Artea: Designing Targeting Strategies

Youssef Lahlou, Zonghan Li, Yixuan Zhu, Ziyi Zhou

Do you believe the team managed to randomize the customers successfully?

The team managed to randomize the customers successfully. We reach this conclusion by comparing the data of the test and control groups before A/B testing. For the categorical variables (channel_acq), we compared the percentage of each channel customers used and the percentage of whether customers added a product to their shopping cart between the test and control groups. For binary and non-categorical variables (num_past_purch, spent_last_purchase, weeks_since_visit, browsing_minutes, shopping_cart). We compared the average of these variables between the test and control groups. The variables' percentages and averages are listed below:

	PERTENTAGE					AVERAGE				
	channel					num past purch	spent last purchase	weeks since visit	browsing minutes	shopping cart
	1	2	3	4	5					
Test Group	41.1%	20.6%	31.4%	4.5%	2.3%	2.0	56.7	3.2	13.7	0.3
Control Group	40.3%	21.7%	30.7%	5.0%	2.3%	2.1	58.1	3.3	13.6	0.3

Table: Differences in Test and Control Group before A/B Testing

As we can see from the table, the maximum of the differences between the test and control group is in the percentage of customers who chose Channel 4, which difference is 0.1. Given the similarities between the data on the test and control group before A/B testing, we consider that the team managed to randomize the customers successfully.

In addition, we also conducted chi-square tests on the categorical and binary variables with the null hypothesis that the difference in proportions of factor levels in each categorical variable is not significant. The resulting p-value of the tests is greater than 0.05, which implies that the null hypothesis is true, as shown in Appendix: Table 1. For the non-categorical variables,

we calculated the 95% confidence interval of the mean value for each and found that all of the confidence intervals include 0, as shown in Appendix Table 2, indicating that there is also no significant difference between the means of the non-categorical variables of the test and control groups. This is further evidence that the team managed to randomize the customers successfully.

Did the coupon increase revenues? Did it increase transactions?

The coupon has no significant increase in revenue and a significant increase in the number of transactions. We reached this result by running regressions of revenue and transaction over the variables. The result is shown in Appendix Table 3: Regression over Revenue and Table 4: Regression over Transaction. As we can see from the results, the coupon has no significant increase in revenue (p-value of test_coupon > 0.05) and a significant increase in transactions (p-value of test_coupon < 0.05). Holding other variables unchanged, coupons increase transactions by 0.026 on average.

Which customers should be targeted in the next targeting campaign?

We target customers with Facebook or Instagram channels, have 0 past purchases, spend 0 on last purchases, but have products in shopping carts.

Our ultimate goal is to increase revenue. We thus run multiple models of the different variables over revenue_after. For customer i, we separately regress revenue over the channel, num_past_purchase, spent_last _purchase, week_since_visit, browsing_minutes, and shopping_cart. Our models and the results of the regressions are listed in Appendix Table 5: Models on revenues and Individual Regression over Revenue Result.

We decide on the target customers by looking at the regression results of the interactive variables in Table 5. We first pick the features of customers we care about by ruling out those whose interactive variables have a high p-value (> 0.15), which are: Channel (Facebook and

Instagram), num_past_purchase, Spent_last_purchase, and shopping_cart. Then, we decided to target the customers described previously by looking at the sign of the coefficients of the features' interactive variables in the regression result.

How many transactions and how much revenue should you expect if Artea targets those customers in the new campaign?

If we target the customers listed in the previous section, we predict revenue of 41422.4 and transactions of 685.

We predict the result first by splitting the A/B test file into two datasets: customers who were given coupons and those who weren't. In these two datasets, we ran regressions of the channel Facebook and Instagram (with neither channel as the base case), num_past_purch, spent_last_purchase, and shopping cart over revenue and transactions, respectively. The models and results are shown in Appendix: 6. Model on Target Customers and Regression Result.

