Analyzing the Impact of Introducing an Online Community at KyngaCell

Student Names:

Anisha Samant

Sheetal Rajgure

Issam (Sam) Tamer

Siwei (David) Ran

School: UC Davis Graduate School of Management

Course Number and Name: BAX 401 Information, Insight and Impact

Instructor: Ashwin Aravindakshan

Date: 2022-11-10

Table of Contents

Executive Summary: This paper examines the short-term and long-term impact of adding an online community feature to *Nicht-Soporific*. We do this by modeling Customer Lifetime Value (CLV), retention rate, and revenue changes to effectively predict and compare groups of customers who did and did not join the online community. Our results indicate a short term positive revenue response but a decline in retention rates in the longer term for those that joined. CLV was at least $145/month higher for customers who joined compared to those who did not. Further analysis on the different customer segments, show a better retention and performance by organic customers compared to campaign acquired customers.

Introduction: KyngaCell, a mobile gaming firm, introduced an online community feature for one of their offered games- *Nicht-Soporific*. As with any new product, it is essential for KyngaCell to measure what value this feature brings to the firm to evaluate its effectiveness. Understanding the kind of impact a new product or feature has on customer behavior and user revenue will not only help the managers at KyngaCell determine its financial value, but also shape further marketing strategies to optimize their business. Our main metric for this analysis would be interpreting the CLV to estimate the change in financial contributions from KyngaCell's customers that attribute to this feature.

Problem Formulation: To understand the direct and short-term impact, we need to look at the difference in the change in user revenues for online community members and that for non-members. To determine the long-term impact, we need to examine the differences in user retention rates and CLVs between them. Quantifying these differences and finding whether they are statistically significant or not is a major part of the problem statement. We also aim to provide actionable insights relating to how the medium through which customers join the game, organically or through campaign, affect the performance metrics.

Data Characteristics: The Game Team has provided data for 199 customers, segmenting them between those who were invited to participate and those who were not. The dataset also contains customer spending one month before and one month after the feature was introduced, along with customer characteristics, churn information, and how the customer was acquired. An advantage of having spends data one month after the feature was introduced is to be able to assess the instant effect the feature has on revenue with fewer possibilities of any external variables.

Assumptions: CLV is calculated at a quarterly level, time period t in quarters. Discount rate is assumed 4% per annum i.e. 1% per quarter (derived from bond yield rate). Acquisition cost is assumed to be $0.

Model Development, Estimates, and Results:

1. To estimate the difference between customer spending before and after the online community network was introduced we carried out a linear regression model. This was done keeping in mind our "joined" qualitative variable corresponding to customers who were invited to join the online community vs customers who were not. The treatment in our model was the introduction of online community, independent variable was 'joined = Yes/No' and we regressed our dependent variable, the difference in spending, on it. From the summary, we see that customers who did not join the online community spent on average $30.8 more the month after the game was introduced compared to the month before. This difference in spending was much more for Customers who joined the community, who on average spent $59.82 ($30.8+$29.02) more the latter month. The p-value of diff-in-diff coefficient is very low, corroborating that the difference is statistically significant. (*1*) We can conclude at any significant level that the online community increased user revenue and the average increase of $29.02 in spending between the two sets of customers can be attributed to this feature.

$Revenue = 70.38 + 17.76*Joined? + 30.87*Treatment + 29.02*Joined?*Treatment$

2. Next, we used data of customers who churned 90 days after introducing the feature in a logistic regression model to predict churn probabilities and (by extension) the retention rate. The independent

variables in our model are ‘joined = Yes/No’, customer age with the firm during the launch of the community and average spend last 3 months of life with the firm. Variable ‘Joined?’ (whether the customer joined the online community or not) is highly significant in the model suggesting that it is associated with the probability of churn rate. The coefficients of customer age and last 3 months' spending are negative indicating that customers who have been with KyngaCell longer or those who have spent more money on their games are less likely to churn. However, both these variables are not significant in the model since the p-values are high.

Using the model and an assumed cutoff of 0.5 for churn, we predicted whether a customer would churn. The accuracy of the model is 62.81%, with a precision of 64.1% and recall/sensitivity is 84.7% (*2.1*).

Next, we determined the retention rate for each customer by subtracting the churn rate from 1. The average retention rate for customers who did not join the online community was 48.71%. This decreased to 29.26% for customers who joined the community, a decrease by 19.44 percentage points, a statistically significant difference (*2.2*). We conclude that the long-term effects (3 months) of introducing the online community were negative, leading to a decrease in the retention rate among customers.

