Introducing an online community at KyngnaCell

Table of Contents

Executive Summary

Online communities have gained in popularity within the gaming sector over recent years. This report assesses the effectiveness of a new online community feature introduced in 2021 by KyngaCell, a mobile gaming powerhouse, for their latest game, *Nicht-Soporific*, by assessing its impacts on the metrics such as user revenue, retention, and customer lifetime value (CLV) through Difference-in-Differences and other regression models.

The data in use contain information about a set of players' in-game spending one month before and after the launch of the online community, their ages with the company at the time of joining the community, as well as their churn status and the average spend within 90 days after the community was introduced. Moreover, information about how users joined the game is also factored into the analysis.

The results indicate that joining the online community produces statistically significant positive impacts on user revenue, but it also leads to an undesired higher probability of customers churning and therefore no significant difference in CLV between people who joined the community and people who did not. Moreover, among those who joined the online community, people who found the game naturally generated more revenue compared with those who played the game due to previous marketing campaigns. The decision of whether to keep the online community should be contingent on what is more important to the company at the moment: if increased revenue is the current business focus, then the community is a useful feature; on the contrary, if increased retention is the current focus, then it is not recommended to keep the online community active.

Introduction

Increasing user revenue, retention, and customer lifetime value (CLV) has been one of KyngaCell's main business focuses. In early 2021, the mobile gaming giant incorporated a creative online community feature in their latest game, *Nicht-Soporific*, which enabled people to connect and interact with each other both during and outside gameplay. The management team believed that the new selling point would stimulate user activities, therefore improving the aforementioned key metrics.

This report intends to quantify the effects of introducing the online community on user revenue, retention, and CLV through various statistical techniques such as Difference-in-Differences and Logistic Regression. The data used for this analysis will be discussed, followed by a detailed explanation of the model development process and interpretations of the results. Specific recommendations will also be provided for the management regarding whether the company should continue with the online community feature in the long run.

Problem Formulation

The goal of this report is to investigate the impact of the new online community feature on key attributes like revenue, user retention and customer lifetime value (CLV) by analyzing certain variables like revenue, whether customers joined this new community or not, churn status, and effectiveness of the game's marketing campaign.

Given the availability of revenue data both before and after the introduction of the online community, we chose to conduct a form of causal analysis called difference-in-differences which compares the difference between those who joined and those who didn't across before and after time periods. This allows us to confidently generate an exact dollar amount for the impact of the community on both customer revenue and CLV. We were also able to extend this analysis to the impact of the game's marketing campaign in relation to the new online community.

For the impact of this new community feature on retention, we used a classification algorithm called logistic regression that can predict the probability of an individual churning or not based on their customer profile. This allows us to take into consideration multiple factors as they relate to churn as well as how the new community impacted the retention rate.

Linear regression was used to analyze the relationship between the community and CLV, mainly because it allows us to investigate whether a relationship between these variables exists, whether it is a positive or negative relationship, as well as whether it is a significant relationship that should influence business decisions.

Data Characteristics

To understand if the online community has led to increase in revenue, we would be looking at the amount spent by the customers before and after a month around the time of introducing this online community feature. We are also given if these customers have joined this community or not. This binary information gives a chance to conduct causal inference through difference-in-differences.

Next, to know the effect of this new feature on retention we will be investigating whether an individual churned after 3 months. Given the customer age with the firm allows us to address their behavior over time. Also, we have the average of the last 3 months spent and monetary metrics are popular for churn predictions (Tamaddoni, 2015).

We are also given if the customer is acquired through inorganic, as in through some campaign run by the firm, or organic manner. With this information, we shall try to analyze the retention and other useful metrics with this additional variable and see how it affects their business.
Although the analysis might not be highly accurate as CLV depends on many other variables, we try to establish how CLV is affected given the presence of the online community.

Model Development, Estimation and Results

For the question of whether the online community has boosted revenue, we used a form of causal inference where we investigate the difference between the community and non-community groups across the before and after time periods. Our results showed that people who joined the online community generated an additional $29.02 in revenue compared to those who didn't. This was also a statistically significant result.

To answer the second question of how the community affected churn, we used a classification algorithm called logistic regression, the equation for which is represented in Figure 1 in the appendix. In this particular case, we found that being a part of the online community made a person 2.5 times more likely to churn than those who didn't join the community holding all other variables constant.

In order to answer the final question of how being a part of the online community affects customer lifetime value, we used a customer margin of 50%, as well as a discount factor of 5%, which corresponds to the inflation rate in May of 2021, around the time this product was released (US BLS, 2021). The margin was derived from customers' spending in the last 3 months, which was converted to a quarterly level because that is the same level at which the retention rate was measured. We then used linear regression to analyze the effect of the community on CLV, the equation for which is in Figure 2. We found that joining the online community did not have a significant impact on CLV, which makes sense because it boosted churn rate which decreases CLV, but it also increased customer revenue which increases CLV.

