

What to read next?

A book recommender system

By Angelica DeGaetano

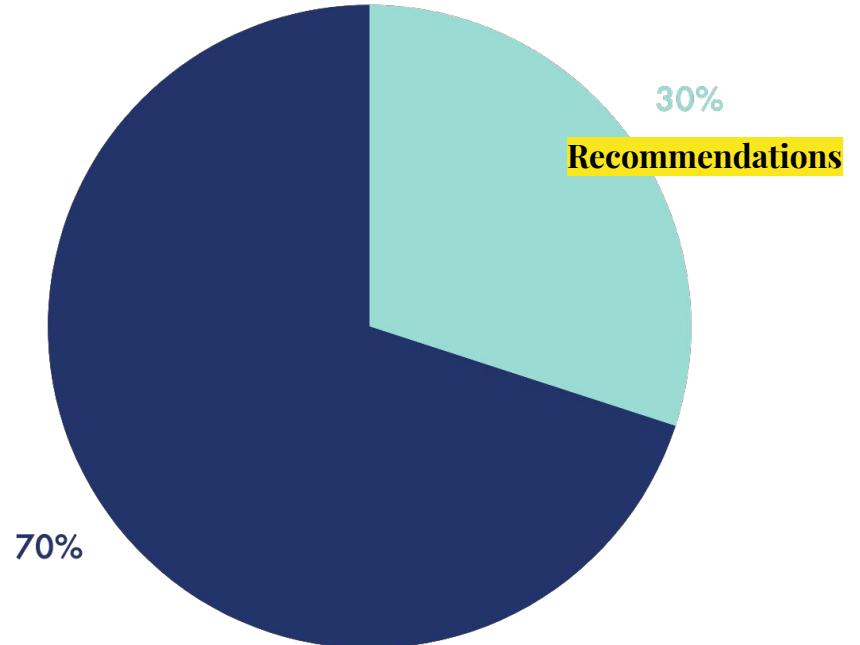


Addressing Choice-Overload with Recommendations

E-Commerce Revenue Worldwide



Photo by Susan Yin on Unsplash





The Data



Books:

484



Reviewers:

28,959

1
didn't like

2
OK

3
liked

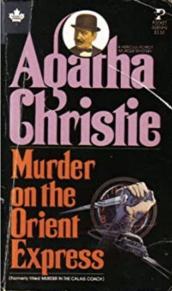
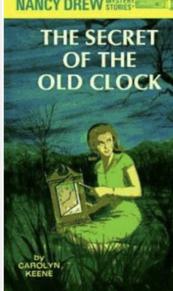
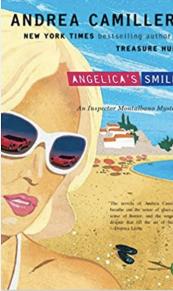
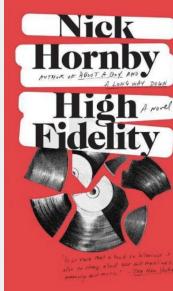
4
really liked

5
amazing

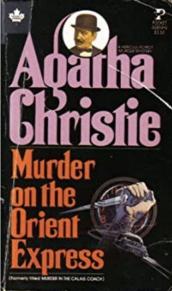
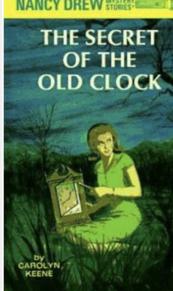
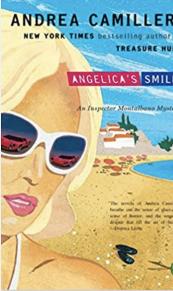
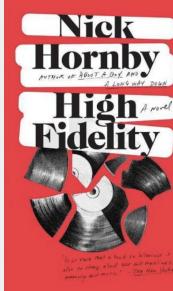
Ratings:

54,387

Because of Others Like You--Collaborative Filtering

	1 	2 	3 	4 	5 
Ana	★★★★		★★★★★	★	
Chris	★★★	★★★★	★★★★		
Sarah	★			★★★★★	★★★★★
Alex		★★★★★	★★★★★	★★	★

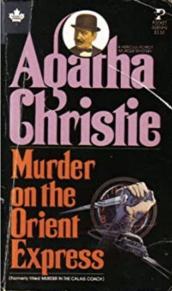
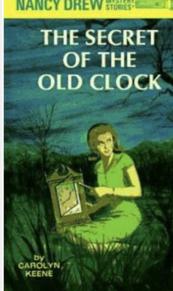
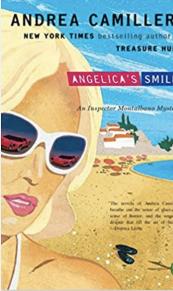
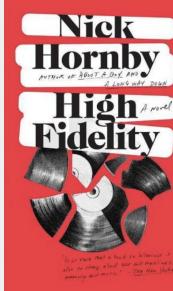
Because of Others Like You--Collaborative Filtering

