FEATURESPACE REPORT: SECONDARY ANALYSIS OF MACHINES DATA

PREPARED FOR THE RESPONSIBLE GAMBLING TRUST

Authors:
David Excell, dave@featurespace.co.uk
Piotr Grudzien, piotr.grudzien@featurespace.co.uk

FEATURESPACE

Table of Contents

Executive summary

A consortium of NatCen Social Research, Featurespace, Geofutures and RTI International conducted the Responsible Gambling Trust's (RGT) Machines Research Programme in 2014. A number of useful datasets were created for this project and Featurespace was asked to answer the following research questions in relation to this data:

- RQ5: *Can the range of linked data set variables be examined through a process of 'reverse engineering' to explore whether any other variables might play a useful role within the development of algorithms?*
- RQ7: *What are the differences in demographics between B2/B3 players? What else can we learn about players' transitions between B2 and B3 content?*
- RQ8: *What further descriptive data can be extracted about the £100 stake?*
- RQ9: *What are the differences in behaviour when players are spending wins vs loading their own new money into the machine?*

In this report we have answered each of these questions and presented the results as four independent chapters. The research has been conducted from what the data has told us from the 4,000 players who were surveyed as part of the 2014 research project. These results are not intended to inform a general understanding of the extent of problem gambling against different factors of gaming machine activity as the survey population biases have not been factored into the presented figures.

However, from the research we have been able to identify the following key findings:

- The accuracy of the problem gambler identification model developed in the 2014 research has been improved. This was principally achieved by including a new marker of harm that measured the diversity on money loaded and money spent by the player.
- The most distinct identifiers of problem gamblers are their chaotic behaviours and the fact that on average they are more successful when playing (they win more often, have higher return rates and more often have winnings to spend).
- Transitions between B2 and B3 bets are not useful when it comes to differentiating between problem and non-problem gamblers.
- Players who place £100 bets are distributed uniformly across problem and non-problem gamblers within the surveyed data. However, players with 100 or more £100 stakes are more likely to be problem gamblers within the data set.
- A typical £100 stake scenario is one where players place the maximum bet several times during a session, it is rarely an isolated single event. £100 stakes happen very rarely at initial stages of sessions and become more common at later stages.
- Variable and intensive activity at early stages of sessions often leads to £100 being staked later.
- When it comes to the differences between playing with winnings and with the player's own money, in the former case players tend to bet higher amounts of money and withdraw money more often. In the latter case on the other hand, players are loading money more often and spending more as a percentage of the balance.

Chapter 1: Research Question 5

Can the range of linked data set variables be examined through a process of 'reverse engineering' to explore whether any other variables might play a useful role within the development of algorithms?

High-Level Findings

Introduction

The purpose of the chapter is to present and discuss the results found when answering research question 5. They concern detailed data on player activity gathered by gambling machines across the United Kingdom. An additional data set was used based on 4,000 players who took part in the survey from the Gambling Machines research project funded by the RGT in 2014. It contained 9 questions from the Problem Gambling Severity Index based on which each player was labelled as a 'problem' or 'non-problem' gambler. The aim was to identify features and aspects of behaviour that are **typical of problem gamblers**. They can be later used as 'markers of harm' for identifying players.

The research was based on the concept of *reverse engineering*. The idea is to considerably expand the scope of features and aspects of player behaviour that are being analysed which are then used as an input to the model. Then, once the model is trained, different techniques may be used to reveal what features best **differentiate between problem and non-problem gamblers**. Results are initially presented from the technical point of view in terms of how discriminatory (informative) they are. The conclusions are then compiled into a high-level description of what types of behaviour are characteristic of problem gamblers.

Findings and insights

The key findings found when answering this research question have been:

- The most informative aspects of player behaviour: **diversity**[1] in terms of money spent (useful by itself[2]) and money loaded (informative in conjunction with other features).
- Behaviour of problem gamblers is more **chaotic**[3]. Diversity measures indicate that they are less consistent with their choice of stake level or amounts of money loaded.
- Overall, problem gamblers do not tend to play more.
- Problem gamblers do not tend to have higher total net losses.
- Sessions of problem gamblers tend to be longer and their durations vary more.
- Problem gamblers have more periods of winning money faster than spending it.
- Problem gamblers play with winnings rather than with their own money more often.

---

[1] When looking at variables describing player behaviour, for example money spent, there are various ways one can aggregate them to obtain a quantitative measure of a certain aspect of player behaviour. If the aspect of interest is **diversity**, measures such as standard deviation or number of distinct levels can be used. They describe how diverse player's behaviour is – whether they tend to consistently stick to one stake value or change it frequently and by a significant margin.

[2] Out of the 3988 players who took part in the survey, 951 (23.8%) are problem gamblers. One of the most discriminatory features is the daily average number of different stakes wagered. For example, setting the threshold at 18 divides the players in two groups: higher than 18 – 452 players 44.0% of whom are problem gamblers; lower than 18 – 3532 players 78.8% of whom are non-problem gamblers. Even though the feature has one of the best discriminatory capabilities, simple thresholding does not perform well in terms of differentiating between problem and non-problem gamblers.

[3] What is meant by **chaotic** behaviour is directly linked to the concept of diversity. A player will be referred to as chaotic if their stake values, amounts of money loaded, etc. are not consistent but rather change frequently and by a significant margin.

- Problem gamblers tend to win more often and their average returns on money spent are higher.
- Problem gamblers have higher one-off losses.

The following characteristics of the data were observed:

- Diversity features (number of different stake levels, how spread out session averages of money spent are, etc.) are the most informative (chaotic behaviour of problem gamblers).
- There are features which are significantly better than others at recognising problem gamblers. There are, however, many aspects of behaviour which should intuitively be characteristic of problem gamblers but are in fact neutral (how much one plays, how much one loses).
- There is little to be learned from manually viewing particular sessions. A more **high-level** view is necessary to be able to differentiate between problem and non-problem gamblers.
- Overall, problem gamblers seem to be **more successful** when playing (they win more often, have higher return rates and more often have winnings to spend)

The following insights were observed when reconstructing the predictive model:

- Extraction of new features and subsequent selection of their most effective subsets has led to an improved performance of the problem gambler classifier: AUC score uplift from the average of 0.6651 (baseline result for the old model definition[4]) for the old 99-feature classifier to the average of 0.**6864** for the new 200-feature classifier. Another interpretation of the uplift is in terms of true positive rate from 21.92% to $\mathbf{24.35}\%$ at a false positive rate of 10% or from 57.15% to $\mathbf{61.90}\%$ at a fixed false positive rate of 30%.
- Valuable features can differentiate well between problem and non-problem gamblers by themselves but also describe diverse aspects of player activity (It is informative to follow a practical example[5])
- A well selected smaller number of features can give comparable performance to large feature sets

Recommendations for future work

The question was very open-ended which has led to many insights and ideas which could be implemented outside of its scope. Most importantly, the findings can be used to improve the performance of the previously developed problem gambler classifier.

Conclusion

The point of the question was to use the process of *reverse engineering* to discover new features which might play a useful role within the development of algorithms. They proved to be the **diversity** ones based on money loaded and money spent by the player. It has been shown how the performance of the classifier in terms of true positive rates and area under curve can be improved (from AUC of 0.67 up to 0.69) by inclusion of extra features and by appropriately selecting among them.

[4] Note that the AUC figure reported in our report published in 2014 was 0.7. The number reported in this report is lower due to a difference in how we have split the data for model 'training' and model 'testing'.

[5] The number of daily different stakes wagered is a good feature – tends to be higher for 'chaotic' problem gamblers. Some other features take on very similar values for both player groups, however it is still useful to include both in the analysis. As an example, while players who on average put higher percentages of their balance at stake might be evenly spread across problem and non-problem gamblers. However those who do that AND choose their stakes chaotically might constitute a relatively homogeneous problem gambler group. Therefore, a feature that is not very useful by itself might work well in conjunction with another one.

The analysis of the model built has led to insights as to what makes a good classifier and what behavioural patterns are indicative of problem gamblers. The most distinct ones are the **chaotic** behaviour of problem gamblers and the fact that on average they seem **more successful** when playing (they win more often, have higher return rates and more often have winnings to spend). Their stakes or amounts of money spent and loaded take on more different values which span wider ranges of numbers. They also win slightly more often as well as have slightly higher average return rates. These general characteristics of problem gamblers have been described in the report along with typical behavioural patterns that can be used as 'markers of harm'.

Reverse Engineering Approach

The approach to answering this research question took the following sets:

1. Feature generation based on:
    a. Previously used 99 features
    b. New ideas
    c. Analysis of behaviour of individual problem gamblers leading to ideas for creation of new features
2. Performance testing for different classification methods
3. Finding the most discriminative features (best at distinguishing between problem and non-problem gamblers)
    a. Selection of the 60 most important features based on the classifiers' feature importance metric
    b. The rest of the experiments aimed at evaluating features' discriminative capabilities were conducted using Logistic Regression as it is a deterministic algorithm which produces more robust performance comparisons
        i. Measuring features' individual discriminative capabilities
        ii. Grouping into classes of highly correlated features
        iii. Manual selection and testing of a small subset of most discriminative features
        iv. Validation by means of a combination-based feature selection scheme
4. Conclusions
    a. Highest true positive rates achieved as a function of the number and types of features
    b. Most important features – what aspects of activity are most useful at recognising problem gamblers?
    c. What makes a subset of features exceptionally discriminative?
    d. How do problem gamblers behave?

Feature Engineering

The process of feature engineering was first of all based on looking at the 99 features that were previously developed. Some new features were added, including those based on the new 'shop visit' based **sessionalisation scheme** and the idea of playing with WIN money (winnings) as opposed to OWN money. Both of these concepts are explained in more detail in Chapter 4.

Manual reviewing of problem and non-problem gambler sessions

Additionally, the behaviour of particular problem gamblers was manually reviewed. This has led to ideas for the following features:

- Percentage of stakes when the player is using money they have won compared to their own money (*WinOwnFrac*)

- The number of times the player switches from playing with their OWN or WIN money, or vice-versa (*Typeswitch*)
- Comparing the rate at which the player loads money into the machine to the rate at which they spend (*LoadFTSpend)*
- Comparing the rate at which the player wins money to the rate at which they spend (*WinFTSpend,*)
- Comparing the rate at which the player wins money to the rate at which they load money into the machine (*WinFTLoad*)

Manual reviewing of sessions was mainly useful as a source of inspiration when constructing features – there are no striking differences that can be spotted when examining a sample of problem and non-problem gambler sessions. It has, for example, been concluded that the proportion of events played with OWN and WIN money could be useful at differentiating between problem and non-problem gamblers. It was not clear, however, whether it is the problem or non-problem gamblers that are more often in the WIN or OWN area. It has indeed turned out that looking at the proportion of time when the player is spending their own money compared to the win money improves classification accuracy – an example of such a feature is 46, *SesAvgSesAvgWinOwnFrac*. This is in spite of it being only slightly positively correlated with the *PG score* (which indeed means that it is the problem gamblers that are on average more likely to be playing with WIN money).

In order to discover general and easily interpretable differences between problem and non-problem gamblers a more high-level approach is necessary. This is provided in the sections to follow by looking at features most useful for recognising problem gamblers as indicated by classification algorithms. A different approach is presented in the Feature correlations with the PG score – high level conclusions section which simply looks at correlations between various metrics and the PGSCORE. This complements the results by providing simple analyses similar to the one in the paragraph above.

Feature themes

The difference between feature themes and features is that the latter is created from the former by using an aggregation and a segmentation. As a result, there are several features corresponding to one feature theme.

Variable	Description
SesCount	The number of sessions
MoneyLoaded	Money loaded (event types: Note in, Coin in, Ticket in, Counter in)
MoneySpent	Money spent (event type: Play)
MoneyWon	Money won (event type: Win)
MoneySpentLevels	Implemented as the count of *MoneySpent* levels - Stake levels (the number of distinct amounts of money spent on bets)
ProbOfWin[6]	Probability of winning (the number of Win events as a fraction of Play events)

[6] The probability of a win for a particular game is fixed but what this parameter says is what proportion of games by the player ended in a win. Its higher value could indicate that the player was often lucky but most likely it means that they have on average been opting for safer games (e.g. 50/50 red/black roulette bet rather than a 1/100 single number bet).

WinOwnFrac	Fraction of events played with WIN money (that is when the player has won more than they've spent so far during the session, the opposite is referred to as playing with OWN money)
TotalEvents	Total number of events
Typeswitch	WIN-OWN money type switch (the number of times the player switches from playing with WIN to OWN money or the other way round)
LoadFTSpend	Comparing speeds of money loading, winning and spending (3 approximate parameters indicating whether the player is currently loading money faster than spending it, winning money faster than spending it or winning money faster than loading it).
WinFTSpend	
WinFTLoad	
NetLoss	Net loss (the difference between money spent and money won)
AvgMoneySpentPerBalance	Average money spent as a fraction of current balance
AvgMoneySpentPerOwnMoney	Average money spent as a fraction of money loaded so far
AvgBalanceWhenMoneyLoaded	Average balance when money loaded
AvgReturn	Average return rate (the difference between money won and spent as a fraction of money spent)
AvgBalance	Average balance
Dur	Session duration (in minutes)

Table 1: Descriptions of high-level themes – parameters which were used to generate features that describe player behaviour

Feature aggregation

Where applicable, the following aggregation functions were used:

1. *Avg* - Average
2. *Max* - Maximum
3. *Min* - Minimum
4. *Levels* - the number of distinct values
5. *Std* - Standard deviation

Feature segmentation

The features have been further segmented using the following data elements.

1. *Pla* - Player
    a. *B2, B3* - Game category (B2 or B3)
    b. *Month* - Month
        i. *Week* - Week
            - *Day* - Day
                - *Ses* - Session

Feature naming convention

Due to the large number of features and in order to allow their generation to be scalable, straightforward and understandable, a feature naming convention has been implemented. Examples:

- *MonthStdMonthMaxMoneyWon* – For each month, find the value of maximum money won, then find the standard deviation for all those values. Meaning: if the value is low, the largest bet the player places each month is fairly consistent.

- *DayAvgDayLevelsMoneySpent* – For each day, find the number of distinct amounts of money spent, then find the average for all those values. Meaning: the value is the average number of different stake levels the player is using daily.

Thanks to the naming convention it is easy to implement and interpret a new feature across all aggregations and segmentations.

Classifier testing

In course of the research, the number of features used has gradually been increased and as a result a new set of 927 features has been constructed. It was very informative to compare the performance of the 927-feature **classifier** with the one built last year which was based on 99 features. There obviously was some level of overlap between the two feature sets but the 927 were constructed using a new scheme as explained before.

The construction of 927 features allows for a thorough exploration of the 17 feature themes. Inevitably however, there is a lot of redundancy within the features – some are highly correlated. As the purpose of this research question is to find aspects that are particularly useful for detecting problem gamblers, the approach adopted was to use various techniques to investigate well-performing feature subsets to identify **features that really matter**.

Below are the results achieved by various feature sets. The 'Top' features have been chosen out of the 927 based on a metric called **feature importance** – explained in the next section. All the results have been averaged over 50 experiments using a randomly selected training sample of 20 % to be used as a test set.

	TPR @ FPR = 10%		TPR @ FPR = 30 %		TPR @ FPR = 50 %		AUC	
	Average	Std	Average	Std	Average	Std	Average	Std
Top200	24.35%	3.62	61.90%	6.83	77.23%	3.96	6864.13	180.53
Top300	24.84%	3.43	61.37%	6.06	77.32%	4.17	6858.77	191.73
Top100	24.86%	3.64	60.70%	7.02	78.00%	4.58	6826.87	208.83
Top150	25.73%	3.27	61.63%	7.40	77.46%	3.47	6809.63	152.49
Top120	25.08%	3.38	62.02%	5.81	77.54%	3.23	6794.69	195.75
Top 60	24.44%	4.39	61.82%	6.80	77.25%	4.38	6749.67	175.88
422 features	24.98%	3.43	62.01%	5.91	75.29%	3.22	6723.04	177.61
927 features	24.79%	3.68	59.78%	8.15	74.53%	3.06	6701.99	221.41
99 features	21.92%	5.00	57.15%	6.92	73.19%	3.67	6650.50	209.97
5 best (tpr)	20.51%	2.97	51.15%	4.91	75.03%	5.52	6279.33	164.51
5 best (genetic)	20.17%	3.16	49.25%	4.95	74.49%	3.85	6201.91	198.01

Table 2: Results obtained using various feature sets and the classifier model as described by true positive rates at different false positive rates and the area under the ROC curve

It is an open question, which classifier performance metric is best and the optimal choice would vary based on application. The metrics provided above are the true positive rates at false positive rates fixed at 10%, 30% and 50% and the area under the curve (AUC) which can be interpreted as an overall performance metric over all false positive rates. For the task at hand it might indeed be useful to look at true positive rates at **lower false positive rate values** to see which feature subsets recognise a higher percentage of problem gamblers while not raising false alarms too often. That approach should, in principle, give preference to features that do reveal some aspects specific to problem gambling as opposed to those that are generally slightly more correlated with problem than non-problem gamblers.

Satisfactorily, the newly created feature sets outperform the old 99-feature classifier when it comes to both true positive rates and AUC. It is also worth noting that the performance when using all 927 features is worse than using only the top 200 or 300 highest importance features. Performance of the two 5-feature classifiers tested is much lower but shows that there are small feature subsets which do provide a lot of **discriminative capability**. Investigation of those will provide useful insights as to what aspects of player activity are indicative of being a problem gambler. The process of choosing the 5 features and their meanings will be discussed in further sections.

Classifier feature importance

The classifier algorithm provides a metric called ***feature importance***. Features can be ranked based on, roughly speaking, how often they are useful when deciding between problem and non-problem gamblers – their discriminative capabilities. Out of the 927 features, the 60 most important were chosen for subsequent experiments. Please refer to Appendix A.1 for the list of the top 60 most important features.

The classification algorithm, however effective at classifying, is not extremely useful when it comes to recognising and quantifying small uplifts in performance created by improvements to feature selection. This is due to the algorithm being non-deterministic – identical training and test datasets might result in slightly different ROC curves. This makes it difficult to **distinguish actual improvements from random noise**.

Logistic Regression classifiers – testing

For the reasons mentioned above, it has been decided to use **Logistic Regression** (LR) classifiers for research purposes. Below are the results of the previously described experiments performed using LR classifiers for raw and normalised (0 mean, unit variance) data.

	AUC		AUC (normalised)	
	Average	Std	Average	Std
Top200	5817.79	210.66	6491.27	203.33
Top400	6035.43	214.40	6367.20	200.85
Top100	5609.39	244.10	6608.58	192.65
Top 60	6624.79	209.95	6662.69	203.55
927 features	6041.05	512.79	6191.28	210.99
99 features	6011.37	232.39	6690.04	216.40
5 best (tpr)	6705.52	196.85	6696.60	198.88
5 best(genetic)	6691.69	203.56	6692.18	202.18
1 best	6469.10	206.69	6479.50	219.51

Table 3: Results achieved using various feature sets (of raw and normalised data) and the Logistic Regression algorithm as described by the area under the ROC curve

60 highest importance features

As mentioned before, the **60 highest importance features** were identified thanks to the classifier algorithm (the full list is available in Appendix A.1). Some of the features have also been assigned to one of the following groups based on their feature theme:

I Levels of money spent
II Standard deviation of money spent
III Average money spent
IV Average money loaded
V Levels of money loaded

Features assigned to the same group stem from the same underlying metric. Not only do they have a lot in common in general terms but they are also **highly correlated**.

In the search for the best (the most discriminative) set of features it is hugely important to appreciate the two fundamental aspects:

- Features' individual **discriminative** capabilities
- **Correlations** between features

Example

The 8 most important features (Appendix A.1) come from the same theme and are extremely highly correlated. Therefore, perhaps rather than using the 8 top features in a classifier, some of them could be replaced by features that have worse individual discriminative capability but are **less correlated** among themselves.

This idea is explored in the following sections. All the Logistic Regression tests are repeated 500 times so that the **average true positive rate** (at the default threshold of 0.5) is a robust indicator of the general performance of the classifier.

Features' individual discriminative capabilities

Features' individual discriminative capabilities were found by testing classifiers which use only that one feature. The **true positive rate** (tpr) and the **threat score** (ts) have been used for evaluating classifiers' performances as suggested in the literature for imbalanced datasets. The **true positive rate** is the percentage of positive samples (problem gamblers) that have been classified correctly. The **threat score** is the number of true positives as a fraction of all samples except the correctly classified non-problem gamblers (all samples but true negatives). Both measures need to be as high as possible. This ensures a high true positive rate (which is what makes a good problem-gambler classifier) while preventing the false positive (false alarm) rate from being too high. Detailed results can be found in Appendix A.2.

The true positive rates found for one-feature classifiers go in line with the feature importances but there are some **exceptions**. The best scores are achieved by features representing group I, group II and group III, lower by group IV and group V. Out of the features that are not associated with any group, some have moderate individual scores while a number of them have a zero score. That means that the feature by itself has no value in terms of differentiating between problem and non-problem gamblers – the classifier's best guess was to say that all players are non-problem gamblers.

Classifier based on top features – testing

A natural approach to selecting an optimal feature subset would be to choose the N highest importance features. The results of testing such classifiers are presented in Appendix A.3 and in the graph below.

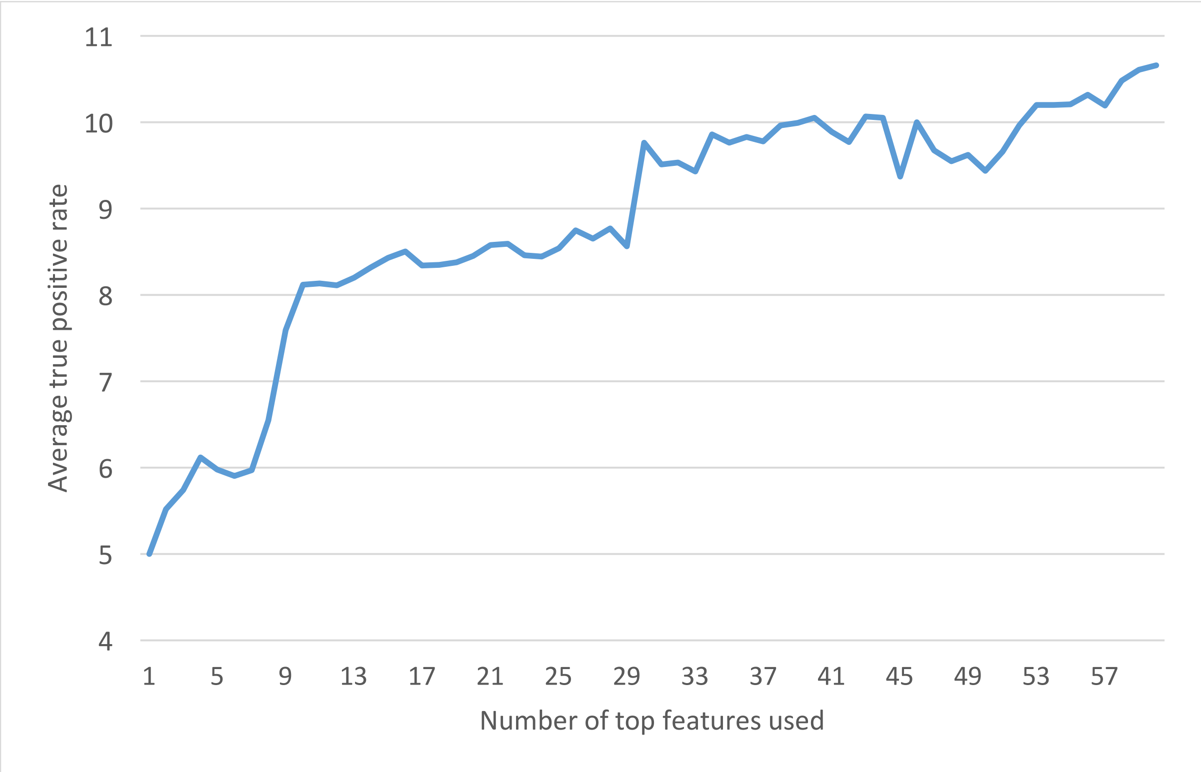


Figure 1: True positive rate for a fixed threshold of 0.5 for a varying number of top features used (e.g. value 29 on the X axis indicates a classifier which uses the top 29 highest importance features)

It is satisfactory to observe that a classifier with 30 – 50 top importance features shows comparable performance to classifiers with a full set of 200 – 300 features. It is, however, important to observe (Appendix A.3) that it is the **inclusion of a feature from a different theme group** that usually produces the highest uplift in performance. This idea is implemented and validated in the following two sections.

Manual selection of an optimal subset of features

The process is a combination of selecting features with the **highest individual discriminative capabilities** alongside the idea of including **features from different theme groups**. Please refer to Appendix A.2 for detailed information on the individual features quoted below.

Feature 1 *DayAvgDayLevelsMoneySpent* is definitely a good starting point because not only does it have the highest feature importance but also the highest individual discriminative capability based on Logistic Regression testing. It is a member of group I, therefore it would be a logical step to try combining it with well-performing features from group II and group III. The results of this trial are shown in the table below:

Features	Average True Positive Rate
1,16	5.89
1,17	6.15
1,14	**6.90**
1,19	5.90
1,10	**6.84**
1,30	5.92
1,36	6.13
1,10,14	**7.45**

Table 4: Results from the trial of combining Feature 1 with well-performing features from group II and group III

The best performing two-feature classifiers indicate that a good choice for a three-feature classifier would be: feature 1 *DayAvgDayLevelsMoneySpent* from group I, feature 10 *DayStdDayAvgMoneySpent* from group II and feature 14 *B2MonthAvgMonthStdMoneySpent* from group III. Indeed, the classifier's average tpr is 7.45, much higher than 5.74 achieved by the classifier using the 3 highest importance features.

It is informative to note that feature 14 *B2MonthAvgMonthStdMoneySpent*, which, together with 1 *DayAvgDayLevelsMoneySpent*, forms the best-performing two-feature classifier out of the ones tested, does not have the highest individual discriminative capability across group II. In fact, both features 16 *WeekAvgWeekStdMoneySpent* and 17 *MonthAvgMonthStdMoneySpent* have higher individual scores. The high performance of the feature 1 and 14 classifier can be attributed to the fact that 1 *DayAvgDayLevelsMoneySpent* is slightly less correlated with 14 *B2MonthAvgMonthStdMoneySpent* than with 16 *WeekAvgWeekStdMoneySpent* or 17 *MonthAvgMonthStdMoneySpent*.

The next step was to find the fourth feature to add to the set. Candidates were considered among members of group IV, group V and among features that have not been assigned to any group.

Features	Average True Positive Rate
1,10,14,18	7.13
1,10,14,31	7.07
1,10,14,32	7.14
1,10,14,22	7.39
1,10,14,35	7.46
1,10,14,27	**7.64**
1,10,14,29	7.42
1,10,14,33	7.49
1,10,14,40	**7.66**
1,10,14,48	7.57
1,10,14,53	7.45

Table 5: Results from the trial to add additional features with our best performing 3 feature classifier.

Two of the features from the last group proved to have provided a noticeable uplift to the performance of the classifier. It is superior to a classifier simply taking the top 4 highest importance feature and comparable with one taking the top 9 features.

The above results confirm the idea that effective feature selection is a combination of **individual discriminative capabilities** and **originating from different theme groups**. The next section is aimed at validating that conjecture further.

Combination-based half-automated feature selection scheme

The purpose of this section is to use a **combination-based** scheme to test a large number of feature subsets and validate the previously stated conjecture. The aim is to arrive at a selection of close-to-optimal 5-feature subsets constructed from the 60 highest importance features.

As it is impractical (and unnecessary) to test all the $\binom{60}{5} = 5\,461\,512$ possible combinations, the scheme has been implemented in 5 steps. At each step one new feature is chosen for a classifier and only 10 – 20 best performing classifiers pass on to the next stage. That way, the number of necessary tests has been reduced from $\binom{60}{5}$ to about $5{\times}15{\times}16 = 1200$ that is by a factor of about $4{,}500$. On the other hand, since only a tiny fraction of the search space is being examined, the classifiers found are by no means sure to be globally optimal. Their performance is, however, far superior to those constructed naively from the highest importance N features. Therefore, they do provide valuable insights into what makes a **good feature subset**.

The results achieved and features used for the best classifiers are presented in Appendix A.4.

Even though the classifiers using the manually selected subsets of features perform slightly worse than those found using the combination-based scheme, they do have a **similar structure**. The structure is best explained by looking closely at some of the best performing 5-feature classifiers (Appendix A.4).

Almost all of the best classifiers use some of the top individually performing features like feature 1 (DayAvgDayLevelsMoneySpent), 4 (B2DayAvgDayLevelsMoneySpent) and 10 (DayStdDayAvgMoneySpent). Their defining characteristic is also containing features from group I (levels of money spent), group II (standard deviation of money spent) and group III (average money spent) as well as some of the features not belonging to any group. This further confirms the conjecture stated in previous sections.

