Player Profiling in Borderlands Science: Mapping Behavioural Clusters to Player Longevity

Name: Hazel Foo Jia Jue (261194705)

Supervisor: Prof. Robert Glew

Abstract

Citizen science games represent a platform for scalable data annotation, however, their long-term viability hinges on understanding and maintaining player engagement. This study analyses player behaviour in Borderlands Science, a citizen science game embedded within a commercial video game, Borderlands 3, to identify factors driving long-term participation. Using a combination of survival analysis and K-means clustering on player activity data, three distinct behavioural archetypes were identified: Speedrunners, Completionists, and Explorers. The results demonstrate that in-game behaviours such as performance consistency and session intensity influence player longevity differently depending on the player's underlying archetype. These findings challenge one-size-fits-all engagement models and lay the foundation for designing targeted interventions to reduce churn and improve the productivity of citizen science platforms.

1 Introduction

Modern artificial intelligence (AI) is rapidly transforming domains ranging from medical diagnosis to self-driving vehicles and robotics. However, the performance of these models is limited by the scale, quality, and representativeness of their training data. To create these large and high-quality labelled datasets, manual annotation by trained experts is required, a process that is prohibitively expensive and time-consuming. As a result, data availability remains a bottleneck in the advancement of AI systems. Studies have shown that developing a single dataset for a complex task can take thousands of person-hours and cost hundreds of thousands of dollars (Weyand et al., 2020). This challenge is particularly acute in specialised domains like medicine, where the need for expert annotators drastically increases costs and creates significant scalability issues (Irvin et al., 2019). Furthermore, automated or semi-automated approaches remain sensitive to label noise and bias (Northcutt et al., 2022).

To address this, crowdsourcing has emerged as a widely adopted strategy to scale annotation efforts cost-effectively, leveraging a global pool of contributors to deliver large volumes of labelled data rapidly (Karger et al., 2014; Vaughan, 2018). However, this approach introduces its own challenges, including inconsistent quality, intentional errors or careless submissions, and skewed results due to uneven participation. These issues must be managed through careful task design, robust aggregation algorithms, and effective incentive structures (Gu et al., 2022; Karger et al., 2014; Vaughan, 2018).

A specialised form of crowdsourcing is the citizen-science game (CSG), which gamifies annotation or problem-solving tasks to harness collective human intelligence at scale. By embedding scientific objectives into engaging game mechanics, CSGs motivate participants through intrinsic enjoyment, competition, or community contribution rather than direct financial reward. Successful examples include Foldit, where players solved complex protein-folding problems that had eluded automated algorithms (Cooper et al., 2010); EyeWire, where players mapped neural connections in the retina (Kim et al., 2014); and EteRNA, which advanced RNA design through community-based gameplay (Lee et al., 2014). These projects demonstrate with a well-defined feedback loop, meaningful incentives, and an appropriate learning curve, non-expert players can also produce scientifically valuable outputs (Cooper et al., 2018; Curtis, 2014; Khatib et al., 2011). In contrast, less successful initiatives often fail to sustain engagement due to poorly balanced reward systems, limited task variety, or insufficient integration between scientific goals and gameplay (Prestopnik and Crowston, 2013). This underscores the importance of not only task design, but also on understanding the behavioural and motivational dynamics of the player communities for the long-term viability of CSGs.

Within this landscape, Borderlands Science marks a breakthrough in cross-platform engagement, embedding a citizen-science task of aligning microbial DNA sequences for biomedical research within a major commercial video game, Borderlands 3. Since its launch, the project has engaged over 4 million players globally who collectively solved over 135 million puzzles, a feat that would otherwise be unachievable through conventional means, which has improved microbial phylogeny estimations beyond state-of-the-art computational methods (Sarrazin-Gendron et al., 2025). Ongoing academic work by Farajollahzadeh and Glew (in preparation) examines how the in-game reward structures, specifically minimum performance targets, affect the quality of solutions and participant longevity within this platform. Their work provides crucial insights into the macro-level levers of engagement.

However, the limitation of such a top-down analysis is its treatment of the player base as a homogeneous group. In reality, player populations are heterogeneous, consisting of distinct segments whose engagement is driven by different motivations and consequently, respond in different ways to fixed interventions (Anand and Peterson, 2000). Interventions designed to improve solution quality or longevity that are based on population-level averages may not be robust, as they can mask the diverse motivations and behaviours of distinct player segments. A one-size-fits-all approach risks being ineffective or even counterproductive for specific, valuable subgroups of players.

Therefore, this study aims to address this gap by conducting a granular behavioural analysis of the Borderlands Science player base. Using a combination of survival analyses and unsupervised clustering, this paper seeks to identify distinct behavioural groups and determine the specific factors that drive long-term engagement for each of them. The findings improve our understanding of the player ecosystem, offering new insights for designing targeted interventions.

2 Method

This section outlines the data and analyses conducted in this study. It first describes the dataset collected from a subset of players who participated in the Borderlands Science experiment, followed by an overview of the survival analysis and clustering methods used to characterise the player profiles within the game platform.

2.1 Data Description

From the pool of over 4 million participants in Borderlands Science, this study examines a subset of 5,000 players who successfully progressed to Level 9 between April 2020 and June 2025. This subset of players represents the experiment's most engaged participants who contributed at least 225 solutions. Understanding the drivers behind their engagement is critical as these players are responsible for a substantial portion of high-quality data.

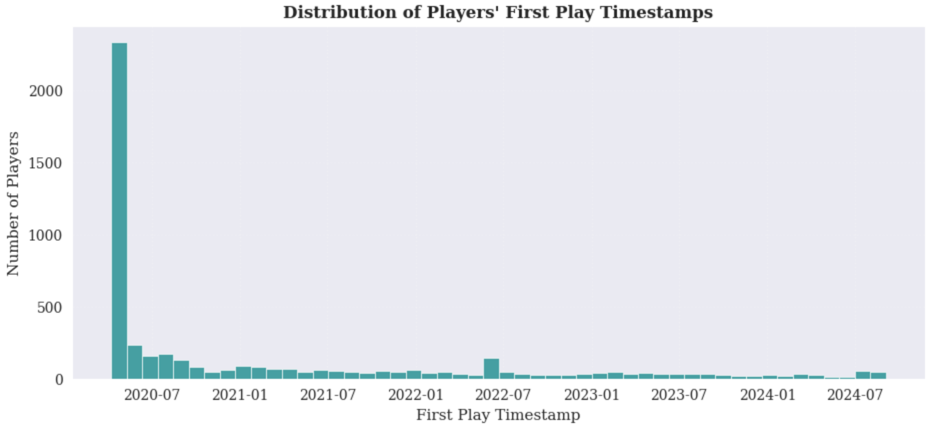

Figure 1: Histogram plot of players' first recorded activity

Figure 1 shows the distribution of players' first recorded activity, with a large proportion (65.0%) of them joining in 2020, within the first year of Borderlands 3 release. To assess the impact of the game's release cycle on long-term engagement, the year and month of first activity were extracted and kept as separate features. The year was normalised relative to the earliest recorded date, while the month was encoded categorically.

2.1.1 Feature Engineering

Using click-trace data collected from 1,213,847 puzzles, extensive feature engineering was performed to capture key behavioural patterns. These features can be grouped into four primary categories, each associated with a different aspect of a player's interaction.

Engagement Patterns and Behavioural Rhythms

Understanding the behaviour of players is crucial for modeling longevity, because the frequency, duration, and intensity of play directly reflect the player's commitment. Analysing this behaviour provides insights into players' engagement habits that are associated with long-term data contribution.

The metrics within this category capture the length and frequency of interactions, independent of performance scores. Puzzles played consecutively within 15 minutes of each other were considered to be part of the same session, and sessions lasting at least 30 minutes were considered ”long”. Key features computed from this process include the total number of days with activity (active days), average time between sessions and puzzles, and variables describing session length, duration and intensity (i.e. number of puzzles completed per minute). Refer to Table 3 in the Appendix for full list of metrics.

Early Commitment and Foundation Building

The foundational phase (Levels 1–4) represents a critical period in shaping long-term player retention and engagement. Behavioural patterns established during this early stage serve as strong predictors of future commitment and success (van der Weiden et al., 2020). This phase functions as a training period where players learn basic puzzle-solving skills. Within this category, two key behaviours are captured: initial overachievement and learning dynamics. The former reflects the tendency to exceed the minimum level requirements, measured by the presence and volume of additional puzzles solved beyond those required for level progression in Levels 1–4. The latter is quantified through the rate of change and predictability of performance metrics (time-per-click and time-per-score), derived from linear regression models tracking each player's initial exposure to the game. Refer to Table 4 in the Appendix for full list of metrics.

