# An analysis of data from the mywellness system at Contours Gym, The Stour Centre

## Chances of sustaining a new exercise regime and the effect of the first month

Jonathan Hobbs

Technogym's mywellness system is a platform for (among other features) tracking exercise. All exercise in and outside of the gym, as well as daily movement, can be recorded either manually using the app or website or automatically using the TGS key or connected devices. All of this exercise is converted into MOVEs, which are a standardised points system.

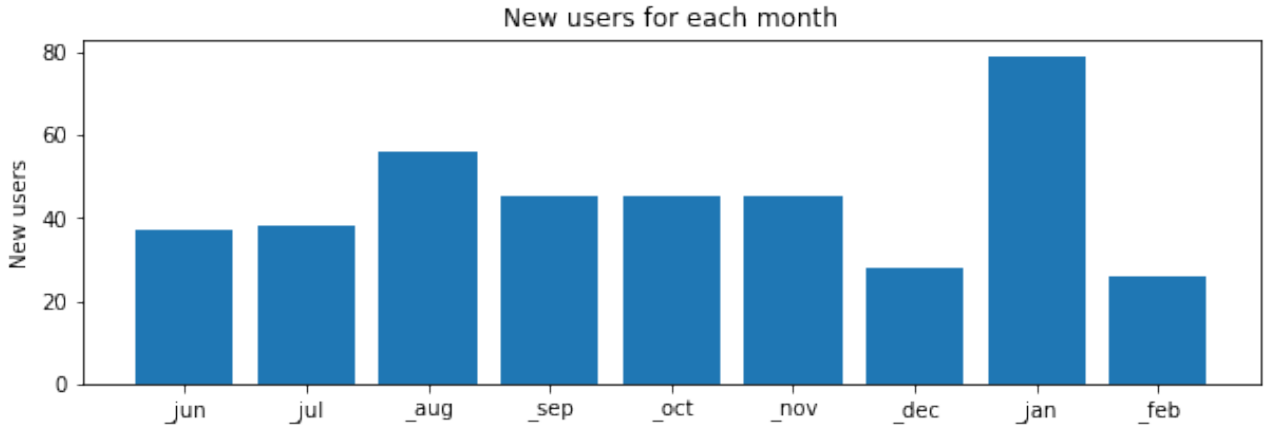
Although the mywellness system is optional, continuous recording of MOVEs is an indicator of whether members are sustaining their exercise regime or not. It is possible that members will drop off of the mywellness system but continue to exercise and conversely that members stop exercising meaningfully in the gym but continue to automatically record MOVEs from movement tracking. Despite these caveats it is an insightful proxy measure for whether members are continuing to exercise or not.