3. The Customer Lifetime Value (CLV) is the net present value of the stream of expected future financial contributions from the customer. To investigate if the online community led to an increase in CLV, we must first calculate the predicted CLV per customer based on their average spend in the last 3 months of life with the firm (3MoAvg) and their predicted retention rate (r). Using an annual discount factor of 4% and assuming no acquisition cost, we can arrive at our CLV model:

$$CLV = 3MoAvg * 3 * 0.5 * (1 + 0.04/4)/(1 + (0.04/4) - r)$$

After running a two-sample t-test to compare the CLV of customers who joined and those who did not, we can conclude that the mean CLV for a customer who joined the online community is expected to be greater than those who did not (2.3). Additionally, we can be 95% confident that the average CLV for customers who joined is at least $145 higher than those who did not.

4. Next, we used acquisition data to investigate if customers who were onboarded organically versus via a campaign customers resulted in a difference in spending or revenue. We ran linear regressions for each segment, regressing the difference between the Month before and Month after spending on the categorical x variable 'Joined'.

$$LRevenue_{DiffCampaign} = 30.8 + 26.9 * Joined?$$

$$LRevenue_{DiffOrganic} = 31 + 32.12 * Joined?$$

On analyzing the coefficients, we see that for customers acquired via campaigns, the revenue increase post the community feature was $26.9 more for those subsets of customers who were invited to join. This increase was slightly more ($32.12) for organically acquired customers *(2.5)* This right off the bat suggests that organically acquired customers reacted better to the community feature than the other segment. On calculating the quarterly retention rate, we see a similar trend wherein retention is only slightly higher (approx. 3.1%) for organically acquired customers. Further, extrapolating the average last 3 month spend to get quarterly revenue, we calculate the CVL for both segments. On performing a two sided t-test to compare the overall mean CLV between both segments, we cannot conclude that there is a difference. However, after segmenting between joined and not joined, for organic customers, we see an increase in CLV of about 51% for those customers who joined. For campaign customers, the increase is ~42%. *(2.52)*.

Recommendations and Managerial Implications: Based on the analysis, we see that launching an online community has pros and cons. Our first model told us the online community helped to increase user revenue. After launching the online community, the average monthly user revenue increased by approximately $29. However, the online community had a negative long-term effect on retention rates. We would suggest KyngaCell to direct effort into making profits from retained customers and invest in improving the online community to increase retention rates.

Organically acquired customers responded better to the online community in terms of revenue growth and CLV. However, our analysis did not take into account the added acquisition cost of Campaign customers; the true CLV for campaign customers is expected to be even lower. To boost the effectiveness of such features, KyngaCell could consider releasing such features primarily to organically acquired customers, who might have more inclination or passion towards their products and community. Also, they could also try focusing more on organic marketing to prioritize natural outreach over promotional efforts, targeting over time growth in retention and CLV.

Limitations: i. Given data lacks long-term details beyond 3 month churn and average revenue, for instance, M-o-m churn and revenue would help us further analyze the behavior of churned customers - whether they return in a later time period or leave permanently. ii. In the model, heterogeneity is introduced because customers in the sample have spent different amounts of time with the firm before the treatment. iii. Our CLV formula accounts for income very far into the future. iv. Introduction of lurking variables such as user experience and login time is another limitation affecting the model.

Conclusion

Although we have built several models and calculated metrics to discover and quantify the effects of launching an online community on customers, there are limitations that may have a potential negative impact. For example, since we only have the data points for only 3 months after launching the online community, it is difficult to investigate the long-term effect of the online community on customers. Moreover, for the churn rate logistic model, the variables we considered might not be enough. There is a possibility of lurking variables to impact the predicted churn rate such as user experience and average login time. There is a scope for further analysis with more details on relevant variables and taking into account the heterogeneity in the model.

