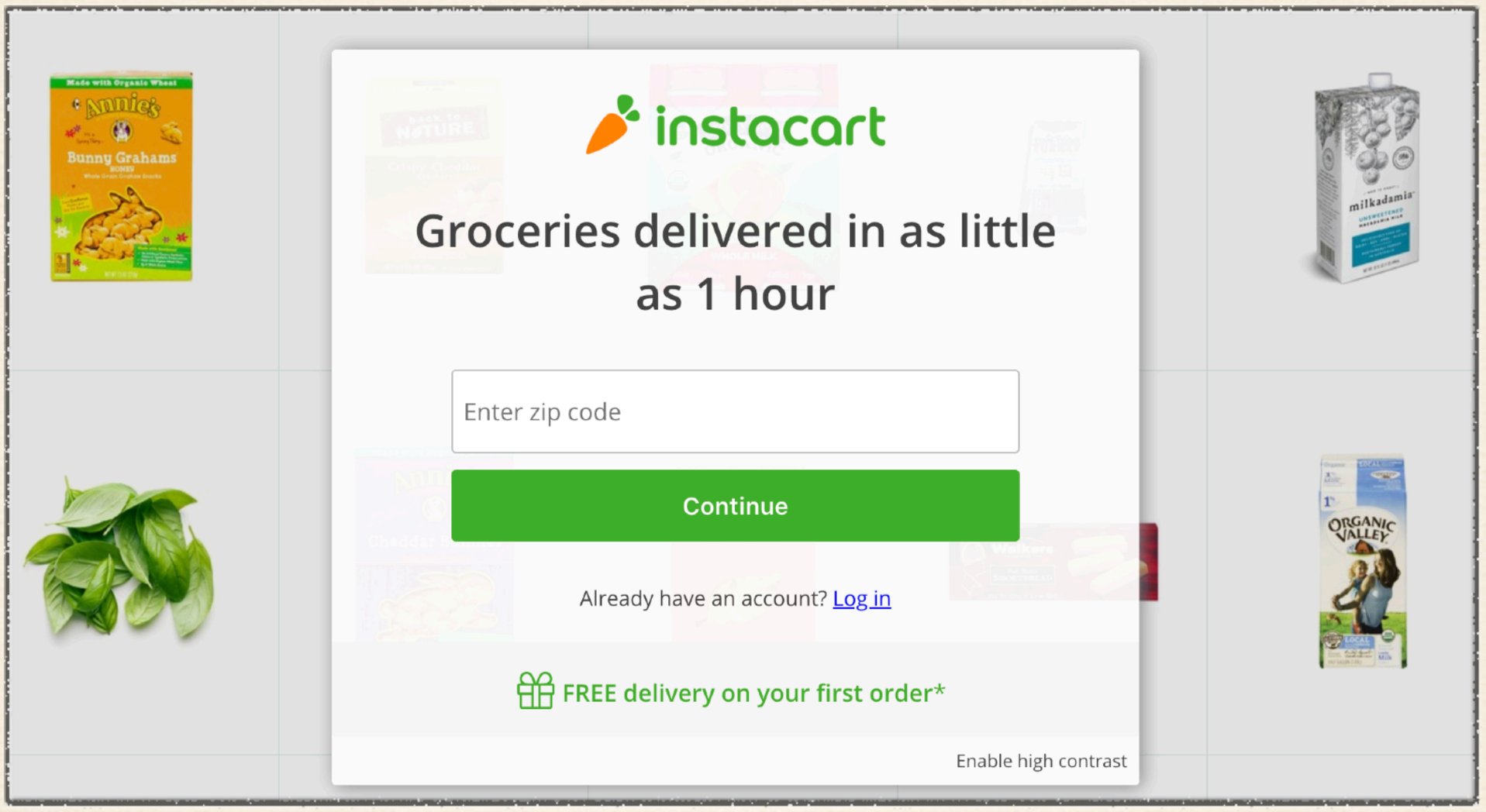
Meat & Drink

An Analysis of Pair Popularity
and
Reordering
among
Instacart Clients

Kaggle



https://www.instacart.com

Instacart Data

Can I help Instacart assess the feasibility of a premium pairing service?
To complete this assessment, I will answer 6 subordinate questions…

Six Questions

- 1. Are there enough clients that would benefit from the service?
    - Operationalized:
      How many clients buy the categories of products in question?
      How many products were sold within the data collection timeframe?
- 2. Do clients reorder those products frequently?
    - Operationalized:
      Is a large proportion of products reordered by clients within 2 weeks?
- 3. Is there room to improve the up-sale of Meat and Alcohol products?
    - Operationalized:
      What proportion of meat sales are not paired with alcohol?
      What proportion of alcohol sales are not paired with meat?