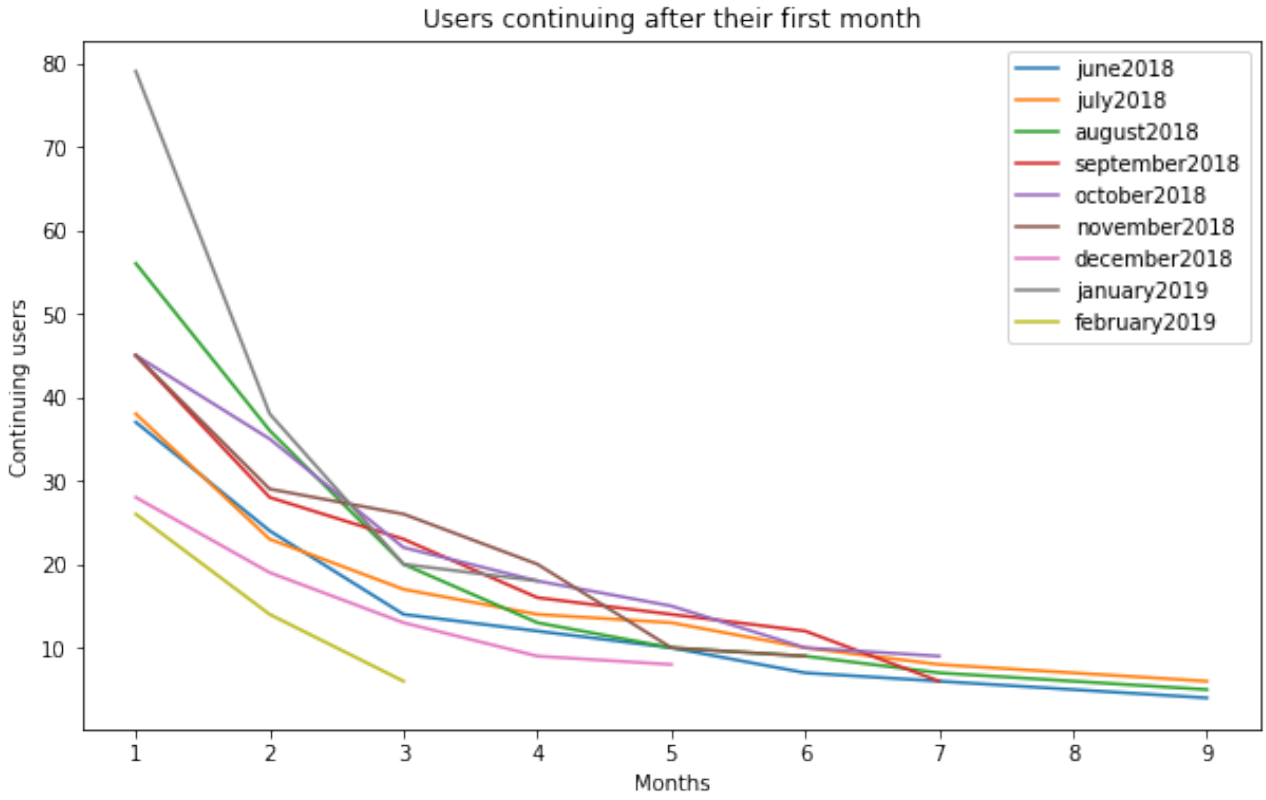
This analysis will focus on the effect of the total MOVEs and number of days recorded in the first month of use. The idea is to establish a helpful target for new members who want to establish and maintain a regular training regime.

The data comes from monthly 'reports' that are generated by mywellness. These monthly reports were then combined and manipulated to find when new users joined, what they did in their first month and how long they continued using the system. Any personal information, such as names or dates of birth, was removed.

This was done using a high-level programming language called Python and its additional libraries to perform calculations, data interpretation, machine learning and graph plotting.

Firstly, it is interesting to take a look at this information.



