Online Grocery (Re)ordering

predicting reorders & restocking from Instacart data

Joel Mott

Flatiron School

Project goals

Present three business recommendations to help **promote new or infrequently purchased products.**

Based on which aspects of an **Instacart** dataset have the strongest correlations with popular, reordered groceries.

These correlations may also help inform:

- restocking
- ways to simplify reordering
- predictions of future orders

Business Context: Grocery Delivery & Curbside-Pickup

Online grocery delivery became mainstream in the mid-2010's through third-party service companies such as **Instacart** and **Shipt**.

During/after the pandemic, many US grocery chains started their own service:

Besides Amazon Fresh, these stores also offer **curbside-pickup service**.

Data Overview

Instacart uploaded this dataset to Kaggle.com for a competition in 2017

data contains over three million records with the following:

- order information
  - product names
  - whether that product is a new item for the user or a reorder
- when orders were placed
  - time of day
  - day of the week
  - number of days since the previous order

Project Overview

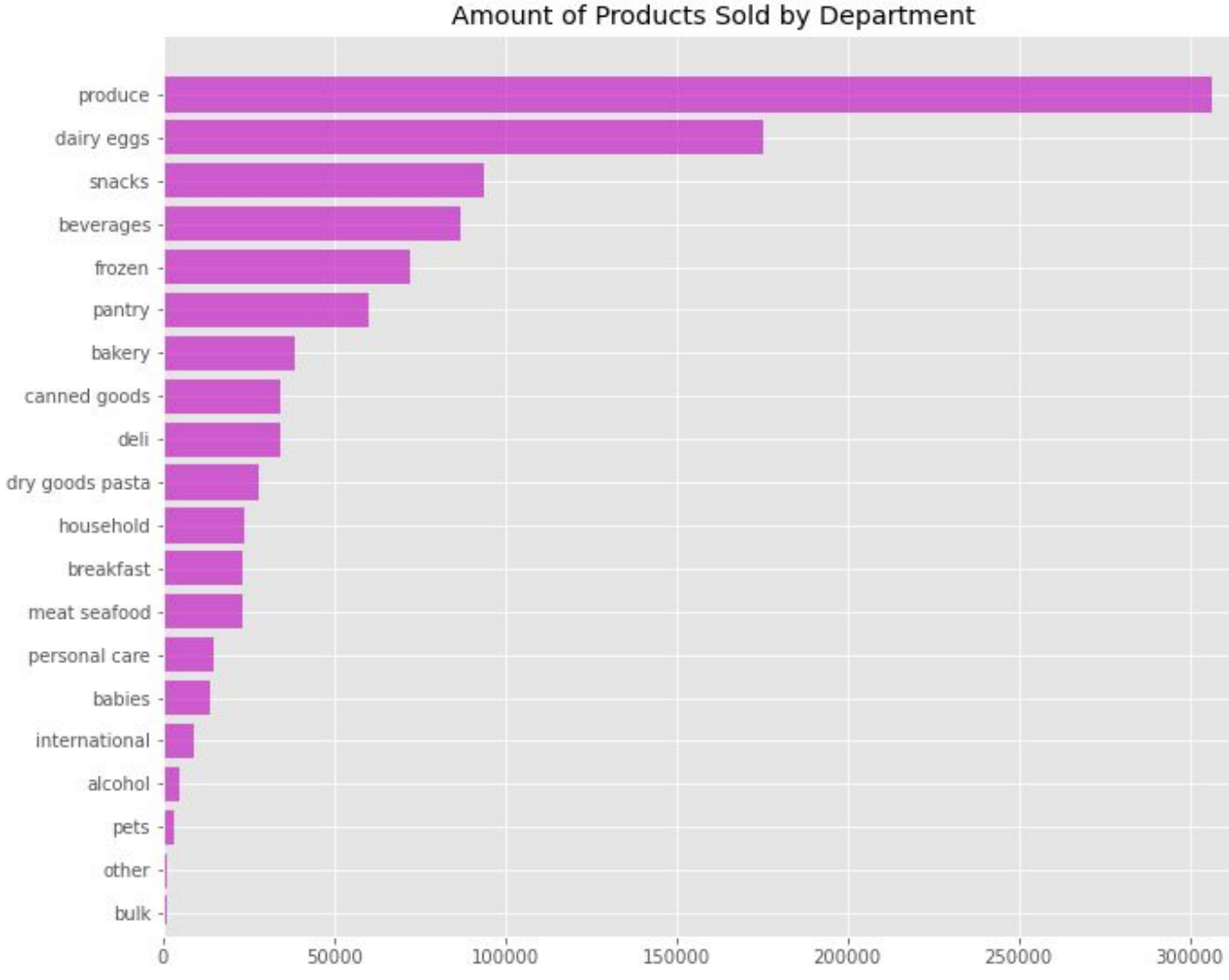
- preliminary analysis
  - popular products & departments
  - when & how often orders are made
  - how products are added to the order
- linear regression analysis
  - isolate different aspects of the data
  - specify which correlate the most with reordering
- business recommendations on promoting new items

