Starbucks Case

Optimize targeting with A/B Testing and Uplift Modeling

Linh Chu

Project Overview

Context

- In the experiment simulated by the data, an advertising promotion was tested to see if it would bring more customers to purchase a specific product priced at $10.
- Since it costs the company $0.15 to send out each promotion, it would be best to limit that promotion only to those that are most receptive to the promotion.

Data: from a randomized experiment

ID	promotion	purchase	V1	V2	V3	V4	V5	V6	V7
1	No	0	2	30.443	-1.165	1	1	3	2
3	No	0	3	32.159	-0.645	2	3	2	2
4	No	0	2	30.431	0.133	1	1	4	2
5	No	0	0	26.588	-0.212	2	1	4	2
8	Yes	0	3	28.044	-0.385	1	1	2	2

- `promotion` – Indicates if the customer was part of treatment or control
- `purchase` – Indicates if the customer purchased the product
- `ID` – Customer ID
- `V1` to `V7` – features of the customer (has been anonymized)

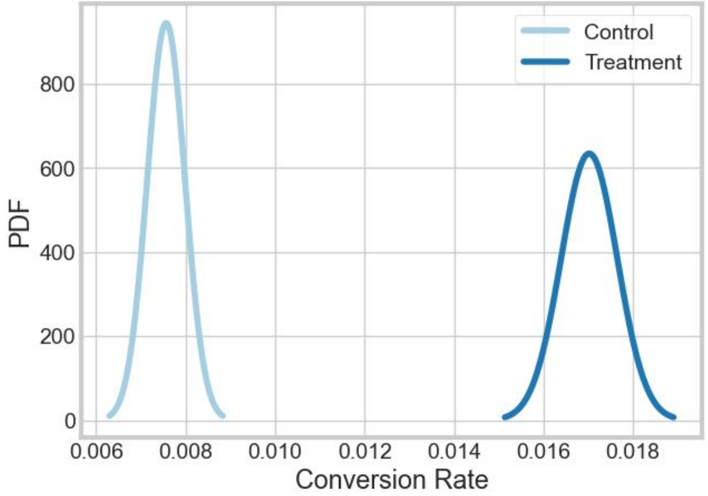
Objectives

1. Analyze the results of the experiment and identify the effect of the promotion on product purchase and Net Incremental Revenue
2. Build a model to select the best customers to target that maximizes the Incremental Response Rate and Net Incremental Revenue

Test Result Analysis

Group	Size	Purchase	Response Rate	Revenue	Cost	Net Revenue
Control	42,170	319	0.76%	$3,190	$0	$3,190.0
Treatment	42,364	721	1.70%	$7,210	$6,354.6	$855.4
Total	84,534	1040	1.23%	$10,400		

- **Incremental Response Rate (IRR)** = $RR_{treat} - RR_{ctrl}$ = 0.95%
- **Net Incremental Revenue (NIR)** = $\text{Net Revenue}_{treat} - \text{Net Revenue}_{ctrl}$ = −$2334.6



Statistical Test

- Run one-sided z-test to detect the increase in response rate
  - Result: P-value = 5.5e-36 < 0.05 → statistically significant

→ Customers receiving the promotion would purchase at a higher rate of 0.95%

- However, NIR < 0 indicates that even though the promotion increased the likelihood of purchase, the cost of sending the promotion outweighed the revenue gained from those extra purchases

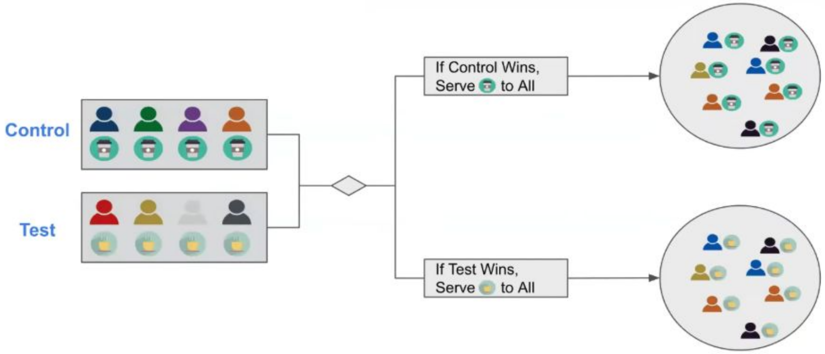
→ The promo works — but **not efficiently** enough when sent to everyone

→ Need to **identify only the customers who are truly influenced by the promotion** instead of mass selection

How can we personalize marketing strategies based on A/B testing results?

