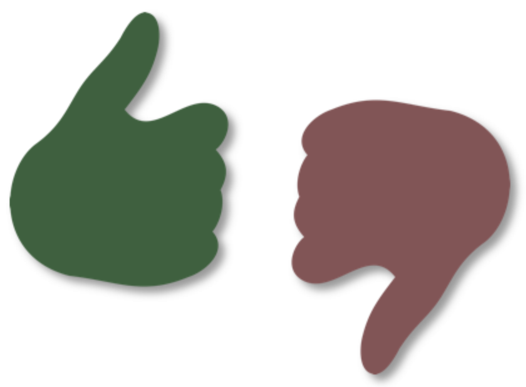
Why learn how to build recommendation engines?

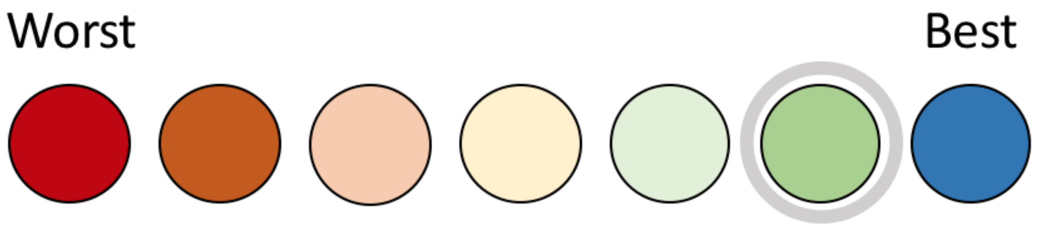
BUILDING RECOMMENDATION ENGINES WITH PYSPARK

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

Movie	Rating
Dark Knight	5
The Notebook	-
Iron Man	5
Finding Nemo	3
Shrek	4
Tangled	2
Jurassic Park	-

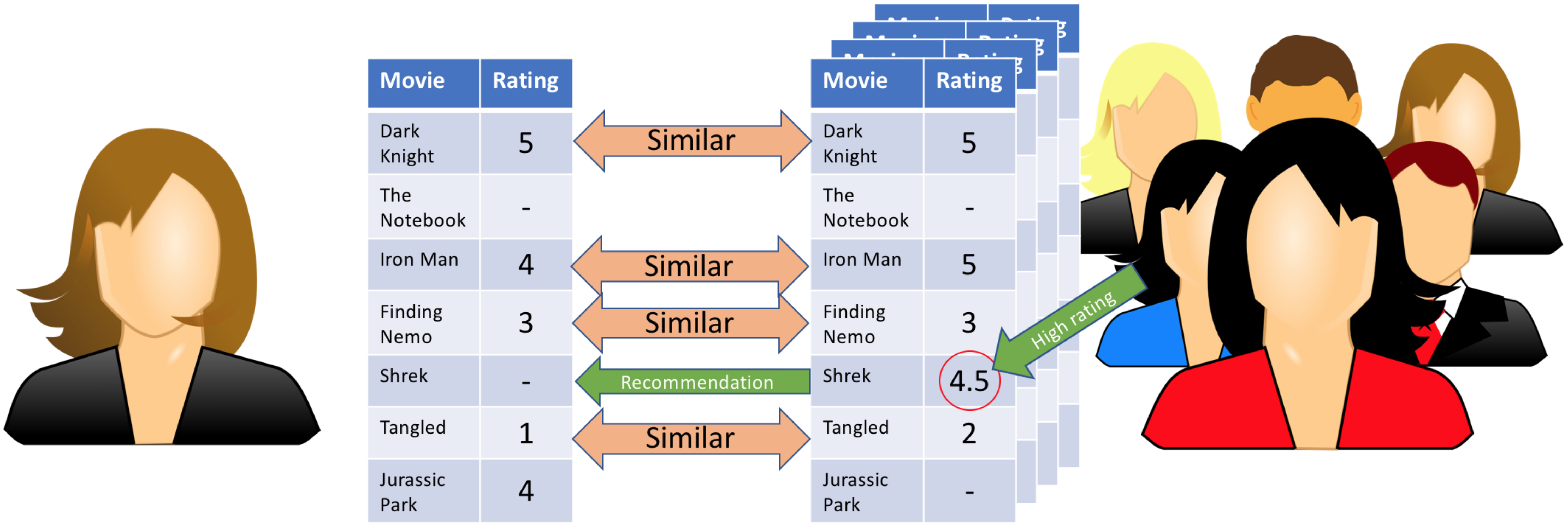
How recommendation engines work

Movie	Rating		Movie	Rating
Dark Knight	5	Similar	Dark Knight	5
The Notebook	-		The Notebook	-
Iron Man	4	Similar	Iron Man	5
Finding Nemo	3	Similar	Finding Nemo	3
Shrek	-		Shrek	4
Tangled	1	Similar	Tangled	2
Jurassic Park	4		Jurassic Park	-