Preliminary findings: Popular departments

Produce dominates the other departments.

Dairy/eggs are a distant second, but still far from third.

These two departments sell oft-used products with relatively shorter shelf lives.

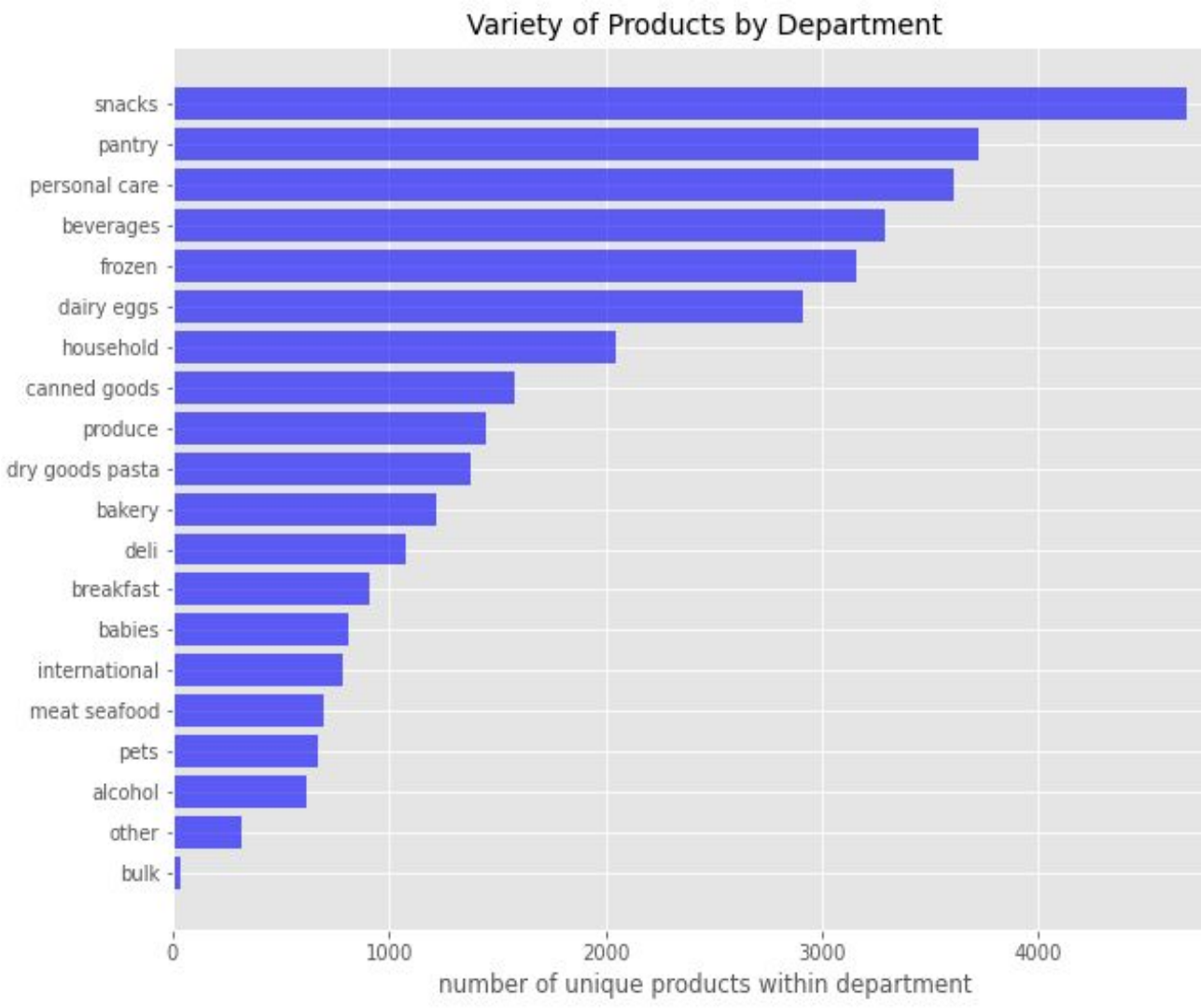


Department size by variety

Produce and dairy/eggs aren't necessarily the *largest* departments.

In fact, they sell fewer unique products than several others.

Scrutinizing the variety of stock for these popular, fresher, short-lived products seems especially vital.