It is very informative to look closely at the top performing 5-feature classifier which uses features 1 and 34 from group I, 10 from group III, 28 and 46. Only features 1 and 10 belong to individually well performing whereas 34, 28 and 46 have close to zero individual discriminative capabilities. However, what makes the difference is the fact that these three features are **almost completely uncorrelated** with 1 or 10. (Even though 1 and 34 belong to the same group, these are fairly uncorrelated, probably because 1 is a general feature whereas 34 refers to B3 type games).

Finally, it is interesting to appreciate the performance of the classifier which uses **as few as 5 features** selected by the above process. As far as Logistic Regression is concerned, its true positive rate is higher than the previously implemented classifier. It is comparable in terms of performance to the extended classifier using 100 – 200 features.

When using the logical approach of choosing the highest importance features to run the classifier on, **as many as 30** would have to be selected for it to perform as well as the 5-feature classifier. This is quite a remarkable result. Please refer to the figure below for a visualisation comparing performances of different Logistic Regression classifiers.

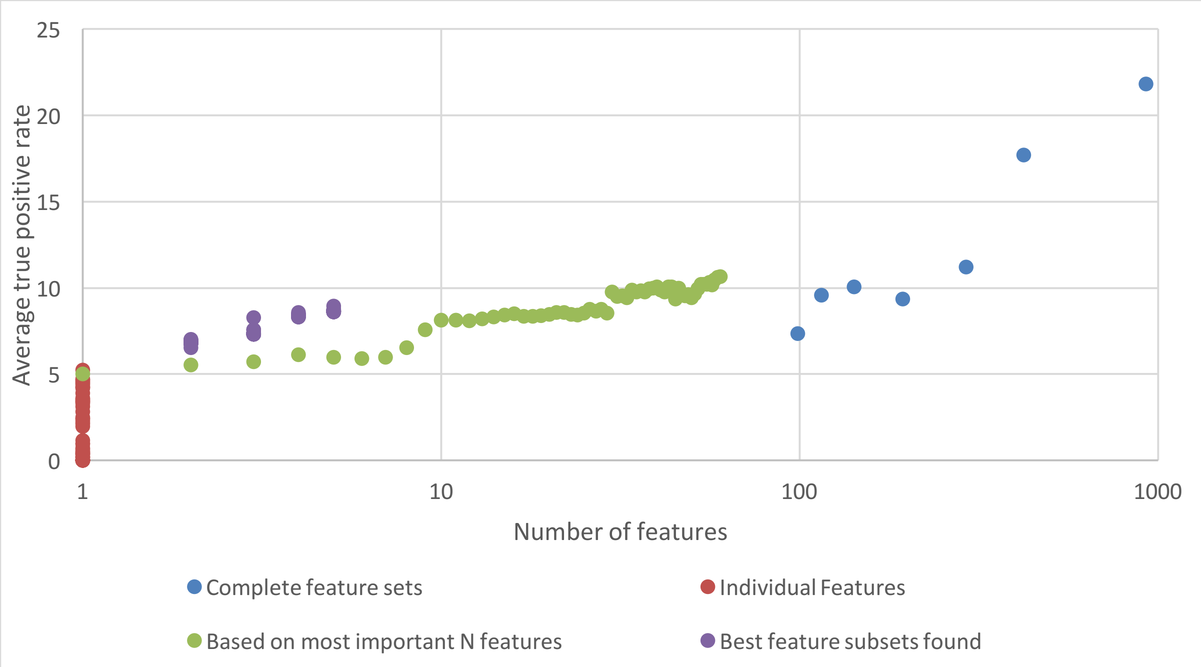


Figure 2: Average true positive rate as a function of the number of features used for different feature selection processes

This fairly technical section can be concluded to be a preliminary confirmation of the approach adopted for selecting subsets of features. In the following section the approach is further confirmed by means of a different objective – true positive rate at a fixed false positive rate of 20 % as opposed to the fixed probability threshold of 0.5.

The technical findings and conclusions are given more practical and **interpretable** context starting from section Non-technical insights – feature meaning where meanings of particular features and feature subsets are investigated.

Optimal feature subset selection – true positive rate at fixed false positive rate

The half-automated combination-based feature set search technique used in the previous section resulted in 5-feature classifiers which are not very diverse. This bias has definitely at least partially been caused by the method itself. Therefore, in the current experiment a method that has been implemented ensures that no structure is artificially imposed on the 5-feature classifiers obtained.

A procedure which somewhat resembles a **genetic** type of algorithm has been implemented. It generates and tests random 5-feature subsets. Only when a well-performing subset is met, it attempts to improve it by only exchanging one or two of its features by a randomly chosen one. That procedure fulfils the following requirements:

- A large subset of the search space is explored
- It is impossible to be stuck with a similar feature subset structure for too long
- Well-performing subsets of the search space are explored more thoroughly

The results of the experiment (best 5-feature classifiers found) are presented in the table below.

					True Positive Rate (at a False Positive Rate of 20%)
1	8	10	16	20	*46.02*
1	6	20	30	37	*45.53*
4	16	28	34	37	*45.49*
4	10	13	24	54	*45.42*
1	8	21	31	37	*45.38*
3	10	15	30	32	*45.37*
4	8	20	31	39	*45.29*
1	8	10	20	31	*45.28*
1	5	8	16	44	*45.21*
3	4	16	20	22	*45.15*
1	4	8	37	47	*45.14*
1	6	19	32	41	*45.12*
1	7	9	14	21	*45.07*
4	12	20	45	54	*45.04*
3	6	9	9	58	*45.00*

Table 6: The best 5-feature classifiers found in terms of the true positive rate at a fixed false positive rate of 20%

It is satisfying to observe that although the particular features of the top-performing classifiers are different than in the previous experiment, their general structure is similar and very much in line with the initial conjecture stating that features from **different feature themes** should be included.

Non-technical insights – feature meaning

Finally, the analysis would not be complete without providing it with non-technical insights based on **feature meanings**. First of all, it is useful to reiterate the meanings behind theme groups that were the most prevalent across the best performing feature subsets found:

I Levels of money spent
II Standard deviation of money spent
III Average money spent
IV Average money loaded
V Levels of money loaded

The following features were found to be most useful (in terms of discriminative capability) based on the best performing 5-feature Logistic Regression classifiers found in the first experiment:

	Feature	Meaning
1	DayAvgDayLevelsMoneySpent	Average daily number of different stakes
10	DayStdDayAvgMoneySpent	How diverse daily average stakes are
28	B2DayMinDayAvgMoneySpentPerBalance	This is a lower bound of how much a player is willing to spend of their current balance on B2 games.
34	B3SesStdSesLevelsMoneySpent	How diverse the numbers of different stakes per session are (B3 games)
46	SesAvgSesAvgWinOwnFrac	How often, on average, the player plays with winnings vs with their own money
11	B2SesAvgSesStdMoneySpent	On average, how diverse stakes within a session are (B2 games)
13	B2DayStdDayLevelsMoneySpent	How diverse daily numbers of different stakes are (B2 games)

5	B2SesAvgSesLevelsMoneySpent	Average number of different stakes per session (B2 games)
4	B2DayAvgDayLevelsMoneySpent	Average daily number of different stakes (B2 games)
39	B2MonthMinMonthAvgMoneySpentPerBalance	The lowest of monthly averages of how big a fraction of current balance stakes are (B2 games)
41	B2SesMinSesAvgMoneySpentPerBalance	The lowest of session averages of how big a fraction of current balance stakes are (B2 games)
15	B2DayAvgDayStdMoneySpent	On average, how diverse stakes are within a day (B2 games)
7	DayStdDayLevelsMoneySpent	How diverse the daily numbers of different stakes are
26	B2DayMinDayAvgMoneySpentPerOwnMoney	The lowest of daily averages of how big a fraction of money put in so far during a session stakes are (B2 games)
24	MonthAvgMonthLevelsMoneySpent	Average monthly number of different stakes
14	B2MonthAvgMonthStdMoneySpent	On average, how diverse stakes are within a month (B2 games)
51	SesStdSesAvgLoadFTSpend	How diverse the per session comparisons between speeds of money loading vs money spending are

Table 7: The most useful features based on the best performing 5-feature classifiers (found using the half-automated feature selection process)

The second experiment has shown the significance of similar features except there were some that do not have a direct counterpart in the above list:

	Feature	**Meaning**
54	MonthMaxMonthAvgWinFTSpend	The highest of monthly averages of how often the player is winning money faster than spending
31	WeekStdWeekAvgMoneyLoaded	How diverse weekly average amounts of money loaded are
32	SesStdSesAvgMoneyLoaded	How diverse per session average amounts of money loaded are
22	DayStdDayLevelsMoneyLoaded	How diverse the number of different daily levels of money loaded are

Table 8: Additional useful features based on the best performing 5-feature classifiers (found using the genetic-like feature selection process)

Practical insights from top-performing feature sets

Further analysis focuses on the insights that can be drawn from the best-performing feature subsets. The following two tables present the feature sets of the best-performing 5-feature classifiers:

	Feature	Meaning
1	DayAvgDayLevelsMoneySpent	Average daily number of different stakes
10	DayStdDayAvgMoneySpent	How diverse daily average stakes are
28	B2DayMinDayAvgMoneySpentPerBalance	This is a lower bound on how much a player is willing to spend of their current balance on B2 games.
34	B3SesStdSesLevelsMoneySpent	How diverse the numbers of different stakes per session are (B3 games)
46	SesAvgSesAvgWinOwnFrac	How often, on average, the player plays with winnings vs with their own money

Table 9: The best performing 5-feature set found using the half-automated feature selection process

	Feature	Meaning
1	DayAvgDayLevelsMoneySpent	Average daily number of different stakes
8	WeekAvgWeekLevelsMoneySpent	Average weekly number of different stakes
10	DayStdDayAvgMoneySpent	How diverse daily average stakes are
16	WeekAvgWeekStdMoneySpent	On average, how diverse stakes during one week are
20	B2WeekAvgWeekLevelsMoneySpent	Average weekly number of different stakes

Table 10: The best performing 5-feature set found using the genetic-like feature selection process

The question that is definitely worth asking is why these two 5-feature sets are performing exceptionally well. Another very interesting and a more general question is why certain sets of features are good at recognising problem gamblers and what values of those features indicate that. The following subsections will present different approaches to addressing these issues through visualisation.

2D feature vs feature scatter plot

One approach to visualising problem and non-problem gamblers would be to represent them as points on a two-dimensional plot with selected features as shown on the X and Y axis. Problem gamblers are represented as red and non-problem as green.

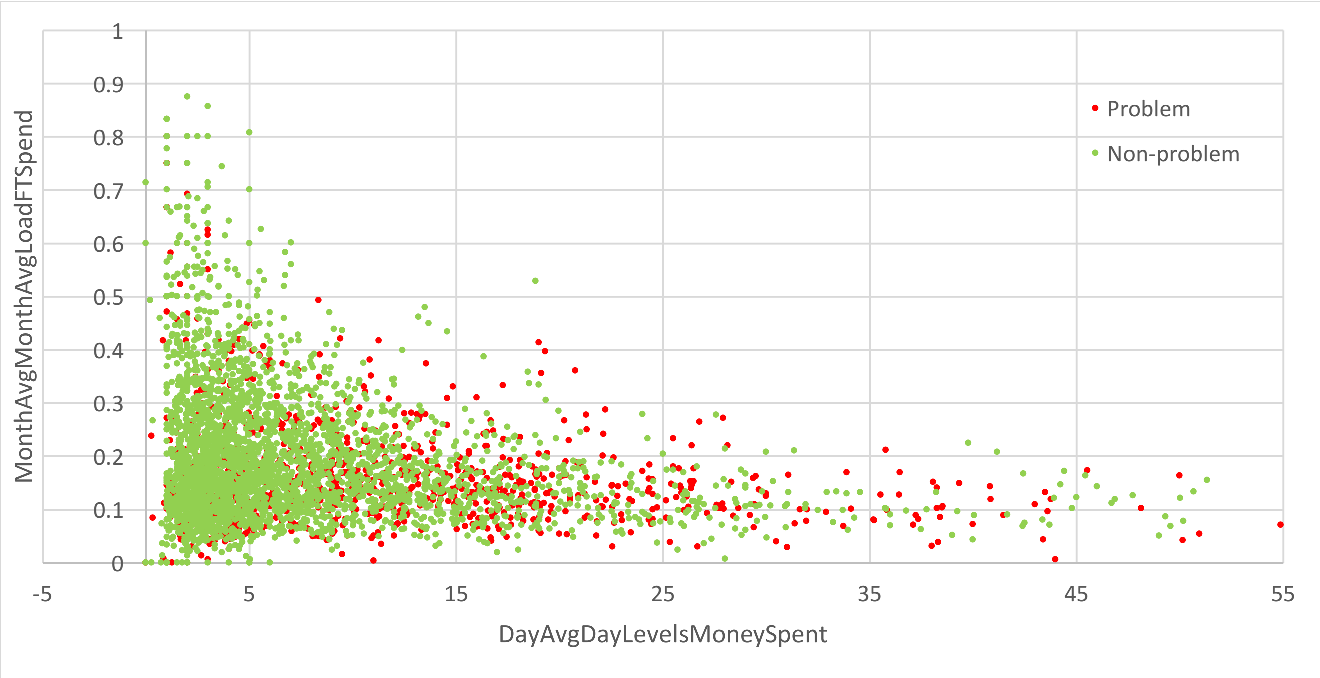


Figure 3: A 2D plot where each point is a gambler and its location is determined by values of two features (as explained below)

DayAvgDayLevelsMoneySpent (X axis): Average daily number of different stakes

MonthAvgMonthAvgLoadFTSpend (Y axis): Comparing the speeds of money loading and spending – aggregated over a time frame of a month

The two features chosen are *DayAvgDayLevelsMoneySpent* (the most positively correlated with the *PG score*) on the X axis and *MonthAvgMonthAvgLoadFTSpend* (the most negatively correlated with the *PG score*) on the Y axis. Therefore, what the X axis shows is the average daily number of stake levels which tends to be higher for problem gamblers. The Y axis shows the monthly average proportion of events when players are loading money faster than spending it which tends to be lower for problem gamblers.

It is tricky, however, to spot any of those patterns on the above plot, not to mention draw any more in-depth conclusions. That is the reason behind developing other visualisation techniques in the sections to follow.

2D feature plot

The plot below is a visualisation of a two-feature classifier prediction. For every point in the plot, its colour indicates the probability of this pair of values corresponding to a problem gambler (the darker, the more likely) according to the best performing classifier. This helps to understand what ranges of feature values correspond to problem gamblers thus showing why certain features are more useful than others.

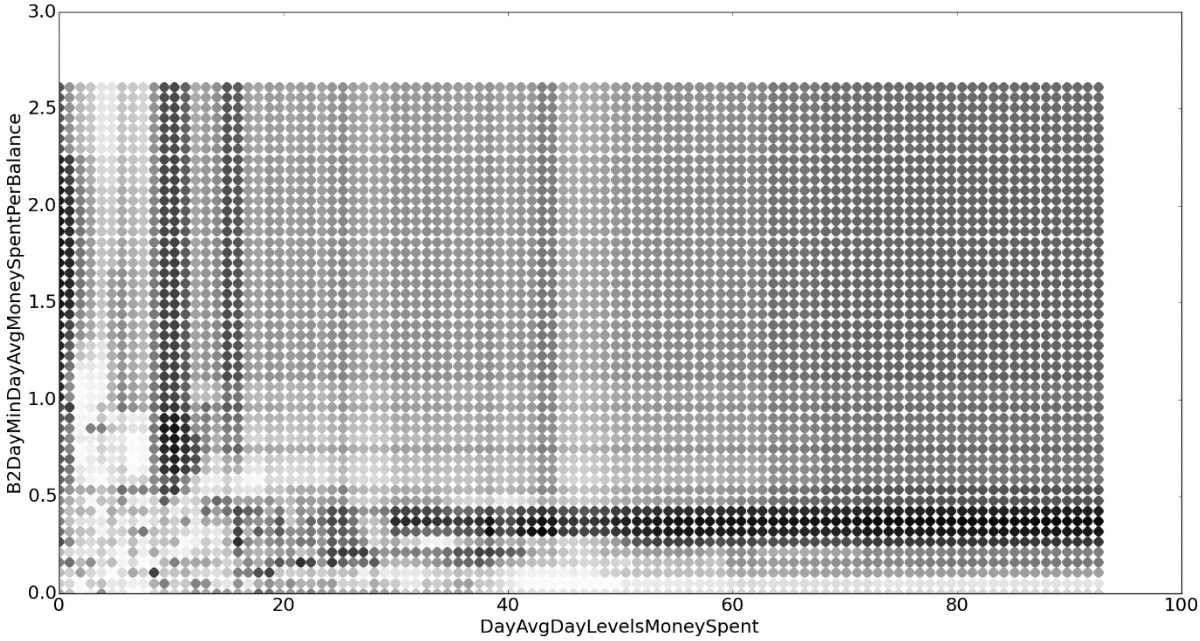


Figure 4: A 2D plot where each point represents a combination of values of two features (as explained below) and its colour represent the probability of this combination of values indicating a problem gambler (darker for higher probabilities)

DayAvgDayLevelsMoneySpent (X axis): Average daily number of different stakes

B2DayMinDayAvgMoneySpentPerBalance (Y axis): Lower bound on how much a player is willing to spend of their current balance on B2 games

It is informative to observe how the two features do not have a great deal of discriminatory capability by themselves (in one dimension). However, when put together to form a **two-dimensional picture**, a certain structure is revealed. The feature on the X axis is the daily average number of stake levels, on the Y axis – this is a lower bound of how much a player is willing to spend of their current balance on B2 games. Problem gamblers are most likely to be found in one of two regions (the values of interest for the money spent per current balance is clearly between 0 and 1):

- Minimum daily average of money spent per current balance between 0.6 and 0.9 and low average number of distinct daily stakes (around 10): **less 'chaotic'** behaviour (smaller number of different stakes daily) but an average bet is putting **more than 50% of the current balance at stake**
- Minimum daily average of money spent per current balance between 0.3 and 0.5 and higher average number of distinct daily stakes: **stake values are more moderate** (less than 50% of current balance on average) but the behaviour is **more 'chaotic'** (60 – 80 different stake values daily)

It is important to note here that, as will be shown later, there is no positive correlation between the *number of sessions* and the *PG score* (it is in fact slightly negative). It is therefore false to explain the higher number of different stakes by problem gamblers using the argument that they tend to play more.

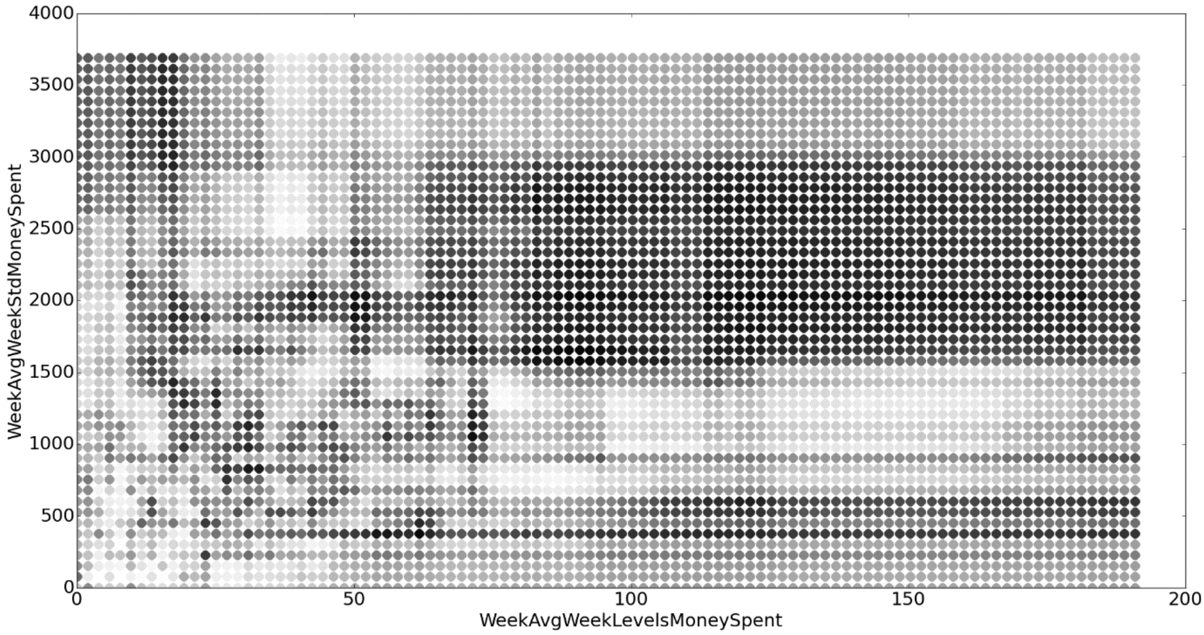


Figure 5: A 2D plot where each point represents a combination of values of two features (as explained below) and its colour represent the probability of this combination of values indicating a problem gambler (darker for higher probabilities)

WeekAvgWeekLevelsMoneySpent (X axis): Average weekly number of different stakes

WeekAvgWeekStdMoneySpent (Y axis): On average, how diverse stakes during one week are

The above plot shows a typical structure arising from theme group I (Levels of money spent), II (Standard deviation of money spent) features. There is a high correlation between the two groups because they are using different metrics to represent the same idea – how diverse the stakes are. There is a region where both X and Y axis values are high where it is most likely to encounter a problem gambler.

Individual features visualisations

Another useful way of measuring feature's discriminative capability is to see how it **correlates with the *PG score***. If the correlation is positive then the higher its value for a player, the more likely they are, on average, to be a problem gambler.

First of all, it is informative to look at some of the simplest metrics.

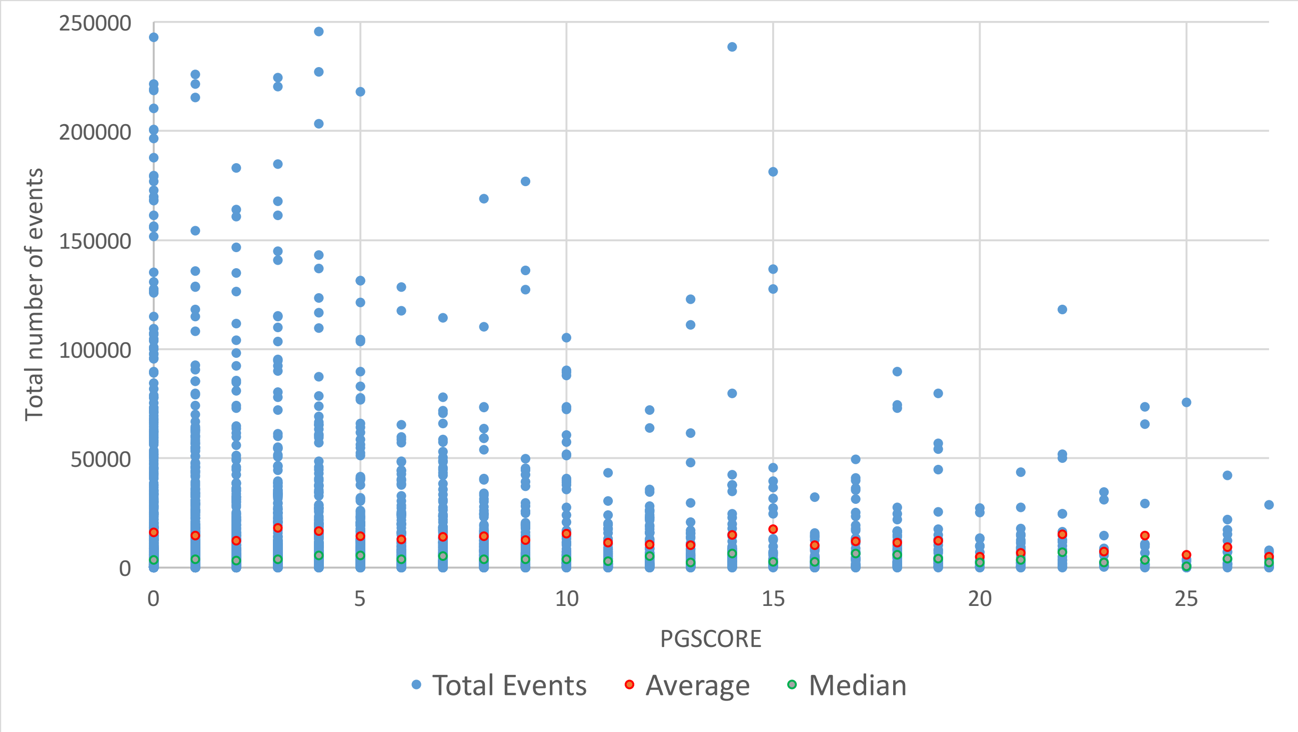


Figure 6: Distribution of the total number of player's events (their total activity) at different PG score levels

The above plot shows the distribution of player's total number of events for different *PG scores*. The correlation between the two metrics is -0.05 which means that **problem gamblers do not tend to play more**. In fact, on average, they tend to play slightly less than non-problem gamblers.

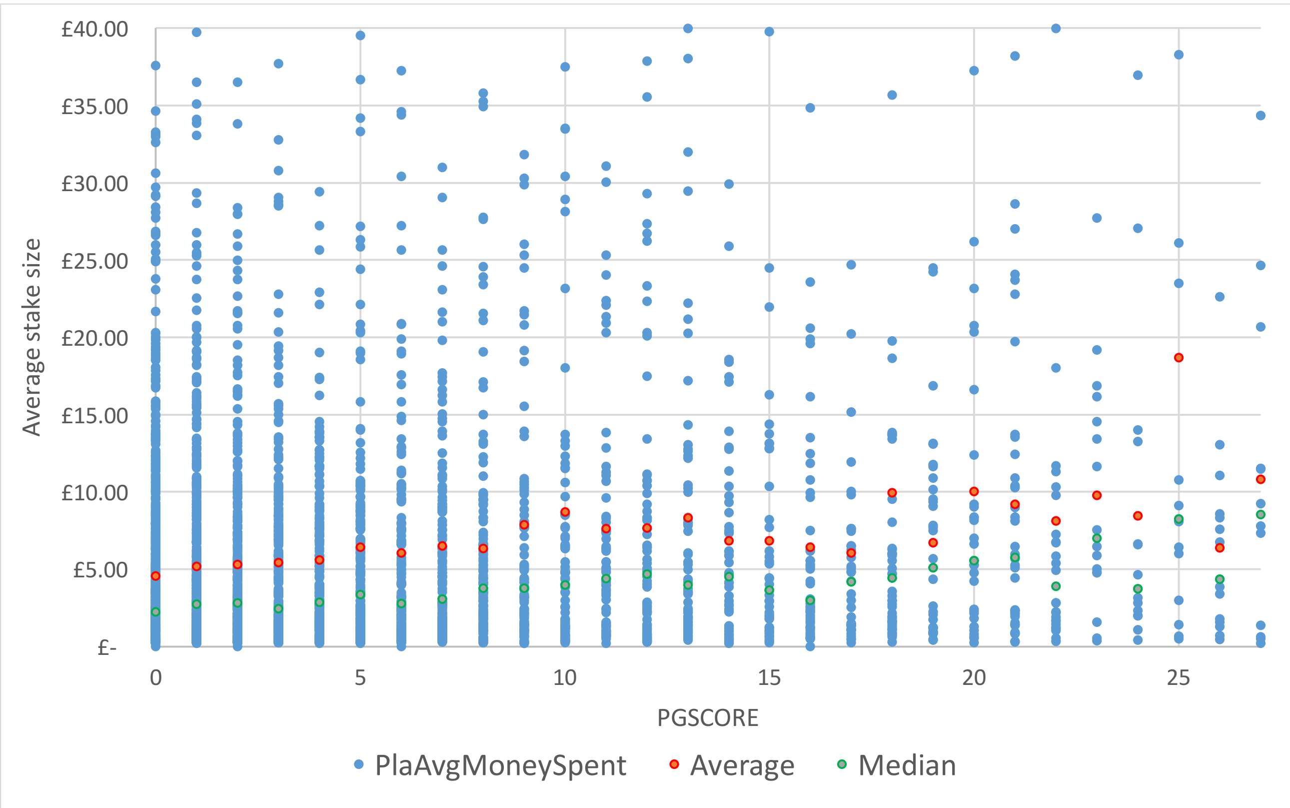


Figure 7: Distribution of player's average stake size at different PG score levels

Above is a similar plot depicting the average stake size whose correlation with *PG score* is 0.16. What this means is that there is a slight tendency for problem gamblers to play with higher stakes.

