University of California, Davis

Introducing an Online Community at KyngaCell

BAX 401: Homework 3

Scarlett Ma, Ethan Meng, Julia Wallace, Xi Wang, and Ran Zhang
Information, Insight, and Impact, Group C-SAT
Professor Ashwin Aravindakshan
November 3, 2021

Table of Contents

Executive Summary

In this report, we explored our firm's experiment of launching an online discussion community for game users to investigate the impact of this new feature. We compared the result of this experiment against three key metrics, user revenue, retention and customer lifetime value (CLV). We used difference-in-difference regression, logistic regression and linear regression to measure the significance and the magnitude of the effect of launching this community to these metrics. We then concluded that this experiment has successfully increased loyal users' engagement and revenue, but without a practical impact on reducing churn rate. Finally, we made recommendations based on our findings to encourage every customer to their full potential lifetime value.

Introduction

The domain of our firm is KyngaCell, a mobile gaming powerhouse, who introduced an online community in mid-2021 that has been introduced to the game *Nicht-Soporific*. The firm had believed that this new feature would improve user revenue as well as the user retention. However, the firm would like to back up their beliefs with numbers. Hence, our goal in this project is to determine whether the online community benefits the firm's user from revenue, retention and Customer Lifetime Value (CLV) perspectives. To do so, we provided insights and possible limitations from the analysis of the current variables, and explored a new variable to discuss whether it may be of benefit to the firm.

Problem Formulation

To analyze the effect of the online community on various performance measures of the firm, we developed several models from different perspectives. We used the Difference in Difference method to calculate the user revenue one month before and one month after the online community was introduced and provided a short-term metric measure to the firm. Then we determined the long term effects by building logistic regressions and using odd ratios to calculate and analyze the user retention rate and Customer Lifetime Value (CLV). Lastly, we built a logistic regression model to explore the relationship between the new variable and the churn rate with provided data.

Data Description

With the information of how much each user spent in the game one month before and one month after the online community was introduced, and also whether they joined the community or not in Data 1, we were able to calculate if the user revenue increased due to the online community. Similarly, churn rate and customer lifetime value can be calculated through Data 2, together with the customer age with firm and average spend variable. Data 3 provides additional information with a new variable on how users join the game (either by campaign or naturally join), which we further examined.

Model Development

I. User Revenue

To determine whether joining the online community increased the revenue per user, we used a difference-in-difference experiment design to find the effect of joining the community. The experiment group is users who joined the online community and those who did not join are the control group. We assume a parallel trend between these two groups. The model is as follows:

$$Revenue = \beta_0 + \beta_1 D_{Treatment} + \beta_2 D_{Joined} + \beta_3 D_{Treatment} * D_{Joined} + \epsilon$$

II. User Retention

Figure 1 shows the counts for user joins and user churn. Join 0 is for the customers who didn't join the community, and join 1 is for the customers who joined the community. Similarly notions apply to churn, where 0 represents not-churn and 1 for churn. Also, we conducted a logistic regression on the joined variable and checked the significance of the joined response variable and other variables in Data 2:

$$log\left(\frac{Churn}{1 - Churn}\right) = \beta_0 + \beta_1 CustomerAge + \beta_2 Joined + \beta_3 AvgSpend + \epsilon$$

III. Customer Lifetime Value (CLV)

Another point is to address whether joining the community has an impact on customer lifetime value (CLV). The CLV for joined and not-joined community customers are calculated separately for comparison. The customers are then classified into eight groups based on age with the company. The total customer lifetime value is calculated for each group and summed up at the end. Hence, the customer lifetime calculation for joined community customers is: $CLV_{joined} = \sum_{i=1}^{8} \frac{\sum_{k=1}^{n} margin_{i,k}}{1-r_i}$, where i represents group 1 to 8, r_i is the retention rate for group i, and $\sum_{k=1}^{n} margin_{i,k}$ represents the sum of margins of n customers in group i. $CLV_{not-joined}$ is calculated similarly.

IV. Campaign/Organic Variable Exploration

We made further analysis and explored whether the new variable would provide more value to the team by running a logistics regression model with the Campaign/Organic variable:

$$log\left(\frac{Churn}{1 - Churn}\right) = \beta_0 + \beta_1 Campaign/Organic + \beta_2 Customer\ Age\ with\ Firm + \beta_3 Joined + \beta_4 Average\ Spend + \epsilon$$

Results

I. User Revenue

From the user revenue model in Figure 6, we found out that the average treatment effect is about \$29.02. The p-value of the coefficient of interaction term is very low, so we can infer that joining the online community has significantly increased user revenue by about \$29.02 per customer.