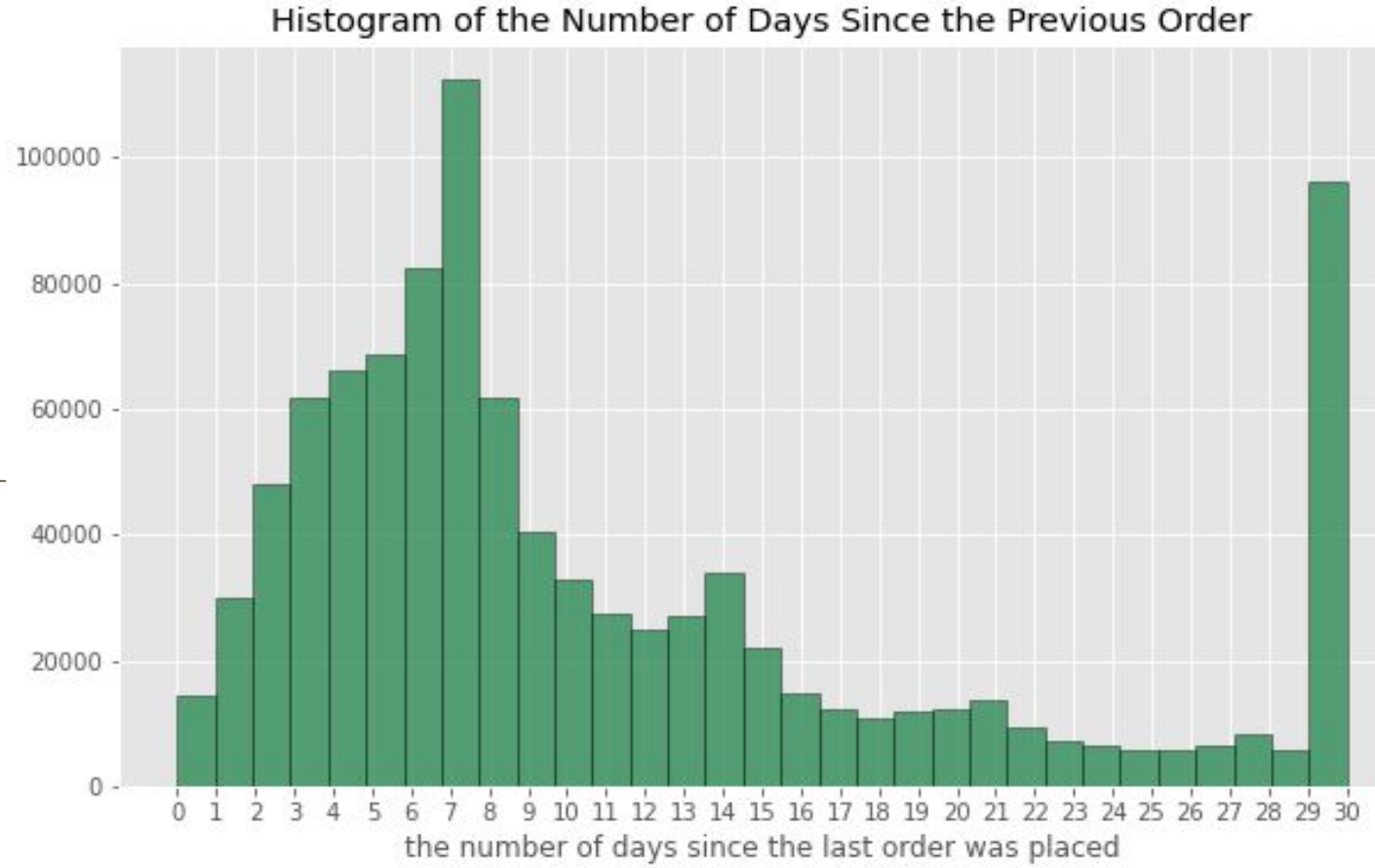


Frequency of orders shows a weekly trend

Many order groceries more often than once a week.

Largest spike is **every seven days** (and then smaller spikes every week thereafter).

"30" days is almost certainly a placeholder for "30 or more days between orders".



