

# Apples to Audiences

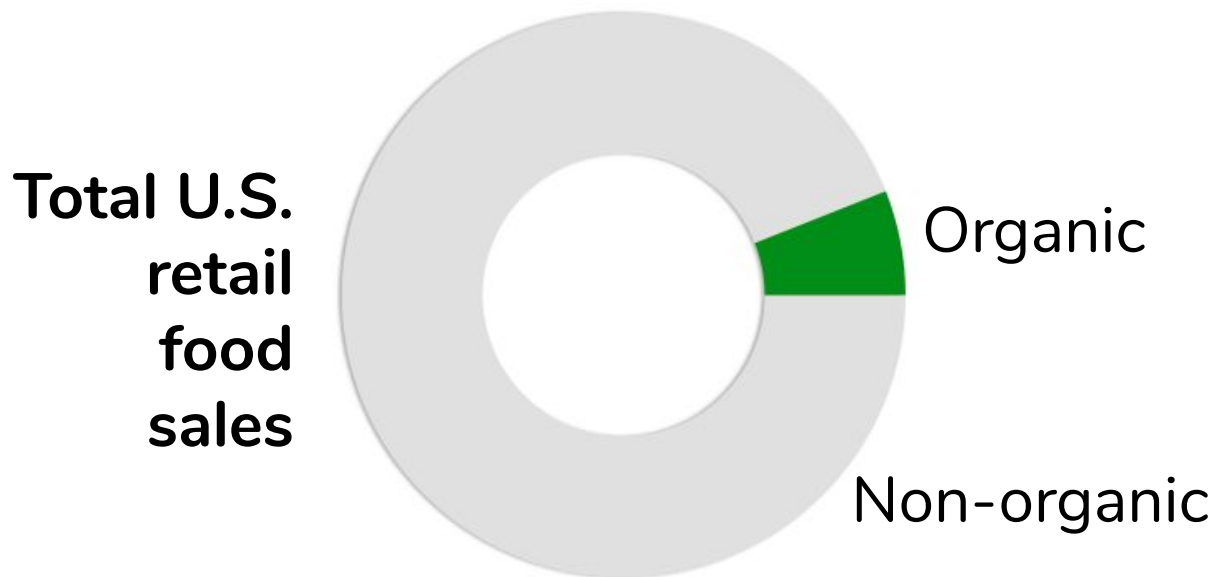
A tool for targeting  
organic produce buyers

Jeff C Ho





# Organic food is **just 6%** of the total U.S. market share



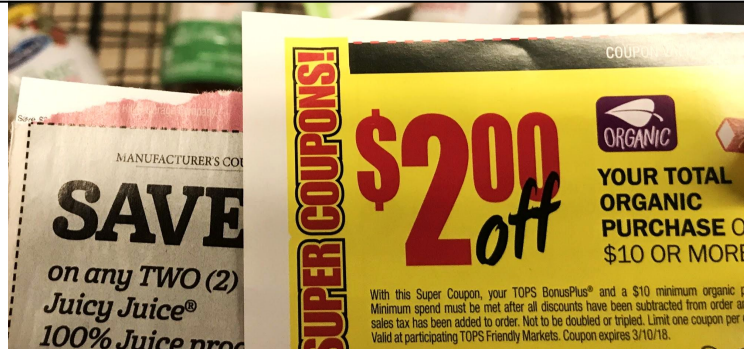


# Existing promotions to grow sales are **not targeted**

not by consumer



not by product

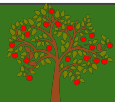
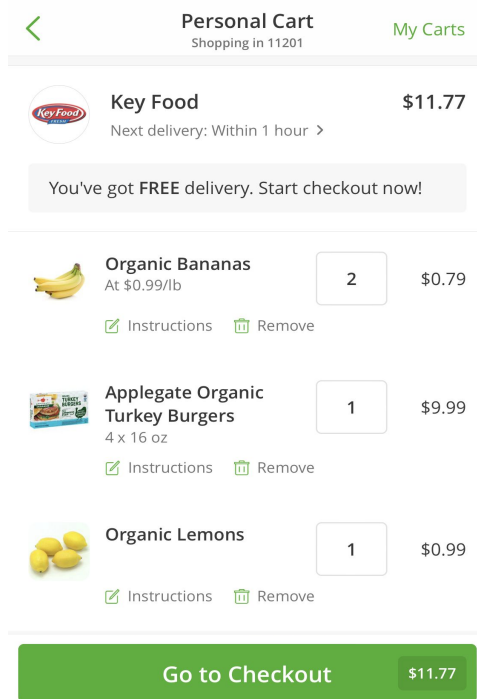




