Introduction to Data Science

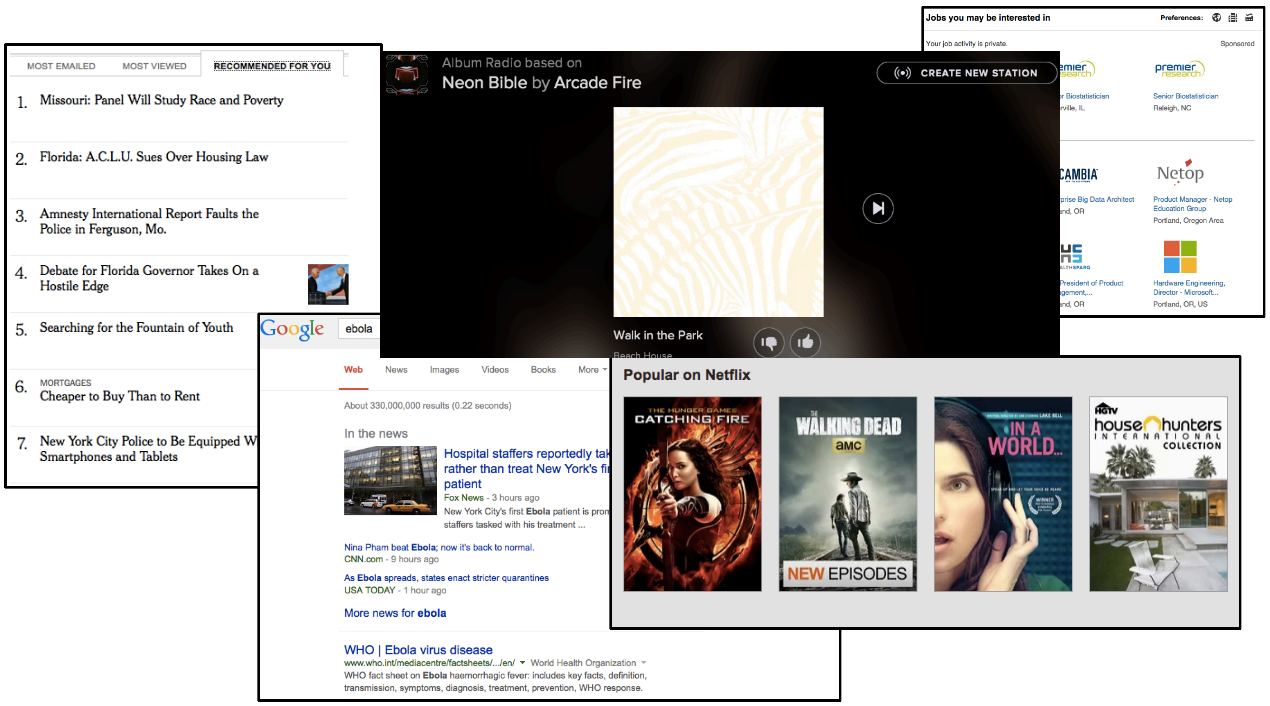
RECOMMENDER SYSTEMS

BRIAN D'ALESSANDRO

RECOMMENDATIONS ARE EVERYWHERE

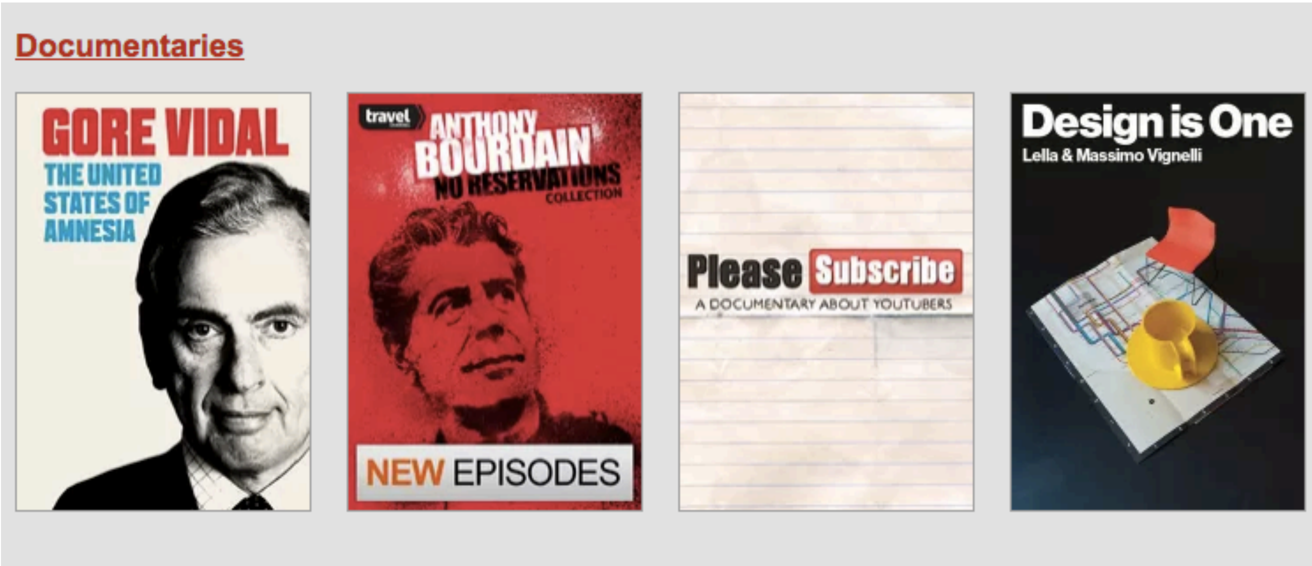
If you use the internet, you likely suffer from this little problem - *too much information and too little time.*

Most companies try to solve this problem for you using data science



RECOMMENDATIONS DRIVE

What you'll watch tonight…



Media content providers rely heavily on recommender systems to create relevant consumption sessions

RECOMMENDATIONS DRIVE

The information I am exposed to…

Google

global warming

Feedback

climate.nasa.gov › resources › global-warming-vs-clim...

Global Warming vs. Climate Change | Resources – Climate ...

Global warming is the long-term heating of Earth's climate system observed since the pre-industrial period (between 1850 and 1900) due to human activities, ...

climate.nasa.gov › causes

The Causes of Climate Change - NASA Climate Change

Scientists attribute the **global warming** trend observed since the mid-20th century to ... Water vapor increases as the Earth's atmosphere warms, but so does the ...

www.nationalgeographic.com › global-warming-overview

What is global warming, facts and information?

We often call the result **global warming**, but it is causing a set of changes to the Earth's climate, or long-term ...

Jan 22, 2019 · Uploaded by Christina Nunez

Search is a type of recommendation problem, where relevance isn't just about query matching.

Given so many options to return, search companies rank results by estimating what a user is most likely to engage with

