Customer Lifetime Value - Group 2C

Animesh Danayak, Dave Arno, Dhwani Mehta, Edwin Liu, Letian (William) Ma and Max Gordon

BAX-401-002 Information, Insight, and Impact

TABLE OF CONTENTS

Executive Summary

KyngaCell, a mobile gaming powerhouse, introduced a new online community feature, *Nicht-Soporfic*. The firm hopes that this new online community feature will allow users to be more engaged and connected with the game and in return, improved user revenue, retention, and ultimately increased customer lifetime value (CLV). We were tasked with testing these hypotheses and quantifying the effectiveness of this new feature. The three datasets that we were given contained customer information such as whether the customer joined the online community, spending before and after, customer age with the firm, average spend of the last three months, whether the customer joined organically or through a campaign, and most importantly, if the customer churned.

As a result of this analysis, we concluded that:

- *The community increased user revenue, on average, by $29/user*
- *The community did not increase retention. Churn rate for users who did not join the community was 50.21% while those who joined had a churn rate of 70.73%*
- *The community led to an increase in CLV, on average, by $14.*
- *How a player joined the game (organically or through a campaign) had very little impact on whether they will churn or the amount they spend.*

Overall, the online community offered significant results in increased revenue and CLV, but did not appear to increase retention rates. More data are needed in order to dive deeper into this category to understand why that happened.

Introduction & Problem Definition

KyngaCell, a mobile gaming company, wants to determine the impact of having an online community on user revenue and retention within its latest game, Nicht-Soporific. Their CFO has approached us to identify three key questions: a) *Whether the community has increased user revenue, b) Whether the community has led to increased retention, c) Whether the community has led to an increase in the CLV.*

We have been provided with details like customer spend before and after the launch of the community, customers who churn within 90-days of community launch, customer age with the firm at the time of joining the community, average spend within the game for these 90 days, and if the customer joined due to ad campaigns or organically. Based on these parameters, we also aim to answer additional questions like the magnitude of change in revenue, retention and CLV due to the online community as well as to answer if ad campaigns helped to gain new users.

Data Selection & Model Development

1. Has the online community increased user revenue?

We ran a diff-and-diff and two sample t-tests to determine the average revenue difference between joined v/s not joined. The logistic regression model (generalized linear model) was run using variable - 'Joined?'. We rejected variables - Age with firm and Average spend as they had negative intercepts and were not significant.*(Refer to Appendix Figure 1.1 and 1.2)* The results showed that joining the online community has

increased user revenue by an average of $29.0184/ user. With p-values being very close to 0, the results are significant.

2. Has the online community led to increased retention?

We predicted the probability of churn rate for all users using a trained logistic regression model with several iterations, and we rejected the variables with negative estimates since they were not significant. The model gives an estimate of 0.83 which suggests that joining the community increases the user churn rate, which is significant at 0.1% significance level. *(Refer to Appendix Figure 1.2 and 2.1)*

3. Has the online community led to an increase in CLV?

We calculated the CLV using the formula $CLV = \sum_{t=1}^{T} (mr^{t-1} / (1+i)^{t-1})$

m = Margin per customer

i = Discount rate (6%) *(See References No - 3)*

r = probability of customer retention

T = Customer lifetime with the firm

The online community has increased the CLV of customers who joined. We computed the CLV for those who joined and for those who did not join and found that the difference is significant, with a p-value of 0.02926. *(Refer to Appendix Figure 3.1)*

4. Magnitude of change:

a. **User revenue:** The results have shown that users who joined the online community had an average user revenue of $59.89/ user, compared to $30.87/user who did not. This is a significant difference to the company, almost doubling the revenue of users who did not participate in the online community. This could suggest that the new feature has provided additional channels of engagements with the users and has been well received by users.
b. **Retention and CLV:** On the other hand, we see an increase in churn rate three months after the introduction of the online community in both our prediction model and actual figures. The churn rate for users who did not join was 50.21% while those who joined had a higher churn rate of 70.73%. CLV for those who joined increased on average, by $14, compared to those who did not join.