# Can I target individuals to grow sales?

## Use Instacart data:

- 3.4 million orders
- 200k users
- 50k products

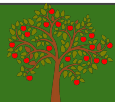
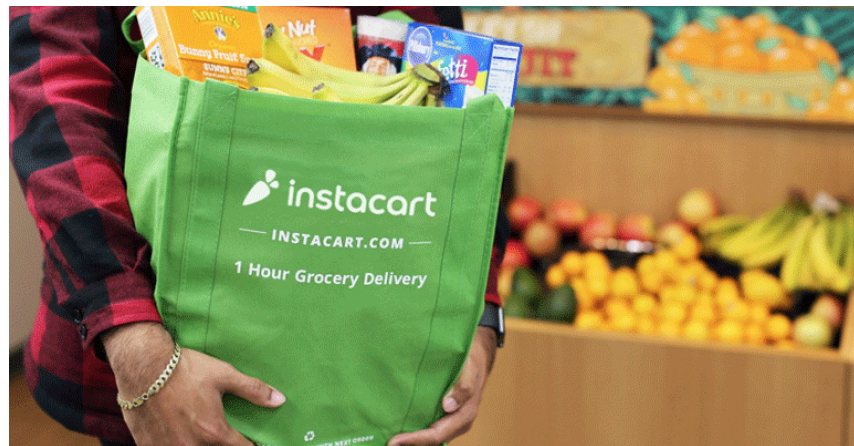




# Can I target individuals to grow sales?

## Use Instacart data:

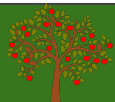
- 3.4 million orders
- 200k users
- 50k products





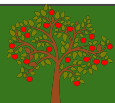
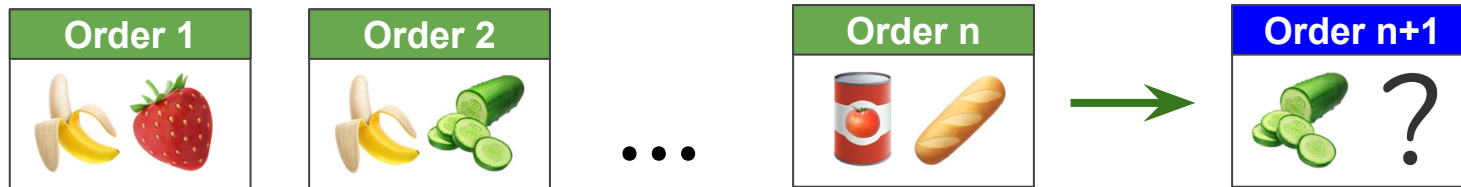
# Can I target individuals to grow sales?

Grow carts of past buyers  
**AND**  
Identify new buyers

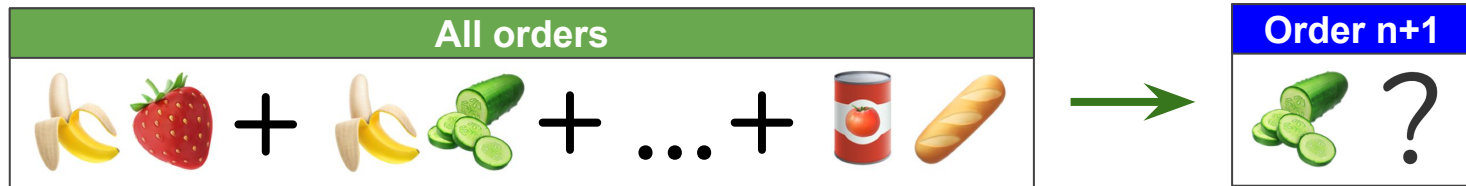




# Predict items in an individual's next order

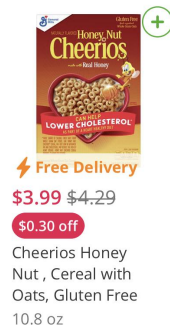


# Use collaborative filtering to surface product predictions



				...		
User 1	2	1	1	...	1	1
...						
User 4314	1	1	?		0	1
...						
User n						

Frequently bought with  
Organic Bananas







# Collaborative filtering results are reasonably good

## User 10 purchased products

Cantaloupe
Parsley, Italian (Flat), New England Grown
Seedless Red Grapes
Small Hass Avocado
Organic Turkey Bacon
Tamari Gluten Free Soy Sauce
Organic Red Cabbage
...



