PREDICTING REPEAT PURCHASES AT INSTACART

MAI ANH LY
DATA SCIENTIST

BIO

@ The University of Sydney

- Education Data Analyst
- Bioinformatician
- PhD in Microbiology

AGENDA

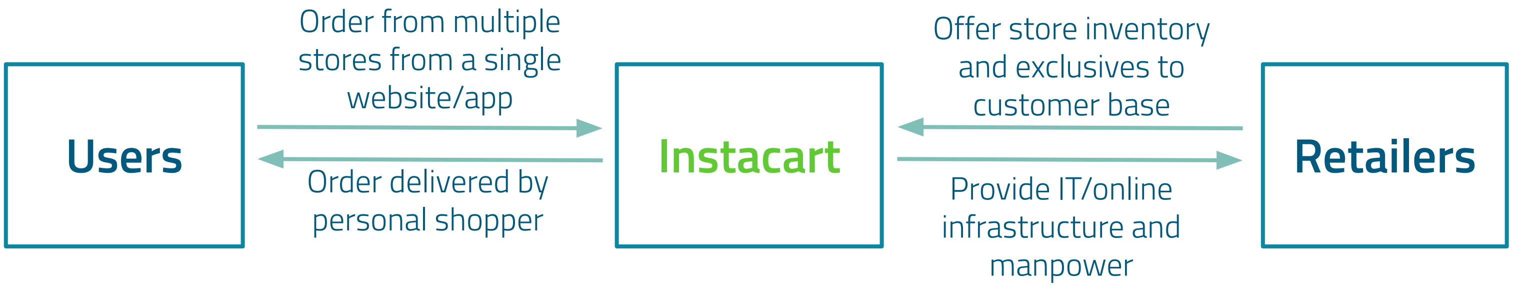
1. Business problem definition
2. Data process design
3. Delivery
4. Next steps and summary

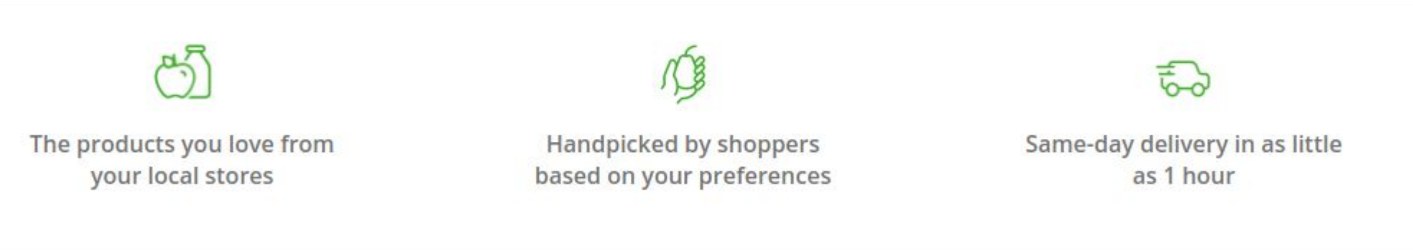
AGENDA

1. **Business problem definition**
2. Data process design
3. Delivery
4. Next steps and summary

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


Users
Order from multiple stores from a single website/app
Order delivered by personal shopper
Instacart
Offer store inventory and exclusives to customer base
Provide IT/online infrastructure and manpower
Retailers