How recommendation engines work

Movie	Rating		Movie	Rating
Dark Knight	5	Similar	Dark Knight	5
The Notebook	-		The Notebook	-
Iron Man	4	Similar	Iron Man	5
Finding Nemo	3	Similar	Finding Nemo	3
Shrek	-		Shrek	4.5
Tangled	1	Similar	Tangled	2
Jurassic Park	4		Jurassic Park	-

How recommendation engines work



The Power of Recommendation Engines

powered by increasingly sop
models that analyze transaction data and dig (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
dations through social networks, user
for the average

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

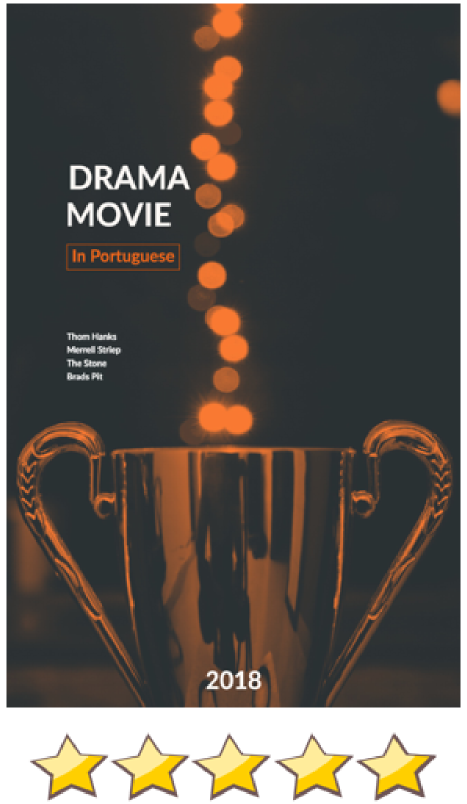
- Genre: Comedy, Action, Drama
- Animation: Animated, Not animated
- Language: English, Spanish, Korean
- Decade Produced: 1950's, 1980's
- Actors: Meryl Streep, Tom Hanks