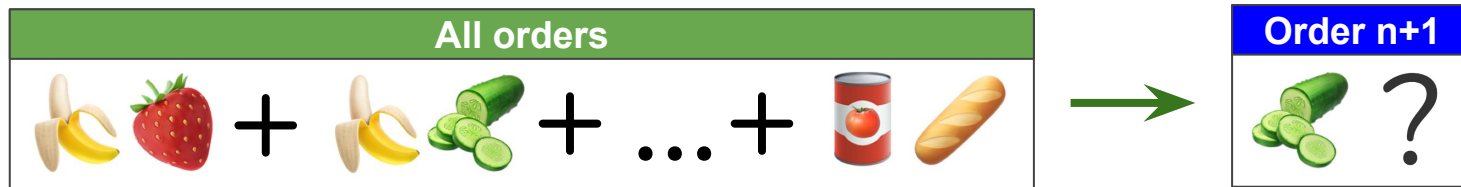
## Top recommended products






Organic Baby Spinach
Organic Strawberries
Organic Garlic
Bag of Organic Bananas
Organic Hass Avocado





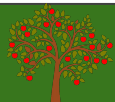
# ...but could still be better



				...		
User 1	2	1	1		1	1
...						
User 4314	1	1	?		0	1
...						
User n						

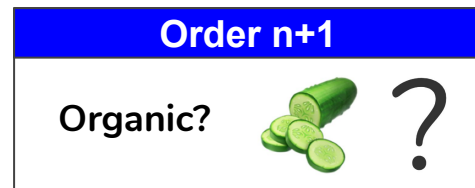


**Just 56% of users  
bought a top  
recommendation**



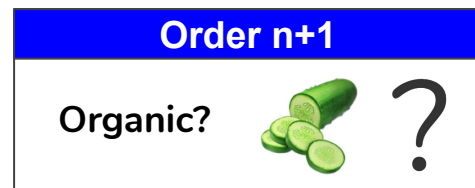


# Solution: Recommendations + buyer likelihood model



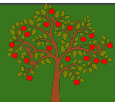


# Solution: Recommendations + buyer likelihood model



Logistic regression based on...	F1 Score
<ul style="list-style-type: none"><li>• Previous organic purchase</li><li>• <b>Percent of recommended products that are organic</b></li></ul>	<b>0.79</b>

Targets products  
toward likely  
buyers





# Let's explore the web app!

➤ [AppletoAudiences.today](https://AppletoAudiences.today)



## A tool for targeting organic produce buyers

- ☒ Target individual users
- ☐ Get likely buyers
- ☐ Get likely buyers by item
- ☐ About
- ☐ Show only users who have not bought organic produce before

Show users with probabilities less than

0.10 0.93 1.00

Select a user to see what products to recommend

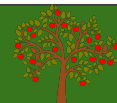
User 13558 (p=0.91) ▼

## Target individual users

User 13558 will next buy organic produce with a probability of 0.91.

- They have bought organic produce before, so recommending new items could grow sales.

	Recommended items	Item popularity rank	New item?
1	Organic Baby Spinach	4	No
2	Organic Garlic	16	No
3	Organic Strawberries	3	No
4	Bag of Organic Bananas	2	No
5	Organic Hass Avocado	7	Yes
6	Organic Baby Carrots	22	Yes
7	Organic Ginger Root	62	No
8	Organic Zucchini	14	No
9	Organic Red Onion	20	No
10	Organic Lemon	25	No
11	Organic Garnet Sweet Potato ...	41	Yes



A tool for targeting organic produce buyers

- ☒ Target individual users
- ☐ Get likely buyers
- ☐ Get likely buyers by item
- ☐ About
- ☒ Show only users who have not bought organic produce before

Show users with probabilities less than

0.10 0.35 1.00

Select a user to see what products to recommend

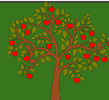
User 77369 (p=0.35) ▼

Target individual users

User 77369 will next buy organic produce with a probability of 0.35.