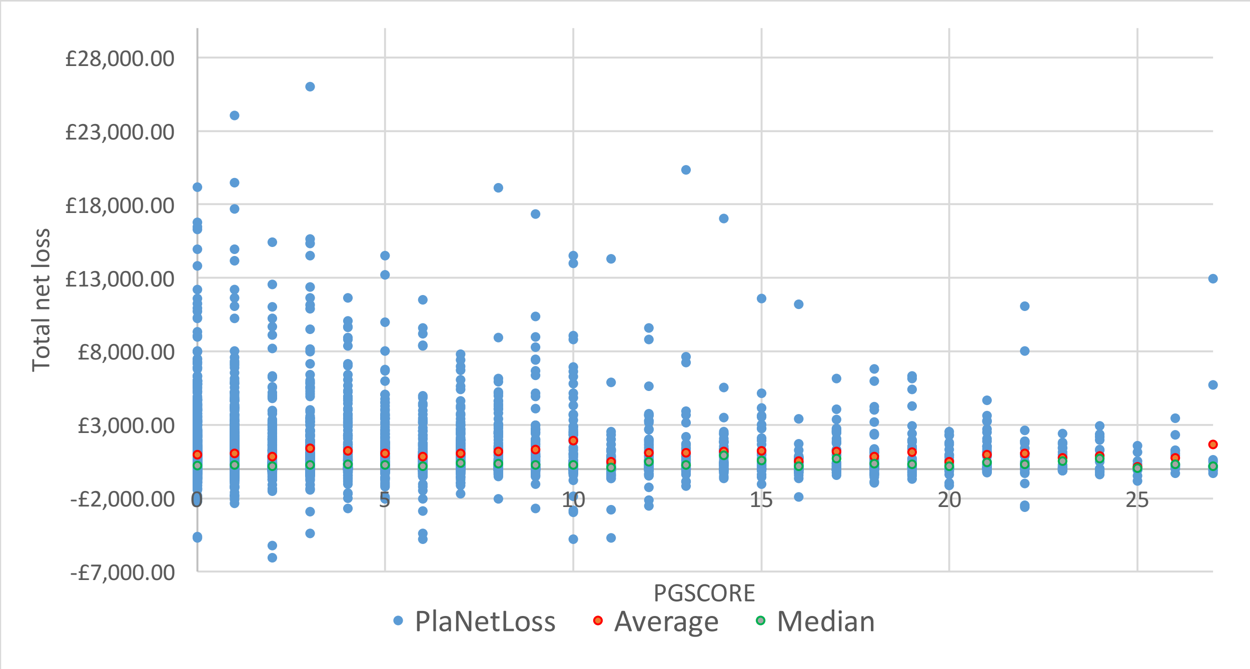


Figure 8: Distribution of player's total net loss at different PG score levels

The graph above shows another counter-intuitive result. It turns out that there is no correlation between the *PG score* and player's total net loss. This means that problem and non-problem gamblers on average lose equal total amounts of money while gambling.

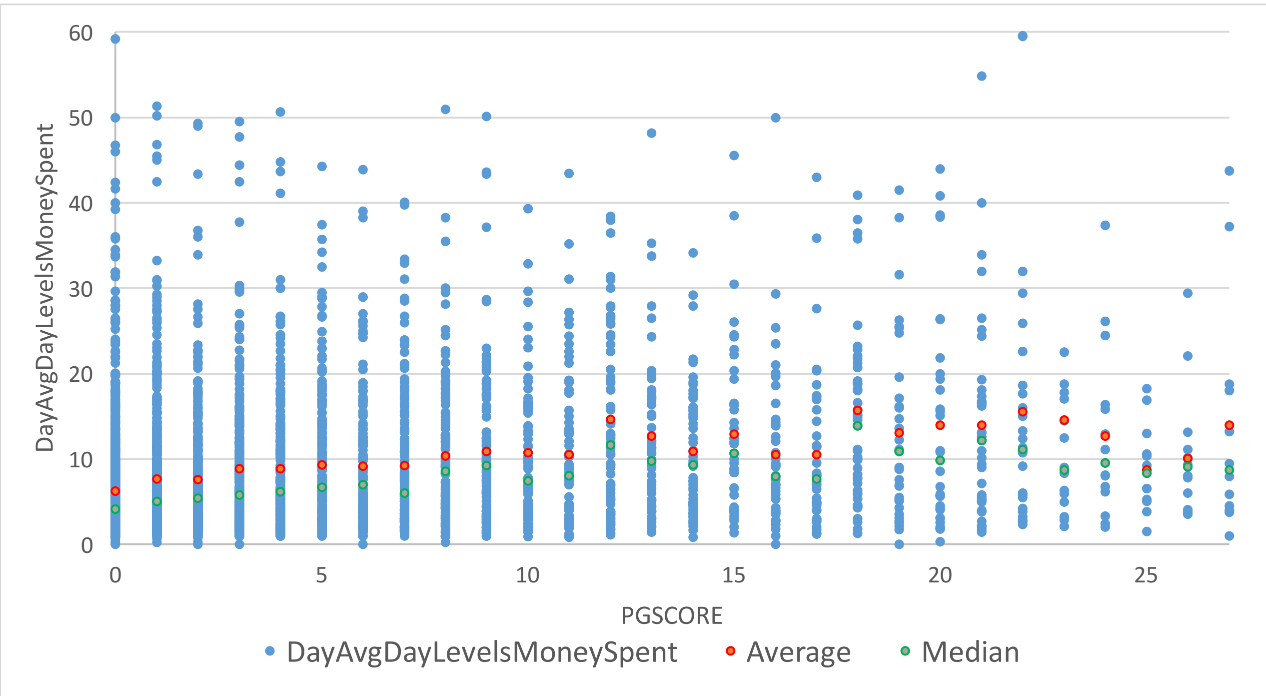


Figure 9: Distribution of player's DayAvgDayLevelsMoneySpent feature value (as explained below) *at different PG score levels*

DayAvgDayLevelsMoneySpent (Y axis): Average daily number of different stakes

The graph above refers to feature 1 *DayAvgDayLevelsMoneySpent* which is the most highly correlated with *PG score* out of all features – correlation of 0.24. The graph indeed shows that there is a tendency for high *PG score* players to use **a higher number of distinct stake levels daily** which can be described as a more **'chaotic'** behaviour.

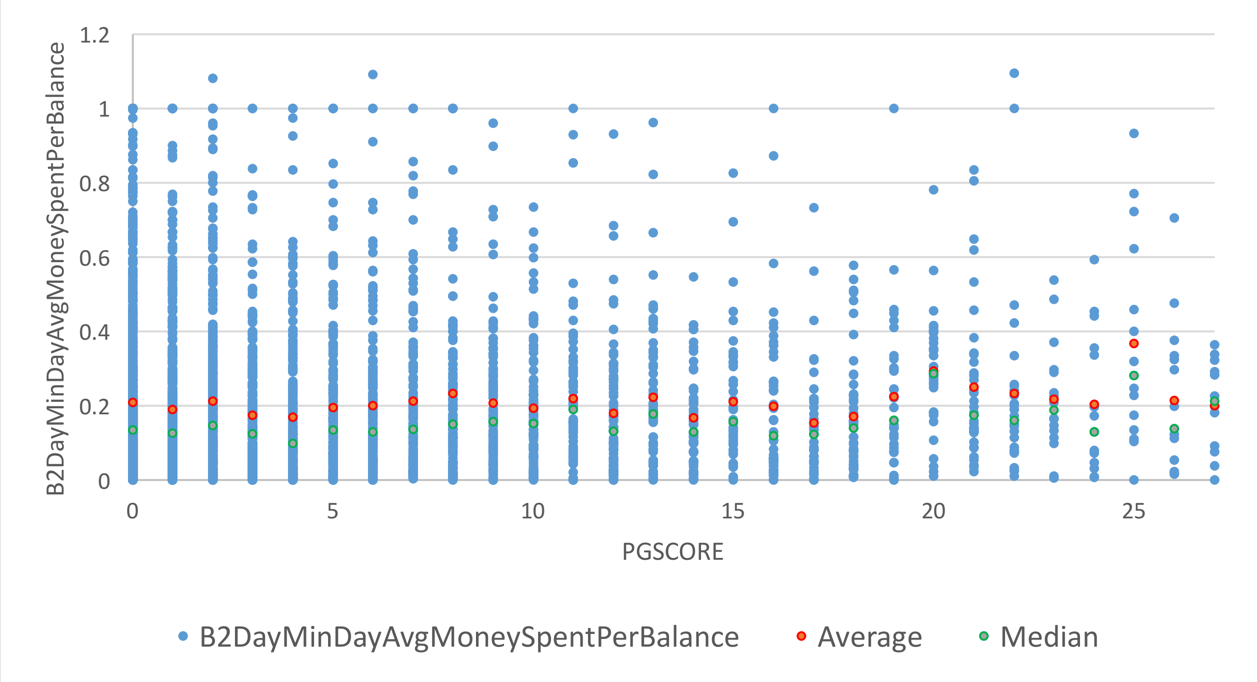


Figure 10: Distribution of player's B2DayMinDayAvgMoneySpentPerBalance feature value (as explained below) *at different PG score levels*

B2DayMinDayAvgMoneySpentPerBalance (Y axis)*: A lower bound on how much a player is willing to spend of their current balance on B2 games*

Finally, the above graph describes feature 28 *B2DayMinDayAvgMoneySpentPerBalance* which is a feature of low individual discriminatory capability. It is, together with the above feature 1 *DayAvgDayLevelsMoneySpent*, a part of the best-performing 5-feature classifier (see Practical insights from top-performing feature sets). As a confirmation of the former, feature 28's correlation with the *PG score* is very close to zero: 0.02.

The two graphs above may give some idea as to why these two particular features would work together well. For *DayAvgDayLevelsMoneySpent*, values are unusually low for players whose *PG score* is of value 25. Judging by that feature alone, such players could easily be confused for non-problem gamblers. The 25 *PG score* players tend to, on the other hand, have a high value of *B2DayMinDayAvgMoneySpentPerBalance*. Therefore, this extra dimension, thanks to it being fairly uncorrelated with the first feature, adds valuable information which the classifier can use to **differentiate between different groups of players**. If both features were very highly correlated, the 25 *PG score* players will have unusually low values in both cases and it would still be very hard to differentiate them from non-problem gamblers.

The above shows the complexity of the problem. It is hard to define exactly what makes features work together well because it is always a product of many different factors. What can, however, be stated are the following two characteristics of useful subsets of features:

- Features of high individual discriminative capability or, in other words, **informative features**. Thanks to how the classifier works, it is not necessary for a feature to align exactly with what constitutes a problem or a non-problem gambler. It needs to be able to be used for creating relatively **homogeneous problem / non-problem player sub-groups** which is connected with being correlated with the *PG score*.

- Features that **complement each other** which, as explained above, is a slightly extended notion of being uncorrelated.

Finally, the following section focuses on the most informative features to paint a picture of what characteristics should and should not be expected to differentiate a problem gambler from a non-problem gambler.

Feature correlations with the PG score – high level conclusions

The two lists below are an attempt at a **high level conclusion** of what can be understood from the data about problem gamblers based on the *reverse engineering* conducted:

Indicative of problem gambling:

1. Standard deviation and number of levels of money spent – how diverse the stakes are. Problem gamblers tend to be more **chaotic,** that is using a higher number of distinct stake levels as well as levels that span larger ranges of values.
2. Standard deviation and maximum value of money won. These types of features again refer to diversity, this time in terms of the money won. Interestingly, it is especially the diversity in the 'upward' direction that is characteristic of problem gamblers – they tend to have **higher extreme wins**.
3. Standard deviation of net losses – problem gamblers' values tend to be more diverse.
4. Standard deviation, minimum and maximum values of session (time) duration – problem gambler **sessions tend to be longer and their duration varies more**.
5. Standard deviation of balance – balance takes on a more diverse range of values for problem gamblers
6. Speed of **winning money** compared to speed of **spending money** – higher for problem gamblers.
7. Problem gamblers tend to **play with winnings** (as opposed to with their own money) more often than non-problem gamblers
8. Speed of **loading money** compared to speed of **spending money** – lower for problem gamblers.
9. Values of **maximum losses** – higher for problem gamblers
10. **Probability of winning** and **average returns** – higher for problem gamblers

Problem and non-problem gamblers – high-level tendencies

As much as it is impossible to find any definitive characteristics of problem gamblers, there are some tendencies which can be identified. The tendencies have been put forward from a more technical perspective in the previous section. This section attempts to give them more context and paint a picture of what, according to the data, **makes a problem gambler**.

The results achieved confirm the fact that a 'problem gambler' is a term that is difficult to define. The most high-level conclusion that can be drawn based on the data is that an average problem gambler seems to be doing **better at gambling** than an average non-problem gambler.

The defining characteristic of problem gamblers would be that they are **chaotic**, less predictable. The amounts of money they bet and win **varies more**, reaches higher maximum values and is spread across wider ranges. When it comes to net losses, problem gamblers' vary more but in general they do not tend to be higher.

There is virtually no way to distinguish between problem and non-problem gamblers based on the number of machine events they have taken part in. Non-problem gamblers tend to have only slightly more B2 machine events (and a bit more so for B3 machine events). In spite of this, problem gambler sessions tend to have longer time duration.

In a very high-level view – problem gamblers tend to seem **more successful**. They have more periods when they win money faster than they spend it. As a consequence, they play with their winnings more often than non-problem gamblers. Their probability of winning and average returns are also higher.

In spite of the above, problem gamblers' chaotic behaviour also manifests itself in their losses. Even though they have similar total values to non-problem gamblers, they tend to have **higher one-off losses**: their maximum session / day / week losses tend to be higher.

Finally, there are some features which might seem like they could be relevant but in fact they behave very similarly for problem and non-problem gamblers. These include: total number of events / sessions, total losses and the numbers of times switching between playing with winnings and with their own money.

Conclusion

The point of the question was to use the process of 'reverse engineering' to discover new features which might play a useful role within the development of algorithms. It has been shown how the performance of the classifier in terms of true positive rates and area under curve can be **improved** by inclusion of extra features and by appropriately selecting them. Furthermore, discriminative capabilities of particular features and groups of features have been analysed to discover aspects of player behaviour that are most effective when it comes to differentiating between problem and non-problem gamblers.

It has been shown how choosing features that have both **high individual discriminative capabilities** and come from **diverse themes** can lead to a well-performing classifier using a much reduced number of features.

Finally, practical insights have been drawn from the findings. By means of more detailed analysis some general tendencies have been found to discriminate between problem and non-problem gamblers. These can be used to both improve the performance of classification algorithms as well as make an attempt at describing what **behaviour types** should be searched for when identifying problem gamblers.

Chapter 2: Research Question 7

What are the differences in demographics between B2/B3 players? What else can we learn about players' transitions between B2 and B3 content?

Note: The data used to answer this research question is from the 4,000 loyalty card customers surveyed in the original research project.

High-level Findings

Introduction

This chapter investigates the distinction between B2 and B3 as games and by the players who play them. On the game level, they are defined by the maximum allowed stake size (£100 for B2 and £2 for B3) and minimum time between stakes (20 seconds for B2 and 2 seconds for B3) whereas players are categorised based on the prevalence of each category of bets as explained in one of the later sections. Additionally, the concept of a 'hybrid' bet is introduced (a bet of more than £2 during a B3 game) together with an analysis in terms of its properties and impact on player behaviour.

General statistics

The following tables summarise high-level statistics concerning B2 and B3 game categories and the transitions between them. An intuitive definition of what constitutes a transition between B2 and B3 content is self-explanatory. The detailed discussion of how precisely it is defined in practice is provided in the sections to follow.

Game category	Count of bets	Proportion of bets	Average stake size	Average number of B2 to B3 transitions	Average number of B3 to B2 transitions	Average money type[7]
B2	9 230 740	26.42%	£ 8.17	1.13%	0.00%	0.37
B3	25 704 001	73.58%	£ 0.68	0.00%	0.43%	0.32

Table 11: Statistics of the B2 and B3 categories on the game level

First of all, almost 75% of all bets belong to the B3 category. They are the bets with a maximum stake size of £2 so they have a much lower average stake size. When playing a B2 game, a player will switch to a B3 game on average once every 88 bets. The opposite transition, when playing B3 games, happens once every 233 bets. One way of measuring how successful players are is to look at how often they are playing with winnings as opposed to their own money (which implies that they have won more money than they have staked so far during the session). That happens slightly more often for B2 category games.

[7] Average money type refers to playing with winnings (1) as opposed to playing with the player's own money (0). The closer the average value is to 1, the more often the player plays with winnings (has won more than they have put at stake).

The two tables below show the same statistics for players put in one of the three categories: 'Most B2', 'Mixed' and 'Most B3'. A player is categorised as 'Most B2' or 'Most B3' if over 90% of all their bets belong to one of the respective categories.

Player category[8]	Game category	Count of bets	Proportion of bets	Average stake size	Average number of B2 to B3 transitions	Average number of B3 to B2 transitions	Average money type
Most B2	B2	2 522 848	7.22%	£ 10.00	0.13%	0.00%	0.40
Most B2	B3	105 582	0.30%	£ 1.21	0.00%	3.11%	0.32
Mixed	B2	6 381 550	18.27%	£ 7.35	1.24%	0.00%	0.35
Mixed	B3	9 544 940	27.32%	£ 0.72	0.00%	0.86%	0.32
Most B3	B2	326 342	0.93%	£ 10.17	6.64%	0.00%	0.40
Most B3	B3	16 053 476	45.95%	£ 0.65	0.00%	0.15%	0.32

Table 12: Statistics of the B2 and B3 categories jointly on the game and player level

Player category	Count of bets	Proportion of bets	Average stake size	Average number of B2 to B3 transitions	Average number of B3 to B2 transitions	Average money type
Most B2	2 628 430	7.52%	£ 9.65	0.13%	0.12%	0.40
Mixed	15 926 490	45.59%	£ 3.37	0.50%	0.52%	0.33
Most B3	16 379 821	46.89%	£ 0.84	0.13%	0.15%	0.32

Table 13: Statistics of the B2 and B3 categories on the player level

Finally, the table below shows that there is very little correlation between the proportion of B2 or B3 games played and the probability of being a problem gambler. It is only marginally higher for the 'Mixed' category players. This finding is in line with what has been found before about problem gamblers: their behaviour tends to be more 'chaotic' and diverse. Please refer to Chapter 1 on research question 5 for more details.

Please note the difference between 'Player category' and 'Normalised player category'. In the former case a player as categorised as 'Most B2' if over 90% of all of their bets are of B2 types. In the latter case, if it is the case for over 90% of their B2 or B3 bets. The slight differences in the statistics stem from the fact that even though over 94.5% of all bets are B2 or B3, there are some players for whom other bets are a significant proportion of all. The statistics in the tables above were provided using the simple (not normalised) player category.

[8] Most B2 – over 90% of the player's bets are B2 category. Most B3 – over 90% of the player's bets are B3 category.

Player category	Number of players	Average PG score	Proportion of problem gamblers
Most B2	1 220	4.74	23%
Mixed	1 936	5.08	26%
Most B3	823	4.56	22%
Undefined[9]	9	3.11	11%

Table 14: PG score statistics based on not normalised player category

Normalised player category	Number of players	Average PG score	Proportion of problem gamblers
Most B2	1 284	4.68	22%
Mixed	1 851	5.15	26%
Most B3	844	4.56	22%
Undefined	9	3.11	11%

Table 15: PG score statistics based on normalised player category

Defining a game-type transition

Over 94.5% of all events are labelled as corresponding to the B2 or B3 stakes category. Most of the other events refer to loading money which in principle does not have any intrinsic category attached to it. Therefore, it is safe to assume that almost all events are either B2 or B3 category.

As far as *Play* events only are concerned, B2 are 26.4% of those and B3 are 73.6%.

The most important definition is that of a transition between B2 and B3 content. There are two extremes possible when looking at this issue:

- Every B2/B3 switch should be counted. In that case every *Play* event of a type different from its following *Play* event would be counted as a transition.
- Very brief, incidental periods of different content should not be counted. In that case a transition would only occur when the player is permanently (that is for a long enough period or until the end of a session) switching from B2 to B3 or the other way round.

It has been concluded that both extremes should be avoided. A moving average filter has been implemented to give a score to every event indicating *the extent to which* it is B2 or B3[10]. It is calculated by looking at events preceding and following it. It takes on values between 0 (B2 event) and 1 (B3 event). Each event is categorised as B2 if its score is below 0.5 and B3 if above. As expected, in a vast majority of cases, these labels match the true game category labels. Examples below:

[9] The 'Undefined' type refers to players whose all bets were neither B2 nor B3 or the game type was not provided in the data.

[10] According to this scheme, an event is labelled as B2 if more than 50% of the set of events including k events preceding it, the event itself and k event following it are of true game category B2. Such a scheme can be described succinctly as smoothing using a moving average filter of length $2k + 1$.

True game category	True Score	Smoothed score	Smoothed game category
B2	0	0	B2
B2	0	0.2	B2
B2	0	0.333333	B2
B2	0	0.428571	B2
B3	1	0.571429	B3
B3	1	0.714286	B3
B3	1	0.857143	B3
B3	1	1	B3

Table 16: Game category data – one-point transition example

True game category	True Score	Smoothed score	Smoothed game category
B2	0	0.25	B2
B2	0	0.4	B2
B3	1	0.333333	B2
B2	0	0.428571	B2
B3	1	0.571429	B3
B2	0	0.714286	B3
B3	1	0.714286	B3
B3	1	0.857143	B3
B3	1	0.857143	B3
B3	1	1	B3

Table 17: Game category data – mixed transition example

The two simplified case studies show how the smoothing does not affect the location of the transition boundary in the case of a one-off, permanent game category switch (which is the most common scenario). The second example shows how smoothing avoids overestimating the number of transitions.

Alongside the event-level transition, a session-level transition has also been defined. A game category label for a session is constructed by averaging the labels of all its *Play* events and setting it to B2 if the results is less than 0.5 or to B3 if otherwise. This completes the definitions of event- and session-level transitions.

The choice of filter length

As mentioned before, the moving average filter used is of length 7 (taking into account the 3 preceding *Play* events, the current event and the 3 following *Play* events). Below are the results of experiments which have led to that choice of filter length by checking how it affects the overall number of transitions (averages are calculated over all players).

Filter length	Event-level				Session-level			
	B2 to B3		B3 to B2		B2 to B3		B3 to B2	
	Average	Total	Average	Total	Average	Total	Average	Total
1	27.53	109656	28.80	114697	4.67	18601	4.62	18384
3	20.43	81380	21.81	86879	4.67	18605	4.62	18383
5	18.17	72375	19.14	76252	4.67	18620	4.62	18394
7	16.87	67178	17.91	71319	4.68	18621	4.62	18396
9	15.56	61959	16.45	65524	4.68	18643	4.62	18419

Table 18: Number of event- and session-level transitions (average per player value and total) as a function of the length of the moving filter used

As a reminder, the choice of filter length was aimed at reducing the effect of short-lived game category transitions which unnecessarily overestimate the numbers. The results above show how increasing the filter length reduces the number of transitions but the effect gradually saturates. It is expected that there would be a relatively large number of short-lived B2 or B3 periods. That would mean that when a player switches from one game category to another they will either switch back almost immediately (the switch is not what the player really wanted in the long run) or keep on playing the game they switched to for a while (because it is what they consciously decided to do). The histogram below shows this quantitatively as the number of game category periods of length 1 or 2 are exceptionally high.

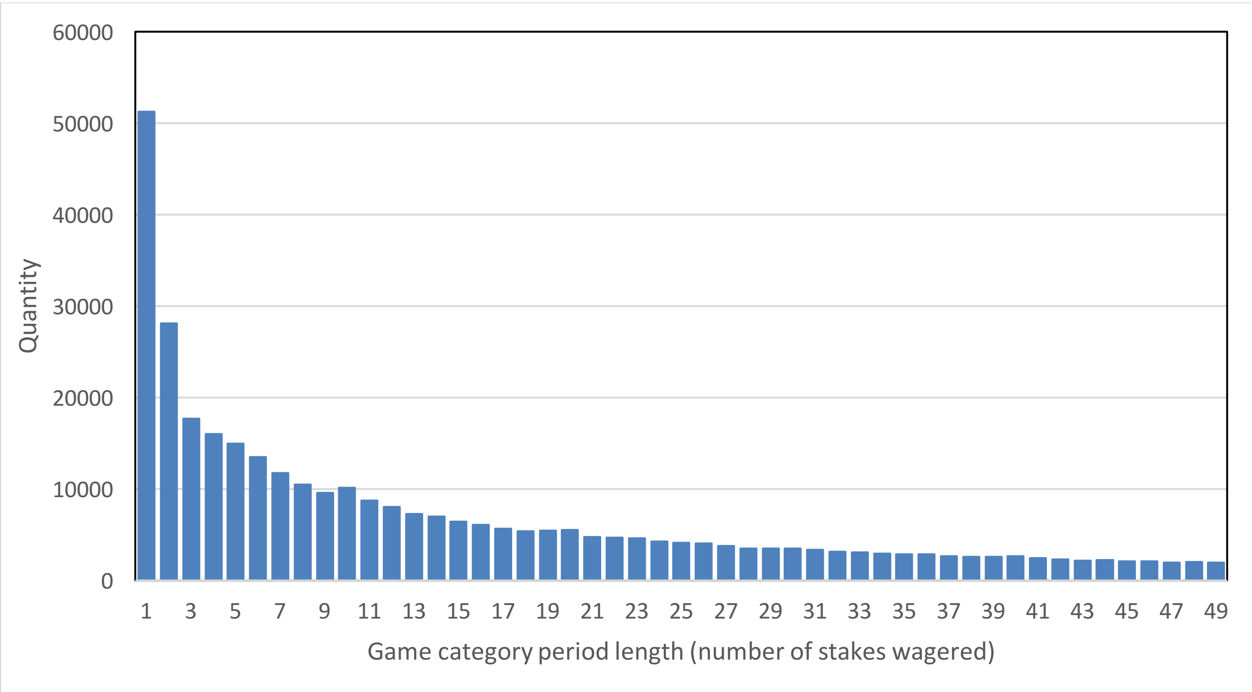


Figure 11: Distribution (histogram) of (B2 or B3) constant game category period lengths (in terms of the number of bets made)

It is to be noted that the choice of filter length 7 is not caused by an assumption that short-lived game category transitions are meaningless. In fact, their predictive value in terms of problem and non-problem gamblers will be examined. The filter of length 7 helps by ignoring many periods of lengths 1, 2 and 3 (if they appear in isolation, please refer to the visual description of the workings of the filter above).

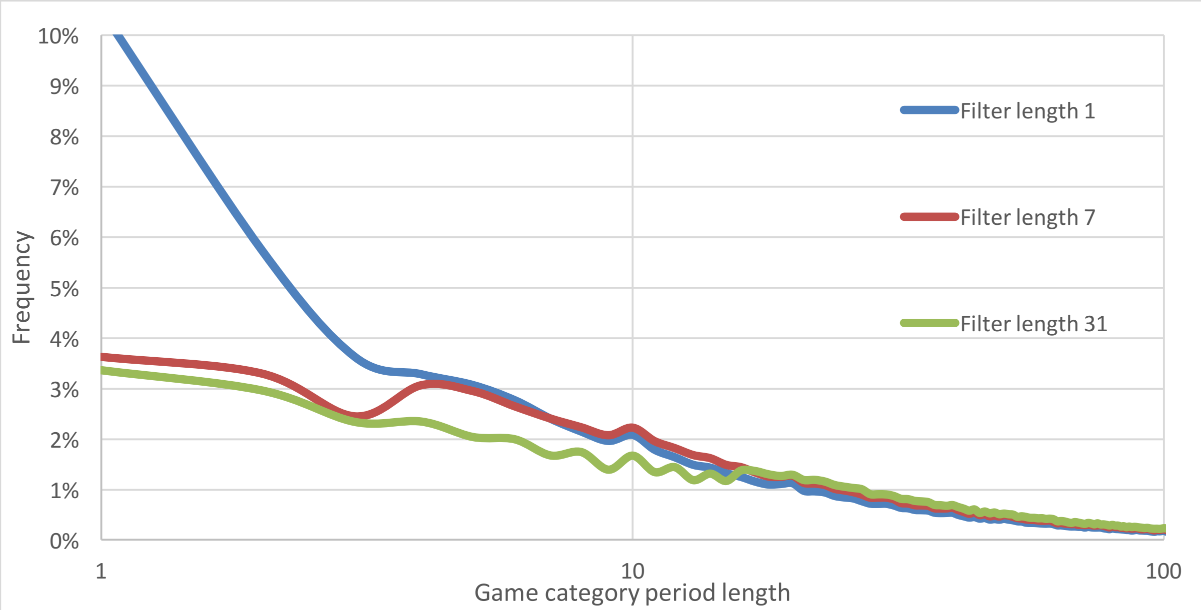


Figure 12: Comparison of distributions of (B2 or B3) constant game category period lengths for different moving average filter lengths used

The graph above shows four distributions of game category period lengths for moving average filter lengths of 1, 5, 7 and 31. As expected, the filter of length 7 significantly reduces the number of very short game category periods compared with the filter of length 1 (most of the reduction is already done when the filter length is 5). What is also shown above is that further increases in the length of the filter do not change the distribution much.

Event and session labels constructed as explained above are used to define player transitions between B2 and B3 content. This allows for tracking transitions on event as well as on session level.