Where A/B Testing falls short

Standard A/B Testing using
ATE (Average Treatment Effect)

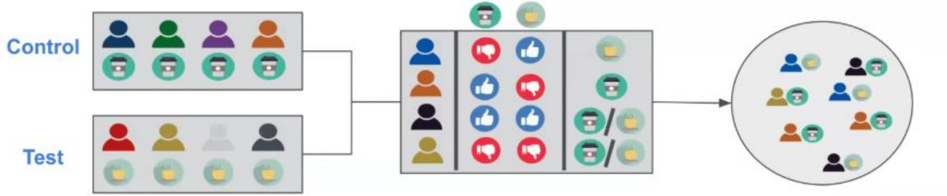


→ Measures overall **average** impact
→ optimize general user experience

Uplift Modeling optimizes for incremental effect and enables personalized treatment

Uplift Model estimates heterogeneous treatment effects with ML algorithm:
Conditional average treatment effect

CATE = E [Y | Intervention, X] − E [Y | No Intervention, X]



Estimating CATE by Uplift Modeling

Launch Personalized Experience

→ Identify differential impact across individuals/segments
→ Allow precise targeting of high-responding segments

Source: Uber CausalML documentation

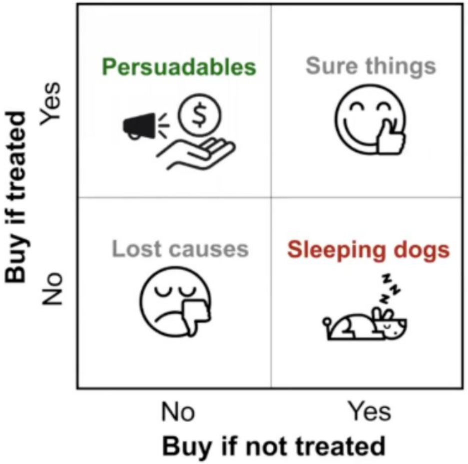
Uplift Model Approach

ID	T (promotion)	Y0	Y1	TE (tau) = Y1 − Y0
1	0	0	?	N/A
2	1	?	0	N/A
3	0	1	?	N/A
4	0	0	?	N/A
5	1	?	1	N/A

use ML model → to impute

Y0 hat	Y1 hat	TE (tau hat)
–	0.23	0.23
0.11	–	-0.11
–	0.9	-0.1
–	0.7	0.7
0.27	–	0.73