- They have not bought organic produce before, so recommending popular organic items could get them to buy organic.

	Recommended items	Item popularity rank	New item?
1	Bag of Organic Bananas	2	Yes
2	Organic Strawberries	3	Yes
3	Organic Baby Spinach	4	Yes
4	Organic Blueberries	11	Yes
5	Organic Avocado	6	Yes
6	Organic Raspberries	10	Yes
7	Organic Hass Avocado	7	Yes
8	Organic Fuji Apple	28	Yes
9	Organic Blackberries	35	Yes
10	Organic Lemon	25	Yes
11	Organic Zucchini	14	Yes



## A tool for targeting organic produce buyers

- ☐ Target individual users
- ☒ Get likely buyers
- ☐ Get likely buyers by item
- ☐ About

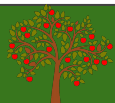
## Get likely buyers

Show users with probabilities greater than



There are **895** users to target. A random sample of at most 20 is shown below:

```
User 39054 <user39054@email.com>,  
User 72565 <user72565@email.com>,  
User 45238 <user45238@email.com>,  
User 64612 <user64612@email.com>,  
User 50168 <user50168@email.com>,  
User 2022 <user2022@email.com>,  
User 10137 <user10137@email.com>,  
User 44986 <user44986@email.com>,  
User 31552 <user31552@email.com>,  
User 61499 <user61499@email.com>,  
User 22842 <user22842@email.com>,  
User 49478 <user49478@email.com>,  
User 75029 <user75029@email.com>,  
User 38291 <user38291@email.com>,
```







# The right organic products for the right consumers

10% new organic produce buyers

= **\$7.2 million in new organic sales** annually

**Reduce spam by >40%** vs no targeting

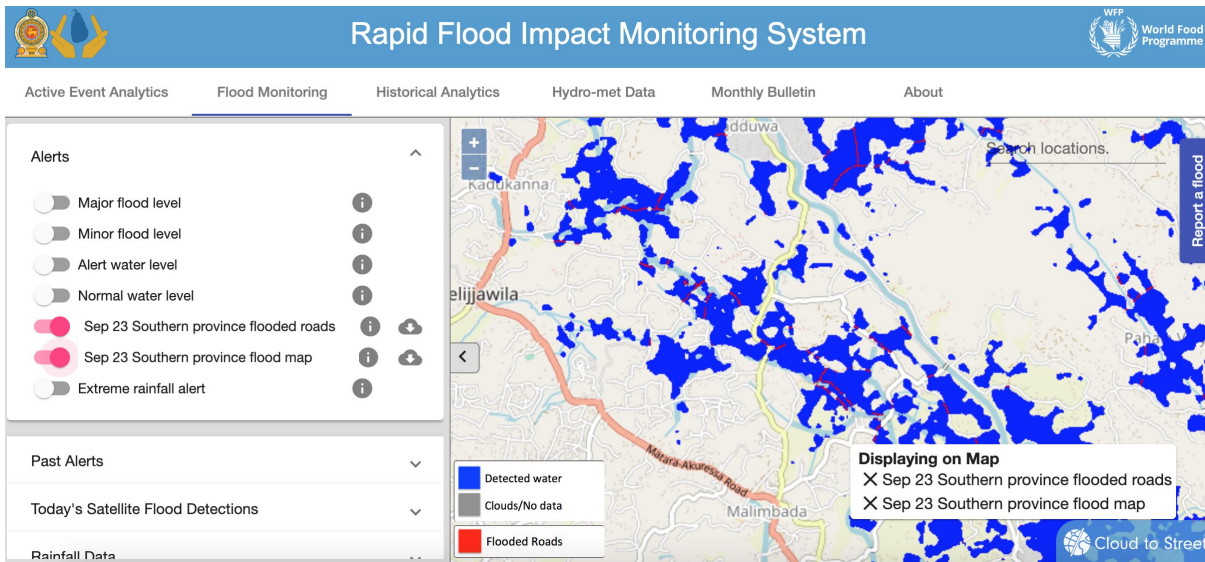




# Jeff C Ho, PhD



## “Building great products with data”



Jeff C. Ho

Apples to Audiences

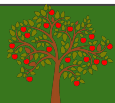


# Collaborative filtering results are reasonably good #2

User 30 purchased products
Clementines
Smartwater
Distilled Water
49 Flavors Jelly Belly Jelly Beans
Tall Kitchen Bag With Febreze odor Shield
Milk Chocolate M&Ms