Feature Engineering

This section looks at what behavioural metrics can be measured in regards to players switching content types and how effective they are at discriminating between problem and non-problem gamblers.

Metrics

The following metrics have been calculated for players (please note that the 'game category score' is the smoothed value calculated to identify switches between B2 and B3 content):

- *PlaAvgSesAvgTrans* – average session event-level transition rate (average number of transitions per event)
- *PlaStdSesAvgTrans* – standard deviation of session average event-level transition rates
- *PlaAvgNoOfTrans* – average number of event-level transitions per session
- *PlaTotalNoOfTrans* – total number of event-level transitions
- *PlaStdNoOfTrans* – standard deviation of the number of event-level transitions per session
- *PlaMinSesAvgGameCatSmooth* – minimum of session averages of event-level game category scores

- *PlaMaxSesAvgGameCatSmooth* – maximum of session averages of event-level game category scores
- *PlaAvgSesAvgGameCatSmooth* – average of session averages of event-level game category scores
- *PlaStdSesAvgGameCatSmooth* – standard deviation of session averages of event-level game category scores
- *PlaMinSesGameCatSmooth* – minimum of session-level game category scores
- *PlaMaxSesGameCatSmooth* – maximum of session-level game category scores
- *PlaAvgSesGameCatSmooth* – average session-level game category score
- *PlaStdSesGameCatSmooth* – standard deviation of session-level game category scores
- *PlaTotalNoOfSesTrans* – total number of session-level transitions
- *PlaAvgNoOfSesTrans* – average number of session-level transitions per session

Metrics useful at discriminating between problem and non-problem gamblers

All of the above metrics have been calculated on the player level. It is informative to check if any of the above features are useful when it comes to differentiating between problem and non-problem gamblers. For that reason, the above features have been added to the existing problem gambler classifier as was described in the previous chapter.

As explained in the aforementioned chapter, the features that dominate the decision of whether a player is a problem gambler are diversity (or volatility) features. They are not the actual values of, for example, money spent or loaded but the numbers of their different levels or standard deviations. It has been shown that problem gamblers' behaviour is more chaotic in that their values of money spent and loaded vary more and span wider ranges of numbers.

The two transition-related features that were among the highest importance were *PlaAvgSesAvgTrans* (average number of session-level transitions per session) and, particularly high – *PlaStdSesAvgTrans* (standard deviation of session average event-level transition rates). The two graphs below show the distribution of the values of the two features across players as a function of the *PGSCORE*. Especially when it comes to the second graph (standard deviation of players' session averages of transition rates) there is a tendency for the values to be on average higher for problem than non-problem gamblers. The tendency, however, is not strong, therefore the conclusion is that transition-related features by themselves are not very good indicators of problem gambling. Consequently, their usefulness lies not in their individual predictive capability but in the fact that they describe a different aspect of a player's behaviour that is fairly uncorrelated with other features, while exhibiting some discriminatory capabilities. The features have been added to the model but no uplift in performance has been observed.

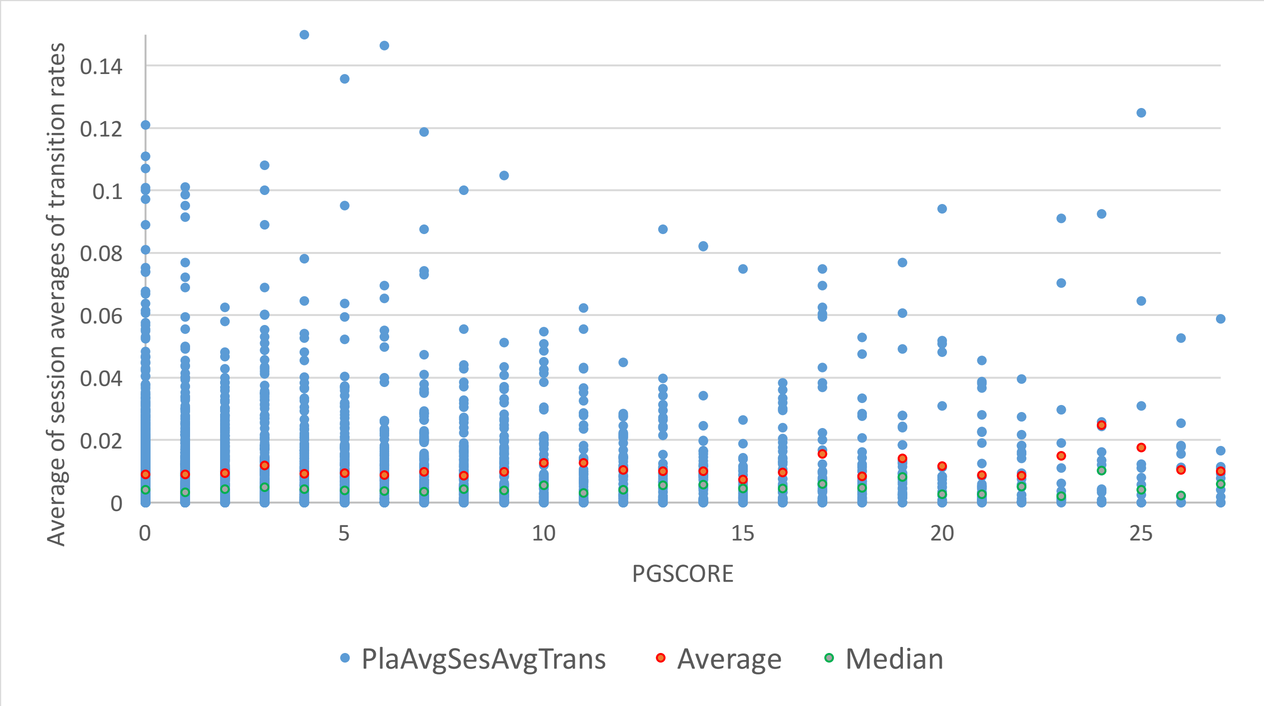


Figure 13: Distribution of a player's PlaAvgSesAvgTrans feature value (as explained below) *at different PG score levels*

PlaAvgSesAvgTrans: average session event-level transition rate (average number of transitions per event)

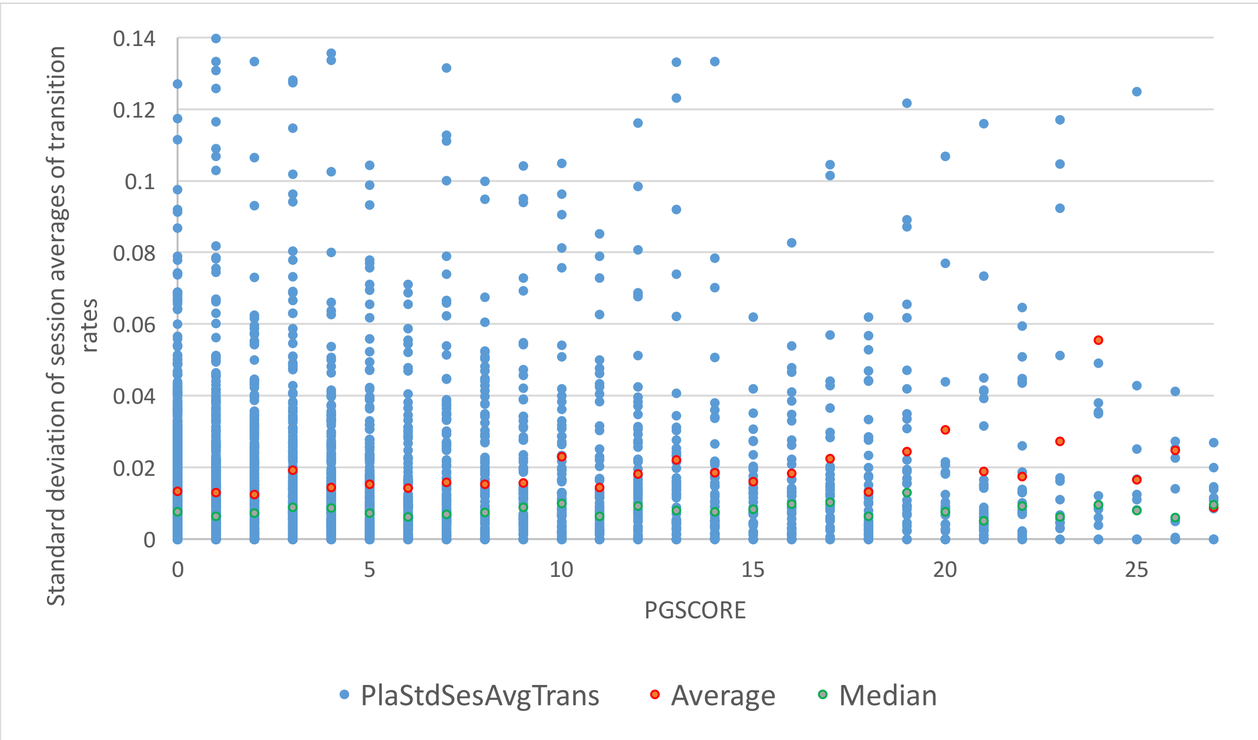


Figure 14: Distribution of a player's PlaStdSesAvgTrans feature value (as explained below) *at different PG score levels*

PlaStdSesAvgTrans: standard deviation of session average event-level transition rates

Event-level B2/B3 transition prediction

Introduction

In order to determine what aspects of players' behaviour are good, predictors of transitions between B2 and B3 games classifiers were built. Four classifiers were trained and achieved the following results in terms of the area under the ROC curve:

- B2->B3 classifier trained on a large number of B2 events
    - Based on filter length 1: AUC = 0.8277
    - Based on filter length 7: AUC = 0.8726
- B3->B2 classifier trained on a large number of B3 events
    - Based on filter length 1: AUC = 0.8161
    - Based on filter length 7: AUC = 0.8299

All the features have been defined on the event level so they are either referring to the particular event or based on all events during the current session up to that point. Therefore, for example, *Hour* refers to when the current event took place whereas *Standard deviation of money spent* refers to how spread out the stake values have been so far during the session. That way the prediction is only based on the current session and all the player history outside of that session is discarded.

Performance

First of all, the performance of transition classifiers is much better than of any realisation of the problem gambler classifier. Players' switching between game types is much more predictable than the survey-based problem/non-problem gambler labels. When a filter length of 1 is used, 80% of the B3 -> B2 transitions can be correctly identified with a false positive rate of 20%. For the B2 -> B3 transitions 80% can be correctly identified for a 30% false positive rate.

Most important features

Feature importance is a metric which can be calculated for a trained model and it indicates how useful a particular feature is when it comes to performing classification (in this case, for example, predicting whether the player will switch to B3 on the next 'Play' event). They can be used to describe a behaviour pattern which indicates when a player who is playing B2 is likely to switch to B3.

The feature that was consistently achieving the highest importance score (across all 4 classifier types) was the standard deviation of money spent so far in the session. What this means is that when it comes to predicting whether a player will switch from one game category to another, what matters most is not really the values of stakes wagered but how spread out (inconsistent or chaotic) they were.

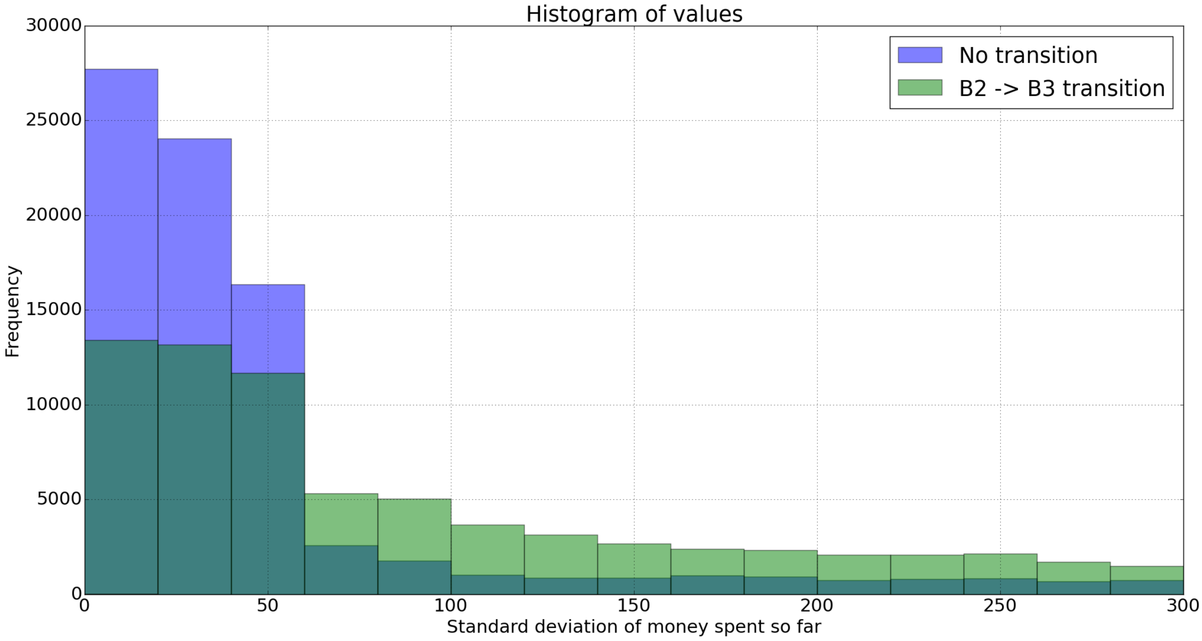


Figure 15: Distribution (histogram) of standard deviation of money spent for B2 to B3 transition events and no transition events (based on a balanced dataset)

The histogram above shows the distribution of the standard deviation of money spent during events for B2 -> B3 transitions. The blue columns indicate no transition and the green indicates when a transition took place. In this balanced dataset, blue events are a majority when standard deviation is below 60 and a minority when above[11]. As far as the real (imbalanced) data is concerned, such simple thresholding would only be enough to divide events into two groups for one of which a B2 -> B3 transition is much more likely to occur.

A histogram for B3 -> B2 transitions is presented below compared to from B2 -> B3 that was shown above. A similar though less extreme effect can be observed.

[11] The dataset is balanced as a result of the data being sampled accordingly so that the proportions of transition and non-transition events are equal. When it comes to the real data, as an example, a B2 -> B3 transition happens on average 6-7 times per 1000 events.

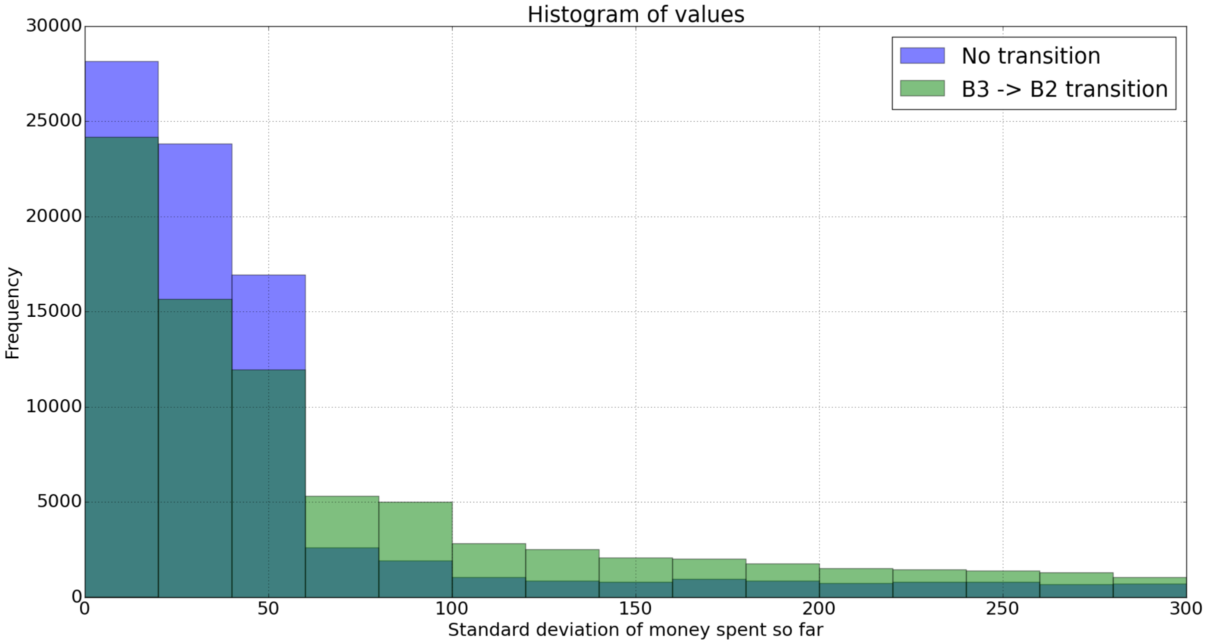


Figure 16: Distribution (histogram) of standard deviation of money spent for B3 to B2 transition events and no transition events (based on a balanced dataset)

Other important features found include the following:

- Number of different stake levels – higher for transition events
- Amount of money loaded so far into the machine – higher for transition events
- Moving average of money spent – higher for transition events
- Current balance – higher for transition events
- Standard deviation of balance – higher for transition events

Therefore, the high-level conclusion is that B2/B3 transitions are more likely to happen at later stages of sessions, especially when many different stake levels have already been used and when the balance has already taken a wide range of values. Some other circumstances when B2/B3 transitions are slightly more likely to occur include situations when large amounts of money have been loaded into the machine and spent recently or when the current balance is high.

Session-level B2/B3 transition prediction

Introduction

Another classifier has been built to classify B2/B3 transitions on a session level. As explained before, a session is labelled as B2 if the majority of its events have been labelled as B2. The aim was to find characteristics of a player's current session that are good predictors of switching to a different game category in the next session.

Results and feature importance

The classifier was able to achieve reasonably good performance (an AUC of 0.795 was obtained) however, little valuable insights were found analysing the most descriptive features. The most relevant feature in terms of predicting whether the player switches to a different game category in the next session turned out to be its time duration.

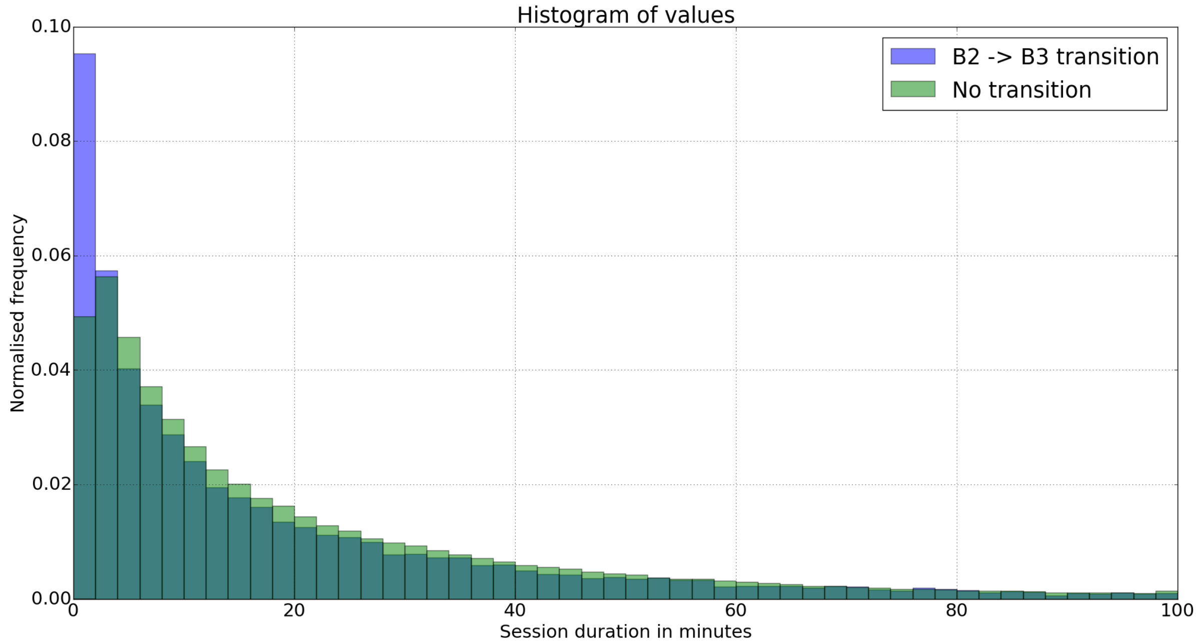


Figure 17: Normalised distribution (histogram) of session duration in minutes for B2 to B3 transition events and no transition events (based on a balanced dataset)

The histogram above shows that even the most relevant feature alone is not very effective when it comes to predicting game category transitions on the session level. Transitions are likely to be preceded by very short sessions. When it comes to longer sessions, they are only marginally less probable. Therefore, effective prediction of transitions can only occur when other features are also taken into account and no insightful conclusions regarding one or two features can be drawn.

Mixed B2/B3 game types

Summary of game types and their meanings

The difference between B2 and B3 games is the maximum stake allowed and the time between stakes: £100 and 30 seconds in the case of B2 and £2 and 2 seconds in the case of B3. The whole idea behind analysing B2/B3 transitions was to capture when and under what circumstances players choose to switch from lower to higher maximum stake size games, or the opposite.

That logic, however, is not strictly correct for all cases because of *hybrid* games.

True game type[12]	Game category	Count	Proportion
Slots (Hybrid)	B3	10 580 288	29%
B3 Slots	B3	10 088 633	28%
Slots	B3	8 162 367	22%
Hybrid Slots	B2	4 797 525	13%
Roulette	B2	3 327 806	9%
Standard Roulette	B2	2 581 596	7%
NULL	NULL	1 392 966	4%
NULL	B3	1 195 110	3%
Poker	B3	830 819	2%
Premium Roulette	B2	787 559	2%
Roulette Feature	B2	747 731	2%
Blackjack	B2	651 730	2%
Hybrid Slots	NULL	463 387	1%
Other	B2	347 416	1%

Table 19: Game category statistics for different game types

Hybrid games are a mixture of B2 and B3 types. In practice, what is of particular interest is a hybrid B3 game where every now and then it is possible to wager a stake higher than £2. That corresponds to the top *True game type* above – *Slots (Hybrid)*. Even though it is a B3 type game, on average 1 in over a 100 bets is higher than £2. What this means is that a player playing this type of game can occasionally be given an opportunity to place a higher bet. This allows for investigating a different aspect of player behaviour: whether when playing a low stake game they will be willing to make an occasional higher stake bet.

The hybrid B3 game

The focus of the analysis in this section is the hybrid B3 game which has an upper bound on the stake size of £2 but every now and then the player has a chance of placing a higher stake bet. These will be treated as special events and the analysis will aim at discovering under what circumstances the players are most likely to take them.

Observations

1.2% of all *Play* events that belong to game category B3, game type *Slots (Hybrid)* are stakes over £2. That gives a total of 109 388 such stakes. It has, however, been discovered that an overwhelming majority of such events are contained within a relatively small number of distinct sessions by a small proportion of players[13]. That group of players has the same proportion as the rest of the population for the number of problem gamblers, therefore, it is not a good indicator or marker of harm.

Even though there were over 60 000 sessions of interest (containing events of game type *Slots (Hybrid)* and which have happened between 01/09/2013 and 01/07/2014 from the 4,000 surveyed customers), only slightly more than 7 000 of them have had any bets of over £2. What is even more significant is that 60% of such bets have all been made in only 1,000 out of the 60,000 sessions.

[12] True game type names come directly from the data provided.

[13] Only 47% of players have ever placed a 'hybrid' bet. 28% of them have done it more than 10 times and 11% more than 100 times. Finally, 1.4% place such bets often – more than 1000 times in total.

The conclusion is that over £2 bets within B3 games do not happen in isolation most of the time. Instead, players (the vast majority of them) either refrain from using them completely or, when they do, spend a significant proportion of their session on them.

The visualisation below shows a typical session of a player who chooses hybrid stakes. The player's behaviour confirms the tendency that hybrid stakes do not tend to appear in isolation. The player was doing well at the beginning of the session – playing with winnings (the green colour). Later on, they lost their winnings and started playing with their own money. At that point, the player loaded more money into the machine and started using hybrid bets only.

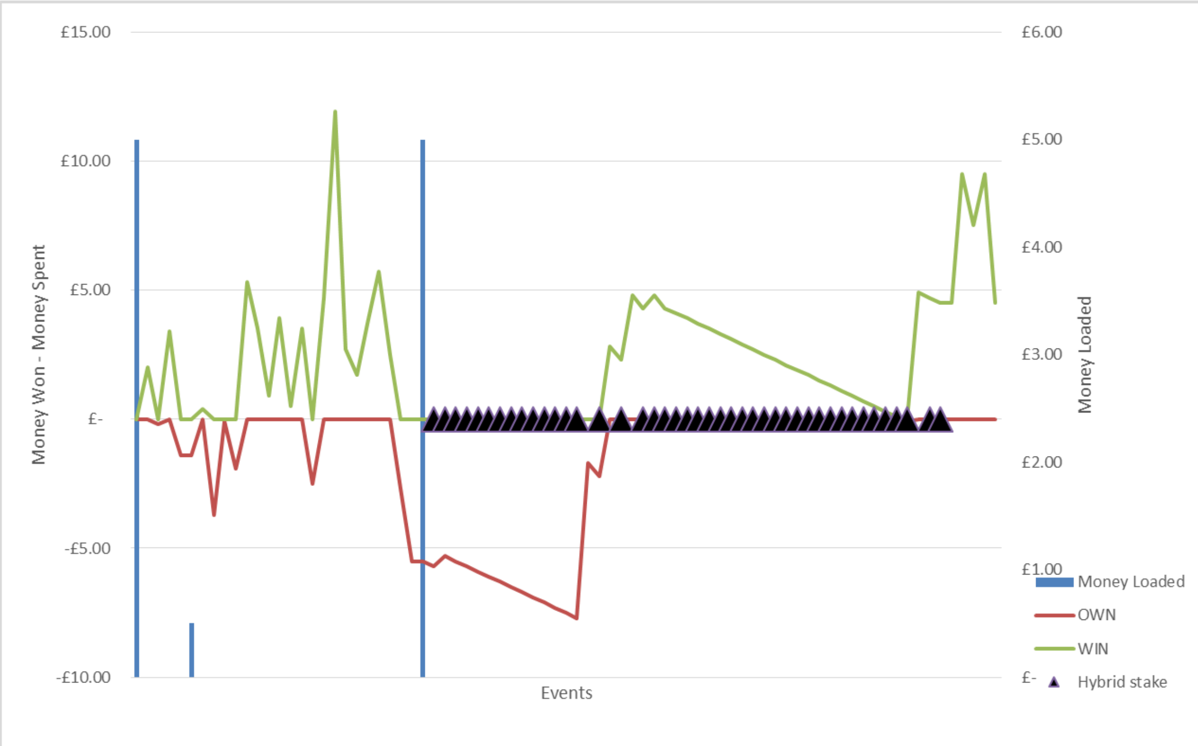


Figure 18: Example of a session including hybrid stakes

The session below is another example of using a number of hybrid stakes in a row. The player started off with hybrid stakes only. Then there was a brief period of using regular stakes which result in the player winning back lost money. Later on, however, they have lost all their winnings and continued using hybrid stakes until the end of the session.

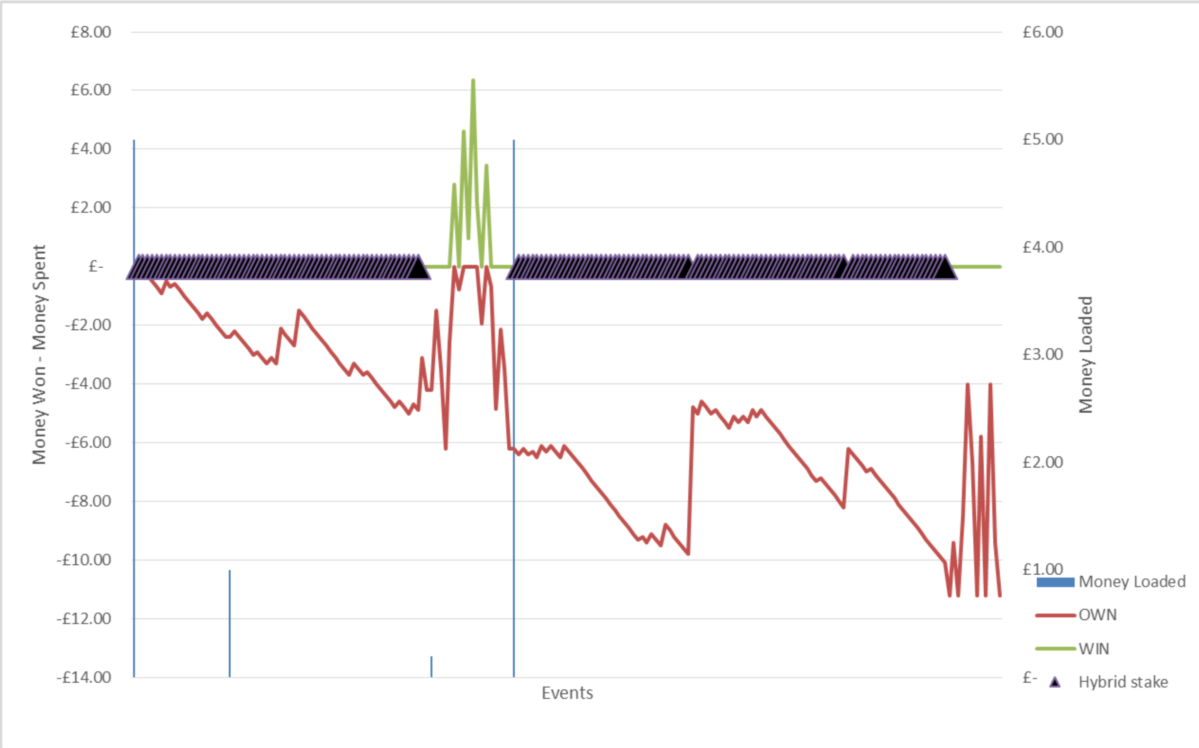


Figure 19: Example of a session including hybrid stakes

Conclusion

Transitions between B2 and B3 bets as well as the use of over £2 bets during B3 games were analysed. What was found was that most aspects of the transition-related player behaviour are not useful when it comes to differentiating between problem and non-problem gamblers. The standard deviation of the numbers of transitions per session and their average were the two reasonably informative features found.