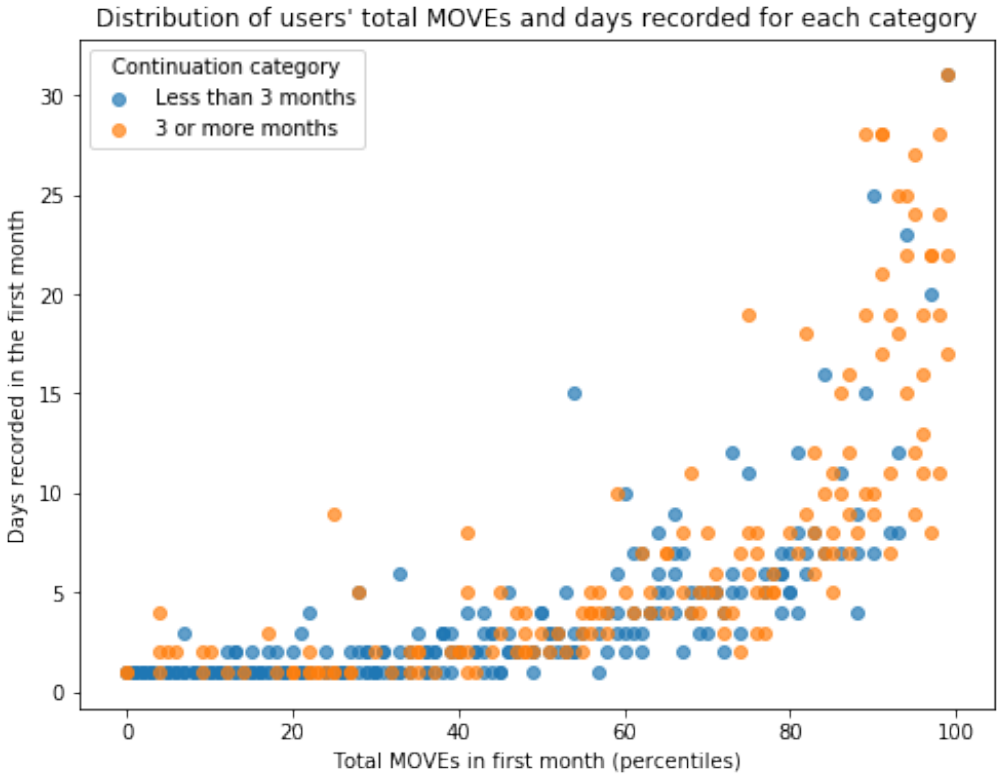



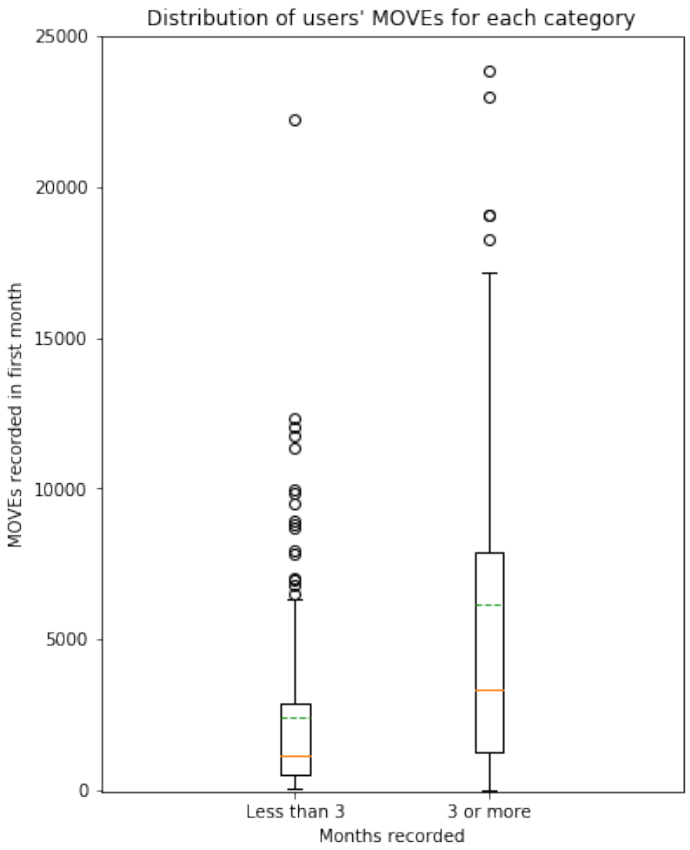

From June to February, the highest number of new starters are in January and the lowest in December and February. Users tend to drop-off in larger numbers within the first few months of use, after which the decline slows. None of this will come as a surprise to anyone who works in the fitness industry.

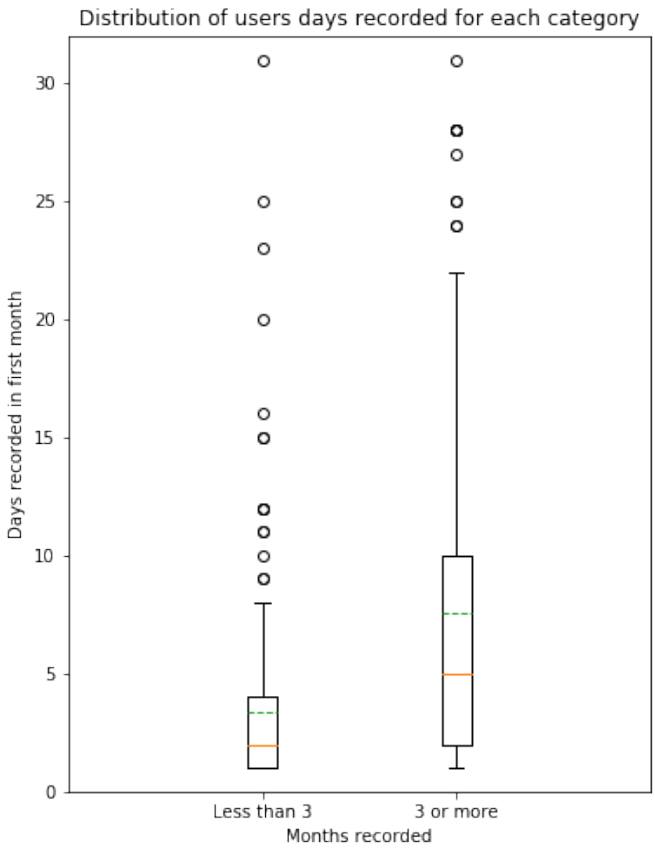
Interestingly, user numbers from most starting months reach about the same level after three months, between 15 and 30. This suggests there is a base number of users who will continue. Above this, any additional new users from January, for instance, tend to drop out quickly. This is perhaps contrary to the pervasive thinking because it implies that the January rush has little effect on long term user numbers. The sample sizes for each month are small, so it is not possible to make this assertion confidently, but it warrants further investigation.