Intrinsic Motivation and "Mastery Drive"

Building on the early foundation-building phase (Levels 1–4), which captures the formation of fundamental skills, and the engagement pattern measures, which describe distinct play styles, the metrics in this category aim to capture intrinsic motivation. According to Self-Determination Theory (Ryan and Deci, 2000), sustained participation arises when players transition from external reinforcement to internal satisfaction. This category therefore captures behaviours associated with self-directed engagement with the puzzles. These features capture two key behavioural dimensions: First, sustained overachievement reflects the player's ability to maintain high-performance standards beyond the minimum requirements. It measures the volume of additional puzzles completed across Levels 5–8. Second, exploratory behaviour quantifies additional actions performed after achieving the target score for puzzle completion; behaviours that indicate experimentation. Refer to Table 5 in the Appendix for full list of metrics.

Performance Consistency and Quality

Another key determinant of long-term player retention is the quality and sustainability of effort. In CSGs, consistent effort is particularly important, as it supports the reliability and integrity of the data generated. Performance quality in this study was quantified with "*excess*": the additional points achieved above the target score for a given puzzle. Excess captures a player's commitment to performance beyond the minimum requirements for puzzle completion. To account for differences in puzzle difficulty and target scores, excess was first standardised as points above target as a ratio relative to the puzzle's target score (*excess ratio*). Thereafter, normalised by the median excess ratio for each given level to enable comparison across levels (*relative excess*).

Once defined, further aggregate features were calculated from the individual puzzle level excess: these features include measures of variability in excess performance, correlations with puzzle count, and consistency of exploratory behaviour. Refer to Table 6 in the Appendix for full list of metrics.

2.1.2 Data transformation

After feature engineering was completed, a data transformation pipeline was implemented to account for skewness in numeric variables while maintaining the interpretability of key metrics. This step was crucial for later analyses: the penalised Cox model used for survival analysis requires features on a comparable scale to produce stable coefficient estimates, while K-Means relies on distance measures that are highly sensitive to feature magnitude.

Square root transformations were applied to *Excess Ratio* and *Relative Excess* to account for their skewed distributions. Figures 2A and 2B illustrate the original distributions of these variables across the player base. Correlation coefficients, binary indicators, and bounded proportions were retained in their original scales to maintain interpretability.

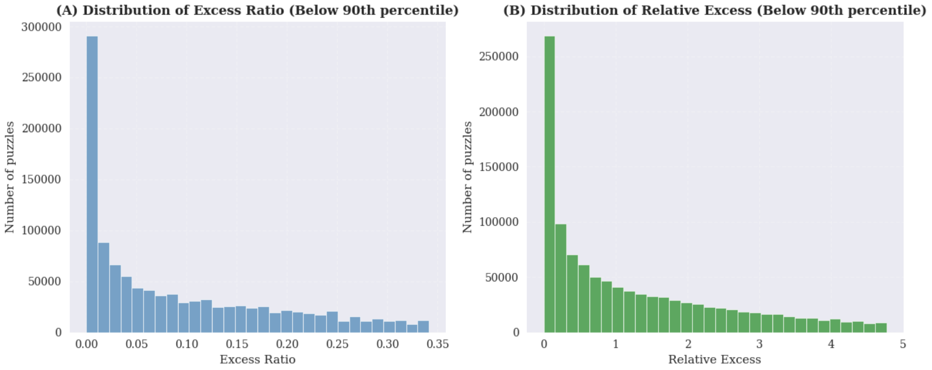

Figure 2: Distribution of (A) excess ratio and (B) relative excess for top 90% players

* Note: Given the extreme right-tail values, only the top 90% of the data are shown for visualisation clarity.

The remaining transformation decision process follows the principles displayed in Figure 3.

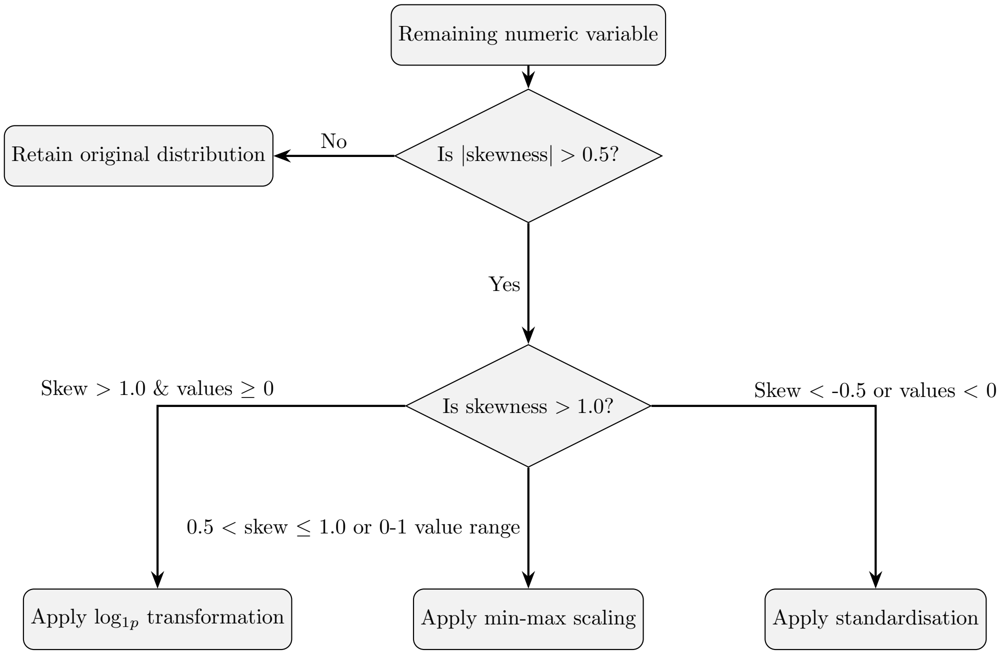
Remaining numeric variable
Is |skewness| > 0.5?
No
Retain original distribution
Yes
Is skewness > 1.0?
Skew > 1.0 & values ≥ 0
0.5 < skew ≤ 1.0 or 0-1 value range
Skew < -0.5 or values < 0
Apply $\log_{1p}$ transformation
Apply min-max scaling
Apply standardisation

Figure 3: Decision logic for skewness-based transformation of numeric variables

2.2 Survival Analysis

Once data transformation was completed, survival analysis was implemented to model player churn. Specifically, a Cox Proportional Hazards regression model was employed, as it provides a flexible semi-parametric method for survival data (Cox, 1972), allowing for the simultaneous assessment of multiple behavioural factors on player churn.

2.2.1 Elastic Net Penalised Cox Proportional Hazards Model

To model player dropout behaviour in Borderlands Science, Cox Proportional Hazards model with Elastic Net was chosen. Elastic net was selected because player engagement features are inherently correlated, and it combines L1 (Lasso) and L2 (Ridge) penalties, making it preferable to alternatives such as Lasso in handling correlated features and high-dimensional data (Zou and Hastie, 2005). This characteristic ensures the model can automatically select among correlated features while maintaining stability. Furthermore, the L2 (Ridge) component shrinks coefficients of redundant features toward zero without complete exclusion and prevents overfitting in the high-dimensional feature space. This process yields a more robust, stable, and interpretable set of coefficients for understanding player dropout risk.

2.2.2 Survival Metrics

In this study, the Cox model was implemented using an events-based survival framework. The survival duration was defined as the cumulative number of Level 9 puzzles completed by each player, rather than chronological time. This provides a more direct measure of engagement than the conventional time-at-risk, which would introduce inaccuracies due to the lack of explicit player timezone information.

The dropout event indicator was assigned using a player-specific censoring approach that accounts for individual engagement patterns. This dynamic method is essential because using a single, static inactivity period (e.g., 30 days) is inadequate for heterogeneous player populations, and risks misidentifying slow-but-consistent contributors as dropouts. In this process, an inactivity threshold was calculated for each player, defined as 2.3× their average time between sessions. The 2.3× multiplier was chosen based on an empirical analysis of the entire player population. The ratio of each player's maximum observed session gap to their average session gap was calculated, which represents their limit of inactivity before their behaviour deviates significantly from their norm. The mean of this maximum-to-average ratio across all players was found to be 2.3 (Figure 4).

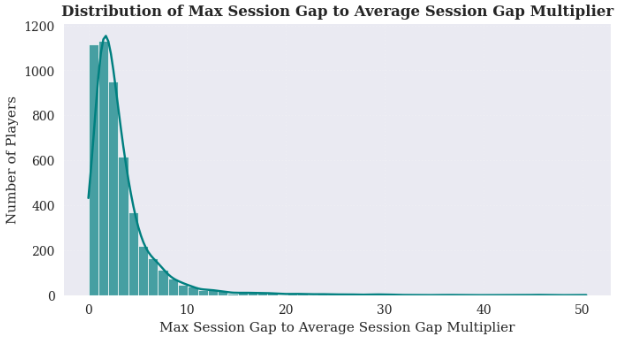

Figure 4: Histogram distribution of maximum session gap to average session gap multiplier

If the time elapsed since their last recorded activity and the final data collection date exceeded their individual inactivity threshold, they were considered to have dropped out (event = 1). Conversely, players below this threshold were considered to be still active and therefore, censored (event = 0). Of the 5,000 players, only two were censored. As shown in Figures 5A and 5B, which illustrate the distribution of level 9 puzzles completed, over 20% of players completed only one. This indicates that once players had obtained all in-game rewards from levels 1–8, a substantial proportion disengaged almost immediately after reaching level 9. These patterns suggest heterogeneity in player engagement and reinforce the importance of clustering to identify distinct behavioural archetypes.