Six Questions

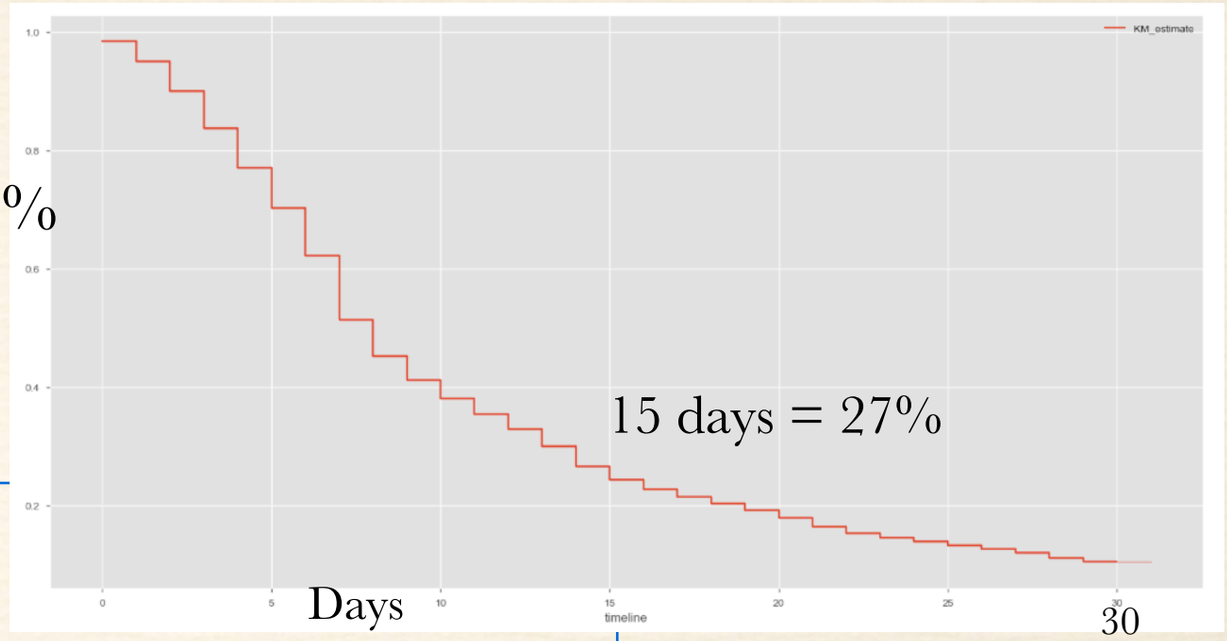
- 4. Which are the most popular pairings? Does the popularity of Instacart combinations reflect popular pairing conventions?
    - Operationalized: What is the ranked order of meat and alcohol pairings by order frequency? Do the most popular combinations match convention?
- 5. Individual Category reorders do not seem to greatly affect popular pairing in the data. Does the healthiness of individual product items affect product reordering?
    - Operationalized:
      How well do product nutrient values and other features predict Reordering?
- 6. How can Product Reordering be predicted best? How can this knowledge be used to inform direct advertising selections based on Sales and ROI?
    - Operationalized: Which supervised classification model type worked best?

