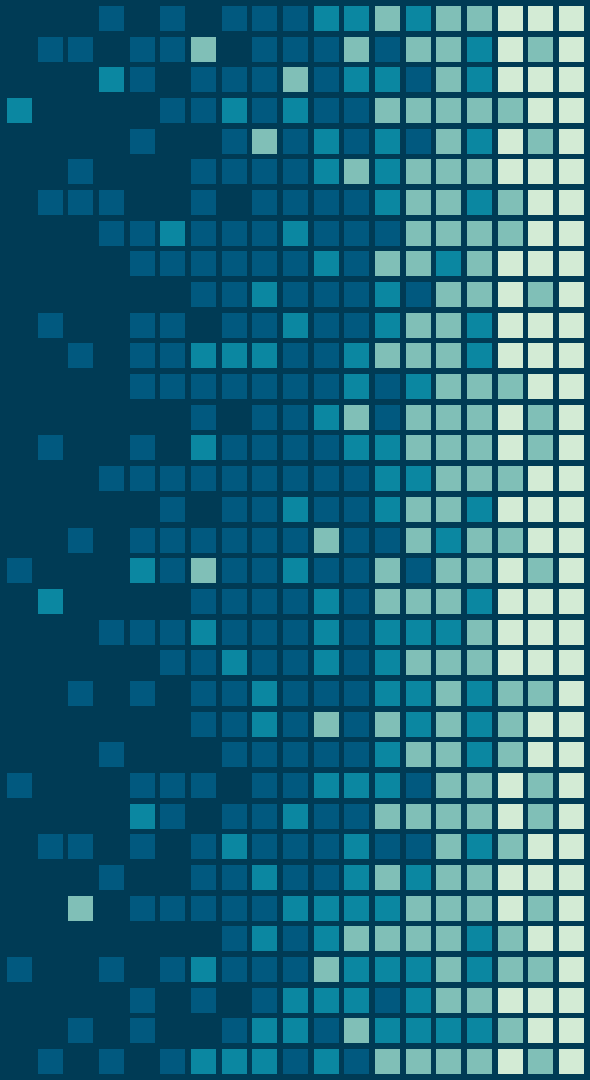
PREDICTING REPEAT PURCHASES AT INSTACART