5. Impact of ad campaigns

The provided variable 'Campaign/Organic' records whether a player joined the game due to the firm's ad campaign, or organically. This allows us to quantify the difference in value between both player groups. One metric we can produce is individual acquisition cost which would equal "Total Campaign Cost" / "Total Customers Gained".

As we see in Figure 5.1, the "Campaign/Organic" variable had very high p-values when modeled with most important success metrics. Therefore we can conclude that whether a player joined the game due to the ad campaign

Response Variable	P-value
Average Customer Spend	0.987
Joined Community	0.534
Churn	0.661

Figure 5.1 - p-values of the Campaign/Organic variable after regressing different variables

has very little effect on whether they will join the community, churn, or the amount they spend. This is important for the firm to know because an ad campaign factors into a customer's acquisition cost, decreasing their Customer Lifetime Value (CLV). We can conclude that organic players are more valuable to the firm and they should factor this into their future decisions regarding advertising. The next step would be to calculate ad costs/user and subtract it from the CLV of the customers where "Campaign/Organic" equals 1. The firm will then be able to see the difference in value between both groups.

Conclusion:

In summary, our analysis found that the introduction of the online community increased user revenue, and customer lifetime value, but reduced the customer retention rate. Whether people joined the online gaming community or not is an important feature to explain the retention rates of the customers.

Further analysis could be done to isolate the impact of the online community by conducting an A/B test. Our analysis could also be enhanced by additional user metrics, such as in game time before and after the introduction of the online community, to draw correlation between user engagement and the introduction of the online community.

References/ Citations:

1. M. Shahbandeh. (2018) Cosmetics Consumer Behavior in the US - Statistics & Facts. Retrieved from: https://www.statista.com.
2. Świątkowski, Wojciech. (2015). Re: What is the meaning of a positive/negative interaction term in a moderation analysis in multiple regression?. Retrieved from: https://www.researchgate.net.
3. Peter J. Coughlan (2018), Re: Note on Home Video Game Technology and Industry Structure Harvard Business Review Case Study. Published by HBR Publications from https://oakspringuniversity.com
4. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: With applications in R.

Appendix:

1.1 Linear Model for the first problem. Joining the Community has a positive impact on user revenue.

```
Call:
lm(formula = df$revenue ~ df$`Joined?`)

Residuals:
     Min       1Q   Median       3Q      Max
-15.8718  -7.8810   0.1098   7.1282  15.1098

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    30.872      0.830   37.20   <2e-16 ***
df$`Joined?`   29.018      1.293   22.44   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.977 on 197 degrees of freedom
Multiple R-squared:  0.7189,    Adjusted R-squared:  0.7174
F-statistic: 503.7 on 1 and 197 DF,  p-value: < 2.2e-16
```

1.2 Ran a generalized linear regression model with independent variables. We rejected the variables with negative estimates since they were not significant.

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

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
mydata$`Joined?`                                                  0.917627   0.355216   2.583  0.00979 **
mydata$`Customer Age with Firm at time of launching the online community` -0.051796   0.073144  -0.708  0.47886
mydata$`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
```

```
> mylogit<-glm(mydata$`Churned at 3 months after launch of the online community`~mydata$`Joined?`,
+              data=mydata,family=binomial(link="logit"))
> summary(mylogit)

Call:
glm(formula = mydata$`Churned at 3 months after launch of the online community` ~
    mydata$`Joined?`, family = binomial(link = "logit"), data = mydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5676  -1.1993   0.8322   1.1557   1.1557

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        0.05129    0.18496   0.277  0.78153
mydata$`Joined?`   0.83110    0.30515   2.724  0.00646 **
---
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: 261.26  on 197  degrees of freedom
AIC: 265.26