Top recommended products
Soda
Clementines
Cereal
Packaged Grape Tomatoes
Baby Cucumbers



## A tool for targeting organic produce buyers

- ☐ Target individual users
- ☐ Get likely buyers
- ☒ Get likely buyers by item
- ☐ About

## Get likely buyers by item

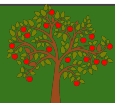
Select an item to get a list of users with likelihood of buying > 0.5 who are recommended that product

Bag of Organic Bananas

There are **1488** users to target who are likely to buy the item [Bag of Organic Bananas].

A random sample of at most 20 is shown below:

```
User 33904 <user33904@email.com>,  
User 941 <user941@email.com>,  
User 70382 <user70382@email.com>,  
User 18641 <user18641@email.com>,  
User 46900 <user46900@email.com>,  
User 30501 <user30501@email.com>,  
User 25362 <user25362@email.com>,  
User 75860 <user75860@email.com>,  
User 55281 <user55281@email.com>,  
User 68502 <user68502@email.com>,  
User 8504 <user8504@email.com>,  
User 17224 <user17224@email.com>,  
User 47007 <user47007@email.com>
```

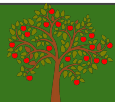




# Next steps: Understanding user segments

**Middle group** could help drive new buyer growth

**High probability group** could help drive old buyer retention





# Next steps: Scaling the app beyond Instacart

1. Translate other product catalogs into Instacart's

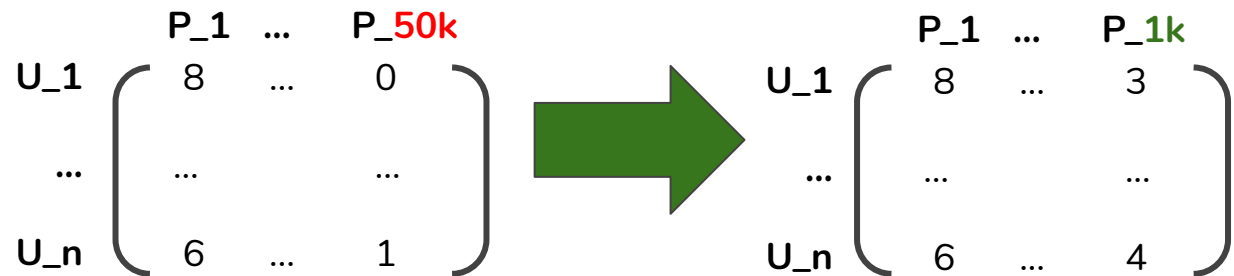


2. Re-train collaborative filter model with new users and unified/translated product catalog
3. Pass new users through pipeline

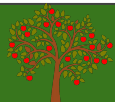


# Next steps: Improving recommendations/predictions

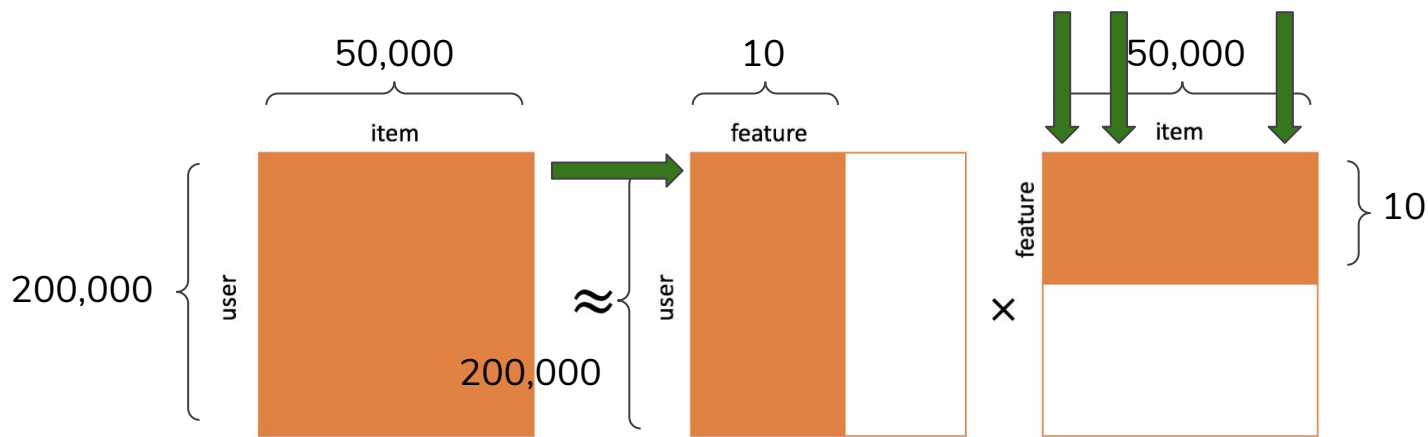
1. Perform matrix factorization using only most frequently-purchased items, rather than full set of 50k items