Rank customers by treatment effect → + identify the potential segments

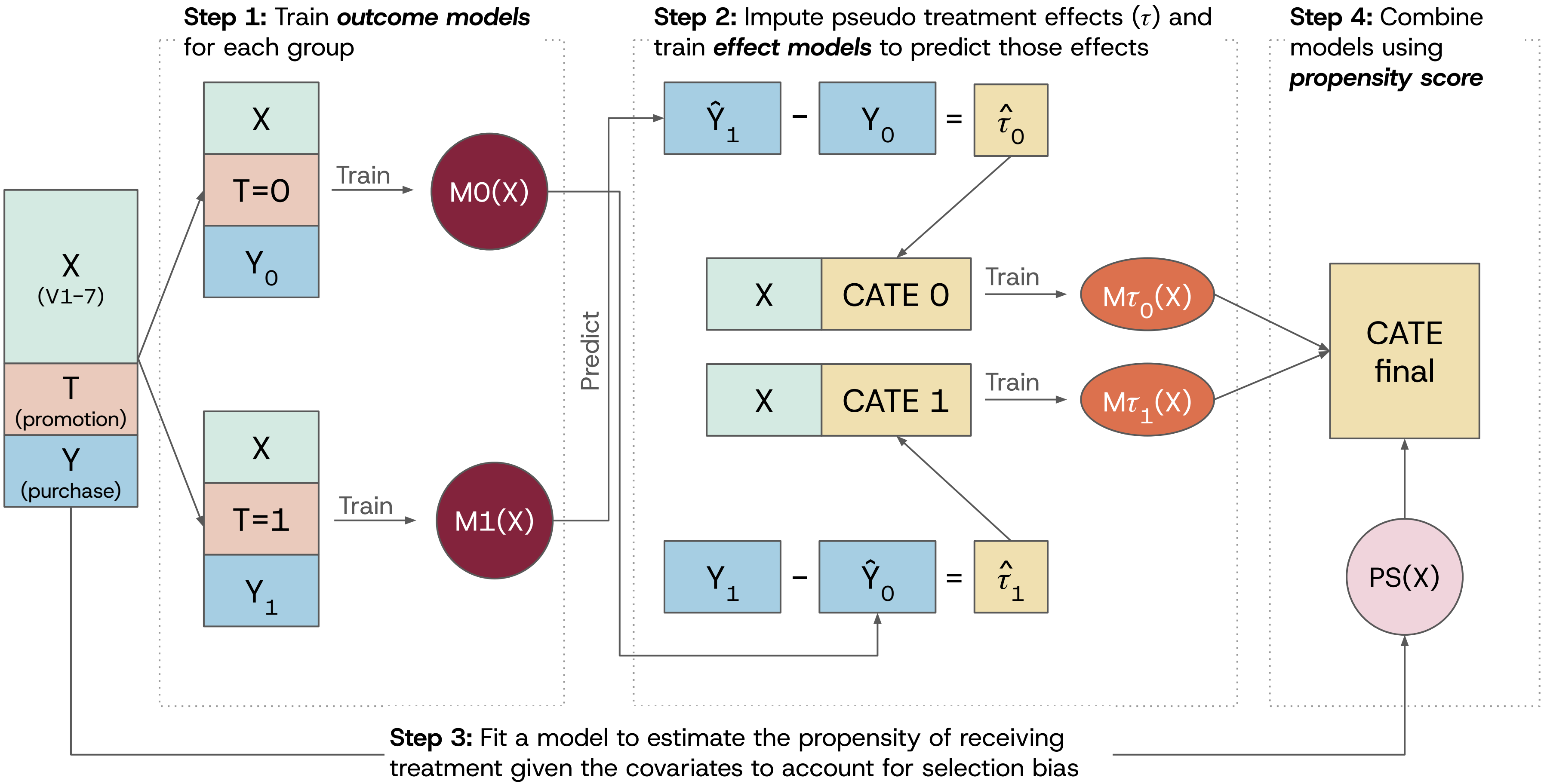


Sure things & Lost causes – Will behave the same no matter what you do. Including them as a target in the model is okay, but will make our targeting inefficient.

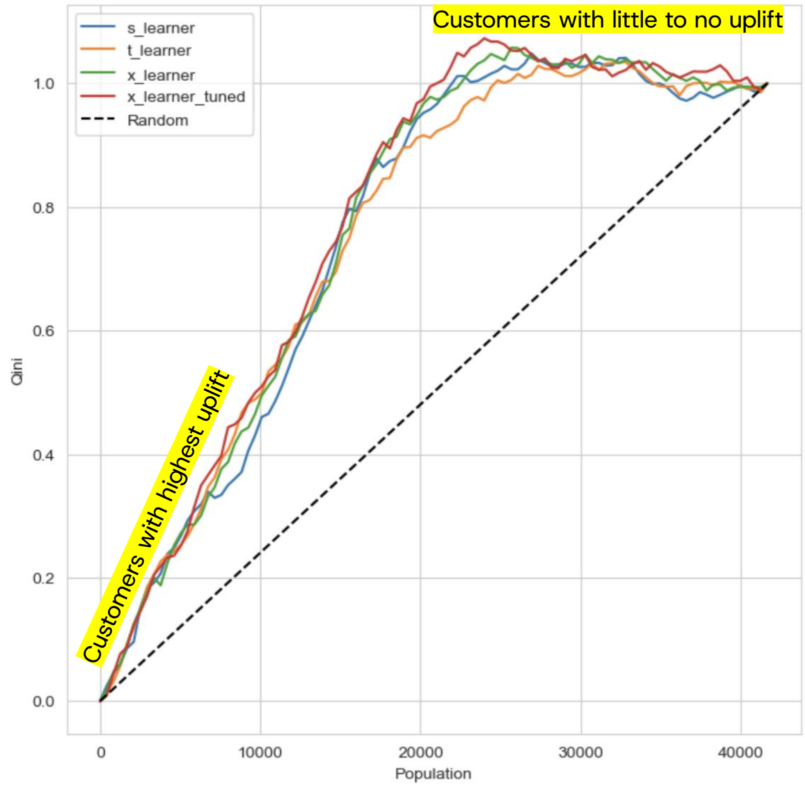
Sleeping dogs - These people are turned off by your intervention. Definitely don't include them, ideally you would even downrank them.

Persuadable - This is the population you actually care about because they exhibit the ideal behavior *because* you intervened. Ideally you uprank them as much as possible ✓

Model Methodology: X-Learner



Model Evaluation on Test Set



Model Evaluation:

- Based on Qini coefficient which quantifies the area under the curve → greater area under the curve implies better model performance
- Based on Net Incremental Revenue (NIR)

Model	Qini score	IRR	NIR
None		0.96%	-$1,132
S-learner	0.2544	2.04%	$453
T-learner	0.2535	1.98%	$377
X-learner	0.2667	2.04%	$465
X-learner (after tuning)	0.2794	2.06%	$541

→ By selecting customers with highest treatment effect, we achieved NIR of $540, improved by $1673 compared to random mass promotion

What’s ‘driving’ the lift?