Introducing an Online Community at KyngaCell

Table of Content

Executive Summary

A mobile gaming firm KyngaCell requests to determine the effectiveness of their new online gaming community feature for the latest game Nicht-Soporific. This new feature aims to increase user revenue and retention. In this report, we discuss the impacts of the online gaming community on the firm's revenue and retention, the evaluation of their Customer Lifetime Value (CLV), the magnitude of probable change within the explanatory variables, and took a step further to explain the campaign effectiveness. Based on our findings, the online community successfully increased revenue in the short term but failed to boost retention rate and CLV in the long run. The same ineffectiveness is observed in the campaign. Thus, KyngaCell should provide more incentives to improve user engagement.

Introduction

In mid-2021, KyngaCell launched a new online gaming community. This enabled users to connect and interact during and outside of gameplay. It is now 2023 and the Chief Analytics Officer (CAO) requested to evaluate this new service. Our three objectives to assess the effectiveness are 1) determine if online gaming increased user revenue, 2) determine if the online community led to an increase in user retention, and 3) determine if the online community led to an increase in CLV.

Problem Formulation

KyngaCell has collected data on the treatment period, the amount spent on the game per month, the average spend of last 3 months of life with the firm, customer age at the time of launching new service, campaign/organic attraction, and whether or not the user churned and joined the online community. For this framework, we performed regression analysis to predict and estimate the effect of a dependent variable based on various independent variables. Specifically, a quasi-experimental design difference-in-difference was employed to estimate the short-term effect of the new feature. Logistic regression was used to predict long-term churn/retention (Cole, 2020).

Data Description

In the first dataset, the dependent variable is the amount of spend, showing how much each of the individuals spent in the game one month before(*Month Before*) and one month after(*Month After*) the introduction of the online community. *Customer ID* is the unique number given to the individuals who play the game. *Joined*? shows who participated in the community or not. We define people who joined the community as the treatment group and recorded them by 1, and the opposite as the control group, recorded by 0. The second dataset consisted of *Customer ID, Joined*?, churn, customer age at the time of joining the online community, and average spend of each user within the last 90 days with the firm. This data was useful in discovering user retention. The third dataset combines the first two datasets and adds an additional categorical variable *Campaign/Organic* based on if the user joined organically (0) or was attracted by a campaign (1). All three datasets were used to address the case objectives and no variables were excluded from the model development except *Customer ID.*

Model Development, Estimation, and Results

Objective 1: Determine if online gaming increased user revenue.

To address the first objective, the difference-in-difference(DID) technique was used. DID is a commonly used technique to appropriately obtain a counterfactual and estimate the causal effect by leveraging the longitudinal data of both control and treatment groups. We used the longitudinal spending data before and after the launch of the online community, classified by whether they joined the online community. This dataset is stored in the datasheet Data1. We constructed a regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon_0$, where Y is the monthly spend, X_1 is the dummy variable indicating the participation of the online community, X_2 is another dummy variable classifying the spending into the month before and after the introduction of the online community. X_3 is the interaction term of the first two variables, used to signify the after-treatment spending of players who joined the community. The results are shown in Table 1 below.

Table 1 Results of Difference-in-Difference

	Estimate	**Std. Error**	**t-value**	**Pr(>|t|)**
X_3 (DID)	29.018	7.859	3.692	0.000253 ***

Note: Data1, significance level = 95%

Based on the output, the online community increased the monthly spending by $29.

Objective 2: Determine if the online community led to increased user retention.

To address this objective, we applied logistic regression to measure the impact on churn, a binary dependent variable. We constructed the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon_0$, with *churn* as the dependent variable, *Joined*, *Age_With_Firm*, and *Avg_Spend*, as the independent variables. According to the model results shown in Table 2 below, we can conclude that at the same level of age with the firm and last-three months of spending, joining the community will increase the odds of churn by 2.5%. In other words, the introduction of the online community reduces the retention rate.

Table 2 Results of the original model

	Estimate	**Exp. Estimate**	**Std. Error**	**Z value**	**Pr(>|Z|)**
Intercept	0.462435	1.5879361	0.535488	0.864	0.38782
Joined	0.917627	2.5033417	0.355216	2.583	0.00979 **
Age_With_Firm	- 0.051796	0.9495225	0.073144	- 0.708	0.47886
Avg_Spend	- 0.002899	0.9971055	0.005657	- 0.512	0.60836

Note: Data 2, significance level = 95%, AIC = 268.54

Objective 3: Determine if the online community led to an increase in CLV.

To address this objective, we first calculated the retention rate based on the Churn data, then derived the CLV for both control and treatment groups, and at last applied the t-test to draw inferences about the magnitude of the two CLVs. The general CLV function is $\sum_{t=1}^{T} \frac{mr^{t-1}}{(1+i)^{t-1}}$. We set T to infinity to simplify the function to $m\frac{1+i}{1+i-r}$, where r is the retention rate, i is the discount rate, m stands for margin.

We assumed a margin of 50% of customers' spend in the last 3 months with the firm and a discount rate at 10%. From the result of the t-test shown in Table 4, we can conclude that people who did not join the online community have a higher CLV than people who did.