2. Normalize user-product matrix by frequency of purchase by user
3. Tune the dimension of the latent factors (currently 10)



# Non-negative Matrix Factorization



**Recommendations** for each user vector are most similar item vectors for each user based on cosine similarity:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$



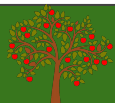
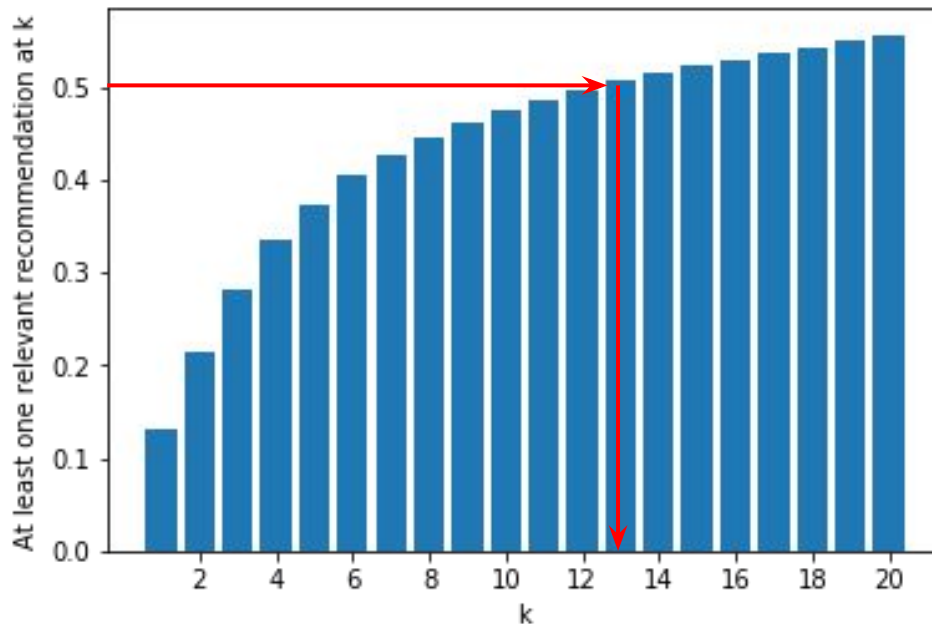




# At least one relevant recommendation at $k$

Over 50% of users with a relevant product purchased at **13** recommendations.

