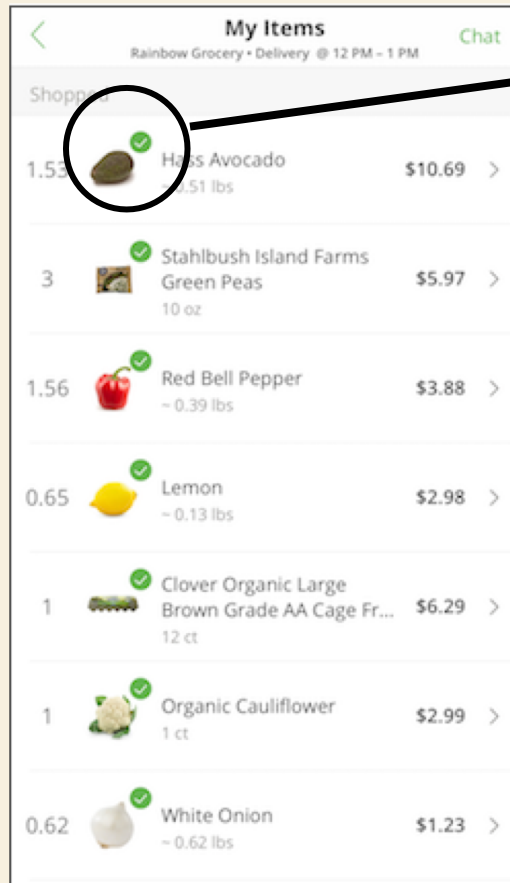




# **WHAT'S IN YOUR SHOPPING CART?**

**INSTACART MARKET BASKET ANALYSIS**

# DATA SOURCE



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)**

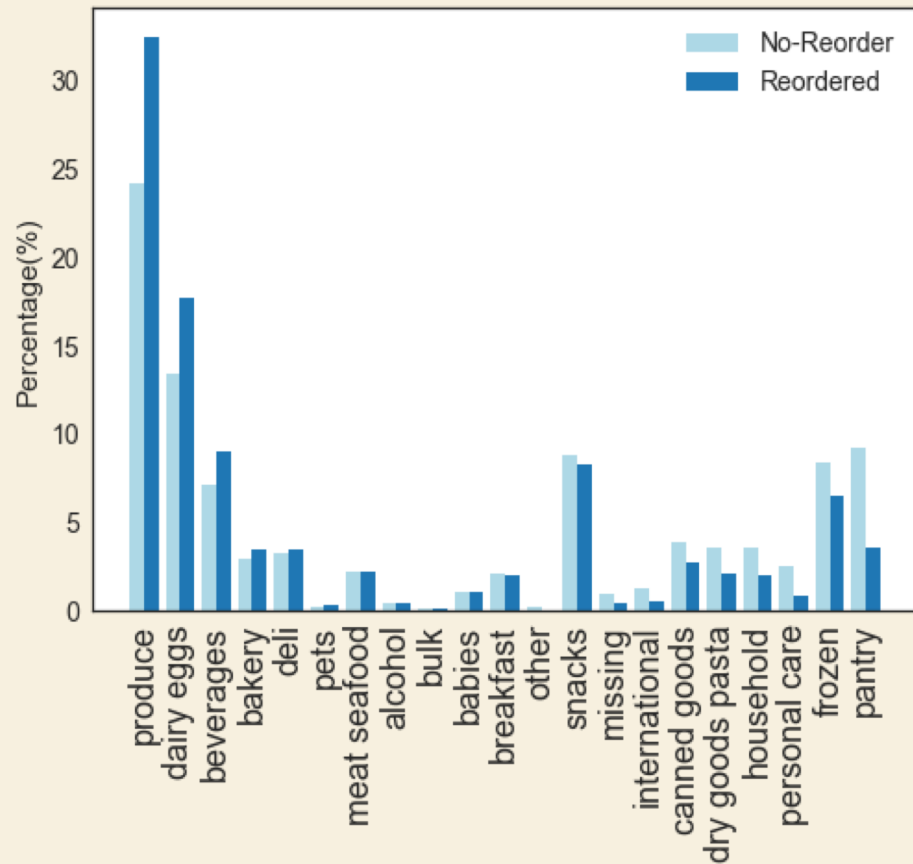
- `n_estimators` = 843
- `max_depth` = 6
- `learning_rate` = 0.01
- `subsample` = 0.3
- `colsample` = 0.7
- Probability threshold = 0.2114
- Metrics - average f1 score on cross validation of subset (6.3%, 1.9M records): 0.437 (std. 0.0078)