RECOMMENDATIONS DRIVE

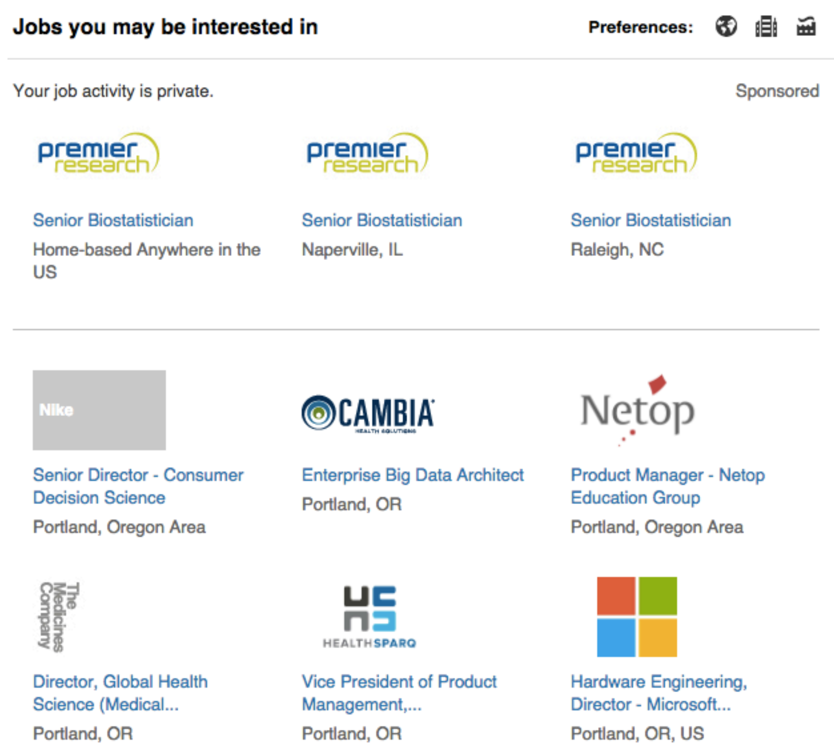
Who gets to be your friend…

Social networks are presumed to be better (and more sticky for the user), the more connections you have.

Therefore, much of their growth strategy relies on encouraging more and more connections, often using recommender systems

RECOMMENDATIONS DRIVE

Where you might work …



Job recommendations are common in many systems. This is where the influence of recommendations on society need to be more carefully examined.

Recommendation systems may learn biases that already exist in the job market data, and through their scale can serve to perpetuate these biases.

RECOMMENDATIONS ARE VERY INTERESTING

There is no single technique, and each problem is unique, though there are some core fundamentals

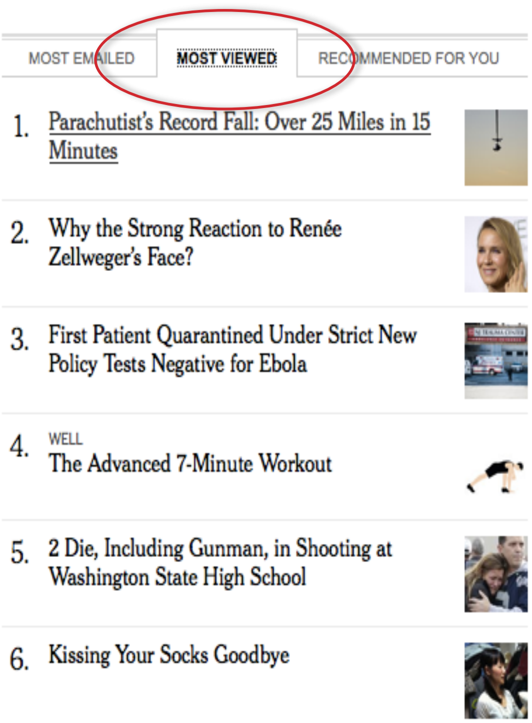
Evaluation is not obvious, requires creativity and ingenuity

Very few machine learning products are this exposed to the public.

- What are the design implications of a recommender system?
- What are the ethical implications of a recommender system?

TWO PHILOSOPHICAL APPROACHES

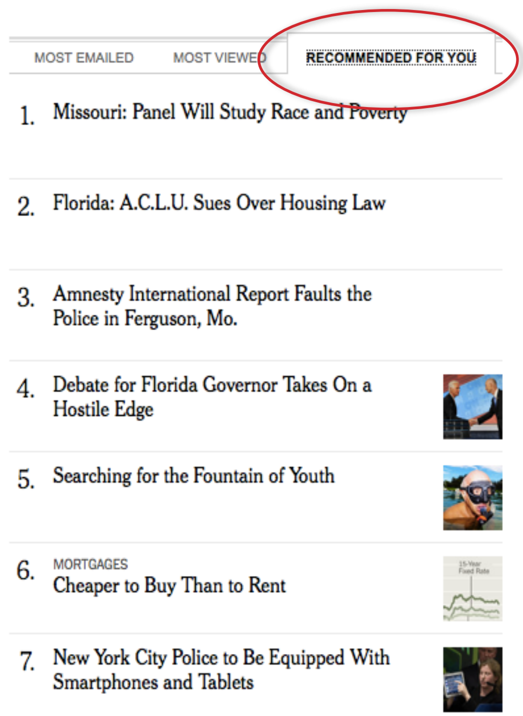
Global:
Recommend what is popular



MOST EMAILED | MOST VIEWED | RECOMMENDED FOR YOU

1. Parachutist's Record Fall: Over 25 Miles in 15 Minutes
2. Why the Strong Reaction to Renée Zellweger's Face?
3. First Patient Quarantined Under Strict New Policy Tests Negative for Ebola
4. WELL
   The Advanced 7-Minute Workout
5. 2 Die, Including Gunman, in Shooting at Washington State High School
6. Kissing Your Socks Goodbye