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

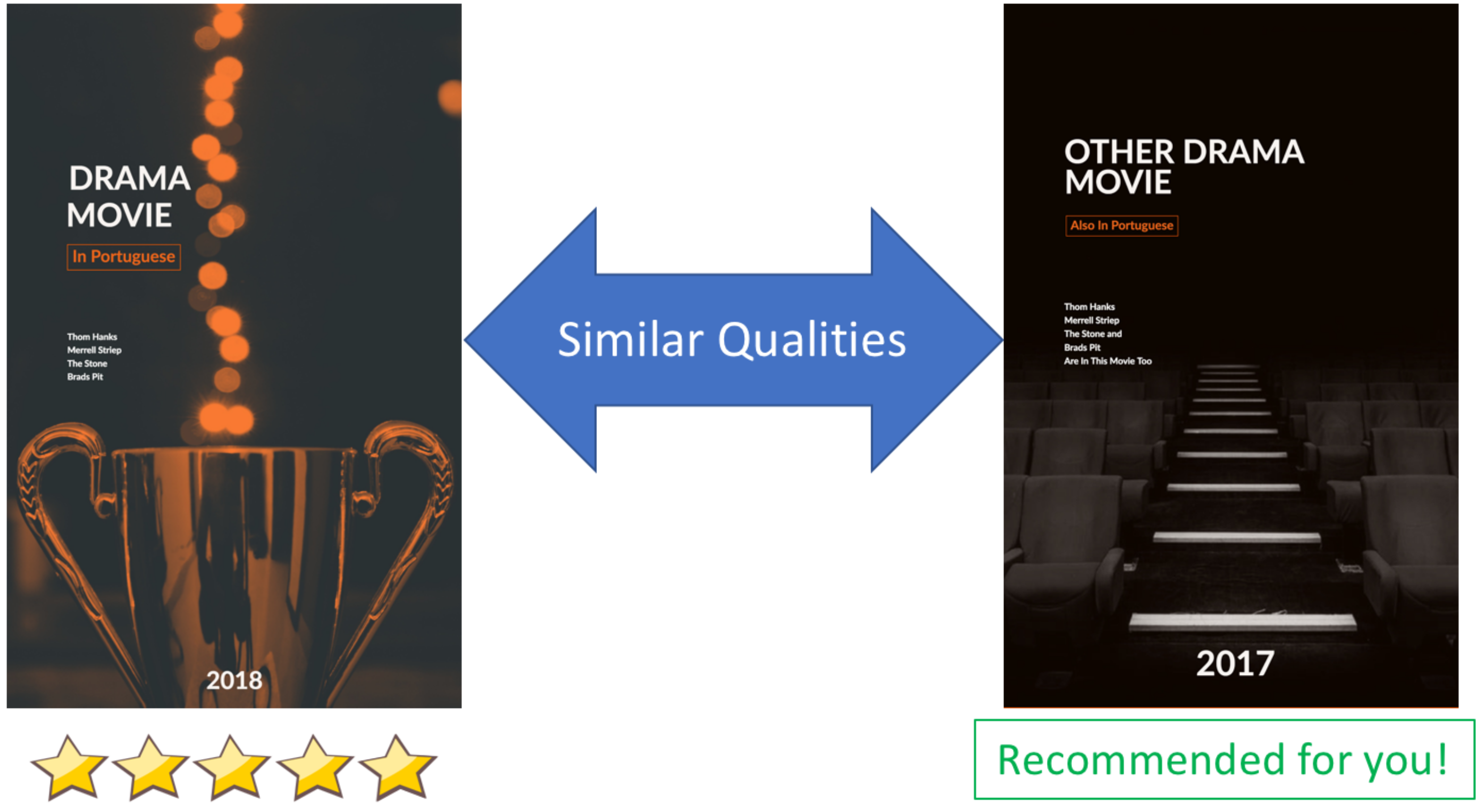
Based on features of items

COLLABORATIVE FILTERING

Two types of recommendation engines

CONTENT-BASED FILTERING

Based on features of items

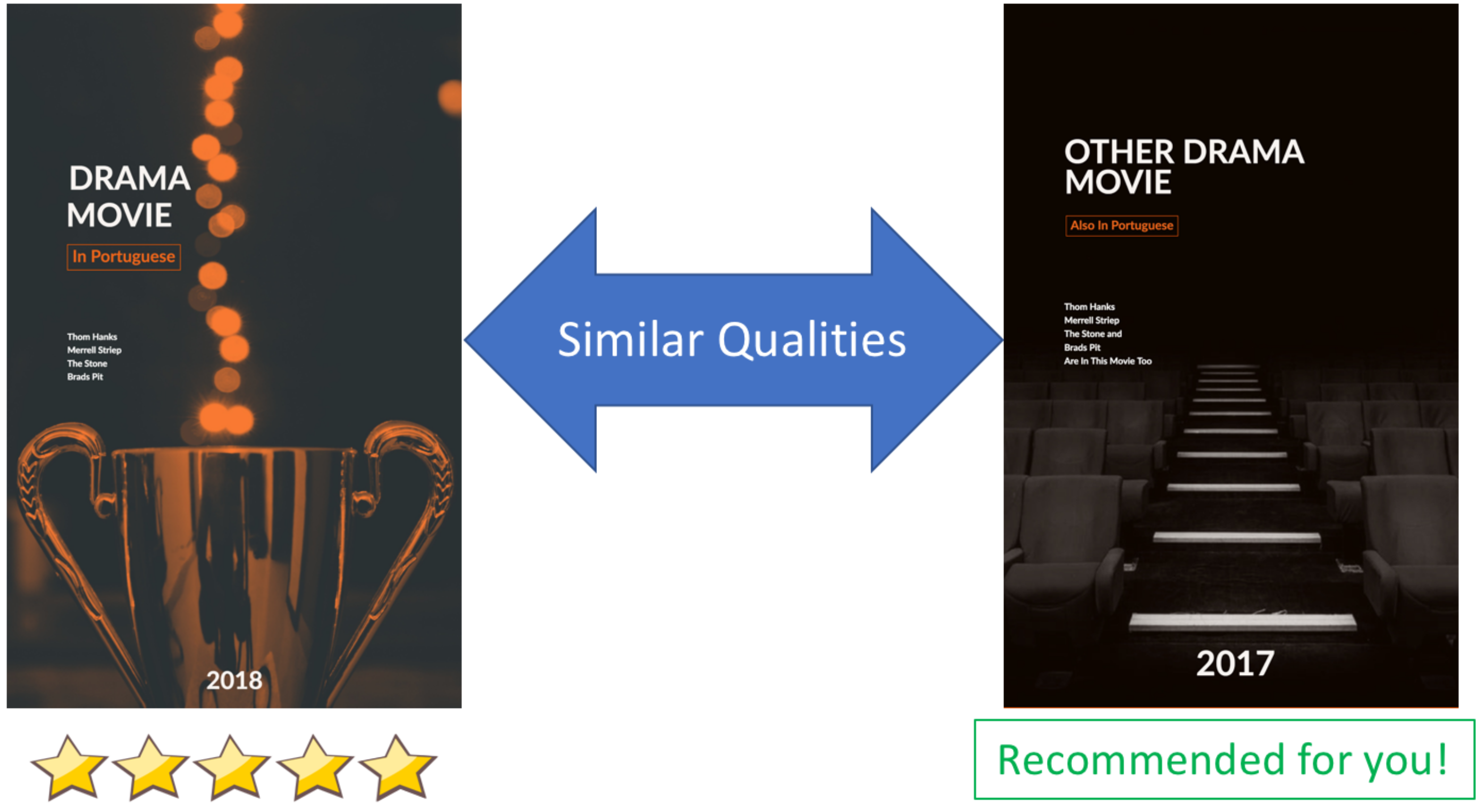


COLLABORATIVE FILTERING

Two types of recommendation engines

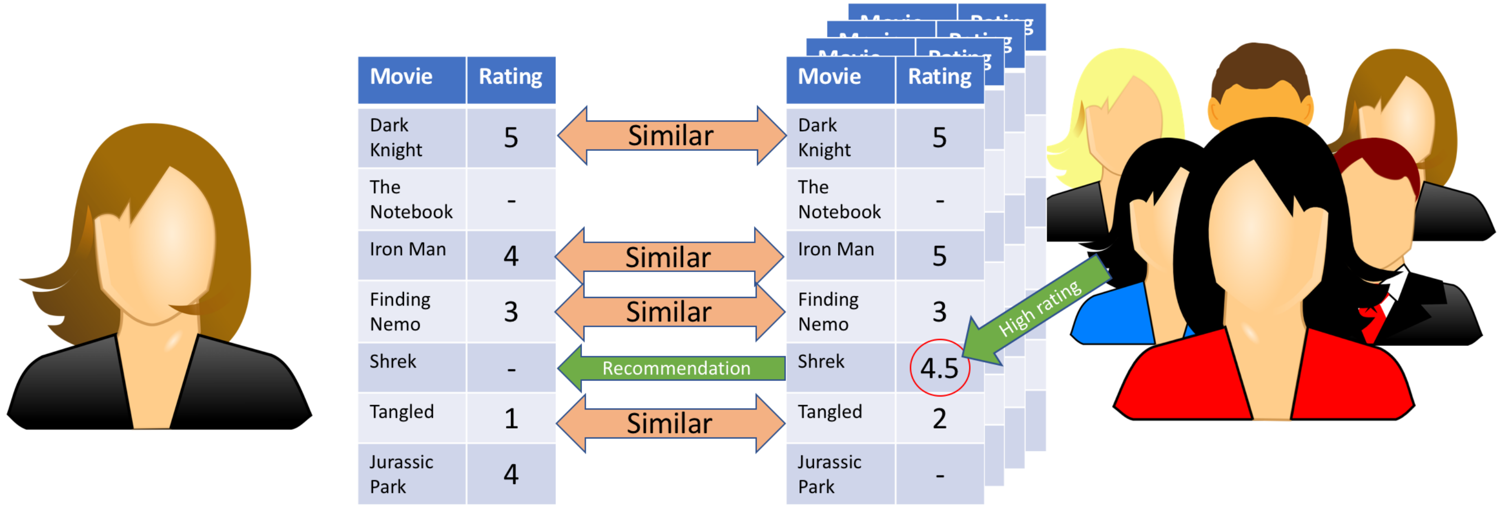
