William Lamazère - Selma Rahib - Emilie Nabet - Lily Renard

Customer Analytics
Pokémon Go Case

Summary

Introduction ***3***

Assignment 1: Creation of a Basetable ***3***

1) Who are the Active customers? 3
2) What are the the different profiles of your customers? 5
3) What's the CLV of your customers? 8

Assignment 2: Lifecycle grids ***9***

1) In the summer 9
2) In the fall 13

Takeaways of the assignments 1 and 2 15

Assignment 3: Churn Analysis ***15***

Recommendations ***17***

1) Acquisition of players 17
2) Boost the transactions 18
3) Retain key players 18
4) Prevent churn 19

Conclusion ***19***

Introduction

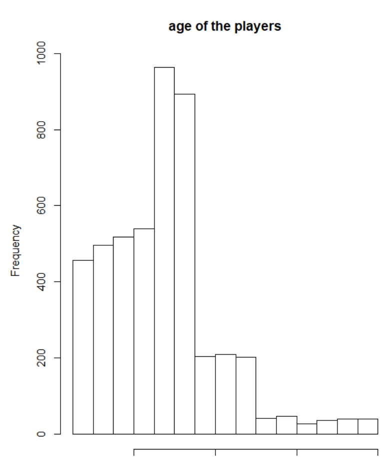
Pokemon Go is most certainly the biggest success in gaming app in the last couple of years. Launched in July 2016, based on augmented reality and displayed through all types of smartphones, the game has rapidly outranked any possible prediction in terms of revenue and engagement. Needless to say, the success of the company was mainly due to the relationship people had to the brand, reminding them of a voluntary and engaged moment during their childhood.
However, thorough analysis in the first months of collected data show naturally that still many people have only tried the application but never came back to it. Retention of Pokemon Go's customers has therefore been an enormous challenge aside the growing acquisition strategy.

This study aims at tackling this topic. We will first explain who our customers are -activity, life-time value- and see how we can make some different groups (assignment 1 & 2). Then, we will look more at time-oriented frame and see what the most probable reasons for them are not to have come back to the application (assignment 3). We will finally advise some recommendations.

Assignment 1: Creation of a Basetable

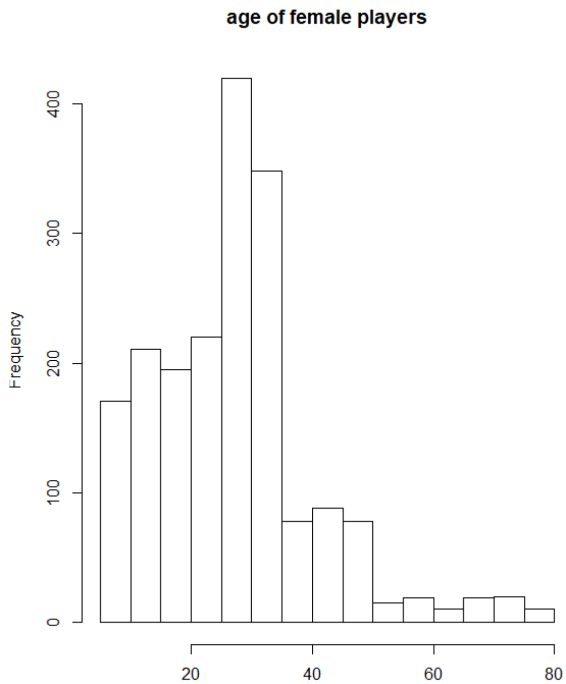
1) Who are the Active customers?

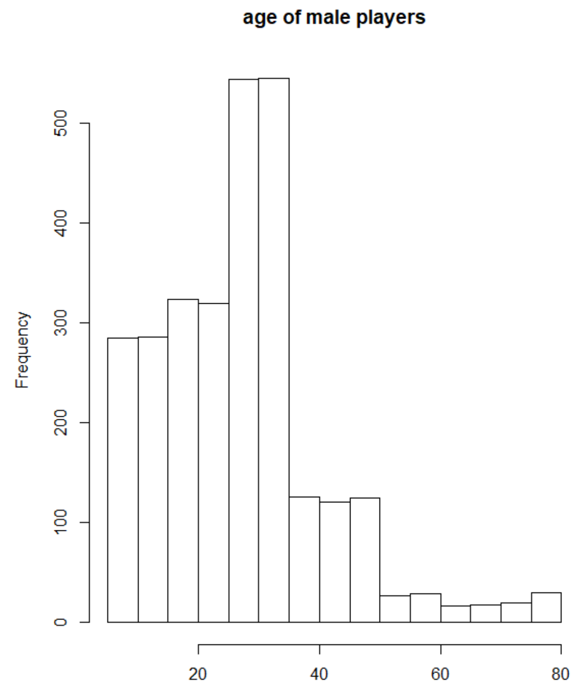
Regarding the fall bonus, the data provided contain 4717 active customers and 20% of them received the fall bonus. The fall bonus has been equally split between men and women (20,46% of the women got the fall bonus and 19,45% of the men). The fall bonus has also been equally split between the 4 customers types.
Thanks to descriptive statistics, we saw that the customers are mostly men: 60% of men and 40% of women. The players are 27,5 years old on average and 75% of them are less than 33 years old.



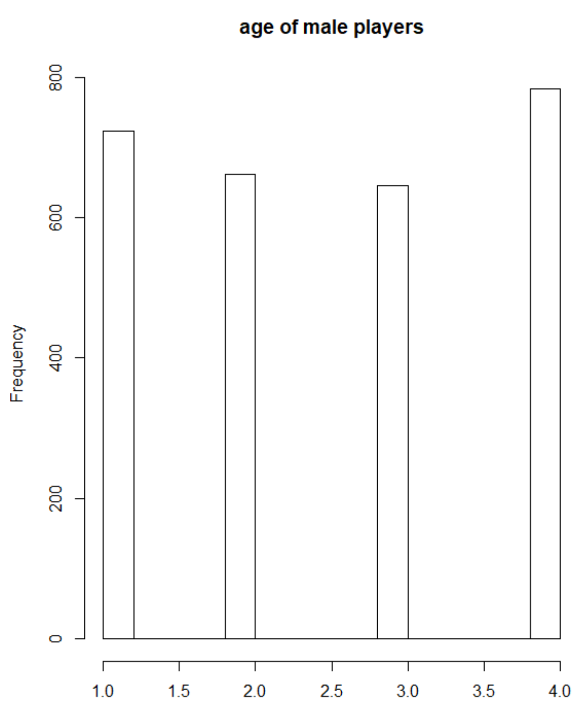

```
     Age
Min.   : 6.00
1st Qu.:18.00
Median :27.00
Mean   :27.58
3rd Qu.:33.00
Max.   :80.00
```

