B2B Sales:
Conversion and Revenue Prediction

Data Science Immersion Program Group 6

Faith Feng, Niall Keleher,
Jinyu Li, Shiya Yi

- Introduction
- Data
- Models & Results
- Future Work

- **Introduction**
- Data
- Models & Results
- Future Work

Project Overview

Goals:

1. Identify which B2B customers will place an order in the next 30 days (Conversion Rate)
2. For converted B2B customers, estimate the amount they will spend in the next 30 days (Revenue)
3. Business insights on how to increase expected revenue (Conversion Rate × Revenue)

Why important to predict B2B conversion rate & revenue?

B2B @ Wayfair

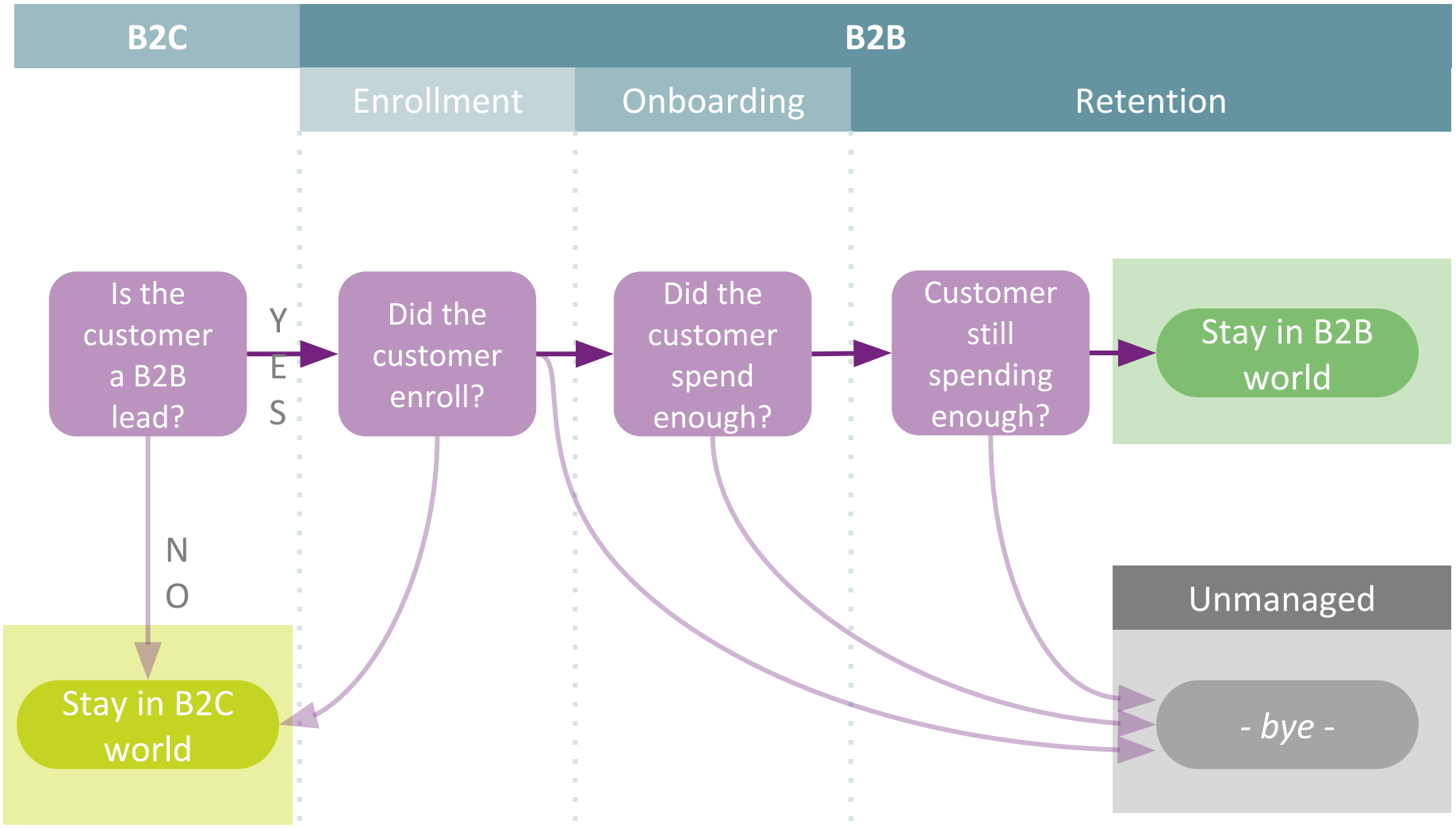
- **Increasing Customer Base**
- ***351K active customers, 114K inactive customers***