We also investigated whether there exists any sort of interaction between the game's marketing campaign and the new online community by implementing the same causal inference from the first business question twice: once for those who found the game naturally, and once for those who found it because of a marketing campaign. We observed that among those who found the game naturally, the online community generated an additional $32.12 in revenue. In those who discovered the game as a result of the campaign, the online community only generated an additional $26.91in revenue.

Recommendation and Managerial Implications

Our analysis reveals the online community generated additional revenue in customers that joined, but it had the potential drawback of significantly increasing a user's probability to churn. As a result, our recommendations are contingent on what is currently more important to the business. If the priority right now is to improve the revenue generated by existing customers, then the community is a valuable feature that should be maintained. However, if the business' goal is more oriented towards maintaining a large customer base, then rolling out the community to the public does not make sense. It should be noted that while we were able to establish a causal link between the community and revenue, our analysis with respect to churn and CLV is not causal but is instead associative. More analysis, such as what we did in business question 1, would need to be done in order to establish a causal link.

Another, potentially better result could be achieved by conducting a deeper investigation into whether or not there is something about the online community that is driving users to churn at a higher rate. For instance, if the communities are becoming toxic, then this would likely drive higher churn rates and hurt retention. If the problem of retention is indeed specific to the communities and can be fixed, then this could potentially allow the company to reap the benefits of increased revenue without damaging retention. The only downside would be that the company would bear the cost of conducting a deeper analysis into the community.

With regard to the marketing campaign that was undertaken to bring awareness to the game, we recommend attempting to drive an increase in people who naturally join the game, such as by implementing a referral reward program. This is because we saw that the increase in user revenue caused by the online community was higher in those who joined the game naturally.

Conclusion

Considering the revenue impact, we found that joining the online community resulted in a statistically significant increase in revenue of $29.02. However, the odds of churning are 150% higher if you joined the community compared with someone who didn't join the online community. Customers who join the online community are willing to spend more in the short term but the retention rate is poor in the long term. If in the long run, the business plans to gain large aggregated value from the customers, it should be focusing on improving the retention. While we do not have important data regarding the online gaming communities, it's recommended that effective measures are taken to improve the user experience in order to improve the retention rate. There are multiple ways in which the business can boost the retention rate in the online gaming community such as by announcing milestones, weekly challenges, rewarding the players, and personalizing the invite and making it attention- grabbing.

In addition to this, the increase in user revenue caused by the community was higher for those who joined the game naturally in comparison to those who joined as a result of the marketing campaign. This can be done by allowing the current players to share their experiences and opinions through social media. Current players will help to spread the word about the game in the most efficient manner which would help the business to increase customer lifetime value i.e profits the game will earn per user.

References

Aravindakshan, Ashwin. “Primer on Customer Analytics”. BAX-401-001. University of California, Davis, 10 Nov, 2022. Lecture 7.

Tamaddoni, A., Stakhovych, S., & Ewing, M. (2015). Comparing churn prediction techniques and assessing their performance. Journal of Service Research, 19(2), 123–141. . https://doi.org/10.1177/1094670515616376

U.S. Bureau of Labor Statistics. (2021, June 16). *Consumer prices increase 5.0 percent for the year ended May 2021*. U.S. Bureau of Labor Statistics. Retrieved from https://www.bls.gov/opub/ted/2021/consumer-prices-increase-5-0-percent-for-the-year-ended-may-2021.htm

Appendix

Figures

$$log(odds_{churn}) = 0.4672 + 0.9196 Joined? - 0.0518 CustomerAgeWithFirm - 0.0030 * AvgSpend$$

Figure 1

$$CLV = 73.2561 + 46.8495 Joined? - 0.7450 CustomerAgeWithFirm - 2.3590 * Churn$$

Figure 2

CLV Calculation

The Customer Lifetime Value was calculated using the following formula:

$$m + \frac{mr}{(1+i)} + \frac{mr^2}{(1+i)^2} + \frac{mr^3}{(1+i)^3} + \ldots = m\left(\frac{1+i}{1+i-r}\right)$$

M represents the margin, i represents the discount rate (in our case we used inflation), and r represents the retention rate. M was calculated using a 50% profit margin which is what was given to us, and it was derived from the average monthly spend of an individual in the last 3 months. This metric needed to be converted to a quarterly level because that is the level on which retention was measured, and we did so by multiplying average monthly spend by 3. The retention rate, r, was calculated using a logistic regression that forecasted the probability of an individual churning. This was important because each individual will have a different churn rate based on factors such as their presence in the community, age with the firm, and average spend.