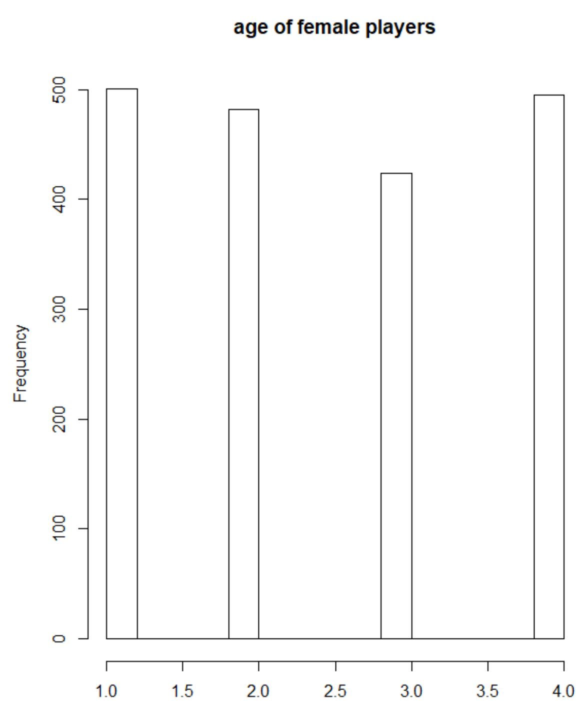
The distribution of the age is slightly different between men and women.





But the distribution of the customer income is almost the same for both men and women.





We have approximately the same distribution for both sexes.

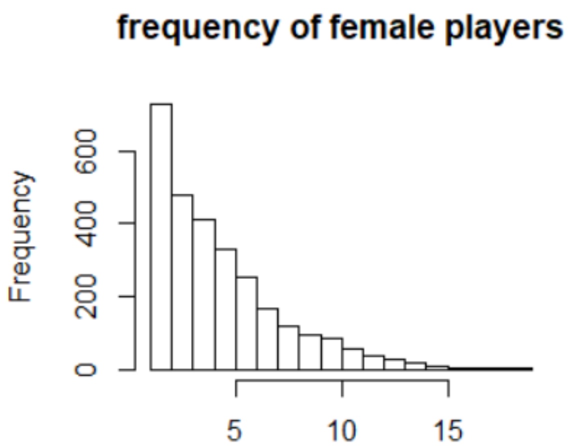
As a result, we learnt that our active customers are rather young. There are more men than women, but they are almost equally split between the different incomes.

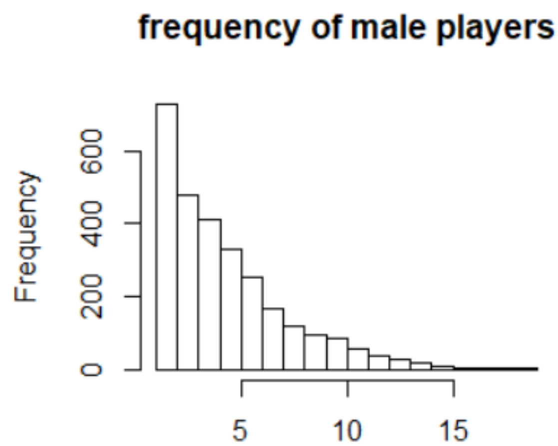
2) What are the the different profiles of your customers?

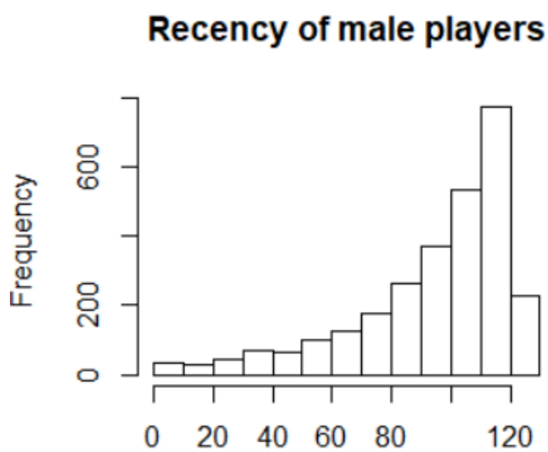
In order to be able to analyze correctly Pokemon Go customers we decided to integrate playing behavior to the RFM - calling it the **Playing value**. Indeed, we noticed that only few people are actually spending money. But to us, even players that are not spending money could be valuable, so we decided to try to evaluate their value as well through this variable. The Playing value is a "potential monetary value" as the players are more likely to buy if they play a lot and it will give us some insights on the customers to target (the best ones, but also the ones we are losing).

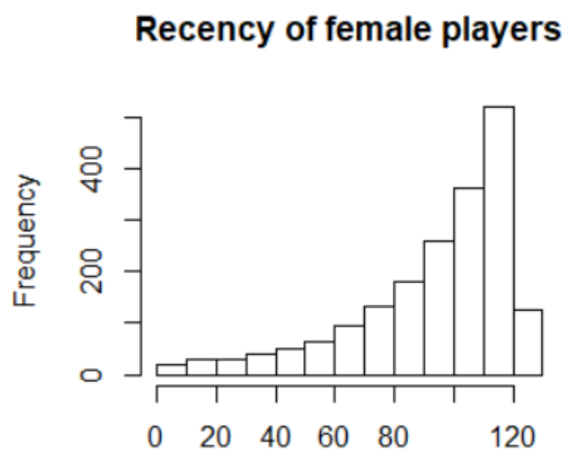
To build a profile of player, we first calculated the correlation between the different variables of our RFM (Frequency, Recency, Monetary value, Play value) and our demographic variables (Age, Sex, Income):

We noticed that indeed men and women have approximately the same behavior in term of Frequency and Recency.