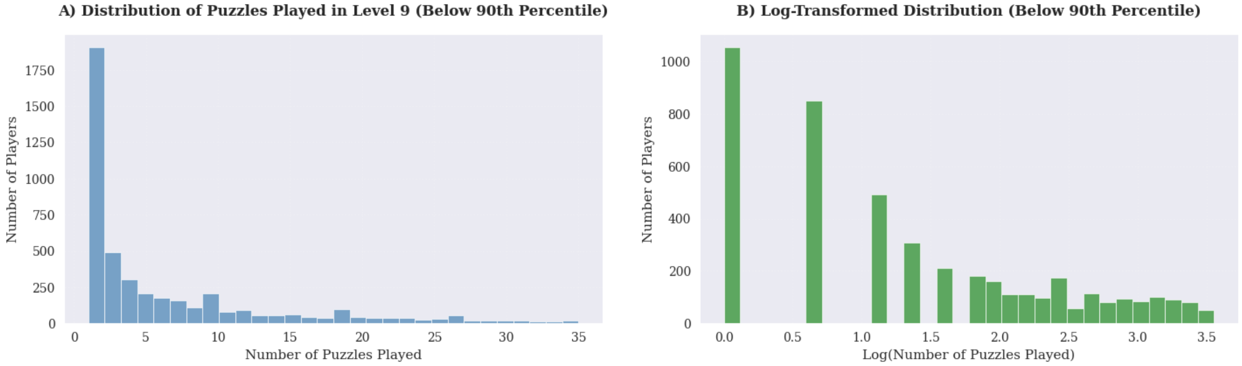

Figure 5: Histogram of number of Level 9 puzzles completed for top 90% players in (A) Original scale (B) After log-transformation

* Note: Given the extreme right-tail values, only the top 90% of the data are shown for visualisation clarity.

2.2.3 Model Specification and Hyperparameter Tuning

In the implementation of the Cox model, hyperparameter tuning was conducted using a repeated train-test validation approach. The regularisation strength (α) was explored across 12 logarithmically spaced values from 10^{-6} and 10, and the L1 ratio controlling the balance between the L1 and L2 penalties was evaluated across five values (0.1, 0.3, 0.5, 0.7, 0.9). Model performance was assessed using repeated random sub-sampling validation. For each iteration, the dataset was randomly partitioned into 80% training and 20% testing subsets. The concordance index (C-index) was computed on each test set to evaluate predictive performance across hyperparameter combinations. This iterative process mitigates variability from any single data partition and ensures robust hyperparameter selection.

2.2.4 Feature Selection and Consensus

Features were selected based on a consensus approach across validation runs. A feature was retained if it had a non-zero coefficient in at least 50% of the runs, ensuring that only stable and reproducible predictors were identified. Among these features, they were ranked by the frequency of their non-zero coefficient, and only the top 20 most frequently selected features were chosen for subsequent clustering analysis. This approach ensured that the identified player groups through clustering were defined exclusively by the behavioural traits found to be the most influential on retention.

2.3 K-Means Clustering

Using the top 20 features selected from the Cox model, along with the cumulative number of Level 9 puzzles completed, player behaviour archetypes were identified using K-Means clustering after feature standardisation. Although Level 9 puzzle count was the duration variable in the survival analysis, it was included in the unsupervised clustering process. The objective of this analysis was to identify distinct player groups rather than to predict performance outcomes; therefore, including the Level 9 puzzle count was essential to capture behavioural differences among players. K-Means was selected for its computational efficiency on large datasets and its interpretability, as the resulting cluster centroids provide a clear and quantifiable representation of each player group.

To establish a robust and interpretable player segmentation, the optimal number of clusters was determined using two widely recognised validation methods: the Elbow Method and the Calinski-Harabasz Index. The Elbow Method plots the model's inertia, which evaluates the internal coherence of the clusters by measuring the sum of squared distances between each data point and its assigned centroid, where a lower score signifies tighter, better-defined clusters. The Calinski-Harabasz Index evaluates the quality of the clustering by measuring the ratio of between-cluster dispersion to within-cluster dispersion, where a higher score signifies a better-defined partition.

To further investigate the behaviours unique to each archetype, a cluster-specific correlation analysis was performed. For each individual cluster, a separate correlation matrix was produced, specifically examining the correlation between each feature and the dependent variable (i.e. the number of Level 9 puzzles completed). The objective was to identify the behaviours associated with longevity that are unique to each group. Though this is not a causal inference analysis, it provides valuable insight into the heterogeneity and associations present within the feature set.

3 Results

3.1 Elastic Net Penalised Cox Regression

A total of 300 penalised Cox models were tested across different hyperparameters. Instead of relying on a single model, features predictive of player longevity were identified based on their consistent selection across the models (Table 1). Features from all four behavioural categories were represented among the top 20 predictors, indicating that player longevity is influenced by multiple facets of player behaviour.

Table 1: Top 20 features in descending order of selection frequency

Feature	**Selection frequency (%)**	**Direction Agreement (%)**	**Average HR**
relative excess puzzle corr	78	100	0.67
active days	77	100	0.80
average session intensity	74	59	1.07
excess ratio puzzle corr	72	100	0.46
level 7 excess indicator	70	100	0.58
average puzzles per session	70	100	1.38
level 1 excess indicator	70	100	0.73
years since 2020	67	100	1.07
level 8 excess indicator	66	100	0.76
average level progress effect	66	86	0.75
max session intensity	65	100	1.38
month 5	65	88	0.96
month 7	65	100	1.09
average session excess std	64	100	1.63
std proportion clicks after target	63	100	0.18
std time per click after target	62	100	1.12
month 9	62	90	0.91
level 2 excess indicator	62	100	0.52
average session relative excess	62	100	0.79
month 2	62	100	1.07

[1] Selection frequency: Number of times a feature was selected relative to the total number of models tested.

[2] Direction agreement: Proportion of times the direction of the Hazard Ratio (HR) agreed across all model iterations.

[3] Level Progression Excess Indicators (Levels 1, 2, 7, and 8): These are binary indicators that reflect whether a player completed more puzzles than the minimum required to advance past the respective level. This is distinct from the Puzzle Score Excess metrics (i.e., relative excess and excess ratio) which quantify the additional points achieved beyond the target score for the completion of individual puzzles.

3.1.1 Protective factors

Metrics associated with high engagement, efficiency, and consistent progress are significantly associated with a significantly reduced likelihood of player dropout.

The strongest protective effect is observed in metrics related to efficient interaction. Specifically, low variability in the proportion of clicks submitted after achieving the target score (std proportion clicks after target) is associated with a 82% reduction in the risk of dropout. Furthermore, players who consistently completed more puzzles than necessary for level progression demonstrated a strong, protective effect against dropout. This reduction in dropout risk is particularly pronounced across Levels 1, 2, 7, and 8, with the strongest effect recorded in Level 2 (48% reduction), followed by Level 7 (42% reduction), Level 1 (27% reduction), and Level 8 (24% reduction).

Several other engagement indicators are also significantly associated with a reduction in HR. Notably, the tendency for a player's quality of effort to improve as their puzzle count increases (excess ratio puzzle corr) is associated with a 54% reduction in dropout risk. Similarly, maintaining a link between the overall volume of play and the quality of performance across sessions (relative excess puzzle corr) is associated with a 33% reduction in risk. Moderate protective factors include the average level progress effect, which is associated with a 25% reduction in risk. This metric, calculated as the average difference between the excess ratio at the end and the start of a level, indicates that players whose quality of effort increases throughout a level are less likely to drop out. Finally, general indicators of activity and efficient play, such as average session relative excess and the number of days with puzzle activity, are associated with lower risk, reducing dropout by 21% and 20% respectively. These findings collectively indicate that both a greater duration of play and achieving higher scores beyond the minimum requirement are associated with a reduced risk in player attrition.

3.1.2 Risk factors

Conversely, factors related to volume, intensity, or variability in player behaviour were associated risk of dropout. The most significant risk indicator is high variability in the amount of additional points scored by a player between sessions, which is associated with an increase in dropout risk by 63%. Metrics related to high session load also increase risk, as both a high average number of puzzles completed in a session (average puzzles per session) and a high peak level of effort exerted in a single session (max session intensity) increase dropout risk by 38%. A high average level of effort exerted across all sessions (average session intensity) shows a slight but noticeable increase in dropout risk of 7%. The finding that the direction agreement for the average session intensity is close to the midpoint (59%) suggests this is a less stable predictor compared to the other factors in this group.

