instacart

Customer Retention

Business Analytics Spring 2020

Nicola Kornbluth | Daniel Moskowitz | Sanchya Sahay | Metika Sikka

Today’s Presentation

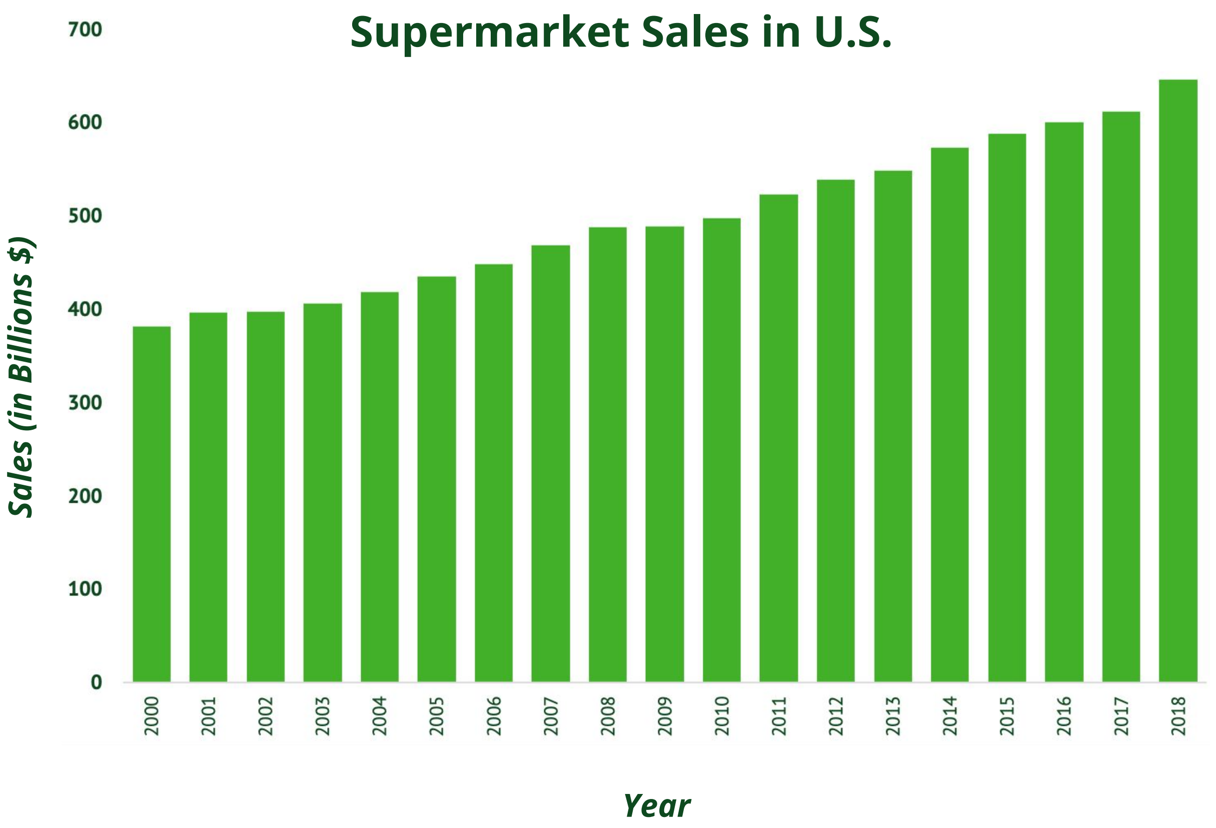
- **The Grocery Industry**
- **The Problem: Driving Retention & Lifetime Value Growth**
- **Our Models + Recommendations**
  - Time Between Orders
  - Reordering Behavior