MAI ANH LY
DATA SCIENTIST

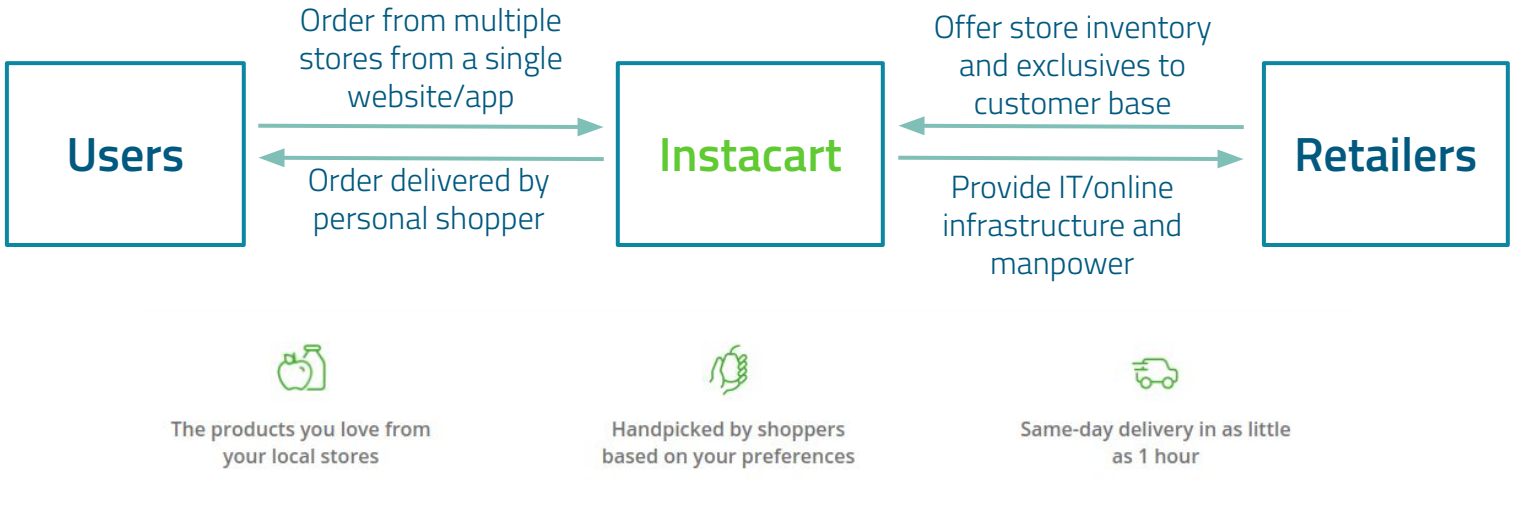


AGENDA

1. **Business problem definition**
2. Data process design
3. Delivery
4. Summary and conclusion

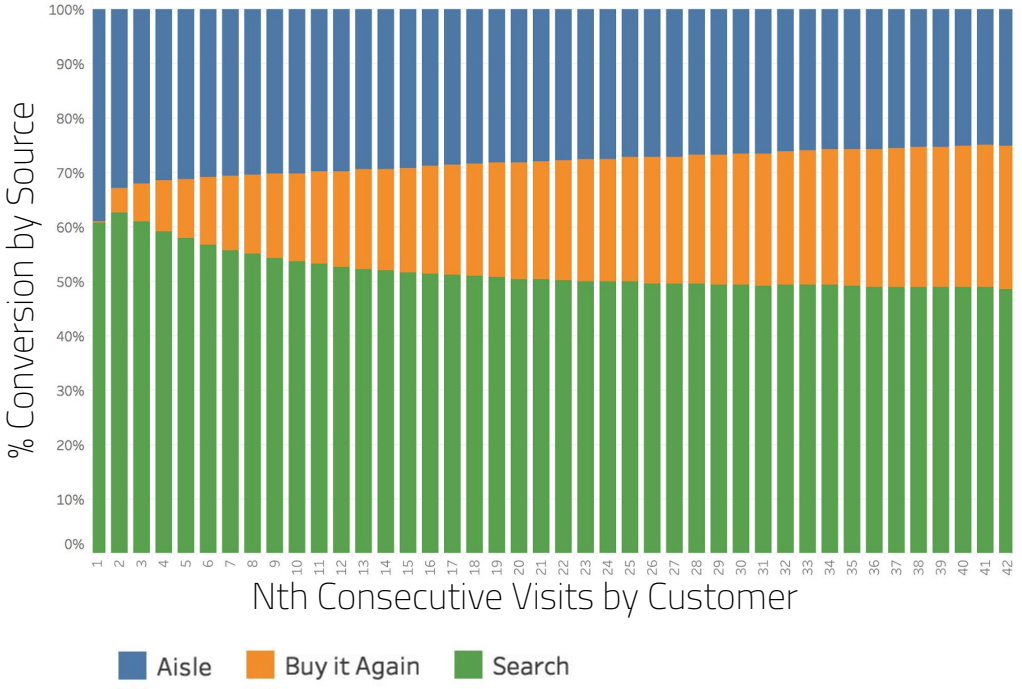
instacart

- Online grocery delivery/pick-up service valued at 4 billion USD
- 2017 revenue: 2 billion USD (Forbes estimate)
- Rely on retail partners (e.g. Costco, Aldi) for inventory management