	1 	2 	3 	4 	5 
Ana	★★★★	?	★★★★★	★	?
Chris	★★★	★★★★	★★★★	?	?
Sarah	★	?	?	★★★★★	★★★★★
Alex	?	★★★★★	★★★★★	★★	★

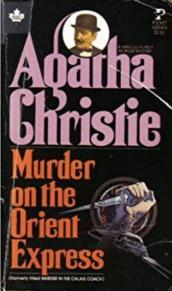
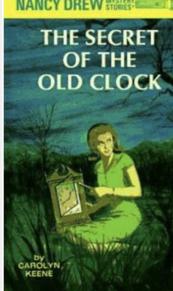
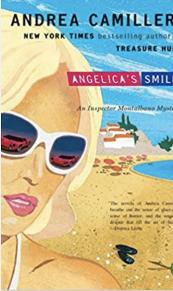
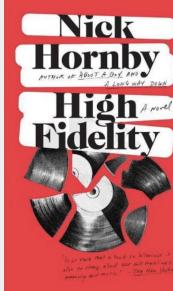
Because of Others Like You--Collaborative Filtering

Ana	★★★★	?	★★★★★	★	?
Chris	★★★	★★★★	★★★★★	?	?
Sarah	★	?	?	★★★★★	★★★★★
Alex	?	★★★★★	★★★★★	★★	★

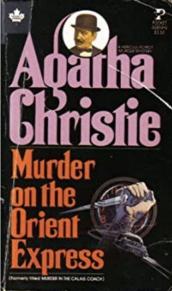
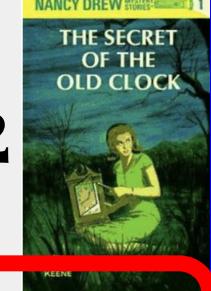
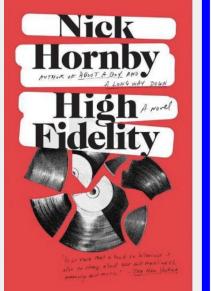
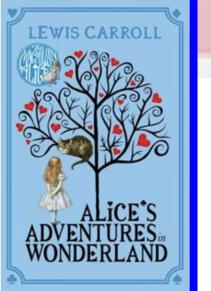
Because of Others Like You--Collaborative Filtering

	1 	2 	3 	4 	5 
Ana	★★★★	?	★★★★★	★	?
Chris	★★★	★★★★	★★★★★	?	?
Sarah	★	?	?	★★★★★	★★★★★
Alex	?	★★★★★	★★★★★	★★	★

Because of Others Like You--Collaborative Filtering

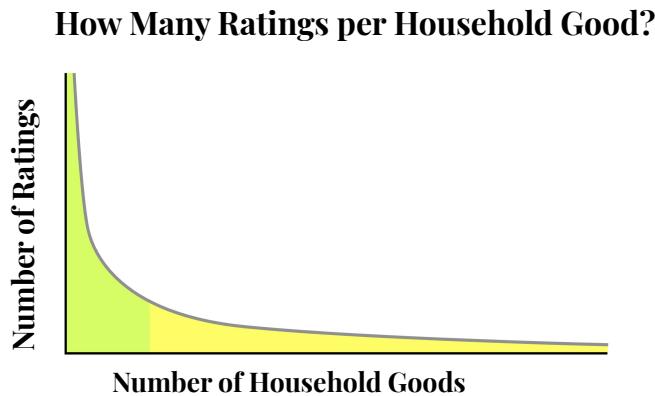
	1 	2 	3 	4 	5 
Ana	★★★★	?	★★★★★	★	?
Chris	★★★	★★★★	★★★★★	?	?
Sarah	★	?	?	★★★★★	★★★★★
Alex	?	★★★★★	★★★★★	★★	★

