Capstone 2 - Instacart

A Predictive Shopping Experience

Not Your Average Shopping Experience

Instacart is an immersive shopping experience

Users compelled to consume a massive variety of products

Predict which products they will buy again

Data Cleaning & Preparation

Removed one of a kind items

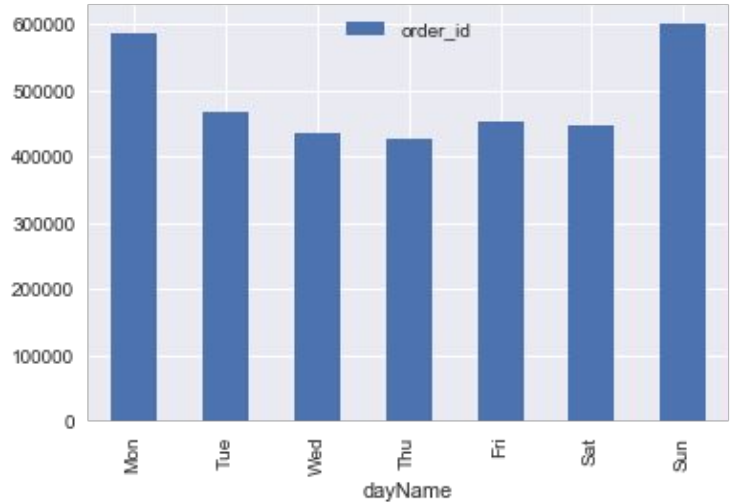
Items with 100% reorder rate or purchased < 40 times in total

Include coding category names for informative plots

Conversion of timestamp data

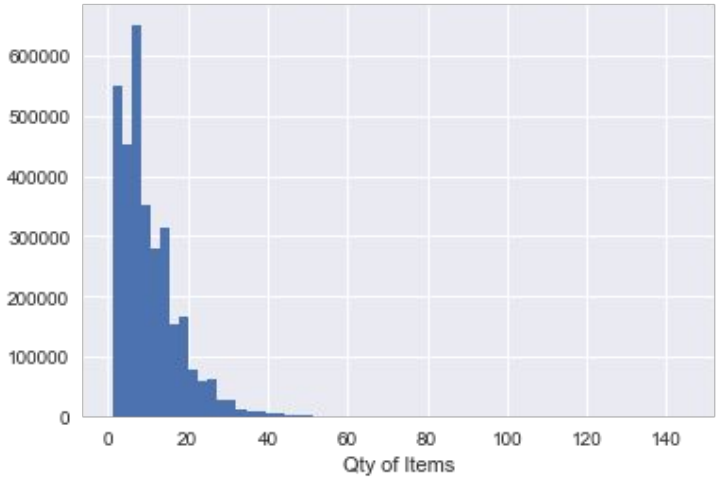
Aggregative calculations

Explorative Analysis



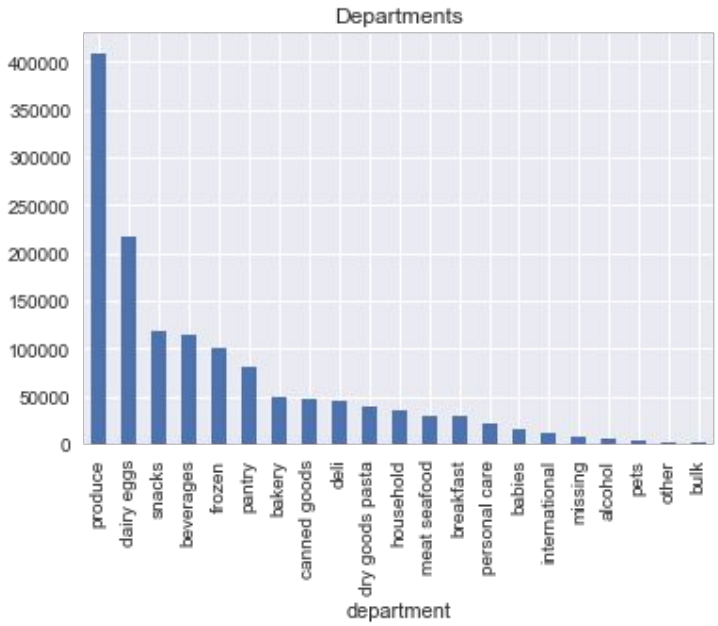
Users are shopping Sunday & Monday, usually midday