- **Higher Average Order Value (AOV)**
- ***B2B AOV $520 vs. B2C AOV $270***

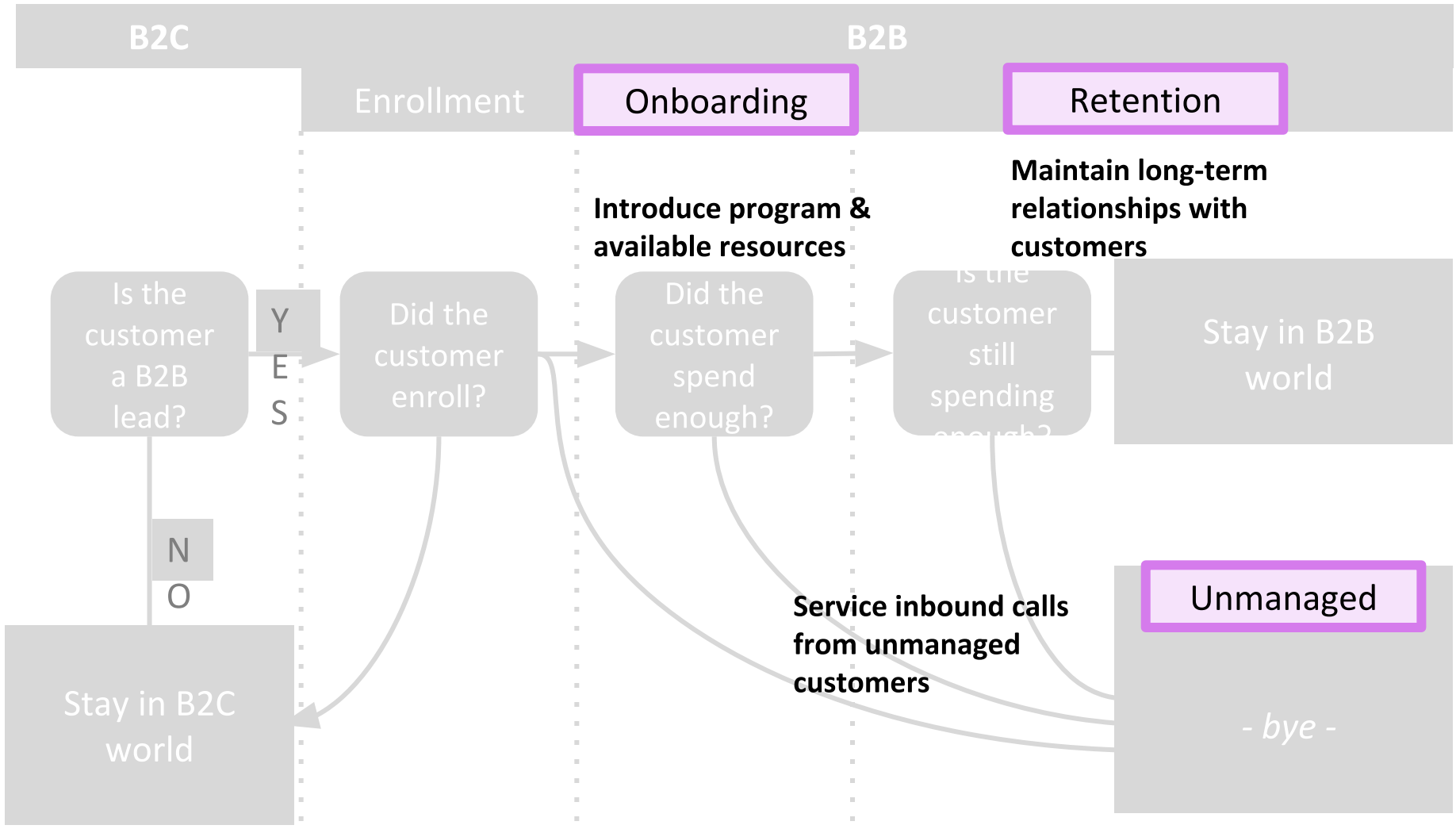
- **Different than B2C Customers**
- ***Purchase pattern --***
  - ***frequency***
  - ***seasonality***

