Wine Retailer Case Study

MSBA Cohort 2 Team 26
Ching-An Chung
Jichao Gui
Zheng Yang
Ruxin Cheng
Weiming Zhu

Methodology

1. Firstly, we figure out the average causal effect to help us get understanding of email's effect in each variable.

2. After finish Slice and Dice Analysis based on different baseline variables, we know that we should target different consumer to send Email.

3. Then, by using the uplift model, we calculate the effect that sending the email has on each customer, and identify our target customers who bring us positive profit

4. Last, using the casual forest model , we identify the target customers who have positive purchase amount after receiving an email.

Average Causal Effect

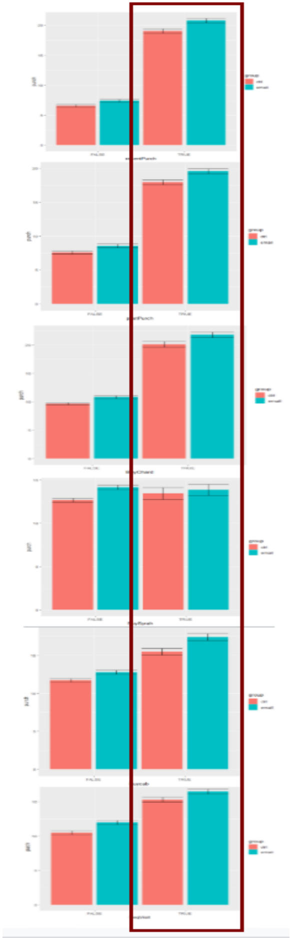
We ran a regression with these baseline variables. Y variable is purch and we want to predict the effect of group.

We also ran a casual forest to find the average effect when people receive our email.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.5269957  0.4363336  33.293  < 2e-16 ***
groupemail   1.2603997  0.3101382   4.064 4.83e-05 ***
chard        0.0346117  0.0007959  43.489  < 2e-16 ***
sav_blanc    0.0433309  0.0020630  21.004  < 2e-16 ***
syrah        0.0240070  0.0149648   1.604    0.109
cab          0.0489413  0.0020948  23.363  < 2e-16 ***
last_purch  -0.0718125  0.0017235 -41.667  < 2e-16 ***
visits      -0.0627548  0.0655217  -0.958    0.338
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> print(cf)
GRF forest object of type causal_forest
Number of trees: 2000
Number of training samples: 78312
Variable importance:
    1     2     3     4     5     6
0.244 0.047 0.293 0.250 0.047 0.118
> average_treatment_effect(cf, method="AIPW")
 estimate   std.err
1.3146168 0.3096089
```

Slice and Dice Analysis



1. Last_purch < 63 days

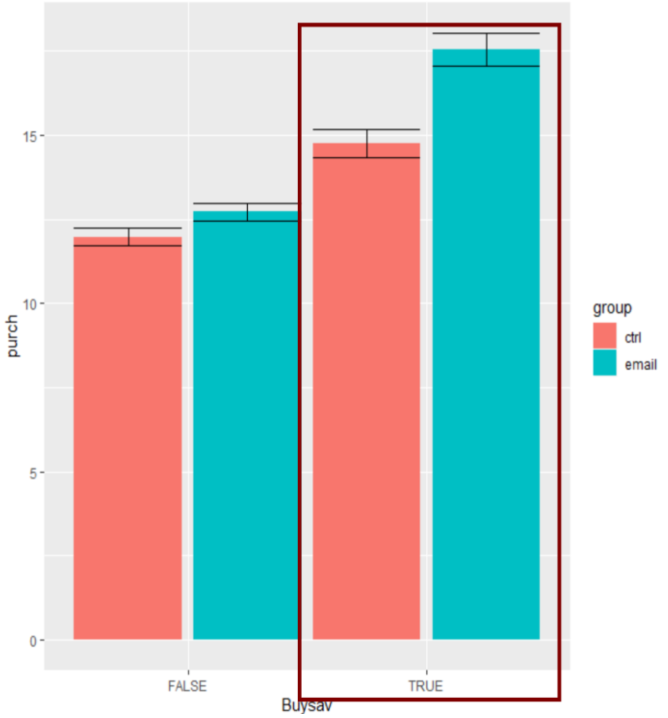
2. Past Purchase > 52

3. Buy Chard before

4. Buy Syrah before

5. Buy Cab before

6. Website visit > 5



Only 1 Significant Interaction:
7. Sav_blanc

BuysavTRUE	2.7750	***
emailTRUE:BuysavTRUE	2.0504	**

Conclusion:

1. Except 4, other subgroup buy more on average (circle by rectangle)
2. E-mail has different effect for different subgroup
3. Target different consumer to send Email.

- save cost
- improve dropped emails or lower responsiveness.

Uplift Model

- Variables:

  Independent variables in the model:

  Number of visiting the website: visits

  Four wines: chard, sav_blanc, syrah, cab

  Group: email

  Dependent variable:

  Purchase: purch

- Calculation:

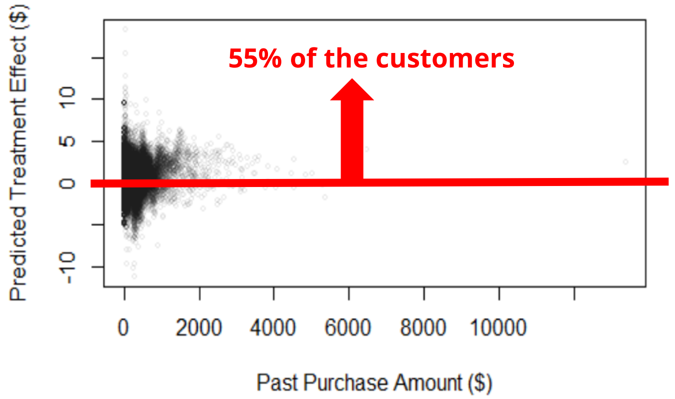
- Holding variables visits and wines fixed for each customer ID, we predict two purchase values in the condition that: a) the email is sent to the customer; and b) the email is not sent to the customer
- Lift: Difference between the two predicted values is the effect that sending email has on customers
- Margin(30%) and cost(10 cent/email) taken into account:

  Lift*Margin-Cost > 0 → Target Customer

Casual Forest Model

Final Model Selection:Casual Forest Model
We chose the causal model as the final model for two reasons.
1.It works better with larger numbers of baseline variables.
2.It works better when there are non-linear relationships.

We target customer with positive purchase amount after receiving the email.



Summary

Analysis：

- Slicing & Dicing
- Uplift
- Casual Forest

Casual Forest Result：

Summary for targeted customers:
55% of the customers
32% bought Chard
34% bought Sav_blanc
10% bought Syrah
32% bought Cab
50% purchased wine 53 days ago
54% had past purchase amount larger than 52
47% visited the company’s website more than 5 times.

Thank you.