According to the criterion of target customer listed in the previous section, we split the Next_campaign datasets into two sub-datasets: the target customers and the non-target customers. We then predict the revenue and transactions of these two data sets by the regression results obtained from Appendix: 6. Model on Target Customers and Regression Result. We predict the revenue of the target and non-target customers as 7966.22 and 33456.18; the total transactions of the two groups as 163 and 522. The sum of the results in listed at the beginning of this section, and the detailed information is listed in Appendix: 7. Prediction.

What are the risks of using your proposed targeting policy?

Firstly, there may exist multicollinearity among predictor variables in the data, which may wrongly lead to some variables appearing as insignificant, leading to us excluding potentially significant predictor variables in our regression model for predicting revenue. Secondly, we chose the target customers by individually regressing revenue on each predictor

variable. This does not allow us to control for the effects of other potential variables on the revenue, which may lead to a less accurate revenue prediction. Thirdly, the threshold of the p-value (0.15) we used to determine whether or not a variable is significant is slightly higher than the commonly used value of 0.05, which may result in us including variables that are, in fact, non-significant in our regression model. Last but not least, the heterogeneous treatment effects for the number of last purchases and spending in last purchase are both negative, while the average treatment effects of those two variables are positive, which implies that coupons will prevent people who bought many products last time from buying this time. This is counterintuitive to our common sense.

Appendix

1. Table 1: Chi-square test of categorical variables between test and control

Variable: channel_acq

Observed						
	channel_acq					
test_coupon	1	2	3	4	5	Row total
0	1027	515	785	113	58	2498
1	1008	544	767	125	58	2502
Col total	2035	1059	1552	238	116	5000

Expected					
	channel_acq				
test_coupon	1	2	3	4	5
0	1016.686	529.0764	775.3792	118.9048	57.9536
1	1018.314	529.9236	776.6208	119.0952	58.0464

H0 there is significant difference
p-value 0.7757469925

Variable: shopping_cart

Observed			
	shopping_cart		
test coupon	0	1	row total
0	1749	749	2498
1	1787	715	2502
col total	3536	1464	5000

Expected		
	shopping_cart	
test coupon	0	1
0	1766.585	731.4144
1	1769.414	732.5856

H0 there is significant difference
p-value 0.2743656238

2. Table 2: 95% confidence intervals of non-categorical variables between test and control

Variable:	num_past_purch		
	mean	sample var	sample size
test	2.091526779	6.895657884	2502
control	2.019615693	6.21467314	2498

95% CI lower bound	0.2138442915
95% CI upper bound	-0.07002211945

Variable:	spent_last_purchase		
	mean	sample var	sample size
test	58.15582334	3199.349383	2502
control	56.68947958	2953.678341	2498

95% CI lower bound	4.541192139
95% CI upper bound	-1.608504623

Variable:	weeks_since_visit		
	mean	sample var	sample size
test	3.256994404	5.078669571	2502
control	3.182946357	5.123505613	2498

95% CI lower bound	0.1992564162
95% CI upper bound	-0.05116032144

Variable:	browsing_minutes		
	mean	sample var	sample size
test	13.67066347	48.87989872	2502
control	13.70736589	48.15261693	2498

