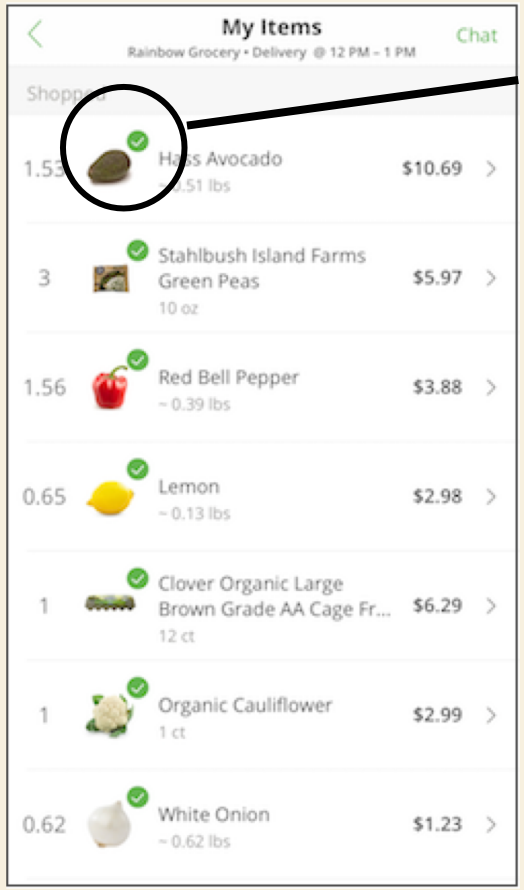
WHAT'S IN YOUR SHOPPING CART?

INSTACART MARKET BASKET ANALYSIS

DATA SOURCE

instacart



Product (Department, Aisle)

Hour of Day

Day of Week

Prior Order

FEATURES

- **User Features**
  - Total Order Count
  - Count of Product Type (Unique Product Count)
  - Average Cart Size
  - . . .

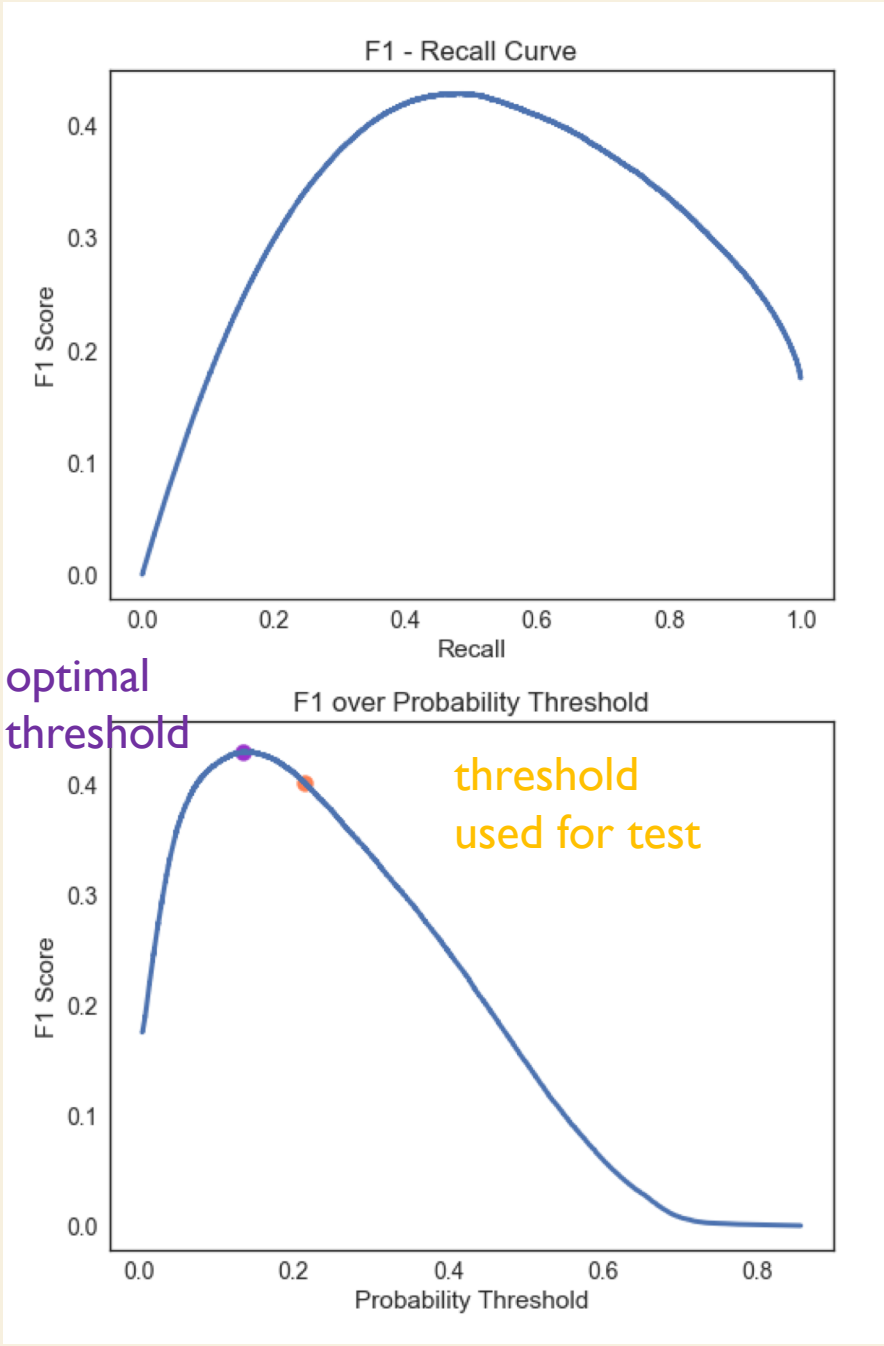
- **Product Features**
  - Total Order Count
  - Total Reorder Percent
  - Department
  - . . .