Conversely, factors related to player behaviour surrounding high volume, intensity, or variability are associated with an increased risk of dropout.

The most significant risk indicator appears to be the high variability in the additional score achieved for each puzzle between sessions (i.e., average session excess std). This variability appears to be associated with a substantial 63% increase in dropout risk. Metrics related to high session load also significantly increase risk. Specifically, both a high average number of puzzles completed in a session (average puzzles per session) and a high peak level of effort exerted in a single session (max session intensity) are associated with a 38% increase in dropout risk. A high average level of effort exerted across all sessions (average session intensity) is also associated with an increased risk of dropout, though only slightly, at 7%. However, this risk factor is considered less stable compared to the other risk factors, as its direction agreement of (59%) is very close to the midpoint (50%) to be a robust predictor.

3.1.3 Temporal and other behavioural factors

Temporal features displayed a mixed impact on dropout risk. On one hand, certain periods, specifically Month 7 and Month 2, are associated with a minor increase in hazard rate, at 9% and 7% respectively. The factor representing a player's enrollment year (normalised relative to the 2020) also indicates a 7% increase in dropout risk, suggesting an overall rising trend in attrition over time. Conversely, specific seasonal periods like Month 9 and Month 5 are associated with a protective effect, reducing the risk of dropout by 9% and 4%, respectively.

Interestingly, behaviour variability after achieving puzzle goals has opposite effects. Players showing greater variability in the number of clicks after achieving minimum requirement are associated with a 82% reduction in dropout risk, while those with greater variability in the timing of these clicks are associated with an increased HR.

However, given the highly inconsistent nature of these effects and the absence of discernible pattern for interpretation, these features were omitted from the clustering analysis.

3.2 Player Archetypes

Analysis across a range of potential cluster size (k values) revealed that $k = 3$ achieved the best trade-off between complexity and precision. Figure 6 shows two plots: (A) The Elbow Method plot indicates the optimal number of clusters as $k = 3$, which corresponds to the point of maximum curvature where the rate of decrease in inertia sharply declines. (B) The Calinski-Harabasz Index plot corroborates this finding, showing that the highest score, which signifies the best-defined partition, also occurs at $k = 3$.

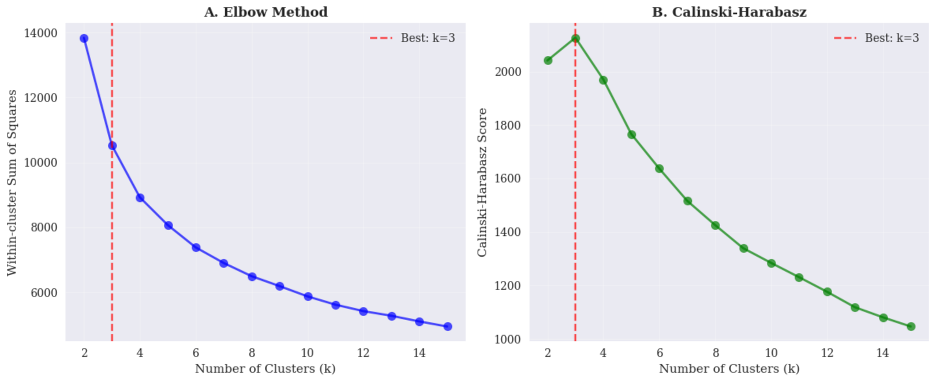

Figure 6: Cluster size determination using (A) Elbow Method, and (B) Calinski-Harabasz index

Tables 2 and 3 illustrate the qualitative and quantitative overview of the defining features of each cluster. Refer to Figure 10 in the Appendix for the full descriptive statistical breakdown of each cluster.

Table 2: Qualitative Overview of Clusters

Cluster	**Cluster Size (%)**	**Defining Features**
0	2680 (53.6%)	High average session intensity and average number of puzzles completed per session. Low average session relative excess and Level 9 puzzle count.
1	1292 (25.8%)	High number of active days, average session relative excess, and Level 9 puzzle count.
2	1028 (20.6%)	High number of players who completed more puzzles than required in Level 8. Moderate average level progress effect and Level 9 puzzle count.

Table 3: Descriptive statistics for distinctive features across clusters

Cluster	**Puzzle rel. excess corr**		**Active Days**		**Avg Puzzles per Session**		**Avg Session Rel. Excess**		**Level 9 Puzzles**	
	Mean (SD)	Min/Max	Mean (SD)	Min/Max	Mean (SD)	Min/Max	Mean (SD)	Min/Max	Mean (SD)	Min/Max
0	0.02 (0.27)	-0.63 0.79	2.98 (2.17)	1 16	81.10 (59.81)	9.38 225	1.00 (0.44)	0.12 2.30	4.22 (5.49)	1 62
1	0.20 (0.25)	-0.60 0.77	8.15 (7.82)	1 104	32.72 (30.17)	1.45 225	1.33 (0.45)	0.09 2.59	49.79 (141.00)	1 3210
2	0.08 (0.27)	-0.72 0.69	4.03 (3.68)	1 31	67.09 (56.36)	4.69 225	1.06 (0.46)	0.13 2.17	9.57 (23.87)	1 551

Note: Values are presented as Mean (Standard Deviation), and Minimum/Maximum which are stacked on top of one another.

3.2.1 Cluster 0

Cluster 0 comprises the majority of players at 53.6%. This cluster is defined by its focus on throughput, exhibiting both the highest average number of puzzles completed per session and the highest average rate of puzzles solved per minute. Players in this cluster engage in few but extended sessions and are generally unlikely to exceed puzzle progression requirements. The relationship between excess performance and puzzle volume was also the lowest of all clusters, suggestive of their focus on rapid completion over quality. Consequently, this cluster exhibits the lowest number of Level 9 puzzles completed. Given these features, especially the high-intensity bursts aimed at rapid progression, these players are designated the "Speedrunners."

Regarding their longevity, four features showed significant correlations with the final Level 9 puzzle count (refer to Figure 7). For the Speedrunners, both the number of active days and the average session relative excess demonstrated a negative correlation coefficient. This indicates that higher values in these metrics are ironically associated with a lower Level 9 puzzle count (i.e., the more effort they exert, shorter the retention). In contrast, the most direct measures of their archetype—a higher number of puzzles played per session and a higher rate of puzzles solved per minute—are associated with a higher Level 9 puzzle count.

3.2.2 Cluster 1

Cluster 1, comprising 25.8% of players, represents the archetype of highly engaged and low-volume players. This cluster completed the highest number of Level 9 puzzles and is the only cluster containing censored players (i.e., those who remained active beyond the analysed timeframe). The cluster is characterised by the lowest average number of puzzles per session but simultaneously the highest average excess score per session, pointing to a quality-driven approach. As a result of a more sustained approach, these players recorded the highest number of active days. A marked behavioural characteristic is the consistent tendency for players in this cluster to complete more puzzles than required for level progression, a behaviour particularly pronounced in Levels 1, 2, and 7. These characteristics define the "Completionists," distinguished by their longevity and a drive to fully exhaust game content.

Regarding longevity, the correlation between a player's excess score and their puzzle volume showed a significant positive correlation with the number of Level 9 puzzles completed (Figure 7). In contrast, the average and maximum session intensity are significantly and negatively correlated with the Level 9 puzzle count.

3.2.3 Cluster 2

Cluster 2 comprises 20.6% of players and generally represents a middle ground in terms of behavioural extremes, containing the most recently joined players. A key characteristic is the stark contrast in exceeding minimum requirements between early and late gameplay: while players showed a low tendency to exceed minimum requirements in the early game (Levels 1 & 2), this behaviour was notably stronger at Level 8, the final level with in-game rewards. These characteristics define the ”Explorers,” whose evolving pattern of over-achievement suggests their learning is a process of building competence before committing extra effort.

The behaviour of completing more puzzles than needed in Level 7 shows a statistically significant positive correlation with the final Level 9 puzzle count (Figure 7), indicating a clear link between late-game effort and long-term participation. While this overachievement behaviour is not statistically significant in the other levels (1, 2, and 8), the positive correlation coefficients for these early and late levels are the highest among all three clusters. Furthermore, the number of active days and the puzzle relative excess correlation (a measure of volume/quality relationship) both demonstrated a significant positive correlation with the Level 9 puzzle count, though the coefficient for the latter was smaller than that of Cluster 1, which aligns with expectations considering that Cluster 1 is defined by its drive for maximum quality and consistent effort, a characteristic that is inherently less pronounced in the exploratory and evolving approach of Cluster 2.

