

Customer's Activation

Liaci Emanuela

2024-06-07

Data

- Data for the analysis are stored in the csv file sample_data.csv
- Assume that the mailing was delivered on October 1st, 2021 (time point is called t_0 in the following)
- The following variables are provided:
 - ID: Customer identifier
 - Activated: Was the customer activated by the mailing within a certain period of time p after t_0 ?
 - Cust_Age: Customer age at t_0
 - Cust_Gender: Customer gender at t_0
 - Cust_NL: Did the customer receive our newsletter at t_0 ?
 - Orders: Number of orders within 12 months before t_0
 - Orders_Voucher: Number of orders in which a voucher was redeemed within 12 months before t_0
 - LastOrder: Date of the last order within 12 months before t_0
- The imported data should look similar to this preview:

##	ID	Activated	Cust_Age	Cust_Gender	Cust_NL	Orders	LastOrder	Orders_Voucher
## 1:	1	1	53	m	0	3	2021-07-24	0
## 2:	2	0	46	f	0	2	2021-04-11	1
## 3:	3	0	28	f	0	1	2021-01-14	0
## 4:	4	1	53	f	0	3	2021-07-20	0

Which measure(s) would you use to evaluate the performance of the mailing?

From the data:

Main KPI

$$\text{Activation Rate} = \frac{\text{Number of Activated Customers}}{\text{Total Number of Customers}}$$

43.49%

Additional measures:

1. **Response Rate:** The proportion of customers who responded to the mailing in any way (e.g., inquiry, website visit) within the specified period. This measure can be broader than activation. Calculation: Similar to activation rate, but includes any form of response.
2. **ROI (Return on Investment)**

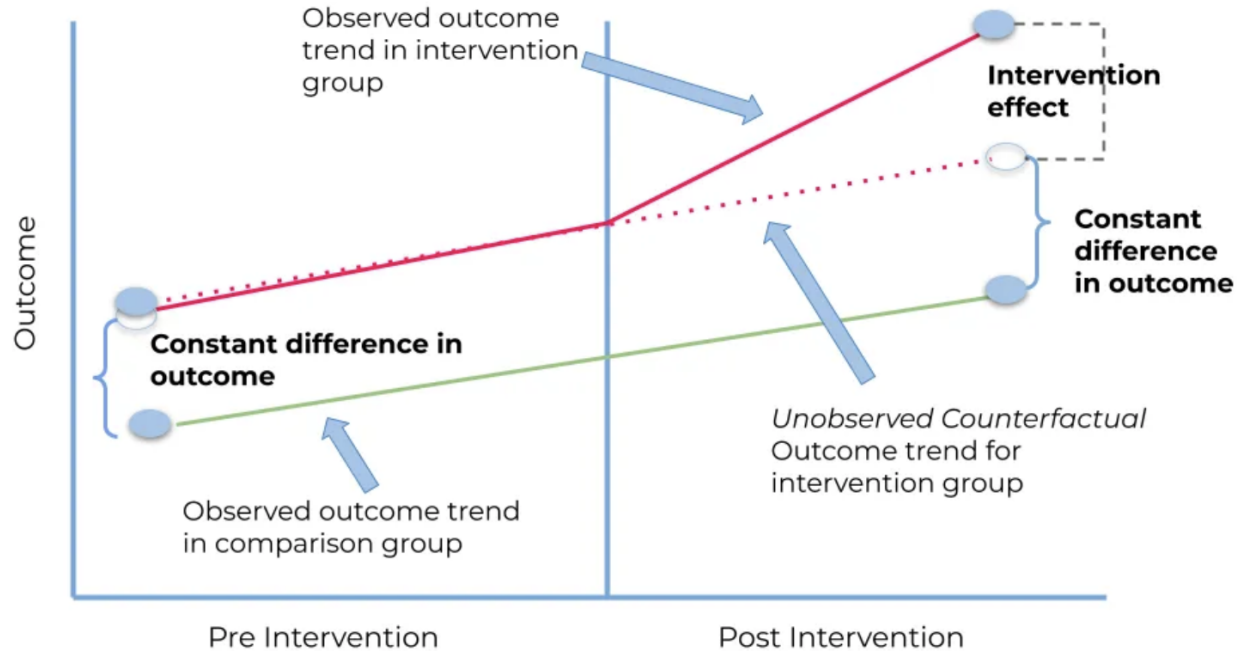
$$\text{ROI} = \frac{\text{Revenue from Activated Customers} - \text{Cost of Mailing}}{\text{Cost of Mailing}}$$



An A/B test can be run on the main KPI (defining MDE on Activation Rate):

- Treatment Group: Customers who receive the print mailing.
- Control Group: Customers who do not receive the print mailing.