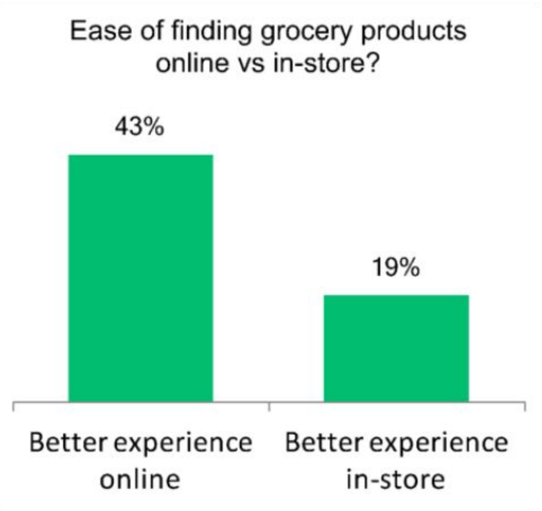

The products you love from your local stores

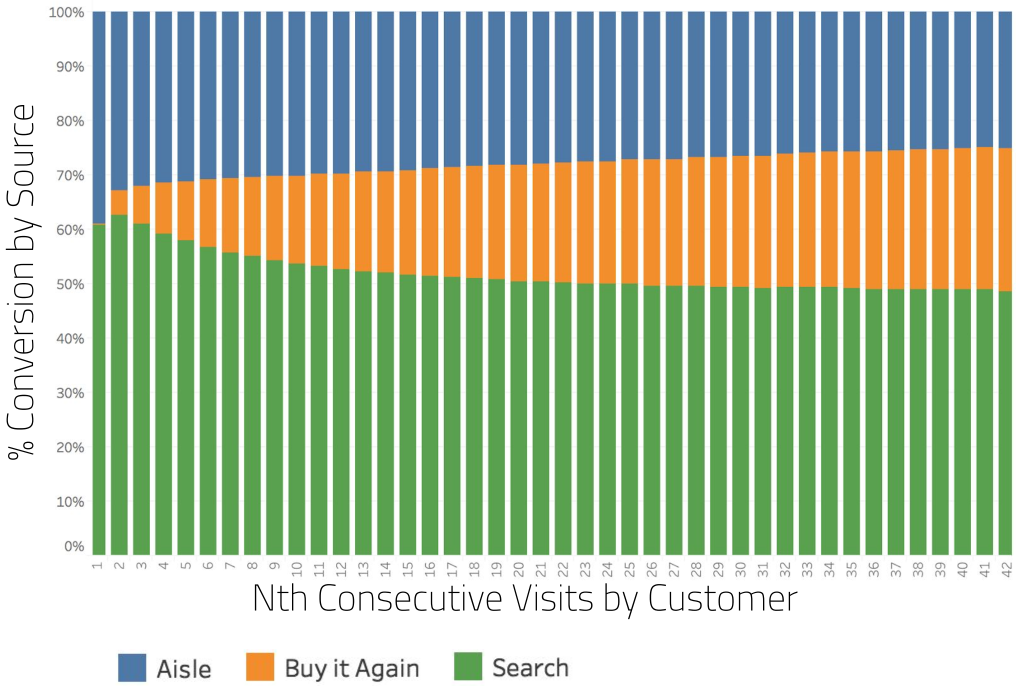
Handpicked by shoppers based on your preferences

Same-day delivery in as little as 1 hour

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

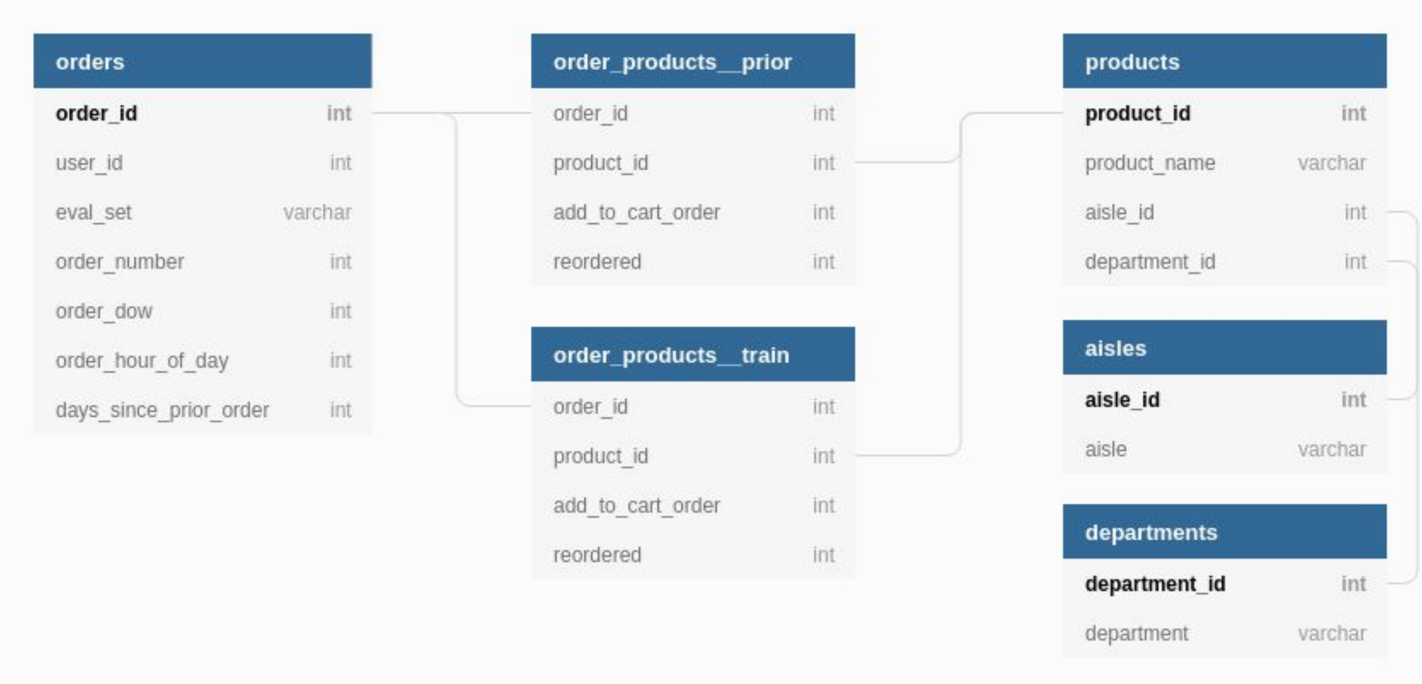
> “*How much value will be added if we choose to increase customer retention by improving the online shopping experience?*”