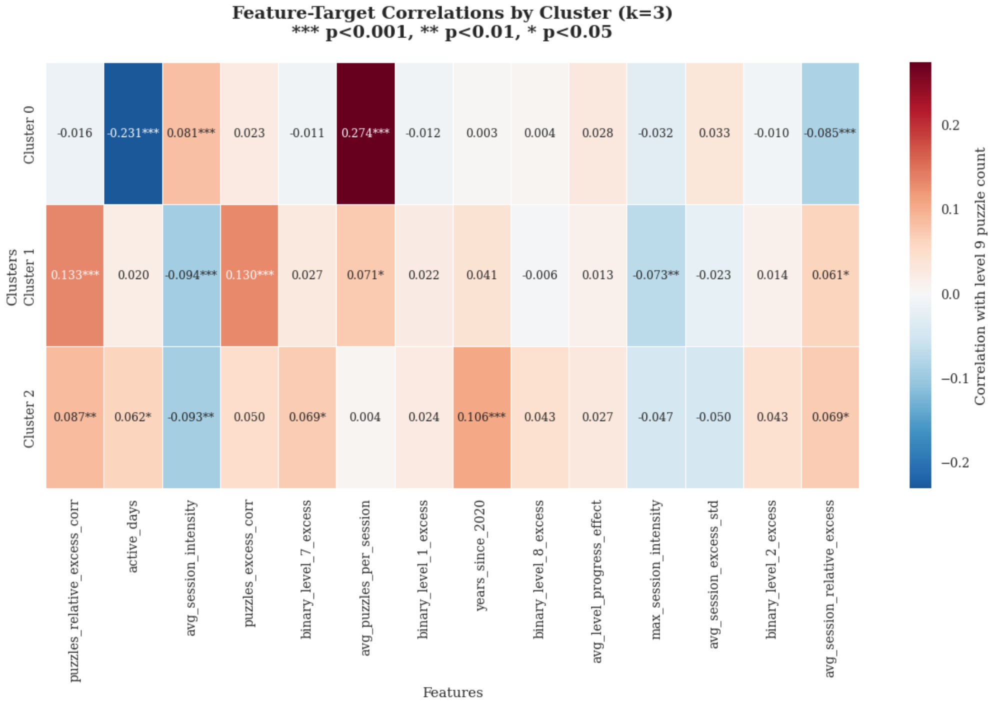

Figure 7: Correlation Coefficient of Features with Level 9 Puzzles, Stratified by Cluster ($k = 3$)

Through this study, the goal was to examine player characteristics associated with longevity, a crucial factor for the sustainability of CSGs where retaining contributors is key to long-term project success (Cooper et al., 2010). The analysis began with Cox Proportional Hazards models to identify features that serve as generalised predictors of player lifetime across the entire population. However, as several of these features were correlated, K-means clustering was employed to understand the relationships between these features and to segment players into distinct behavioural archetypes. A comparison of the two methods reveals an important nuance: while the Cox model identifies population-level risk factors, the cluster analysis demonstrates that the underlying drivers of longevity are not uniform. The general hazard ratios mask significant heterogeneity, as the same feature can exhibit a different relationship with longevity depending on the player's overall behavioural pattern.

4 Discussion

Based on the identified player segmentation, this discussion analyses the unique behavioural characteristics of the three player groups, and proposes reasons behind the counter-intuitive relationships observed. Finally, these heterogeneous findings are then translated into a set of targeted strategy proposals aimed at improving retention and engagement.

4.1 Performance Consistency & Quality Metrics

With the player archetypes established, a closer examination of how performance consistency and quality metrics drive engagement reveals that the overall Cox model results are highly segment-dependent.

The Cox model identified excess performance (relative excess–puzzle correlation) as a significant protective factor against dropout at the population level (HR = 0.67). However, the cluster analysis suggests that this average effect masks substantial variation across segments. For the intrinsically motivated segments—the Completionists (Cluster 1) and the Explorers (Cluster 2)—the protective effect anticipated by the Cox model is evidently observed. This is reflected by their strong, significant positive correlation with the final Level 9 puzzle count ($+0.133^{***}$ and $+0.087^{**}$ respectively). This relationship is consistent with their archetypes: for these players, scaling effort with volume reflects a mastery-oriented drive that intrinsically rewards their engagement. As the Completionists seek exceptional quality and the Explorers successfully build competence, the resulting over-achievement establishes a positive feedback loop that reinforces their intrinsic motivation and guards against attrition.

In sharp contrast, The Speedrunners (Cluster 0) show a negative correlation (-0.016) with Level 9 puzzle count, indicating that the protective behaviour for others is instead, detrimental to them. This inverse relationship was observed because, unlike the other two groups, Speedrunners are primarily motivated by rapid progression rather than mastery or quality. Consequently, engaging in ”excess performance” which requires focusing on quality or non-essential effort creates a goal conflict and acts as a source of friction, leading to accelerated saturation or burnout. From a game design standpoint, this means that imposing quality or volume requirements beyond the minimum necessary for progression is counter-productive for retaining high-intensity players. Though these players tend to contribute the least compared to the other two groups, they still represent a substantial proportion of the player base. Therefore, it is essential to understand the mechanisms for extending this group's longevity while simultaneously maintaining engagement across the other player segments to secure the long-term success of Borderland Science.

The risk associated with erratic performance (average session excess standard deviation, HR = 1.63) also requires segmented interpretation. For Clusters 1 and 2, the observed negative correlation with longevity aligns with expectations and is integral to their behavioural archetypes. For Completionists, who prioritise exceptional quality and methodical engagement, performance variability directly contradicts their mastery-oriented drive. Similarly, for Explorers, high variability signals a failure to build the consistent skill base necessary to maintain their sustained engagement and sense of increasing competence. Thus, for both of these clusters, consistent performance reflects successful focus and skill acquisition, while erratic performance is a reliable signal of impending disengagement.

On the other hand, Cluster 0 displays the inverse relationship, indicating that for Speedrunners, performance variability may paradoxically be associated with increased engagement. This inconsistency, likely stemming from intermittent high-score bursts on easier puzzles, provides random bursts of positive reinforcement (a variable reward schedule). Such stochastic rewards sustain their motivation and high-volume strategy without requiring the sustained, high-quality effort that leads to burnout. Therefore, for Speedrunners, performance inconsistency reflects the flexibility of their volume-driven approach rather than a decline in commitment, allowing continued participation even in the absence of stable performance. From a design perspective, maintaining a degree of variability in puzzle difficulty may be beneficial, as it preserves the unpredictable reward patterns that sustain their engagement while periodically encouraging deeper effort on more challenging puzzles.

4.2 Engagement Patterns and the Duality of Intensity

Beyond performance quality, the analysis of player engagement patterns reveals further nuances in the relationship between behaviour and longevity. While the Cox model identified the number of active days as a population-level protective factor, this effect varies across clusters. For the Completionists and Explorers (Clusters 1 & 2), the strong positive correlations with the Level 9 puzzle count align with expectations, supporting the notion that distributing engagement over time is a hallmark of sustainable, long-term participation.

In contrast, Cluster 0 (The Speedrunners) shows a statistically significant negative correlation (-0.273^{***}), indicating that for this group, more active days are associated with lower Level 9 completion. This paradox aligns with their high-intensity, task-oriented nature. For Speedrunners, an increase in number of active days is likely a consequence of struggling to complete puzzles efficiently, which may be a sign of frustration rather than commitment. Since their primary goal is to rapidly progress through the levels to obtain the in-game rewards efficiently, this resistance generates friction, leading to burnout and eventual disengagement shortly after obtaining all in-game rewards at the end of Level 8.

From a design standpoint, this finding reinforces the earlier principle: the game's difficulty curve must be carefully balanced. Designers should avoid making progression prohibitively difficult in an attempt to obtain higher quality outputs from the Mastery segments (Clusters 1 and 2). Failing to do so risks alienating the majority of the player base (Cluster 0), as the resulting struggle accelerates burnout and minimises their overall high-volume contribution.

Finally, the contradictory findings extend to the intensity of player engagement. Although the Cox model identified solving speed (average session intensity) as a population-level risk factor (HR = 1.07), the cluster-level results indicate that this relationship is not inherently detrimental. In fact, solving speed provides the clearest segmentation of the three player archetypes, though its effect must be interpreted with caution given its moderate direction agreement (59%). For Speedrunners (Cluster 0), faster puzzle solving aligns perfectly with their goal-oriented, efficiency-driven strategy ($+0.081^{***}$), demonstrating that high speed is consistent with, and necessary for, their short but productive sessions. In contrast, for Completionists and Explorers (Clusters 1 and 2) (-0.094^{***} and -0.093^{**} respectively), higher speed appears to undermine long-term engagement. This negative correlation suggests that rapid puzzle completion limits the cognitive depth and reflective processing required for these players to feel a sense of mastery or quality control. For these intrinsic groups, rushing disrupts the deliberate process that sustains their motivation over time.