VALUE PROPOSITION

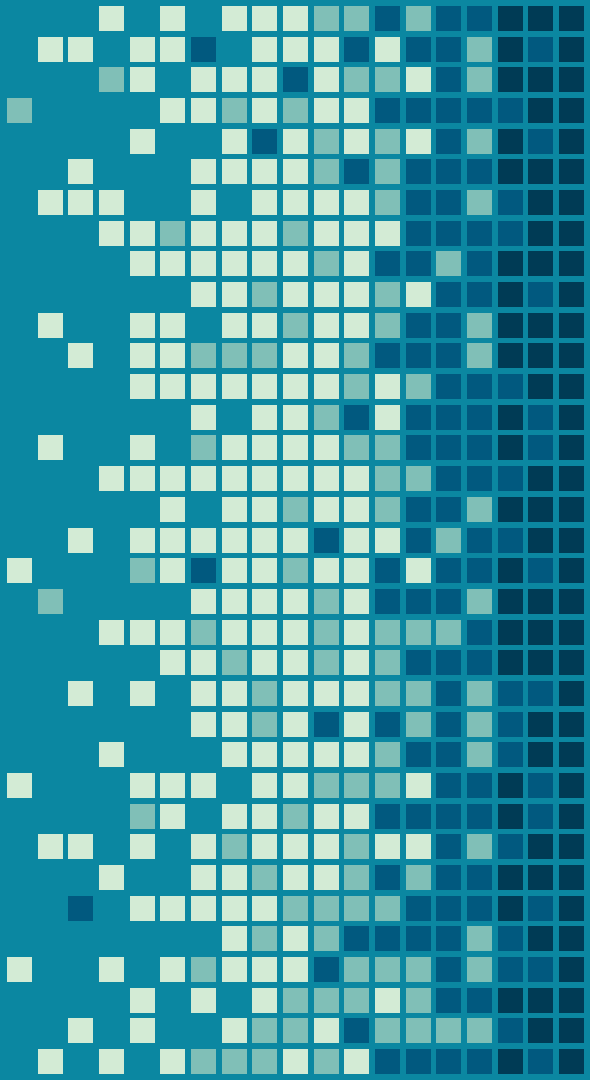
How can Instacart improve retention of their customer base?



Source: "Winning the digital shelf: it's all about the customer" Instacart blog

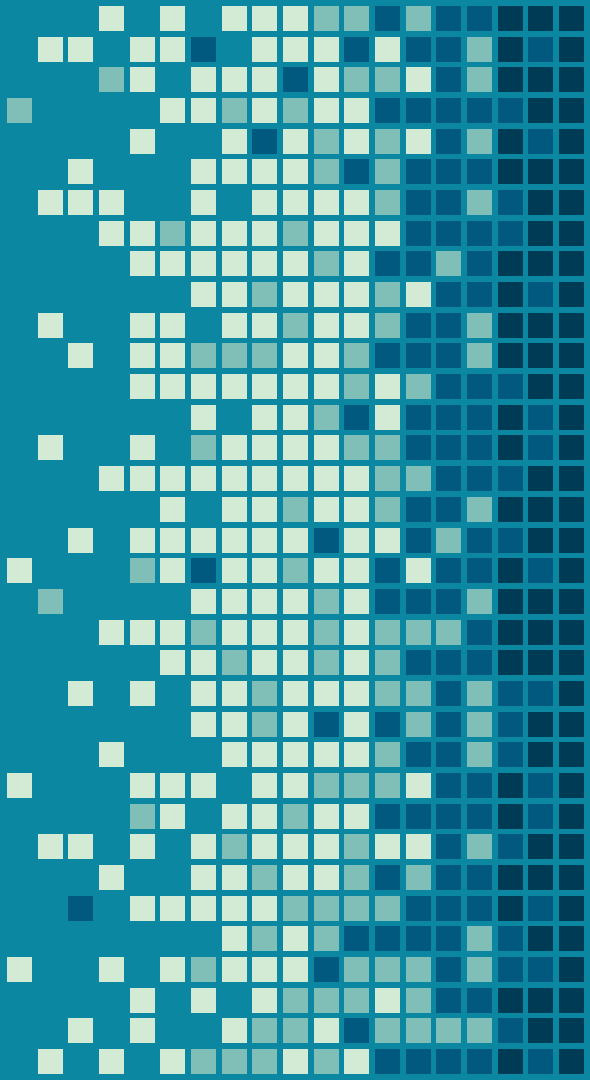
THE BUSINESS QUESTION

“How much value will be added if we choose to increase customer retention via improving the online shopping experience?”