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



orders	
order_id	int
user_id	int
eval_set	varchar
order_number	int
order_dow	int
order_hour_of_day	int
days_since_prior_order	int

order_products__prior	
order_id	int
product_id	int
add_to_cart_order	int
reordered	int

order_products__train	
order_id	int
product_id	int
add_to_cart_order	int
reordered	int

products	
product_id	int
product_name	varchar
aisle_id	int
department_id	int

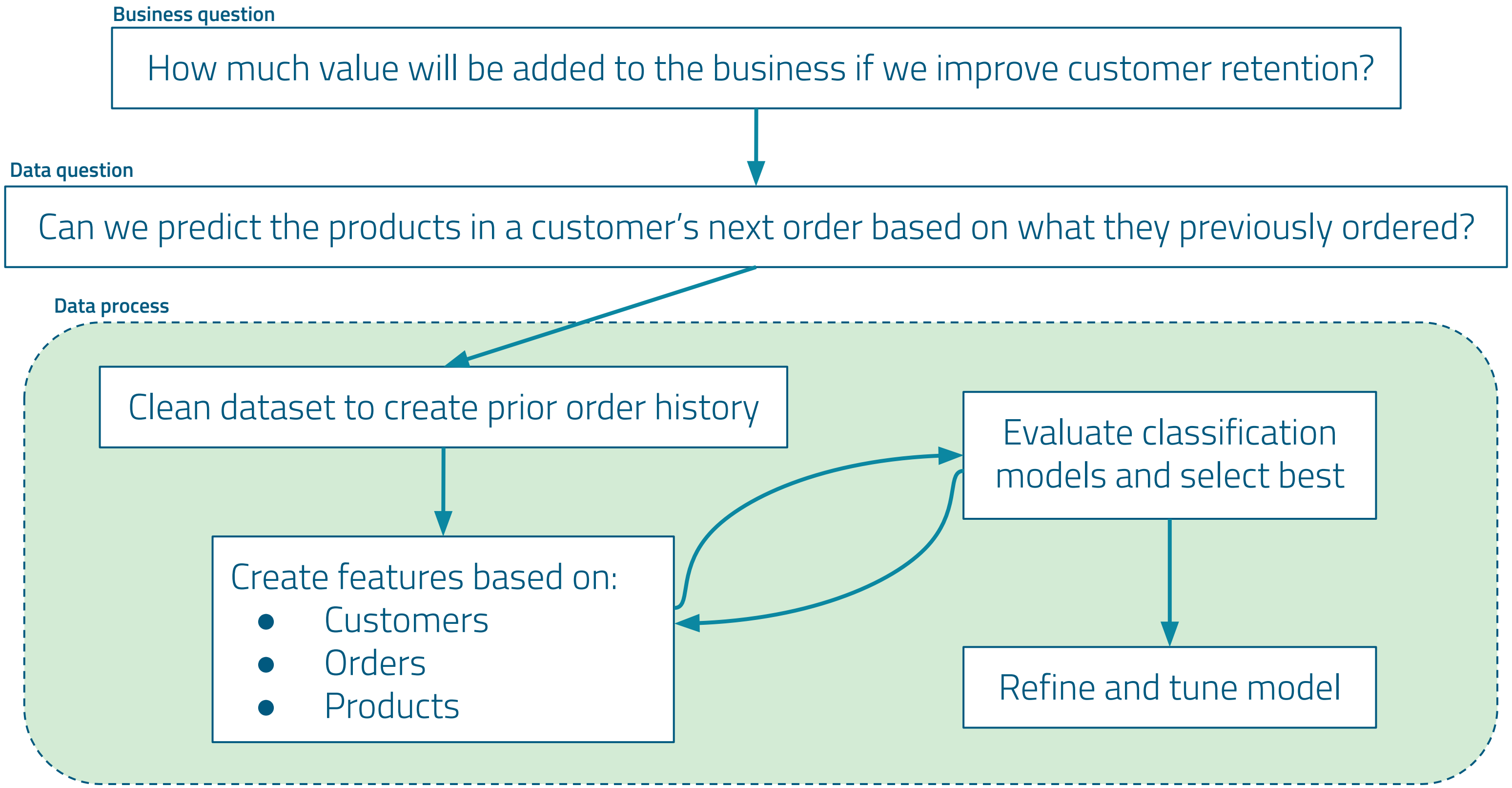
aisles	
aisle_id	int
aisle	varchar

departments	
department_id	int
department	varchar

AGENDA

1. Business problem definition
2. **Data process design**
3. Delivery
4. Next steps and summary

PROCESS WORKFLOW



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

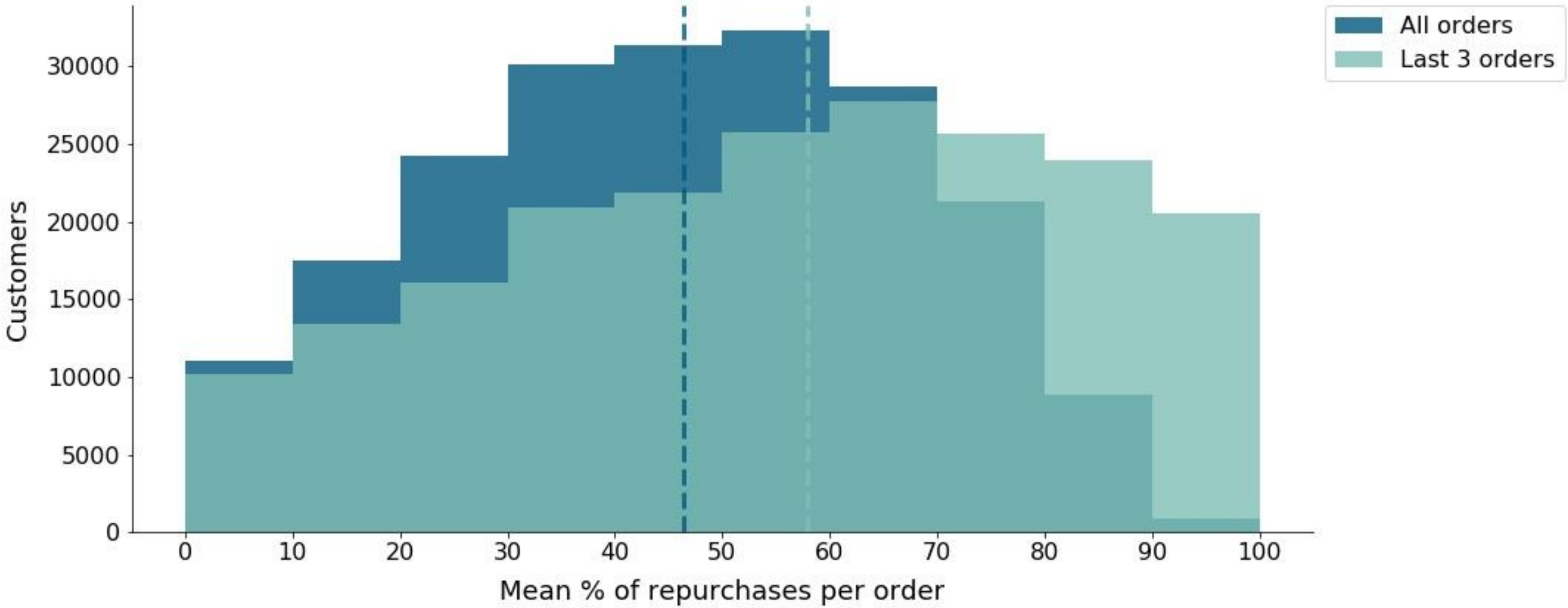
Evaluate classification models and select best