Number of Fisher Scoring iterations: 4
```

2.1 Confusion matrix to identify accuracy of the model

```
> confMat2<-confusionMatrix(data = as.factor(predchurn),reference = as.factor(mydata$`Churned at 3 months after launch of the online community`),positiv
e = "1")
Warning message:
In confusionMatrix.default(data = as.factor(predchurn), reference = as.factor(mydata$`Churned at 3 months after launch of the online community`),  :
  Levels are not in the same order for reference and data. Refactoring data to match.
> confMat2
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0   0   0
         1  81 118

               Accuracy : 0.593
                 95% CI : (0.5212, 0.6619)
    No Information Rate : 0.593
    P-Value [Acc > NIR] : 0.5305

                  Kappa : 0

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 1.000
            Specificity : 0.000
         Pos Pred Value : 0.593
         Neg Pred Value :   NaN
             Prevalence : 0.593
         Detection Rate : 0.593
   Detection Prevalence : 1.000
      Balanced Accuracy : 0.500

       'Positive' Class : 1
```

3.1 Tests to check in CLV due to joining the community

$H_o: \mu_{joined} = \mu_{didnotjoin}$

$H_1: \mu_{joined} > \mu_{didnotjoin}$

```
wilcox.test(clv_joined$CLV_final, clv_no_join$CLV_final, alternative = c('g'))
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  clv_joined$CLV_final and clv_no_join$CLV_final
## W = 5554, p-value = 0.02926
## alternative hypothesis: true location shift is greater than 0
```

```
clv_no_join <- finaldata %>% filter(`Joined?` == 0)
clv_joined <- finaldata %>% filter(`Joined?` == 1)

mean(clv_no_join$CLV_final) - mean(clv_joined$CLV_final) #CLV for those who joined increased by ~14
```

```
## [1] -13.90326
```

```
Call:
lm(formula = df$clv2 ~ df$`Joined?`)

Residuals:
    Min      1Q  Median      3Q     Max
-42.017 -17.938  -1.255  16.923  44.067

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    56.814      1.995  28.482  < 2e-16 ***
df$`Joined?`   10.270      3.107   3.305  0.00113 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21.58 on 197 degrees of freedom
Multiple R-squared:  0.05253,   Adjusted R-squared:  0.04772
F-statistic: 10.92 on 1 and 197 DF,  p-value: 0.001129
```

5.1 - Identifying avg spend of customers who joined organically v/s those who joined through ad campaigns

```
"Avg Spend"

Call:
glm(formula = df$avg_spend ~ df$`Campaign/Organic`)

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            80.30222    3.51519  22.844   <2e-16 ***
df$`Campaign/Organic`   0.07412    4.45312   0.017    0.987
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


"Joined Community"

Call:
glm(formula = df$`Joined?` ~ df$`Campaign/Organic`, family =
binomial(link = "logit"))

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)             -0.2412     0.2326  -1.037    0.300
df$`Campaign/Organic`   -0.1845     0.2964  -0.622    0.534


"Churn"

Call:
glm(formula = df$churn ~ df$`Campaign/Organic`, family =
binomial(link = "logit"))

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)             0.2955     0.2335   1.266    0.206
df$`Campaign/Organic`   0.1302     0.2971   0.438    0.661
```

5.2 - Mean retention rate and Revenue for campaign/organic

```{r}
finaldata %>% filter(`Campaign/Organic` == 0) %>% select(retention) %>% unlist %>% mean
finaldata %>% filter(`Campaign/Organic` == 1) %>% select(retention) %>% unlist %>% mean
```

```
[1] 0.4008525
[1] 0.4107747
```

```{r}
finaldata %>% filter(`Campaign/Organic` == 0) %>% select(revenue) %>% unlist %>% mean
finaldata %>% filter(`Campaign/Organic` == 1) %>% select(revenue) %>% unlist %>% mean
```

```
[1] 45.13333
[1] 41.43548
```