Transitions between B2/B3 content as well as the use of over £2 bets when playing B3 games proved to be more predictable than problem/non-problem gambler labels. It has been found that B2/B3 transitions are more likely to occur later on in a session – when the player has already placed a number of bets and, importantly, when they have tried a range of different stake sizes.

It has been found that the *hybrid* over £2 bets with B3 games occur rarely and there is only a small proportion of players (proportionally spread between problem and non-problem gamblers) who ever make use of them. Therefore, the hybrid bets are almost never a one-off event that happens once or twice during a session. Instead, players either refrain from using such bets altogether or they constitute a significant proportion of their activity.

Chapter 3: Research Question 8

What further descriptive data can be extracted about the £100 stake?

High-Level Findings

Introduction

The aim of the report is to analyse players' behaviour when it comes to the **highest possible stake of £100**. Statistics are provided to characterise it as well as some insights into what the use of the maximum stake can tell about the player's gambling behaviour and under what circumstances it occurs. The analysis is based on data gathered for the 4,000 players who took part in the survey from the Gambling Machines research project funded by the RGT in 2014.

General statistics

First of all, it is important to point out that £100 stakes are extremely **rare** – only 0.26% of all stakes that is one in every 388 bets. Over 96% of all such stakes have been recorded to be roulette bets. The distribution of such bets is unusual in that they are very strongly focused on a **small proportion of sessions**. They rarely take place in isolation – it is relatively uncommon for a session to contain only a few maximum value bets. Most of the sessions which do contain £100 bets contain a higher number of them (over 80% of such sessions contain more than one £100 maximum value bet). These characteristics are summarised in the following statistics:

- 10% of all players are responsible for over 52% of all £100 stakes
- All £100 stakes have been wagered in 2% of all sessions
- 50% of £100 stakes have been wagered in as few as 450 out of nearly 280 000 sessions (0.16%)

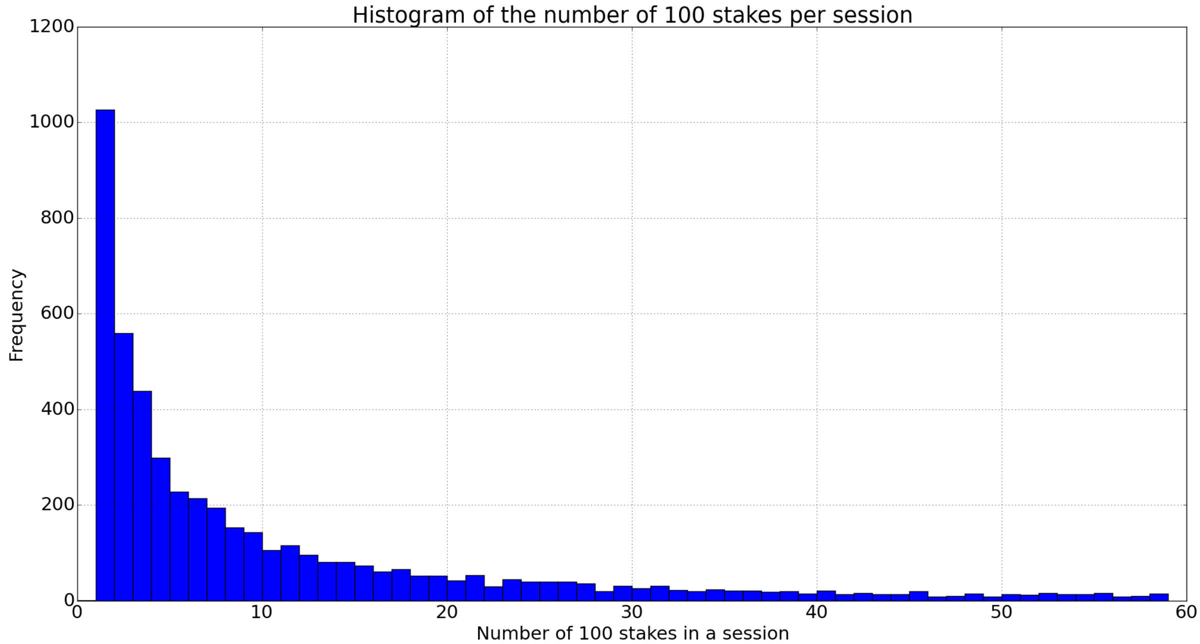


Figure 20: Distribution (histogram) of the number of £100 stakes per session

The histogram above shows the distribution of the numbers of £100 stakes per session (the current analysis only concerns sessions that do contain at least one £100 stake[14] and the frequency on the Y-axis is the number of such sessions which contained the particular number of £100 stakes). As expected, sessions with more £100 stakes are less frequent but the frequency values are decaying relatively slowly. In particular, only 19.6% of sessions (1026 out of the 5240 sessions which contain any £100 stakes) contain exactly one £100 stake. Therefore, it is wrong to interpret £100 stakes as occasional occurrences that happen extremely rarely and in isolation. In fact, **only 5.7% of all £100 stakes have not been preceded by another £100 stake in the current session** and **only 1.1% were the only £100 stake in its session**.

These statistics show how the usage of £100 stakes should be interpreted. Occasions when a player all of a sudden decides to interrupt their normal course of play and throw in a single £100 stake in the session are unusual. A more accurate interpretation would be one where certain players happen to have sessions during which they gradually increase their bet values and eventually wager several £100 stakes. Altogether, 26% of the surveyed players placed at least 1 £100 stake.

The table below summarises some general statistics differentiating players who have (26% of all) and who have not (74% of all) ever made £100 bets.

	£100 players (26%)	Non-£100 players (74%)
Average return rate	-0.006	0.030
Average number of sessions	120.7	52.2
Average total loss (£)	1,857.00	524.46
Average session duration (minutes)	26.9	20.0
Average proportion of events played with winnings	34.8%	31.8%
Average number of wins per bet	0.38	0.34

Figure 21: Comparison of high-level statistics between player who have ever used a £100 stake and those who have not

The average return is comparable, slightly lower for the first group. Players who have made £100 bets have on average had over twice as many sessions as those who have not – they are a lot more frequent players. The average total loss of the former group is much higher – £1857 on average, over the period between September 2013 and July 2014. Their sessions are also slightly longer on average – almost 27 minutes compared with 20 for non-£100 stake players. Finally, £100-stake players play more often with winnings and have a higher average number of wins per bet. The latter should indicate that they tend to use safer bets.

Player-level statistics

Distribution of £100 stakes across players

The following statistics will focus on the 26% of players who, as mentioned above, have used the £100 stake.

[14] Sessions with no £100 stakes were ignored for clarity since they represent 98% of all data and would dwarf the information displayed in the chart.

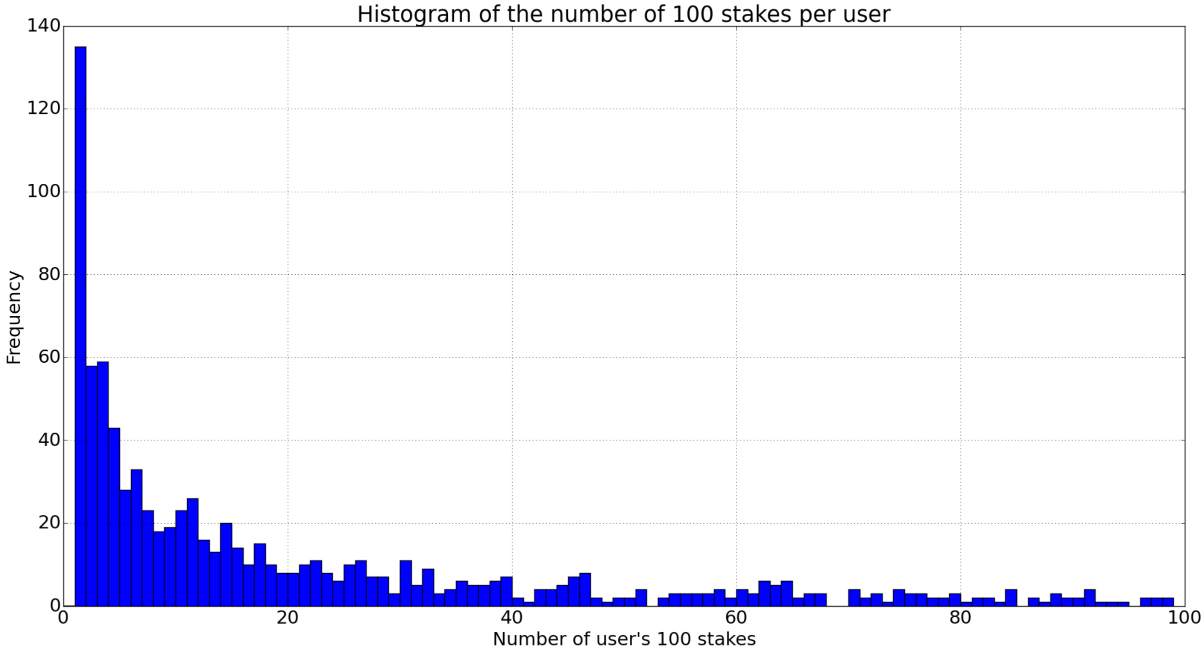


Figure 22: Distribution (histogram) of the number of £100 stakes per player

The histogram above shows the distribution of the total number of £100 stakes across players. As expected, lower total numbers of £100 stakes are more common, however, similarly to the session-level distribution, the values decay relatively slowly. As a result, there are only 175 (4.4%) players who have bet £100 more than 100 times and as few as 16 (0.4%) players who have done it over 1000 times.

Distribution of £100 stakes across problem gamblers

Number of £100 stakes	Total	Problem gamblers	Conditional proportion of problem gamblers
Any	3988	1142	28.64%
> 0	1042	249	23.90%
> 10	603	151	25.04%
> 100	172	50	29.07%
> 200	92	36	39.13%
> 500	30	13	43.33%
> 1000	16	7	43.75%

Table 20: Distribution of problem gamblers conditioned on the number of £100 stakes made

The table above shows how problem and non-problem gamblers are distributed among players who have ever wagered a £100 stake (under the condition that the number of such stakes is above a certain value). First of all, problem gamblers are 28.64% of all players contained in the dataset. When it comes to £100 players, only 23.90% of them are problem gamblers. That already shows that the £100 stake as such might not be a good problem gambler indicator. However, as the above table indicates, reducing the group of interest to players who have wagered £100 at least 100 or 200 times increases the proportion of problem gamblers it contains. Therefore, there is a slight trend for players who use the £100 stake frequently to be problem gamblers more often but, by definition, it can only be tested on a small player sample (less than 200 of them have wagered more than 100 £100 stakes). Therefore, even though the trend is clearly visible, because of the sample size it should not be trusted and, for obvious reasons, it does not generalise well enough to produce conclusions applicable to all players. The trend must be coupled with a slight tendency for problem gamblers to wager higher stakes. Please refer to Appendix B.1 for more details and a visualisation of the relationship.

When are £100 stakes being wagered?

Focus has been put on finding aspects of player behaviour which are typical of £100 stakes being wagered. It has not been straightforward for reasons mentioned before – £100 stakes are concentrated on a small number of sessions. A vast majority of such sessions contain a number of £100 stakes which by itself makes them **unique**, dominates and disguises any other potential signals which could differentiate them from the bulk of the data.

What has been found is that there are some stages of sessions for which £100 stakes are more likely than for others. In particular, **£100 stakes tend to happen at later stages of sessions**.

Please refer to Figure 23 for a histogram of which minute in a session bets other than £100 are made by players. The average value is 38.4 minutes and median 21 minutes. The histogram in Figure 24 shows the same distribution for £100 stakes. There the average value is as much as 56.2 minutes and median 37 minutes. By comparing the averages (or the histograms), it is clear that £100 bets happen at later stages of sessions.

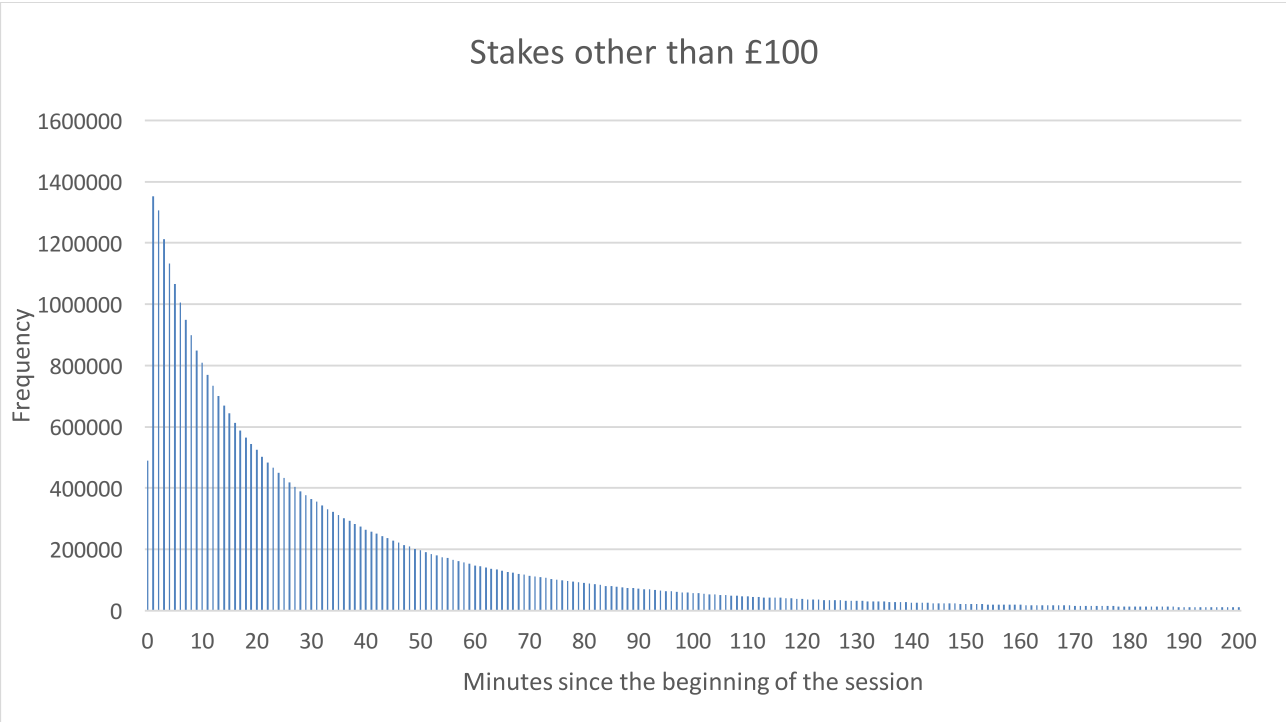


Figure 23: Distribution of which minute of a session a bet has been placed for all stakes other than £100 across all sessions. Average 38.4 minutes. Median 21 minutes.

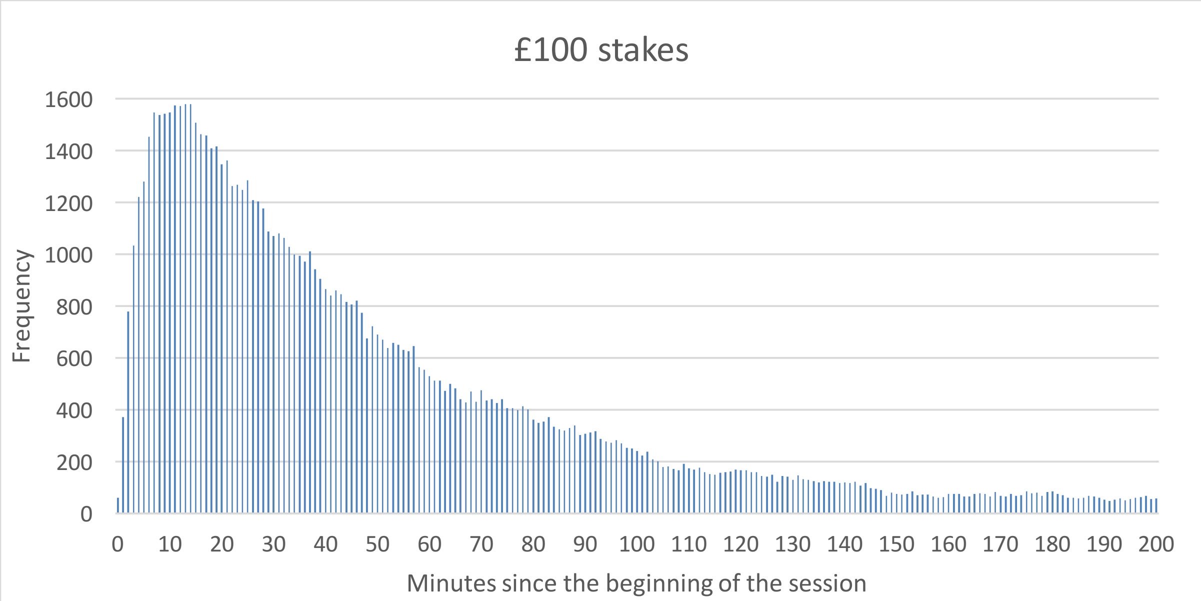


Figure 24: Distribution of which minute of a session a bet has been placed for £100 stakes. Average 56.2 minutes. Median 37 minutes.

Figure 25 and Figure 26 show which normalised position within the sessions are more likely to contain all other bets and £100 bets, respectively. In these figures each session has been normalised so that the events at the beginning, middle and end of the sessions can be compared irrespective of session length. Figure 25 confirms what would be a natural conjecture – bets (*Play* events), as being the fundamental components of every player activity, are very evenly spread throughout session lengths[15]. It is interesting to observe that the distribution of £100 bets are very different (Figure 26). Such bets happen extremely rarely at beginnings of sessions and are becoming more prevalent at their later stages.

[15] Peaks at values such as 0, 50 and 100 come naturally from the way the data has been pre-processed and should be ignored.

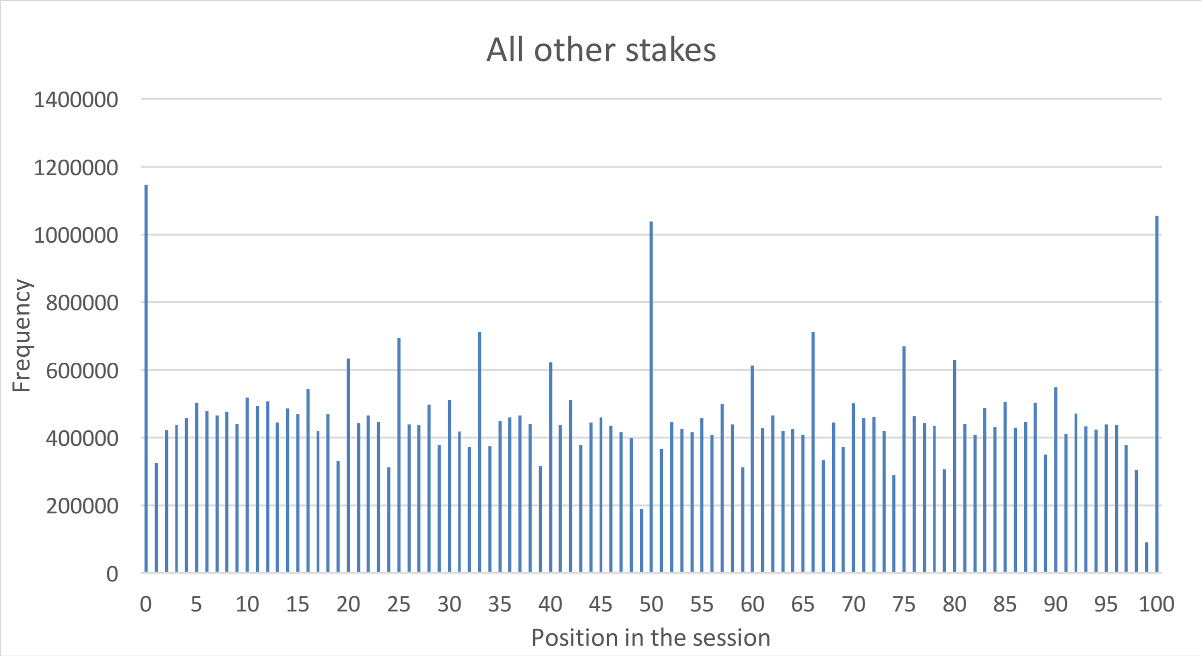


Figure 25: Time distribution of bets in a session. The shape is uniform if the noise caused by data pre-processing is ignored. This confirms the natural conjecture that bets, as the fundamental building block of sessions, are evenly spread in time in the course of the session. Please use as a point of reference for the graph in Figure 26.

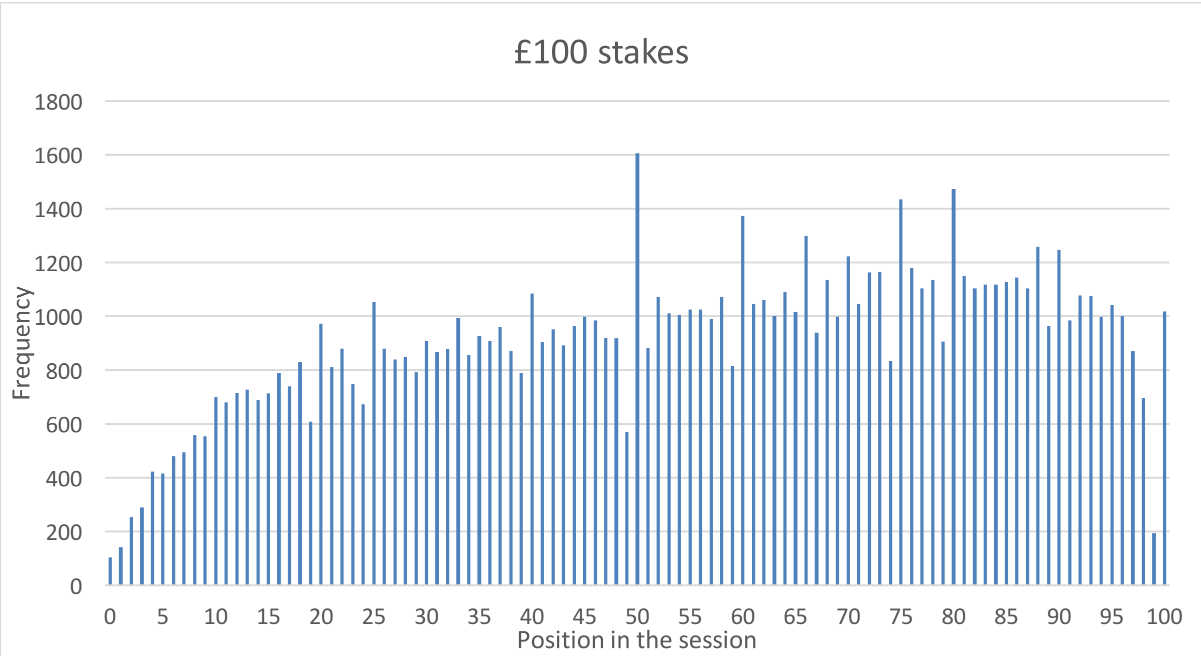


Figure 26: Time distribution of £100 bets in a session. There are more such bets at later stages of sessions.

This finding adds to the previously found conclusion. Players rarely start off sessions by going straight to placing the maximum bets. There usually is a lead-up consisting of lower bets and only then do players start betting the maximum stake, almost always more than once. Please refer to Appendix B.2 for example visualisations of such sessions.

Behaviour that leads to placing £100 bets

It is interesting to investigate what type of behaviour leads to £100 bets being placed – the behaviour that precedes them. A natural approach would be to characterise a player's current state before every bet and see what characteristics are **good predictors** of whether their next bet will be the maximum value one.

Drawing useful conclusions using such an approach would be very difficult because, as mentioned before, only 5.7% of all £100 stakes have not been directly preceded by another one. Therefore, the characteristics of a player's state which truly indicate what leads them to placing £100 bets would be **obscured** by the recently placed maximum value bets.

In order to mitigate such unwanted effects, a classifier has been trained which takes as an input characteristics of player's state before their 5th bet[16] in the session and predicts whether a £100 will be placed later in the session. As a rough indication of its performance: it was able to correctly predict that a £100 bet would happen in a session **68% of the time while raising false alarms 18% of the time**. Importantly, however, it provides information on early stage predictors for the maximum stake bets.

The following four features turned out to be the best predictors:

1. Standard deviation of money spent (bets placed) so far
2. Moving average of money spent (bets placed) so far
3. Standard deviation of balance
4. Total money loaded into the machine so far

Practical conclusions can be drawn from the respective predictors. Below are the characteristics of early-session states which are later likely to lead to £100 bets:

1. The values of bets placed **vary** a lot and span large ranges of values.
2. Values of bets, especially the most recent ones, are **high**.
3. Player's balance varies a lot and spans a large range of values. (A direct consequence of points 2 and 4)
4. The total amount of money loaded so far into the machine is **high**.

The above findings are in line with earlier conclusions. £100 bets usually do not appear unexpectedly but are rather preceded by a period of more intensive activity (higher amounts of money loaded, spent). Naturally, as a consequence, the amounts of money won during these periods would also be higher. However, the money won has been found to be a much weaker predictor of future £100 bets than money put at stake or loaded.

The typical behaviour would therefore be for a player to relatively **steadily increase their stakes** rather than suddenly jump to the maximum value bet. Please refer to session visualisations in Appendix B.2 for examples of such behaviour.

[16] Sessions where players place a £100 bet in one of the first four bets, which are a minority, were ignored.

Conclusion

In course of the research it has been found that maximum stakes of £100 are being wagered very **infrequently**. It is 26% of all players surveyed that have placed at least one maximum stake bet. They are **uniformly distributed** across problem and non-problem gamblers. However, those who have placed more 100 or more £100 stakes are more likely to be problem gamblers within the data set.

A typical £100 stake scenario is one where players place the maximum bet **several times** during a session - it is rarely an isolated, single event (please refer to Appendix B.2 for visualisation of example sessions). What has also been found is that £100 stakes happen very rarely at initial stages of sessions and become **more common at later stages**. It is the **variable and intensive activity** at early stages of sessions that often leads to £100 being placed later. They are usually preceded by **gradually increasing stake sizes** rather than appearing unexpectedly.

Chapter 4: Research Question 9

What are the differences in behaviour when players are spending wins vs loading their own new money into the machine?

High-Level Findings

Introduction

The purpose of this chapter is to investigate how gamblers' behaviour changes depending on how well they are doing. The framework that was proposed in the research question was to divide players' activity into periods when they are playing with their own money versus spending their winnings.

Playing with WIN/OWN money

When the player is spending their winnings, it is referred to as playing with **WIN** money. It happens when the player has won more than they have staked so far during the session. Alternatively, the player is playing with **OWN** money when they have spent (staked) more money than they have won so far during the session. Please refer to Figure 27 for a graph that visualises these concepts.