The Grocery Industry

DISCLAIMER

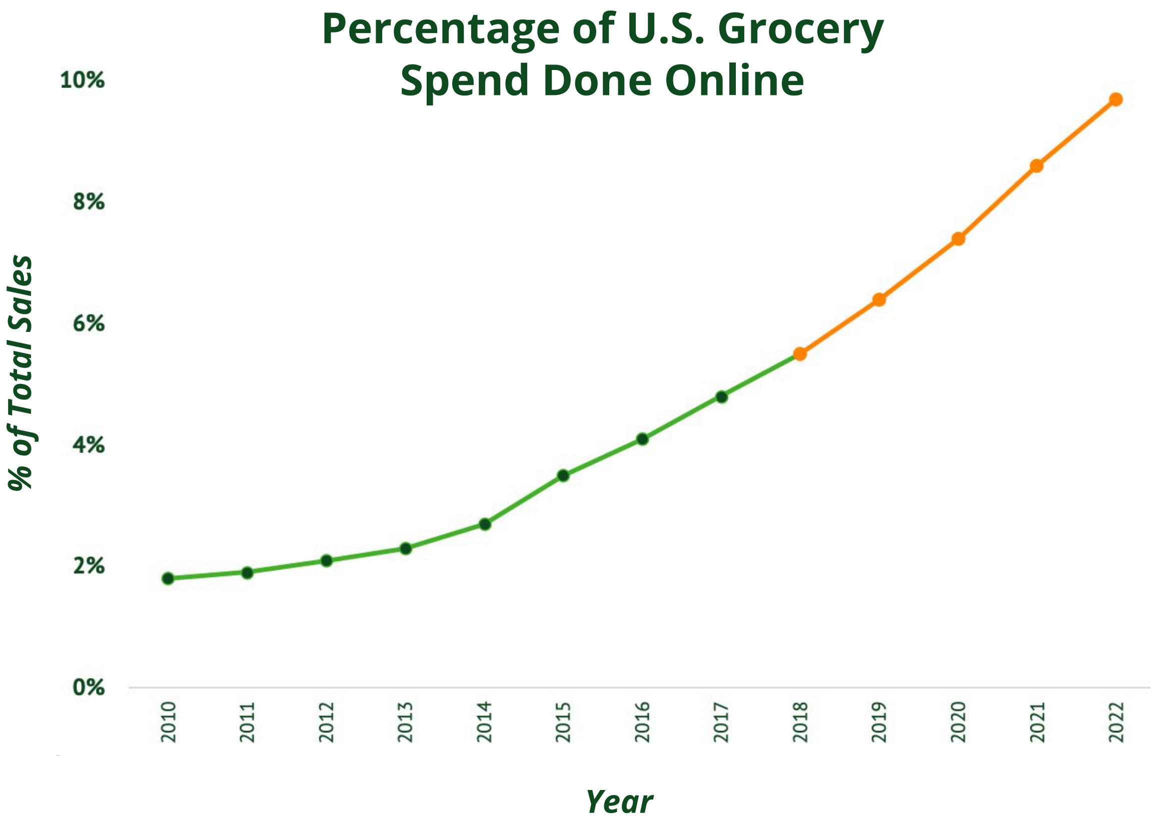
Shopping behaviors studied here are <u>not</u> applicable during COVID-19