- **Current Order Features**
  - Order Time (Hour, Day)
  - Days since Prior Order
  - . . .

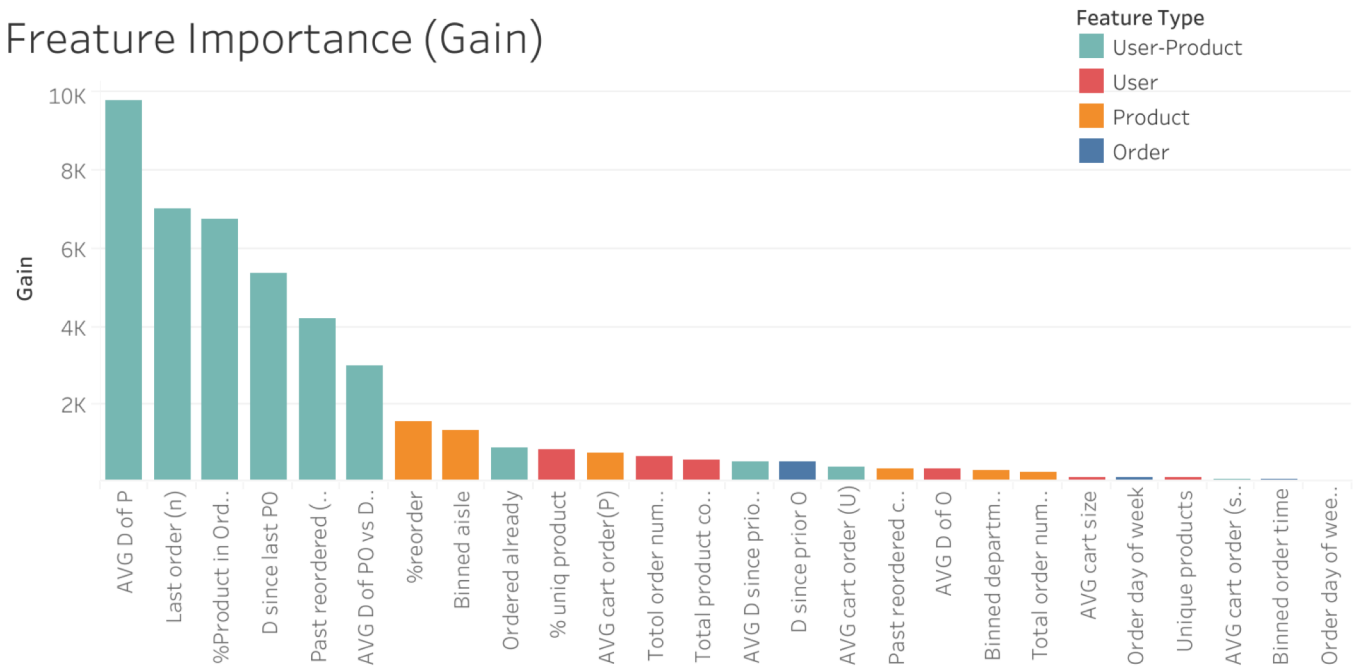
- **User - Product Features**
  - Average Days Between Purchasing the Product
  - Total Reordered Count
  - Last Order Purchase the Product (Order)
  - Percent of Orders Containing the Product
  - . . .