CONTENT-BASED FILTERING

Based on features of items



COLLABORATIVE FILTERING

Based on similar user preferences



Two types of ratings

Two types of ratings

EXPLICIT RATINGS

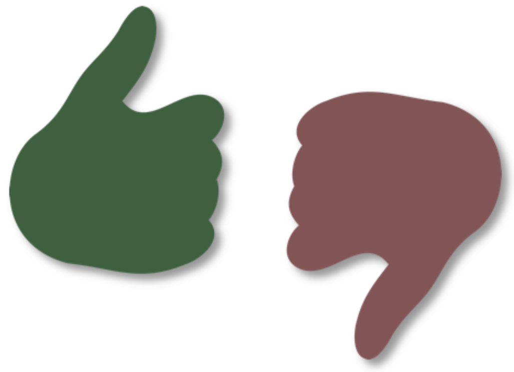
Two types of ratings

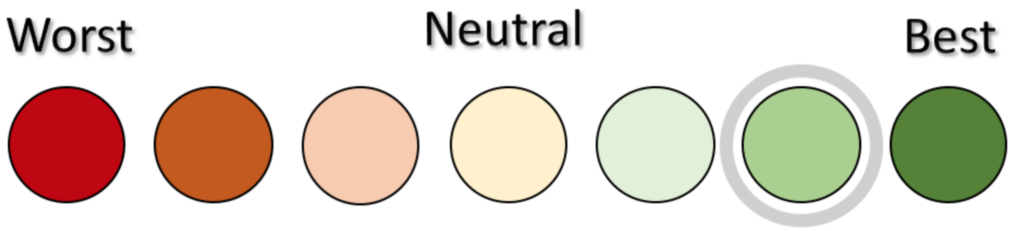
EXPLICIT RATINGS

IMPLICIT RATINGS

Two types of ratings

EXPLICIT RATINGS

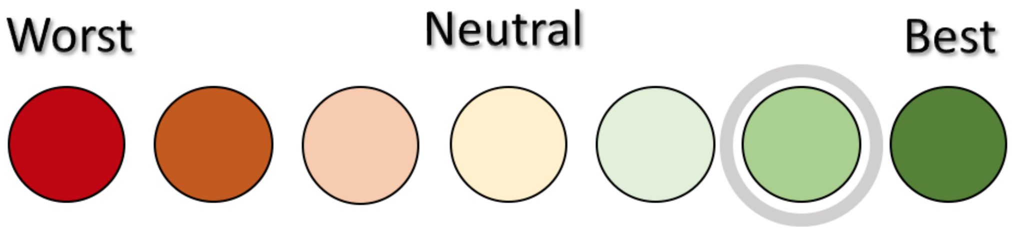