- **Business Account Manager (BAM)**
- ***Onboarding BAM (OBAM)***
- ***Retention BAM (RBAM)***
- ***Unmanaged BAM (Ninja)***

B2B customer journey



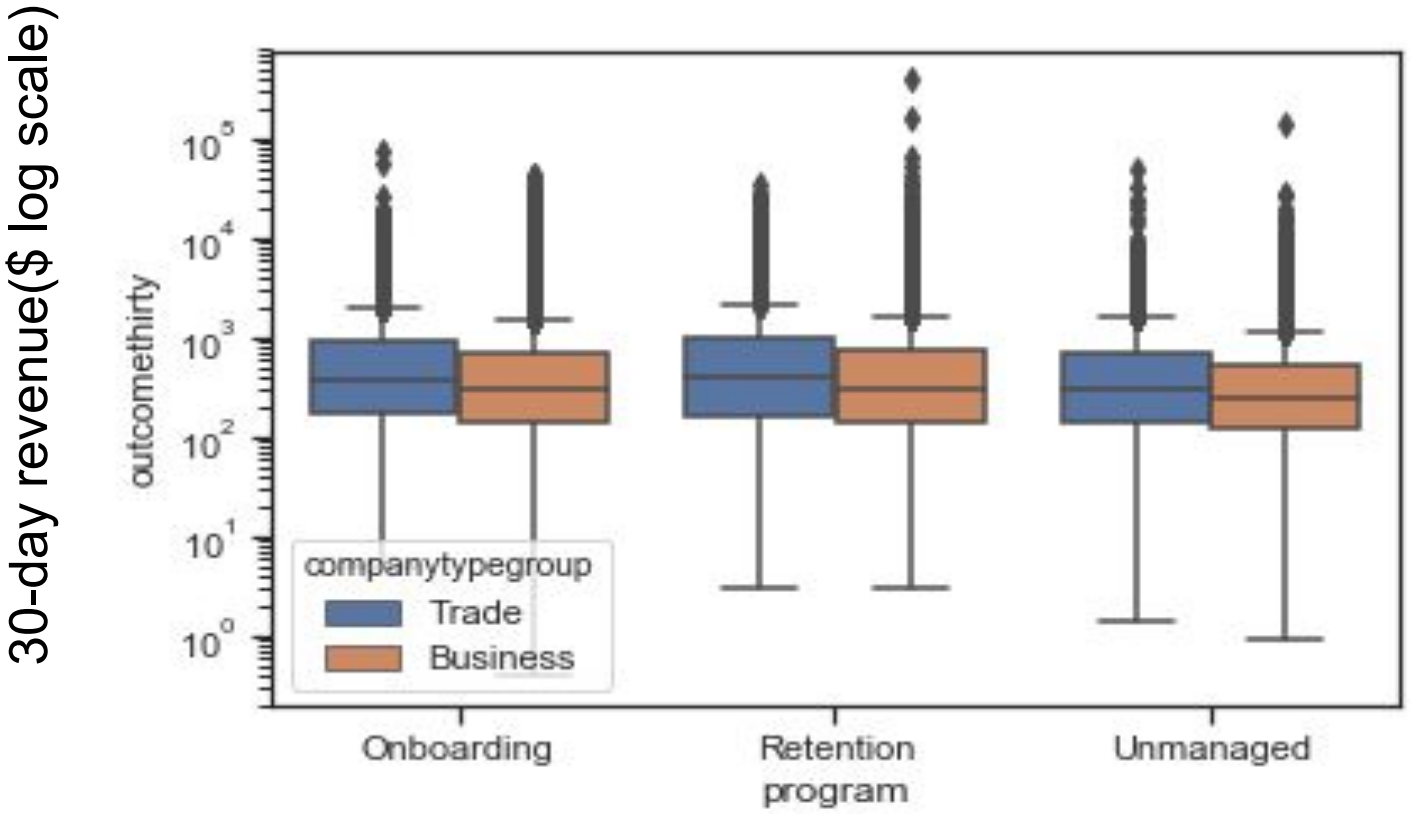
B2B Customer Journey & 3 Main Programs



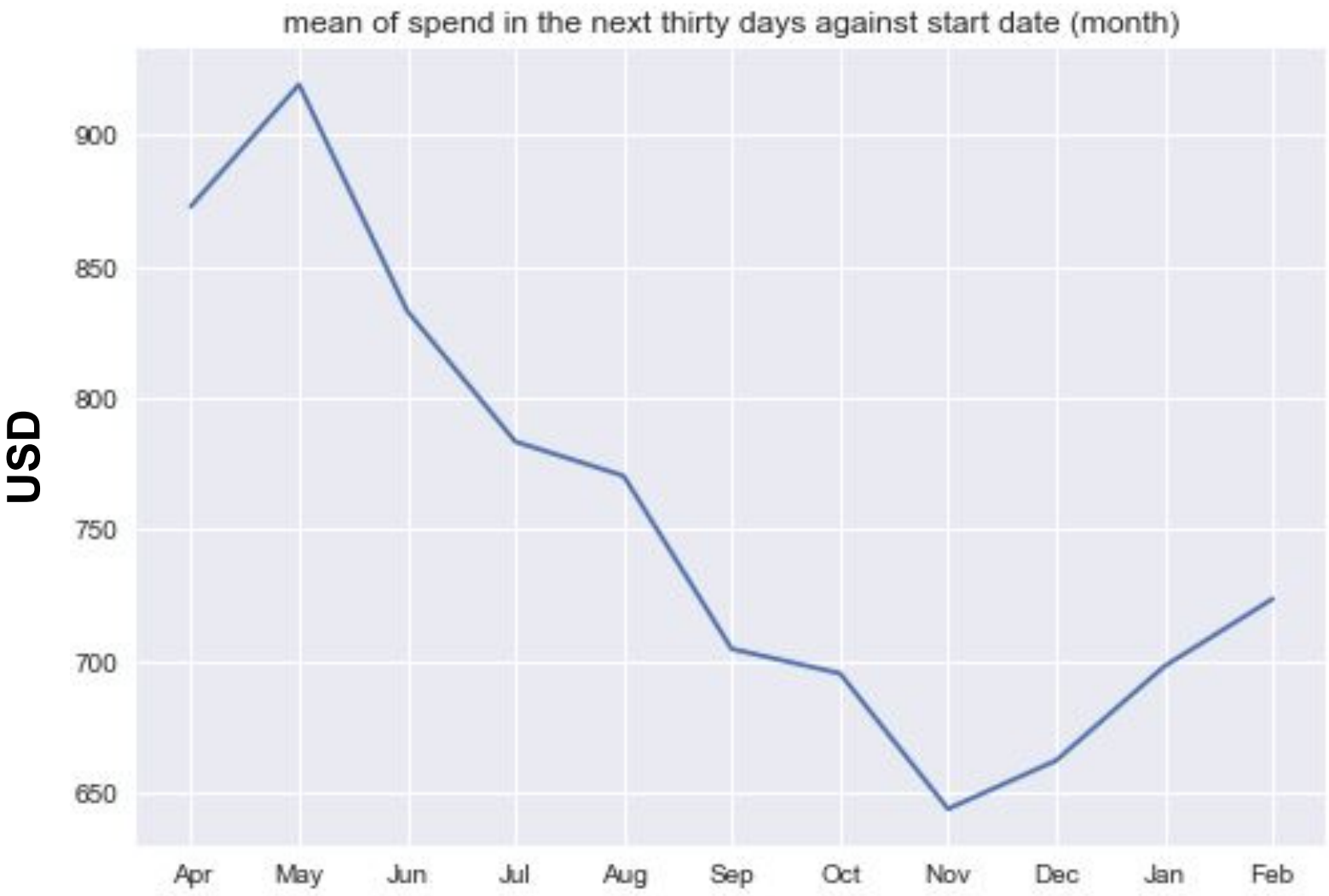
- Introduction
- **Data**
- Models & Results
- Future Work