II. User Retention

Figure 2 portrays that the retention rate of the customers who did not join the community is 48.71%, while the retention of joined users is 29.26%. These percentage numbers show us that customers who joined the community are not likely to retain. We can then conclude that the community has not been able to increase the retention rate. Figure 7 is a summary of the logistic regression model, by fitting churn, customer age, and average spend with logistic regression, we found that the predictor joined has a p-value of 0.00979. At the 5% significance level we infer that the coefficient of a joined variable is not equal to zero. The p-value of customer age and average spend within the game is larger than 0.05, so they do not significantly influence the odds ratio. We noticed from Figure 8 that on average the customer who joined the community has exp(0.918) = 2.503 higher odds ratio to churn than the customers who did not join. We can conclude that the new community is increasing the churn rate, which means it is not helpful for increasing retention rates.

III. Customer Lifetime Value (CLV)

From Figure 9 we can see that the overall CLV model is valid, since the p-value is 4.134e-08. The coefficient of joining the community is significant at the 0.1 level, suggesting that joining the community will bring 21.3737 units to customer lifetime value. Customer age, however, is not significant and has a limited effect on customer lifetime value. Figures 3 and 4 show that joining the community increases the CLV to 5813.795, compared to the CLV of 88883.368 for the not-joined group. Figure 5 compares the CLV of different cohorts. The CLV improvements mainly exist in groups 2 and 4, where many customers fall into group 2. The CLV improvements for groups with less people are ignorable and even negative. Exploring whether joining the community has a relationship with CLV is ideal, so we applied regression analysis: $Churn = 0.6101704 - 0.0129183\, CustomerAge + 0.2131131\, Joined - 0.006436\, AvgSpend + \epsilon$

IV. Campaign/Organic Variable Analysis

From Figure 10, we noticed that only "joined" is statistically significant among all of the variables used for logistic regression. By analyzing the different ways users joined the mobile game *Nicht-Soporific*, we can see in which way players are more loyal and have higher retention rate. From the odds ratio in Figure 11 from the logistic regression analysis, we may infer that customers who joined the game from a campaign churn 1.4272 times more than users who joined organically. The overall model shown in Figure 12 accuracy is 0.6281, which is better than random (0.5).

Recommendations and Managerial Implications

I. Online Community

The launch of the online community is not helpful in increasing user retention rate. But the user online community increased user revenue by approximately $29.02 per customer and joining the community has a positive impact on customer lifetime value by our models. Although customer churn rates are high, the retained customer has a high expenditure on the game. Thus we recommend that the company focuses on increasing the expenditure of retained users, since the community does not impact overall retention rates.

II. New Model with Campaign/Organic Variable

For our new model, the hypothesis is that the users who joined the game through the campaign are joining on a whim; once the rewards are gone or the novelty has worn off, the number of users who joined the game through the campaign would sharply reduce. It is supported by our model in figure 10: the larger the value of the campaign variable, the larger the churn rate will be. Even though the coefficient of Campaign/Organic variable was not statistically significant in the model, Hyndman states, "It is possible to have an insignificant coefficient associated with a variable that is useful for forecasting" (Hyndman, 2011). Moreover, we have sufficient evidence to keep this variable in the model.

Conclusion

While it is helpful to create these models to quantify the effect of introducing the online community to KyngaCell users, there are limitations. We developed several multiple linear regression models that do not contain all variables. In addition, for user revenue, creating a time series model with historical data for joining and not-joining users can be a more conclusive accurate measure. With the churn model, there could be several lurking variables that are affecting the churn rate, including user gaming time, customer satisfaction, and so on. For the CLV model, customers are classified by their age with the company, however this may not be the optimal method. Also, sample sizes within groups are small, thus more customers should be studied, or regrouping should be conducted.

The Campaign/Organic variable is not statistically significant in the model developed using the current data; however if we obtained updated data or historical data from the firm, the variable may become significant. On the other hand, users who are naturally interested in games of similar type have a higher retention rate. Furthermore, the company may want to change its promotion strategy by advertising on platforms with related themes, thus attracting users who are interested in this genre to join the game.