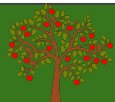
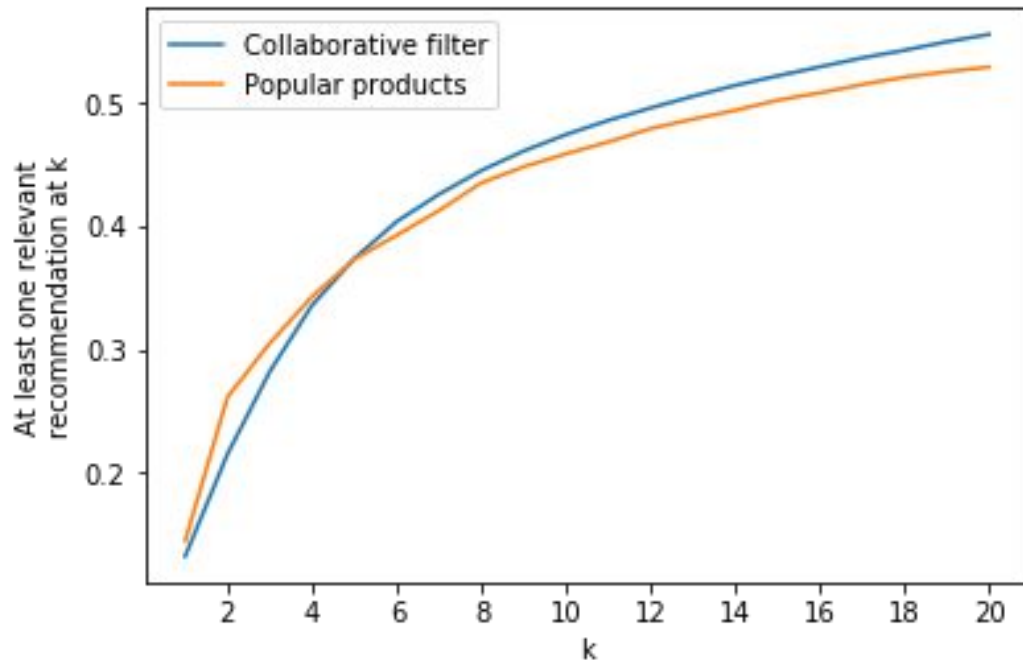
Precision at  $k$  was lower, between **0.06** and **0.13**



# 1-4% better than recommending popular products

Collaborative filter takes over >5 recommendations.

2000-8000 people with better recommendations.