Customer Type by B2B Program Status

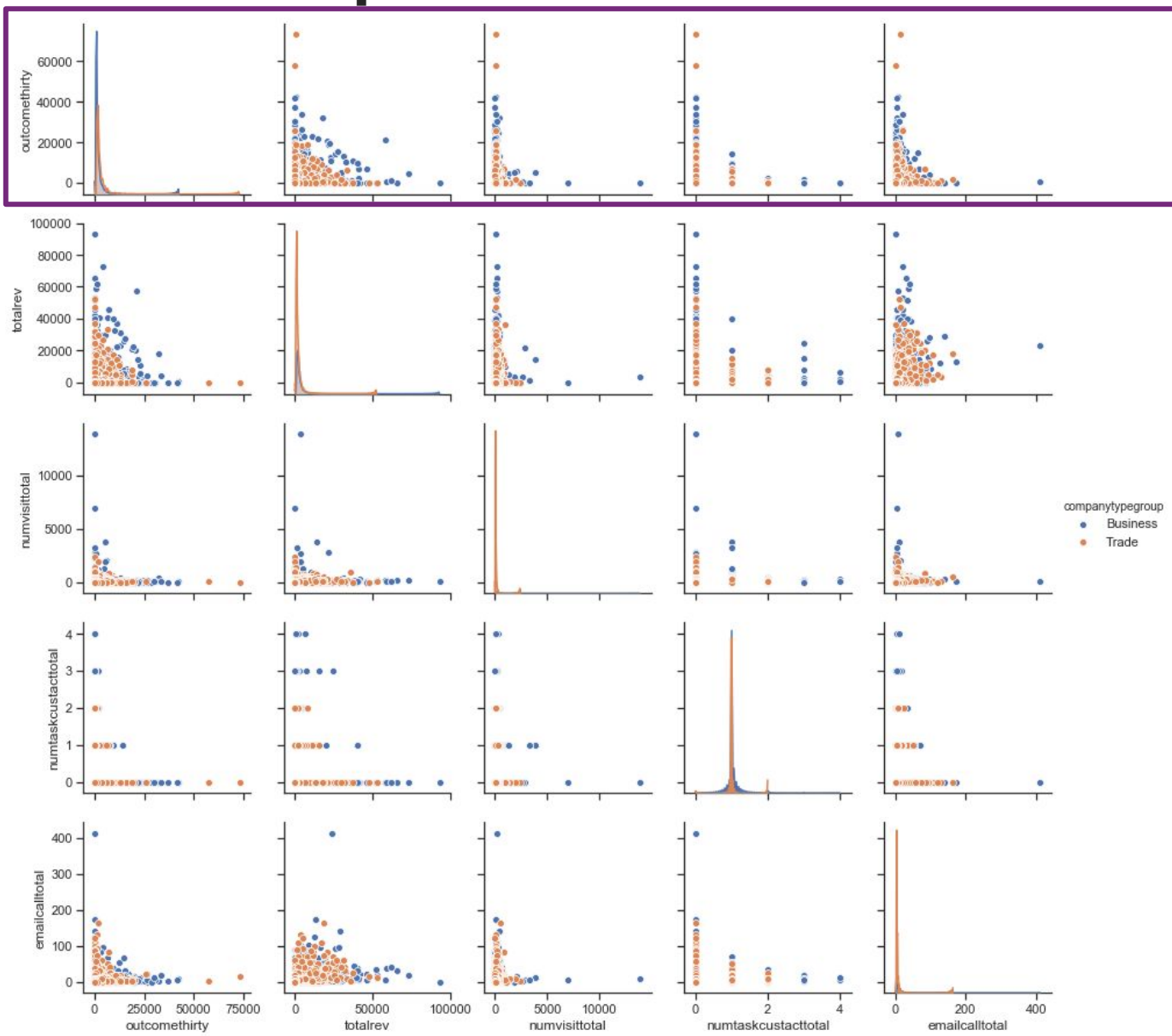
Onboarding	437,555	532,167
Retention	104,479	unique
Unmanaged	278,839	customers