References

1. *Classification, An Introduction to Statistical Learning*, by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani
2. Customer Lifetime Value: Marketing Models and Applications, (1998), by Paul D. Berger and Nada I. Nasr, Journal of Interactive Marketing 12 (1). (https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.390.2785&rep=rep1&type=pdf)

Appendix:

Initial EDA: Correlation matrix

	Customer ID	Month Before	Month After	Joined?
Customer ID	1.00000000	-0.05212729	-0.06727209	-0.08015299
Month Before	-0.05212729	1.00000000	0.93240367	0.22643957
Month After	-0.06727209	0.93240367	1.00000000	0.50675359
Joined?	-0.08015299	0.22643957	0.50675359	1.00000000

1. Data 1: Regression Output for Dif-in-Dif

Model: y = intercept + b1*JoinedCommunity + b2*Treat + b3*JoinedCommunityxTreat

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	70.37606838	3.6	19.7	0.000000000	63.4	77.4	63.4	77.4
X Variable 1 - JoinedCommunity	17.75807797	5.6	3.2	0.001507997	6.8	28.7	6.8	28.7
X Variable 2 - Treatment	30.87179487	5.0	6.1	0.000000002	21.0	40.8	21.0	40.8
X Variable 3 - JoinedCommunity xTreatment	29.01844903	7.9	3.7	0.000253448	13.6	44.5	13.6	44.5

	Pre	Post	Dif (Post-Pre)	Dif-in-Dif
Joined Community	88.13	148.02	59.89	
Did Not Join Community	70.38	101.25	30.87	**29.02**

So compared to Customers who did not join the community, Customers who joined the community spent 'on average' $29.02 more. The p-value is very very low, corroborating that the difference is statistically significant.

2.1 Data 2: Logistic Regression Output

Model: p = p(Churn|X) : Probability of customers churning given whether they join the community or not, their age with the firm at the launch of online community and average spend last 3 months of life with the firm

log(p/(1-p)) = intercept + b1*JoinedCommunity + b2*CustAge + b3*Avg3MonthSpend

Variable ‘Joined?’ (whether the customer joined the online community or not) is highly significant in the model which means that the probability of customer churn is affected by whether the customers join the community or not

```
Call:
glm(formula = Data_2$Churn ~ Data_2$Joined + Data_2$CustAge +
    Data_2$Avg3MonthSpend, family = binomial(link = "logit"),
    data = Data_2)

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
Data_2$Joined          0.917627   0.355216   2.583  0.00979 **
Data_2$CustAge        -0.051796   0.073144  -0.708  0.47886
Data_2$Avg3MonthSpend -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
```

Confidence intervals:

```
                            2.5 %      97.5 %
(Intercept)           -0.58342556 1.524159597
Data_2$Joined          0.23199636 1.629262848
Data_2$CustAge        -0.19585547 0.091807565
Data_2$Avg3MonthSpend -0.01406909 0.008176207
```

Confusion matrix:

```
            Actual 1 Actual 0 Total
Predicted 1      100       56   156
Predicted 0       18       25    43
Total            118       81   199
```

2.2 T-test to check if difference in retention is significant

```
> t.test(join$retention, no_join$retention, alternative = "two.sided",

	Welch Two Sample t-test

data:  join$retention and no_join$retention
t = -47.062, df = 185.83, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2026497 -0.1863435
sample estimates:
mean of x mean of y
0.2926829 0.4871795
```

2.3 Calculating CSLV

Question 3 code:

```{r}

Data_3['QuarterlyChurnRate'] <- churnprob

Data_3['MonthlyChurnRate'] <- 1-((1-churnprob)^(1/3))

Data_3['MonthlyRetentionRate'] <- 1-Data_3$MonthlyChurnRate
```

```
Data_3['CustomerLifetime'] <- 1/Data_3$MonthlyChurnRate ##Lifetime in months

Data_3['CustomerLifetime'] <- round(Data_3$CustomerLifetime, digits = 0) ##Lifetime in months

Data_3


## Calculate CLV

discount_rate <- 0.04/4

rev_nojoin <- Data_3$`Average Spend Last 3 months of Life with the firm`[which(Data_3$`Joined?`==0)]*3

rev_join <- Data_3$`Average Spend Last 3 months of Life with the firm`[which(Data_3$`Joined?`==1)]*3

Data_3['clv_nojoin'] <- rev_nojoin*0.5*(1+discount_rate)/(1+discount_rate-(1-Data_3['MonthlyChurnRate']))

Data_3['clv_join'] <- rev_join*0.5*(1+discount_rate)/(1+discount_rate-(1-Data_3['MonthlyChurnRate']))


summary(Data_3['clv_nojoin'])

summary(Data_3['clv_join'])


clvtest <- t.test(Data_3['clv_join'],Data_3['clv_nojoin'], alternative = "greater")

clvtest
```