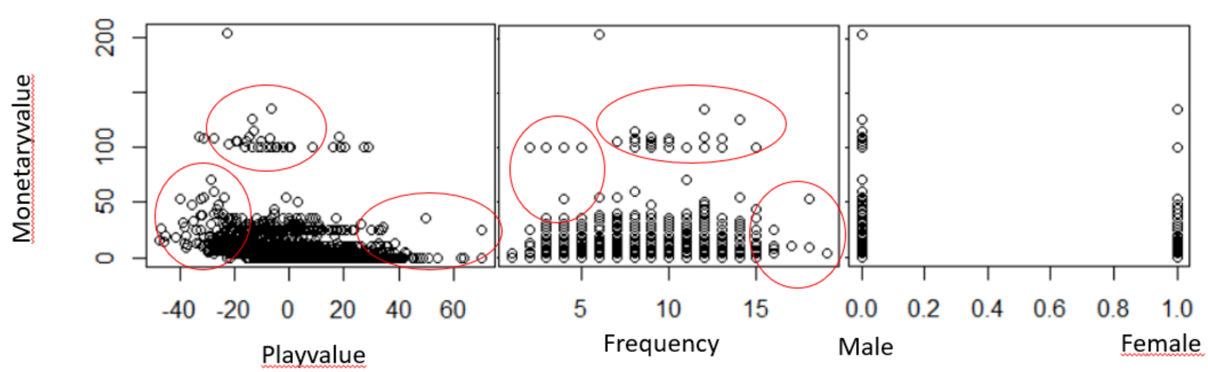








We then focused more on the Monetary Value and tried to identify groups of individuals on this variable.



We noticed that there are 3 groups : one that spend a lot of money but is average players (in term of frequency and play value), one that spend a lot but is a "bad player", and the last one would be the one that doesn't pay but plays a lot.

We first analyzed the characteristics of the “big spender” group (i.e. monetary value > 50).

```
> summary(generalprofil)
   CustomerID     Frequency         Recency        Monetaryvalue       Playvalue              Age             Sex             Income     
 Min.   :   1   Min.   : 1.000   Min.   :  1.00   Min.   :  0.00   Min.   :-47.48998   Min.   : 6.00   Min.   :0.0000   Min.   :1.000  
 1st Qu.:1252   1st Qu.: 2.000   1st Qu.: 82.00   1st Qu.:  0.00   1st Qu.: -6.08194   1st Qu.:18.00   1st Qu.:0.0000   1st Qu.:1.000  
 Median :2506   Median : 4.000   Median :103.00   Median :  0.00   Median :  0.09482   Median :27.00   Median :0.0000   Median :2.000  
 Mean   :2502   Mean   : 4.603   Mean   : 94.52   Mean   :  4.19   Mean   :  0.00000   Mean   :27.58   Mean   :0.4032   Mean   :2.014  
 3rd Qu.:3751   3rd Qu.: 6.000   3rd Qu.:115.00   3rd Qu.:  5.00   3rd Qu.:  5.14269   3rd Qu.:33.00   3rd Qu.:1.0000   3rd Qu.:3.000  
 Max.   :5000   Max.   :19.000   Max.   :122.00   Max.   :203.00   Max.   : 70.51541   Max.   :80.00   Max.   :1.0000   Max.   :3.000  
> summary(generalprofil[generalprofil$Monetaryvalue>50,])
   CustomerID     Frequency         Recency        Monetaryvalue      Playvalue           Age             Sex             Income     
 Min.   : 371   Min.   : 2.000   Min.   : 33.0   Min.   : 53.00   Min.   :-40.153   Min.   : 7.00   Min.   :0.0000   Min.   :1.000  
 1st Qu.:1143   1st Qu.: 6.000   1st Qu.:107.0   1st Qu.:100.00   1st Qu.:-22.234   1st Qu.:20.00   1st Qu.:0.0000   1st Qu.:1.000  
 Median :1963   Median : 8.000   Median :114.0   Median :100.00   Median :-11.478   Median :28.00   Median :0.0000   Median :3.000  
 Mean   :2421   Mean   : 8.659   Mean   :108.5   Mean   : 98.93   Mean   : -8.970   Mean   :30.24   Mean   :0.1951   Mean   :2.244  
 3rd Qu.:3598   3rd Qu.:12.000   3rd Qu.:118.0   3rd Qu.:105.00   3rd Qu.: -1.139   3rd Qu.:34.00   3rd Qu.:0.0000   3rd Qu.:3.000  
 Max.   :4912   Max.   :18.000   Max.   :122.0   Max.   :203.00   Max.   : 28.725   Max.   :75.00   Max.   :1.0000   Max.   :3.000  
```

Men are highly represented in this group (as in there is only 20% of women in this group against 40% in general) and they are not playing very well (they score on average at -9 against 0 in general), but they do play frequently (8,6 on average against 4,6 in general) and finally they earn more money (2,244 on average against 2 in general). There is an improvement to make regarding women spending.

