Artea Case Study

Data Ethics Assignment

Soumyadeep Pal
M2022ANLT028

Problem Statement

Part 1

Define strategies for targeting specific customer groups in marketing,
considering factors like behavior and preferences for Artea based on certain coupons
circulated among randomly chosen individuals out of 5000 customers.

Part 2

Identify if the model developed for trageting strategies have algorithmic bias in marketing,
which leads to unintended discrimination due to biased or incomplete data.

Questions

1. Did the coupon increase revenues? Did it increase transactions?
2. In the future, should Artea give the discount to everyone, no one, some people?
3. Which customers should be targeted in the next targeting campaign?
4. How many transactions and revenues should you expect if Artea targets those customers?
5. What are the risks of using your proposed targeting policy?
6. What do you learn from the demographic data?
7. How would you proceed with the targetting of customers now?

Did the coupon increase revenues?
Did it increase transactions?
In the future, should Artea give the discount to everyone, no one, some people?

Variables Involved : trans_after, revenue_after

Procedure of Analysis

With the given data we were able to test multiple hypothesis in order to figure out if the coupons circulated among the test group had any significant increase in the number of transactions or increase in revenue, as compared to the control group.

Hypothesis Testing

H11 : The number of **transactions** for target group is greater than the number of transactions for control group after experiment

P value :0.01367

Results significant 0.05 level of sig.

H21 : The **revenue** for target group is greater than to the revenue for control group after experiment

P value :0.3589

No significant results at 0.05 level of sig.

Further Segmentation

Further segmentation of data was done based on **Channels**

Individuals from different channels had different results when tested for the same hypothesis.

Instagram and Facebook were the only channels that showed significant results for increase in number of transactions.
P_values 0.021 and 0.007 respectively.
No significant results for increase in revenue.

Conclusion

We can infer that the significant increase in number of transactions in the target group was due to individuals from Instagram and Facebook.

We didn’t observe and increase in revenue because the revenue after accounts for transaction net of discounts.

Suggestions : Increased Frequency of coupons but reduced discount rate for each coupon targeting Instagram and Facebook channel customers.

Which customers should be targeted in the next targeting campaign?

Procedure of Analysis

Developed a Logistic Regression model with the AB_TEST data, by converting trans_after columns into categorical column did_transact which implies if an individual made a transaction after the experiment, irrespective of if the individual received a coupon or not.

Model Summary

```
Optimization terminated successfully.
         Current function value: 0.285143
         Iterations 7
                           Logit Regression Results
==============================================================================
Dep. Variable:           did_trasnact   No. Observations:                 5000
Model:                          Logit   Df Residuals:                     4990
Method:                           MLE   Df Model:                            9
Date:                Thu, 17 Aug 2023   Pseudo R-squ.:                 0.2027
Time:                        19:02:17   Log-Likelihood:                -1425.7
converged:                       True   LL-Null:                       -1788.3
Covariance Type:            nonrobust   LLR p-value:                2.703e-150
==============================================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const             -4.2735      0.191    -22.359      0.000      -4.648      -3.899
test_coupon        0.2167      0.099      2.195      0.028       0.023       0.410
num_past_purch     0.2993      0.016     18.230      0.000       0.267       0.331
weeks_since_visit -0.1842      0.022     -8.221      0.000      -0.228      -0.140
browsing_minutes   0.0324      0.007      4.471      0.000       0.018       0.047
shopping_cart      1.3950      0.103     13.602      0.000       1.194       1.596
channel_acq_2      1.1943      0.139      8.591      0.000       0.922       1.467
channel_acq_3      1.1858      0.131      9.086      0.000       0.930       1.442
channel_acq_4      1.3795      0.216      6.389      0.000       0.956       1.803
channel_acq_5      1.6095      0.273      5.895      0.000       1.074       2.145
==============================================================================
```

Which customers to target ?

According to the model the 66 customer should be targeted whose ids can be obtained from the code.

[Assumption: Probability that the individual will make a transaction >= 0.7]

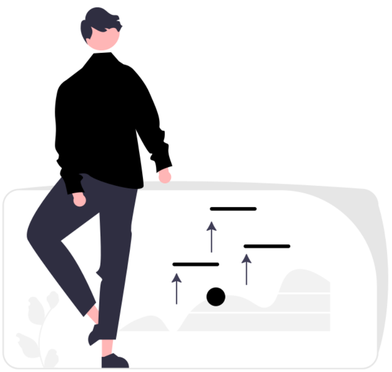
How many transactions and revenues should you expect if Artea targets those customers?

Assuming that every individual will make at least one transaction we can expect at least 66 transactions although a linear model suggests 33 transactions, with and expected revenue of 5575

[Assumptions: The revenue estimate has been made based on the previous purchase made]

What are the risks of using your proposed targeting policy?

1. Dependency on channels : The targeting policy relies heavily on the assumption that the significant increase in transactions is primarily driven by individuals from Instagram and Facebook. If customer behavior shifts, and these channels become less effective, the policy might fail to produce the expected results.
2. Discount Rate Impact: Reducing the discount rate might not always lead to increased revenue. It could potentially discourage customers from making purchases or lead them to expect even greater discounts in the future.
3. Revenue Estimation Assumptions: Estimating revenue based solely on previous purchase behavior might not accurately predict future revenue.



What do you learn from the demographic data?

Procedure of Analysis

We applied hypothesis testing to check if the model so suggested is getting fetching biased results based on minority or gender.

Contingency Table

Predictions	0	1
non_male		
0	2499	19
1	3435	47

p_value : 0.0397

Predictions	0	1
minority		
0	4691	62
1	1243	4

p_value : 0.0049

Results

The model is biased towards providing coupons to Female population as well as non_minority population.

Suggestions : We need to have additional features which captures the variance in the data better and removes the biasness towards a certain population. Also attributes like channel which serve as a proxy for gender or minority attributes can be removed.

Thank you