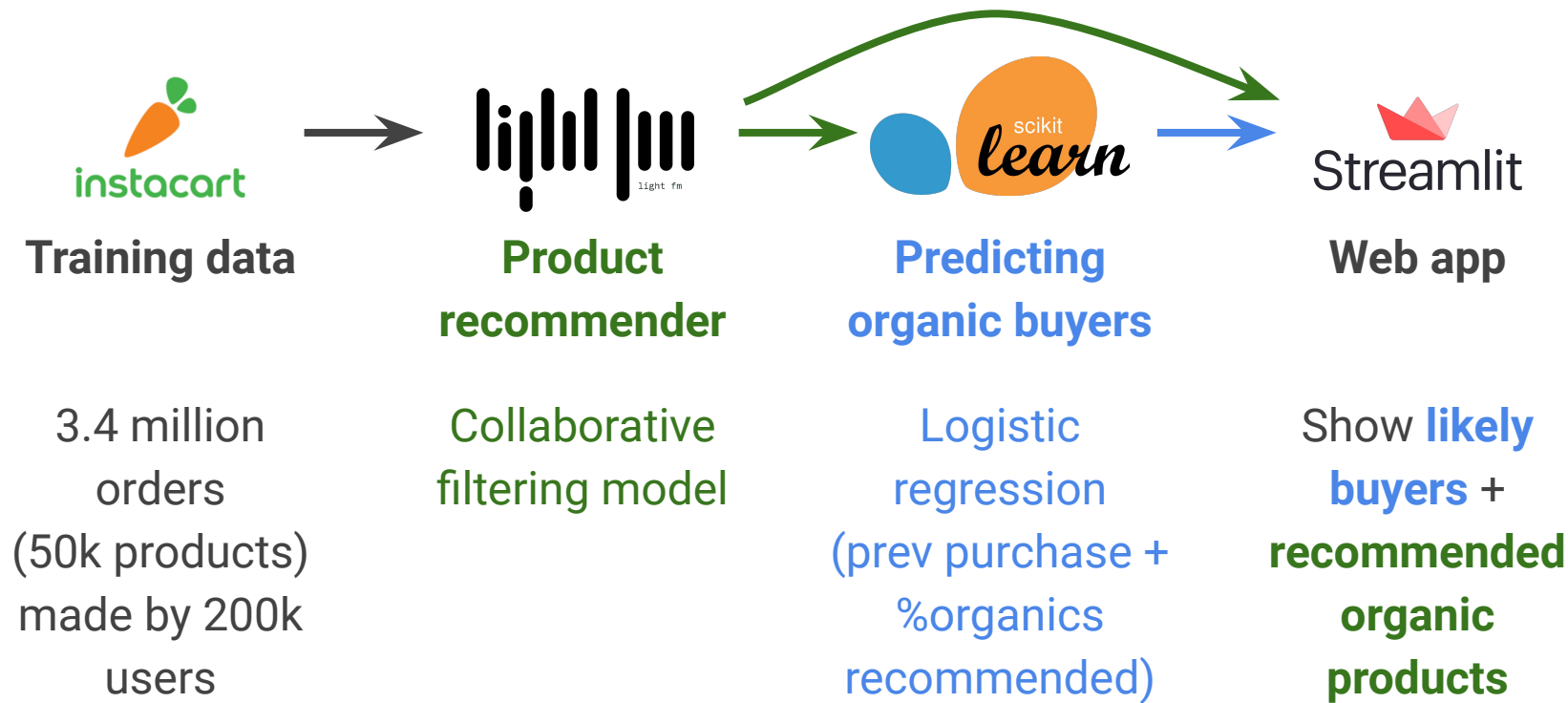
# The value of targeting

Targeting model **identified 10% new organic produce buyers**,  
As a percentage of all users, 6% of them would be new.

- On average, households spend \$30 on organic produce/year (Nielsen, 2016) or \$3 per order based on the Instacart data
- Instacart has ~4 million users, and 240k of them could be identified with the model
- $\$3 \times 240\text{k users} = \$720\text{k in sales per next order,}$   
or **\$7.2million per year for organic producers.**
- Based on a 5% service fee, meaning targeting would be worth \$36,000 per order or **360,000 per year to Instacart.**



# Full product pipeline





# Feature engineering

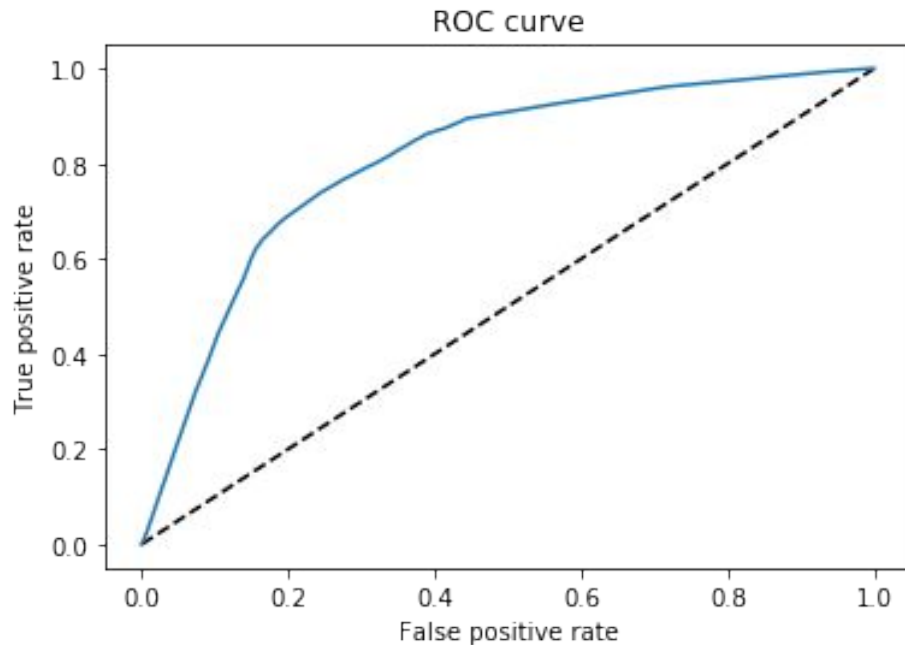
	Original variable(s)	Engineered feature(s)
User's next order	Order day of week	Weekend or not
	Order time of day	Morning or not
User's most recent previous order	Days since most recent previous order	Categories of days: [0-4], [5-8], [8-15], [15-22], [22-30]
	Number of previous orders	--
	List of products	Whether organic produce was purchased
		Departments (e.g., frozen) of purchased products
User's historical order history	Recommended products from collaborative filtering	Principal components of departments
	List of products/frequency	Percent of products that are organic
		Whether organic produce was purchased





# ROC curve of logistic regression model

Classifier performing better than random





# Who is growing the organic market?



- National Organic Trade Association represents more than **9,500 organic business in 50 states.**
- Members include: growers, shippers, processors, certifiers, farmers' associations, distributors, importers, exporters, consultants, and retailers.