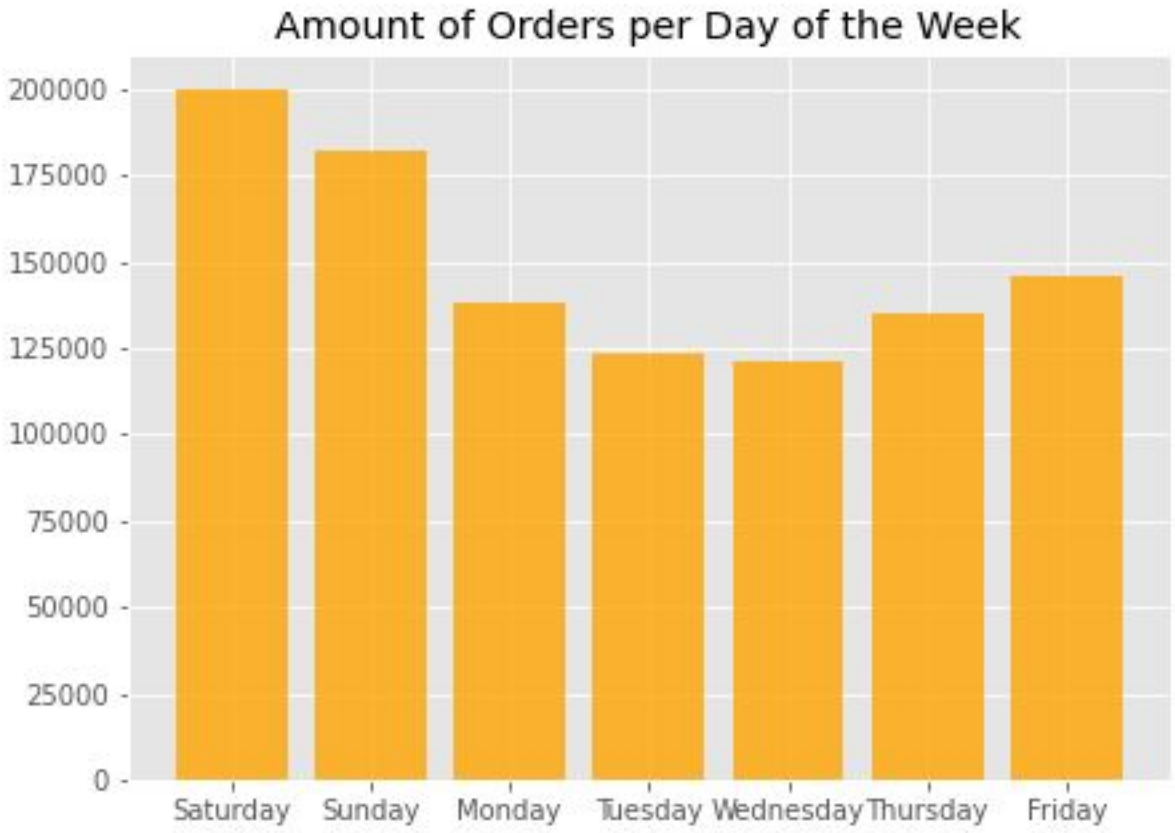
Variance among days of the week

Weekend grocery shopping is quite popular.

Monday-Wednesday may be ideal days for deeper restocking and lighter staffing.

Thursday already begins to ramp up to the weekend.