Industry Growth



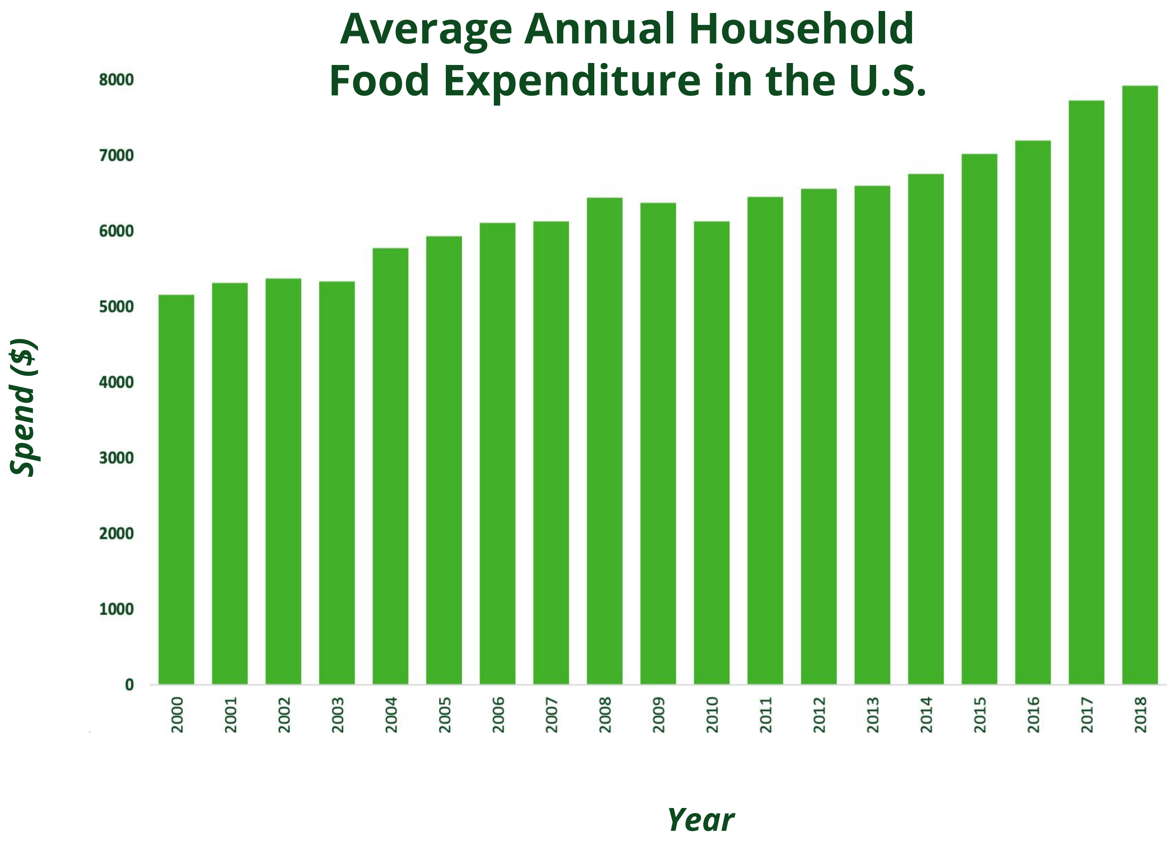
Source: Statista

Industry Growth



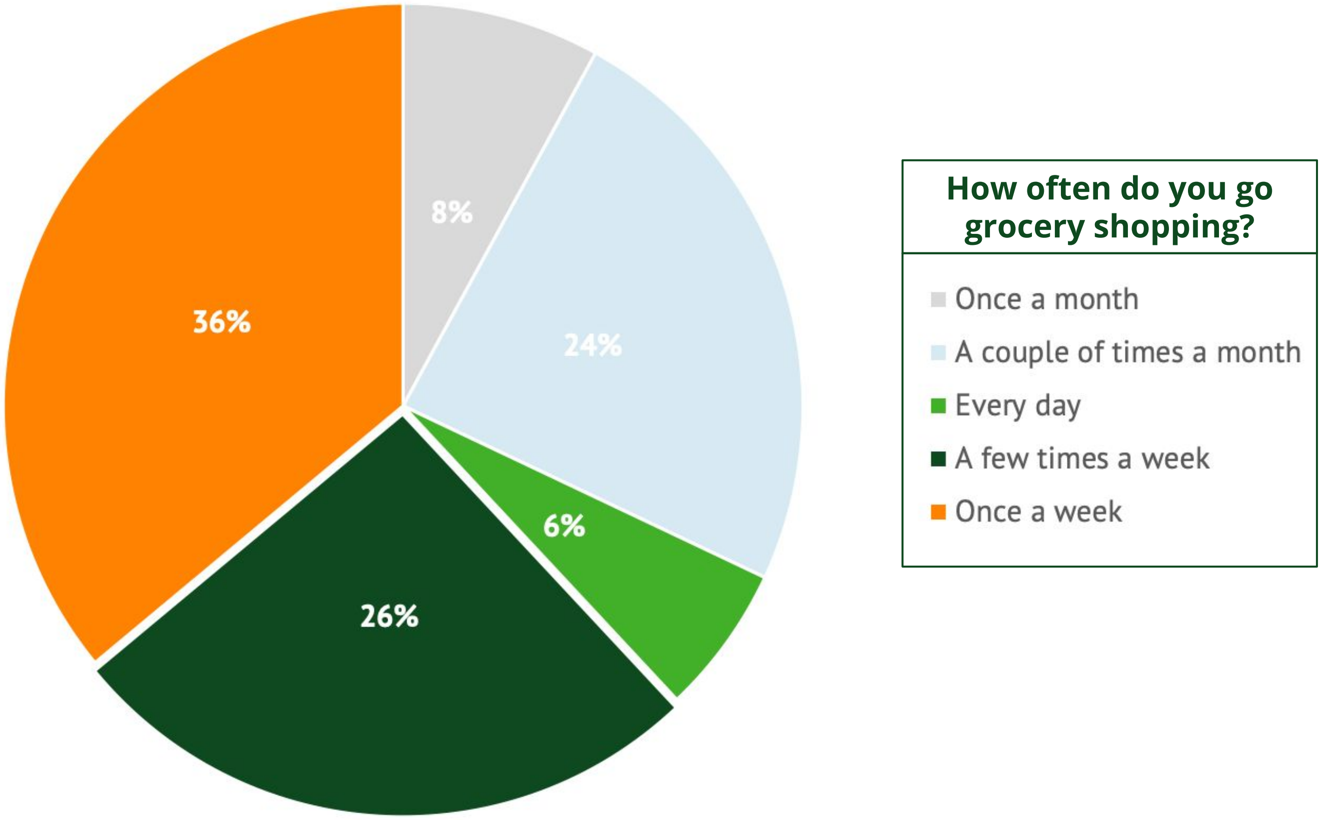
Source: OneSpace

Individual Spend



Source: Statista

Individual Frequency



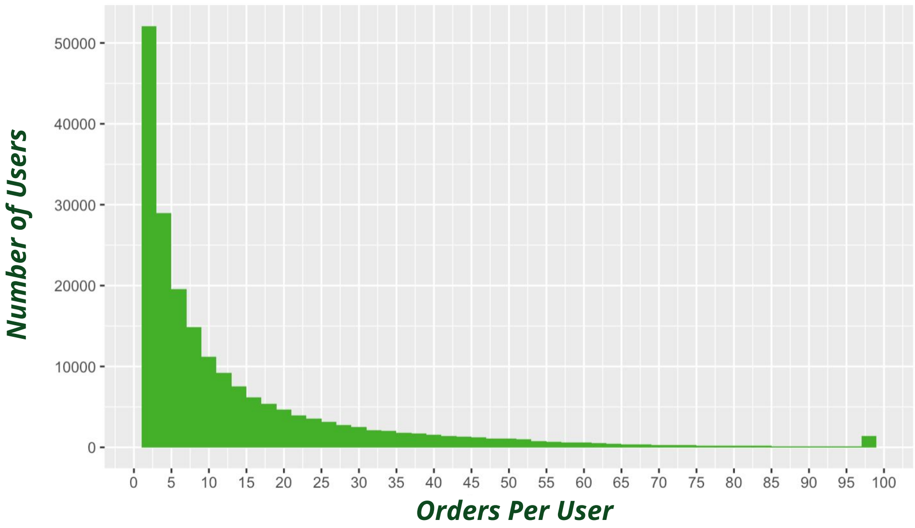
Source: Brandon Gaille

Our Focus:
Drive Retention & Lifetime Value

Instacart Customer Behavior

User Value