We then focused on the group of “very good players” (i.e. play value > 30).

```
> summary(generalprofil[generalprofil$Playvalue>30,])
   CustomerID     Frequency        Recency       Monetaryvalue      Playvalue          Age             Sex             Income    
 Min.   :  23   Min.   : 2.00   Min.   : 53.0   Min.   : 0.000   Min.   :30.02   Min.   : 6.0   Min.   :0.0000   Min.   :1.00  
 1st Qu.:1778   1st Qu.: 9.00   1st Qu.:104.0   1st Qu.: 0.000   1st Qu.:31.74   1st Qu.:16.0   1st Qu.:0.0000   1st Qu.:1.00  
 Median :2954   Median :10.00   Median :113.0   Median : 0.000   Median :35.60   Median :23.0   Median :0.0000   Median :2.00  
 Mean   :2721   Mean   :10.27   Mean   :108.9   Mean   : 3.567   Mean   :38.34   Mean   :23.4   Mean   :0.3582   Mean   :1.91  
 3rd Qu.:3668   3rd Qu.:12.00   3rd Qu.:118.0   3rd Qu.: 3.000   3rd Qu.:40.47   3rd Qu.:29.0   3rd Qu.:1.0000   3rd Qu.:3.00  
 Max.   :4965   Max.   :16.00   Max.   :122.0   Max.   :35.000   Max.   :70.52   Max.   :55.0   Max.   :1.0000   Max.   :3.00  
> summary(generalprofil)
   CustomerID     Frequency         Recency        Monetaryvalue       Playvalue              Age             Sex             Income     
 Min.   :   1   Min.   : 1.000   Min.   :  1.00   Min.   :  0.00   Min.   :-47.48998   Min.   : 6.00   Min.   :0.0000   Min.   :1.000  
 1st Qu.:1252   1st Qu.: 2.000   1st Qu.: 82.00   1st Qu.:  0.00   1st Qu.: -6.08194   1st Qu.:18.00   1st Qu.:0.0000   1st Qu.:1.000  
 Median :2506   Median : 4.000   Median :103.00   Median :  0.00   Median :  0.09482   Median :27.00   Median :0.0000   Median :2.000  
 Mean   :2502   Mean   : 4.603   Mean   : 94.52   Mean   :  4.19   Mean   :  0.00000   Mean   :27.58   Mean   :0.4032   Mean   :2.014  
 3rd Qu.:3751   3rd Qu.: 6.000   3rd Qu.:115.00   3rd Qu.:  5.00   3rd Qu.:  5.14269   3rd Qu.:33.00   3rd Qu.:1.0000   3rd Qu.:3.000  
 Max.   :5000   Max.   :19.000   Max.   :122.00   Max.   :203.00   Max.   : 70.51541   Max.   :80.00   Max.   :1.0000   Max.   :3.000  
```

They play more frequently than in general (10,27 against 4,6). They are not particularly big spenders and follow almost the proportion of men and women, but they are under the general average in term of incomes (1,91 against 2,014).