Personalized:
Try to understand your tastes



MOST EMAILED | MOST VIEWED | RECOMMENDED FOR YOU

1. Missouri: Panel Will Study Race and Poverty
2. Florida: A.C.L.U. Sues Over Housing Law
3. Amnesty International Report Faults the Police in Ferguson, Mo.
4. Debate for Florida Governor Takes On a Hostile Edge
5. Searching for the Fountain of Youth
6. MORTGAGES
   Cheaper to Buy Than to Rent
7. New York City Police to Be Equipped With Smartphones and Tablets

Realistically, one doesn't have to choose a single approach. Products can offer users a choice between global and personalized, or global and heuristic methods can be used as initial baselines to make the system operational and to collect more data for future ML based methods.

METHODS FOR PERSONALIZATION IN RECOMMENDER SYSTEMS

Collaborative Filtering

Recommend based on user or item similarity, with similarity measured across vectors in the user-item matrix

Matrix Factorizations

Recommend based on similarity between user and item embedding vectors, where the embeddings are vectors pulled from a factorized user-item matrix

Neural Rec Sys

Recommend based on learned non-linear embedding combinations, learned through back propagation in neural networks

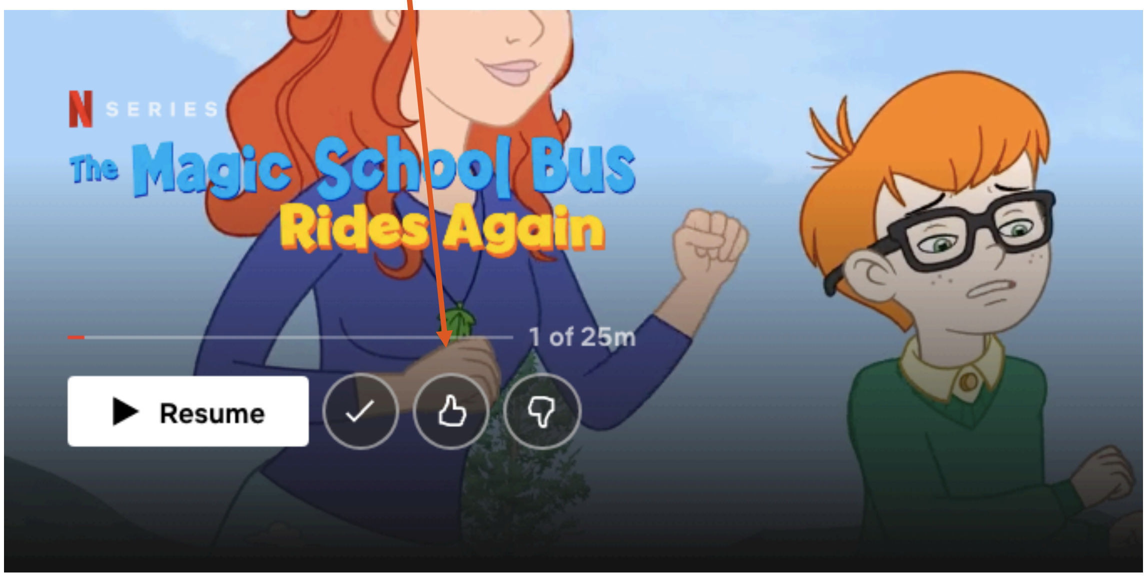
GOING BEYOND TRADITIONAL MACHINE LEARNING

With recommendation systems we have to approach many aspects of the ML process with new tools and considerations

Data Collection	Modeling	Evaluation	Humans Impact
• Implicit vs Explicit feedback • Ranking bias • Exploration vs exploitation	• Need to choose loss functions • Models not supported by common ML packages (SkLearn)	• New, information retrieval specific metrics • Better served with counterfactual evaluation	• Scale, ubiquity and exposure require careful consideration of human impact • User interface of recommendations have large performance impact

LABELING YOUR DATA

Each product will have different label considerations. Ratings, thumps up/down, purchases & conversions are considered an ***explicit*** type of feedback.