Explorative Analysis



They’re shopping every week and buying ~10 items per shop

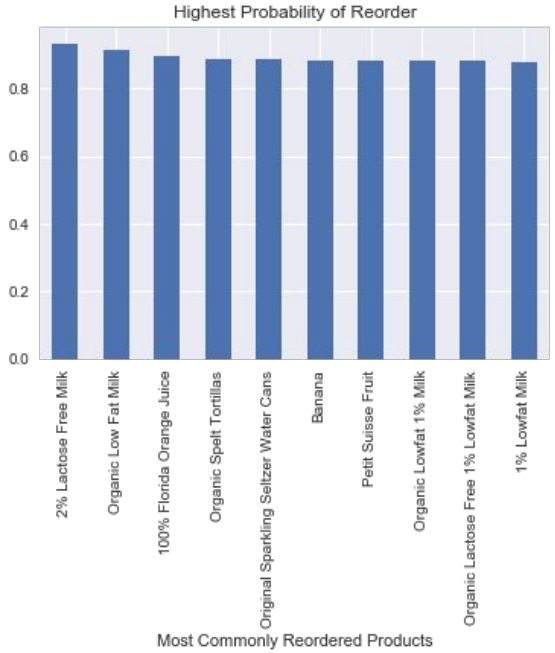
This is bananas!



Shoppers love produce, dairy, eggs and snacks...

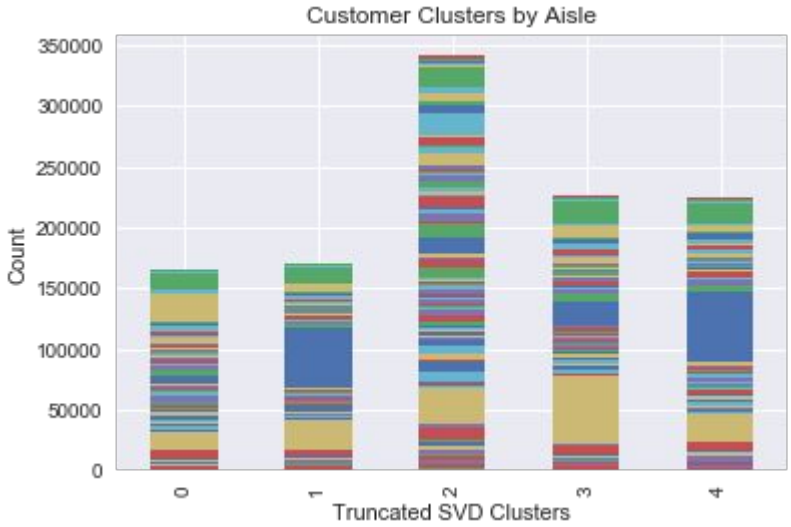
Explorative Analysis

What is most likely to be reordered?



Clustering for customer types

Implemented K-Means on PCA and TruncatedSVD Purchasing Data



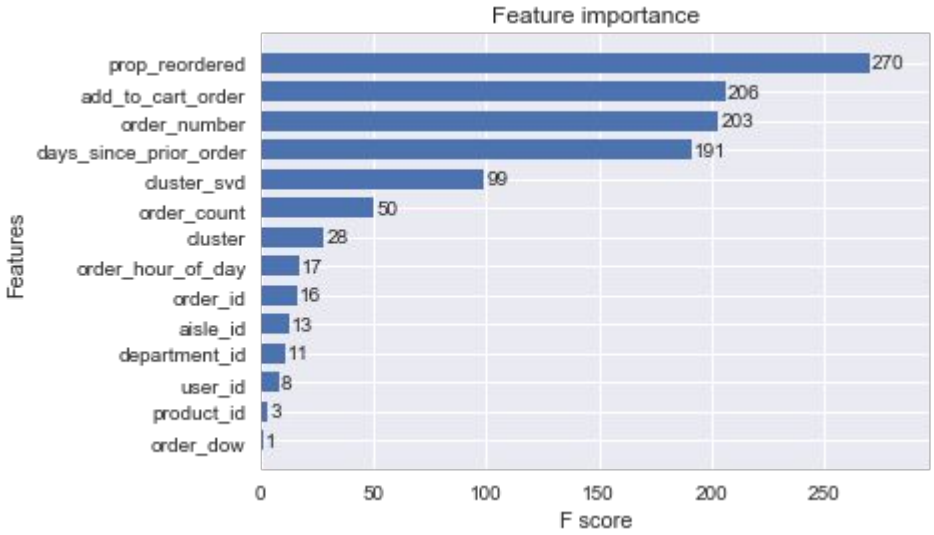
5 Clusters

Similar Product Counts by Aisle

Further Analysis Required...

Modelling & Predictions

XGBoost classifier outperforms them all...



	Logistic	Random Forest	XG Boost
Precision	0.40	0.67	0.70
Recall	0.63	0.67	0.71
F1 Score	0.49	0.67	0.69

References

Appendix:
Calculations
https://github.com/andrewcmilne/capstone1_instaCart

References
https://www.instacart.com/