Seasonality in Spending



Feature space



Target: 30-day revenue

Previous Yr Revenue

Site Visits

BAM Tasks

Customer Engagement

Preprocessing

- Missing value imputation
- Remove duplicate and irrelevant features
- One hot encoding of categorical features
- Create date-specific features (month, week, day of year)
- Feature aggregation and interactions

154 features as inputs to our models

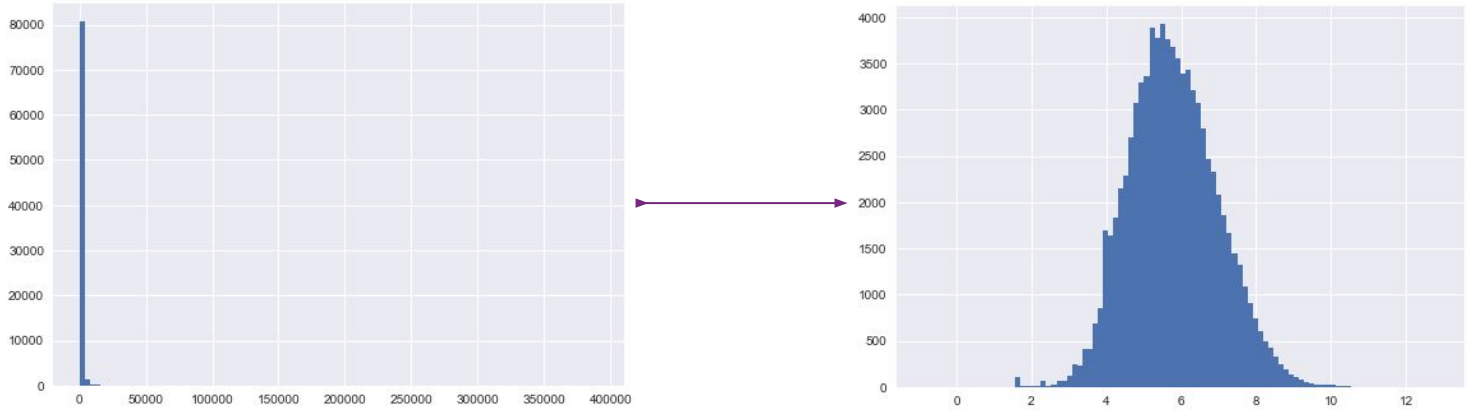
- Introduction
- Data
- **Models & Results**
- Future Work

Model

1. **Binary classification:** Predict which customers will place an order in the next 30 days
2. **Regression:** For those customers who placed an order in the next 30 days, predict how much revenue they will generate