MODEL

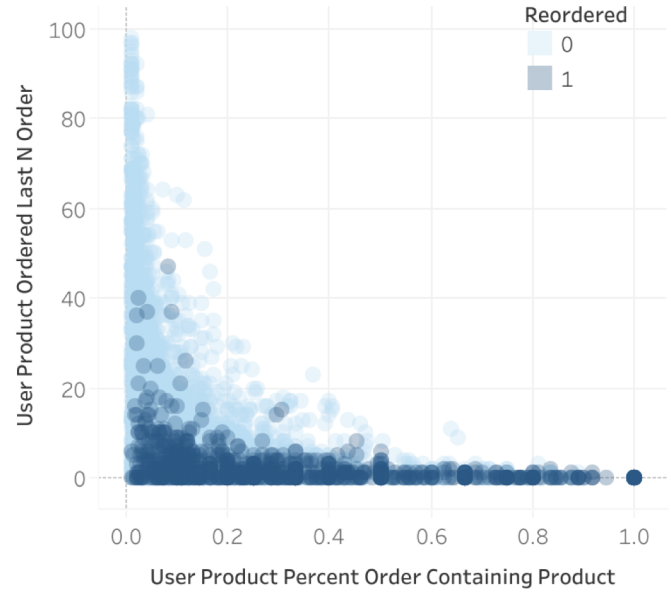
- **Gradient Boosting (XGB)**
  - n_estimators = 339
  - max_depth = 4
  - learning_rate = 0.025
  - subsample = 0.7
  - colsample = 0.6
  - Probability threshold = 0.2138
  - Metrics - average f1 score on cross validation of subset (2.5%, 0.2M records): 0.429 (std. 0.014)
- **Test Result**
  - **f1 Score: 0.401**
  - **Recall: 0.383**
  - **Precision: 0.486**



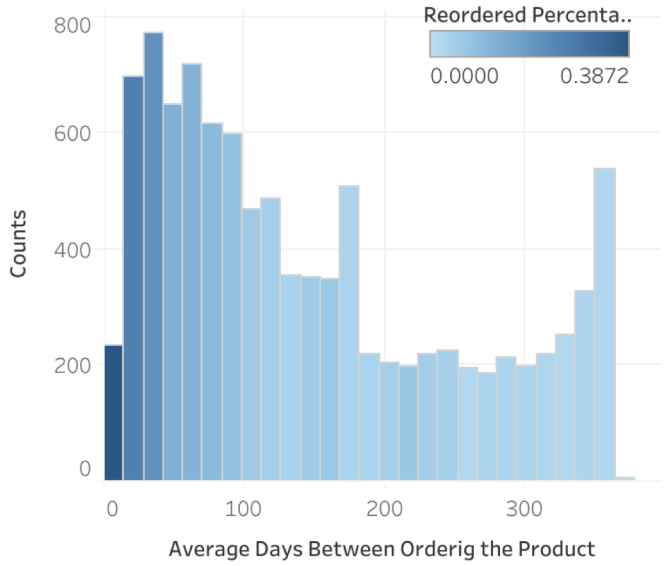
Freature Importance (Gain)