References

Hyndman, Rob J. "Statistical Tests for Variable Selection." Rob J Hyndman, 14 Mar. 2011, https://robjhyndman.com/hyndsight/tests2/.

Appendix

	churn 0	churn 1	Total
Join 0	57	60	117
Join 1	24	58	82
Grand Total	81	118	199

Figure 1: Number of Individual Retained /Churned

	# of customer not churn at three month after lauch community (retention rate)	# of customer churn at three month after lauch community(churn rate)	Total
Join 0 (control)	0.487179487	0.512820513	1
Join 1(treatment)	0.292682927	0.707317073	1
Grand Total	0.407035176	0.592964824	1

Figure 2: Retention Rate and Churn Rate

Group number	Churn rate	margin	Not joined CLV
1	44.44%	325.0000	731
2	61.54%	840.6667	1366
3	56.25%	629.0000	1118
4	75.00%	434.3333	579
5	36.84%	544.0000	1477
6	47.06%	590.0000	1254
7	30.00%	366.6667	1222
8	50.00%	219.0000	438
		Not-joined CLV total	**8185**

Figure 3: CLV table for not-joined customers

Group number	Churn rate	margin	Joined CLV
1	66.67%	133.6667	200.5000
2	68.00%	1186.0000	1744.1176
3	80.00%	477.3333	596.6667
4	62.50%	412.3333	659.7333
5	70.59%	919.6667	1302.8611
6	66.67%	260.3333	390.5000
7	57.14%	350.3333	613.0833
8	100.00%	306.3333	306.3333
		Joined CLV total	**5813.795425**

Figure 4: CLV table for joined customers

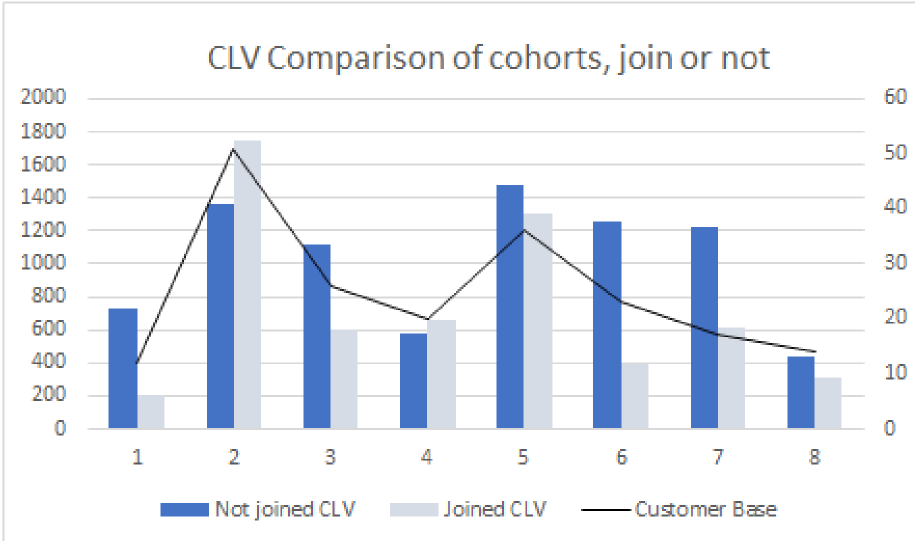

Figure 5: CLV Comparison of Different Cohorts

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.583745137
R Square	0.340758386
Adjusted R Square	0.335738779
Standard Error	38.58521917
Observations	398

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	303207.6137	101069.2046	67.88548183	2.09234E-35
Residual	394	586594.7406	1488.819138		
Total	397	889802.3543			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	70.37606838	3.567204774	19.72863147	9.38493E-61	63.36293237	77.38920438	63.36293237	77.38920438
Post	30.87179487	5.044789372	6.119540896	2.26723E-09	20.95372282	40.78986693	20.95372282	40.78986693
Treatment	17.75807797	5.557092842	3.195569782	0.001507997	6.832815647	28.68334028	6.832815647	28.68334028
Post*Treat	29.01844903	7.858916065	3.692423839	0.000253448	13.56779489	44.46910317	13.56779489	44.46910317