User Order Behavior

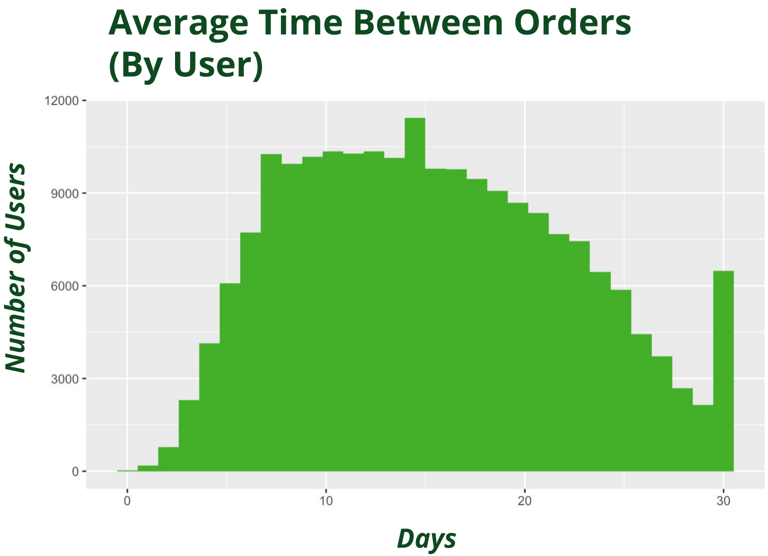


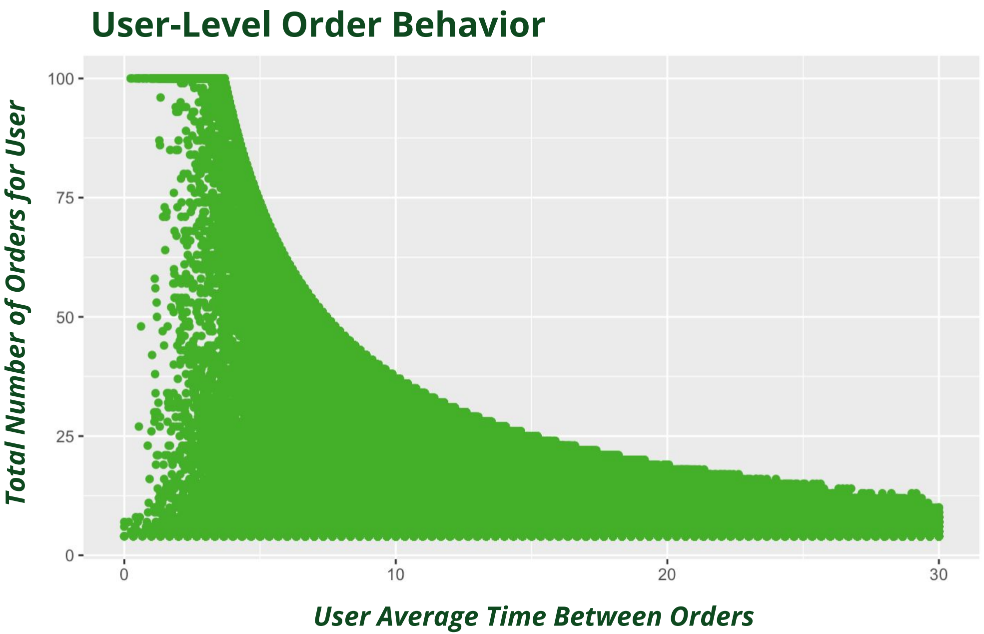
Most Purchased Products

Product	***Reorder Rate***
Bananas	83%
Bag of Organic Bananas	70%
Organic Strawberries	78%
Organic Baby Spinach	78%
Organic Hass Avocado	84%
Organic Avocado	68%
Large Lemon	83%
Strawberries	80%
Limes	70%
Organic Whole Milk	76%

Instacart Customer Behavior