Because of Others Like You--Collaborative Filtering

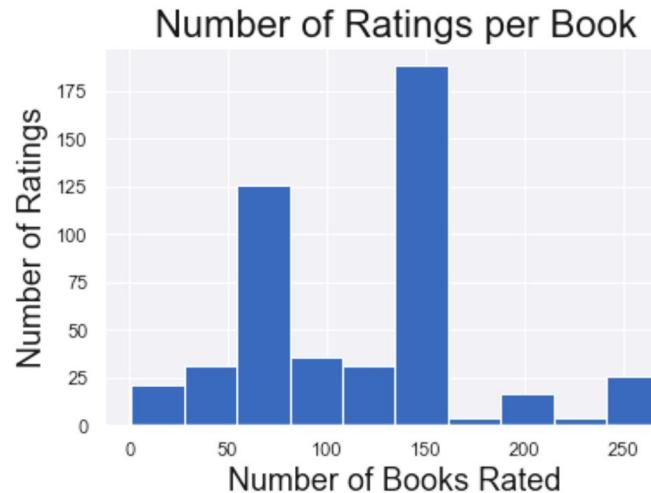
	1 	2 	3 	4 	5 
Ana	★★★★	★★★★★	★★★★★	★	?
Chris	★★★	★★★★	★★★★	?	?
Sarah	★	?	?	★★★★	★★★★★
Alex	?	★★★★★	★★★★★	★★	★