When an A/B Testing is not possible... Quasi-experiment



Difference-In-Difference is a method of estimating a causal effect by comparing the change between two groups over time.

What are the factors which explain if the customer gets activated by the mailing?

Logit Regression Results

```
=====
Dep. Variable:          Activated    No. Observations:          10000
Model:                  Logit        Df Residuals:              9994
Method:                 MLE          Df Model:                  5
Date:                   Thu, 30 May 2024    Pseudo R-squ.:            0.1602
Time:                   18:50:03           Log-Likelihood:           -5749.4
converged:              True            LL-Null:                  -6846.5
Covariance Type:        nonrobust        LLR p-value:              0.000
=====
```

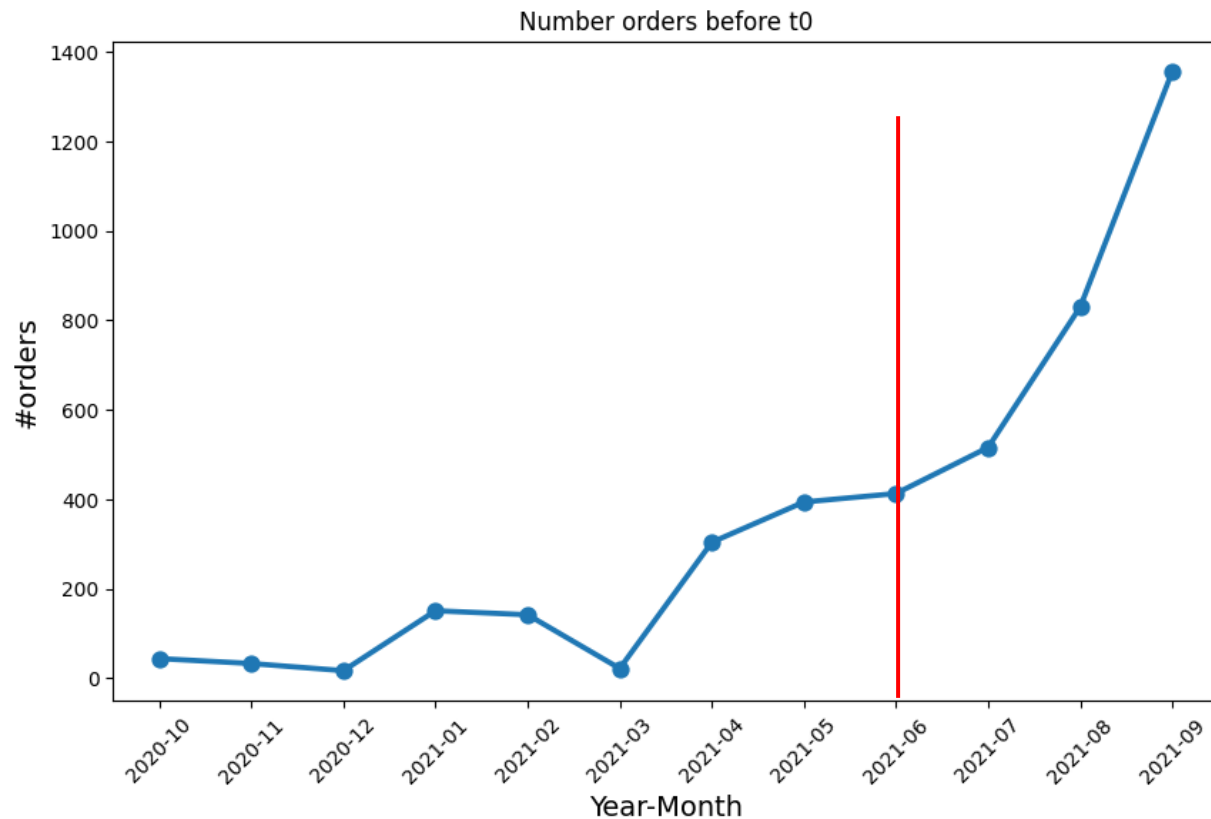
```
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
const          -1.9412      0.098     -19.734      0.000     -2.134     -1.748
→ Cust_Age       0.0065      0.002       3.655      0.000       0.003      0.010
Cust_NL         0.0348      0.061       0.575      0.565     -0.084      0.154
→ Orders         0.3965      0.012     32.088      0.000       0.372      0.421
→ Orders_Voucher 0.1193      0.024       4.939      0.000       0.072      0.167
→ is_female      0.1083      0.051       2.124      0.034       0.008      0.208
=====
```

What are the factors which explain if the customer gets activated by the mailing?

