This study used the platform jumping game Zool, gathering user information for a month following the game's initial release (from May 16 to June 6).The game was released on the playstation platform, and the data include 520 players and 6511 rows of game records. The game automatic records the player's performance, time, and other data every times when a player exits or completes a level. Game data is also recorded when a player runs out of lives to pass a level and fails.

data Cleaning Preparation

Player id is encrypted in our situation.At the start of the game, we check to see if the player has an encrypted player ID assigned to them, if not we create one for them. This Encrypted ID is MD5 encoded and salted with the time of creation.  It allows us to tell each player apart but it is impossible to reverse or even re-create on the original console and so would be considered anonymous in the eyes of GDPR.

In preprocessing, firstly the encrypted player id from the original data is converted into a simple numerical serial number so that the same player can be identified and labelled using the same number.

Second, change the name and data format. Make the original data easier to evaluate by converting it from being categorised by various items to being categorised by player game records. Time data was formatted differently, and various items were given new names.

In the data cleaning section, the data of one player and 122 lines of game records were deleted in this study. By looking at the data, it was found that one player (player\_id: 358) repeated the same level 109 times (the first level of the first world) and did not have any game records for other levels, while the player other than him who repeated the same level the most played it 19 times, so the data of this player was deleted and cleaned. In addition, 10 pieces of data with mostly missing content and 1 piece of data with significantly more than normal play time (>2000s) were removed.

The cleaned data contained 519 players and 6389 lines of gameplay.

Variables included in data

1.Warning

In this game, the warning setting indicates whether the player has enabled a reminder function. This function will cause the joystick to vibrate when the player has only one life remaining in a particular level. Players who have deactivated this feature will not be impacted. The player is free to decide whether or not to activate the feature, and they can make adjustments to it at any point while they are playing the game.

2.Churn Rate

For the purpose of this study, player churn is defined as a player not continuing to play the game. For a defined observation time, the player's last game data will be defined as the point in time when the player churns which means that the player no longer continues to play the game after attempting the level. We also take into account player churn that occurs naturally and in general while analysing this data. Churn is when a player completes all levels and abandons the game, while general churn is when a player abandons the game in the midst of a play session. 48 participants (9.23%) completed the entire jam, while the remaining players experienced churn during the middle of the game. Since the time node of our model was defined as the number of levels in that study, players who churned naturally were a subset of general churn players, which means that they churned at the end of the final level at the end of the level's time, and were not labelled in any special manner. The last time of game data for the previous level of the same participant is defined as general churn and marked with a 1 in the data. The remaining data, which is marked with a zero, indicates that the participant will continue to play the game.

3.Degree

This game has 7 worlds, each with 4 levels, re-encoded as 28 different degrees. This was used as a time variable in the model in that study.

4.Inputs

Total count of controller inputs the player produced during that level.

5.Deaths

The number of times the player has lost all their health and died within the level.

1. dup\_flag

Indicates whether the player has replayed the level, with 1 for yes and 0 for no.

7.Dups\_count

The number of times the same player repeats the same level is marked in chronological order. For example, the first time player1 plays degree 1 is marked as 1; the second time he play degree 1 is marked as 2, and so on.

8.Times

Total time to complete the level (or game over) in seconds.

9.Total\_times

The total time spent by an individual player on all games and all level.

**10. Collecteds**: The number of giant collectibles collected within each level. 0-3

**11. Collectible Badge**: Awarded for collecting over a target percentage of collectibles. Boolean; 0 for false or 1 for true.

**12. Time Badge**: Awarded for beating a target time. Boolean; 0 for false or 1 for true.

**13. Game Over**: Player dies without completing the level; 0 for completing the level or 1 for death.

3 Hypothesis

This study uses **Cox proportional hazards model** to explore the impact of various factors on player churn rate. In this research, 6 covariates are considered, with the hypothesis that Collectible Badge, Time Badge, gameover, warning, dup\_flag, and total\_time will all have an effect on the Churn Rate.