From a strategic standpoint, the design principle is to ensure that while the Speedrunners' rapid play remains unhindered, the slower, more deliberate engagement characteristic of the Mastery segments (Clusters 1 and 2) is equally rewarded. This balance ensures that neither speed nor reflection is inadvertently penalised.

4.3 Early Commitment & Foundation Building

The early commitment indicators, specifically the completion of additional puzzles in Levels 1 (HR = 0.73) and 2 (HR = 0.52), were found to be strong protective factors against dropout, reflecting the importance of early investment in sustaining engagement. This relationship is likely primarily driven by Clusters 1 and 2, both of which display positive correlations with these indicators. Their behaviour exemplifies an intrinsic form of engagement established early in the gameplay, where additional effort beyond the minimum requirement signals the development of enduring commitment and skill. In contrast, Cluster 0 exhibits negative coefficients for the same measures, consistent with its extrinsically motivated, completion-oriented group, where early overperformance does not translate into long-term retention.

These findings suggest that players who demonstrate higher voluntary effort during the early stages whether through additional puzzles or excess are more likely to persist through the puzzles in Level 9. In the context of CSGs, such early engagement behaviours may serve as predictive signals for identifying contributors likely to generate reliable, high-quality data over the long term.

4.4 Intrinsic Motivation and Mastery Drive

Continued persistence in later stages depends on whether players derive satisfaction from the task itself rather than its external rewards. The HR for the average level progress effect (HR = 0.75) reinforces this finding, as this metric captures players' tendency to exceed baseline puzzle requirements even when approaching the reward threshold at each level, where motivation would typically wane due to the anticipation of completion and in-game reward.

The correlation coefficients for this metric are positive across all clusters, but the underlying motivational mechanisms differ significantly. For the Completionists (Cluster 1) and Explorers (Cluster 2), this sustained excess effort near reward thresholds may result from intrinsic motivation and mastery-oriented goals. They continue putting in additional effort because the value is derived from the quality of the task itself, not the external prize. In contrast, this sustained effort presents a counter-intuitive pattern for Speedrunners (Cluster 0). Their positive coefficient may reflect brief, intermittent episodes of flow — a state in which players experience effortless competence and deep engagement despite being primarily extrinsically motivated (Csikszentmihalyi, 1990). This suggests that when the task momentarily allows for effortless progress and uninterrupted performance, the rewarding experience of flow temporarily overrides their extrinsic goal of rapid completion.

Based on these findings, the core game design principle is the decoupling of task-based intrinsic incentive from the extrinsic completion threshold. To sustain player effort across all segments near reward limits, the game must ensure that the reinforcement schedule remains localised to the active process. For example, implementing immediate, mastery-oriented feedback loops that consistently signal competence may be required, thereby maintaining the incentive gradient of the task and preventing the anticipation of the final reward from diminishing the current action's value.

4.5 Potential Strategies

The preceding discussion delineates distinct behavioural trajectories and their associated engagement risks. Drawing on these empirical insights, the next section presents targeted design interventions intended to enhance sustained participation and mitigate hazard rates among the identified player groups.

The positive correlation between average progress level and long-term participation across all clusters points to a strategic design opportunity: moderating the micro-difficulty curve within each level to make high-performance outcomes more consistently achievable. The core intervention is to establish a single, optimised progression track that gradually decreases in difficulty toward the end of each level, while incorporating non-linear variations to sustain engagement. For the high-intensity Speedrunners (Cluster 0), this subtle modulation ensures that the overall flow remains achievable and rewarding, while intermittent fluctuations provide periodic bursts of challenge and positive reinforcement. At the same time, this mitigates the frustration associated with sustained difficulty, thereby reducing the risk of early burnout within this group. In parallel, these fluctuations create opportunities for the Mastery groups (Clusters 1 and 2) to engage in deeper problem-solving and accrue excess performance without facing prolonged difficulty barriers. This approach rewards their mastery-oriented drive through the psychological satisfaction of achieving a higher excess score rather than demanding an unnecessary barrier of difficulty. Therefore, it supports dual motivational pathways, rewarding both efficiency-oriented and mastery-driven play through a strategically moderated difficulty gradient, thereby improving long-term participation and enhancing the reliability of citizen science data annotation (Sweetser and Wyeth, 2005).

The second intervention addresses the risk that high session intensity poses to the intrinsically motivated players (Clusters 1 and 2). It introduces an optional reflective tool designed to counter the detrimental effects of rushing by redirecting focus from speed to solution quality. Upon puzzle completion, players can access a post-session review interface that presents key performance indicators related to problem-solving efficiency rather than completion time. Specifically, the interface displays (1) the player's personal efficiency score (number of clicks or yellow tiles used to solve the puzzle, weighted against the puzzle's score) and (2) the lowest corresponding value recorded for that specific puzzle, normalised against its inherent difficulty parameters. By providing quality-oriented feedback, the tool encourages players to redefine success in terms of strategic efficiency and precision. This approach transforms speed-oriented play into a reflective challenge centred on mastery and optimisation. In doing so, it supports sustained engagement among intrinsically motivated players by reinforcing a slower, more deliberate play style—one associated with greater activity over time and higher rates of Level 9 puzzle completion—while ensuring that progress is not hindered for players in the Speedrunner group.

The final intervention leverages the observation that excess performance (relative excess–puzzle correlation) is significantly positively correlated with the Level 9 puzzle count for the intrinsically motivated groups (Clusters 1 and 2). The central objective is to strengthen the behavioural consistency that underpins long-term engagement by rewarding sustained, consecutive high-quality performance. To implement this, a supplementary scoring framework explicitly designed to recognise and reinforce consistent excellence across sequential puzzles would be introduced. It is structured around two core components.

First, the feedback system is enhanced to compute and display an ”Excess Score” for each puzzle. Making this metric visible encourages players to prioritise performance quality over mere task completion, aligning feedback with their intrinsic mastery orientation. Second, a Quality Streak Multiplier is applied dynamically to the ”Excess Score” based on performance consistency. For every consecutive puzzle in which a player achieves a positive ”Excess Score”, the system increments a Quality Streak counter, applying a progressive multiplier (e.g., 1.1×, 1.2×, 1.3×) to the subsequent Excess Score. The streak resets if a player fails to exceed the minimum threshold, creating a clear behavioural incentive to sustain above-target performance, especially in the later stages of the game. The points accumulated through this mechanism are tracked independently from the standard high score metric, providing a distinct mastery-based progression pathway. The points accrued through this system constitute a parallel currency, termed ”Mastery Points”.

The primary application of this is to grant early access to advanced ”Level 9” puzzles, irrespective of a player's standard progression. This direct link between the behavioural driver (consistent excellence) and a high-value reward (elite content) provides a compelling, mastery-based incentive designed to solidify the engagement of high-potential players early in their participation. Nonetheless, the broader impact of this early-access reward on overall progression and long-term retention warrants further empirical evaluation.

5 Conclusion

Through this study, we have provided a deeper understanding of the player base that exists within the Borderlands Science universe. However, there are still several limitations that provide direction for future research. First, the analysis was conducted exclusively on players who successfully progressed to Level 9. Consequently, the identified clusters and their relationships with longevity describe only the highly engaged segment of the player population. The factors predicting dropout for the majority of players who never reached this stage remain unexamined. A critical future direction is to replicate this analysis on the entire player cohort to determine if the same archetypes emerge and to identify the distinguishing factors between those who churn early and those who persist.

Second, the analysis performed was inherently exploratory. The Cox model, while effective for identifying generalised risk factors, assumes linear relationships. The predictive power and potential non-linear interactions of these features could be further validated using other modeling techniques, such as tree-based models (e.g., random survival forests), which may capture more complex decision boundaries for player dropout.

Third, the analysis was constrained by the available data. The absence of player demographics limits the ability to contextualise behaviours within broader socio-cultural frameworks. Furthermore, the use of a standardised timestamp instead of local time imprecisely measures ”active days” and session timing, potentially misrepresenting true engagement rhythms. Finally, while the feature set was comprehensive, other nuanced behavioural metrics or interaction terms between metrics from different categories were not captured and could yield further insights.

Despite these limitations, this study establishes a foundational framework for examining player longevity in the Borderlands Science CSG. By identifying three distinct behavioural clusters, we examined the heterogeneous drivers of engagement and retention across player types. The results demonstrate that the relationship between in-game behaviour and longevity is non-uniform; behaviours that promote engagement in one segment may be neutral or detrimental in another. This nuanced, segment-specific understanding provides a critical basis for developing targeted strategies to enhance participant retention and ensure the long-term success of the citizen science project.

References

Anand, N., & Peterson, R. (2000). When market information constitutes fields: Sensemaking of markets in the commercial music industry. *Administrative Science Quarterly*, *45*(1), 171–180.