- **Test Result**

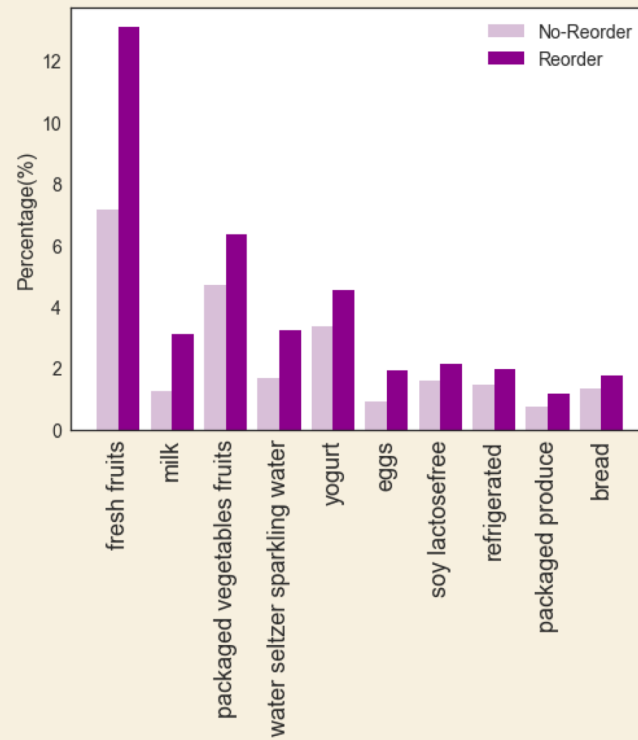
- **f1 Score: 0.436**
- **Recall: 0.506**
- **Precision: 0.383**

# PRODUCT TYPE

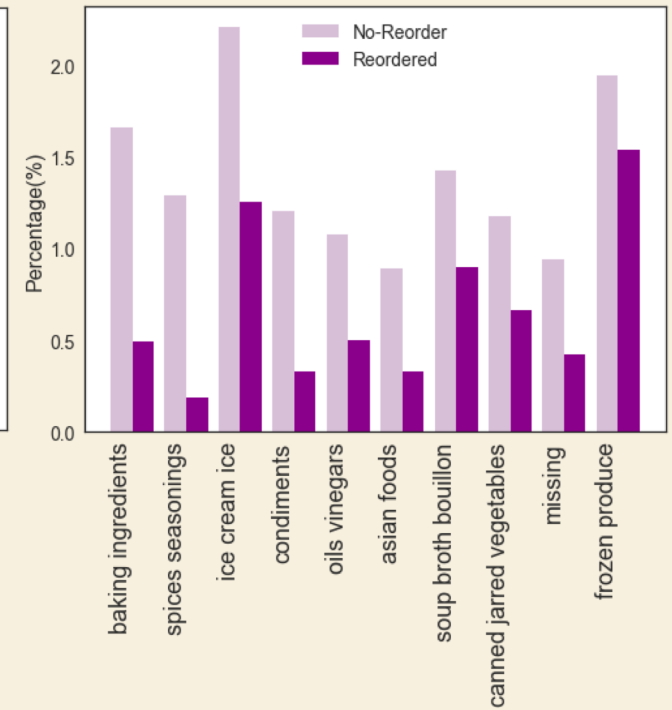
Percentage of Reorder vs No-reorder by Dept



Percentage of Reorder vs No-reorder by  
Top 10 Aisle



Bottom 10



# FURTHER IMPROVEMENT

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

*Thank You!*

