Analysis of Community Feature and Insight Into Customer Analytics: KyngaCell

Anirudh ‘Ani’ Menon

Kexin ‘Shera’ Huang

Sanka ‘Nitesh’ Naga Nitesh

Yashi Tiwari

EXECUTIVE SUMMARY

The problem statement relates to a firm known as Kynga Cell, a mobile gaming powerhouse. The firm has launched a new online community feature in their game that allows users to connect and interact within the game. The firm aims to assess the impact of this community under the assumption that this new feature will lead to an improved user related revenue and retention in the game.

Conducting an analysis of the feature to assess the metrics in place to quantify the impact of the community feature is very important. The primary step related to the analysis of the feature would be to assess whether the feature has generated an increased revenue when compared to the users who have not received access to this online community feature.

Further, the online community feature must be assessed to comprehend the customer retention rate by creating a logistic regression model that measures the relationship between the churn of the customer and whether the user has been invited and joined and use the community feature.

The Customer Lifetime Value between the users who joined versus the users who have not joined the community feature to gain insight about the potential customer contributions related to the subsets. We finally created a logistic regression model that measured the differences in relationships between churn related to customers who have joined due to the community feature versus customers who joined 'organically'.

The following studies reveal while the community feature may be used by the organization to increase and grow revenue, factors such as the retention rate and the customer lifetime value show that the feature does not aid in the growth of these factors. Opting in for another strategy to assess and reduce churn would be the recommendations in the following case.

INTRODUCTION

KygaCell introduced a new feature - online community - to the existing mobile game. We want to help the organization to assess the impact and magnitude of the new online community feature to measure whether the introduction of the online community leads to increase in revenue and further leads to positive impact on retaining users in the long run.

PROBLEM FORMULATION

To explore the true impact of the new feature, we will initially focus on 3 main aspects: whether the online community had a positive impact on revenue, retention, and customer lifetime value. Further assessing the impact, we aim to quantify the effect and discuss the potential limitations of the analysis. Measures such as the retention rate of the customers and the Customers' Lifetime Value (CLV) have to be assessed to determine if the feature has increased these respective measures. In the end, we will explore any new variable that will add value to our assessment (e.g. Insights related to other variables such as whether the customer joined organically or through the new feature implemented in the game).

DATA DESCRIPTION

All 3 datasets consist of quantitative data and binary qualitative data. We denoted 0 as "False/No" and 1 as "True/Yes". The first dataset presented users' dollar spending for the month before and after launching the new feature, and whether or not users got invited to the game and accepted an invitation ("Joined?" in Column D)(Appendix[1]). The second dataset added extra information on Column C as how long customers stayed with the firm at the time of launching new feature in months, Column D as whether customers dropped ("churned") the game 3 months after launching the new feature, and Column E as customers' average dollar spending for their last 3 months with firm. For the third dataset, we introduced "Campaign/Organic" on top of the second dataset to indicate whether customers joined the game by marketing campaign or they just joined the game naturally/organically without the campaign influence.

MODEL DEVELOPMENT

We stated two assumptions regarding the characteristics of users to determine whether the revenue has increased after launching an online community. The primary assumption is that users with value of 1 in the dataset are invited to use the new feature and have accepted the invite. The secondary assumption is that all users listed in the dataset possess similar characteristics. To assess whether the new feature has led to an increase in user revenue, differences in differences (diff-in-diff) will be analyzed. First, we calculated the averages of users' spending for those who joined in the prior period, joined in the post period, not joined in the prior period, and not joined in the post period. We then calculated the difference between prior and post periods for joined and not joined. Given the results, we calculated the diff-in-diff between prior and post periods for joined and not joined (Appendix [2, 3]). We also ran a regression for "diff-diff" model: with average spending as Y, X variables are treatment period, "whether joining or not", and the interaction effect between "joined and treated" was an extra variable since we'd like to separate people who joined and treated from people who weren't treated but joined (Appendix [3]). A test of equal variances was run showing that probability of variances being equal is not significant and that the variances must be considered individually. To assess whether there is an increment in retention rate, we built a logistic regression model (Appendix[5]) using all the variables provided. The predictor variable is churn probability whereas the response variables are whether the user has been invited to use the new feature, customer's age with the firm, and average spending within 3 months with the firm. To determine customers' churning status, we started by comparing the probability of individuals with the given margin of 0.5. Above 0.5, customers tend to churn and vice versa. Thus, we subtracted churn probability from 1 to get retention probability. Assessing CLV measures the business worth of the customer relationship over a period of time. We assumed that the discount rate amounts to 10% and no customer acquisition cost present. We conducted CLV assessment on users who did not churn at 3 months. A difference between the mean CLV between users who were invited to use the online community and those who were not invited to use the feature. A test of equal variances was run which shows that the probability of the variances being equal is not significant and that the variances must be considered individually. To establish a relationship between Campaign/Organic and customers' churning status at 3 months, users were classified as who joined the game by new-feature campaign and who joined organically, aligning with previous assumptions. We created a logistic regression model (Appendix[9]) containing the predictor variable as the churn at 3 months of life and the response variables as whether the user joined due to the new feature or 'organically', and the user's age at the time of launching the new feature and user's average spending in the last 3 months of life with the firm.