This pattern informs the peaks in our previous graph on days since the prior order.



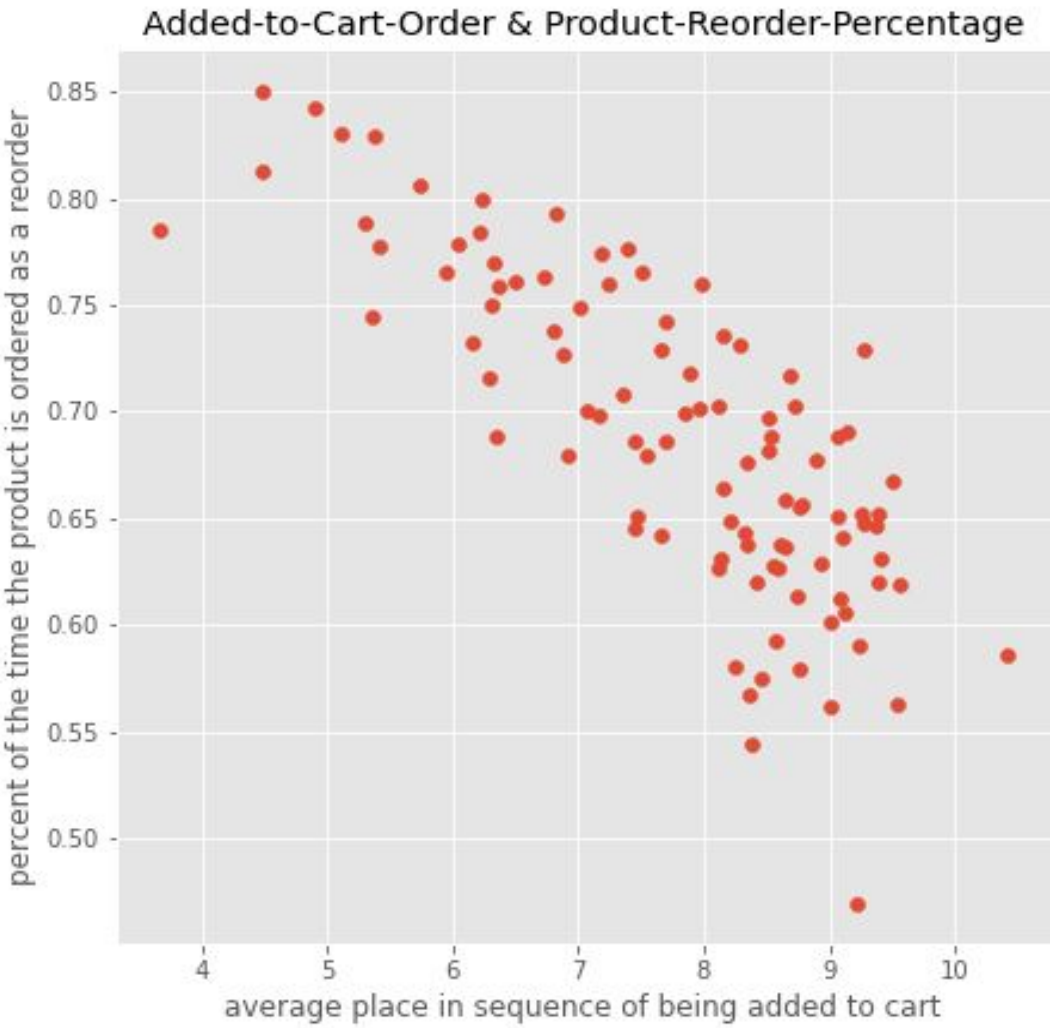
Product details and cart order

Reordered products are placed in the cart earlier in the order process.