Rating Distribution

Long Tail Problem
Example

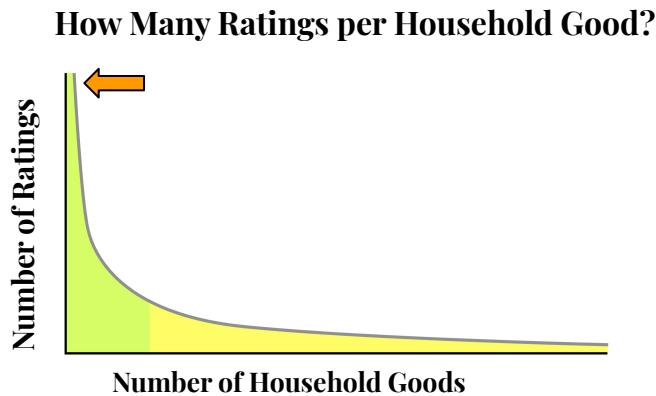


This Dataset

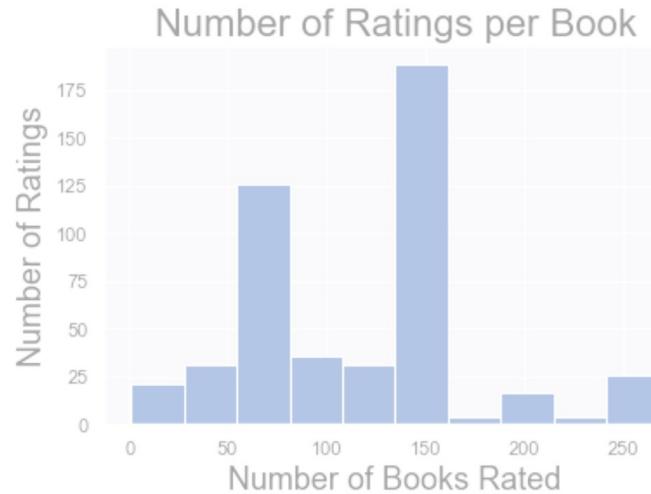


Rating Distribution

Long Tail Problem
Example

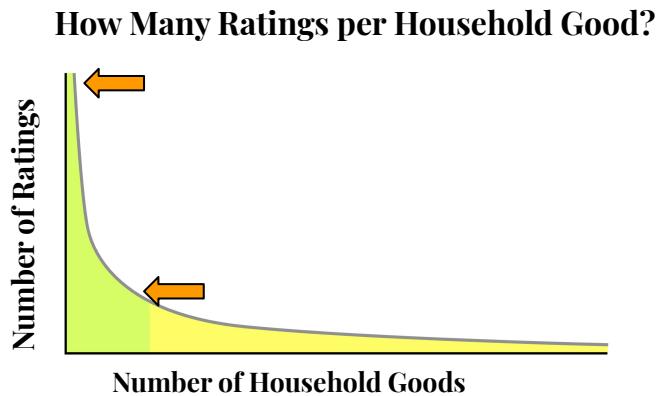


This Dataset

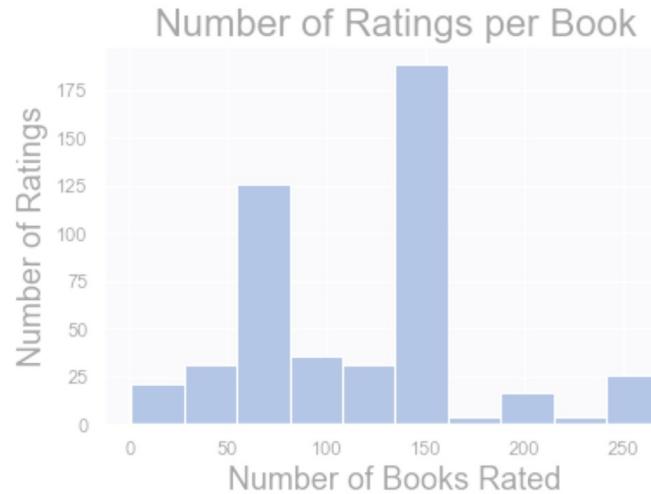


Rating Distribution

Long Tail Problem
Example

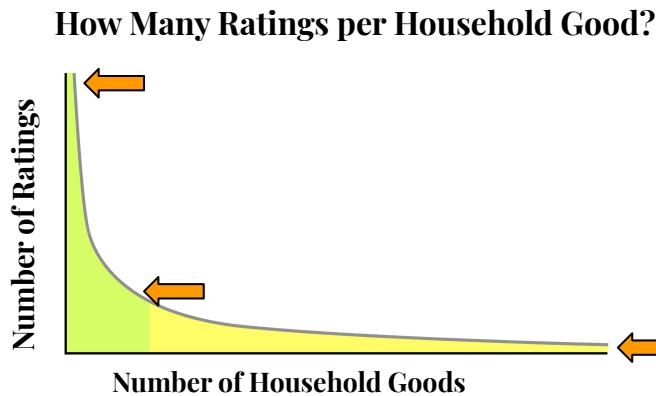


This Dataset

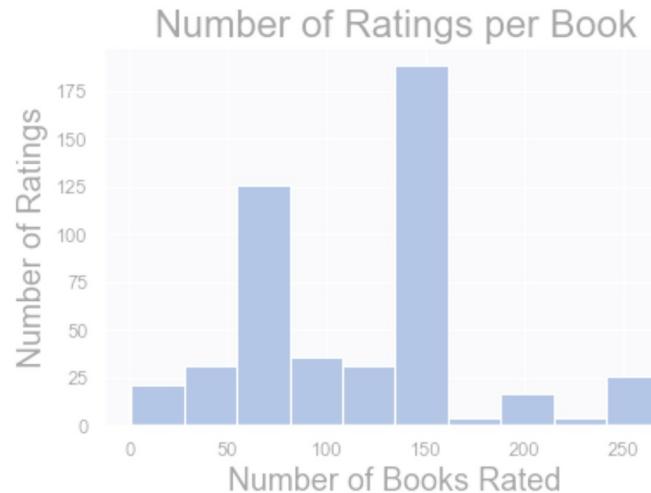


Rating Distribution

Long Tail Problem
Example

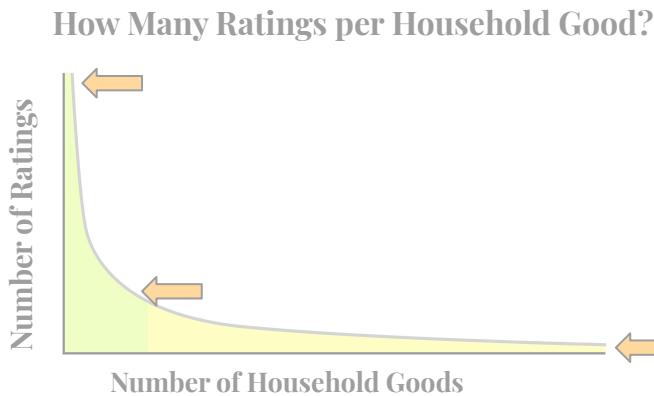


This Dataset

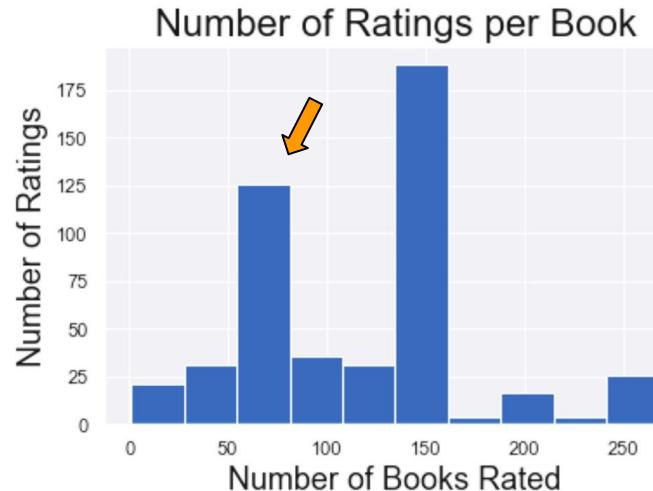


Rating Distribution

Long Tail Problem
Example

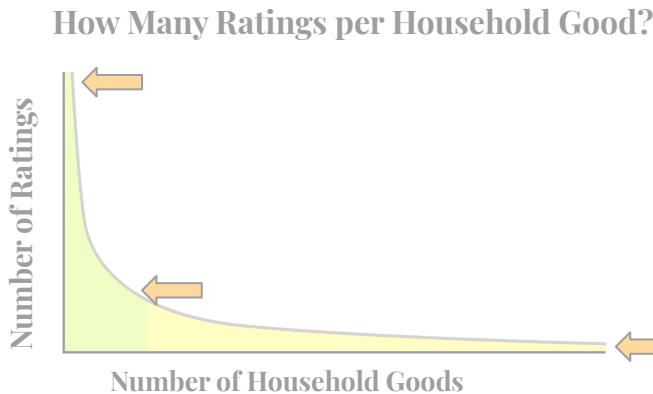


This Dataset

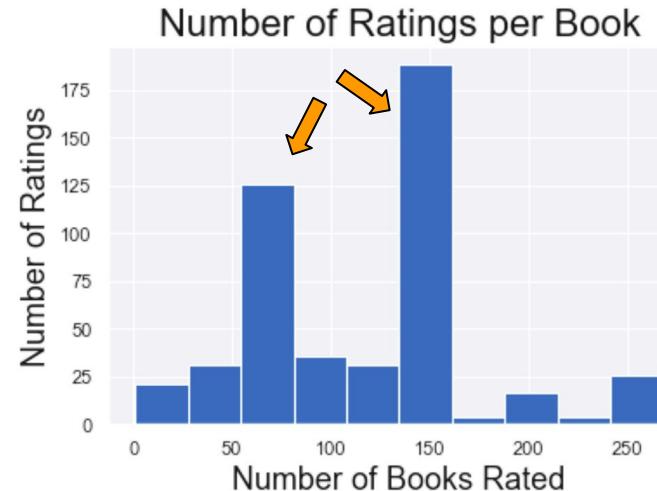


Rating Distribution

Long Tail Problem
Example



This Dataset





The Model

Model: Spark ALS

