Taylor & Francis
Taylor & Francis Group

Addiction Research & Theory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iart20

An empirical attempt to identify binge gambling utilizing account-based player tracking data

Michael Auer & Mark D. Griffiths

To cite this article: Michael Auer & Mark D. Griffiths (10 Oct 2023): An empirical attempt to identify binge gambling utilizing account-based player tracking data, Addiction Research & Theory, DOI: 10.1080/16066359.2023.2264763

To link to this article: https://doi.org/10.1080/16066359.2023.2264763

Published online: 10 Oct 2023.

ADDICTION RESEARCH & THEORY
https://doi.org/10.1080/16066359.2023.2264763

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

An empirical attempt to identify binge gambling utilizing account-based player tracking data

Michael Auer[a] and Mark D. Griffiths[b]

[a]Neccton GmbH, Lienz, Austria; [b]Psychology Department, International Gaming Research Unit, Nottingham Trent University, Nottingham, UK

ABSTRACT

Binge gambling is a relatively under-explored area and the few published studies have all used self-report data (i.e. surveys and interviews). The use of account-based tracking data has increasingly been used to identify indicators of problem gambling. However, no previous study has ever used tracking data to operationalize and explore binge gambling. Therefore, the present study investigated whether it is possible to identify behavioral patterns that could be related to binge gambling among a real-world sample of online gamblers. The authors were given access to an anonymized secondary dataset from a British online casino operator comprising 150,895 online gamblers who gambled between January and March 2023. Using 14 parameters of gambling (e.g. total number of gambling days, total number of gambling sessions, average amount of money spent per game), six distinct clusters of gamblers were identified. Two clusters – Cluster 2 ($n = 22{,}364$) and Cluster 5 ($n = 12{,}523$) – gambled on a relatively low number of days during three months, but displayed a high gambling intensity on those days compared to the other four clusters. These two profiles could potentially match the habits of binge gamblers. The majority of players retained their behavior in the following three months between April and June 2023 and were consequently assigned to the same cluster in the latter time period. A total of 17% of gamblers in Cluster 3 and 29% of gamblers in Cluster 5 stopped gambling entirely between April and June 2023. The findings suggest that binge gambling may be able to be identified by online gambling operators using account-based tracking data and that targeted interventions could be implemented with binge gamblers.

ARTICLE HISTORY
Received 12 April 2023
Revised 23 September 2023
Accepted 26 September 2023

KEYWORDS
Binge gambling; problem gambling; behavioral tracking; account-based data; cluster analysis

Introduction

The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association 2013) and the eleventh revision of the *International Classification of Diseases* (ICD-11; World Health Organization 2019) describe gambling disorder (GD) as persistent and recurrent maladaptive gambling that results in gambling-related harms (e.g. financial problems, psychosocial problems, health problems, occupational/educational problems). In the DSM-5 (American Psychiatric Association 2013) GD was re-classified from an impulsive disorder to a behavioral addiction within the revised category of Substance-Related and Addictive Disorders (Petry et al. 2014).

Binge gambling

The DSM-5 also includes new provisions for the specification of persistent versus episodic forms of gambling disorder. An episodic progression means that symptoms subside between gambling episodes. The episodic nature of problem gambling was also supported by Williams et al.'s (2015) longitudinal study of a sample of 4211 adult Canadians. They found that intermittent periods of problem gambling were common when considered across annual intervals. Nower and Blaszczynski (2003) discussed the difficulties associated with the attempt to diagnose problem gambling based on two case studies. One was a 30-year-old Asian male bank employee and the other was a 26-year-old female factory worker. Both exhibited gambling problems. However, the first gambler met the criteria for problem gambling in both the DSM-IV (American Psychiatric Association 2013) and SOGS (Stinchfield 2002), but the second gambler did not.

In a conceptualization similar to binge drinking and binge eating, Nower and Blaszczynski (2003) posited that there was a sub-group of gamblers who participate in intermittent binge episodes of gambling during which they experience a transient loss of control that results in acute harmful consequences. Both case studies described by Nower and Blaszczynski (2003) reported the emergence of transient harmful economic and social consequences associated with each episode of time-limited excessive gambling. In between the episodes there was not only abstinence from

CONTACT Mark D. Griffiths mark.griffiths@ntu.ac.uk Psychology Department, International Gaming Research Unit, Nottingham Trent University, 50 Shakespeare Street, Nottingham, NG1 4FQ, UK

gambling but also an absence of any drive to resume gambling. Nower and Blaszczynski (2003) described the following criteria for the identification of binge gambling: (i) sudden onset of an irregular and intermittent period of sustained gambling, (ii) excessive monetary expenditure relative to income, (iii) rapid spending of money over a discrete interval of time, (iv) gambling that is accompanied by a sense of urgency and impaired control, (v) gambling resulting in marked intra-personal and inter-personal distress, and (vi) the absence, between bouts of gambling, of any rumination, preoccupation or cravings to resume participation in gambling.

A case study of problem binge gambling was also reported by Griffiths (2006). He interviewed a 31-year-old male slot machine addict three times over a three-month period preceding a court case regarding the individual's gambling-related crimes. The gambler's intense episodes of gambling were typically caused by very specific 'trigger' incidents (e.g. relationship break-up, not being allowed to see his young daughter). Griffiths also noted that in this case, the man's gambling problem was the symptom of other underlying problems. When these underlying problems were dealt with, his problem gambling all but disappeared. Griffiths concluded that binge problem gambling appeared to be less serious than chronic problem gambling but could still cause significant problems in the lives of people it affects.

Cowlishaw et al. (2018) developed a nine-item Binge Gambling Screening Tool (BGST) to provide a preliminary means of operationalizing the binge gambling criteria of Nower and Blaszczynski (2003) in clinical settings. If players indicate intermittent and irregular gambling, they are asked eight core questions regarding their periods of gambling (e.g. '*During this episode, did you feel you lost control over gambling?*'). In a sample of 214 patients entering a specialist treatment clinic for gambling problems, 32% reported gambling episodes that were irregular and intermittent. Those individuals with irregular episodic gambling demonstrated lower levels of problem gambling severity and comorbidity. Based on a strict interpretation from Nower and Blaszczynski (2003), 3.7% of players were classified as binge gamblers. The criteria for binge gambling (taken verbatim from Cowlishaw et al. 2018) were that: (i) gambling was irregular and intermittent, (ii) gambling episodes began suddenly after a period of abstinence, (iii) during gambling episodes, there were losses of control over gambling, (iv) outside of episodes, there was absence of pre-occupation with gambling, and (v) outside of episodes, there was absence of gambling urges. When excluding the fifth criterion, the results indicated 13.6% of patients were potential binge gamblers ($n = 29$). Cowlishaw et al. (2018) also reported that binge gamblers were less likely to report any psychiatric comorbidities, and also reported lower anxiety.