Additional Exploration

With the additional variable, Campaign/Organic, we explored the effect of initial attraction methods on retention. The difference in retention rate (if any) can in turn signal the effectiveness of the campaign held to attract new users. To investigate, we built a logistic model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon_0$, with *Joined*, *Age_With_Firm*, *Avg_Spend*, and *Campaign/Organic* as the regressors, and *Churn* as the regressand. The results are summarized in Table 3 below. First, we found the "*Joined*" coefficient slightly increased from 0.9176 in the original model to 0.9300 in this model, while the exponentiated coefficient increased from 2.5033 to 2.5345. We can conclude that the online community is more ineffective in increasing retention when considering Campaign/Organic. Next, we looked at the Campaign/Organic variable - with a coefficient value of 0.1797 and exponentiated coefficient value of 1.1969, it is implied that for people who joined the game through campaigns, the probability of Churn increases by 1.2%. Alternatively, we conducted a proportion test on retention % between the control group (organic attraction) and treatment group (campaign.) With a p-value of 0.66, we do not have enough evidence to conclude that there is a difference between the retention rate of Campaign and Organic users. Last, we evaluated the overall model and achieved an accuracy rate of 0.5779. We found that the campaign may have attracted users initially, but it did not contribute significantly to the retention rate in the long term.

In addition, we quantified the effect of Campaign/Organic on CLV, and determined that there were no significant differences between the control and treatment groups' CLV (refer to Table 5).

Table 3 Results of adding a new variable 'Campaign/Organic'

	Estimate	**Exp. Estimate**	**Std. Error**	**Z value**	**Pr(>|Z|)**

Interception	0.355687	1.4271603	0.565412	0.629	0.52930
Joined	0.930000	2.5345081	0.356426	2.609	0.00907 **
Age_With_Firm	- 0.052077	0.9492559	0.073166	- 0.712	0.47661
Avg_Spend	- 0.003006	0.9969981	0.005669	- 0.530	0.59588
Campaign/Organic	0.179729	1.1968931	0.304496	0.590	0.55502

Note: Data 3, significance level = 95%, AIC = 270.19

Recommendation and Managerial Implications

We recommend that KyngaCell act on these next insights to improve user revenue and retention. First, we strongly recommend KyngaCell to increase engagement and participation in the game. Although some may think that the online community resulted in increased user revenue and was effective, there is still a large room to boost long-term retention rate and CLV. KyngaCell can launch loyalty programs to reward long-term users and daily sign-in, while new features can also be introduced to enable players to challenge their friends in real life, increasing overall engagement.

Regarding campaigns, KyngaCell should focus on improving organic engagement since campaigns are not necessarily successful in increasing acquisition and long-term retention. It may seem more challenging to observe noticeable growth from organic engagement; however, in the long run, engaging users organically will establish a compounding effect. Furthermore, a potential explanation for the campaign ineffectiveness is that the new joiner incentives provided by the campaign appealed to the group of users who were interested in the game but needed nudging. It is harder to retain these customers as their overall interest in the game was generally lower than those who joined organically. This information is critical to the management, as they can adjust the campaign accordingly to achieve the desired effectiveness, develop loyalty among users, and encourage spending to increase overall revenue. They can act on these insights to plan marketing activities, promotions, and outreach to increase and improve their user retention and revenue.

Conclusion

Based on the analysis, the introduction of the online community successfully increased revenue in the short term but failed to boost retention rate and CLV in the long run. The online community may affect beyond the existing customers and attract some new players in the beginning, yet unsatisfying interactions fail to retain them and make the existing customers lose interest at the same time. The campaign shows the same ineffectiveness in retaining customers. More robust measures should be taken to motivate user engagement. The company should rely on the analysis to make sure the team is utilizing resources effectively.

References

Cole, A. (2020, May 13). *Predicting Customer Churn Using Logistic Regression.* towards data science. https://towardsdatascience.com/predicting-customer-churn-using-logistic-regression-c6076f37eaca

Appendix

To confirm the CLV dividing the customers into Joined and Not Joined, we established Not Joined as our control group and Joined as our treatment group.

Table 4 CLV for dividing customers into Joined and Not Joined

Var.test for comparison of variances between Joined and Not Joined	
H0: Variances are equal	H1: Variances are not equal
Result: Since the p-value(=0.0001183) < alpha(=0.05), we reject the null. Variances of the control group and treatment group are not equal.	

One tailed t-test	**Two tailed t-test**
H0: CLV_Control group >= CLV_Treatment group H1: CLV_Control group < CLV_Treatment group	H0: CLV_Control group = CLV_Treatment group H1: CLV_Control group =/= CLV_Treatment group
Result: Since the p-value(=1) > alpha(=0.05), we fail to reject the null. The CLV of the control group is larger than or equal to the CLV of the treatment group.	Result: Since the p-value(=4.312e-12) < alpha(=0.05), we reject the null. The CLV of the control group is not equal to the CLV of the Treatment group.
Conclusion: The control group has a larger CLV compared to the treatment group. The variable 'Joined/Not Joined' is not effective.	

To confirm the CLV dividing the customers into Campaign and Organic, we established Organic as our control group and Campaign as our treatment group.

Table 5 CLV for dividing customers into Campaign and Organic

Var.test for comparison of variances between Campaign and Organic	
H0: Variances are equal	H1: Variances are not equal
Result: Since the p-value(=0.5205) > alpha(=0.05), we fail to reject the null. Variances of control group and treatment group are equal.	

One tailed t-test	**Two tailed t-test**

H0: CLV_Control group >= CLV_Treatment group H1: CLV_Control group < CLV_Treatment group	H0: CLV_Control group = CLV_Treatment group H1: CLV_Control group =/= CLV_Treatment group
Result: Since the p-value(=0.8858) > alpha(=0.05), we fail to reject the null. The CLV of the control group is larger than or equal to the CLV of the treatment group.	Result: Since the p-value(=0.2284) > alpha(=0.05), we fail to reject the null. The CLV of the control group is equal to the CLV of the Treatment group.
Conclusion: The variable 'Campaign/Organic' is not effective.	