This graph just shows the top 100 products.

Including all 35,449 unique products makes these patterns hard to see.

Later, regression analysis will help us specify the exact correlation with more (but not all) products.



Linear Regression Analysis

turning preliminary findings
into
business recommendations

Overall process:

1. Baseline model - includes only the single strongest aspect of our data that correlates with reorders.

2. Investigate the pros & cons of that model.

3. Improve on the baseline by adding and/or transforming more aspects of the data.

Contextualizing regression analysis

- How correlation is measured:
  - via an “R-Squared score”
  - measures how well certain dataset aspects explain differences among reorders
  - ranges from **0** (poor fit) to **1** (perfect fit)

- Only more popular products make for useful insights
  - A model based on all 35k unique products doesn’t explain reordering.
  - Some products were ordered over 150,000 times while others only once.
  - Subsequently, only products ordered over 100 times are considered here.

product aspects considered:

1. product name
2. the product's department
3. the product's aisle
4. the average day of the week the product is ordered
5. the average hour of the day the week the product is ordered
6. the average amount of days between orders for this product
7. the average order in which this product is placed into the cart
8. the amount of times the product was ordered
9. the amount of times the product was reordered

strongest correlation

= aspects that show any kind of correlation with reordering

Baseline model

This is essentially **a correlation between:**

the average order in which a product is added to the cart
&
that product's percent of orders *as a reorder*.

This gives us a useful baseline R-squared score of **0.32**, which means it explains some of the variation in the data, but it could be better.

Improving the baseline model

We can add more data aspects to the model as long as they:

- make sense from a logical, business standpoint
- follow a few regression analysis requirements
- don’t introduce problems

The next strongest aspect of the data (still relating back to reorders):

- **average number of days since a product was ordered**
- adding this improves the model (R-squared: **0.383**)

Approaching the final model

The next two strongest aspects regard order **frequency**:

- order **hour of day** & **day of the week**
- however, we cannot include both because they are too interrelated
- this means including both makes it hard to see their individual relation to reordering
- the **hour of day** aspect proves problematic for other reasons

Solution:

- exclude hour of day
- just keeping **day of the week** works well
- R-squared score improves to **0.434**

business recommendations based on reordering correlations

correlation: As the **average add-to-cart order** of a product increases by one, the likelihood of that product being ordered as a reorder **decreases by 4.7%**.

recommendation 1: Advertising a new or as-of-yet-unordered product online or in-app may be more effective towards the end of the online ordering process.

recommendation 2: Finding a way to cluster reorders may allow greater user patience for newer product promotion; this is likely part of the reason “click to reorder previous” options are available in some existing apps/desktop sites.

(recommendations continued)

correlation: Over the course of a week, as a product's average order **day of the week** increases *away from Saturday by one whole day*, the likelihood of that product being ordered as a reorder **decreases by 15.2%**.

Recommendation 3: Promoting new products in commercials, online, or in-app on Thursdays or Fridays may help increase interest in new products before the weekend rush.

Conclusions and future applications

Instacart's helpful dataset can also inform grocery chains' own delivery, curbside, and restocking operations.

Future or similar projects *tailored towards specific stores* may help:

- optimize aisle layouts
- quantify reserved parking spaces for curbside pickup
- inform allocation of staff to delivery & curbside services
- improve grocery ordering app design

Future projects may also *include data on in-store shopping* to help refine these correlations with restocking.

This current project may also help inform online-only stores (aka "online fulfillment centers").

Thank You

joel.mott8@gmail.com

Flatiron School