Sklar et al. (2010) conducted interviews with adolescents and young adults in a residential drug treatment facility and found some evidence for binge gambling. Participants described episodic gambling as a way of to raise funds for drug habits or as a substitute form of excessive risk-taking when drugs were not available. Furthermore, it appeared that gambling, drug use, and alcohol use frequently served as triggers for each other. Binge gambling precipitated or was precipitated by the use of alcohol or drugs.

Hing et al. (2012) derived gambling profiles from interviews of 169 indigenous Australians. Profiles were based on seven dimensions of gambling involvement (importance/interest, pleasure, centrality, self-expression, social bonding, risk probability, and risk consequence). One respondent in the study stated that '*many people are binge gamblers and Indigenous people have more triggers or reasons to prompt a binge*' (p. 224). Binge gamblers were said to be mostly males who usually gambled socially with family or social groups in a contained way with lower expenditure, but on occasions gambled intensively in a continuous style spending large amounts of money. Hing et al. (2012) also noted that binge gambling was motivated by seeking escape from stress, for time out and time away from home, and as a distraction from recurring problems.

Gupta and Derevensky (2011) surveyed 1254 participants aged between 17 and 23 years and 16.4% responded that they experienced binge gambling episodes over a sustained period of time (in answer to the item: '*I have episodes of gambling, over sustained period of time, that seems to have a clear beginning and end*'). The study also found that gambling severity was positively correlated with the occurrence of experiencing gambling episodes that had a clear beginning and end. Gupta and Derevensky (2011) also asked participants if gambling episodes began with a sudden uncontrollable urge to gamble but the majority of binge gamblers did not agree to that statement. Binge gamblers were identified as those who endorsed the aforementioned item. The gamblers who agreed with the item were more likely to be problem gamblers and less likely to be social gamblers.

Operationalizing problem gambling behavior using tracking data

A number of previous studies have attempted to apply account-based tracking data to criteria of disordered gambling in an attempt to operationalize indicators of problem gambling (e.g. Perrot et al. 2018; Challet-Bouju et al. 2020; Catania and Griffiths 2022). Catania and Griffiths (2022) argued that most of the DSM-5 criteria for gambling disorder can be operationalized (at least to some extent) using actual transaction data. In a sample of 982 online gamblers, they assessed customer service contacts, number of hours spent gambling, number of active days, deposit amounts and frequency, the number of times a responsible gambling tool (such as deposit limit) were removed by the gamblers themselves, number of canceled withdrawals, number of third-party requests, number of registered credit cards, and frequency of requesting bonuses through customer service. During a subsequent cluster analysis of the 982 players Catania and Griffiths identified non-problem gamblers, at-risk gamblers, financially vulnerable gamblers, and emotionally vulnerable gamblers. However, this study was exploratory and it was not known whether the gamblers were problem gamblers or not. Their conclusions regarding problem gambling status were purely made on the basis of observed gambling behavior.

Auer and Griffiths (2023a) attempted to operationalize chasing losses in a sample of 16,771 online gamblers from the UK, Spain, and Sweden. The study developed five metrics indicative of chasing losses. These were (i) within-session chasing, (ii) across-session chasing, (iii) across-days chasing, (iv) regular gambling account depletion, and (v) frequent session depositing. Auer and Griffiths (2023a) concluded that frequent session depositing reflected chasing losses better than any of the other four metric operationalizations used.

Perrot et al. (2018) and Challet-Bouju et al. (2020) assessed chasing losses in studies of account-based player-tracking data. Both studies operationalized chasing losses as either three or more deposits within a 12-h period or a deposit less than one hour after a previous bet. Based on 10,000 French online lottery players Perrot et al. (2018) identified a cluster of players which was characterized by a high gambling activity and a high probability of chasing behavior. Based on the first 6 months of gambling from a sample of 1152 French online lottery players Challet-Bouju et al. (2020) identified a subgroup of gamblers which stood out in terms of much higher gambling activity and breadth of involvement, and the presence of chasing behavior.

A number of other studies have also utilized player tracking data for the understanding of gambling behavior. Dragičević et al. (2011) analyzed the trajectory, frequency, intensity, and variability of gambling behavior in a sample of online gamblers during the first four months after registration. They found subgroups of gamblers who were spending more time and gambled more frequently especially on online slots games. Other studies have used player tracking data to predict voluntary self-exclusion (VSE) (i.e. Percy et al. 2016; Finkenwirth 2021; Hopfgartner et al. 2023). Percy et al. (2016) compared several machine learning algorithms and concluded that Random Forest delivered the highest prediction accuracy based on a sample of 845 gamblers who self-excluded. Based on a sample of 2157 Canadian online gamblers, Finkenwirth et al. (2021) found that the variance in money bet per session contributed the most to the prediction of self-reported problem gambling. Hopfgartner et al. (2023) used a sample of 25,720 online gamblers out of which 414 self-excluded. A greater probability of future self-exclusion across countries was associated with a higher number of previous voluntary limit changes and self-exclusions, higher number of different payment methods for deposits, higher average number of deposits per session, and a higher number of different types of games played

The present study

The aforementioned studies concerning binge gambling have either been based on interviews or information from surveys. None of the previously published studies on binge gambling have ever utilized account-based player tracking data. Moreover, the previous studies that have tried to operationalize aspects of problem gambling and markers of harm have never tried to operationalize binge gambling. Therefore, the present study investigated whether it is possible to identify behavioral patterns that could be related to binge gambling among a real-world sample of online gamblers. Given that binge gambling could evade detection algorithms because of its episodic nature, these efforts could help identify this subgroup of gamblers and assist prevention efforts by online gambling operators.

In the present study, the authors examined a sample of British online gamblers to study patterns of gambling which could be indicative of binge gambling. There are currently no definitions of binge gambling with respect to the frequency of episodes, intensity or gambling habits. For that reason, there were no specific hypothesis (as the study was exploratory) other than the investigation of gambling patterns over a three-month period. It was anticipated that the findings will be helpful for policymakers and regulators, as well as for online gambling operators.