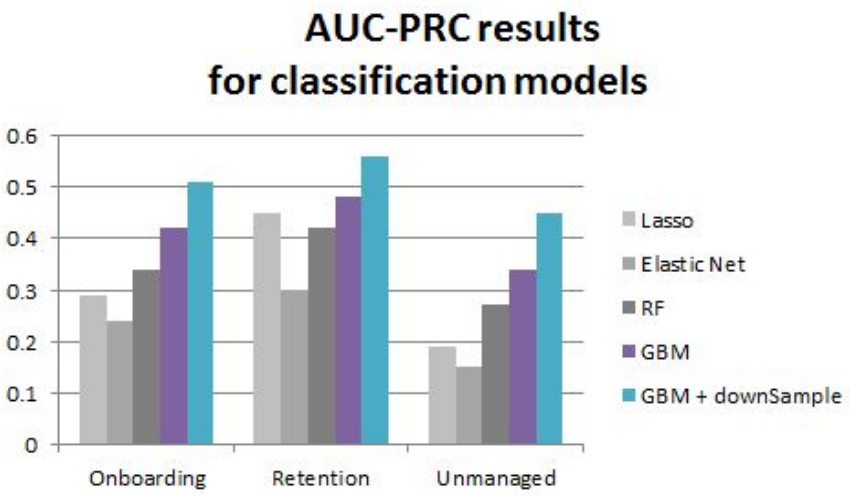
   non-normal distribution of target:

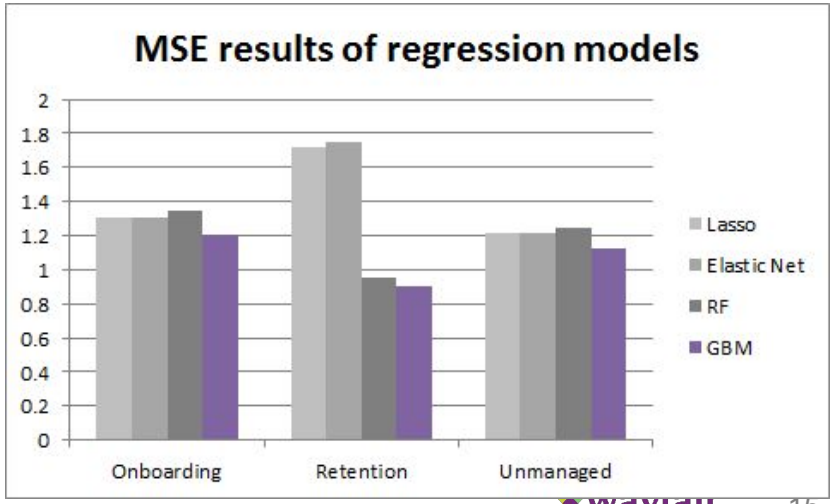
   - violate the assumptions of linear model
   - possible to get negative prediction values
   - solution: **log transformation**



Performance Comparison

- Compare models for performance:
  (classification: AUC-PRC, regression: MSE)
  - Lasso
  - ElasticNet
  - Random Forest
  - GBM
- Model improvements
  - Tuning parameters
  - Down sample/over sample
  - Model stacking (future work)