THE DATA QUESTION

> *Based on a customer's purchase history, how accurately can we predict the products that will be in their next order?*



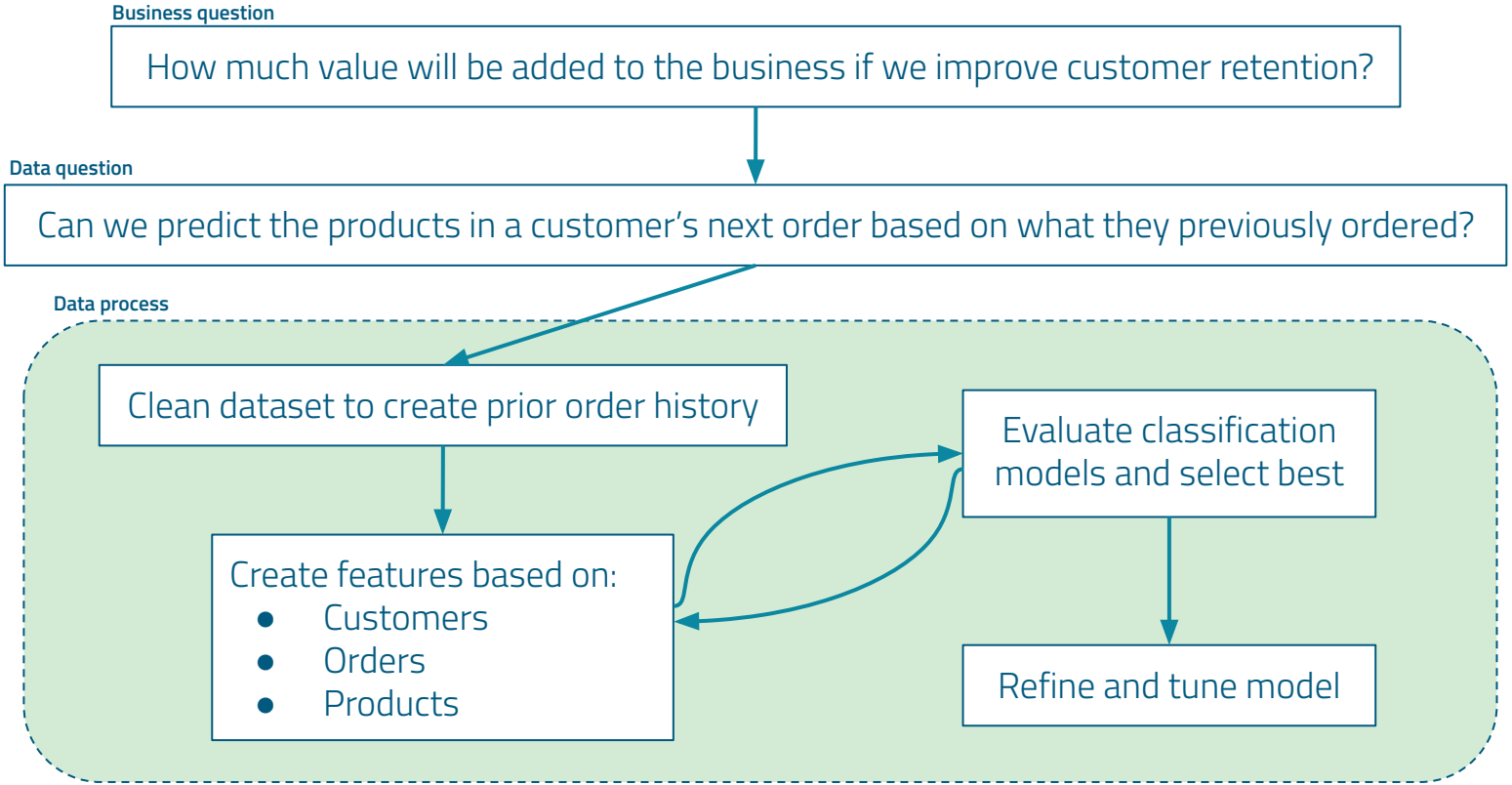
THE CUSTOMER ORDERS DATASET

- Three million orders by 200 thousand users

AGENDA

1. Business problem definition
2. **Data process design**
3. Delivery
4. Summary and conclusion

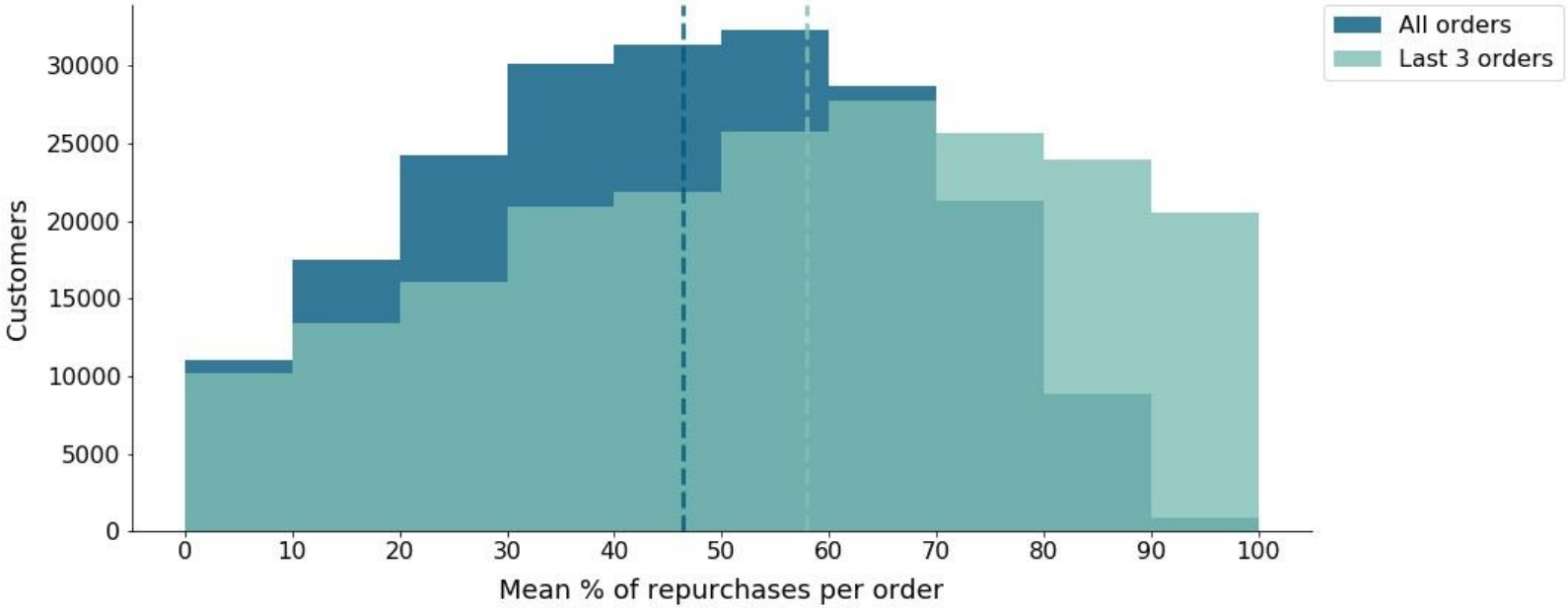
PROCESS WORKFLOW



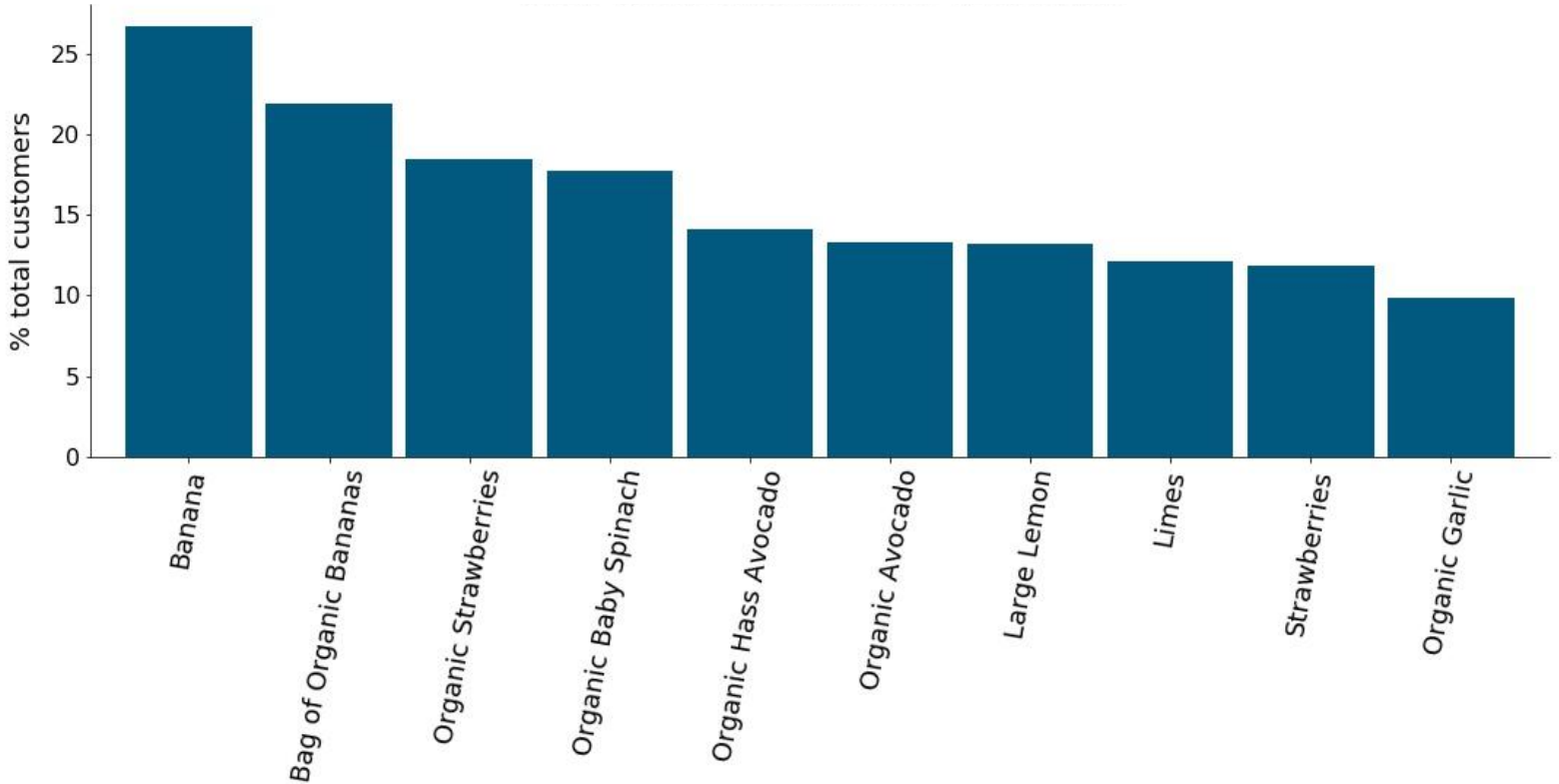
AGENDA

1. Business problem definition