Method

The data for the present study comprised an anonymized secondary dataset provided by an online casino operator from the UK. The dataset comprised all types of games played, details of every game that was won, all deposits that were made, as well as all financial transfers from the players' online gambling accounts to their bank accounts (i.e. monetary withdrawals). All the transactions were assigned to single player accounts. Other information in the secondary dataset included the players' gender, age, and registration details. The dataset comprised player data from January 1, 2023 to June 30, 2023 (inclusive).

Study design

The dataset comprised all gamblers who placed at least one bet between January-June 2023 and registered before January 2023. The data from January to March were used to carry out a cluster analysis and the derived cluster analysis was applied to data from April to June. The goal of the cluster analysis with the data from January to March was to identify groups of players which were similar with respect to a number of player tracking features. The goal of the application of the cluster model with the data from April to June was to determine the stability of the cluster memberships. For each gambler, the gambling behavior during January to March was calculated (see Appendix 1 for a list of all the variables). Apart from two demographic variables (i.e. age and gender), 14 player tracking variables were calculated (e.g. number of gambling days, average amount of money bet per game, etc.). The variables assessed monetary gambling intensity as well as impulsivity which can result in chasing losses. The latter were captured by the number of deposits per day and per session, as well as the percentage of sessions terminating with low balance. This is in line with the findings of Auer and Griffiths (2023a) who argued that frequent depositing within a session or on a day and regular depletion of the gambling account

Table 1. Mean (and standard deviation) and median (and range) for each variable which was calculated based on data from January to March 2023.

	Variable	Mean (SD)	Median (range)
1	Age (in years)	39.12 (11.74)	37 (77)
2	Female	45%	0 (1)
3	Number of gambling days in total	10.02 (13.32)	4 (89)
4	Average number of days between two gambling days	6.61 (9.77)	3.33 (89)
5	Standard deviation of the number of days between two gambling days	4.54 (6.47)	2.07 (60.81)
6	Number of gambling sessions in total	15.98 (30.93)	5 (860)
7	Average amount of money bet per game (£)	0.72 (9.54)	0.21 (1390.95)
8	Standard deviation of the amount of money bet per game (£)	0.54 (6.90)	0.05 (869)
9	Average monetary deposit amount (£)	31.86 (76.38)	14.05 (4956)
10	Standard deviation of the average monetary deposit amount (£)	2.32 (13.50)	0.1 (1012)
11	Average amount of money deposited per day (£)	49.04 (129.27)	13.33 (6885)
12	Average amount of money bet per day (£)	440.30 (3280.55)	69.5 (599,886)
13	Average number of deposits per day	1.06 (1.29)	0.81 (37)
14	Average number of deposits per session	0.78 (0.93)	0.62 (37)
15	Average number of gambling sessions per day	1.38 (0.72)	1.11 (13)
16	Percentage of sessions terminating with a low monetary balance	0.79 (0.29)	0.97 (1)

may be indications of chasing and not being able to stop gambling. The balance refers the amount of money that players have in their gambling account at a specific point in time. If there was less than £5 in the gambling account at the end of a session, this was considered as a low balance (Auer & Griffiths, 2023a).

Griffiths (2006) concluded that the intense gambling episodes among problematic binge gamblers were caused by very specific 'trigger' incidents (e.g. relationship break-up) which would indicate an irregular number of days between the binges. For that reason, the authors also calculated the average number of days between to gambling days and the standard deviation of the number of days between two gambling sessions. If a player only gambled on one day, the standard deviation was not defined because there was only one observation.

Statistical analysis

A k-means cluster analysis was performed (Likas et al. 2003). The 14 player tracking features were used to identify clusters of players. Age and gender were not used to compute the clusters. K-means cluster analysis minimizes the sum of squares and is therefore sensitive to the range of the input variables. The larger the range, the bigger the influence of a variable on the cluster formation. In order to assign equal importance to each variable, it is recommended that each variable is standardized (Mohamad and Usman 2013). The min-max standardization was applied. In this procedure, first, the minimum value is subtracted from an observation and then divided by the range. The range is the difference between the maximum and the minimum value. Consequently, each scaled variable has a range from 0 to 1 and the influence on the clustering is equal for each variable. The programming language *Python* (Van Rossum 2007) was used to analyze the dataset. The *sklearn* package was used to perform the k-means clustering. The coding for the analysis is publicly available at: https://osf.io/4ub25/.

The number of clusters was determined by a scree-plot (Brusco and Cradit 2001) and silhouette scores (Lletı et al. 2004) A scree plot displays the within sum of squares for each cluster solution. The within sum of squares is the distance between each data point and the respective cluster center and decreases with an increasing number of clusters. If every record is its own cluster, the within sum of squares is zero. The silhouette score is a measure of how similar a record is to its own cluster center compared to other cluster centers. The value ranges between −1 and +1 where +1 indicates that the data record is well matched to its own cluster center and badly matched to other cluster centers.

In order to determine the stability of the players' cluster membership, the cluster model was applied to data from April to June, 2023. Based on the gambling behavior during April to June, every player was assigned to one of the clusters. Then the percentages of players remaining in each cluster and transitioning to each of the other clusters were calculated.

Results

A total of 150,895 online gamblers registered before January 2023 and placed at least one bet between January and March 2023. The average age was 39 years (SD = 11.74) and 67,681 online gamblers were female (46%). Table 1 reports the mean and standard deviation for each variable which was calculated based on data from January to March 2023. On average, individuals gambled for 16 sessions on ten days and there were approximately seven days between two gambling days. On average they bet £0.72 per game and the average amount of money deposited was £32. The average daily deposit amount was £49 and the average amount bet per day was £440. On average gamblers deposited once per day and less than once per session.