The main purpose of this analysis is to look at how behaviour in the first month affects whether a person will continue exercising or not. In particular, it will look at the number of MOVEs collected, which measures the amount of exercise, and the number of training days recorded. It groups new users into two groups – those who continue for less than three months and those who do 3 months or more.

Evidently, there is a wide spread across each variable in both categories, but it is generally higher for those who continue. It can also be asserted with more than 99.9% confidence that the two groups are distinct in terms of each variable.

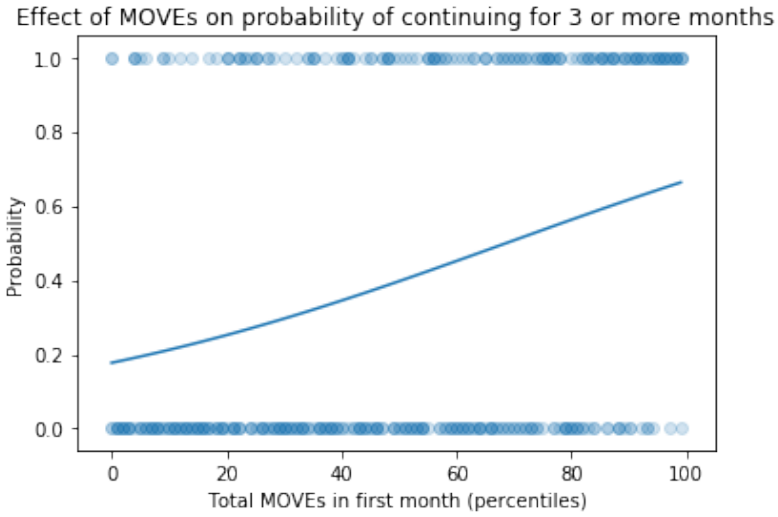


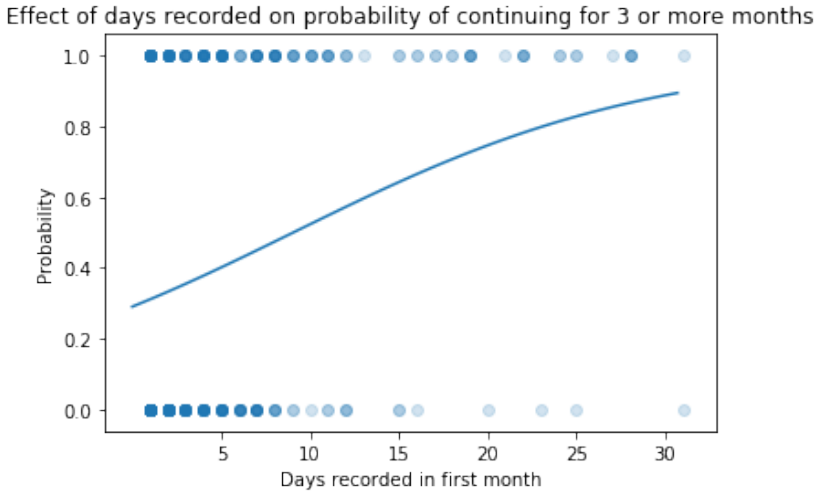



It is clear that more MOVEs and more days are associated with continuing, but the next step is to quantify this. Five different models were made to attempt this. The first two deal with one variable at a time, while the other three consider both variables concurrently.