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., & Popović, Z. (2010). Predicting protein structures with a multiplayer online game. *Nature*, *466*(7307), 756–760. https://doi.org/10.1038/nature09304

Cooper, S., Sterling, A. L. R., Kleffner, R., Silversmith, W. M., & Siegel, J. B. (2018). Repurposing citizen science games as software tools for professional scientists. https://doi.org/10.1145/3235765.3235770

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.

Curtis, V. (2014). Online citizen science games: Opportunities for the biological sciences. *Applied and Translational Genomics*, *4*, 90–94. https://doi.org/10.1016/j.atg.2014.07.001

Gu, Z., Bapna, R., Chan, J., & Gupta, A. (2022). Measuring the impact of crowdsourcing features on mobile app user engagement and retention: A randomized field experiment. *Management Science*, *68*(2), 1297–1329. https://doi.org/10.1287/mnsc.2020.3943

Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R. L., Shpanskaya, K. S., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., & Ng, A. Y. (2019). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *CoRR*, *abs/1901.07031*. http://arxiv.org/abs/1901.07031

Karger, D. R., Oh, S., & Shah, D. (2014). Budget-optimal task allocation for reliable crowdsourcing systems [Originally published in Operations Research (2014)]. *Operations Research*.

Khatib, F., Cooper, S., Tyka, M. D., Lu, H., Xu, L., Mulligan, V. K., Kim, Y., Popovic, Z., Gingeras, T. R., Das, R., Baker, D., Contenders, F., & Scientists, F. (2011). Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences*, *108*(47), 18949–18953. https://doi.org/10.1073/pnas.1115898108

Kim, J. S., Greene, M. J., Zlateski, A., Lee, K., Richardson, M., Turaga, S. C., Purcaro, M., Balkam, M., Robinson, A., Behabadi, B. F., et al. (2014). Space-time wiring specificity supports direction selectivity in the retina. *Nature*, *509*(7500), 331–336. https://doi.org/10.1038/nature13240

Lee, J., Kladwang, W., Lee, M., Cantu, D., Azizyan, M., & Das, R. (2014). Rna design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences*, *111*(1), 212–217. https://doi.org/10.1073/pnas.1313039111

Northcutt, C. G., Jiang, L., & Chuang, I. L. (2022). *Confident learning: Estimating uncertainty in dataset labels*. https://arxiv.org/abs/1911.00068

Prestopnik, N., & Crowston, K. (2013). Motivation and data quality in a citizen science game: A design science evaluation. *Proceedings of the 46th Hawaii International Conference on System Sciences*. https://doi.org/10.1109/HICSS.2013.413

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sarrazin-Gendron, R., Ghasemloo Gheidari, P., Butyaev, A., Keding, T., Cai, E., Zheng, J., Mutalova, R., Mounthanyvong, J., Zhu, Y., Nazarova, E., Drogaris, C., Erhart, K., Bélanger, D., Bouffard, M., Davidson, J., Falaise, M., Fiset, V., Hebert, S., Hewitt, D., ... players, B. S. (2025). Improving microbial phylogeny with citizen science within a mass-market video game. *Nature Biotechnology*, *43*(1), 76–84. https://doi.org/10.1038/s41587-024-02175-6

Sweetser, P., & Wyeth, P. (2005). Gameflow: A model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)*, *3*(3), 3–3.

van der Weiden, A., Benjamins, J., Gillebaart, M., Ybema, J. F., & de Ridder, D. (2020). How to form good habits? a longitudinal field study on the role of self-control in habit formation. *Frontiers in Psychology*, *11*, 560. https://doi.org/10.3389/fpsyg.2020.00560

Vaughan, J. W. (2018). Making better use of the crowd: How crowdsourcing can advance machine learning research. *Journal of Machine Learning Research*, *18*, 1–46.

Weyand, T., Araújo, A., Cao, B., & Sim, J. (2020). Google landmarks dataset v2 - A large-scale benchmark for instance-level recognition and retrieval. *CoRR*, *abs/2004.01804*. https://arxiv.org/abs/2004.01804

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendix

Table 4: Engagement Patterns & Behavioural Rhythm Metrics

Metric	**Description**
active_days	Total number of unique days with puzzle activity.
engage_duration	Total involvement period of each player measured in days.
activity_ratio	Ratio of number of days that a player was active to their total engagement duration.
avg_time_between_sessions	Average time between consecutive sessions in minutes.
avg_time_between_puzzles	Average time between individual puzzle solves in minutes.
prop_long_sessions	Proportion of long sessions relative to total number of sessions recorded.
avg_session_duration	Average duration of each session in minutes.
avg_puzzles_per_session	Average number of puzzles completed per session.
avg_session_intensity	Average number of puzzles solved per minute for each session.
max_session_intensity	Maximum number of puzzles solved per minute across all sessions.
longest_session_duration	Length of longest recorded session in minutes
longest_session_length	Number of puzzles completed during the longest recorded session
intensity_consistency	Standard deviation of puzzles solved per minute across all recorded sessions, measuring intensity stability.
avg_level_progress_effect	Mean change in excess ratio between the last and first third of puzzles completed per level

Table 5: Early Commitment & Foundation Building

Metric	Description
learning_click_slope	Slope of the regression line describing the change in time per click as players approach the target score.
r2_learning_click	R^2 from the regression model capturing the change in time per click.
learning_score_slope	Slope of the regression line describing the change in time per score as players approach the target score.
r2_learning_score	R^2 from the regression model capturing the change in time per score.
prop_level_1_excess	Ratio of additional puzzles completed to the required number of puzzles in level 1.
prop_level_2_excess	Ratio of additional puzzles completed to the required number of puzzles in level 2.
prop_level_3_excess	Ratio of additional puzzles completed to the required number of puzzles in level 3.
prop_level_4_excess	Ratio of additional puzzles completed to the required number of puzzles in level 4.
binary_level_1_excess	Binary indicator of whether a player completed more puzzles than required in level 1.
binary_level_2_excess	Binary indicator of whether a player completed more puzzles than required in level 2.
binary_level_3_excess	Binary indicator of whether a player completed more puzzles than required in level 3.
binary_level_4_excess	Binary indicator of whether a player completed more puzzles than required in level 4.

Learning dynamics were modelled using the first n puzzles of each player's activity. An optimal window size of 45 puzzles was selected empirically by evaluating the coefficient of determination (R^2) and its incremental change (R^2) across varying puzzle counts. The cut-off point was identified where additional puzzles yielded minimal improvement in model fit, after accounting for the trade-off in R^2 reduction between the two linear regression models used in the learning dynamics assessment—time per click and time per score. Refer to Figures 8 and 9.

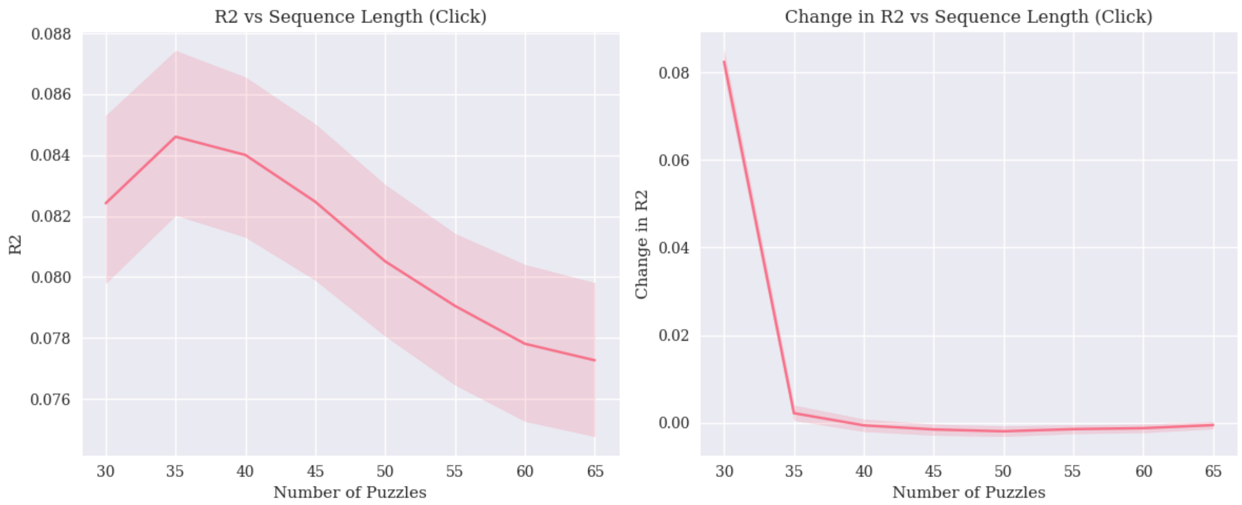

Figure 8: Relationship between puzzle count and (A) the model's coefficient of determination (R^2) and (B) the incremental change in R^2 (R^2) for the linear regression of time per click against puzzle count.