Table 1 also reports the median value and the range (difference between maximum and minimum). Metrics such as amount bet, deposit, and number of sessions can potentially be very large and lead to skewed distributions. Skewed distributions can be determined based on the difference between the mean and median. The mean average number of gambling sessions was 15.98 and the median number of gambling sessions was 5. The range was 860 which means that some players had a very large number of gambling sessions. The mean average amount of money deposited per day was £49 and the median amount of money deposited per day was £13. The range of the same metric was £6,885

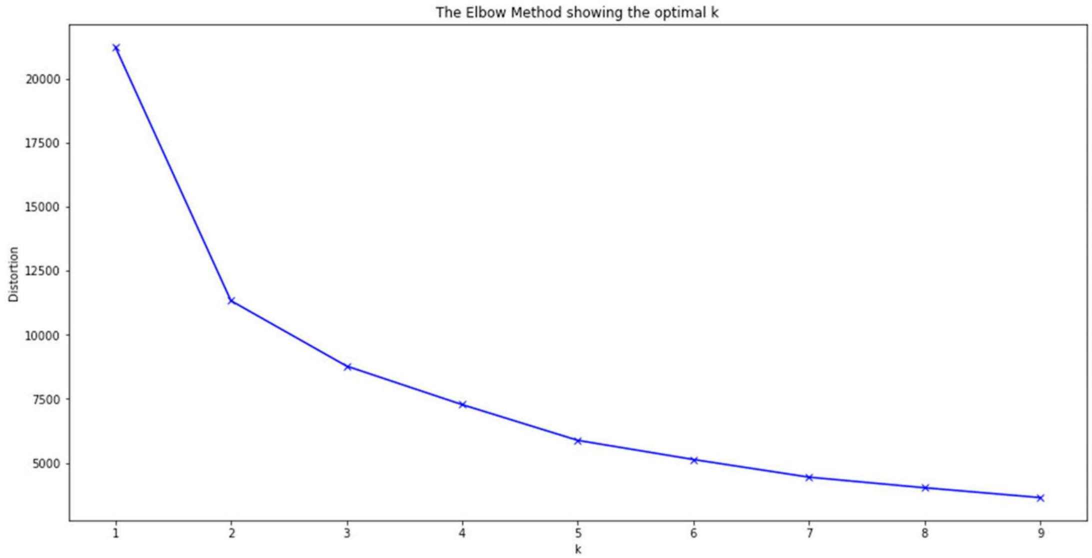

Figure 1. Elbow plot which visualized the within sum of squares for a one-cluster up to a nine-cluster solution.

Table 2. Average silhouette value for a two-cluster up to an eight-cluster solution.

Number of clusters	Silhouette score
2	0.515
3	0.406
4	0.398
5	0.424
6	0.442
7	0.414
8	0.424

which also indicates that some gamblers deposited large amounts of money per day. The same relationship between median and mean average can be observed for the mean average amount of money bet per day.

Various cluster solutions were calculated and the authors decided on a solution with six clusters. The size of clusters, uniqueness, and interpretability were the deciding factors. Furthermore, a scree plot (Brusco and Cradit 2001) which displays the optimal number of clusters was used during the decision process. The y-axis value decreases much more between the one-cluster and the three-cluster solution, but there was little difference when the number of clusters was further increased. For that reason, three clusters would be recommended by the scree plot (Figure 1).

An examination of three clusters did not provide any indications of a binge gambling group and the authors suspected that binge gamblers might comprise a smaller group of players which would only surface if the number of clusters was increased. Therefore, a silhouette score (Lletı et al. 2004) was subsequently computed for different cluster solutions ranging from two clusters to eight clusters (Table 2). The highest silhouette score was generated by two clusters and the second highest silhouette score was generated by a six-cluster solution. For that reason, the authors chose to continue the analysis with six clusters because binge gamblers were not identified by the two-cluster solution. k-means clustering can also be used as a method for detecting outliers (Yoon et al. 2007). The potentially small group of binge gamblers were regarded as outliers in the present study.

Table 3 reports the average profile per cluster and the average across all players. Cluster 1 included the highest number of playing days (49 days) as well as the highest number of total sessions (103 sessions). If there was a subgroup of binge gamblers it would be expected to include a relative low number of playing days due to the episodic nature of binge gambling. Cluster 1 can be ruled out based on this assumption. The maximum of playing days from January 1 to March 31 was 90. The average number of player days across all six clusters was 10. Gamblers in Clusters 2 played on approximately as many days as the average player. Players in Clusters 3, 5 and 6 played on fewer days than average. Cluster 4 gamblers played on average 16 days.

During a low number of playing days, binge gamblers should have a high intensity of gambling. Gamblers in Cluster 2 (17 sessions) and Cluster 5 (10 sessions) had more sessions than any of the other low gambling days clusters. Of these, 15% of the sample were classified in Cluster 2 and 8% were in Cluster 5. The average amount of money deposited per deposit between January and March in Cluster 2 (£51.98) and Cluster 5 (£43.66) was higher than in any other cluster. On average, players in Cluster 2 bet £1.20 per game and players in Cluster 5 bet £1.18 per game. This was higher than the average bet per game in any other cluster. The average bet per game in the entire sample was £0.72. Also, the average amount of money deposited per day in Cluster 2 (£85.91) was higher than in any other cluster. The average amount of money deposited per day in Cluster 5 (£50.67) was about average. The average amounts of money bet per day for Cluster 2 and Cluster 5 (£1021.08 and £827.48) were higher than in any other cluster. The average number of deposits per day and per session were not highest in either Cluster 2 or Cluster 5. Moreover, players in Cluster 2 and Cluster 5 did not deplete their sessions more frequently than the average (of the total sample). On average, 6.61 days elapsed between two sessions in Cluster 2 and the standard

Table 3. Average values for of each of the six clusters calculated based on data from January to March 2023.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Average
Age (in years)	42.37	40.63	37.89	39.16	39.95	38.86	39.12
Female	53%	44%	43%	48%	41%	46%	45%
Number of playing days	49.27	9.58	3.14	16.82	5.15	3.89	10.02
Average number of days between two gambling days	1.78	5.64	3.64	4.49	3.61	24.79	6.61
Standard deviation of the number of days between two gambling days	1.66	5.92	1.61	4.85	2.35	14.14	4.54
Number of sessions in total	103.32	17.17	3.18	22.00	10.42	3.89	15.98
Average amount of money bet per game	1.10	1.20	0.34	0.86	1.18	0.57	0.72
Standard deviation of the amount of money bet per game	1.01	0.91	0.24	0.68	0.86	0.36	0.54
Average amount of money deposited per deposit	32.31	51.98	22.66	33.75	43.66	24.30	31.86
Standard deviation of the amount of money deposited per deposit	8.05	3.86	0.56	3.25	1.94	0.94	2.23
Average amount of money deposited per day	55.07	85.91	36.78	51.86	50.67	33.25	49.04
Average amount of money bet per day	537.86	1021.08	206.50	414.12	827.48	190.21	440.30
Average number of deposits per day	1.75	1.46	0.80	1.38	0.61	0.76	1.06
Average number of deposits per session	0.89	0.85	0.73	0.98	0.36	0.73	0.78
Average number of gambling sessions per day	2.18	1.90	1.11	1.47	1.17	1.13	1.38
Percentage of sessions terminating with low monetary balance	67%	52%	100%	85%	6%	97%	79%
N	9215	22,364	56,686	31,479	12,523	18,628	
N Percentage	6%	15%	38%	21%	8%	12%	