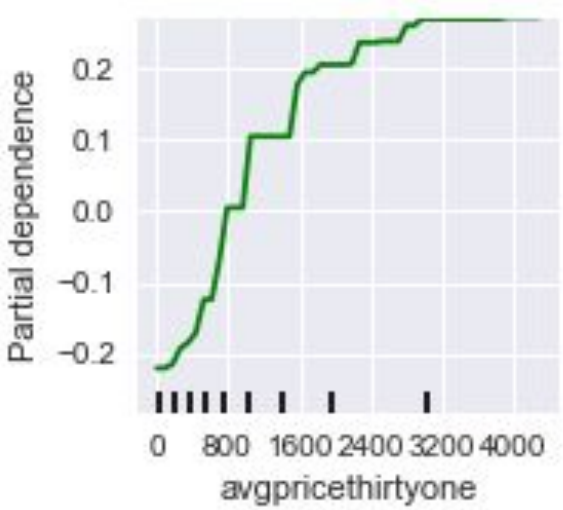


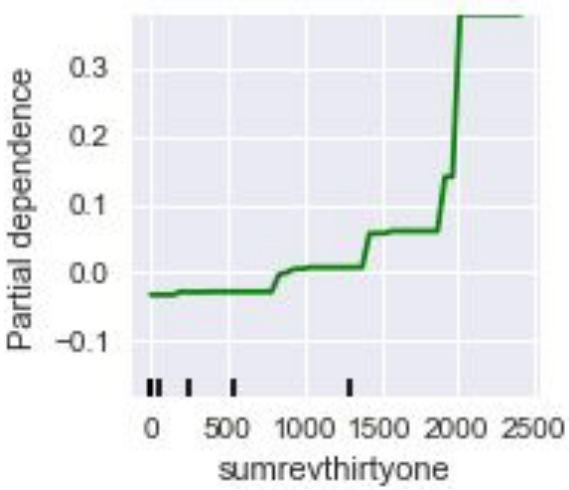


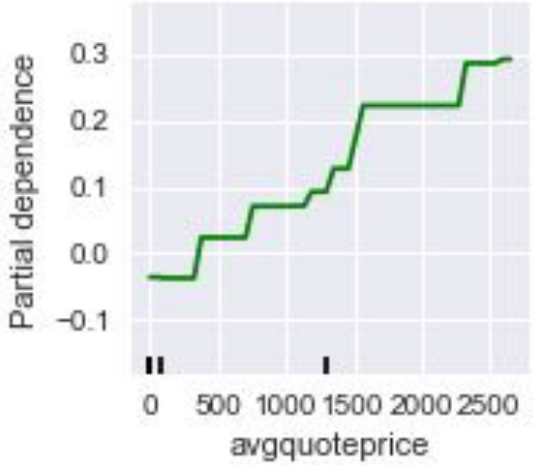
Results

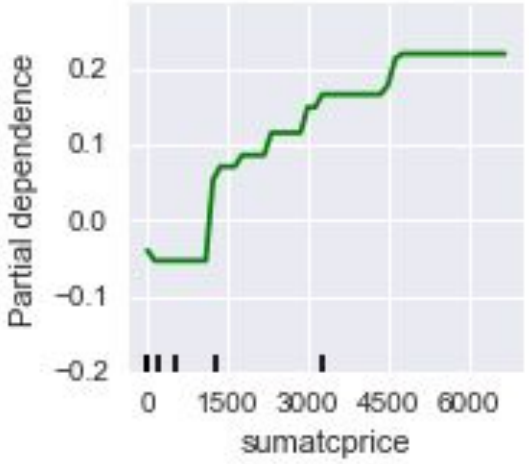
- **Important predictors for all program categories:**
  - Interaction with webpage (SKU viewed, ATC, click, idea board)
  - Number/amount of orders in the past year
  - Communication with Wayfair (email list, past calls)
- **Important predictors for:**
  - *Onboarding*: price of items viewed, whether the main orderer in a company, percentage of large orders
  - *Retention*: conversion rate of quotes
  - *Unmanaged*: days since last visit

Partial Dependence Plots









- Introduction
- Data
- Models & Results
- **Future Work**

Future Work

- **Caveat**: this does not suggest causality of the feature:
  - *Self-selection*: customers who are more interested in Wayfair might be more likely to subscribe to maillist and communicate with Wayfair.
- This project tells us the correlation between features and revenue. Useful for projecting expected revenue.
- Not appropriate to evaluate counterfactuals.
- Need more causal inference to tell which channel we should focus on in order to increase profit.

Future Work: Experiment

	Call	Not Call
Convert	Complier	Always-taker
Not Convert	Never-taker	Defier

- ***Experiment:***
  - Making calls to customers is **expensive ($16/call)**
  - Don’t waste money on customers who are never going to convert and who will convert no matter what.
  - Estimate local average treatment effect (LATE) on compliers.
  - Infer whether a potential customer can be influenced, predict customer lifetime value.

Future Work: Model and Data

- ***Model refinements:***
  - Different customer programs have different rate of conversion and expected revenue.
  - Model the transition among those groups, e.g. Markov-switching model.

- ***Additional data:***
  - More detailed categories of *business type* and *business size*.
  - Fast-growing companies and large companies might tend to buy more.

Thank you!
Questions?

Faith Feng, Niall Keleher,
Jinyu Li, Shiya Yi