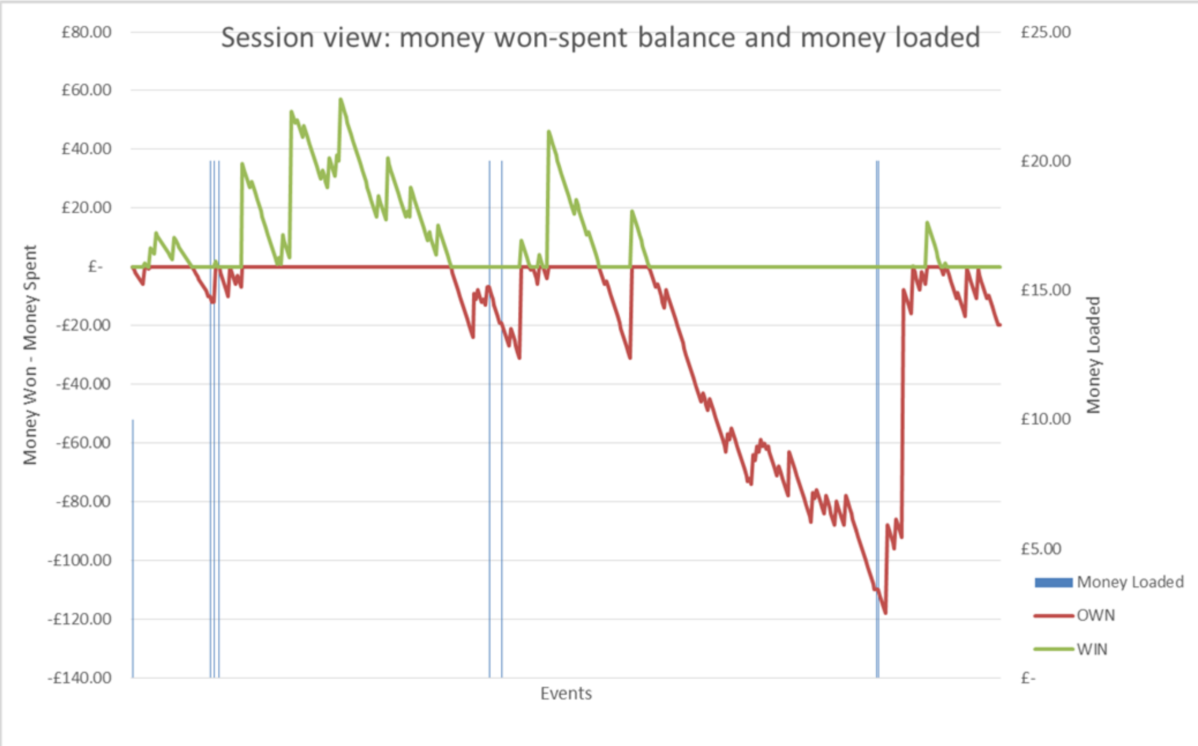


Figure 27: The visualisation below shows an example of a session by a player. The curve shows the difference between money won and money staked by the player. It is green when the player is playing with winnings (has won more than they have staked so far during the session) and red when the opposite is true.

Machine session

The sessionalisation scheme which has previously been created will be referred to as the **machine session**. It is based on a classifier trained to classify which events are most likely to be session boundaries.

Shop visit

It has been found that when using the *machine session* scheme a session boundary is on many occasions assigned to an event which should not be interpreted as one (for example a player switching from one machine to another or a player taking a break from playing and then inserting money). That is why a new, additional sessionalisation scheme was introduced which will be referred to as a **shop visit**. Improved sessionalisation is hugely important especially in terms of merging sessions that belong to the same player and time period. Playing with WIN or OWN money can be interpreted as a *state of mind* of a player which might preserve for longer periods of time.

The new sessionalisation scheme: shop visit

A new scheme has been devised which focuses on merging sessions which have been split by the previously used sessionalisation algorithm and are not in line with the shop visit definition. Session boundaries where the player stays in the same shop and takes a break no longer than 30 minutes have been removed.

The differences between the two sessionalisation schemes is the average number of machines used during a session. For the *machine session* scheme it is, as expected, equal to 1 because that scheme does not accept switching machines during sessions. Using the *shop visit* scheme, the average number of machines per session is 1.26. It is therefore clear that the new scheme includes cases where the player switches from one machine to another, while staying in the same shop, as one session.

General statistics and the effect of session length

The analysis is based on data gathered for the 4,000 players who took part in the survey from the Gambling Machines research project funded by the RGT in 2014

The machine session scheme

The following statistics have been created by averaging values over all events played with either OWN or WIN money. The *machine session* sessionalisation scheme was used.

Money type	OWN	WIN
Average money loaded	£9.41	£10.36
Average money staked	£2.14	£3.65
Average win size	£9.03	£13.07
Average return	-9.36%	-8.72%
% of wins	23.48%	26.95%

Table 21: Comparison of statistics referring to playing with players' own money (OWN) versus playing with winnings (WIN)

The table above indicates that when playing with winnings players on average load more money into the machine. Average stake sizes are higher and, as a consequence, average win sizes are higher as well. There is a slight tendency for players who play with winnings to be winning more often which would imply that they would more often choose less risky[17] bets.

[17] What is meant by less risky here is games where the probability of winning is higher. An example of that would be betting on black/red when playing roulette as opposed to betting on a single number.

The *shop visit* scheme

The following statistics have been created by averaging values over all events played with either OWN or WIN (the *shop visit* sessionalisation scheme was used).

Money type	OWN	WIN
Average money loaded	£9.19	£11.03
Average money staked	£2.28	£3.32
Average win size	£9.45	£12.23
Average return	-8.24%	-8.36%
% of wins	23.78%	26.19%

Table 22: Comparison of statistics referring to playing with players' own money (OWN) versus playing with winnings (WIN)

It is satisfactory to observe that most of the above statistics are very similar when using the *shop visit* sessionalisation scheme. There are, however, some differences which show the advantages the new scheme has introduced. When using the *shop visit* scheme, the average returns when playing with OWN and WIN money are very close to equal which is what would be expected. This is the sign that the new sessionalisation scheme is making the data less noisy and more reliable.

Machine session and shop visit compared

Therefore, it is possible to use the above statistics to appreciate the differences between the two sessionalisation schemes. The table below provides some additional metrics.

Sessionalisation	Machine session	Shop visit
OWN events	68.66%	67.17%
WIN events	31.34%	32.83%
Number of sessions	839 565	479 959
Average session length (number of events)	90	168
OWN event: average location[18] in a session	233.2	383.2
WIN event: average location in a session	230.3	379.7

Table 23: Comparison of statistics calculated using different sessionalisation schemes

The *shop visit* sessionalisation scheme nearly halves the number of sessions and, consequently, almost **doubles their average length**. That means that session boundaries inserted when the player stays in the same shop and takes a break no longer than 30 minutes have been common.

When do players play with winnings?

As shown in the table above, the average chronological event number when playing with winnings is slightly lower than when playing with own money. This puts forward a conjecture that players tend to play with **WIN money, only slightly more often, earlier on** in a session.

This, however, does not tell the full story. Please refer to the figure below for the visualisation of the relationship. What can be observed is a slight overall trend for players to be less likely to be playing with winnings as sessions go along. This is to be expected – players, on average, slowly lose money when playing. At early stages of sessions players play with winnings around 34% of the time and that proportion decreases and plateaus at close to 32% at later stages of sessions.

[18] What is meant by the event's location in a session is simply when in the chronological order it is.

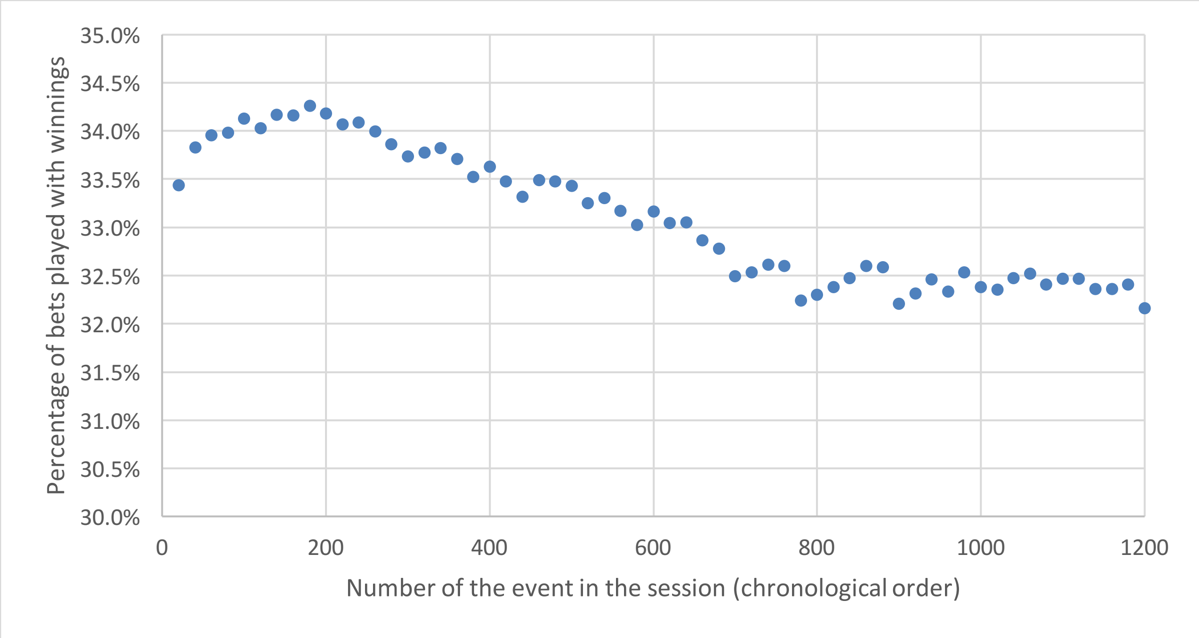


Figure 28: The graph shows the proportion of events played with winnings (as opposed to player's own money) as sessions progress. A slight downward trend can be observed. This suggests that players are slightly less likely to be playing with winnings at later stages

When do players load money?

The two tables presented at the beginning of this section contain metrics calculated using the two sessionalisation schemes. According to the *shop visit* scheme, the average amount of money loaded during WIN events increases more significantly. Consequently, the average amount of money loaded during OWN events is lower. This would suggest that **larger amounts of money are loaded at later stages of sessions**[19]. A high positive (0.48) correlation between the average amount of money loaded and the number of events so far in the session confirm that. Please refer to Appendix C.1 for a visualisation and detailed analysis of that relationship.

When do players wager higher stakes?

A similar effect, except in the reverse direction, can be observed when it comes to money spent (stakes wagered). The visualisation in Appendix C.2 confirms that on average **players wager lower stakes at later stages of sessions** and the analysis conducted outlines the causes of that relationship.

When do players win more often?

Another question worth asking is whether it is at early or late stages of sessions that players win more often. That directly reflects whether they are choosing riskier bets (for example, betting on a single number rather than on red/black when playing roulette). The graph in Appendix C.3 indicates that less risky bets are more prevalent at early stages of sessions, and more so if players are playing with winnings rather than with their own money. That might also indicate that higher risk bets are more common during longer sessions at their later stages. Short sessions would contain a relatively high proportion of low-risk bets.

---

[19] This is because, as shown before, using the *shop visit* scheme, the average session length almost doubles.

Playing with winnings vs own money based on machine session

The following statistics have been calculated for each player separately and then averaged over all players. Thanks to that approach slightly different aspects of the data are highlighted: each player has an equal influence on the end result. This is due to the fact that values are averaged over players, not over sessions and therefore players who have had many sessions do not have an overwhelmingly large influence on the end result.

Reference	Money type	OWN	WIN
1	Average time to next play (seconds)	24.75	26.34
2	Average time to next money loaded (seconds)	301.03	533.59
3	Average time to next money withdrawn (seconds)	1010.69	844.63
4	Average amount of money loaded	£8.80	£9.13
5	Average amount of money staked	£5.28	£7.06
6	As a % of balance	34.57%	21.36%
7	Average balance when money loaded	£14.05	£36.80
8	Average win size	£12.78	£15.04
9	Average return	-9.36%	-8.72%
10	% of wins	35.35%	36.58%

Table 24: Comparison of statistics referring to playing with players' own money (OWN) versus playing with winnings (WIN) using the machine session scheme

The following set of basic conclusions has been derived from the metrics:

- [Ref 1, Ref 2] *Playing with WIN money*: there is more **uninterrupted play** because players do not load money that often
- [Ref 2, Ref 4] *Playing with OWN money*: players are **loading money more often** and in slightly **smaller amounts**
- [Ref 3] *Playing with WIN money*: players are **withdrawing money more often** – please refer to Appendix C.4 for a detailed analysis
- [Ref 7] *Playing with WIN money*: the **average balance** when loading money is much **higher**
- [Ref 5] *Playing with WIN money*: players are spending more in absolute terms (**betting higher amounts of money**, staking with larger amounts) – please refer to Appendix C.5 for detailed analysis
- [Ref 6] *Playing with OWN money*: players are spending **more as a percentage of balance**
- [Ref 8] *Playing with WIN money*: average **win size** is **higher**
- [Ref 9] Average **returns** are roughly **constant** – please refer to Appendix C.6 for detailed analysis
- [Ref 10] The percentage of wins, both when playing with winnings and own money is higher when the result is averaged over all players (each player carries the same weight as opposed to averaging over all events, hence putting higher weights on players who play more). This result suggests that it is the **players who bet more** (have placed more bets in their lifetime according to the dataset being analysed) choose **higher-risk bets more often**. Please refer to Appendix C.7 for an analysis of the percentage of wins in the context of playing with winnings and players' own money.

Playing with winnings vs own money based on shop visit

For reference, this section provides equivalent metrics as the ones presented in the previous section based on the *shop visit* sessionalisation scheme. The results in both tables are very similar.

Money type	OWN	WIN
Average time to next play (seconds)	24.83	25.38
Average time to next money loaded (seconds)	313.93	491.80
Average time to next money withdrawn (seconds)	1001.29	855.79
Average amount of money loaded	£8.69	£10.20
Average amount of money spent	£5.44	£6.84
As a % of balance	33.87%	22.53%
Average balance when money loaded	£13.99	£27.04
Average win size	£13.10	£14.69
Average return	-9.34%	-8.83%
Proportion of wins	35.33%	36.23%

Table 25: Comparison of statistics referring to playing with players' own money (OWN) versus playing with winnings (WIN) using the shop visit scheme

Conclusion

The new 'shop visit' sessionalisation scheme allowed for more accurate definitions of differences in player behaviour when playing with winnings versus with their own money. It was shown that players tend to play with winnings slightly more often at the beginnings of sessions and then slowly lose money as sessions go along. Another conclusion was that at later stages of sessions, players tend to wager lower stakes but also choose higher risk bets.

When it comes to the differences between playing with winnings and with player's own money, in the former case players tend to bet higher amounts of money and withdraw money more often. In the latter case on the other hand, players are loading money more often and spending more as a percentage of balance.

Appendix A: Additional RQ5 Supporting Information

A.1

Below is the list of the top 60 most important features out of the 927 used in the classifier. The colours indicate features belonging to the same theme group (no colour indicates no theme group assigned).

1	DayAvgDayLevelsMoneySpent
2	SesAvgSesLevelsMoneySpent
3	SesStdSesLevelsMoneySpent
4	B2DayAvgDayLevelsMoneySpent
5	B2SesAvgSesLevelsMoneySpent
6	B2SesStdSesLevelsMoneySpent
7	DayStdDayLevelsMoneySpent
8	WeekAvgWeekLevelsMoneySpent
9	DayAvgDayStdMoneySpent
10	DayStdDayAvgMoneySpent
11	B2SesAvgSesStdMoneySpent
12	B2WeekAvgWeekStdMoneySpent
13	B2DayStdDayLevelsMoneySpent
14	B2MonthAvgMonthStdMoneySpent
15	B2DayAvgDayStdMoneySpent
16	WeekAvgWeekStdMoneySpent
17	MonthAvgMonthStdMoneySpent
18	DayStdDayAvgMoneyLoaded
19	SesStdSesStdMoneySpent
20	B2WeekAvgWeekLevelsMoneySpent
21	B2WeekStdWeekLevelsMoneySpent
22	DayStdDayLevelsMoneyLoaded
23	SesAvgSesStdMoneySpent
24	MonthAvgMonthLevelsMoneySpent
25	WeekStdWeekLevelsMoneySpent
26	B2DayMinDayAvgMoneySpentPerOwnMoney
27	DayAvgDayMaxDur
28	B2DayMinDayAvgMoneySpentPerBalance
29	MonthAvgMonthAvgDur
30	SesStdSesAvgMoneySpent
31	WeekStdWeekAvgMoneyLoaded
32	SesStdSesAvgMoneyLoaded
33	MonthStdMonthMinDur
34	B3SesStdSesLevelsMoneySpent
35	WeekStdWeekLevelsMoneyLoaded
36	B2PlaAvgMoneySpent
37	DayStdDayStdMoneySpent
38	B2MonthAvgMonthLevelsMoneySpent
39	B2MonthMinMonthAvgMoneySpentPerBalance
40	WeekStdWeekMinDur
41	B2SesMinSesAvgMoneySpentPerBalance

42	DayAvgDayStdMoneyLoaded
43	DayStdDayAvgLoadFTSpend
44	B3WeekStdWeekLevelsMoneySpent
45	SesMaxSesAvgBalanceWhenMoneyLoaded
46	SesAvgSesAvgWinOwnFrac
47	WeekStdWeekAvgMoneySpent
48	WeekAvgWeekAvgDur
49	B2MonthStdMonthAvgWinOwnFrac
50	WeekMinWeekAvgMoneySpentPerBalance
51	SesStdSesAvgLoadFTSpend
52	B2MonthAvgMonthAvgMoneySpent
53	WeekStdWeekAvgDur
54	MonthMaxMonthAvgWinFTSpend
55	MonthMinMonthAvgBalance
56	SesStdSesAvgBalanceWhenMoneyLoaded
57	MonthAvgMonthMinDur
58	B2WeekAvgWeekAvgMoneySpent
59	DayAvgDayAvgBalance
60	MonthMaxMonthAvgWinFTLoad

A.2

The table contains average results of 500 Logistic Regression classifier tests. Each classifier used one feature and the true positive rate can be interpreted as the feature's individual discriminative capability. The following results are provided (at a fixed threshold of 0.5):

- Average true positive rate
- Standard deviation of the true positive rate
- Normalised standard deviation of the true positive rate as a percentage of the average
- Average threat score (the number of true positives as a fraction of all samples but true negatives)
- Standard deviation of the threat score
- Normalised standard deviation of the threat score as a percentage of the average

	Feature	Avg(tpr)	Std(tpr)	NStd(tpr)	Avg(ts)	Std(ts)	NStd(ts)
1	DayAvgDayLevelsMoneySpent	5.10	1.36	27%	4.83	1.27	26%
2	SesAvgSesLevelsMoneySpent	3.52	1.19	34%	3.37	1.13	34%
3	SesStdSesLevelsMoneySpent	4.55	1.49	33%	4.36	1.40	32%
4	B2DayAvgDayLevelsMoneySpent	4.74	1.40	30%	4.50	1.33	30%
5	B2SesAvgSesLevelsMoneySpent	3.23	1.07	33%	3.11	1.02	33%
6	B2SesStdSesLevelsMoneySpent	4.44	1.38	31%	4.26	1.30	31%
7	DayStdDayLevelsMoneySpent	4.84	1.31	27%	4.63	1.25	27%
8	WeekAvgWeekLevelsMoneySpent	3.34	1.22	37%	3.20	1.15	36%
9	DayAvgDayStdMoneySpent	3.78	1.25	33%	3.57	1.17	33%
10	DayStdDayAvgMoneySpent	4.51	1.35	30%	4.29	1.27	30%
11	B2SesAvgSesStdMoneySpent	3.62	1.24	34%	3.43	1.16	34%

12	B2WeekAvgWeekStdMoneySpent	4.36	1.36	31%	4.12	1.27	31%
13	B2DayStdDayLevelsMoneySpent	4.67	1.31	28%	4.47	1.24	28%
14	B2MonthAvgMonthStdMoneySpent	4.24	1.35	32%	4.02	1.26	31%
15	B2DayAvgDayStdMoneySpent	4.24	1.31	31%	4.01	1.22	30%
16	WeekAvgWeekStdMoneySpent	4.51	1.28	28%	4.29	1.19	28%
17	MonthAvgMonthStdMoneySpent	4.49	1.34	30%	4.27	1.25	29%
18	DayStdDayAvgMoneyLoaded	0.27	0.38	141%	0.27	0.38	141%
19	SesStdSesStdMoneySpent	2.05	1.02	50%	1.97	0.97	49%
20	B2WeekAvgWeekLevelsMoneySpent	3.18	1.12	35%	3.05	1.06	35%
21	B2WeekStdWeekLevelsMoneySpent	3.40	1.16	34%	3.27	1.11	34%
22	DayStdDayLevelsMoneyLoaded	0.13	0.24	185%	0.12	0.24	200%
23	SesAvgSesStdMoneySpent	3.45	1.18	34%	3.29	1.11	34%
24	MonthAvgMonthLevelsMoneySpent	2.26	1.00	44%	2.18	0.95	44%
25	WeekStdWeekLevelsMoneySpent	3.39	1.18	35%	3.26	1.13	35%
26	B2DayMinDayAvgMoneySpentPerOwnMoney	0.03	0.11	367%	0.02	0.11	550%
27	DayAvgDayMaxDur	1.01	0.62	61%	0.99	0.61	62%
28	B2DayMinDayAvgMoneySpentPerBalance	0.00	0.00	N/A	0.00	0.00	N/A
29	MonthAvgMonthAvgDur	0.51	0.47	92%	0.50	0.46	92%
30	SesStdSesAvgMoneySpent	2.50	1.00	40%	2.40	0.95	40%
31	WeekStdWeekAvgMoneyLoaded	0.66	0.56	85%	0.65	0.55	85%
32	SesStdSesAvgMoneyLoaded	0.40	0.42	105%	0.39	0.42	108%
33	MonthStdMonthMinDur	0.34	0.36	106%	0.34	0.36	106%
34	B3SesStdSesLevelsMoneySpent	0.00	0.00	N/A	0.00	0.00	N/A
35	WeekStdWeekLevelsMoneyLoaded	0.17	0.27	159%	0.17	0.27	159%
36	B2PlaAvgMoneySpent	3.27	1.13	35%	3.15	1.08	34%
37	DayStdDayStdMoneySpent	2.07	0.94	45%	1.99	0.89	45%
38	B2MonthAvgMonthLevelsMoneySpent	2.13	0.92	43%	2.06	0.89	43%
39	B2MonthMinMonthAvgMoneySpentPerBalance	0.00	0.00	N/A	0.00	0.00	N/A
40	WeekStdWeekMinDur	0.97	0.64	66%	0.96	0.64	67%
41	B2SesMinSesAvgMoneySpentPerBalance	0.00	0.00	N/A	0.00	0.00	N/A
42	DayAvgDayStdMoneyLoaded	0.63	0.56	89%	0.62	0.55	89%
43	DayStdDayAvgLoadFTSpend	0.00	0.00	N/A	0.00	0.00	N/A
44	B3WeekStdWeekLevelsMoneySpent	0.05	0.15	300%	0.05	0.15	300%
45	SesMaxSesAvgBalanceWhenMoneyLoaded	0.39	0.42	108%	0.39	0.42	108%
46	SesAvgSesAvgWinOwnFrac	0.02	0.13	650%	0.02	0.13	650%
47	WeekStdWeekAvgMoneySpent	2.94	1.08	37%	2.83	1.02	36%
48	WeekAvgWeekAvgDur	0.61	0.54	89%	0.60	0.53	88%
49	B2MonthStdMonthAvgWinOwnFrac	0.00	0.00	N/A	0.00	0.00	N/A
50	WeekMinWeekAvgMoneySpentPerBalance	0.00	0.00	N/A	0.00	0.00	N/A
51	SesStdSesAvgLoadFTSpend	0.00	0.00	N/A	0.00	0.00	N/A
52	B2MonthAvgMonthAvgMoneySpent	2.40	0.92	38%	2.32	0.89	38%
53	WeekStdWeekAvgDur	1.16	0.65	56%	1.14	0.64	56%
54	MonthMaxMonthAvgWinFTSpend	0.00	0.00	N/A	0.00	0.00	N/A
55	MonthMinMonthAvgBalance	0.45	0.46	102%	0.44	0.45	102%
56	SesStdSesAvgBalanceWhenMoneyLoaded	0.23	0.31	135%	0.23	0.31	135%
57	MonthAvgMonthMinDur	0.16	0.26	163%	0.16	0.26	163%

58	B2WeekAvgWeekAvgMoneySpent	2.51	1.01	40%	2.44	0.97	40%
59	DayAvgDayAvgBalance	0.67	0.59	88%	0.66	0.57	86%
60	MonthMaxMonthAvgWinFTLoad	0.00	0.00	N/A	0.00	0.00	N/A

A.3

The table below shows the results of testing classifiers based on a subset of best-performing (according to the *feature importance* metric) features. The following results are provided (at a fixed threshold of 0.5):

- Average true positive rate
- Standard deviation of the true positive rate
- Normalised standard deviation of the true positive rate as a percentage of the average
- Average threat score (the number of true positives as a fraction of all samples but true negatives)
- Standard deviation of the threat score
- Normalised standard deviation of the threat score as a percentage of the average

# top features	Avg(tpr)	Std(tpr)	NStd(tpr)	Avg(ts)	Std(ts)	NStd(ts)
1	5.00	1.38	28%	4.74	1.30	27%
2	5.52	1.46	26%	5.19	1.37	26%
3	5.74	1.52	26%	5.41	1.41	26%
4	6.12	1.57	26%	5.75	1.45	25%
5	5.98	1.54	26%	5.61	1.41	25%
6	5.90	1.64	28%	5.53	1.51	27%
7	5.97	1.61	27%	5.60	1.48	26%
8	6.55	1.67	25%	6.13	1.53	25%
9	7.59	1.81	24%	6.97	1.61	23%
10	8.12	1.82	22%	7.40	1.61	22%
11	8.13	1.90	23%	7.42	1.69	23%
12	8.11	1.72	21%	7.40	1.52	21%
13	8.20	1.77	22%	7.48	1.57	21%
14	8.32	1.83	22%	7.59	1.63	21%
15	8.43	1.76	21%	7.68	1.57	20%
16	8.50	1.83	22%	7.75	1.62	21%
17	8.34	1.84	22%	7.60	1.63	21%
18	8.35	1.78	21%	7.61	1.58	21%
19	8.38	1.89	23%	7.63	1.67	22%
20	8.45	1.89	22%	7.68	1.67	22%
21	8.58	1.93	22%	7.79	1.69	22%
22	8.59	1.89	22%	7.82	1.68	21%
23	8.46	1.82	22%	7.69	1.61	21%
24	8.44	1.86	22%	7.66	1.63	21%

25	8.54	1.86	22%	7.75	1.63	21%
26	8.75	1.79	20%	7.92	1.57	20%
27	8.65	1.93	22%	7.84	1.68	21%
28	8.77	1.79	20%	7.95	1.59	20%
29	8.56	1.84	21%	7.76	1.63	21%
30	9.76	2.08	21%	8.83	1.82	21%
31	9.51	1.99	21%	8.59	1.74	20%
32	9.53	2.00	21%	8.61	1.74	20%
33	9.43	1.99	21%	8.51	1.72	20%
34	9.86	1.94	20%	8.89	1.70	19%
35	9.76	2.07	21%	8.78	1.81	21%
36	9.83	1.98	20%	8.86	1.75	20%
37	9.78	1.88	19%	8.81	1.65	19%
38	9.96	2.02	20%	8.97	1.77	20%
39	9.99	1.99	20%	8.98	1.74	19%
40	10.05	2.07	21%	9.03	1.80	20%
41	9.89	2.08	21%	8.89	1.80	20%
42	9.77	1.96	20%	8.78	1.72	20%
43	10.07	2.14	21%	9.01	1.87	21%
44	10.05	2.17	22%	9	1.88	21%
45	9.37	2.11	23%	8.43	1.84	22%
46	10.00	2.14	21%	8.98	1.84	20%
47	9.67	2.11	22%	8.68	1.84	21%
48	9.55	2.12	22%	8.58	1.81	21%
49	9.62	2.01	21%	8.63	1.74	20%
50	9.44	1.92	20%	8.49	1.67	20%
51	9.66	1.99	21%	8.66	1.73	20%
52	9.96	1.91	19%	8.93	1.65	18%
53	10.20	1.99	20%	9.13	1.74	19%
54	10.20	2.05	20%	9.12	1.77	19%
55	10.21	1.99	19%	9.13	1.73	19%
56	10.32	2.17	21%	9.21	1.89	21%
57	10.19	2.07	20%	9.13	1.79	20%
58	10.48	2.00	19%	9.37	1.74	19%
59	10.61	1.97	19%	9.53	1.71	18%
60	10.66	2.00	19%	9.54	1.73	18%

A.4

Top results (in terms of the true positive rate at a fixed threshold of 0.5) for 2-, 3-, 4- and 5-feature classifiers.