Table 4. Median and 90th percentile values for of each of six clusters calculated based on data from January to March 2023.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Age (in years)	41(59)	39(58)	36(54)	37(56)	38(59)	37(55)
Number of playing days	46(69)	7(22)	2(8)	16(30)	1(15)	3(7)
Average number of days between two gambling days	1.76(2.38)	3.86(12.25)	1(10.5)	3.87(7.82)	1(8.89)	20.5(45)
Standard deviation of he number of day between two gambling days	1.46(2.86)	4.1(14.29)	0(6.36)	3.94(9.8)	0(7.91)	13.69(28.29)
Number of sessions	83(180)	11(40)	2(8)	19(41)	1(32)	3(7)
Average amount of money bet per game	0.36(1.64)	0(1.15)	0(0.23)	0.12(1.21)	0(0.58)	0(0.57)
Standard deviation of the amount of money bet per game	0.17(1.46)	0(0.58)	0(0.05)	0(0.8)	0(0.13)	0(0.19)
Average amount of money deposited	20.62(63.9)	23.95(113.62)	11.33(50)	19.22(68.24)	11.35(103.95)	11.51(53.32)
Standard deviation of the amount of money deposited	1.41(21.7)	0(6.39)	0(0)	0(7.65)	0(0)	0(0)
Average amount of money deposited per day	32.86(125.15)	28.46(216.67)	5.73(80.51)	21.79(125.04)	3.8(115.6)	7.63(71.76)
Average amount of money bet per day	238.24(1116.8)	235.35(2048.75)	20.82(367.38)	131.44(908.13)	32.61(1308.67)	27.24(359.42)
Average number of deposits per day	1.44(3.23)	1(3)	0.43(2)	1(2.79)	0.2(1.64)	0.5(1.86)
Average number of deposits per session	0.81(1.54)	0.67(1.67)	0.5(2)	0.85(1.79)	0(1)	0.5(1.67)
Average number of session per day	1.96(3.26)	1.73(3)	1(1.43)	1.33(2)	1(3.12)	1(1.5)
Percentage of sessions terminating with low balance	0.7(0.7)	0.5(0.5)	1(1)	0.86(0.86)	0(0)	1(1)

deviation of the number of days between two gambling days was 4.54 days. The respective numbers for Cluster 5 were 3.61 days and 2.35 days.

k-means clustering is based the sum of squares and is therefore a parametric approach. The amount of money bet, amount of money deposited, and number of gambling sessions can potentially be very large. This could lead to skewed distributions which could in turn impact the k-means clustering. Table 4 reports the median values and the 90th percentile for each of the six clusters. Clusters 2 and 5 appeared to match the assumed patterns of binge gamblers most closely based on the analysis of the average cluster profiles. In Cluster 2, the median number of gambling days was 3.86 and the median number of sessions was 11. Therefore, players in Cluster 2 also had relatively many sessions on just a few gambling days based on median values. In Cluster 2, the 90th percentile for the average amount of money deposited per day was £216.67 and therefore higher than in any other cluster. This was also the case for the amount of money bet per day. The analysis based on the median and 90th percentile also support the assumption that Cluster 2 comprised gamblers who played relatively rarely but intensely.

The second cluster which appeared indicative of binge gambling based on the average values was Cluster 5. The median number of playing days and the median number of gambling sessions in Cluster 5 were both 1. However, the 90th percentile of the number of sessions was 32. This means that a few extreme players in this cluster were responsible for the high average number of sessions reported in Table 3. The same extreme ratio between the median value and 90th percentile can be observed for the average amount of money deposited and the amount of money bet per day. Cluster 5's median value for the average amount of money bet per day was £3.80, but the 90th percentile was £115.60. The respective values for the average amount of money deposited per day were £32.61 and £1308.67. Cluster 5's analysis of the median and 90th percentile also supported the analysis based on the median values. However, Cluster 5 was a more heterogeneous group than Clusters 1, 3, 4 and 6 because a few gamblers were responsible for the large number of sessions, amount of money deposited, and amount of money bet.

Figure 2 shows the deviation of Cluster 2 from the average for each player tracking feature in Table 3. The deviation was calculated as the difference between the cluster average and the total average, and the resulting value was divided by the total average. Cluster 2's number of playing days was only slightly lower than the total average. The largest deviation from the total average was observed for the average amount of money bet per day, and the average amount of money deposited per day.

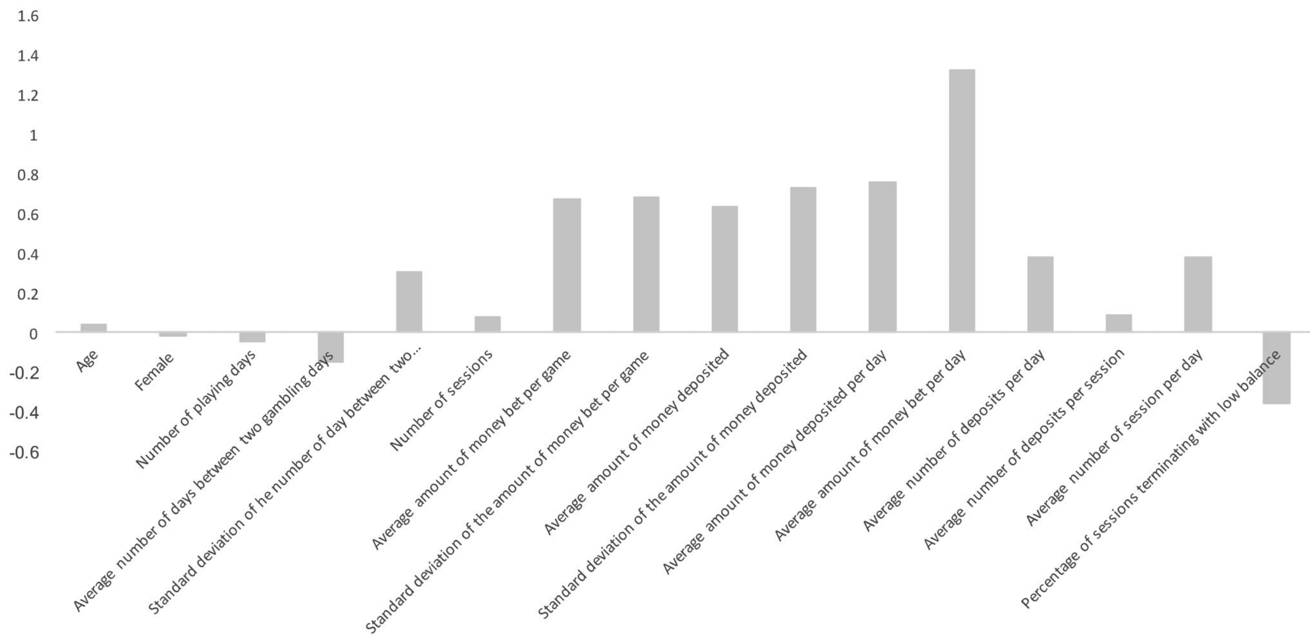

Figure 2. Deviation of Cluster 2 from the average values for each player tracking feature.