95% CI lower bound	0.3494364715
95% CI upper bound	-0.4228413185

3. Table 3: Regression over Revenue

SUMMARY OUTPUT: Regression over revenue								
Regression Statistics								
Multiple R	0.41375836							
R Square	0.17119598							
Adjusted R Square	0.16953472							
Standard Error	21.5316618							
Observations	5000							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	10	477760.582	47776.0582	103.051713	0			
Residual	4989	2312962.56	463.61246					
Total	4999	2790723.14						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.88853815	1.03145295	-2.80045556	0.00512269	-4.91063936	-0.86643694	-4.91063936	-0.86643694
test_coupon	-0.28245507	0.60936471	-0.46352384	0.64300915	-1.47707778	0.91216764	-1.47707778	0.91216764
num_past_purch	3.17267295	0.13113256	24.1943957	0	2.9155955	3.42975041	2.9155955	3.42975041
spent_last_purchase	-0.01441292	0.00605199	-2.38151834	0.01727872	-0.02627747	-0.00254836	-0.02627747	-0.00254836
weeks_since_visit	-1.11275869	0.13490559	-8.24842518	0	-1.37723295	-0.84828442	-1.37723295	-0.84828442
browsing_minutes	0.16060922	0.04463954	3.5979136	0.00032389	0.07309609	0.24812235	0.07309609	0.24812235
shopping_cart	9.30864701	0.6779304	13.7309774	0	7.97960541	10.6376886	7.97960541	10.6376886
channel_fb	6.04480639	0.81655886	7.40278094	0	4.44399206	7.64562071	4.44399206	7.64562071
channel_insta	5.97656031	0.73005466	8.18645596	0	4.54533225	7.40778837	4.54533225	7.40778837
channel_referral	6.65682675	1.47717809	4.50644834	6.7429E-06	3.76090833	9.55274517	3.76090833	9.55274517
channel_other	8.69589227	2.05583109	4.22986709	2.3802E-05	4.66555959	12.7262249	4.66555959	12.7262249

4. Table 4: Regression over Transaction

SUMMARY OUTPUT: Regression over Transaction								
Regression Statistics								
Multiple R	0.421298699							
R Square	0.177492594							
Adjusted R Sq	0.175843952							
Standard Error	0.380823498							
Observations	5000							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	10	156.1354071	15.61354071	107.6598879	0			
Residual	4989	723.5373929	0.145026537					
Total	4999	879.6728						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.06490489	0.018242973	-3.55780218	0.000377447	-0.10066913	-0.02914064	-0.10066913	-0.02914064
test_coupon	0.025618587	0.010777635	2.377013662	0.017490936	0.004489683	0.04674749	0.004489683	0.04674749
num_past_pur	0.057102055	0.002319299	24.6203951	0	0.052555209	0.0616489	0.052555209	0.0616489
spent_last_pu	-0.00036724	0.00010704	-3.43086138	0.00060655	-0.00057708	-0.00015739	-0.00057708	-0.00015739
weeks_since_v	-0.02039294	0.002386031	-8.5468042	0	-0.02507061	-0.01571527	-0.02507061	-0.01571527
browsing_min	0.003086672	0.000789525	3.909529467	9.37029E-05	0.001538855	0.004634488	0.001538855	0.004634488
shopping_cart	0.17279141	0.011990334	14.41089156	0	0.149285083	0.196297736	0.149285083	0.196297736
channel_fb	0.113074199	0.014442211	7.829424366	0	0.084761116	0.141387281	0.084761116	0.141387281
channel_insta	0.108299771	0.012912239	8.387373304	0	0.082986106	0.133613436	0.082986106	0.133613436
channel_referı	0.129540313	0.026126368	4.958221168	7.34994E-07	0.078321146	0.18075948	0.078321146	0.18075948
channel_other	0.158679477	0.036360816	4.364024114	1.30271E-05	0.087396294	0.22996266	0.087396294	0.22996266

5. Models on revenues and Individual Regression over Revenue Result

$$Revenue_i = \alpha + \beta * I(coupon_i) + \gamma * I(Other_i) + \delta * I(Facebook_i) + \epsilon * I(Instagram_i) + \zeta * I(referral) + \eta * I(coupon_i) * I(Other_i) + \theta * I(coupon_i) * I(Facebook_i) + \iota * I(coupon_i) * I(Instagram_i) + \kappa * I(coupon_i) * I(referral)$$

$$Revenue_i = \alpha + \beta * I(coupon_i) + \gamma * (weeks\ since\ visit_i) + \delta * I(coupon_i) * (weeks\ since\ visit_i)$$

$$Revenue_i = \alpha + \beta * I(coupon_i) + \gamma * (browsing\ minutes_i) + \delta * I(coupon_i) * (browsing\ minutes_i)$$

$$Revenue_i = \alpha + \beta * I(coupon_i) + \gamma * (num\ past\ purchas_i) + \delta * I(coupon_i) * (num\ past\ purchas_i)$$

$$Revenue_i = \alpha + \beta * I(coupon_i) + \gamma * I(shopping\ cart_i) + \delta * I(coupon_i) * I(shopping\ cart_i)$$