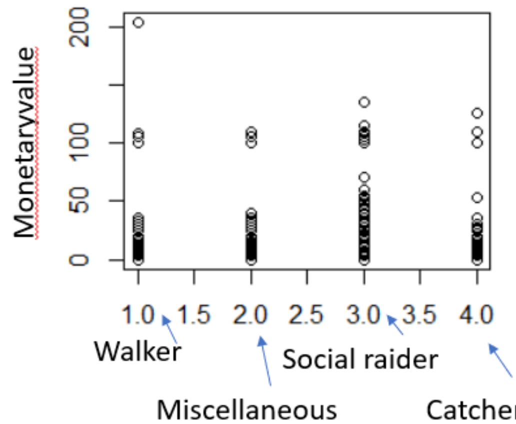
We made a regression on the Monetary Value in order to understand statistically the variables that are significant.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.568e+00  8.346e-01  -1.879  0.06034 .  
CustomerID  -6.064e-05  1.034e-04  -0.587  0.55754    
Frequency    1.458e+00  5.829e-02  25.014  < 2e-16 ***
Recency     -7.071e-03  6.324e-03  -1.118  0.26353    
Playvalue   -3.230e-01  1.363e-02 -23.694  < 2e-16 ***
Age          1.241e-02  1.099e-02   1.129  0.25899    
Sex         -9.000e-01  3.038e-01  -2.962  0.00307 ** 
Income       1.776e-01  1.808e-01   0.983  0.32581    
Type        -1.873e-01  1.324e-01  -1.414  0.15743    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.23 on 4708 degrees of freedom
Multiple R-squared:  0.1985,	Adjusted R-squared:  0.1971 
F-statistic: 145.7 on 8 and 4708 DF,  p-value: < 2.2e-16
```

The variables that are significant are the frequency, the play value and the gender, confirming the results above.

We finally tried to focus on the profile of the players paying the most.



Monetaryvalue	General	Walker	Miscellaneous	Social raider	Catcher
Min.	0.00	0.000	0.000	0.00	0.000
1st Qu.	0.00	0.000	0.000	0.00	0.000
Median	0.00	0.000	0.000	0.00	0.000
Mean	4.19	3.104	2.595	7.67	3.749
3rd Qu.	5.00	3.000	3.000	10.00	5.000
Max.	203.00	203.000	110.000	135.00	125.000

Social raiders are the ones spending the most with a mean much higher than the average. It could be a category to focus on.

3) What's the CLV of your customers?

The customer life time value (or CLV) represents the sum of discounted profits expected on average over the life time of a client. It represents some kind of prediction over historical values. Many assumptions need therefore to be taken for its calculation in our case:

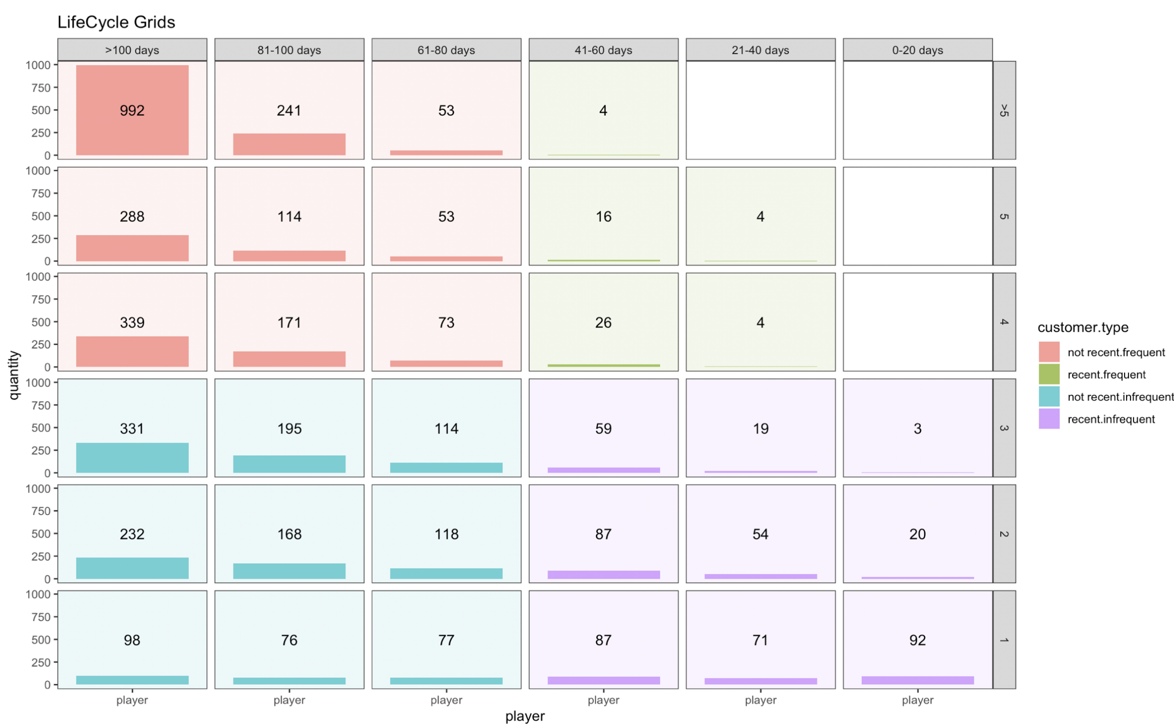
- the margin, or the expected profit of a customer consists in a weighted sum of the monetary value and the potential monetary value (i.e. play value)
- The average life time of a customer, we here suppose it is 12 months
- The retention rate (the user returns to the app at least one time in the time period) is 43% after 1 month (source localytics 2017), and further work would be to have a different value per customer thanks to a machine learning estimation
- 3,75 dollars/customer for the acquisition cost for app install (source mobile gaming apps report 2018)

The majority of customers are not worth the money spent acquiring them, according to our assumptions. The obvious strategy would be then to keep customers who show a positive NPV for the marketing investment and drop the ones who don't.
Naturally, the result would be more positive with a longer life-value of our average customer, higher retention rate and better margin per customer.

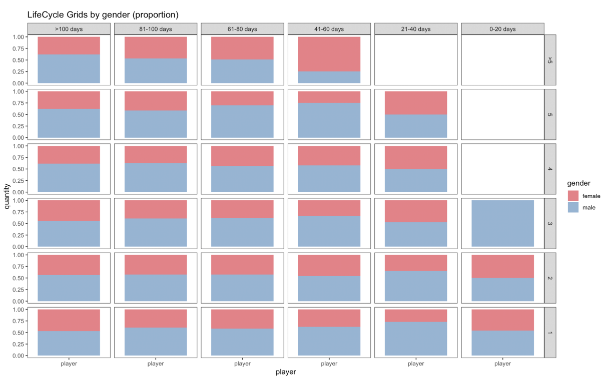
Assignment 2: Lifecycle grids

1) In the summer

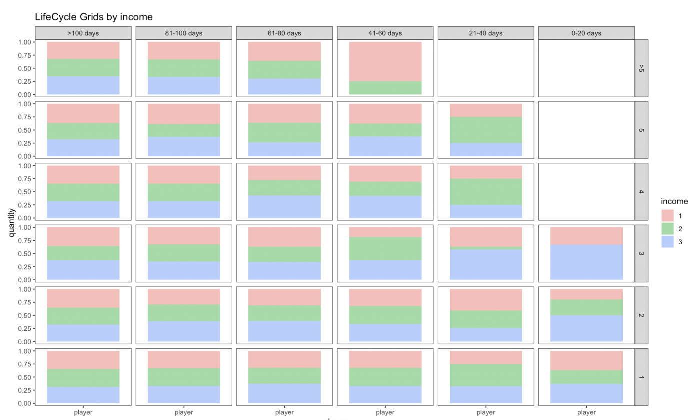
To better understand playing patterns, we first decided to focus on the summer time, when the game play peaks occur. We built LifeCycle Grids on the two following dimensions: the frequency and the recency of play time.



We can clearly observe that the majority of players does not tend to play again very often and, as such, the mean of the recency of play time is around 3 months. The matrix shows a low number of best players (recent and frequent) compared to the number of one-time players (not recent and not frequent) and the number of former best players (not recent but frequent). It shows that the game is not retaining its players, that are more tempted to play massively once as the disproportionate number of players playing more than 6 times in more than 3 months. The number of new players (recent and not frequent) shows that the game is still attractive.