Dataset Sparsity: 99.60%

Best model RMSE: 1.2

1
didn't like

2
OK

3
liked

4
really liked

5
amazing

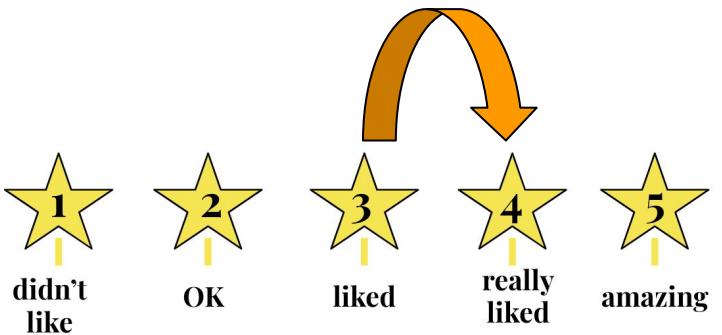


The Model

Model: Spark ALS

Dataset Sparsity: 99.60%

Best model RMSE: 1.2



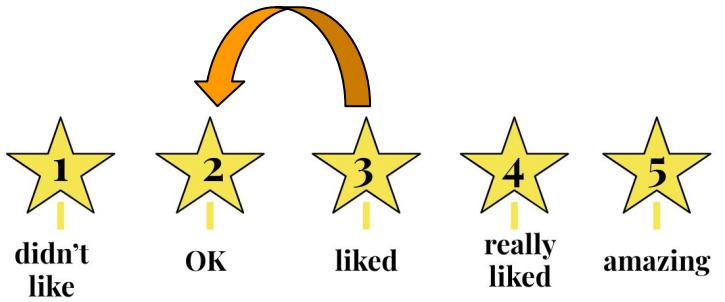


The Model

Model: Spark ALS

Dataset Sparsity: 99.60%

Best model RMSE: 1.2



The Road to Further Recommendations

- ❑ Adding more data
- ❑ Natural Language Processing on text reviews
- ❑ Content-based filtering
- ❑ Deployment of model



Photo by Laura Käpfer on Unsplash

Contact



github.com/angelicacodes



linkedin.com/in/degaetanoma



AngelicaCodes@gmail.com



Cover: Photo by [Patrick Tomasso on Unsplash](#)

Title: Photo by [Nick Fewings on Unsplash](#)

Book clipart: books by [Rocket000](#)

Group clipart: People by [OpenClipart](#)



Photo by [Jaredd Craig on Unsplash](#)