Refine and tune model

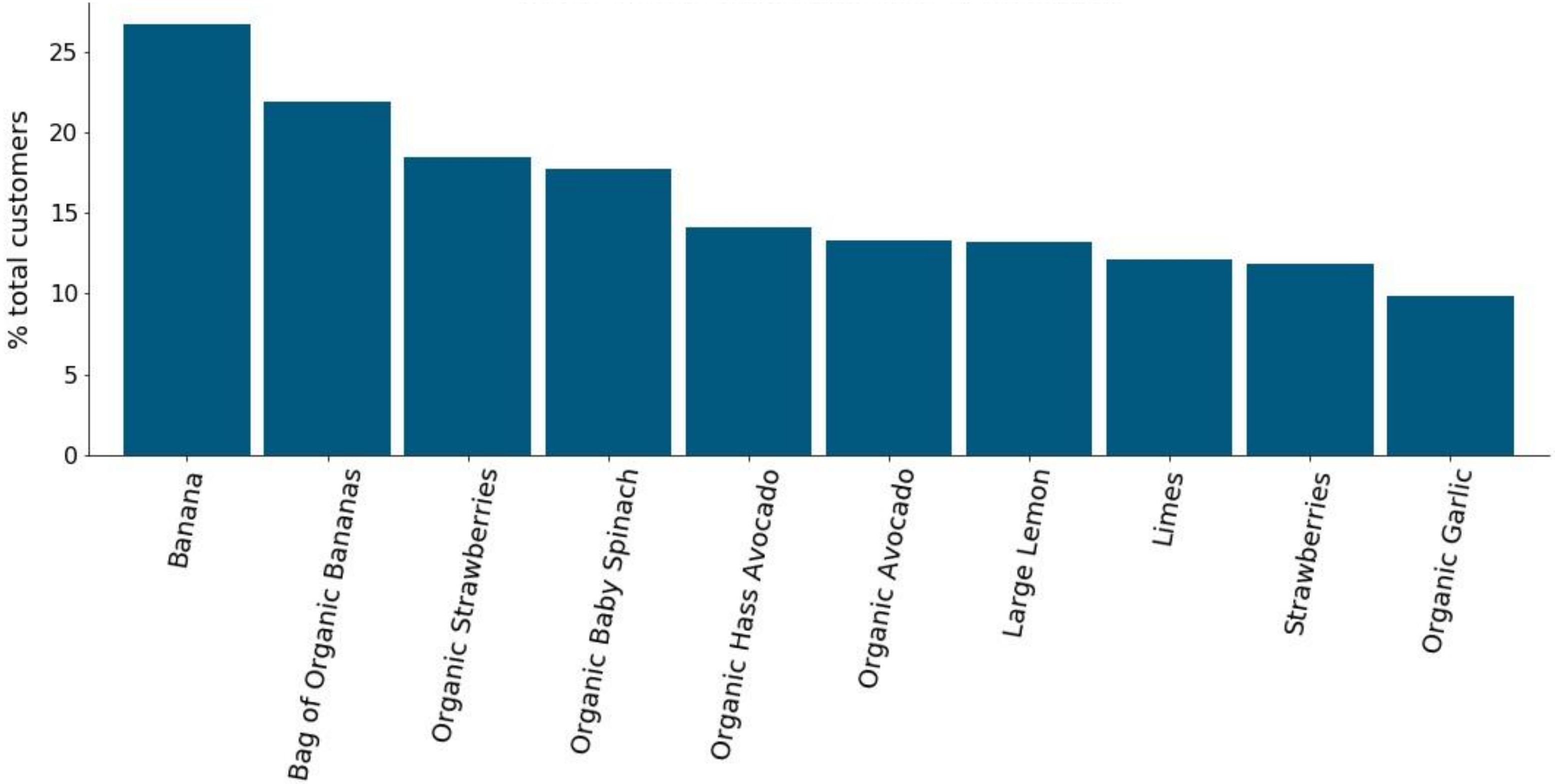
AGENDA

1. Business problem definition
2. Data process design
3. **Delivery**
4. Next steps and summary

PRODUCT REPURCHASE % PER ORDER



TOP 10 REPURCHASES BY CUSTOMER



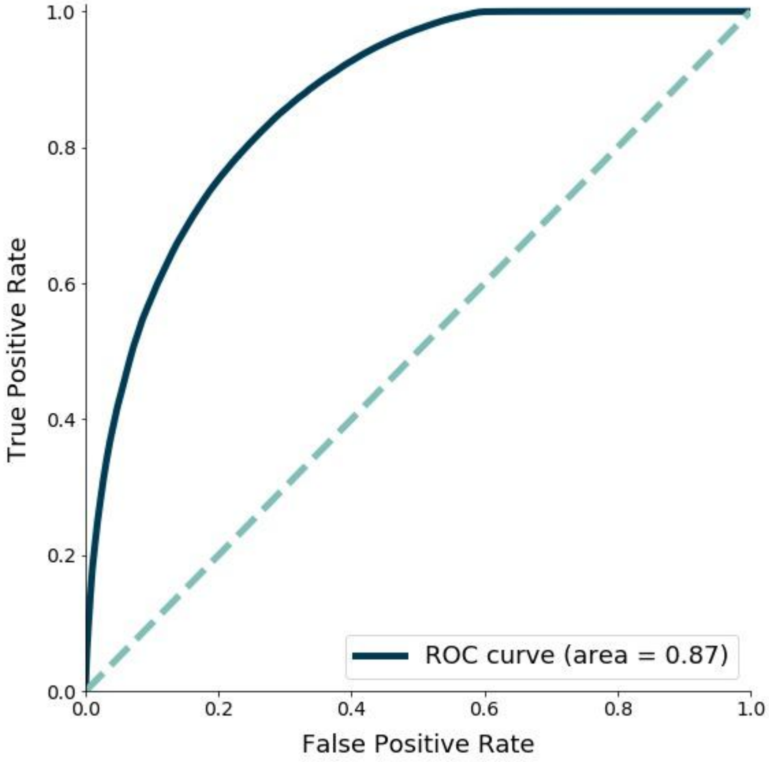