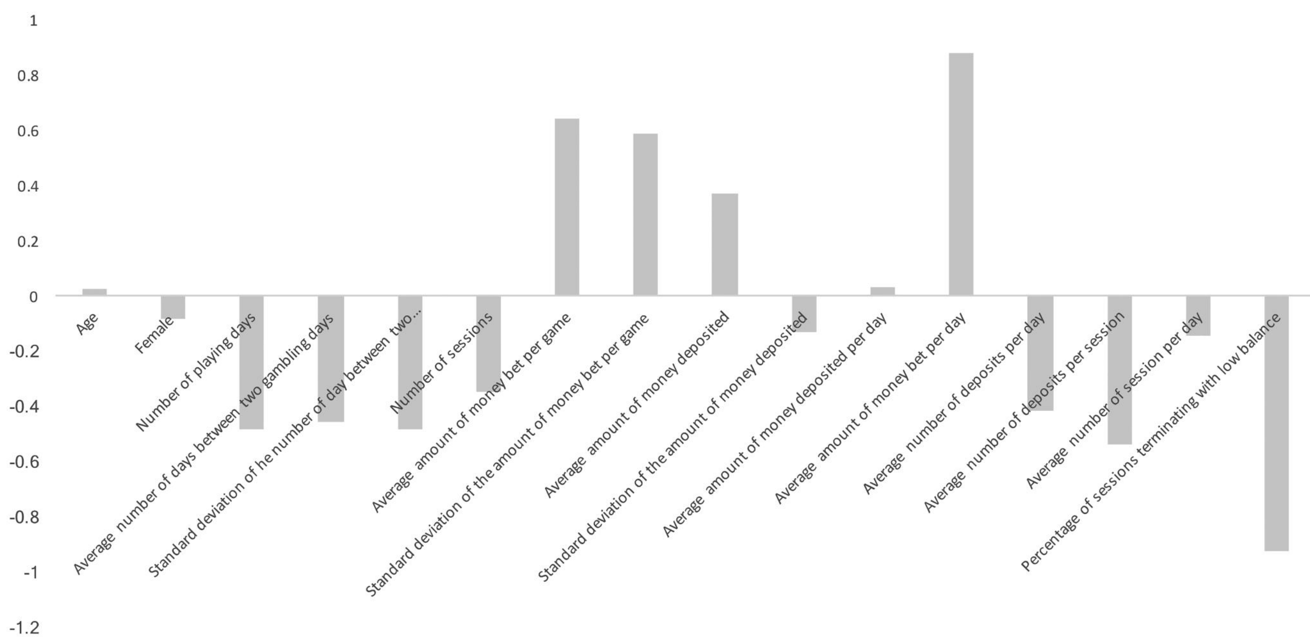

Figure 3. Deviation of Cluster 5 from the average values for each player tracking feature.

Figure 3 shows the deviation of Cluster 5 from the average for each player tracking feature. As with Cluster 2, the deviation was calculated as the difference between the cluster average and the total average, and the resulting value was divided by the total average. Cluster 5's number of playing days was smaller than the total average. The largest deviation from the total average was observed for the average amount of money bet per day, the amount of money bet per game, and the standard deviation of the amount of money bet per game. The percentage of sessions terminating with low balance was smaller than the average in Cluster 5.

The k-means solution with six clusters was applied to the behavior of the 150,895 players between April and June 2023. If all players behavior stayed the same, they would have been assigned to the same cluster. Table 5 reports the number of players which moved away or stayed in their cluster during April to June 2023. The rows indicate the original cluster and the columns report the cluster membership based on the playing behavior between April and June 2023. The main diagonal reports the number of players who remained in their respective cluster. A total of 6,450 players who were in Cluster 1 from January to March 2023 remained in Cluster 1. Moreover, 577 players from Cluster 1 did not place a single bet during April to June 2023. Out of the 150,895 players who placed at least one bet between January and March 2023, 25,867 players became inactive between April and June 2023 (17%).

Table 5. Number of players moving from each cluster to clusters 1 to 6 or inactivity cluster in April to June 2023.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	inactive	Total
Cluster 1	**6450**	721	8	1393	66	–	577	9215
Cluster 2	408	**13,578**	846	1869	1070	758	3835	22,364
Cluster 3	1	1241	**33,559**	2769	406	5210	13,500	56,686
Cluster 4	922	2714	2,838	**20,462**	97	1197	3249	31,479
Cluster 5	56	960	492	72	**6894**	418	3631	12,523
Cluster 6	–	1029	6108	1237	455	**8715**	1084	18,628
Total	7837	20,243	43,851	27,802	8988	16,298	25,876	150,895

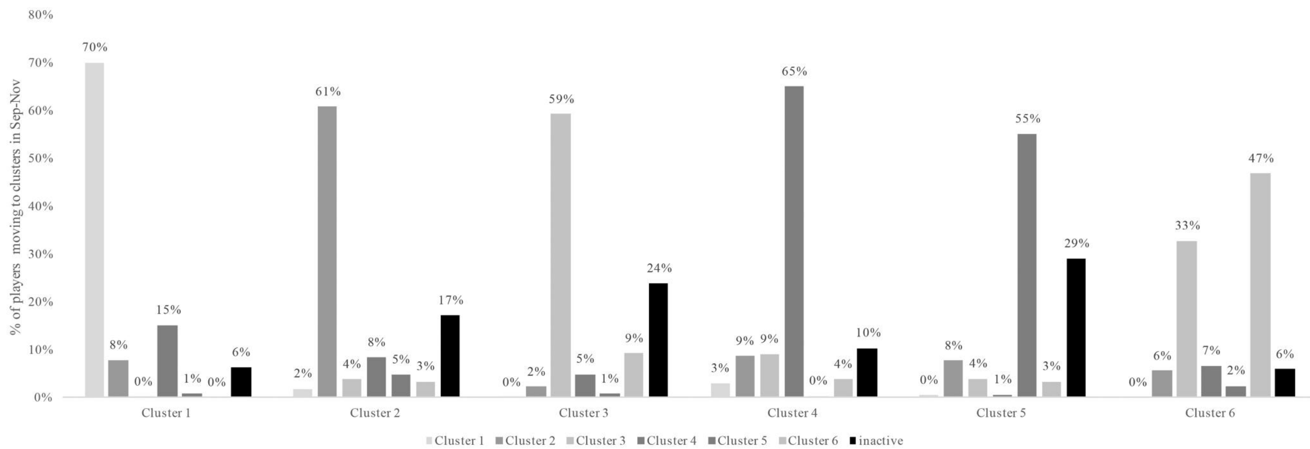

Figure 4. Percentage of players moving from each cluster to cluster 1–6 or inactivity in April to June 2023.