IMPLICIT RATINGS

Two types of ratings

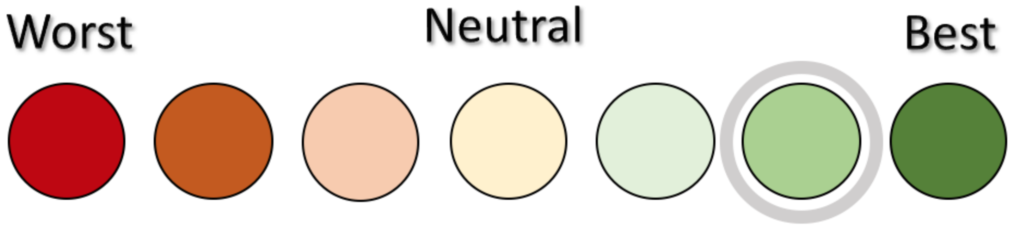
EXPLICIT RATINGS



IMPLICIT RATINGS

Two types of ratings

EXPLICIT RATINGS

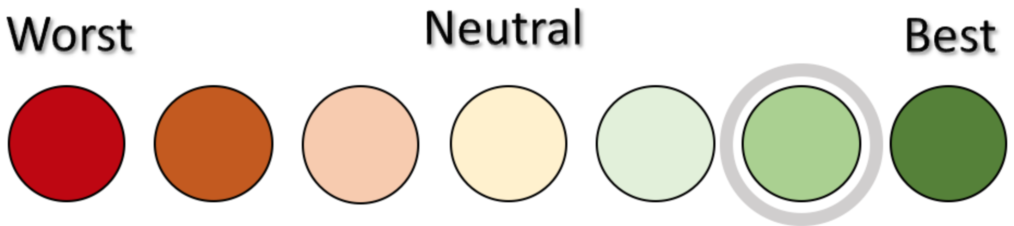


IMPLICIT RATINGS

≈ 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

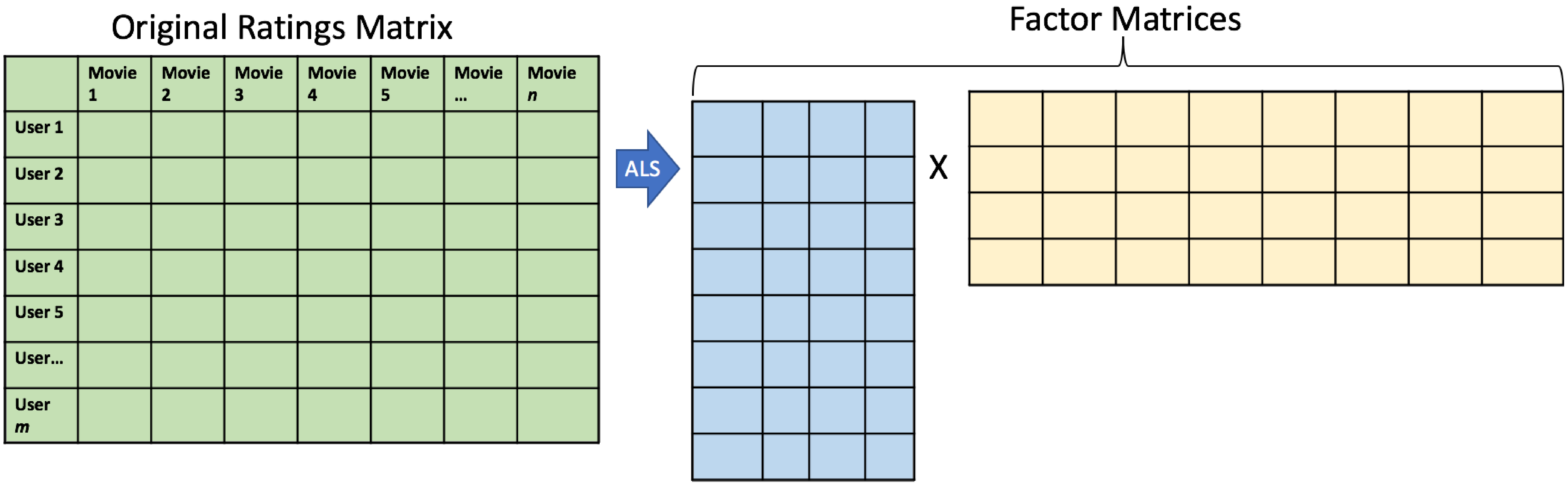
Uses for recommendation engines

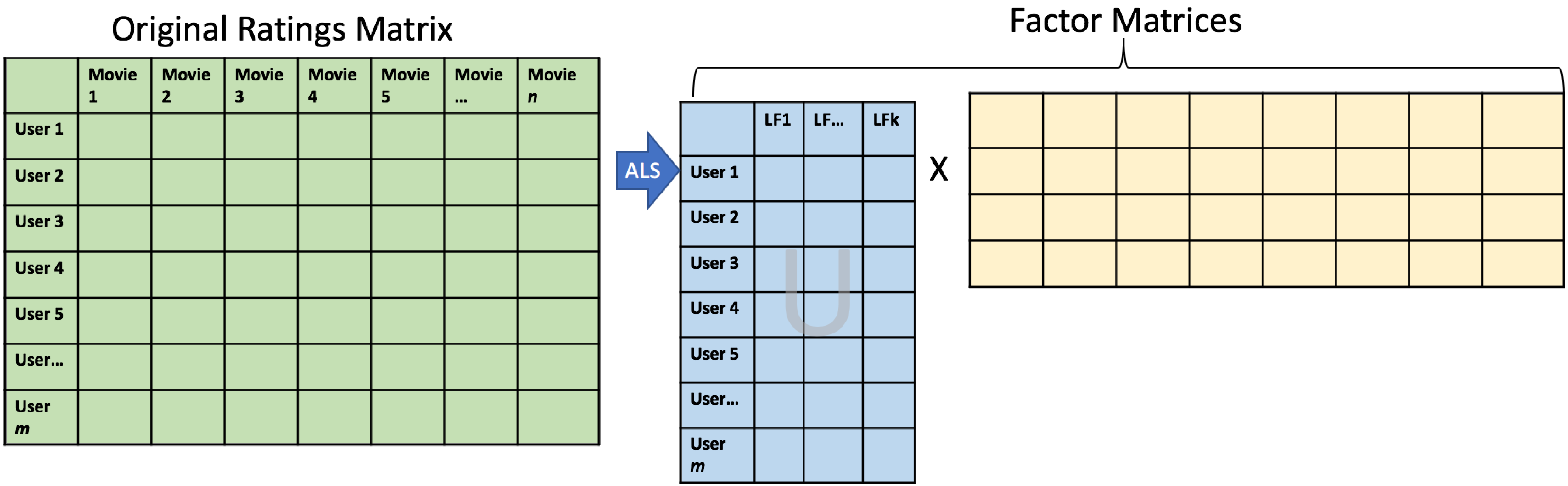
BUILDING RECOMMENDATION ENGINES WITH PYSPARK

Jamen Long
Data Scientist at Nike

Original Ratings Matrix

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


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
X
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
Factor Matrices
LF1
LF...
LFk
User 1
User 2
User 3
User 4
User 5
User...
User m
U
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