MODEL DEVELOPMENT & EVALUATION

- Models trained on 1.3 million observations and 14 predictors

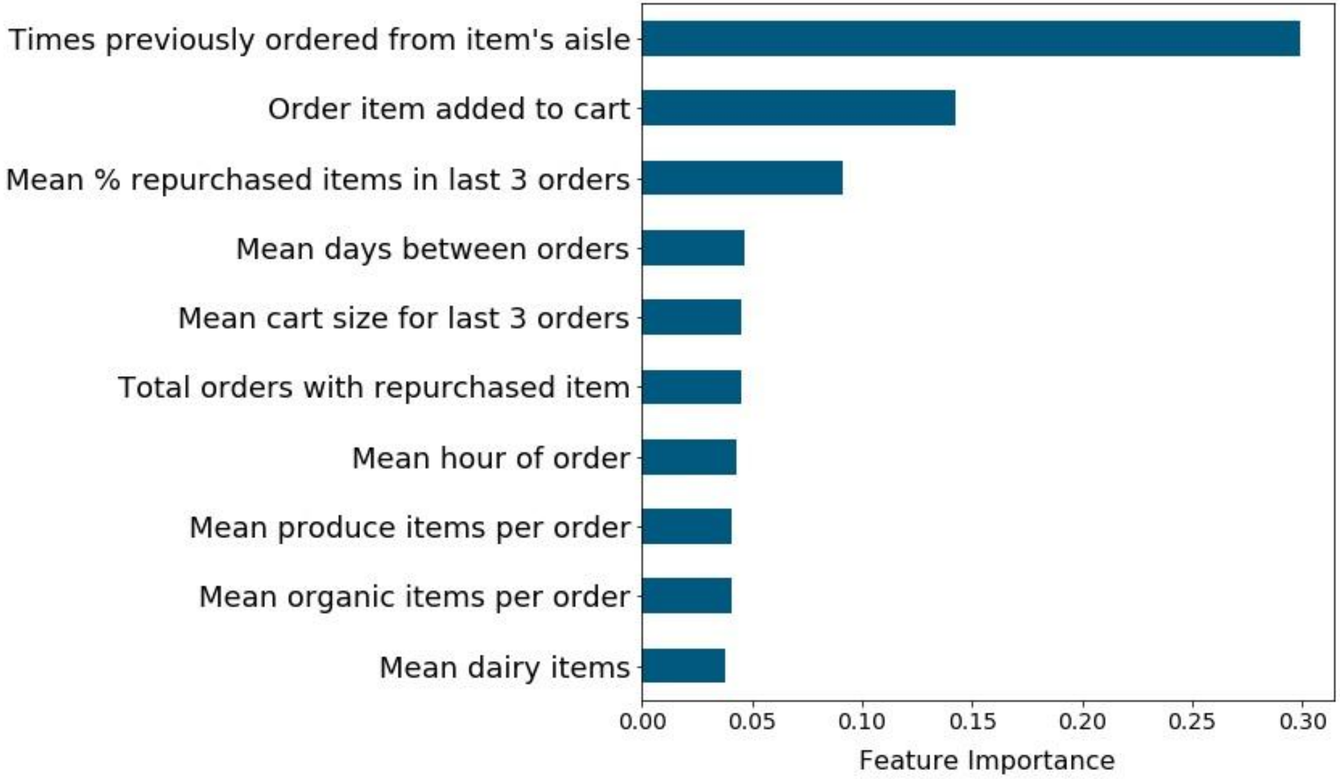
COMPLEXITY

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

- RF model (9 predictors) → **79%** accuracy