1. **Orders:** Significant positive effect (for each additional order in the past 12 months, the odds of getting activated increase by approximately 48.65%)
2. **Redeeming order Voucher:** Moderate positive effect (customers who used vouchers in their past orders are more likely to get activated by about 12.66% for each voucher-using order)
3. **Being female:** Moderate positive effect (female customers are about 11.44% more likely to get activated compared to male customers).
4. **Customer's age:** Small positive effect (for each year older a customer is, the likelihood they get activated by the mailing increases by about 0.65%.)

Other factors derived from the information in the sample?

1. Day lag from the last order to the first October 2021



Other factors derived from the information in the sample?

1. Day lag from the last order to the first October 2021
2. Customer made an order after June 2021 (binary variable)

Other factors derived from the information in the sample?

Logit Regression Results

```
=====
Dep. Variable:          Activated    No. Observations:          10000
Model:                  Logit        Df Residuals:              9993
Method:                 MLE          Df Model:                  6
Date:                   Sun, 02 Jun 2024    Pseudo R-squ.:           0.1677
Time:                   13:02:44           Log-Likelihood:          -5698.4
converged:              True           LL-Null:                 -6846.5
Covariance Type:        nonrobust        LLR p-value:             0.000
=====
```

```
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
const          -1.6386      0.079     -20.863      0.000     -1.793     -1.485
Cust_NL         0.0773      0.061       1.273      0.203     -0.042      0.196
→ Orders        0.3299      0.014     24.323      0.000      0.303      0.356
→ Orders_Voucher 0.1182      0.024       4.963      0.000      0.072      0.165
is_female       0.0698      0.051       1.378      0.168     -0.029      0.169
day_lag_from_last_order -0.0006      0.000     -1.689      0.091     -0.001      8.92e-05
→ Order_after_June 0.4944      0.065       7.642      0.000      0.368      0.621
=====
```

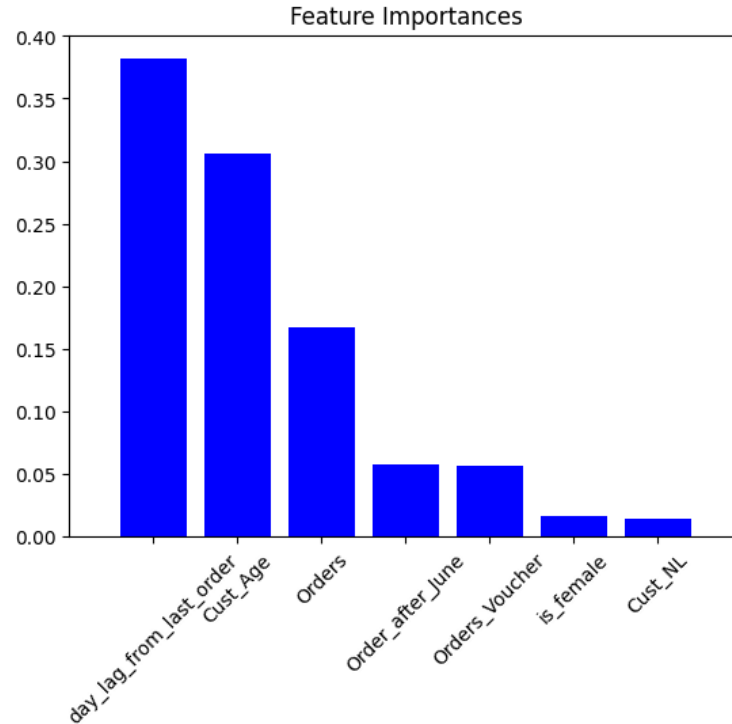
What are the factors which explain if the customer gets activated by the mailing?

1. **Making a purchase after June 2021:** Significant positive effect. Recent customers making an order after June 2021 are substantially more likely (by approximately 63.95%) to get activated.
1. **Orders:** Significant positive effect (for each additional order in the past 12 months, the likelihood of getting the customer activated increase by approximately 39.08%).
1. **Redeeming order Voucher:** Moderate positive effect (customers who used vouchers in their past orders are more likely to get activated by about 12.54% for each voucher-using order).

What are the factors which explain if the customer gets activated by the mailing?

When searching for non-linear relationship between the predictors and the outcome, a more complex model, like random forest should be used.

Based on the results of running the random forest on the sample data, the most important first third features are the following:



What are the factors beyond the sample data?

- #visits on the website (e.g. #visits per month)
- #items added in the cart
- defective experience in the customer journey:
 - on the website (e.g. poor navigation, slow loading times, broken links, etc.)
 - defects in customer fulfillment (e.g. long reimbursement time, long delivery time, poor customer service)