RESULTS

The regression related to differences in differences provides an insight that the online community feature showed an increase in user revenue for the immediate following month post-launch of the new feature. The new feature tends to have a positive upside on users' revenue. However, the 3-month retention rate shows that the new feature is not capable of retaining customers. In other words, users were interested to explore the new feature when it first launched but the feature seemed unappealing later on. After doing EDA, we see that the retention rate for users who didn't join is higher than that of users who joined the online community. This is further supported by the odds ratio and T-test (Appendix[4, 6]). Differences between the Customer Lifetime Values of the users who joined versus those who didn't join amounted to $30. So to summarize, the CLV and user revenue increased and retention dropped for users who joined w.r.t users who didn't join.

Quantifying the results for the above mentioned statements, we see that the new feature shows an increase in the average user revenue by $29 and also an average increase in the CLV by $30 for customers who joined the online community. However, the retention did not increase. In-fact the 3-month retention rate of users who joined the online community is just about 30% when compared to 48.7% for the users who didn't join (Appendix[7, 8]). Speaking of facts, we know that the "CLV and retention rate are positively correlated. However, our results depict that the retention rate and CLV are negatively correlated." *(Customer Gauge, 2022, Reference [1])*. This would only be possible if and only if our sample has "whales" or in other words high paying customers.) It is because of this reason that we see an increase in CLV and user revenue despite the retention dropping. In this analysis the sample data given to us is very small and also skewed. Also, we haven't taken into consideration the gaming market, the competitors, the seasonality and many other factors as such which add up to the potential limitations. When customers were classified based on the mode of their entry into the game(Campaign/Organic), we noticed that the retention rate of customers who have not joined is higher when compared to those who haven't joined the community(Appendix[13, 14]). This leads to the result that the campaign is ineffective and not significant in terms of retention of customers. However the CLV of customers is higher for those who have joined compared to those who haven't joined irrespective of the mode of entry. So we can say that the mode of entry is not significant.

RECOMMENDATIONS AND MANAGERIAL IMPLICATIONS

Based on the above, we can say that the online community as a feature has no long term upside. Except for the initial surge in the user revenue and CLV, this feature leads to drop in retention and hence would also result in drop in CLV, which is not a good sign for any firm. As a next step, we would ask KyngaCell to pause this feature and provide a bigger sample to conduct deeper funnel analysis (cohort level analysis) In the meantime, we would also recommend KyngaCell to test new features that would increase retention which would translate to increase in CLV.

CONCLUSION

The interpretation of metrics used to assess the impact of introduction of the online community feature gave results explaining that the feature did lead to an increase in revenue but did not solve the problem of customer retention. Since we have behavioral trends of the users for only 3 months, we cannot definitely conclude that the significance of these results would remain unchanged for an extended period of time. We know for a fact that an increase in retention will lead to an increase in CLV, which is not the case here. Hence we can conclude that this scenario is an exceptional case and also a good learning for the firm.

APPENDIX

[1] Basic EDA

Composition of users (joined vs not joined)

Online Community Feature	no. of users	% compositon
Joined	82	41.2%
Not Joined	117	58.8%
Total Users	199	

Actual Churn and Retention rates (joined vs not joined)

	No. of Users	Churned	Retained	Churn Rate	Retention Rate
Joined	82	58	24	70.7%	29.3%
Not Joined	117	60	57	51.3%	48.7%
Total	199	118	81		

Composition of users based on mode of entry (organic vs campaign)

Mode of Entry	no. of users	% compositon
Organic	75	37.7%
Campaign	124	62.3%
Total	199	100.0%

Actual Churn and Retention rates of users who have come organically

ORGANIC

	No. of Users	Churned	Retained	Churn Rate	Retention Rate
Joined	33	22	11	66.7%	33.3%
Not Joined	42	21	21	50.0%	50.0%
Total	75	43	32		

Actual Churn and Retention rates of users who have come through campaign

CAMPAIGN

	No. of Users	Churned	Retained	Churn Rate	Retention Rate
Joined	49	36	13	73.5%	26.5%
Not Joined	75	39	36	52.0%	48.0%
Total	124	75	49		

[2] Diff -in - Diff (normal method)

Online Community/ Time of Launch	Pre	Post	Diff (Pre - Post	Difference-in-Difference
Joined	88.13	148.02	59.89	
Not Joined	70.38	101.25	30.87	**29.02**

[3] Diff -in - Diff (regression method)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.58
R Square	0.34
Adjusted R Square	0.34
Standard Error	38.59
Observations	398

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	3,03,207.61	1,01,069.20	67.89	2.09234E-35
Residual	394	5,86,594.74	1,488.82		
Total	397	8,89,802.35			

		Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
	Intercept	70.38	3.57	19.73	0.00	63.36	77.39	63.36	77.39
TREATMENT [BEFORE]	X Variable 1	30.87	5.04	6.12	0.00	20.95	40.79	20.95	40.79
JOINED	X Variable 2	17.76	5.56	3.20	0.00	6.83	28.68	6.83	28.68
TREATMENT [BEFORE] X JOINED	X Variable 3	29.02	7.86	3.69	0.00	13.57	44.47	13.57	44.47

[4] F-test and T-test results for user revenue

```
        F test to compare two variances

data:  didnot_join_diff and join_diff
F = 1.0394, num df = 116, denom df = 81, p-value = 0.8603
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6890303 1.5452111
sample estimates:
ratio of variances
          1.039431


        Two Sample t-test

data:  didnot_join_diff and join_diff
t = -22.444, df = 197, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -31.56823 -26.46867
sample estimates:
mean of x mean of y
 30.87179  59.89024

[1] 29.01845
```

[5] Logistic Regression for Retention

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

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

[6] Odds ratio for Churn Probability

```
                                                                     OddsRatio  2.5 % 97.5 %
(Intercept)                                                             1.5879 0.5580 4.5913
`Joined?`                                                               2.5033 1.2611 5.1001
`Customer Age with Firm at time of launching the online community`      0.9495 0.8221 1.0962
`Average Spend Last 3 months of Life with the firm`                     0.9971 0.9860 1.0082
```

[7] CLV and average retention for users joined vs not joined

Metric/Online Community Feature	joined	not joined
avg retention	29.3%	48.7%
clv	213.00	183.05

[8] F-test and T-test results for user CLV

```
        F test to compare two variances

data:  joined_CLV and notjoined_CLV
F = 0.39552, num df = 23, denom df = 56, p-value = 0.01672
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2067291 0.8386249
sample estimates:
ratio of variances
         0.3955202


        Welch Two Sample t-test

data:  joined_CLV and notjoined_CLV
t = 2.0504, df = 66.897, p-value = 0.02212
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 5.586564      Inf
sample estimates:
mean of x mean of y
 213.0007  183.0487
```

[9] Logistic Regression for Retention on introduction of mode of entry (organic vs campaign)

```
Call:
glm(formula = organic_df$Churned.at.3.months ~ ., family = binomial(link = "logit"),
    data = organic_df)

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
Joined.                                              0.930000   0.356426   2.609  0.00907 **
Average.Spend.Last.3.months.of.Life.with.the.firm   -0.003006   0.005669  -0.530  0.59588
Customer.Age.with.Firm.at.time.of.launching.the.online.community -0.052077   0.073166  -0.712  0.47661
---
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
```

[10] Odds ratio for Churn probability on introduction of mode of entry (organic vs campaign)

```
                                                                  OddsRatio  2.5 % 97.5 %
(Intercept)                                                          1.4272 0.4717 4.3680
Campaign.Organic                                                     1.1969 0.6578 2.1767
Joined.                                                              2.5345 1.2742 5.1779
Average.Spend.Last.3.months.of.Life.with.the.firm                    0.9970 0.9859 1.0081
Customer.Age.with.Firm.at.time.of.launching.the.online.community     0.9493 0.8218 1.0958
```

[11] F-test and T-test results for user CLV (for users who have come organically)

```
        F test to compare two variances

data:  joined_organic_CLV and notjoined_organic_CLV
F = 0.40893, num df = 10, denom df = 20, p-value = 0.1474
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1474342 1.3979589
sample estimates:
ratio of variances
         0.4089341


        Welch Two Sample t-test

data:  joined_organic_CLV and notjoined_organic_CLV
t = 1.9174, df = 28.58, p-value = 0.03262
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 5.331354      Inf
sample estimates:
mean of x mean of y
 226.1437  179.1356
```

[12] F-test and T-test results for user CLV (for users who have come through campaign)

```
        F test to compare two variances

data:  joined_campaign_CLV and notjoined_campaign_CLV
F = 0.39366, num df = 12, denom df = 35, p-value = 0.08667
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1681857 1.1537931
sample estimates:
ratio of variances
           0.39366


        Welch Two Sample t-test

data:  joined_campaign_CLV and notjoined_campaign_CLV
t = 0.97414, df = 34.236, p-value = 0.1684
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -13.30813       Inf
sample estimates:
mean of x mean of y
 203.5408  185.4461
```

[13] CLV based on Mode of Entry/Online Community Feature

CLV

Mode of Entry/Online Community Feature	Joined	Not Joined
Organic	$226.14	$179.13
Campaign	$203.54	$185.44

[14] Retention Rate for Mode of Entry/Online Community Feature

RETENTION

Mode of Entry/Online Community Feature	Joined	Not Joined
Organic	31.8%	51.7%
Campaign	27.9%	47.1%

Reference List

[1] CustomerGauge. (n.d.). Retrieved November 11, 2022, from *https://customergauge.com/solutions*