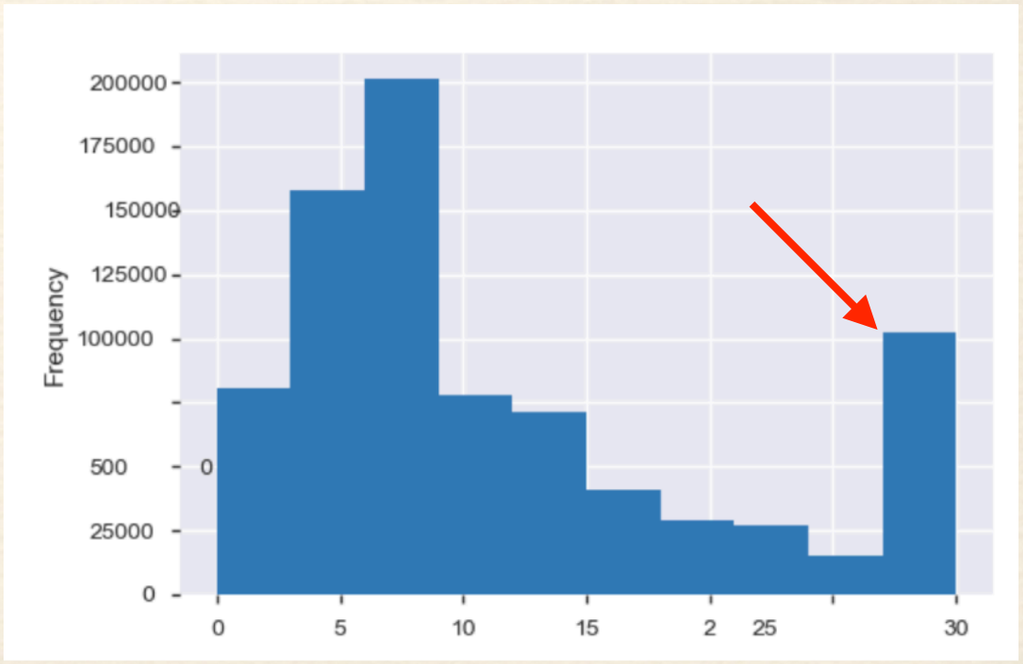
Survival Analysis

Enough volume, frequency?

- Q1. 862,000+ orders; 120,000+ clients
- Q2. Time to next order
- Proportion **not** reordered
- Non-parametric approach
- ***73% reordered by 15 days***
- Censored data
  - Looking at distribution…
  - Huge spike at 30 days, at upper limit of range

Days to Next Order





https://Flaketravislifestyle.com

Q3.

The Drink

Instacart perhaps best known for groceries… Alcohol has a high profit margin. 90% of Meat orders were not paired with Alcohol - an opportunity for upselling

Where's the beef?

https://www.kolfoods.com

Q3.

	Drink	Group	Count
1	beers coolers	none	26085
2	white wines	none	21293
3	red wines	none	21245
4	spirits	none	19131
5	specialty wines champagnes	none	8492

Instacart also has an opportunity to up-sell meat when clients purchase alcohol

LA Times on Pairing Food & Drink

Whiskey with beef, gin with shrimp: A guide to pairing spirits, food

- **Barbecue/Grilling + Beer, Coolers**
- **Red meat + Red wine**
- **White meat, seafood + White wine**
- **Spirits:**
  - **Whiskey + Beef**
  - **Gin + Shrimp, Seafood**
  - **Vodka + Smoked/cured Fish**
  - **Tequila + guacamole, tacos**
  - **Rum + Pork**
  - **Bourbon + Barbecued meats**

LA Times | By S. IRENE VIRBILA MAY 15, 2015 | 11:00 AM

Flight of Vodka Oyster Shooters, from left to right, The Red Devil, The Sunset and The Kentucky, with an oyster platterat Tipple & Brine. (Ricardo DeAratanha / Los Angeles Times)

http://www.latimes.com/food/drinks/la-fo-food-cocktails-20150516-story.html

Sommelier

vs

The Data-Driven Approach

https://www.1001freedownloads.com

Do the data reveal those generally held guidelines?

vs

Newer convention - pair what you like!