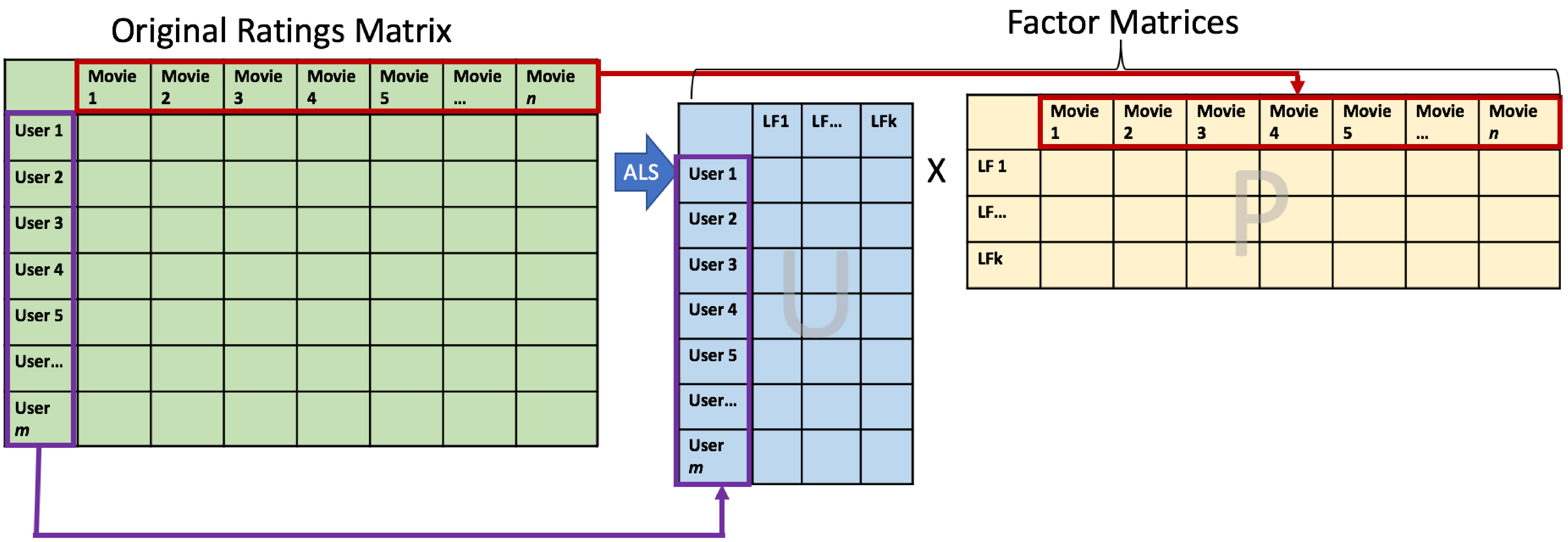
U

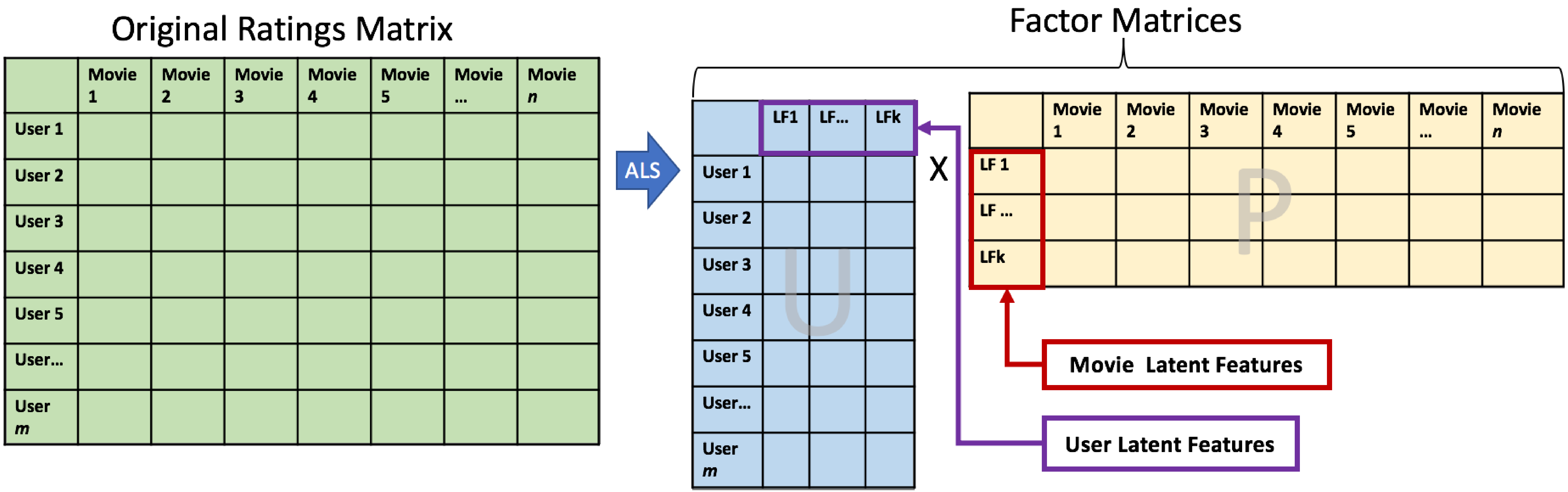
	LF1	LF...	LFk
User 1			
User 2			
User 3			
User 4			
User 5			
User...			
User *m*			

X

P

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


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
U
X
Factor Matrices
Movie 1
Movie 2
Movie 3
Movie 4
Movie 5
Movie ...
Movie n
LF 1
LF...
LFk
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
LF1
LF...
LFk
U
X
P
LF 1
LF ...
LFk
Movie Latent Features
User Latent Features

Original Ratings Matrix

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Drama Mov*N*
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 *m*	1	1	1	4	5	...	5

ALS

Factor Matrices

U

	LF1	LF...	LFk
User 1	2.2	...	.00
User 2	1.9	...	.00
User 3	2.1	...	.01
User 4	.8	...	.01
User 5	.01	...	2.3
User...	....	...	...
User *m*	.00	...	2.2

X

P

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Movie *n*
LF 1	2.1	2.2	.01	.2	.04	...	.1
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 Mov*N*
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 *m*	1	1	1	4	5	…	5

ALS

Factor Matrices

U

	LF1	LF…	LFk
User 1	2.2	…	.00
User 2	1.9	…	.00
User 3	2.1	…	.01
User 4	.8	…	.01
User 5	.01	…	2.3
User…	….	…	…
User *m*	.00	…	2.2

X

P

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie …	Movie *n*
LF 1	2.1	2.2	.01	.2	.04	…	.1
LF …	…	…	…	…	…	…	…
LFk	.1	.01	.04	2.0	2.1	…	2.2

Original Ratings Matrix

[No Title]

	ror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Drama Mov*N*
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 *m*	1	1	1	4	5	...	5

ALS

Factor Matrices

U

	LF1	LF...	LFk
User 1	2.2	...	.00
User 2	1.9	...	.00
User 3	2.1	...	.01
User 4	.8	...	.01
User 5	.01	...	2.3
User...	....	...	...
User *m*	.00	...	2.2