		Coefficients	Standard Error	t Stat	P-value
Channel	Intercept	4.922288218	0.73068608	6.736529344	0
	test_coupon	-2.118391393	1.038204183	-2.040438121	0.04135924
	other	9.937194541	3.160324705	3.144358719	0.001674285
	facebook	4.892624403	1.264356094	3.869656995	0.000110379
	instagram	4.249495221	1.110133228	3.827914626	0.000130816
	referral	6.257269304	2.320833783	2.696129877	0.007038573
	other_treat	0.015598289	4.47049976	0.00348916	0.997216198
	face_treat	3.198772889	1.774964423	1.802161693	0.071580295
	ins_treat	3.283953455	1.578368617	2.080599817	0.037521444
	refer_treat	2.456913871	3.211973735	0.7649234	0.444353332
		Coefficients	Standard Error	t Stat	P-value
num_past_purch	Intercept	0.325887593	0.572714787	0.569022488	0.569366458
	test_coupon	1.973430271	0.806932792	2.445594343	0.014495637
	num_past_purch	3.690940098	0.178529058	20.67416998	0
	num_past_purchase_treat	-1.185901553	0.246072566	-4.819316393	1.48318E-06
		Coefficients	Standard Error	t Stat	P-value
spent_last_purchase	Intercept	4.511435419	0.678912312	6.645092951	0
	test_coupon	0.793069459	0.956186681	0.829408602	0.406912862
	spent_last_purchase	0.057660306	0.008645799	6.669170378	0
	spent_last_purchase_treat	-0.019243376	0.011985384	-1.605570245	0.108431569
		Coefficients	Standard Error	t Stat	P-value
weeks_since_visit	Intercept	11.60730315	0.810880612	14.31444158	0
	test_coupon	-0.323308248	1.156960843	-0.279446145	0.77991403
	weeks_since_visit	-1.202387531	0.207627534	-5.791079372	7.42107E-09
	weeks_since_visit_treat	0.052455609	0.29415882	0.178324106	0.858475678
		Coefficients	Standard Error	t Stat	P-value
browsing_minutes	Intercept	6.394602691	1.046784953	6.108802646	1.07985E-09
	test_coupon	0.169301573	1.473844881	0.114870687	0.908552236
	browsing_minutes	0.101081817	0.068136256	1.483524668	0.137998147
	browing_minutes_treat	-0.029778123	0.095962012	-0.310311573	0.756336975
		Coefficients	Standard Error	t Stat	P-value
shopping_cart	Intercept	5.481686678	0.556997246	9.841496909	0
	test_coupon	-0.73953111	0.783514251	-0.943864275	0.345284652
	shopping_cart	7.665696499	1.017204696	7.536041204	0
	shopping_cart_treat	2.120159121	1.448193123	1.46400303	0.143256035

6. Model on Target Customers and Regression Results

$$Revenue_{ic} = \alpha + \beta * I(shopping\ cart_{ic}) + \gamma * I(num\ past\ purchase_{ic}) + \delta * I(spent\ last\ purchase_{ic}) + \epsilon * I(Channel - Facebook_{ic}) + \zeta * I(Channel - Instagram_{ic})$$

$$Transaction_{ic} = \alpha + \beta * I(shopping\ cart_{ic}) + \gamma * I(num\ past\ purchase_{ic}) + \delta * I(spent\ last\ purchase_{ic}) + \epsilon * I(Channel - Facebook_{ic}) + \zeta * I(Channel - Instagram_{ic})$$

$$Revenue_{inc} = \alpha + \beta * I(shopping\ cart_{inc}) + \gamma * I(num\ past\ purchase_{inc}) + \delta * I(spent\ last\ purchase_{inc}) + \epsilon * I(Channel - Facebook_{inc}) + \zeta * I(Channel - Instagram_{inc})$$