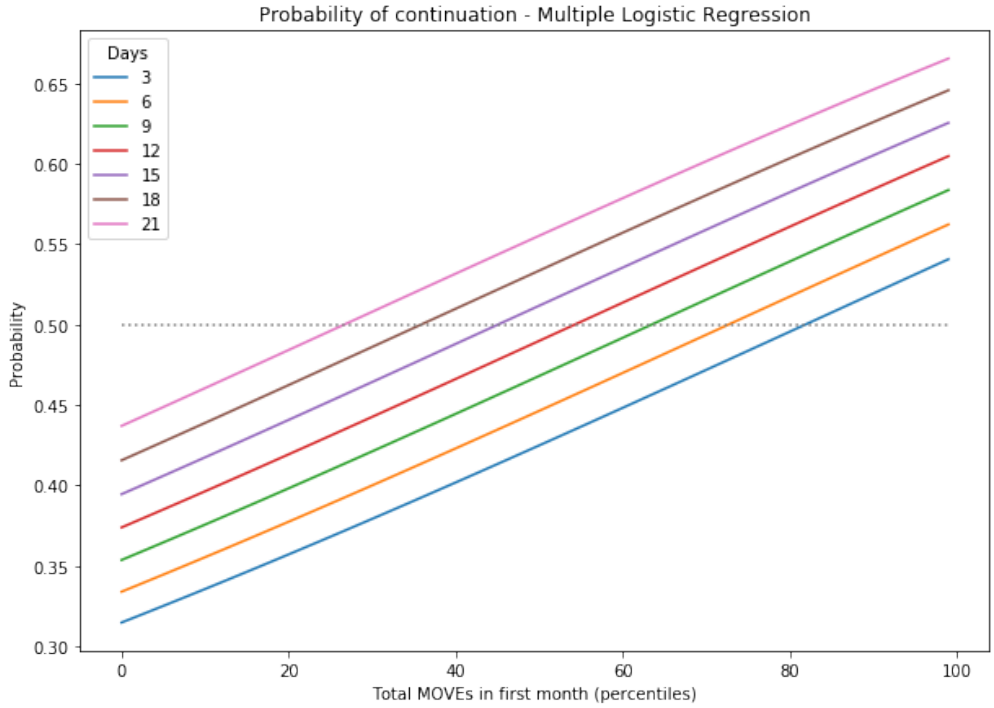
Logistic regression for MOVEs only. $R^2$ of 0.677 (The model explains 67.7% of the variance.)