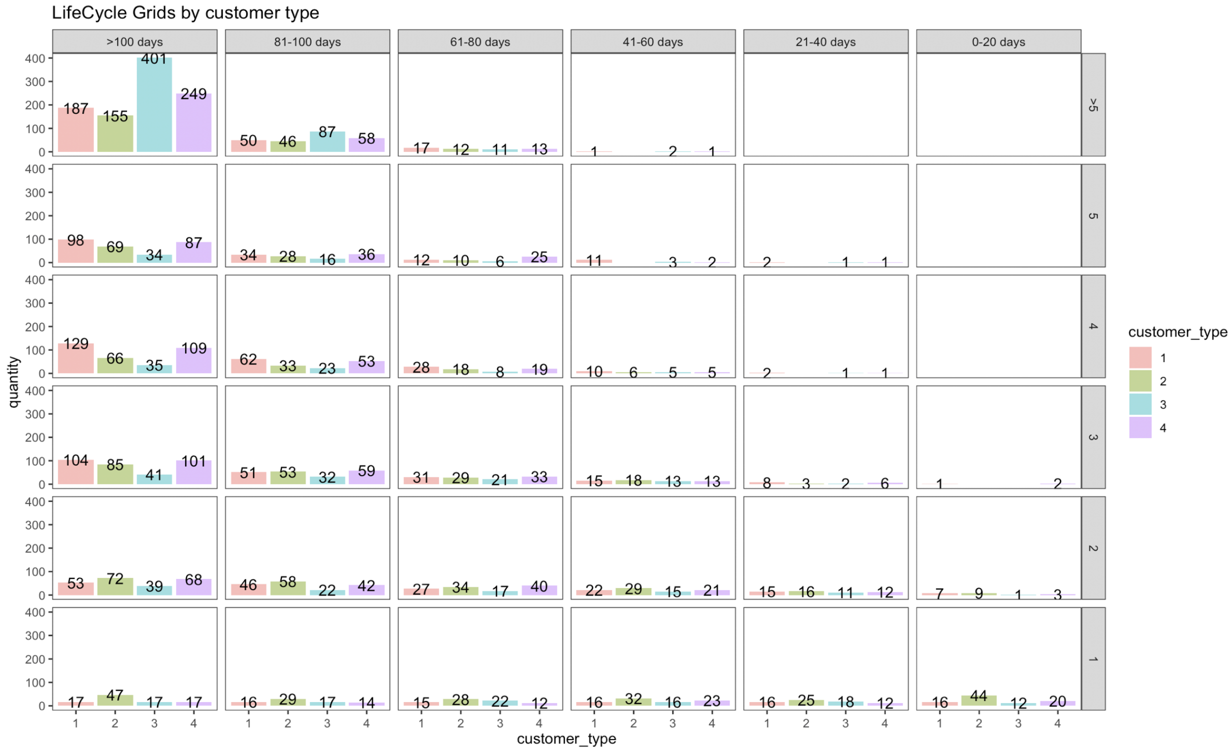


If we only observe the proportion of man and women playing to Pokémon Go, we cannot observe a major trend. This trend is confirmed when we decided to couple this information with the customer type.



When looking at the income of the players, we see that the revenues of the players do not play a major role and that the three categories are equally represented all along the cycle. During the fall, the same conclusions can be made on the gender and the income of the players.

We focused on the customer type to better analyze the segment of players.



The mean of customer type is around 2.5, showing that the players profiles are quite equilibrated, with a slightly higher number of players being walkers and catchers to the expense especially of social raiders. It can be explained by the fact that the initial purpose of the game is not in favor of social raiders, that the gym experience is one of the weaknesses of the game and the limitations of playing with friends.

If we analyze this LifeCycle Grids according to the four categories of players mentioned higher, we can see that the best players are mostly walkers (i.e. people that do more than 3.9 km with the app). This category of players is indeed encouraged to play a lot in order to explore the world around them and to play each time they go out. The recent players as well as the one-time players are mostly miscellaneous players. They are still in the testing phase of the game - no specific profile can be determined. Our former best players are in majority the catchers. They played a lot of time as they probably tried to catch new pokemons but not very often. It can be linked to the fact that the variety of pokemons they encountered did not satisfied them leading to a loss of attractivity of the game. In this category, the social raiders outperform the other type of players when we look at the highest frequency of play. We could try to reactivate this category.

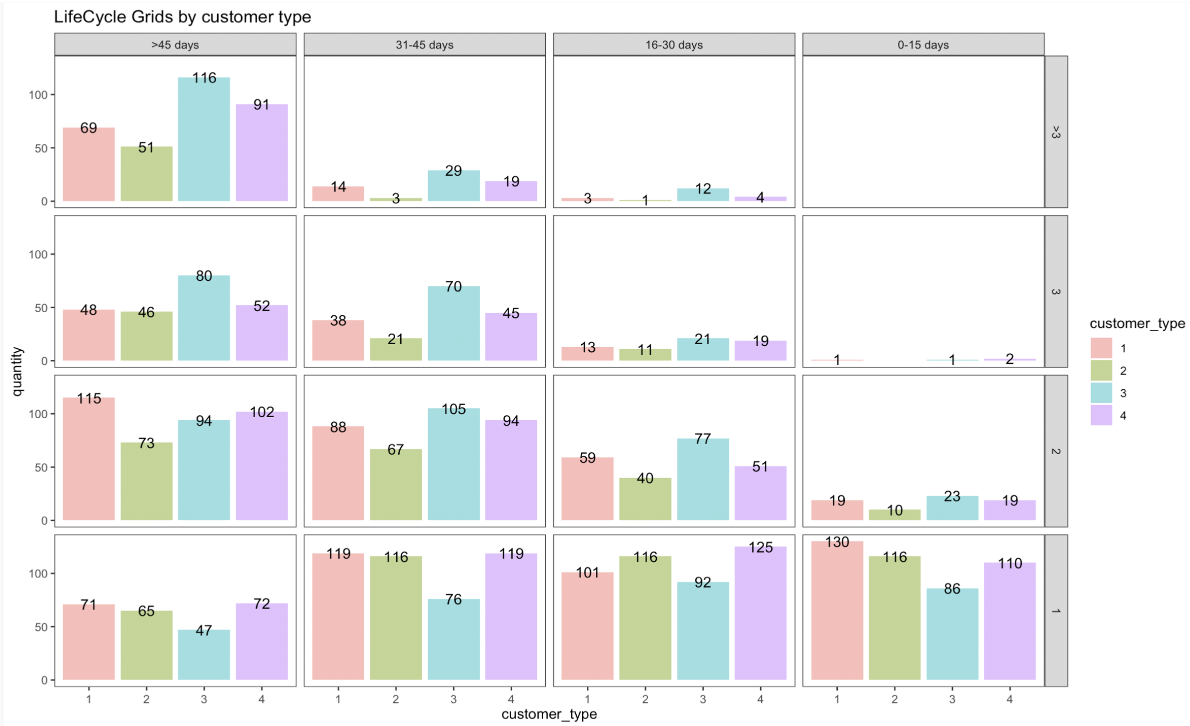
We decided then to build LifeCycle Grids on the two following dimensions: the frequency and the recency of pay. We can clearly observe that people tend to do one or two purchases maximum in general. Our best buyers are very limited (frequent and recent).

2) In the fall

We then focused on the fall time, with the following two variables: frequency and recency of play.

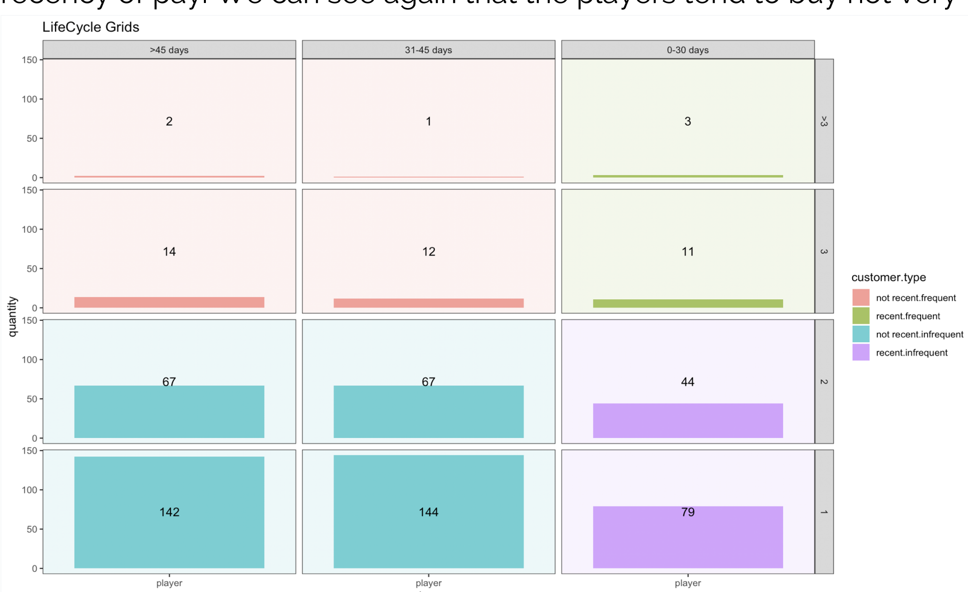
As in summer timer, we have a low number of best players (recent and frequent players) proportionally to the other categories of players. However, the number of best players in the fall is much higher than in the summertime. The number of former best players is also significantly lower in the fall, indicating a better retention in the fall than in the summer. It can be explained by the “fall bonus” but also by the fact that the core players play all year long. The number of first players is also higher in the fall than in the summer. The LifeCycle in the fall is reduced compared to the summer due to the data but also that people tend to play less frequently. It can be caused by the end of the summer holidays.

We decide to focus again on the customer type to better analyze the customer segments.



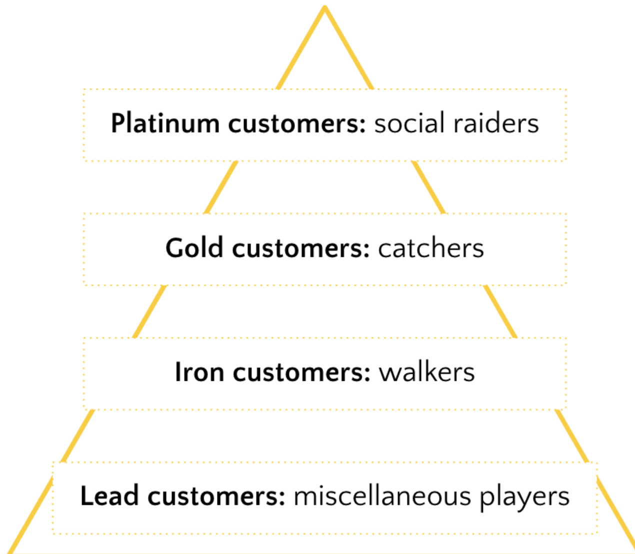