$$Transaction_{inc} = \alpha + \beta * I(shopping\ cart_{inc}) + \gamma * I(num\ past\ purchase_{inc}) + \delta * I(spent\ last\ purchase_{inc}) + \epsilon * I(Channel - Facebook_{inc}) + \zeta * I(Channel - Instagram_{inc})$$

revenue with coupon								
	Coefficients	*Standard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-3.6310811	0.80627202	-4.5035434	6.99E-06	-5.2121119	-2.0500503	-5.2121119	-2.0500503
shopping_cart	10.285643	0.92221471	11.1531977	3.17E-28	8.47725842	12.0940275	8.47725842	12.0940275
facebook	5.84188374	1.07630077	5.42774279	6.26E-08	3.73134956	7.95241792	3.73134956	7.95241792
instagram	6.66096587	0.96307488	6.91635306	5.87E-12	4.77245803	8.54947372	4.77245803	8.54947372
num_past_purch	2.61891627	0.17286766	15.1498336	1.15E-49	2.27993751	2.95789504	2.27993751	2.95789504
spent_last_purchase	-0.0096164	0.00802044	-1.1989886	0.23064618	-0.0253438	0.00611098	-0.0253438	0.00611098
transaction with coupon								
	Coefficients	*Standard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.0666681	0.01589457	-4.1943962	2.83E-05	-0.097836	-0.0355002	-0.097836	-0.0355002
shopping_cart	0.20637731	0.01818022	11.3517495	3.72E-29	0.17072745	0.24202718	0.17072745	0.24202718
facebook	0.11683974	0.02121782	5.50667982	4.03E-08	0.0752334	0.15844608	0.0752334	0.15844608
instagram	0.12882008	0.01898572	6.78510227	1.44E-11	0.09159069	0.16604947	0.09159069	0.16604947
num_past_purch	0.05377134	0.00340785	15.7786546	1.55E-53	0.04708883	0.06045385	0.04708883	0.06045385
spent_last_purchase	-0.0003059	0.00015811	-1.9345791	0.05315484	-0.0006159	4.16E-06	-0.0006159	4.16E-06
revenue without coupon								
	Coefficients	*Standard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-3.13635142	0.88252184	-3.55385133	0.00038669	-4.86690297	-1.40579987	-4.86690297	-1.40579987
num_past_purch	3.8839756	0.20211278	19.2168727	7.42E-77	3.48764933	4.28030187	3.48764933	4.28030187
spent_last_purchase	-0.02218684	0.00927558	-2.39196393	0.01683219	-0.04037547	-0.00399821	-0.04037547	-0.00399821
shopping_cart	7.55463341	0.98963313	7.63377173	3.23E-14	5.61404559	9.49522124	5.61404559	9.49522124
channel_fb	4.26663387	1.19226744	3.57858794	0.00035205	1.92869711	6.60457063	1.92869711	6.60457063
channel_insta	3.7719212	1.04029701	3.62581183	0.00029383	1.73198574	5.81185666	1.73198574	5.81185666
transaction without coupon								
	Coefficients	*Standard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.04468319	0.01398827	-3.19433334	0.00141904	-0.07211301	-0.01725337	-0.07211301	-0.01725337
num_past_purch	0.06245769	0.00320356	19.4963667	6.64E-79	0.05617579	0.06873959	0.05617579	0.06873959
spent_last_purchase	-0.00046081	0.00014702	-3.13434418	0.00174243	-0.00074911	-0.00017252	-0.00074911	-0.00017252
shopping_cart	0.12593635	0.01568602	8.02857455	1.50E-15	0.09517739	0.15669532	0.09517739	0.15669532
channel_fb	0.06857313	0.01889784	3.62862284	0.00029067	0.03151605	0.10563021	0.03151605	0.10563021
channel_insta	0.05876961	0.01648906	3.5641586	0.00037188	0.02643595	0.09110327	0.02643595	0.09110327

7. Prediction

Target Group	
average rev	5.836058608
average txn	0.1196769457
sum rev	7966.22
sum txn	163.39

Non-Target Group	
average rev	7.873896001
average txn	0.1228232683
sum rev	33456.18411
sum txn	521.8760669