Most Popular Combinations

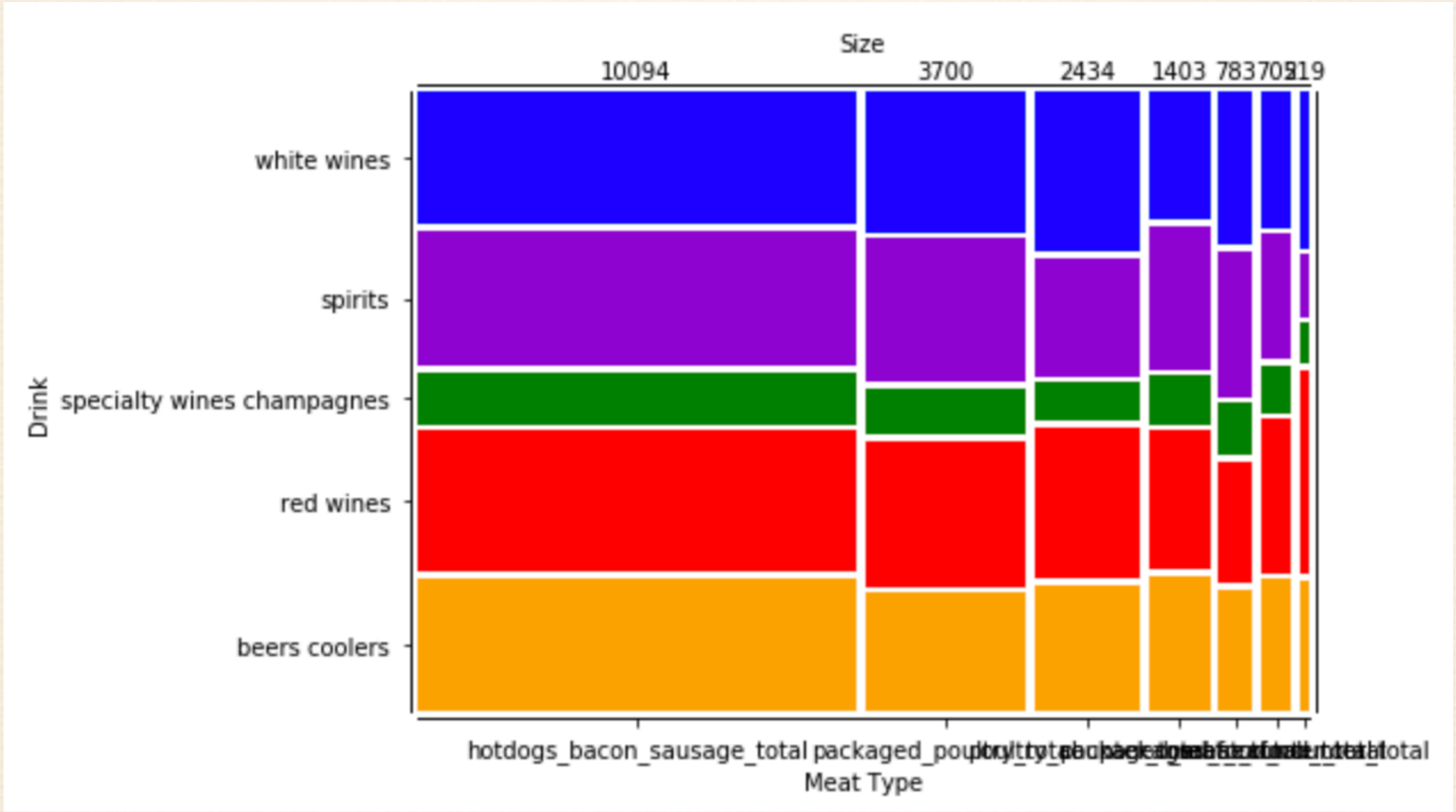
Meat type strong effect

Lower-Priced Meat sold at Higher Frequency

Pairing less important

	Drink	Group	Count
1	red wines	hotdogs_bacon_sausage_total	2387
2	spirits	hotdogs_bacon_sausage_total	2278
3	white wines	hotdogs_bacon_sausage_total	2275
4	beers coolers	hotdogs_bacon_sausage_total	2273
5	red wines	packaged_poultry_total	902
6	spirits	packaged_poultry_total	889
7	specialty wines champagnes	hotdogs_bacon_sausage_total	881
8	white wines	packaged_poultry_total	878
9	beers coolers	packaged_poultry_total	737
10	white wines	poultry_counter_total	657
11	red wines	poultry_counter_total	615
12	beers coolers	poultry_counter_total	520
13	spirits	poultry_counter_total	484
14	spirits	packaged_meat_total	339
15	red wines	packaged_meat_total	328
16	beers coolers	packaged_meat_total	319
17	white wines	packaged_meat_total	303
18	specialty wines champagnes	packaged_poultry_total	294
19	white wines	packaged_seafood_total	203
20	spirits	packaged_seafood_total	191
21	red wines	meat_counter_total	181
22	white wines	meat_counter_total	162
23	beers coolers	packaged_seafood_total	161
24	red wines	packaged_seafood_total	160
25	specialty wines champagnes	poultry_counter_total	158
26	beers coolers	meat_counter_total	157
27	spirits	meat_counter_total	149
28	specialty wines champagnes	packaged_meat_total	114
29	red wines	seafood_counter_total	75
30	specialty wines champagnes	packaged_seafood_total	68
31	white wines	seafood_counter_total	58
32	specialty wines champagnes	meat_counter_total	56
33	beers coolers	seafood_counter_total	48
34	spirits	seafood_counter_total	23
35	specialty wines champagnes	seafood_counter_total	15