Welch Two Sample t-test

data: Data_3["clv_join"] and Data_3["clv_nojoin"]

t = 9.0409, df = 395.59, p-value < 2.2e-16

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

 145.5993   Inf

sample estimates:

mean of x mean of y

 587.2348 409.1617

2.4 Calculating revenue growth difference between two customer segments

Question 5: Segment 1 = Organic - Customers who joined the game organically
Segment 2 = Campaign - Customers who joined through the advertising campaign

campaign<-subset(Data_3, Data_3$`Campaign/Organic` == 1)

ORGANIC<-subset(Data_3, Data_3$`Campaign/Organic` == 0)

For Campaign

urevenue_lmcampaign <- lm(campaign$`Month After`-campaign$`Month Before` ~ campaign$`Joined?`, data = campaign)

```
Call:
lm(formula = campaign$`Month After` - campaign$`Month Before` ~ 
    campaign$`Joined?`, data = campaign)

Residuals:
     Min       1Q   Median       3Q      Max 
-15.8000  -6.9857  -0.2571   7.2857  17.2857 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)          30.800      1.004   30.67   <2e-16 ***
campaign$`Joined?`   26.914      1.597   16.85   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.696 on 122 degrees of freedom
Multiple R-squared:  0.6994,    Adjusted R-squared:  0.697 
F-statistic: 283.9 on 1 and 122 DF,  p-value: < 2.2e-16
```

For Organic

urevenue_lmorg<- lm(ORGANIC$`Month After`-ORGANIC$`Month Before` ~ ORGANIC$`Joined?`, data = ORGANIC)

```
lm(formula = ORGANIC$`Month After` - ORGANIC$`Month Before` ~
    ORGANIC$`Joined?`, data = ORGANIC)

Residuals:
    Min      1Q  Median      3Q     Max
-18.121  -8.561   1.879   7.000  14.000

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         31.000      1.407   22.03   <2e-16 ***
ORGANIC$`Joined?`   32.121      2.122   15.14   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.121 on 73 degrees of freedom
Multiple R-squared:  0.7584,	Adjusted R-squared:  0.7551
F-statistic: 229.2 on 1 and 73 DF,  p-value: < 2.2e-16
```

2.51 Retention and Churn Rates

	retention rate	churn rate
Organic	0.4266667	0.5733333
Campaign	0.3951613	0.6048387

2.52 CLV comparison

rev_organicnojoin <- ORGANIC$`Average Spend Last 3 months of Life with the firm`[which(ORGANIC$`Joined?`==0)]*3

rev_organicjoin <- ORGANIC$`Average Spend Last 3 months of Life with the firm`[which(ORGANIC$`Joined?`==1)]*3

rev_campaignnojoin <- campaign$`Average Spend Last 3 months of Life with the firm`[which(campaign$`Joined?`==0)]*3

rev_campaignjoin <- campaign$`Average Spend Last 3 months of Life with the firm`[which(campaign$`Joined?`==1)]*3

ORGANIC['clv_nojoin'] <- rev_organicnojoin*0.5*(1+discount_rate)/(1+discount_rate-(1-ORGANIC['QuarterlyChurnRate']))

ORGANIC['clv_join'] <- rev_organicjoin*0.5*(1+discount_rate)/(1+discount_rate-(1-ORGANIC['QuarterlyChurnRate']))

campaign['clv_nojoin'] <- rev_campaignnojoin*0.5*(1+discount_rate)/(1+discount_rate-(1-campaign['QuarterlyChurnRate']))

campaign['clv_join'] <- rev_campaignjoin*0.5*(1+discount_rate)/(1+discount_rate-(1-campaign['QuarterlyChurnRate']))

```
> mean(ORGANIC$clv_nojoin)
[1] 175.3274
> mean(ORGANIC$clv_join)
[1] 263.5305
> mean(campaign$clv_nojoin)
[1] 178.6049
> mean(campaign$clv_join)
[1] 254.8963
```

2.53 T-test to check for revenue difference between organic and campaign customers post online community

```
> clvtest <- t.test(ORGANIC['clv'],campaign['clv'], alternative = "two.sided")
> clvtest

        Welch Two Sample t-test

data:  ORGANIC["clv"] and campaign["clv"]
t = 0.67304, df = 155.56, p-value = 0.5019
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -14.63501  29.76207
sample estimates:
mean of x mean of y
 208.8060  201.2425
```