2-feature classifier		
Features		Avg(tpr)
13	23	7.01
13	15	6.99
9	13	6.97
1	14	6.96
11	13	6.91
13	16	6.87
1	10	6.84
7	9	6.80
7	16	6.80
7	23	6.79
7	15	6.78
7	11	6.75
1	12	6.54

3-feature classifier			
Features			Avg(tpr)
1	4	10	8.29
1	10	34	7.58
1	5	10	7.57
1	10	14	7.41
1	10	11	7.41
7	15	44	7.39
11	13	27	7.36
1	10	27	7.34
1	10	44	7.34
1	10	48	7.33
7	9	27	7.32
1	12	34	7.32
7	16	34	7.32
1	7	11	7.31
7	15	27	7.31
7	15	29	7.31
9	13	44	7.30

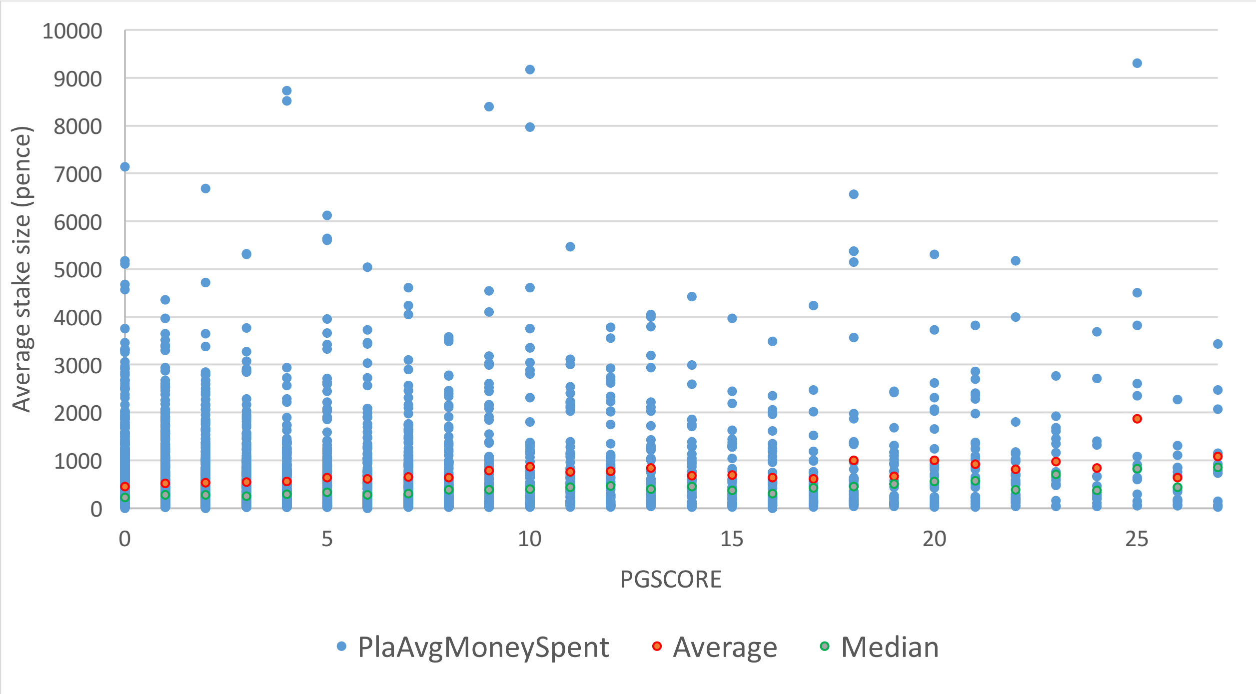
4-features classifier				
Features				Avg(tpr)
1	4	10	39	8.58
1	4	10	11	8.46
1	10	28	34	8.45
1	4	10	28	8.42
1	4	10	41	8.42
1	4	10	15	8.41
1	4	10	46	8.40
1	4	9	10	8.38
1	4	10	23	8.38
1	4	10	11	8.36
1	4	10	12	8.34
1	4	10	16	8.33
1	2	4	10	8.32
1	4	10	43	8.32
1	4	10	60	8.32

5-feature classifier					
Features					**Avg(tpr)**
1	10	28	34	46	8.94
1	4	10	11	13	8.76
1	4	10	28	46	8.72
1	4	5	10	39	8.70
1	10	28	34	38	8.70
1	4	5	10	41	8.69
1	4	10	11	28	8.68
1	4	10	15	39	8.68
1	4	7	10	11	8.66
1	4	10	11	26	8.66
1	2	4	10	41	8.66
1	4	10	11	13	8.65
1	4	10	11	39	8.64
1	4	10	41	46	8.64
1	4	7	10	39	8.63
1	4	10	13	39	8.63
1	10	24	28	34	8.63
1	4	10	14	41	8.63
1	4	10	26	46	8.63
1	4	10	28	34	8.62
1	7	10	28	34	8.62
1	4	10	15	41	8.62
1	4	10	39	51	8.61
1	4	10	15	46	8.61
1	2	4	10	11	8.61

Appendix B: Additional RQ8 Supporting Information

B.1

Distribution of players' average stake sizes as a function of their PGSCORE. There is a small positive correlation (0.16) between the two which means that there is a slight tendency for problem gamblers to wager higher stakes.



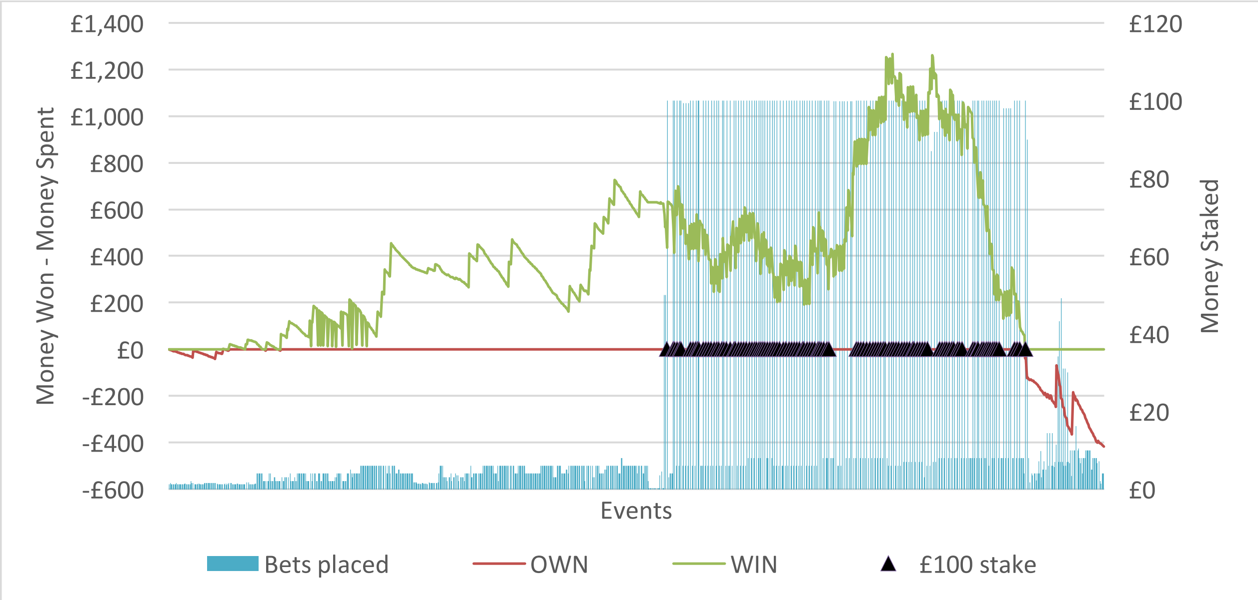
B.2

In the visualisations below, the green / red curves represent the difference between the money won and spent by the player so far during the session. The curve is green when the player is playing with winning – has won more so far during the session than they have spent. If the opposite is true the curve is red. Blue bars correspond to the secondary axis on the right hand side and indicate the amounts of money staked. Black triangles indicate £100 stakes wagered.

A random selection of sessions containing different numbers (ordered from high to low) of £100 stakes from various players are presented below. As it is only a small sample of the whole data set conclusions drawn from them do not necessarily generalise very well. It is, however, possible to observe some trends concerning the circumstances in which maximum stakes are wagered, in particular the fact that they are more prevalent at later stages of sessions and rarely happen in isolation.

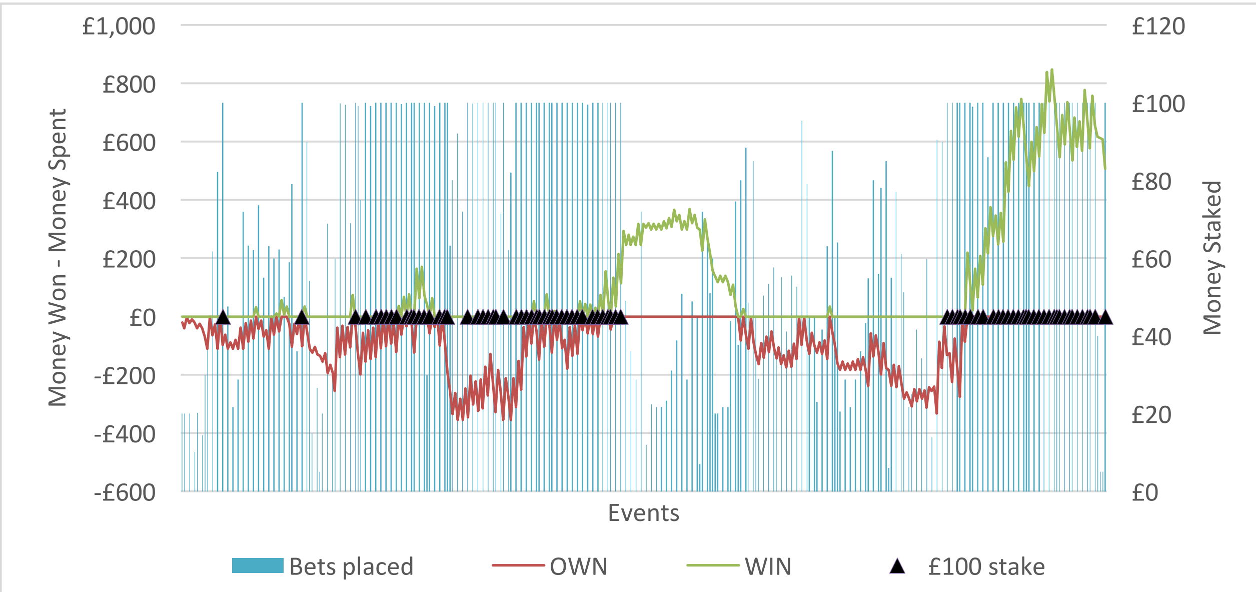
PGSCORE = 5 (Non-problem gambler), session duration: 13:47 – 19:17, total bets: 855, £100 bets: 116

For the first half of the session, the player played low bets. Eventually, they decided to switch to £100 bets. Even though they were in the winning zone for most of the time (have won over £1000 more than spent, at one point), they ended up losing close to £400.



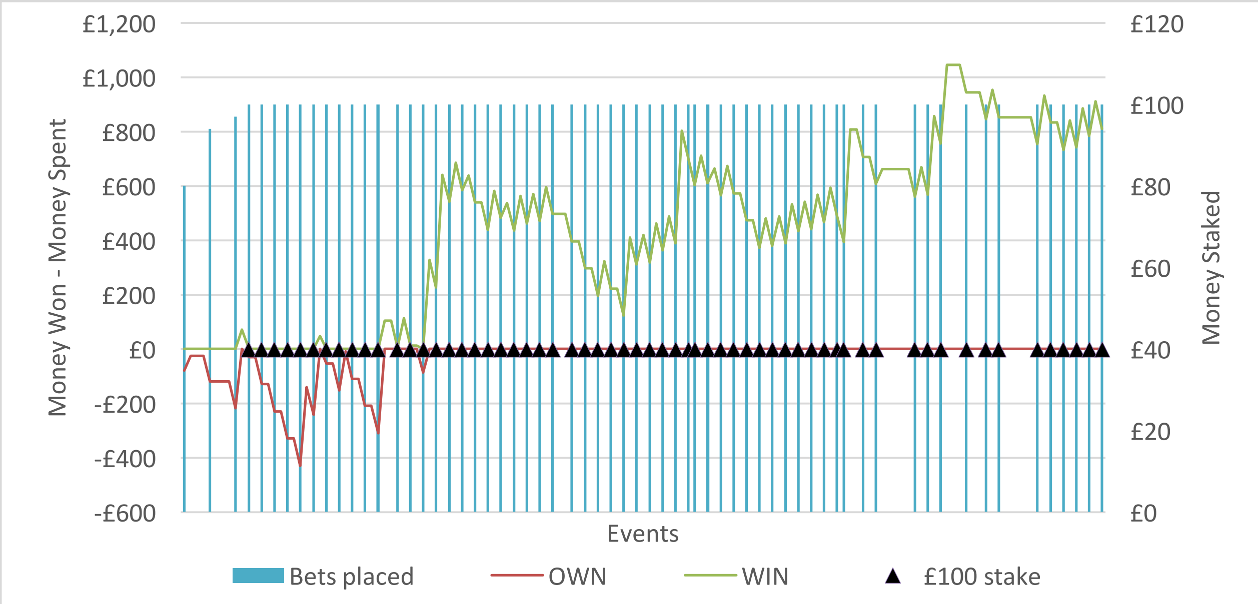
PGSCORE = 18 (Problem gambler), session duration: 12:31 – 16:00, total bets: 212, £100 bets: 80

In this session the player had three periods when they placed £100 bets one after another. Lower bets were used in between the three periods.



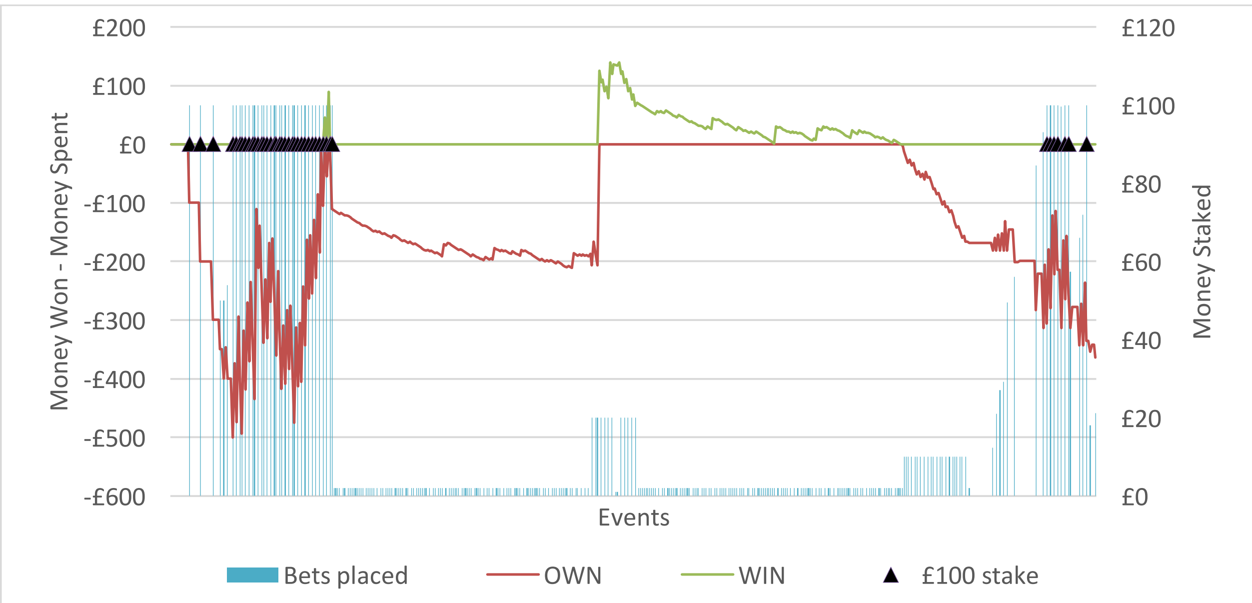
PGSCORE = 4 (Non-problem gambler), session duration: 12:07 – 12:39, total bets: 64, £100 bets: 61

Almost all of the bets in the session were £100. Still, there were three bets of lower, gradually increasing value at the beginning of the session.



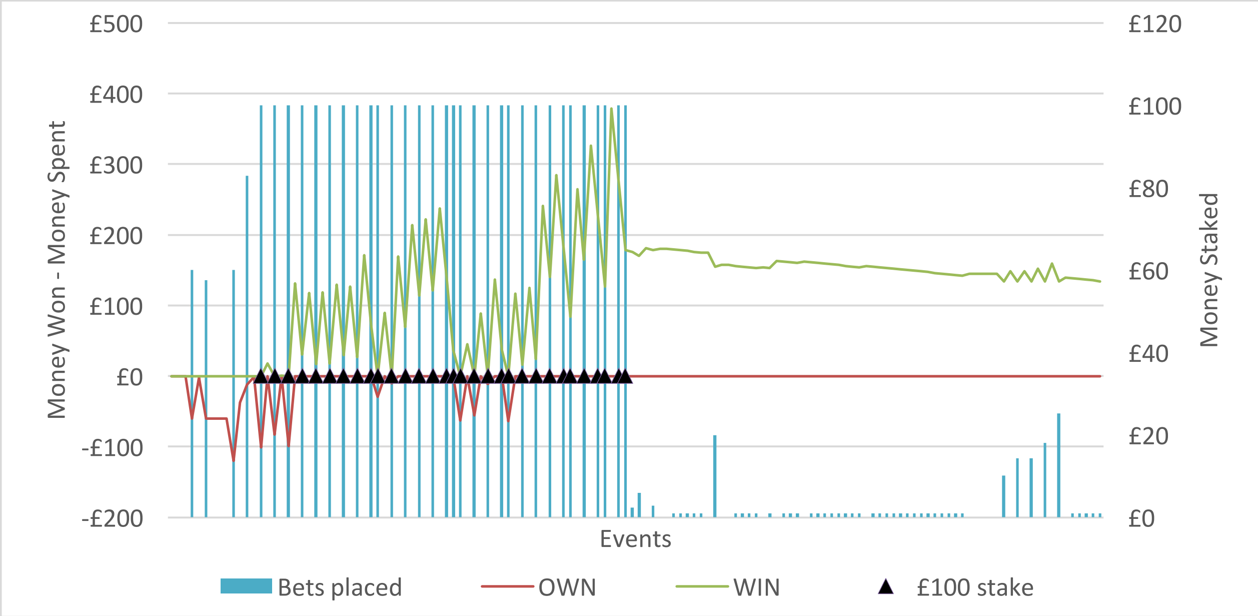
PGSCORE = 1 (Non-problem gambler), session duration: 17:27 – 19:06, total bets: 317, £100 bets: 42

The player has started off the session with £100 bets. After a while, they switched to lower bets and continued using them for most of the session. At the end of the session, the player again placed a number of £100 bets.



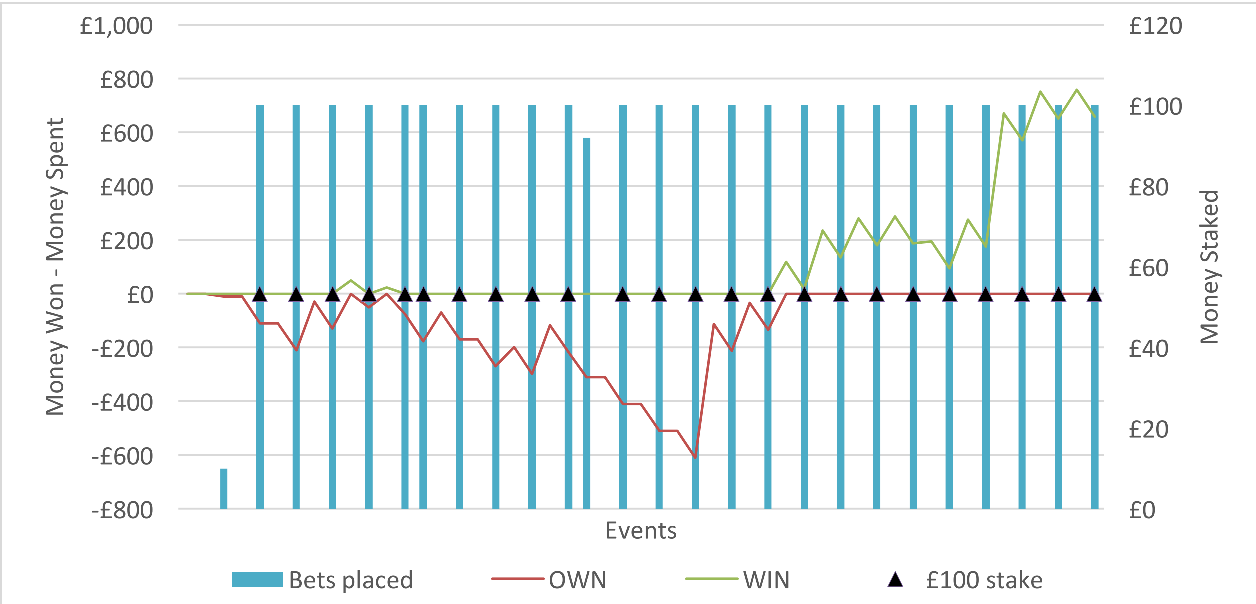
PGSCORE = 13 (Problem gambler), session duration: 20:33 – 21:05, total bets: 84, £100 bets: 31

After a few bets gradually increasing in value the player placed a number of £100 bets in a row. Later on they played throughout the rest of the session using smaller bets.



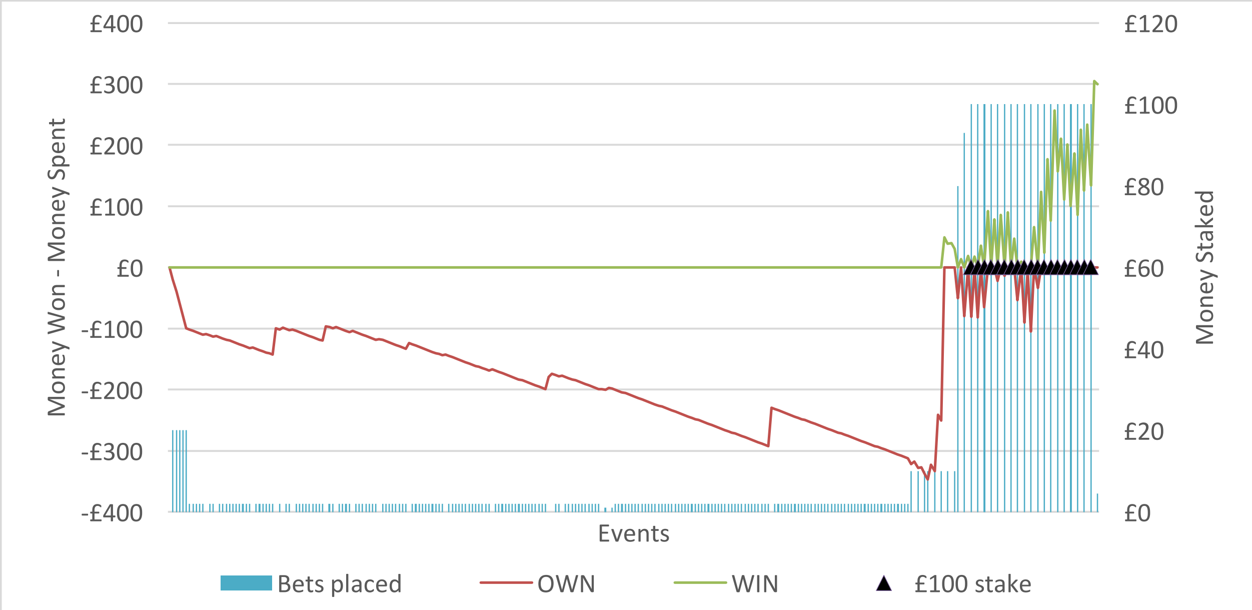
= 1 (Non-problem gambler), session duration: 10:56 – 11:16, total bets: 26, £100 bets: 24

The player has been consistently using £100 throughout the session (apart from the very first bet and one half-way through the session).



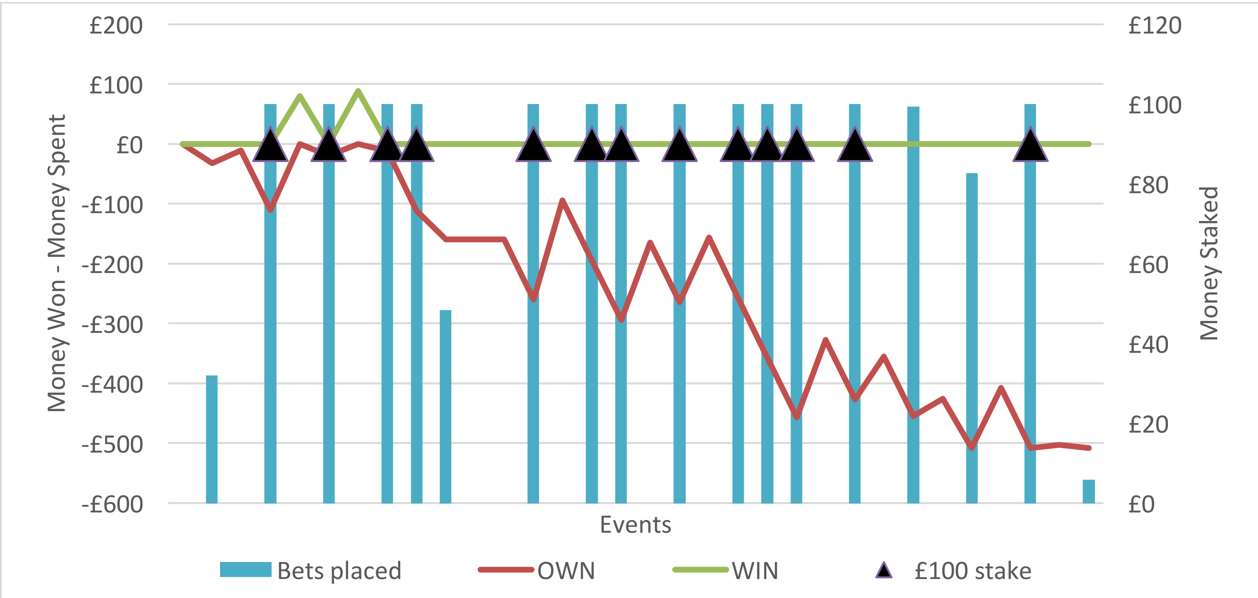
PGSCORE = 0 (Non-problem gambler), session duration: 17:38 – 18:23, total bets: 233, £100 bets: 19

Throughout the majority of the session, the player has been placing lower value bets. Only later did they gradually start increasing the values of bets until reaching £100 and sticking to that value for the rest of the session.



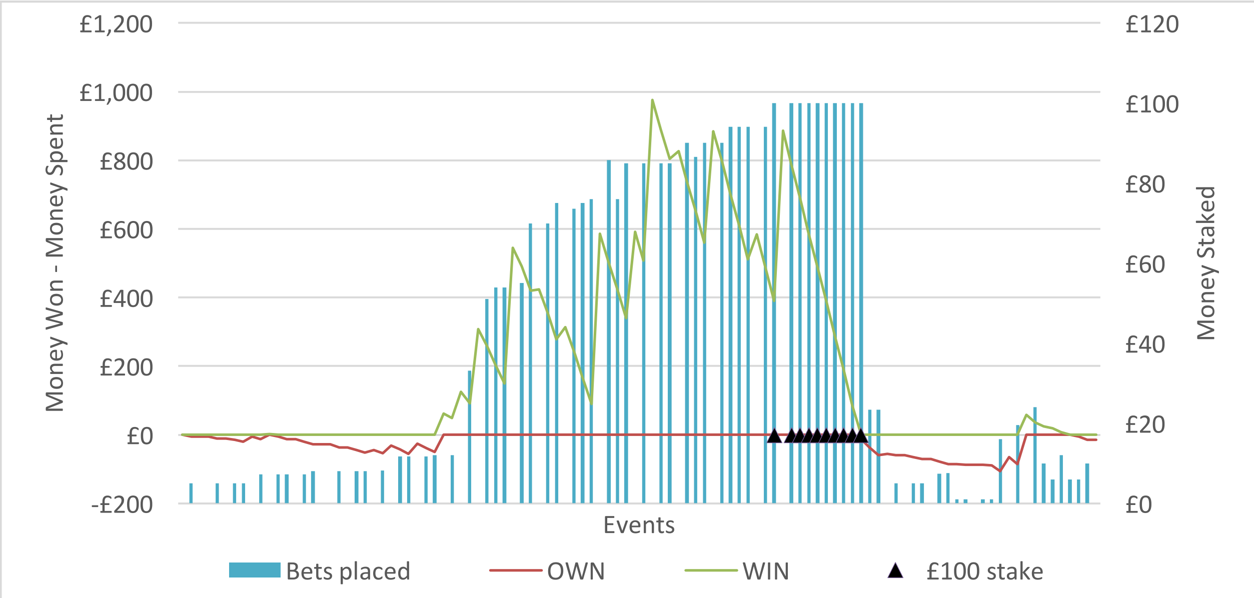
PGSCORE = 18 (Problem gambler), session duration: 10:23 – 10:40, total bets: 18, £100 bets: 13

The player has been placing mostly £100 bets throughout the session.



