

Data Fest

Team Tie

Andy Wulu, Bridge Hu, Cara Yang, Andy Liu, Jacqueline Ma

4/3/2022

Investigation into Elm City Stories' MiniGame Design

In this project, we are given the game-playing data on 166 students and their subsequent test results to see whether we can interpret meaningful insights on drug usage from the way they play the game. More specifically, our research focused on the completion time of each minigame and how that reflects the game design then gave advice on future optimization.

After an extensive amount of time playing the game, we discovered that the main game's tasks were mainly composed of using the scopes to find objects, which does not involve any decision making or learning. The MiniGames, however, asked the players to use their own judgments to make decisions. Therefore we believe that data from the minigames are more useful. More specifically, we want to use the duration of playtime in each Minigame to cluster and discover.

Cleaning the data, we used the start point and endpoint indicated by "event_id" and "event_time" to find the duration of playtime in each MiniGame. Then, we deleted those that did not finish the game and fast-forwarded prematurely, which is a large chunk of the original data.

From a preliminary graph of the game time played, we found that zero is the most common value across all minigames. We suggest the research team improve the gameplay experience to increase the incentive to finish games. To that end, we analyzed the data of game time played and hope to generate insight on how the game can be improved. We focused on those players who finished most of the games and analyzed how they did in the games.

Analyzing the data, we want to discover whether the games are designed to reflect their associated categories as meant instead of correlating to one another. We used Principal Components Analysis to look into this. The loadings of the PCA show differences across categories in Refuse, Knowledge, Priority, and People which justifies the game's separation of these games.

Afterward, we investigated if there is any room to improve the game's user experience, which could help solve the fast-forward problem. We used the K-means method to segment players into different groups. The analysis split them into three groups: fast, slow, and those who are only slow in priority games. The individual differences in game-playing speed might be caused by video game skills and knowledge in the category, but a change in pace could indicate that these games (Priority) need to improve on their setup.

There are some limitations to our findings. Perhaps the most important one is using only time played as an independent variable. In using time played to perform PCA, we are assuming the underlying nature of each category can be represented by time. We also tried to perform a regression of time on students' s5 score (risk on the drug), to see which game is more efficient. However, the sample size is too small, causing our result to overfit. Moving forward, we also suggest the research team collect more samples for further analysis.