KEY FACTORS INFLUENCING REPEAT PURCHASES

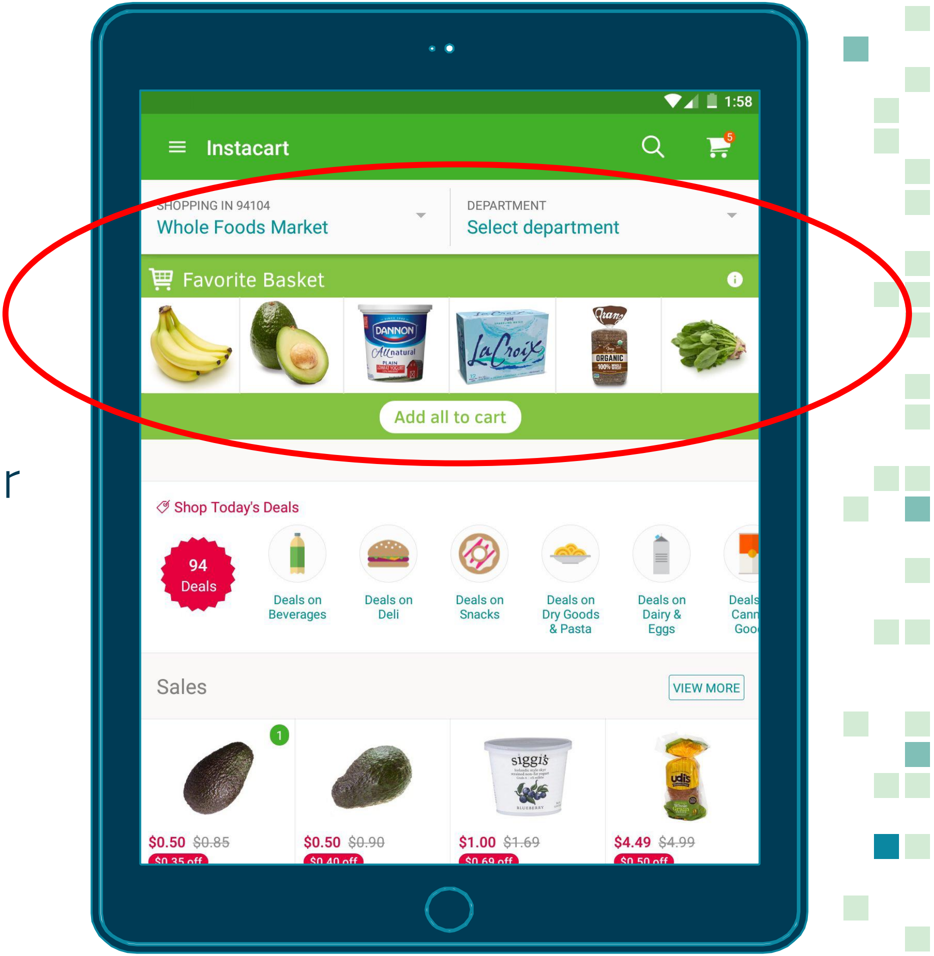


AGENDA

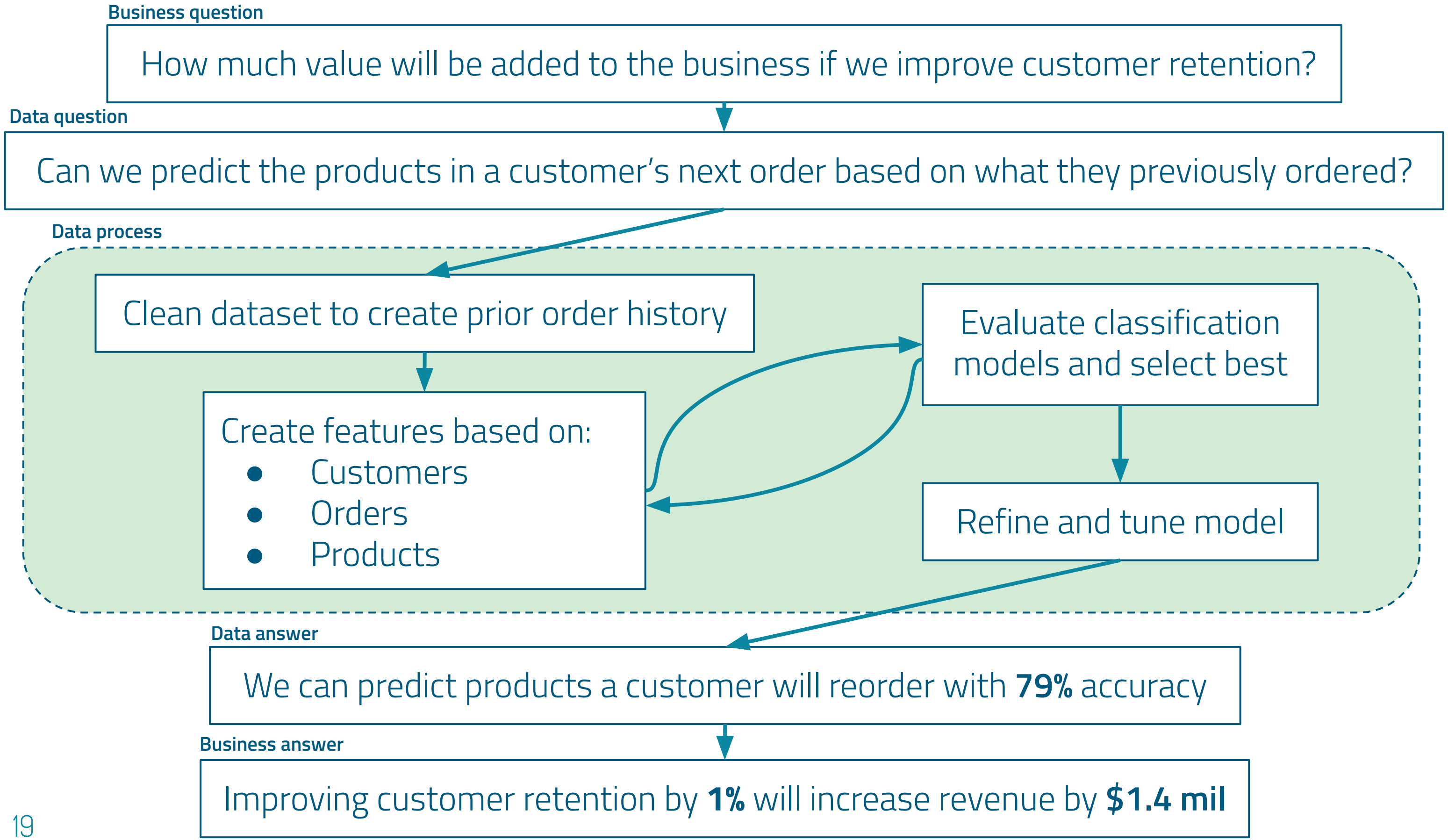
1. Business problem definition
2. Data process design
3. Delivery
4. **Next steps and summary**

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



SUMMARY



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 **79%** accuracy

Business answer

Improving customer retention by **1%** will increase revenue by **$1.4 mil**

THANKS!

Any questions?

Supporting documentation

@ github.com/maianhly/instacart_repurchases