PGSCORE = 0 (Non-problem gambler), session duration: 10:55 – 11:30, total bets: 73, £100 bets: 10

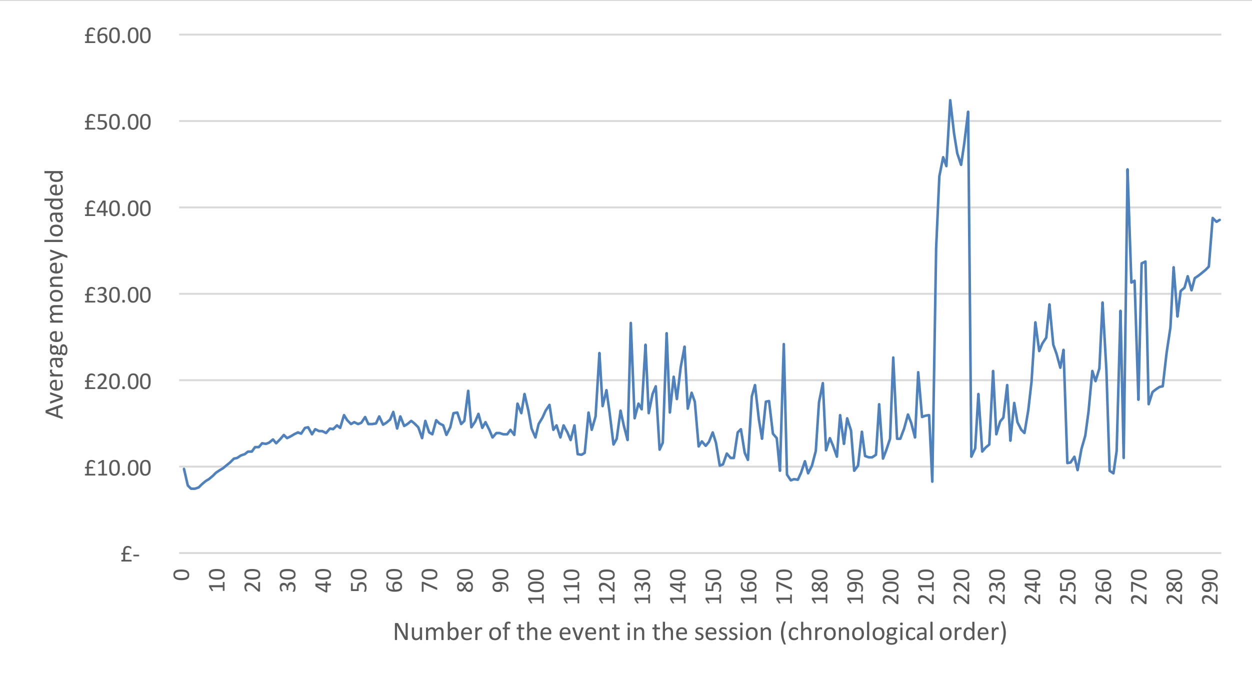
The player has been gradually more successful in terms of the difference between money won and spent and has been increasing the value of bets. Finally, they reached the maximum value of £100 and stuck to it until they have lost all their winnings. During the last stage of the session, the player lowered their stakes.



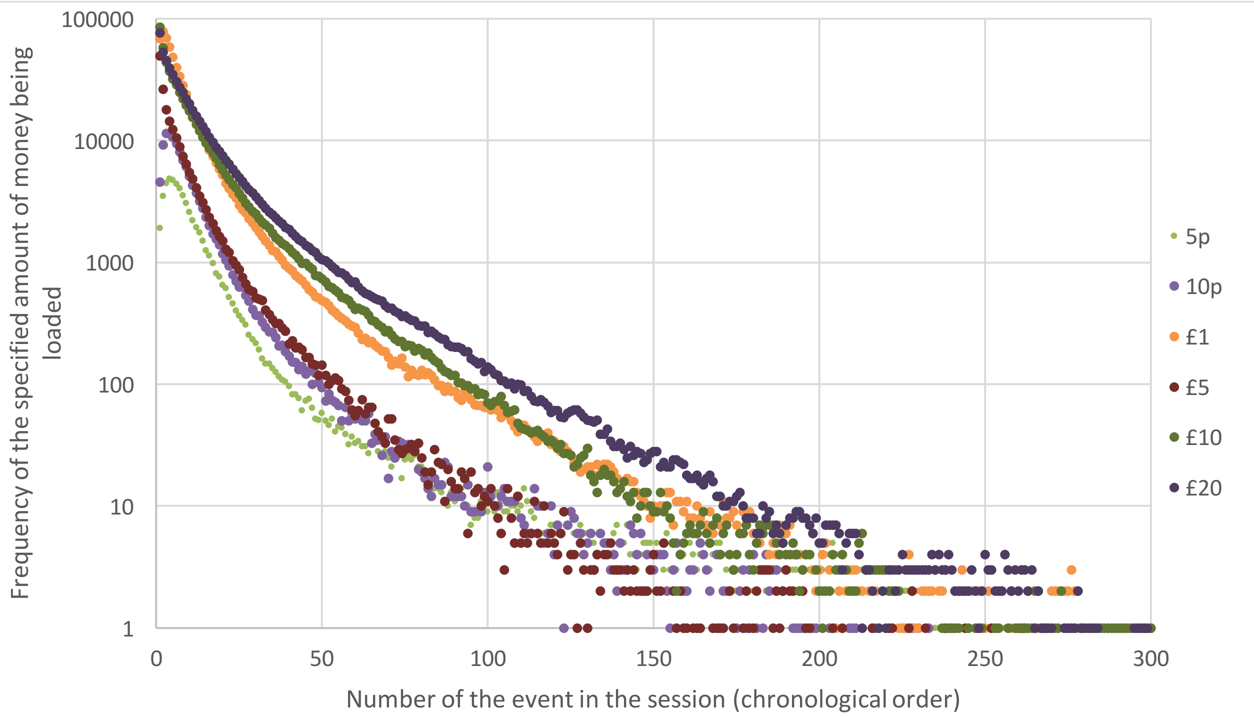
Appendix C: Additional RG9 Supporting Information

C.1

The graph presents the average amount of **money loaded** as a function of which event, in chronological order, it is in the session. There is a significant positive (0.48) correlation between the two variables and the noisy curve below clearly follows an upward trend.



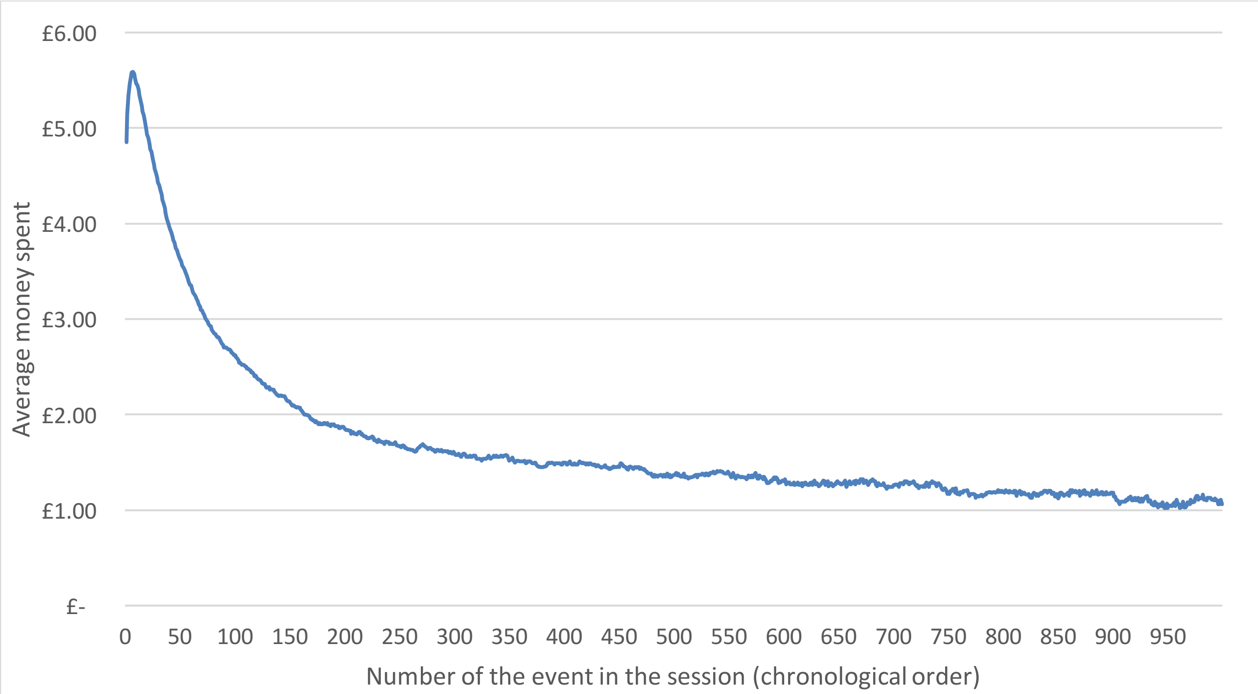
Therefore, the conclusion is that players tend to **load more money** at **later stages** of sessions. The following graph gives a bit more detail on how this is happening.



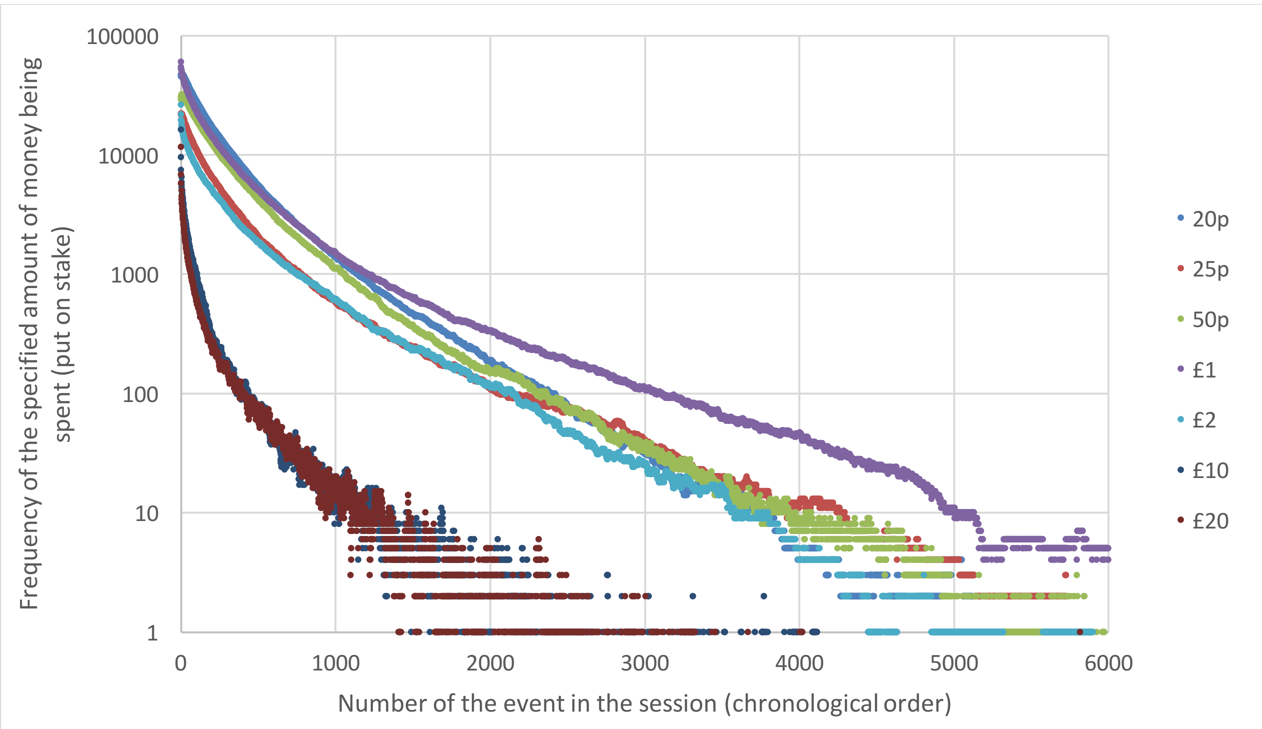
The most commonly used values for loading money in are the £10 and £20 notes, the £1 coin and (only during the first few events of the session) the £5 note. During the early session events £1 is the most common value loaded, followed by £10 and £20. As sessions progress, however, players tend to use fewer £1 coins, relatively more £10 notes and, particularly, £20 notes. This effect is first visible around the 20th event of a session and slowly reinforces itself across subsequent session events. This explains the **slow increase in average money loaded** as sessions progress.

C.2

The graph presents the average amount of **money spent** (staked) as a function of which event in chronological order it is in the session. There is a clear negative correlation (-0.74): the stakes wagered on average get **lower** the further into a session players are.



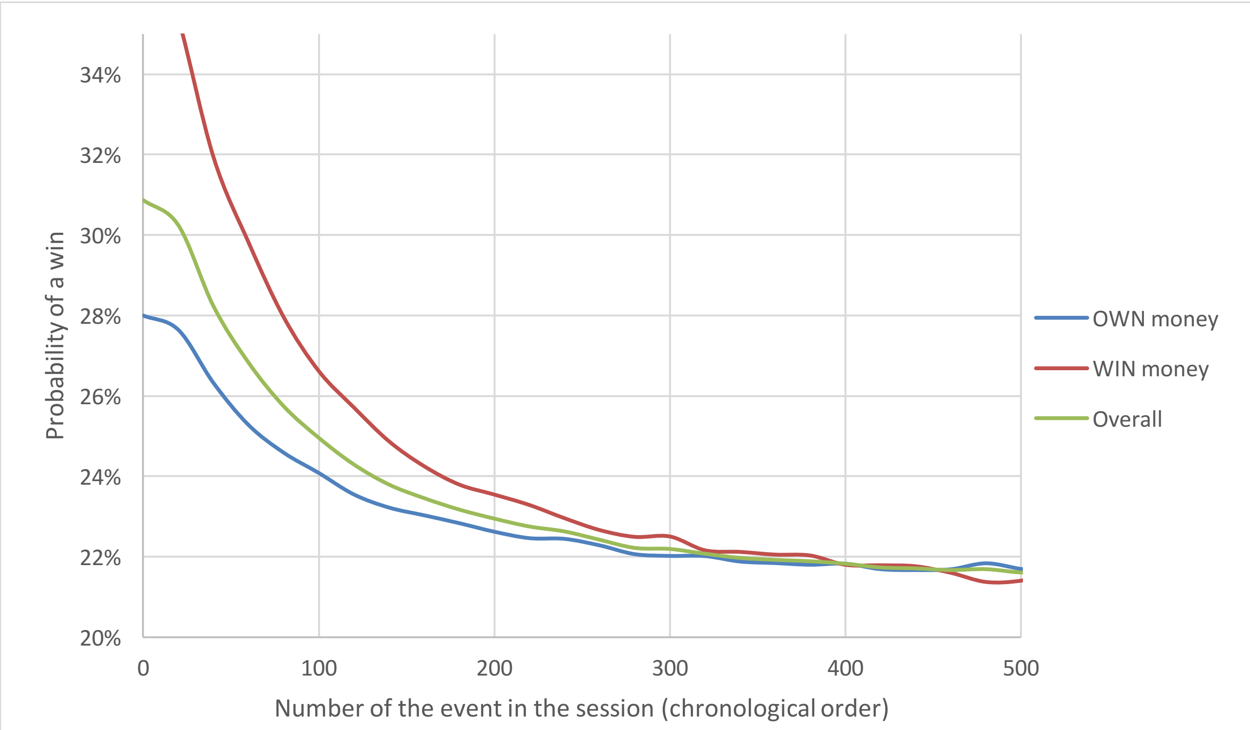
There is a clear trend visible which indicates that stakes wagered are of lower values at later stages of sessions. The graph below decomposes the trend into particular (the most common) stake values.



Throughout the length of sessions the most common stakes are the low ones – up to £2. Interestingly, when sessions reach the 1000th event, £1 stakes become the most common by a significant margin. That is why, in the previous graph, the average stake size **slowly tends** towards £1 for very long sessions. The **initial rapid decrease** in the average stake value can be attributed to the high stakes of £10 and £20 being relatively popular only at early session stages. Their popularity quickly decays as sessions progress.

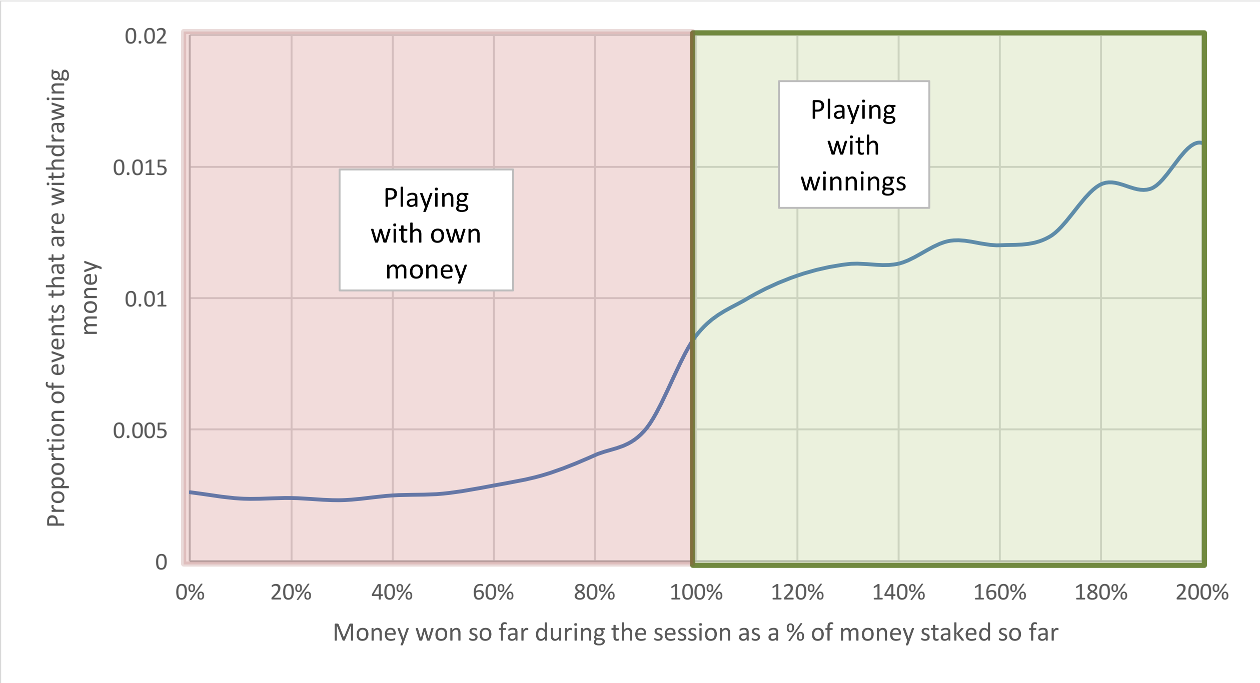
C.3

The graph below shows the **average probability of a player winning a bet** when playing with winnings, with their own money and overall in red, blue and green, respectively. It can be observed that players are more likely to place lower-risk stakes (higher probability of a win) at early stages of sessions (or during short sessions).



C.4

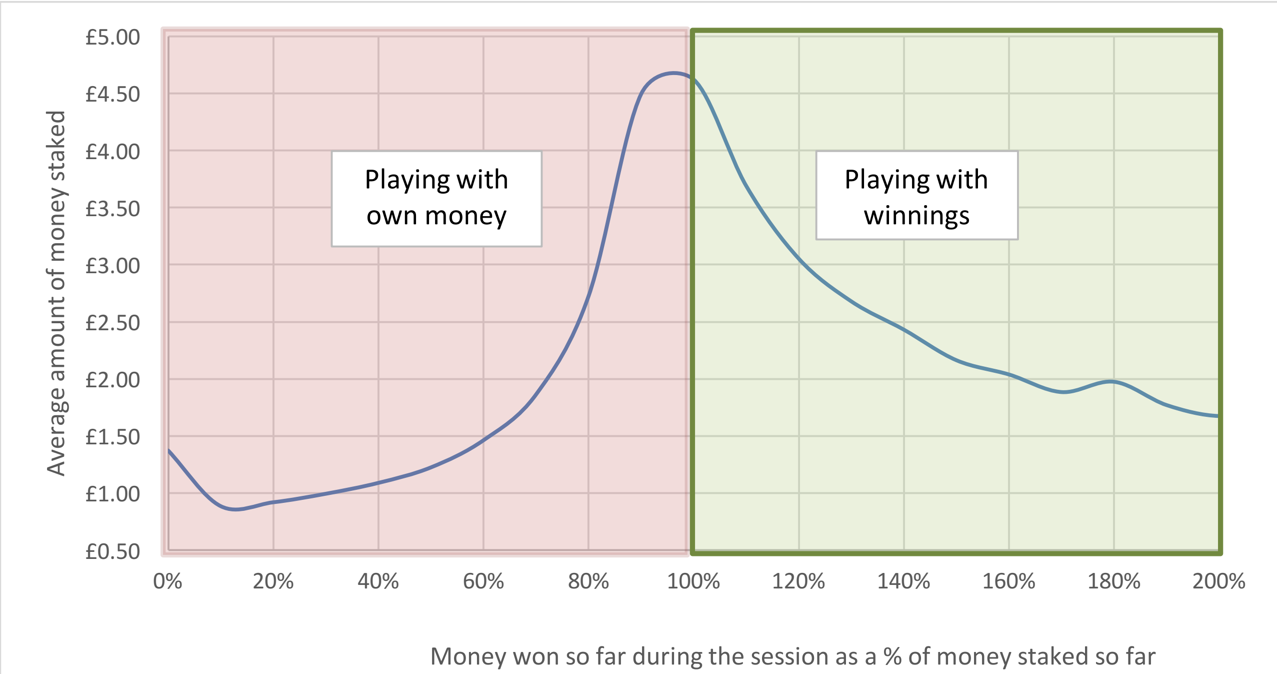
The graph presents the proportion of events that are **withdrawing money** as a function of the money won so far during the session as a percentage of money spent so far.



Clearly, the left hand side of the graph (on the left from the 100% mark) refers to playing with OWN money, the right hand side with WIN money. There is a tendency for players to withdraw money more often as their winnings become higher than the sum of money they have staked (in the green region). When far in the OWN money region, players tend to withdraw money around **3 times less often** than in the WIN zone. That means that once a player has lost a significant proportion of the money they are unlikely to withdraw what is left of it. On average, they would more often choose to keep on playing to go back to the WIN region or end up losing all the money.

C.5

The graph presents the average amount of **money staked** as a function of the money won so far during the session as a percentage of money spent so far.

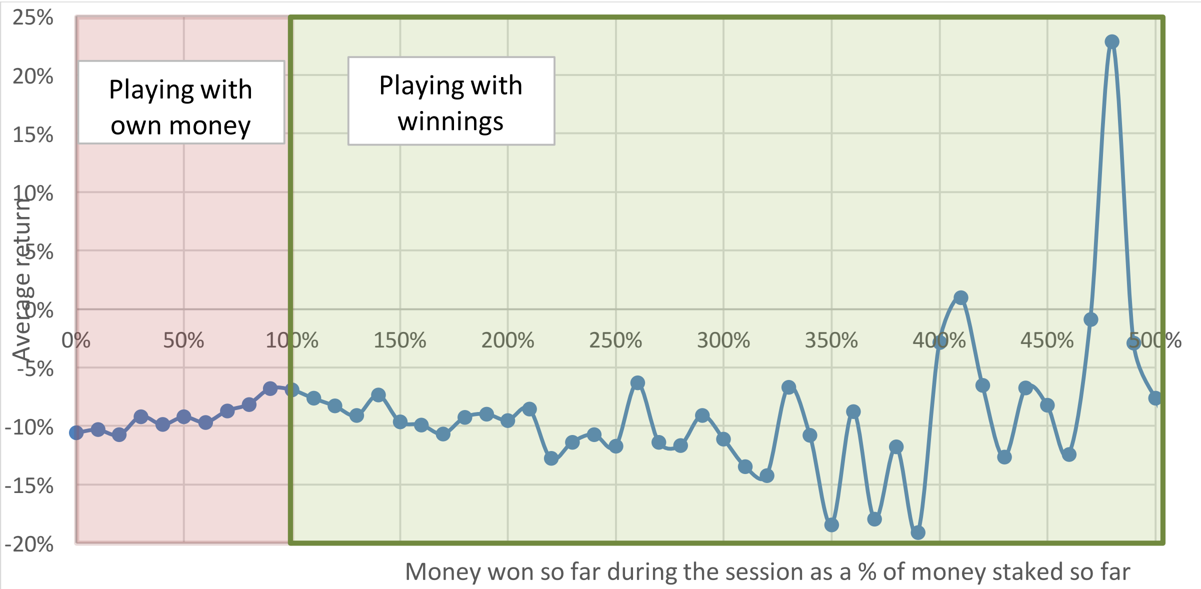


Simply looking at the graph confirms that players wager higher stakes when playing with WIN compared with OWN money. What is interesting, however, is that the peak of the curve is slightly to the left of the point where money spent so far during the session is equal to the money won (the 100% point). In Appendix B an observation was made that the **highest stakes wagered happen at early stages** of a session. These two facts are consistent with one another because at the beginning of a session the player is just to the left of the 100% in the graph above – they have not yet won and have spent a small amount of the money loaded on their initial bet(s).

Interestingly, the average stake value drops as winnings constitute a higher percentage (over 100%) of the money spent so far. This means that the tendency for players to stake more as they have more winnings to spend is a relatively **weak** one. It is **overshadowed** by the tendency to put less at stake later on in a session. It is the relationship between those two tendencies that is depicted by the graph above and the one presented in Appendix B.

C.6

The graph below presents a player's **average return rate** as a function of the money won so far during the session as a percentage of money spent so far. In order to correctly interpret the graph, it is necessary to understand how it was built. Every 10%-wide window of the ratio between money won and spent so far during the session has been treated separately. For all the *Play* and *Win* events in that window an average return rate has been calculated for each player, averaged across all players and plotted on the graph.

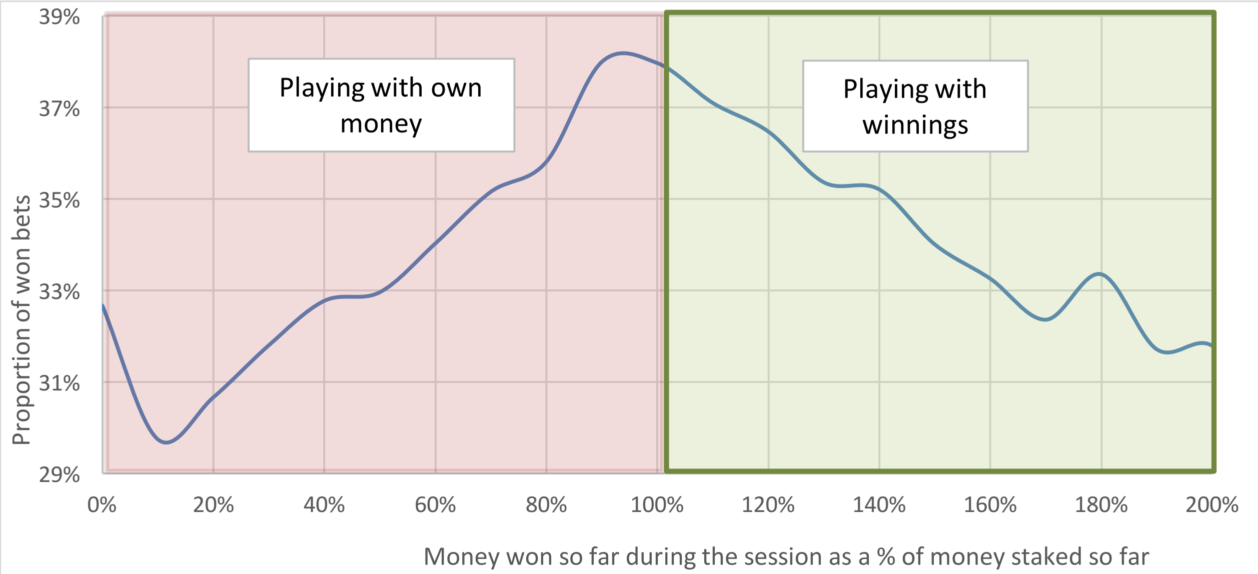


The noisy curve in the graph above confirms the expected result that the average return rate does not vary across different player groups but is fairly constant around a value close to -9%. The curve becomes more noisy further to the right because the number of bets in each 10%-wide bin gets smaller.

A very high and unexpected peak of almost $+25\%$ on the right hand side of the graph is based on a bin which contains over $8\,000$ bets. One of them, however, is an extremely unlikely bet which won £500 when £0.20 was staked which gives a $250\,000\%$ return rate. This single result causes the average value for the bin to be so disproportionally high.

C.7

The graph below shows the **proportion of bets won** depending on whether the player is playing with winnings (green) or with their own money (red).



The graph shows that players tend to place the least risky bets when the total money they have won is close in value to the total money they have staked so far. On the other hand, risky bets are more prevalent further from the 100% boundary (either to the left of to the right). This would indicate that players who are either far in the 'winnings' zone (are doing very well) or far in the 'own money' zone (are doing poorly) are more likely to place risky bets. Finally, the shape of the above curve is somewhat similar to the shape of the 'money staked' curve in Appendix C.5. That shows that there is a general tendency for players to stake higher values of money on less risky bets and vice versa which is what would be expected.