Last Order vs Order Frequency

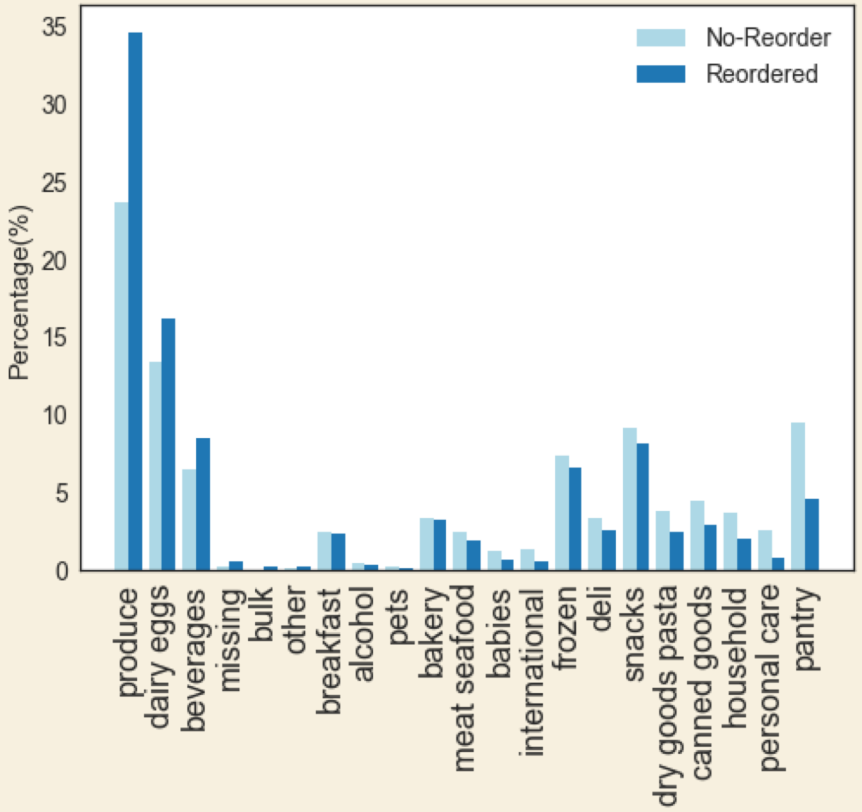


Average Days Between Ordering the Product with Reordering Percentage



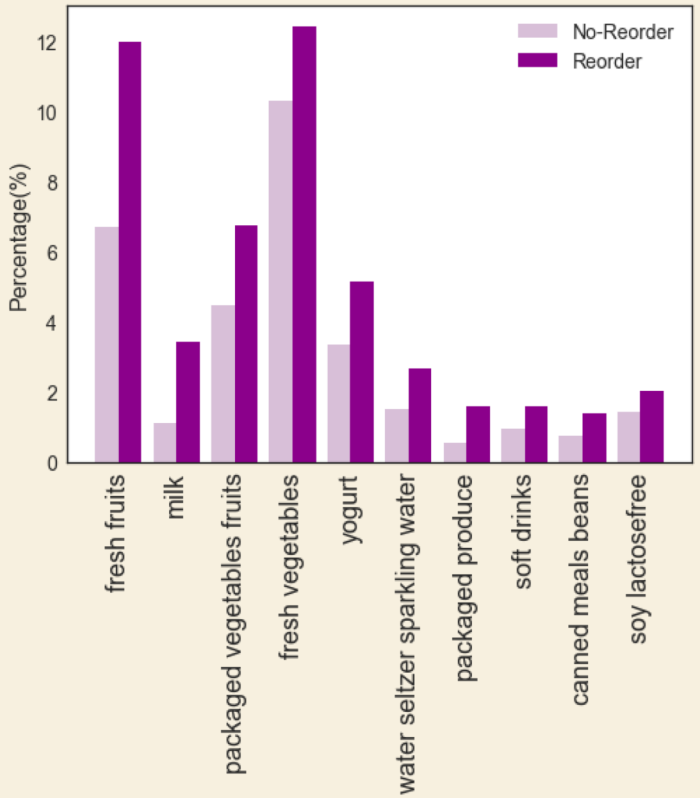
PRODUCT TYPE

Percentage of Reorder vs No-reorder by Dept

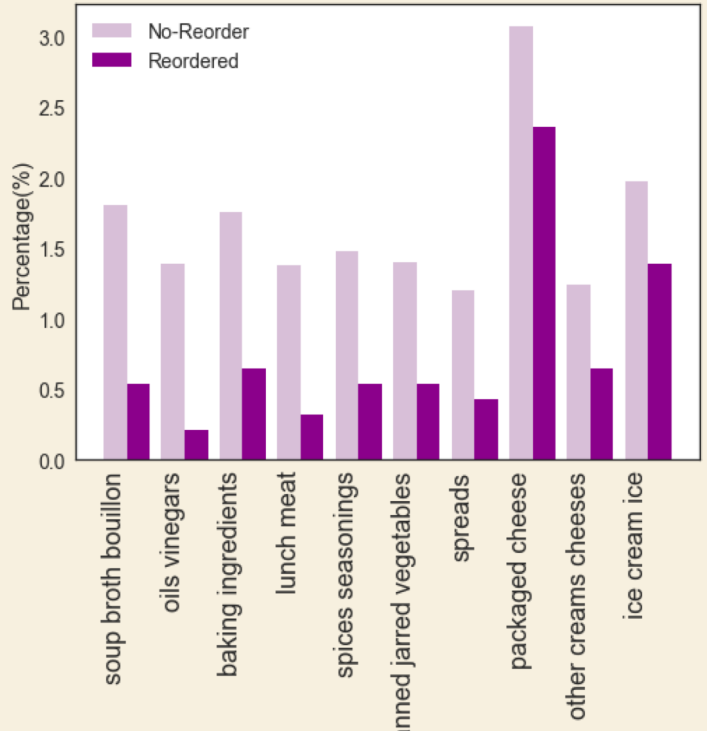


Percentage of Reorder vs No-reorder by Aisle

Top 10



Bottom 10



FURTHER IMPROVEMENT

- Implement the product recommendation system
- Incorporate time series factor in the model
- Adding external data source, eg. product similarities
- Select model on a larger subset

Thank You!