2. Data process design
3. **Delivery**
4. Summary and conclusion

PRODUCT REPURCHASE % PER ORDER



TOP 10 REPURCHASES BY CUSTOMER

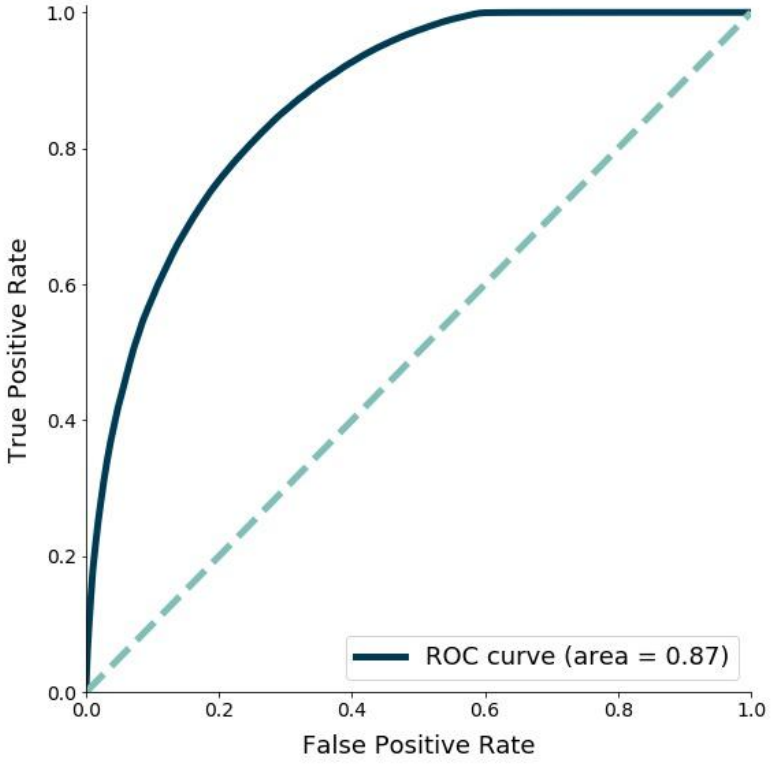


MODEL DEVELOPMENT & EVALUATION

- Model trained on 1.3 million observations and 14 predictors

COMPLEXITY

Model	Accuracy
Baseline (most frequent)	**0.60**
Logistic regression	**0.71**
Random forest	**0.80**
XGBoost	**0.77**



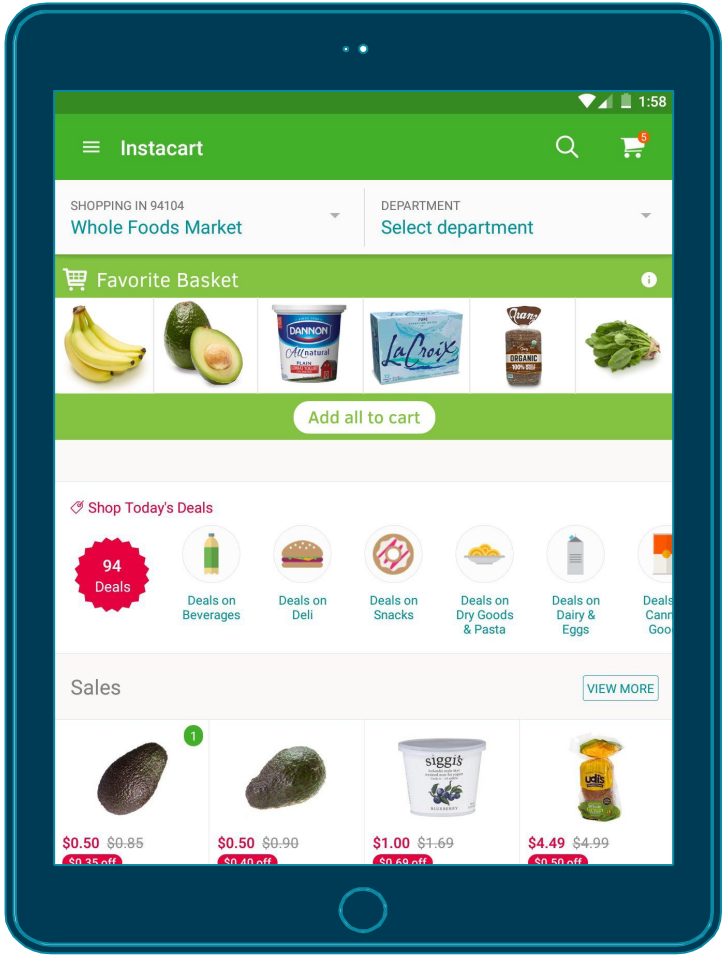
KEY FACTORS INFLUENCING REPEAT PURCHASES

AGENDA

1. Business problem definition
2. Data process design
3. Delivery
4. **Summary and conclusion**

NEXT STEPS

1. Deploy model
2. Develop ‘Favorite Basket’ feature
3. A/B test and measure month-to-month customer retention/churn
4. Decide on implementing feature site-wide



CONCLUSION

Business question

How much value will be added to the business if we improve customer retention?

Data question

Can we predict the products in a customer's next order based on what they previously ordered?

Data process

Clean dataset to create prior order history

Create features based on:

- Customers
- Orders
- Products

Evaluate classification models and select best

Refine and tune model

Data answer

We can predict products a customer will reorder with **80%** accuracy

Business answer

Improving our online experience and customer retention will increase revenue by **$1.4 mil**

THANKS!

Any questions?

Supporting documentation
@ github.com/maianhly/instacart_repurchases