Figure 4 shows the percentage of players in each cluster moving to one of the six clusters or inactivity between April and June 2023. With 70%, Cluster 1 showed the largest percentage of players who remained in their cluster. Players in Cluster 1 had the largest number of gambling days and sessions. Only 6% of players in Cluster 1 stopped gambling between April and June 2023. Moreover, 29% of players in Cluster 5 and 24% of players in Cluster 3 became inactive. Cluster 5 was identified as potentially fitting the pattern of binge gambling. Here, 55% of players in Cluster 5 remained in the cluster. Cluster 2 was also identified as another potential group of binge gamblers. Here, 17% of gamblers in Cluster 2 became inactive and 61% remained in Cluster 2.

Discussion

The present study was carried out in an attempt to identify patterns indicative of binge gambling. Player tracking data from a secondary dataset of 150,895 British online gamblers were used. Every single bet, win, deposit, etc. for the time period between January and June were available in the secondary dataset. The average age was 38 years which is in line with samples from other online gambling studies with other gambling operators (e.g. Auer and Griffiths 2023a, 2023b). The authors attempted to compute player tracking variables which could reflect the episodic events of extreme gambling characteristic of binge gambling. Binge gamblers were hypothesized to gamble intensively on relatively few days. Gambling intensity was measured using the volume of deposits as well as frequency and the volume of the amount of money bet.

The 14 player tracking variables (excluding gender and age) were used to compute a k-means cluster analysis and a six-cluster solution was chosen because of the size of clusters and the uniqueness of the profiles of the clusters. Furthermore, a scree plot and silhouette scores were computed. The scree plot suggested only two clusters. Silhouette scores suggested a six-cluster solution with an average silhouette score of 0.44. This does not represent a very good homogeneity of clusters, but it was the maximum value apart from the one-cluster solution.

Two clusters (Cluster 2 and Cluster 5) gambled on a relatively low number of days, but displayed a high gambling intensity on those days. Their profiles could potentially match the habits of binge gamblers. Gamblers in Cluster 2 and Cluster 5 deposited and bet more money per day than in any other cluster. Moreover, their single bets per game and single deposits were also higher than in any other cluster. An analysis based on the median and 90th percentile confirmed the findings based on average values. This also provides additional support for the validity of two assumed binge gambling clusters.

Three other metrics (number of deposits per day, number of deposits per session, and percentage of sessions terminating with low balance) which have previously been reported as potentially indicative of loss of control and impulsivity (Auer and Griffiths 2023a) were also calculated. However, Cluster 3 and Cluster 5 did not display high values with respect to the three metrics. Players in the two clusters (if they are genuine binge gamblers) did not deposit frequently when active and also seemed to be able to stop gambling before they lost all their funds. This contradicts the assumptions that binge gamblers suffer from intermittent episodes of uncontrolled gambling (Nower and Blaszczynski 2003). On the other hand, there are no clear definitions of binge gambling, other than it is episodic and not continuous. In order to determine the stability of the gamblers' cluster membership, the cluster model was applied to data from April to June 2023. Each of the 150,895 gamblers cluster membership was calculated based on the behavior between April to June 2023.

The largest percentage of players (70%) who remained in their cluster was observed for Cluster 1. This was also the cluster with the highest gambling frequency. Gamblers in Cluster 1 gambled on average on 49 out of 90 possible days. The high gambling frequency carried on during the following three months. Only 6% of the gamblers in Cluster 1 stopped gambling entirely between April and June 2023. However, 17% and 29% of players in the two binge gambling clusters (i.e. Cluster 2 and Cluster 5) stopped gambling entirely between April and June 2023. Moreover, 61% and 55% of gamblers in Cluster 2 and Cluster 5 remained in their cluster, respectively. This does not contradict the nature of binge gambling because there are no pre-defined time periods between the gambling episodes. The authors simply wanted to test how stable the cluster membership was over time. Griffiths (2006) described a player who did not gamble for years and only started to develop problems again after relationship problems. In their longitudinal study of Canadian players, Williams et al. (2015) also found that problem gambling can cease for years after it reappears again.

Limitations

The present study has a number of limitations. First, given that the data only came from one operator, it is unlikely that the data represent the totality of an individual's gambling behavior because many individuals will have gambled with other operators both online and offline. Second, binge gambling might not be visible over such short periods as investigated in the present study. Third, there is always a possibility that more than one individual might be using a single account (e.g. a married couple) although the number of accounts where more than one individual was using it are likely to be low. Future replication studies should be conducted with data from different operators with different types of gamblers and utilize longitudinal data preferably covering several years.

Conclusions

The present study is the first to have used account-based player tracking data to study the concept of binge gambling. The findings presented here will be of interest to a number of different stakeholders including academic researchers in the gambling field, gambling regulators, and the gambling industry. The findings suggest that binge gambling may be able to be identified by online gambling operators using account-based tracking data. Given that binge gambling appears to have some association with problem gambling, gambling operators who have account-based data from online players and/or data from player cards can look for binge gambling behavior and provide interventions that they are already providing for other types of gamblers showing markers of gambling harm.

Ethical approval

Ethical approval was provided by the ethics committee of Nottingham Trent University.

Informed consent

Not applicable. Secondary data analysis.

Disclosure statement

The second author's university has received funding from *Norsk Tipping* (the gambling operator owned by the Norwegian Government). The second author has received funding for a number of research projects in the area of gambling education for young people, social responsibility in gambling and gambling treatment from *Gamble Aware* (formerly the *Responsibility in Gambling Trust*), a charitable body which funds its research program based on donations from the gambling industry. Both authors undertake consultancy for various gambling companies in the area of player protection and social responsibility in gambling.

