.9
User 2	4.5	4.3	4.4	...	4.3
User 3	4.8	4.6	4.7	...	4.6
User 4	2.1	1.98	2	...	2
User 5	4.1	4	4	...	4
User...
User <i>m</i>	4.6	4.4	4.5	...	4.4

ALS



	LF 1	LF...	LF <i>k</i>
User 1	2.2	...	0.0
User 2	1.95	...	0.0
User 3	2.1	...	0.01
User 4	0.90	...	0.01
User 5	1.8	...	0.03
User...
User <i>m</i>	2.0	...	0.02

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
LF 1	2.3	2.2	2.23	...	2.21
LF...
LF <i>k</i>	0.1	0.01	0.02	...	0.02

Factor Matrices

Original Matrix

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
User 1	5	4.8	4.9	...	4.9
User 2	4.5	4.3	4.4	...	4.3
User 3	4.8	4.6	4.7	...	4.6
User 4	2.1	1.98	2	...	2
User 5	4.1	4	4	...	4
User...
User <i>m</i>	4.6	4.4	4.5	...	4.4

ALS



	LF 1	LF...	LF <i>k</i>
User 1	2.2	...	0.0
User 2	1.95	...	0.0
User 3	2.1	...	0.01
User 4	0.90	...	0.01
User 5	1.8	...	0.03
User...
User <i>m</i>	2.0	...	0.02

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
LF 1	2.3	2.2	2.23	...	2.21
LF...
LF <i>k</i>	0.1	0.01	0.02	...	0.02

Factor Matrices

Original Matrix

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
User 1	5	4.8	4.9	...	4.9
User 2	4.5	4.3	4.4	...	4.3
User 3	4.8	4.6	4.7	...	4.6
User 4	2.1	1.98	2	...	2
User 5	4.1	4	4	...	4
User...
User <i>m</i>	4.6	4.4	4.5	...	4.4

ALS



	LF 1	LF...	LF <i>k</i>
User 1	2.2	...	0.0
User 2	1.95	...	0.0
User 3	2.1	...	0.01
User 4	0.90	...	0.01
User 5	1.8	...	0.03
User...
User <i>m</i>	2.0	...	0.02

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
LF 1	2.3	2.2	2.23	...	2.21
LF...
LF <i>k</i>	0.1	0.01	0.02	...	0.02

Movie 1: "The Lion King"
Movie 2: "10 Things I Hate About You"
Movie 4: "West Side Story"
Movie *n*: "She's the Man"

Factor Matrices

Original Matrix

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
User 1	5	4.8	4.9	...	4.9
User 2	4.5	4.3	4.4	...	4.3
User 3	4.8	4.6	4.7	...	4.6
User 4	2.1	1.98	2	...	2
User 5	4.1	4	4	...	4
User...
User <i>m</i>	4.6	4.4	4.5	...	4.4
Avg Rating	4.2	4.0	4.1	...	4

ALS



	LF 1	LF...	LF <i>k</i>
User 1	2.2	...	0.0
User 2	1.95	...	0.0
User 3	2.1	...	0.01
User 4	0.90	...	0.01
User 5	1.8	...	0.03
User...
User <i>m</i>	2.0	...	0.02

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
LF 1	2.3	2.2	2.23	...	2.21
LF...
LF <i>k</i>	0.1	0.01	0.02	...	0.02

Movie 1: "The Lion King"
Movie 2: "10 Things I Hate About You"
Movie 4: "West Side Story"
Movie *n*: "She's the Man"

Factor Matrices

Original Matrix

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
User 1	5	4.8	4.9	...	4.9
User 2	4.5	4.3	4.4	...	4.3
User 3	4.8	4.6	4.7	...	4.6
User 4	2.1	1.98	2	...	2
User 5	4.1	4	4	...	4
User...
User <i>m</i>	4.6	4.4	4.5	...	4.4
<i>Avg Rating</i>	4.2	4.0	4.1	...	4

ALS

The diagram illustrates the decomposition of the original matrix into two factor matrices. A large blue arrow labeled "ALS" points from the original matrix to the factor matrices. To the right of the factor matrices is a large "X" symbol.

Original Matrix:

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
User 1	5	4.8	4.9	...	4.9
User 2	4.5	4.3	4.4	...	4.3
User 3	4.8	4.6	4.7	...	4.6
User 4	2.1	1.98	2	...	2
User 5	4.1	4	4	...	4
User...
User <i>m</i>	4.6	4.4	4.5	...	4.4
<i>Avg Rating</i>	4.2	4.0	4.1	...	4

User Latent Feature Matrix (LF):

	LF 1	LF...	LF <i>k</i>
User 1	2.2	...	0.0
User 2	1.95	...	0.0
User 3	2.1	...	0.01
User 4	0.90	...	0.01
User 5	1.8	...	0.03
User...
User <i>m</i>	2.0	...	0.02

Item Latent Feature Matrix (LF):

	Movie 1	Movie 2	Movie3	Movie...	Movie <i>n</i>
LF 1	2.3	2.2	2.23	...	2.21
LF...
LF <i>k</i>	0.1	0.01	0.02	...	0.02

Annotations:

- Movie 1:** "The Lion King" ("Hamlet")
- Movie 2:** "10 Things I Hate About You" ("Taming of the Shrew")
- Movie 4:** "West Side Story" ("Romeo and Juliet")
- Movie *n*:** "She's the Man" ("Twelfth Night")

Let's practice!

BUILDING RECOMMENDATION ENGINES WITH PYSPARK