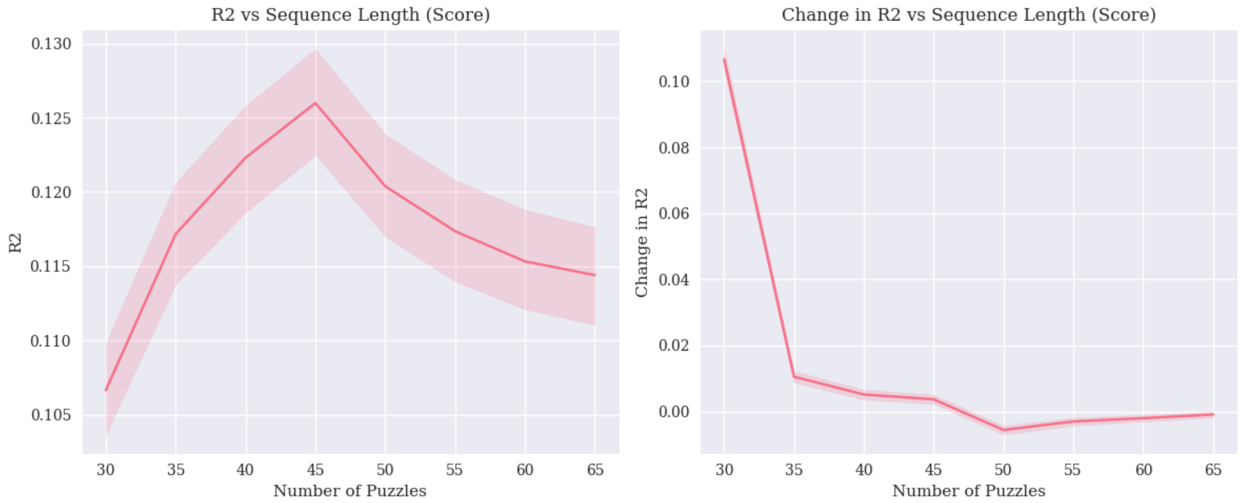

Figure 9: Relationship between puzzle count and (A) the model's coefficient of determination (R^2) and (B) the incremental change in R^2 (R^2) for the linear regression of time per score against puzzle count.

Table 6: Intrinsic Motivation & Mastery

Metric	**Description**
prop_level_5_excess	Ratio of additional puzzles completed to the required number of puzzles in level 5.
prop_level_6_excess	Ratio of additional puzzles completed to the required number of puzzles in level 6.
prop_level_7_excess	Ratio of additional puzzles completed to the required number of puzzles in level 7.
prop_level_8_excess	Ratio of additional puzzles completed to the required number of puzzles in level 8.
binary_level_5_excess	Binary indicator of whether a player completed more puzzles than required in level 5.
binary_level_6_excess	Binary indicator of whether a player completed more puzzles than required in level 6.
binary_level_7_excess	Binary indicator of whether a player completed more puzzles than required in level 7.
binary_level_8_excess	Binary indicator of whether a player completed more puzzles than required in level 8.
exploration_intensity	Composite metric capturing players' post-target score engagement. Calculated as the product of (i) the proportion of puzzles with clicks after achieving the target score, and (ii) the average proportion of such clicks relative to total clicks per puzzle.
avg_time_per_click_after_par	Overall mean of the average time per click recorded after players achieved the target score for each puzzle.
avg_level_progress_effect	The average change in mean excess ratio between the first 25% of puzzles and the last 25% of puzzles in a given level. This measures any change in performance consistency or engagement decay within a given level, considering that there is an incentive (game reward) at the end of each level.
total_excess_puzzles	Total number of additional puzzles completed across all levels 1-8.
overperf_rate	Proportion of additional puzzles completed relative to total required number for levels 1-8.
freq_any_clicks_after_par	Proportion of puzzles in which at least one click was made after reaching the target score, indicating how frequently players engaged in post-goal exploration.
avg_prop_clicks_after_par	Average proportion of clicks made after reaching the target score relative to total clicks per puzzle, representing the extent of exploratory behaviour beyond goal completion.

Table 7: Performance Consistency & Quality

Metric	**Description**
consistency_excess	Correlation between excess ratio of puzzle n and excess ratio of n-1. Analyses for any relationship between past and present excess behaviour.
relative_consistency_excess	Correlation between relative excess of puzzle n and relative excess of n-1. Analyses for any relationship between past and present excess behaviour.
excess_variation	The variability in a player's excess ratio across all puzzles.
relative_excess_variation	The variability in a player's relative excess across all puzzles.
avg_session_excess	Mean excess ratio across all puzzles within each session.
avg_session_relative_excess	Mean relative excess across all puzzles within each session.
avg_session_excess_std	The variability in a player's average excess ratio between puzzle sessions.
avg_session_relative_excess_std	The variability in a player's average relative excess between puzzle sessions.
std_prop_clicks_after_par	The variability in proportion of clicks recorded after attaining target score across puzzles.
std_time_per_click_after_par	The variability in time recorded per click after attaining target score across puzzles.
longest_excess_streak	The greatest number of consecutive puzzles with excess > 0 for each player.
avg_excess_streak	The average streak length of each player.
total_excess_streaks	Total number of streaks obtained across all puzzles.

Continued on next page

Table 7 – continued from previous page

Metric	**Description**
effort_excess_corr	Correlation between effort ratio and excess ratio. This looks at the relationship between the effort required to complete a level and additional effort given for puzzle completion.
relative_effort_excess_corr	Correlation between effort ratio and relative excess.
puzzles_excess_corr	Correlation between the number of puzzles played and excess ratio. This measures the relationship between the total volume of puzzles played and the degree to which a player exceeds the minimum puzzle requirements for completion.
puzzles_relative_excess_corr	Correlation between the number of puzzles played and relative excess. This measures the relationship between the total volume of puzzles played and the degree to which a player exceeds the minimum puzzle requirements for completion.
longest_session_avg_time_per_click	Mean time per click (after reaching target score) during the player's longest session.
longest_session_avg_time_per_score	Mean time per score (after reaching target score) during the player's longest session.
longest_session_std_time_per_click	Standard deviation of time per click (after reaching target score) during the player's longest session.
longest_session_std_time_per_score	Standard deviation of time per score (after reaching target score) during the player's longest session.
longest_session_raw_decline	Proportional change in average excess ratio from the first third to the last third of the longest session. This metric captures the decline in performance over the course of a session.

Continued on next page

Table 7 – continued from previous page

Metric	Description
longest_session_relative_decline	Proportional change in average relative ratio from the first third to the last third of the longest session. This metric captures the decline in performance over the course of a session.
longest_session_time_score_corr	Correlation between time per score after reaching the target score and puzzle position within the session.
longest_session_time_click_corr	Correlation between time per click after reaching the target score and puzzle position within the session.

** Effort ratio: Ratio of score required to complete a puzzle relative to the target score*

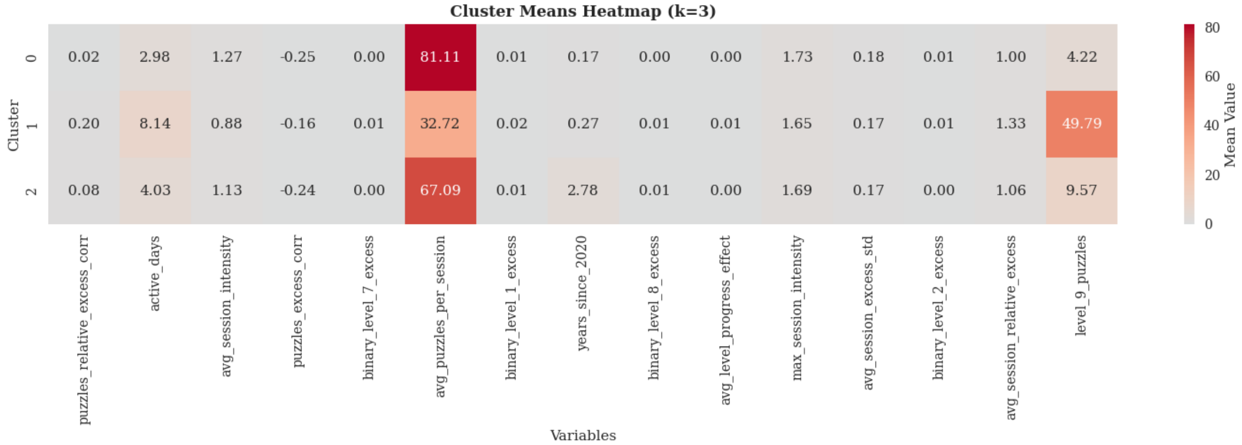

Figure 10: Cluster means heatmap