Out[191]:

OLS Regression Results

Dep. Variable:	clv	**R-squared:**	0.014
Model:	OLS	**Adj. R-squared:**	-0.002
Method:	Least Squares	**F-statistic:**	0.8927
Date:	Thu, 10 Nov 2022	**Prob (F-statistic):**	0.446
Time:	20:20:41	**Log-Likelihood:**	-1134.1
No. Observations:	199	**AIC:**	2276.
Df Residuals:	195	**BIC:**	2289.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	186.0318	14.003	13.285	0.000	158.415	213.648
Joined?	15.9927	10.720	1.492	0.137	-5.149	37.134
Customer Age with Firm at time of launching the online community	1.8077	2.574	0.702	0.483	-3.268	6.883
Churned at 3 months	-3.6765	10.749	-0.342	0.733	-24.876	17.523

Omnibus:	33.902	**Durbin-Watson:**	1.961
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	8.317
Skew:	0.022	**Prob(JB):**	0.0156
Kurtosis:	1.999	**Cond. No.**	14.0

Difference-in-Differences

The method used for evaluating the impact of the online community on revenue as well as the interaction between the campaign and the online community was a form of causal inference known as difference in differences, diff in diff for short. The basic principle behind diff in diff is to evaluate two different realities: one in which the online community was introduced, and one in which it wasn't. However, because we cannot actually observe two different realities, we create two different groups, one in the community and one out. We then observe the revenue in both categories before and after the community was introduced. We then take the difference within each group to see how revenue has changed over time within the community and non-community group. This gives us how revenue has changed over time, but it does not isolate the impact of the community on revenue. In order to do this, we then take the difference between these two differences and analyze the statistical significance using linear regression. The result is a clear picture of how the community affected the revenue per customer. We assume composition of groups being studied are stable over the time period we are concerned about. We also assume that both groups have parallel trends in their outcomes. That means, if no treatment was given, the difference between the two groups would be consistent over time.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              total_rev   R-squared:                       0.341
Model:                            OLS   Adj. R-squared:                  0.336
Method:                 Least Squares   F-statistic:                     67.89
Date:                Wed, 09 Nov 2022   Prob (F-statistic):           2.09e-35
Time:                        13:00:53   Log-Likelihood:                -2016.6
No. Observations:                 398   AIC:                             4041.
Df Residuals:                     394   BIC:                             4057.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     70.3761      3.567     19.729      0.000      63.363      77.389
g             17.7581      5.557      3.196      0.002       6.833      28.683
t             30.8718      5.045      6.120      0.000      20.954      40.790
gt            29.0184      7.859      3.692      0.000      13.568      44.469
==============================================================================
Omnibus:                      424.099   Durbin-Watson:                   1.982
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               27.202
Skew:                          -0.057   Prob(JB):                     1.24e-06
Kurtosis:                       1.724   Cond. No.                         6.45
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Note: the gt coef represents the diff in diff value.

Logistic Regression

In order to assess the impact of the community on churn, we used a classification algorithm known as logistic regression which outputs the probability of an event happening given certain independent variables. The accuracy score of our model, which is simply the number of observations the model correctly predicted/total number of observations, was 61.3%. We also evaluated the roc-auc score, which is the area under the curve when our model outputs are plotted. In order for our model to be better than random guessing, the roc-auc score must be greater than 50%. Our score was 58.8%.

Logit Regression Results

Dep. Variable:	Churned at 3 months after launch of the online community	**No. Observations:**	199
Model:	Logit	**Df Residuals:**	195
Method:	MLE	**Df Model:**	3
Date:	Thu, 10 Nov 2022	**Pseudo R-squ.:**	0.03133
Time:	13:04:07	**Log-Likelihood:**	-130.26
converged:	True	**LL-Null:**	-134.48
Covariance Type:	nonrobust	**LLR p-value:**	0.03796

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	0.4672	0.536	0.872	0.383	-0.583	1.517
Joined?	0.9196	0.355	2.589	0.010	0.223	1.616
Customer Age with Firm at time of launching the online community	-0.0518	0.073	-0.709	0.479	-0.195	0.092
Average Spend Last 3 months of Life with the firm	-0.0030	0.006	-0.524	0.600	-0.014	0.008

```
                                                          5%       95%  \
const                                               0.558242  4.560641
Joined?                                             1.250391  5.031874
Customer Age with Firm at time of launching the...  0.822677  1.095848
Average Spend Last 3 months of Life with the firm   0.986036  1.008163

                                                    Odds Ratio
const                                                 1.595600
Joined?                                               2.508348
Customer Age with Firm at time of launching the...    0.949489
Average Spend Last 3 months of Life with the firm     0.997038
```