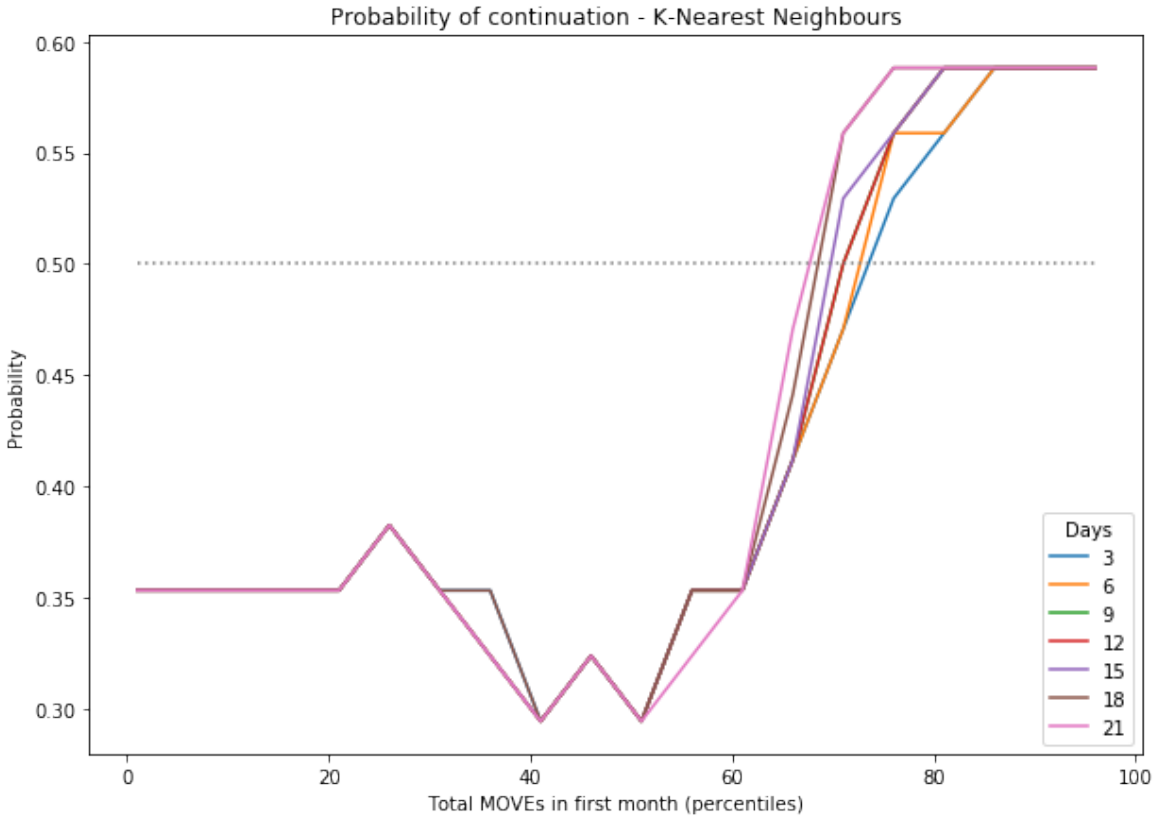

Logistic regression for days only. $R^2$ of 0.669



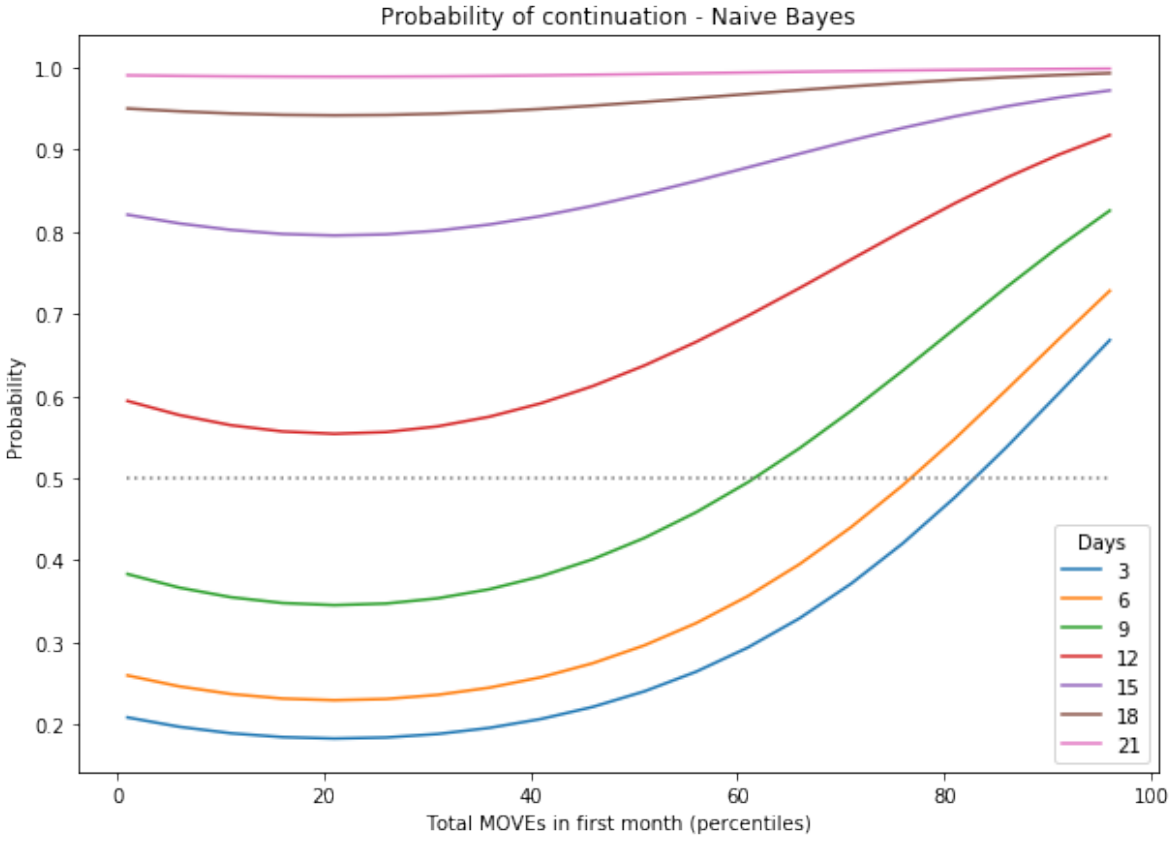

Multiple logistic regression. $R^2$ of 0.391



K-Nearest Neighbours algorithm. $R^2$ of 0.678.




