WeR Project Summary

Introduction & Background

Our project analyzes the data collected from Elm City Stories, an educational video game that aims to increase middle to high school students' perception of risk and therefore prevent their risky behaviors. The experiment records the logs of 166 players (11-14 years-old) with their actions and time spent in each event. Our goal is to investigate the engagement level of participants so that they can learn more and complete more events. Our research question is to understand the probability of completion the game from a survival analysis perspective.

Data Cleaning & Statistical Analysis

We kept the proportion of games completed at the largest time elapsed grouped by player and event and removed missing values. We created a binary response variable "event_new" based on the proportion of game completed compared to the mean completion and classify them into 2 groups. We select the most relevant variables using principal component analysis and propose two survival models to understand the behavior of game completion.

Principal Component Analysis

We performed principal component analysis to reduce the dimensionality. By computing the principal components for numerical variables, we want to explain most of the variability using a relatively small number of factors that are uncorrelated with each other. The ten most important components are: elapsed time (in seconds), player's current "know", "priority", "people", "refusal", and "me" skill level, advancement level, previous and current points at this level and the amount of time watched for animation.

Model 1 – Survival Random Forest

Survival Random Forest is an ensemble tree method for analysis of survival data. We introduced all variables from PCA and player's age to estimate the probability of completing the game above the threshold. Event though the mean predictive probability is 50% as expected, individual cases vary much. We also provide a list of the most important variables: player's current "know", "priority", "people", "refusal" skill level, age and the amount of time watched for animation.

Model 2 – Cox Proportional-hazards Model

We fitted a Cox Proportional-hazards Model to provide estimates of the effects of the covariates on the probability of the event completion. The model measures the association between the playing time of participants and whether the event is complete. We obtained Hazard Ratio (HR), the proportion of the completion rate among groups. Our results show that the previous points at this level of the game, age, player's current skill level have significant positive effects on event completions.

Conclusion

We found Player's current "know" , "priority", "people", "refusal", and "me" skill level affect the event completion rates most. Therefore, the game designers could think about how they can train these skills of participants to help them better utilize this game. The Survival Random Forest Model predicts the probability of completion to be only around 50%, but the probability varies much depending on the player. Therefore, there is still room to adjust the focuses and the overall design of the game. Additionally, the Cox Proportional-hazard model results suggest that the game designer should consider lowering the game difficulty, to help the participants get a sense of achievement, which is shown to increase the event completion rate.

Limitation

The lack of health outcome data from players makes us unable to build a prediction mode, which can be very helpful in terms of understanding the effects of the game. Also, we only analyzed the log dataset instead of incorporating more information, which might limit us from building a more accurate model. In the future, we can include a more holistic range of categorical variables. In addition, we could utilize the additional datasets to investigate the mean score of students in the treatment group to analyze the effects of the game.