In the fall, there are more players that are social raiders and catchers than walkers. It could indicate that the social raiders are a solid base of customers, confirming the results of the Assignment 1. Indeed, if we look at our best players, social raiders are more represented that the other category of players. However, our former best players are also social raiders indicating that efforts could be made on this category. Compared to the summer, in the fall there are very few miscellaneous players, showing that in the fall we are left with the core of our customers. The one-time players as well as the recent players tend to be walkers.

We focus then on LifeCycle Grids on the two following dimensions: the frequency and the recency of pay. We can see again that the players tend to buy not very frequently.



Takeaways of the assignments 1 and 2

Thanks to the first two assignments, we are able to build the following customer pyramid:



Our platinum customers are without any doubt the social raiders. Our gold customers are the catchers based on the monetary value as well as the results of the LifeCycle Grids on frequency/recency of play in the summer. Indeed, the catchers are placed as former best players and it will be easier to do specific actions to reactivate those players and incentivize them to purchase and play rather than walkers. Finally, our lead customer are the miscellaneous players, they are the under belly of our customer force.

Assignment 3: Churn Analysis

For the churn analysis, we decided to compute and compare 2 levels of churn between summer and fall:

- The financial churn (defined, for a summer player performing microtransactions during summer, as not performing any microtransactions in fall 2018)
- The playing churn (defined, for a summer player, as not playing in fall 2018)

The financial churn rate is 66%, this is based on a 6 months period (summer & fall). The interpretation of this churn rate remains tricky as a quarter period is maybe not the most adapted for gaming apps, as acquisition & retention are very unstable from to month. Nevertheless, more than one customer out of two do not purchase any more after a maximum period of 6 months, which needs to be taken into consideration especially if we compare this financial churn rate to the general playing churn rate.

About the playing churn rate, engagement is naturally a lot more important (73% have at least played once in a maximum period of 3 months after summer ended). If many users have dropped purchasing (66% of purchasers) between summer and fall, a large majority of them have kept playing (73% of overall players). Needless to say, people are more loyal and keen in playing than purchasing.

Factors of the financial churn: we use a linear model (the Logistic Regression) to find the variables that have statistically impacted the churn rate variable.

Good news: Fall Bonus is one of them, meaning that there is a link between the fall bonus package sent at the of the end of summer 2018 and the decrease in retention rate (the marginal effect of the variable is negative as shown below). Indeed, the attribution of the fallbonus would decrease the log odds of having churned by 1.38 and this result is highly significant (p value < 2e-16).
The other significant variable for explaining why customers have churned are the frequency and the play value (which is variable defined as the sum of scaled variables that are linked to the potential of players - i.e. number of pokemons catched, pokemons, gyms, raids, social interactions, distance and duration). This variable we built for our model in the calculation of RFM appears here as being significant to predict churn behavior which gives more weigh to our RFM analysis.