Mosaic Plot Combinations



Reordering Ranked

Individual Categories

	Category	Reordered
1	white wines	0.63
2	poultry counter	0.61
3	packaged poultry	0.60
4	spirits	0.57
5	beers coolers	0.57
6	hot dogs bacon sausage	0.56
7	red wines	0.55
8	packaged meat	0.54
9	meat counter	0.53
10	seafood counter	0.52
11	packaged seafood	0.51
12	specialty wines champagnes	0.49

Does Health Influence Reordering?

- Multifactor Nutrition Information https://www.ars.usda.gov
- Merged several tables
- Calories, CHO, Protein, Fat, Sugar, Alcohol content
- Nutrition feature development
- Mapped category inclusions based on group T-tests for each continuous feature in USDA data

https://baskentweb.com

Predicting Product Repurchase

- Naive Bayes 0.66 (Test AUC; recall & precision fairly balanced)
- Decision tree 0.70
- Random forest 0.75
- Gradient Boosted Tree 0.75 AUC - Misclassification Rate 29.8%
  - 20% holdback, 5-fold cross validation
  - Change threshold to boost recall or precision, cost adjustment (e.g. if advertising mode is inexpensive - website advertisement)

Predicting Product Repurchase

- Most Influential factors (by model contributions):
    - Order number, Days since prior order, Meat/Drink category
- Other Influential factors:
    - Protein, Fat, Calories, ETOH content
- Less important:
    - Day of Week, Time of Day, others…

Six Questions Answered

- 1. Are there enough clients that would benefit from the service?
    - Operationalized: How many clients buy the categories of products in question?
      How many products were sold within the data collection timeframe?
    - Answer: Yes! 862,000+ products; 120,000+ clients (Summary Statistics)
- 2. Do clients reorder those products frequently?
    - Operationalized: Is a large proportion of products reordered by clients within 2 weeks?
    - Answer: Yes! About 73% within 15 days (Survival Analysis)
- 3. Is there room to improve the up-sale of Meat and Alcohol products?
    - Operationalized: What proportion of meat sales are not paired with alcohol?
      What proportion of alcohol sales are not paired with meat?
    - Answer: Yes! 90% of meat sales, and 85% of alcohol sales are not paired.

Six Questions Answered

- 4. What are the most popular pairings?
  Does the popularity of Instacart combinations reflect popular pairing conventions?

  - Operationalized: What is the ranked order of meat and alcohol pairings by frequency?

  - Answer: No! Data do not adhere strictly to popular pairing conventions. (Summary Stats)
    For details, refer to frequency list in descending order. (by Order)

- 5. Individual Category reorders do not seem to greatly affect popular pairing in the data. (Summary Stats)
  Does the healthiness of individual product items affect product reordering? (by Product)

  - Operationalized: How well do product nutrient values and other features predict Reordering?

  - Answer: Yes! Product nutrients and other features do contribute value to predict reordering.

- 6. How can Product Reordering be predicted best?
  How can this knowledge be used to inform direct advertising selections based on Sales and ROI?

  - Operationalized: Which supervised classification model type worked best?

  - Answer: Gradient Boosted Tree. Adjust threshold to boost recall for inexpensive web advertising.
    Use features with high model contributions in DOE to test different combinations in advertising.

Take Home + Next Steps

- **Instacart <u>should</u> attempt to launch the new pairing service, including healthy options.**
- Offer portal for Beta testing of new pairing service
- Designed Experiments:
    - Test direct marketing approaches
        - Optimize Sales, Return on Investment
        - Increase revenue
    - Test advertising methods/combinations against two groups
        - Those clients who repurchased products
        - Those who did not
        - Adjust confusion matrix to boost recall or precision, respectively

https://www.kolfoods.com

https://www.istockphoto.com

References

- LA Times Article: http://www.latimes.com/food/drinks/la-fo-food-cocktails-20150516-story.html

- Instacart Datasets: https://www.kaggle.com/datasets/

- Food & Drink Nutrition: https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/nutrient-data-laboratory/docs/usda-compiling-food-composition-data-for-over-115-years/