Figure 6:Result of Difference-in-Difference Regression of Community on Customer Value

```
Call:
glm(formula = commu$`Churned at 3 months` ~ commu$`Joined?` +
    commu$`Customer Age with Firm at time of launching the online community` +
    commu$`Average Spend Last 3 months of Life with the firm`,
    family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6641  -1.2094   0.8045   1.1049   1.2815

Coefficients:
                                                                         Estimate Std. Error z value Pr(>|z|)
(Intercept)                                                              0.462435   0.535488   0.864  0.38782
commu$`Joined?`                                                          0.917627   0.355216   2.583  0.00979 **
commu$`Customer Age with Firm at time of launching the online community` -0.051796   0.073144  -0.708  0.47886
commu$`Average Spend Last 3 months of Life with the firm`               -0.002899   0.005657  -0.512  0.60836
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 268.95  on 198  degrees of freedom
Residual deviance: 260.54  on 195  degrees of freedom
AIC: 268.54

Number of Fisher Scoring iterations: 4
```

Figure 7: Summary of Logistic Model

```
Waiting for profiling to be done...
                                                                          coefficient       2.5 %      97.5 %
(Intercept)                                                               0.462435100 -0.58342556 1.524159597
commu$`Joined?`                                                           0.917626522  0.23199636 1.629262848
commu$`Customer Age with Firm at time of launching the online community` -0.051796003 -0.19585547 0.091807565
commu$`Average Spend Last 3 months of Life with the firm`               -0.002898694 -0.01406909 0.008176207
Waiting for profiling to be done...
                                                                         Odds Ratio     2.5 %   97.5 %
(Intercept)                                                               1.5879361 0.5579837 4.591283
commu$`Joined?`                                                           2.5033417 1.2611151 5.100114
commu$`Customer Age with Firm at time of launching the online community`  0.9495225 0.8221310 1.096154
commu$`Average Spend Last 3 months of Life with the firm`                 0.9971055 0.9860294 1.008210
```

Figure 8: Confidence Interval of Odds Ratio Change

	(1)
	CLV model
Joined	21.3737***
	(7.37e-09)
Customer Age	-0.4765
	(0.582)
Intercept	58.4338***
	(<2e-16)
R^2	0.1593
Adjusted-R^2	0.1507
n	117

Figure 9: Regression Results for CLV Model

Significance levels: * = significant at 10%; ** = significant at 5%; *** = significant at 1%.

```
Call:
glm(formula = Churn ~ campaign_organic + cust_age + joined +
    avg_spend, family = binomial(link = "logit"), data = mydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6987  -1.1934   0.7953   1.0871   1.3342

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       0.355687   0.565412   0.629  0.52930
campaign_organic  0.179729   0.304496   0.590  0.55502
cust_age         -0.052077   0.073166  -0.712  0.47661
joined            0.930000   0.356426   2.609  0.00907 **
avg_spend        -0.003006   0.005669  -0.530  0.59588
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 268.95  on 198  degrees of freedom
Residual deviance: 260.19  on 194  degrees of freedom
AIC: 270.19

Number of Fisher Scoring iterations: 4
```

Figure 10: R output for Logistics Regression Model with Campaign/Organic Variable

```
oddsr=round(exp(cbind(OddsRatio=coef(mylogit),confint(mylogit))),4)
oddsr
```

```
Waiting for profiling to be done...
                 OddsRatio  2.5 % 97.5 %
(Intercept)         1.4272 0.4717 4.3680
campaign_organic    1.1969 0.6578 2.1767
cust_age            0.9493 0.8218 1.0958
joined              2.5345 1.2742 5.1779
avg_spend           0.9970 0.9859 1.0081
```

Figure 11: Odds Ratio Calculation and Confidence Intervals

```
               Accuracy : 0.6281
                 95% CI : (0.557, 0.6954)
    No Information Rate : 0.593
    P-Value [Acc > NIR] : 0.1743

                  Kappa : 0.1579

 Mcnemar's Test P-Value : 5.773e-07

            Sensitivity : 0.8729
            Specificity : 0.2716
         Pos Pred Value : 0.6358
         Neg Pred Value : 0.5946
             Prevalence : 0.5930
         Detection Rate : 0.5176
   Detection Prevalence : 0.8141
      Balanced Accuracy : 0.5722

       'Positive' Class : 1
```

Figure 12: Logistic Regression Model Statistics