Time Between Orders



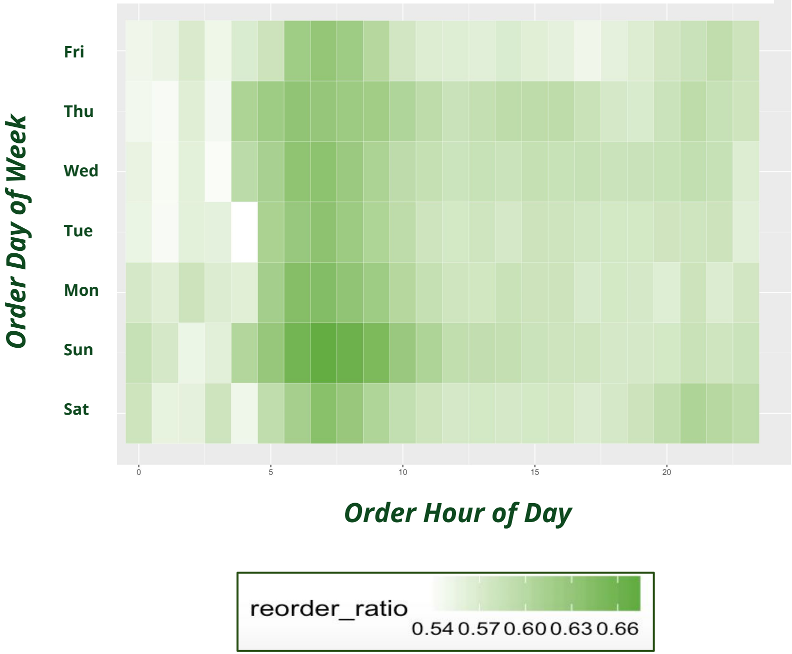


	< 10 Days	11-16 Days	>16 Days
Avg # of Orders	30.5	14.8	8.2

Instacart Customer Behavior

Reorder Behavior

Reorder Ratio by Order Time



Add to Cart Order by Reorder Tendency

User Reorder Rate	< 30%	30-49%	50-74%	> 75%
Avg # of Orders	5.6	10.7	23	43.8

Our approach

Key objectives & impact:

Predict order timing

>> Find drivers of frequent shopping to try increasing order frequency

>> Well-timed retention marketing ("win back" for lapsing customers)

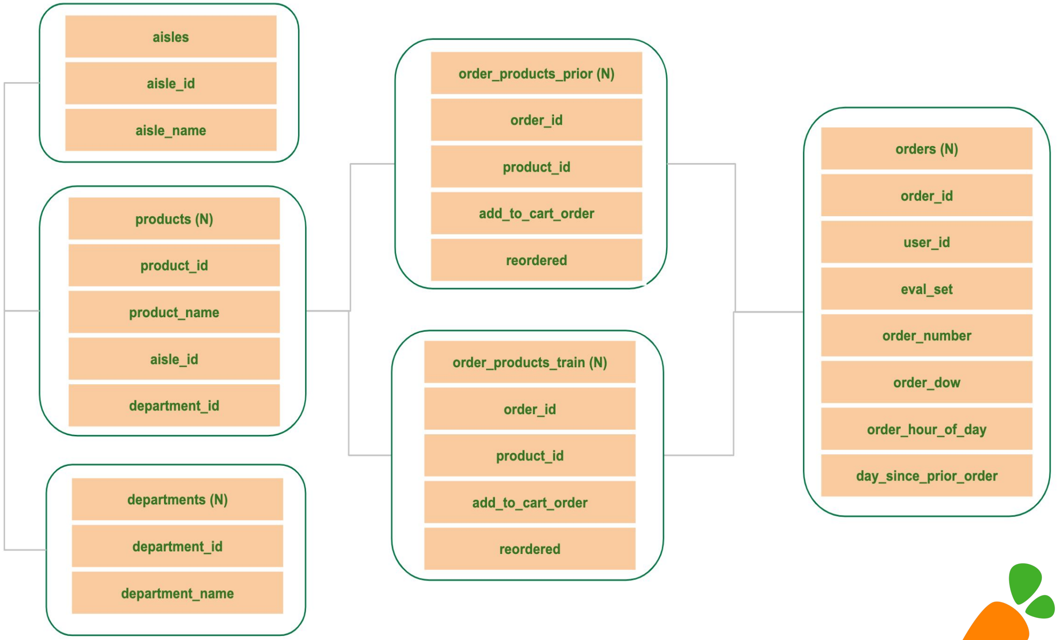
Predict product reordering

>> Increase loyalty and develop 'habits' by marketing high reorder products

>> Maximize profits by increasing frequency of returning customers

>>Improve consumer experience with "easy-ordering" on the platform

Instacart dataset:



Predicting Order Timing

The Model

Feature Engineering

1 — Previous Order Aggregate Metrics

- # of Products
- # of Departments
- # of Aisles
- % from Each Department

2 — Shopper Type "Flags"

- Kitchen Supplies
- Beauty
- Health
- Cleaning
- Junk Food

3 — Cumulative Behaviors

- Average Time Between Past Orders
- Total Previous Products Bought
- Total # of Previous Orders

Model Exploration

- Linear Regression
- Decision Tree

Target Variable: # of days between each order

What We Learned

Drivers of *longer* time between orders

- Longer user-level average time between orders
- Previous order in the evening/night (6pm-5am)
- % of all previous purchases that was cleaning products or kitchen products
- % of previous order that is “long-lasting” - canned goods, frozen goods, pasta/dry goods
- % of previous order that is pet food

Drivers of *shorter* time between orders

- Number of previous orders
- Higher rates of reordered products
- Smaller order sizes (fewer products)
- Products from more departments in previous order
- Previous order made in the morning (6am-12pm)
- Previous order made on any non-Saturday day
- % of all previous purchases that was junk food

Key Takeaways

- Focus on driving weekday morning orders
- Eliminate minimum order size for first few orders to build habit
- Try to introduce users to new-to-them departments
- Feature impulse purchase products in marketing materials

Predicting Product Reorders

The Model

Feature Engineering

1 User level metrics

- Average & mode size of the orders
- No. of times the user continuously ordered new products
- Mode shopping hour of day and day of week
- Mode of no. of days since prior order