Naive Bayes equation. $R^2$ of 0.675



When reading the multiple-variable graphs, each line represents a specified number of days and then shows the probability of continuing depending on the number of MOVEs. For example, the green line means that a user recorded 6 days and looking at the 60$^{th}$ percentile for MOVEs shows a probability of around 0.5. MOVEs are converted to percentiles so that they are not distorted by outliers where users have recorded a very high number – for example, the 60$^{th}$ percentile means it is higher than 60% of all users.

These graphs show some interesting features of each model, which can be used to select the best one. Of the multiple-variable models, the multiple logistic regression is unreasonably flexible. In reality, it is not likely that a user could achieve a very high number of MOVEs but a very low number of days, so the trade off in this model is meaningless. The K-Nearest Neighbours model, conversely, is too rigid as it is almost only affected by the number of MOVEs. The Naive Bayes model, however, shows a more realistic relationship between MOVEs and days, where at the extremes of one the other is less important (ie. there is a high probability for 12 days regardless of MOVEs), but in the middle ground there is a trade off. There is an unexpected downward curve at the low end of MOVEs, which is probably explained by the wide spread of data. Overall, it is not too flexible like the first model nor too rigid like the second so, according to Goldilocks' judgement, it is 'just right'.

The final question that arises is which pair of MOVEs and days to select. The loop below shows a list of possible combinations that give a probability of approximately 50% using Naive Bayes. Since there is a positive correlation between MOVEs and days, it is unlikely that someone could sit at opposite of extremes (high MOVEs and few days or low MOVEs and many days) so a middle value is more sensible.

The values selected were the 69$^{th}$ percentile of MOVEs (which is **3578 MOVEs**) and **8 days**. Using the Naive Bayes model, this gives a probability of continuing of 50.8%.

The multiple logistic regression gives 52.0% and single logistic regression for MOVEs gives 50.1%, though the K-Nearest Neighbours model gives 44.1% and the logistic regression for days gives 47.4%. While they do not all agree, this is adequate to support the Naive Bayes model, in addition to its 67.5% accuracy.

A final exercise is to illustrate how gaining more MOVEs and days in the first month can increase chances of continuing. As an example, 5000 MOVEs and 10 days gives a probability of 71.2% that the user will continue for at least 3 months.

This analysis has demonstrated positive independent and interdependent relationships between total MOVEs and days recorded in the first month and the likelihood of continuing to use the system for 3 months or more. Given what we know about establishing exercise regimes, this is unsurprising as it requires a considerable impetus and commitment from the start. Motivation is usually highest when starting out, so if it is not possible to capitalise on this then it is difficult to continue as motivation starts to decline.

The Naive Bayes classifier model is able to predict the chance of a new user continuing, given their first month MOVEs and days, with a 67.5% accuracy. From this, a target of 3578 MOVEs and 8 days was found, that gives a new user a 50% chance of continuing. In other words, if they achieve more than this, they are more likely to continue than not. A draft communication to new members (which could be send automatically using mywellness) reads as follows:

*"What you do in your first month has a big effect on your chances of successfully sustaining a new exercise regime. To give you the best chance, you should set some targets for this first month. As a minimum, we suggest a target of 3600 MOVEs over at least 8 days. If you can achieve this, you will have a much better chance of keeping it going.*

*“Our research has revealed that users who hit 3578 MOVEs and exercise on 8 days in their first month have a 50% chance of continuing for at least 3 months. Gaining 5000 MOVEs and doing 10 days in the first month gives new users a 71% likelihood of continuing with their exercise regime.”*

If members can be helped to continue regularly using the facilities, this will make them more likely to maintain their memberships, rather than cancelling. Setting reasonable, informed goals can help with this. A gym that helps its members to stick to their regime and achieve their goals sets itself apart from other companies who might only appear interested in short-term memberships. In short, it shows that the staff care about the members.