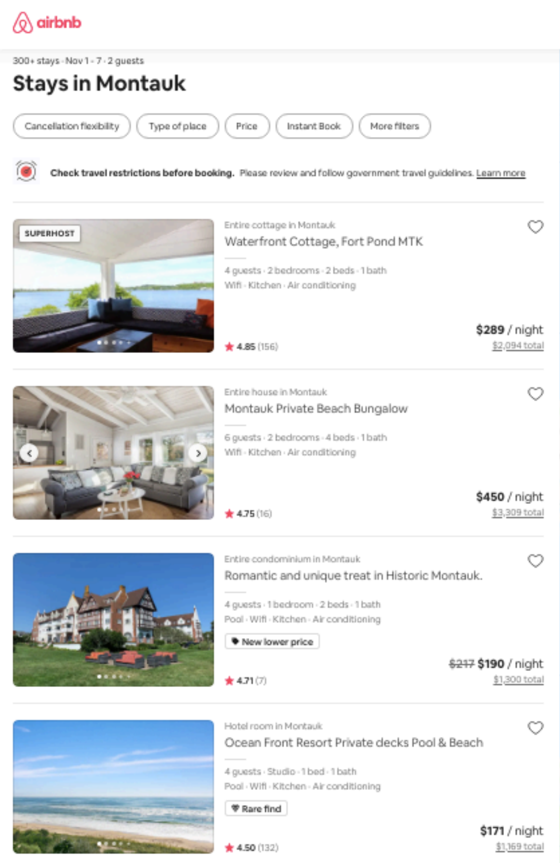


Consumption of an item (particularly media) is also feedback that can used in a RecSys. This is called ***implicit*** feedback.

Usually explicit feedback is better for understand a user's true taste, but implicit feedback is often more abundant due to the ease of collecting it.

BIAS IN YOUR DATA COLLECTION

Operational data will have two types of bias to consider:



Rank bias: items higher up in the rankings are likely to get more engagement, independent of how relevant the items are. When training off of this data, it is important to correct for the rank/position bias.

- Can randomize top K results to average out rank effect
- Can use rank as a feature and control for this

Exploitation bias: we typically recommend items that our model considers relevant. We only receive feedback on the items shown, so we risk only being able to learn (and re-learn) on items we've already ranked highly. To unbias our data we need to perform more exploration (showing less relevant or newer items) with users.

METRICS – PRECISION VARIANTS

Very often we consider precision over recall in recommendations. It is common to have many relevant items so our emphasis is in having the small set of displayed items be relevant, as opposed to recovering all relevant documents

Precision@K

For a given query/recommendation…

Set a rank threshold ***k*** → Compute % relevant in top ***k***

Notes:
- Allows for multiple relevant items
- Relevance is binary
- Ignores items past rank k
- Need to aggregate across queries/recommendations

Mean Avg Precision

Compute rank position for each relevant item → Compute Precision@K for each relevant item → Average Precision for each query/rec = mean(Precision@K) across relevant items → Mean Avg Precision = mean(Average Precision) across queries / recommendations

METRICS – OTHER

Other common metrics are intuitively similar to precision, in that they measure how well more relevant documents are ranked higher than non-relevant documents

Mean Reciprocal Rank

For a ranked list of items

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

Notes:

- Usually $rank_i$ taken at first relevant document
- Mean is taken over all queries/recs of size |Q|
- With implicit feedback, only applicable to recs/queries that had an engagement (click, conversion, etc.)

Discounted Cumulative Gain

For a ranked list of items, at a given rank *p*

$$\text{DCG}_{\text{p}} = \sum_{i=1}^{p} \frac{rel_i}{\log_2(i+1)}$$

Notes:

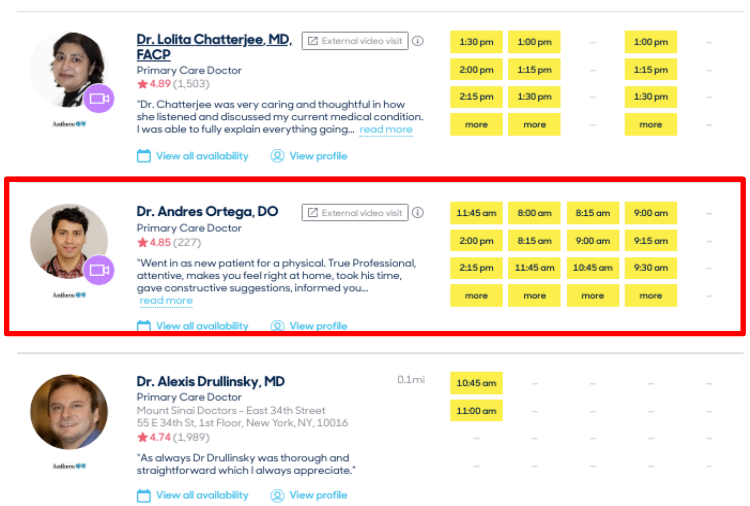
- This metric is used typically when we have non-binary relevance (like star rating)
- Can normalize by dividing DCG by the DCG of a perfect ranker
- This is for single rec/query

COUNTERFACTUAL EVALUATION

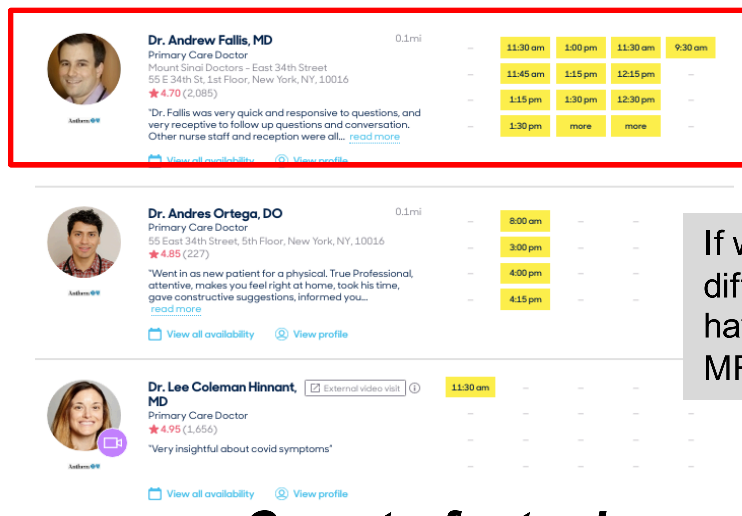
When it comes to model selection comparisons, we often can not evaluate alternative scenarios due to the fact that we don't have exposure and engagement data on user-item pairs that have not been observed in our data

We are only shown these items, so our feedback data is biased to what our previous recommendations considered relevant

What would engagement have looked like if we showed this group instead (which is likely under a different algorithm)