2 User-Product level metrics

- Reorder tendency (# of reorders/ # of orders)
- Mode of add to cart order
- Was the product in the user's penultimate cart
- Count of streaks of ordering the product by the user

3 Product level metrics

- Overall reorder ratio of the product
- Department associated to the product

Model Exploration

Log-Lasso Regression

Decision Tree
Random Forest
XG Boosting

Data Map: Utilized User-Product level data with **8.5 mn rows and 27 features**
Target Variable: Products from prior orders that is present in the latest order

Model Comparison - Test Set Performance

Models	Accuracy	Precision	Recall	F1 Score	AUROC	AUPRC
Log-Lasso Regression	0.91	0.62	0.16	0.24	0.82	0.38
Decision Tree	0.90	NA	0.00	NA	0.76	0.28
Random Forest	0.91	0.61	0.21	0.32	0.82	0.42
Random Forest (adjusted for imbalanced classes)	0.87	0.36	0.51	**0.43**	0.82	0.40
XGB Tree	0.87	0.36	0.51	**0.43**	0.82	0.41

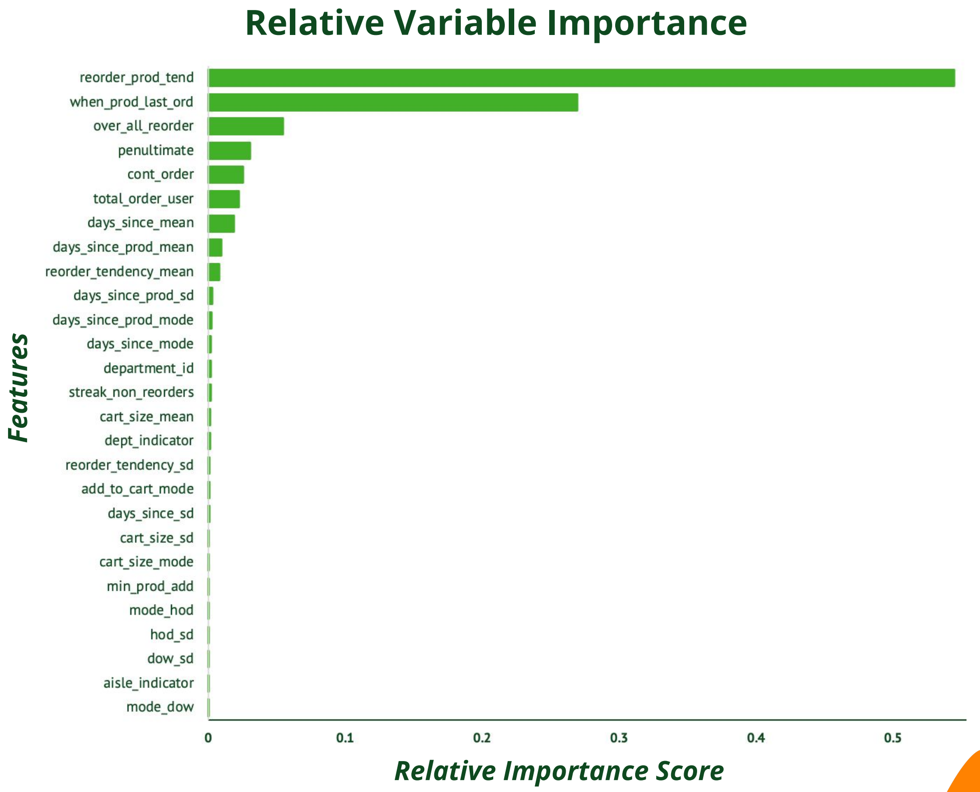
What We Learned

Features that *drive* reordering

- User tendency to reorder that product
- Overall reorder tendency of the product
- Mean Reorder tendency of the user
- Was the product in the user's penultimate cart?
- Number of times the user continuously ordered the product

Features that *deter* reordering

- Days since the product was last ordered
- Total orders of the user
- Mean of the intervals at which the product was reordered



Use Cases for Predictions

- Create subscription-based model for products prone to reordering
- Assist partner stores with inventory management
- Send promotions for most reordered & complimentary products
- Suggest high-margin substitutes for frequently reordered items

Questions?