X

P

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Movie *n*
LF 1	2.1	2.2	.01	.2	.04	...	.1
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 Mov*N*
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 *m*	1	1	1	4	5	...	5

ALS

Factor Matrices

U

	LF1	LF...	LFk
User 1	2.2	...	.00
User 2	1.9	...	.00
User 3	2.1	...	.01
User 4	.8	...	.01
User 5	.01	...	2.3
User...	....	...	...
User *m*	.00	...	2.2

X

P

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Movie *n*
LF 1	2.1	2.2	.01	.2	.04	...	.1
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 Mov*N*
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 *m*	1	1	1	4	5	…	5

ALS

Factor Matrices

U

	LF1	LF…	LFk
User 1	2.2	…	.00
User 2	1.9	…	.00
User 3	2.1	…	.01
User 4	.8	…	.01
User 5	.01	…	2.3
User…	….	…	…
User *m*	.00	…	2.2

X

P

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie …	Movie *n*
LF 1	2.1	2.2	.01	.2	.04	…	.1
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 Mov*N*
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 *m*	1	1	1	4	5	...	5

ALS

Factor Matrices

U

	LF1	LF...	LFk
User 1	2.2	...	.00
User 2	1.9	...	.00
User 3	2.1	...	.01
User 4	.8	...	.01
User 5	.01	...	2.3
User...	....	...	...
User *m*	.00	...	2.2

X

P

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Movie *n*
LF 1	2.1	2.2	.01	.2	.04	...	.1
LF ...	...	...	...	...	...	...	...
LFk	.1	.01	.04	2.0	2.1	...	2.2

Horror Movies Scored Low

Dramas Scored High

Original Ratings Matrix

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Drama Mov*N*
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 *m*	1	1	1	4	5	...	5

ALS

U

	LF1	LF...	LFk
User 1	2.2	...	.00
User 2	1.9	...	.00
User 3	2.1	...	.01
User 4	.8	...	.01
User 5	.01	...	2.3
User...	....	...	...
User *m*	.00	...	2.2

X

Horror Movies Scored High

Dramas Scored Low

P

	Horror Mov1	Horror Mov2	Horror Mov3	Drama Mov1	Drama Mov2	Movie ...	Drama MovN
LF 1	2.1	2.2	.01	.2	.04	...	.1
LF ...	...	...	...	...	...	...	...
LFk	.1	.01	.04	2.0	2.1	...	2.2

Horror Movies Scored Low

Dramas Scored High

Original Matrix

	Movie 1	Movie 2	Movie3	Movie...	Movie *n*
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 *m*	4.6	4.4	4.5	...	4.4

ALS

Factor Matrices

U

	LF 1	LF...	LF k
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 *m*	2.0	...	0.02

X

P

	Movie 1	Movie 2	Movie3	Movie...	Movie *n*
LF 1	2.3	2.2	2.23	...	2.21
LF...	...	...	...	...	...
LF k	0.1	0.01	0.02	...	0.02

Original Matrix

	Movie 1	Movie 2	Movie3	Movie...	Movie *n*
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 *m*	4.6	4.4	4.5	...	4.4

ALS

Factor Matrices

U

	LF 1	LF...	LF k
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 *m*	2.0	...	0.02

X

P

	Movie 1	Movie 2	Movie3	Movie...	Movie *n*
LF 1	2.3	2.2	2.23	...	2.21
LF...	...	...	...	...	...
LF k	0.1	0.01	0.02	...	0.02

Factor Matrices

Original Matrix

	Movie 1	Movie 2	Movie3	Movie...	Movie *n*
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 *m*	4.6	4.4	4.5	...	4.4

ALS

U

	LF 1	LF...	LF k
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 *m*	2.0	...	0.02

X

P

	Movie 1	Movie 2	Movie3	Movie...	Movie *n*
LF 1	2.3	2.2	2.23	...	2.21
LF...	...	...	...	...	...
LF k	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”

Original Matrix

	Movie 1	Movie 2	Movie3	Movie...	Movie *n*
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 *m*	4.6	4.4	4.5	...	4.4
Avg Rating	4.2	4.0	4.1	...	4

ALS

Factor Matrices

U

	LF 1	LF...	LF k
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 *m*	2.0	...	0.02

X

P

	Movie 1	Movie 2	Movie3	Movie...	Movie *n*
LF 1	2.3	2.2	2.23	...	2.21
LF...	...	...	...	...	...
LF k	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"

Original Matrix

	Movie 1	Movie 2	Movie3	Movie...	Movie *n*
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 *m*	4.6	4.4	4.5	...	4.4
Avg Rating	4.2	4.0	4.1	...	4

ALS

Factor Matrices

U

	LF 1	LF...	LF k
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 *m*	2.0	...	0.02

X

P

	Movie 1	Movie 2	Movie3	Movie...	Movie *n*
LF 1	2.3	2.2	2.23	...	2.21
LF...	...	...	...	...	...
LF k	0.1	0.01	0.02	...	0.02

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