Funding

None received.

Data availability statement

The data for this study are not available due to commercial sensitivity.

References

American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington (VA): American Psychiatric Publishing.

Auer M, Griffiths MD. 2023a. An empirical attempt to operationalize chasing losses in gambling utilizing account-based player tracking data. J Gambl Stud. doi: 10.1007/s10899-022-10144-4.

Auer M, Griffiths MD. 2023b. Using artificial intelligence algorithms to predict self-reported problem gambling with account-based player data in an online casino setting. J Gambl Stud. 39(3):1273–1294. doi: 10.1007/s10899-022-10139-1.

Brusco MJ, Cradit JD. 2001. A variable-selection heuristic for k-means clustering. Psychometrika. 66(2):249–270. doi: 10.1007/BF02294838.

Catania M, Griffiths MD. 2022. Applying the DSM-5 criteria for gambling disorder to online gambling account-based tracking data: an empirical study utilizing cluster analysis. J Gambl Stud. 38(4):1289–1306. doi: 10.1007/s10899-021-10080-9.

Challet-Bouju G, Hardouin JB, Thiabaud E, Saillard A, Donnio Y, Grall-Bronnec M, Perrot B. 2020. Modeling early gambling behavior using indicators from online lottery gambling tracking data: longitudinal analysis. J Med Internet Res. 22(8):e17675. doi: 10.2196/17675.

Cowlishaw S, Nespoli E, Jebadurai JK, Smith N, Bowden-Jones H. 2018. Episodic and binge gambling: an exploration and preliminary quantitative study. J Gambl Stud. 34(1):85–99. doi: 10.1007/s10899-017-9697-z.

Dragičević S, Tsogas G, Kudic A. 2011. Analysis of casino online gambling data in relation to behavioural risk markers for high-risk gambling and player protection. Int Gambl Stud. 11(3):377–391. doi: 10.1080/14459795.2011.629204.

Finkenwirth S, MacDonald K, Deng X, Lesch T, Clark L. 2021. Using machine learning to predict self-exclusion status in online gamblers on the PlayNow.com platform in British Columbia. Int Gambl Stud. 21(2):220–237. doi: 10.1080/14459795.2020.1832132.

Griffiths MD. 2006. A case study of binge problem gambling. Int J Ment Health Addict. 4(4):369–376. doi: 10.1007/s11469-006-9035-7.

Gupta R, Derevensky J. 2011. Defining and assessing binge gambling. In: Derevensky JL, Shek DT, Merrick J, editors. Youth gambling: the hidden addiction. Berlin: De Gruyter. p. 79–97.

Hing N, Breen H, Gordon A. 2012. A case study of gambling involvement and its consequences. Leisure Sciences. 34(3):217–235. doi: 10.1080/01490400.2012.669682.

Hopfgartner N, Auer M, Griffiths MD, Helic D. 2023. Predicting self-exclusion among online gamblers: an empirical real-world study. J Gambl Stud. 39(1):447–465. doi: 10.1007/s10899-022-10149-z.

Letı R, Ortiz MC, Sarabia LA, Sánchez MS. 2004. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. Anal Chim Acta. 515(1):87–100. doi: 10.1016/j.aca.2003.12.020.

Likas A, Vlassis N, Verbeek JJ. 2003. The global k-means clustering algorithm. Pattern Recognit. 36(2):451–461. doi: 10.1016/S0031-3203(02)00060-2.

Mohamad IB, Usman D. 2013. Standardization and its effects on K-means clustering algorithm. RJASET. 6(17):3299–3303. doi: 10.19026/rjaset.6.3638.

Nower L, Blaszczynski A. 2003. Binge gambling: a neglected concept. Int Gambl Stud. 3(1):23–35. doi: 10.1080/14459790304589.

Percy C, França M, Dragičević S, d'Avila Garcez A. 2016. Predicting online gambling self-exclusion: an analysis of the performance of supervised machine learning models. Int Gambl Stud. 16(2):193–210. doi: 10.1080/14459795.2016.1151913.

Petry NM, Blanco C, Auriacombe M, Borges G, Bucholz K, Crowley TJ, Grant BF, Hasin DS, O'Brien C. 2014. An overview of and rationale for changes proposed for pathological gambling in DSM-5. J Gambl Stud. 30(2):493–502. doi: 10.1007/s10899-013-9370-0.

Perrot B, Hardouin JB, Grall-Bronnec M, Challet-Bouju G. 2018. Typology of online lotteries and scratch games gamblers' behaviours: a multilevel latent class cluster analysis applied to player account-based gambling data. Int J Methods Psychiatr Res. 27(4):e1746. doi: 10.1002/mpr.1746.

Sklar A, Gupta R, Derevensky J. 2010. Binge gambling behaviors reported by youth in a residential drug treatment setting: a qualitative investigation. Int J Adolesc Med Health. 22(1):153–162. doi: 10.1515/ijamh.2010.22.1.153.

Stinchfield R. 2002. Reliability, validity, and classification accuracy of the South Oaks Gambling Screen (SOGS). Addict Behav. 27(1):1–19. doi: 10.1016/s0306-4603(00)00158-1.

Van Rossum G. 2007. Python programming language. [accessed 2022 June 25]. https://www.python.org.

Williams RJ, Hann RG, Schopflocher D, West B, McLaughlin P, White N, King K, Flexhaug T. 2015. Quinte longitudinal study of gambling and problem gambling. Guelph (ON): Ontario Problem Gambling Research Centre.

World Health Organization. 2019. International statistical classification of diseases and related health problems (11th revision). Geneva: World Health Organization.

Yoon KA, Kwon OS, Bae DH. 2007. An approach to outlier detection of software measurement data using the k-means clustering method. First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007) (pp. 443–445). IEEE. doi: 10.1109/ESEM.2007.49.

Appendix 1. List of variables carried out for January-March 2023

Number	Variable
1	Age (in years)
2	Female (percentage)
3	Number of gambling days
4	Average number of days between two gambling days
5	Standard deviation of the number of days between two gambling days
6	Total number of gambling sessions
7	Average amount of money bet per game (£)
8	Standard deviation of the amount of money bet per game
9	Average amount of money deposited across all the single deposits (£)
10	Standard deviation of the amount of money deposited (£)
11	Average amount of money deposited per day (£)
12	Average amount of money bet per day (£)
13	Average number of deposits per day
14	Average number of deposits per session
15	Average number of gambling sessions per day
16	Percentage of sessions terminating with a low monetary balance