```
Call:
glm(formula = churn ~ ., family = "binomial", data = generalprofil_finchurners)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0581  -1.0653   0.7598   0.8450   1.5712

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.942e+00  3.682e-01   5.275 1.33e-07 ***
CustomerID    -6.244e-05  3.746e-05  -1.667   0.0956 .
Frequency     -4.826e-02  2.076e-02  -2.324   0.0201 *
Recency       -3.117e-03  2.975e-03  -1.048   0.2947
Monetaryvalue  6.929e-03  3.686e-03   1.880   0.0602 .
Playvalue      1.048e-02  4.658e-03   2.249   0.0245 *
Age           -1.878e-03  3.859e-03  -0.487   0.6265
Sex            8.454e-03  1.093e-01   0.077   0.9383
Income        -4.795e-02  6.420e-02  -0.747   0.4551
CustomerType  -3.091e-02  4.857e-02  -0.636   0.5245
fallbonus     -1.385e+00  1.293e-01 -10.713  < 2e-16 ***
```

Factors of the session churn: it is mainly the customer type that has affected the retention rate of customers between summer and fall. As only 22 % have churned, we see again that engagement in playing is the strongest naturally. This means also that these 22% of customers that are leaving the application should be treated differently from the ones that have financially churned. The latter is about monetization of people that play, but former is about not interested any more in the service. Therefore, if the customer type explains partly why they leave, we should maybe conduct further analyses in finding which customer type drives the most of these drop-outs.

Min. 1st Qu. Median Mean 3rd Qu. Max. 0.0000 0.0000 0.0000 0.2759 1.0000 1.0000 Min. 1st Qu. Median Mean 3rd Qu. Max. 0.0000 0.0000 0.0000 0.3934 1.0000 1.0000 Min. 1st Qu. Median Mean 3rd Qu. Max. 0.0000 0.0000 0.0000 0.1506 0.0000 1.0000 Min. 1st Qu. Median Mean 3rd Qu. Max. 0.0000 0.0000 0.0000 0.2666 1.0000 1.0000	• Churn rate for Customer Type 1 • Churn rate for Customer Type 2 • Churn rate for Customer Type 3 • Churn rate for Customer Type 4

As expected, customers with type 2 which have no specific preferences for a particular aspect of the game play are the biggest churners: 40% against 22% overall.

The results of this churn analysis provide of few interesting insights for our final recommendation.

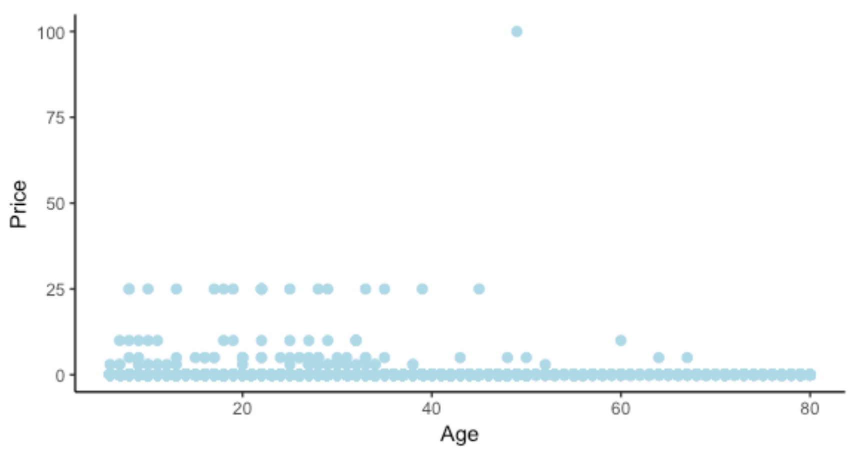
Recommendations

Our recommendations follow four major steps of the customer cycle:

- Acquisition of players
- Boost the transactions
- Retain key players
- Prevent churn

1) Acquisition of players

Considering that the Pokemon Go game is already attracting customers thanks to the awareness of the brand, the aim here is to attract loyal customers. That's why our first recommendation is to create a referral program targeted at young customers. Indeed, it will not only increase the retention of this category but also, we are strongly convinced of the power of word-of-mouth among the youngster that can attract family and friends. The referral program could for example only work after 10 hours of play in a row and will allow the player to win special items.



Our second recommendation is to expand in a new product category by designing an external battery to be attached to the phone to give more visibility in the streets. A similar action has been taken with the Pokemon Go Plus, a connected bracelet that notifies you when there is a Pokemon nearby.
The cross-channel strategy is already a success in Europe with the Pokemon games on the Switch and could be expanded like in Asia with the amusement parks to be constructed and the Pokemon Centers. A collaborative action could be taken on those geographical sites with special events in the Pokemon Go Games with rare pokemons.

2) Boost the transactions

If we focus on our gold customers, our recommendation to retain them and encourage them make more purchases naturally, would be to increase the variety of pokemons you can encounter. One of the flaws underlined by the Pokemon Go community is indeed the fact that the same monsters keep showing up.
Regarding our iron customers, our recommendation is to create more items in the shop for them. Indeed, the majority of the items you can purchase in the shop are for walkers and catchers. An example would be a special item you can buy after a certain distance.

In the Assignment 1, we discovered that women buy less in proportion to men. Another recommendation would be then to push transactions for women during the game to help them progress more rapidly during the quests for example and encourage them to purchase.
From the Assignment 2, we saw that the frequency of the transactions is low. A recommendation to boost this frequency could be to develop compatible items to unlock with a transaction.

3) Retain key players

Our key players are the social raiders. They value the social aspects of the game that can be improved.
Our first recommendation would be to allow the trading of monsters between the players, an old feature that will thrill the fans. It would also create synergies between the catchers and the social raiders and improve these two categories of players.
Along with the trading of monsters, the whole gym experience needs to be redesigned as it is considered in general as a failure. In order to do so, we recommend allowing player-on-player battles outside the gyms and expand the basic gym structure by adding badges, leagues, type constraints... to make it more rewarding. A key shift could be also to change the team mechanism. Today, there are only three teams worldwide, but it would make more sense to enable the creation of smaller team to make the pokemon go experience more local to enhance engagement.
Another recommendation would be to increase inbound marketing and content creation by developing a specific forum on the Pokemon Go website for the players to communicate and exchange more easily.

Finally, we recommend betting on the customer knowledge value by giving a dedicated space for the players to indicate local events either to place ephemeral pokestops or special pokemons.

4) Prevent churn

As concluded in the Assignment 3, the fall bonus had a positive impact on the churn and could thus be reiterated the following year to prevent financial churn.
Regarding the churn of session, we recommend launching challenges for the miscellaneous players and tailor-made touch points depending on the player type.

Conclusion

The Pokemon Go game is not churning as fast as some would think and benefits from the huge fanbase of the Pokemon world. Our recommendations rely on the feeling of warm familiarity of the whole Pokemon experience, that has followed the players since their childhood.
However, it would be interesting to observe in the coming years if there is an effect of seasonality on the data. To improve the recommendations, a better tracking of personal information of the payers would be useful to determine their profile. The reporting of the geolocalisation could also give a useful insight on the player profiles.