Factual

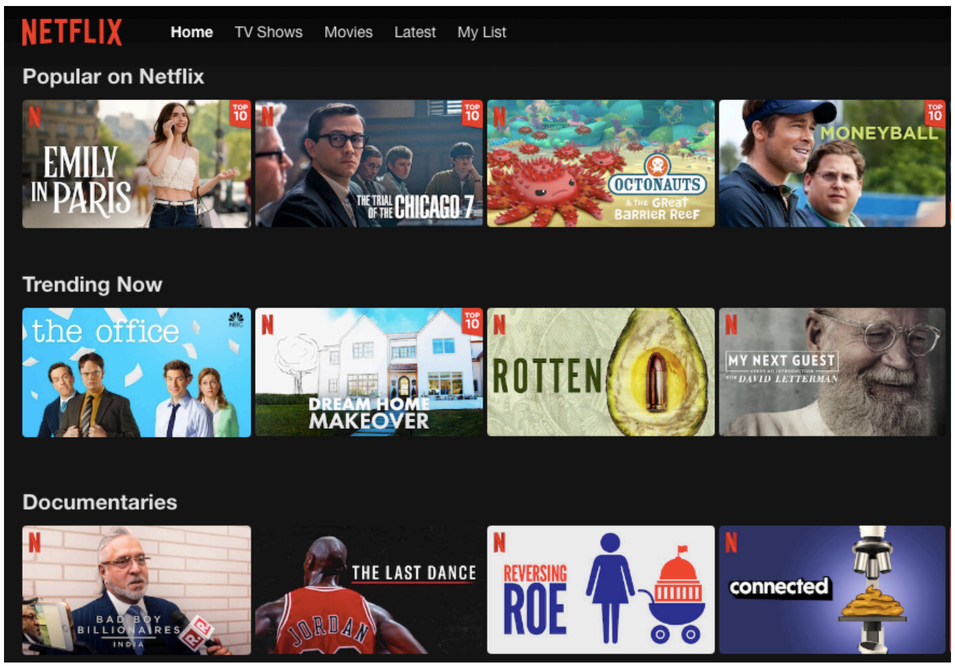


Counterfactual

If we had shown a different set, would we have observed a higher MRR?

RECOMMENDING W/ EMPATHY

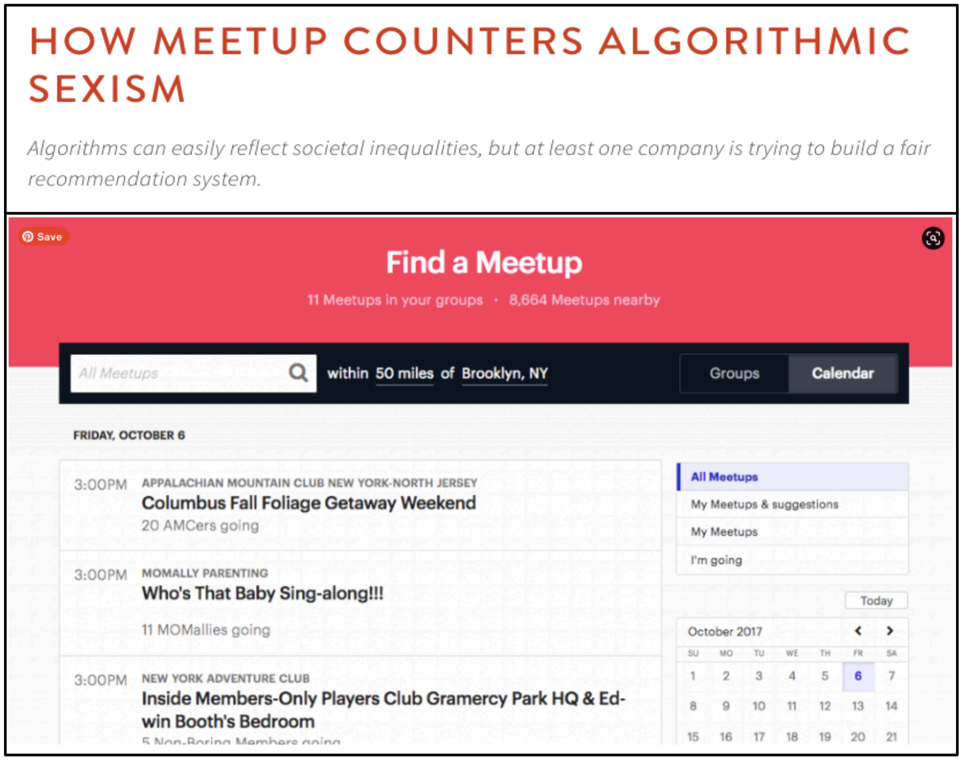
How you present your recommendations is often as important as how you rank your recommendations. Many of the important design and UI elements are outside of traditional DS concerns. Most organizations will have a cross functional team of designers, data scientists and product managers to think and test through an optimal experience.



Things to think about…

- How many items should you expose in one page?
- Should you explain the rationale of your recommendations?
- Is the data you are using in compliance with legal and user expectations?
- How do you design the UI to get better and more feedback?
- Should you retrain your model when the UI changes?
- What is the right balance of exploration vs exploitation

ON BEING FAIR



"If you let the algorithm auto-optimize, it would see that men are more likely to join tech groups than women," *Hodgson said. The site would then learn to recommend tech groups to men more often than women.*

To mitigate discriminatory biases…

- Be careful about what features go into your model, and the interactions between them
- Be aware of proxies for protected attributes
- Bias yourself towards more explainable and auditable systems
- Leverage appropriate testing techniques
- Make sure a diverse team is including in overall evaluation

Source: https://civichall.org/civicist/meetup-counters-invisible-sexism/