Introduction & Motivation

We looked at the Play2Prevent game data hoping to help the researchers at Yale University to come up with predictions about the potential real-life behaviors of the students when they are exposed to the use of these harmful stimulants. We hope that with these predictions, schools and the authorities may accurately identify the students at risk and design the appropriate lessons to help educate the students.

Summary of analysis

1. **Variables of interest**

After having played some trial rounds of the game, we hypothesized that there might be a correlation between the students' performance in the Play2Prevent game and their score in the clinical study on drug resistance. Particularly, we believed that the Knowledge Minigame and Refuse Power Minigame would impact the students' performance in the study the most as they got to learn about the harm of the stimulants in these games.

We believe that the time a player spent on the knowledge and refusal minigame categories in each session would be a useful indicator of the level of their stimulant resistance. Thus, we introduced a new variable *"time lapse"* that computes the amount of time the player spent on each task. We want to look at the average time spent by these players. Additionally, we are also interested in comparing the average S5-scores of the two groups that spent the most and least amount of time in the knowledge and refusal stack.

2. **Data summary**

We simplified the logs dataset with a focus on "time-lapse". The data can be found in the file named "logs_delta_sec.csv". Individual datasets of players 6607011, 6486029, and 6427031 were also simplified to only focus on Knowledge and Refuse Power Minigame. Each of the three datasets for the players consists of approximately 2600 data points as players went through 9 sessions of the game. We also found out that there were only 64 distinct players' scores shown in the dataset as opposed to the 166 players mentioned in the description.

3. **Data analysis**

When looking at the data, we focused on the different game sessions, the types of events, the event category specifically the knowledge minigame and refusal power minigame, the time taken for each game session, and the S5 scores of the players. We tried to answer the following two research questions.

RQ1: Does spending more time on the Knowledge/Refusal minigame result in a lower S5-score(Higher resistance)?

RQ2: Does the average S5-score of the two groups (One who spent relatively more time on the Knowledge/Refuse stack than the other) differ?

Results and Conclusion

For RQ1, we hypothesized that more time spent on the knowledge and refusal stack would result in lower S5-scores. But the results showed otherwise. We found that on average, the participants who spent more time on the knowledge stack also spent more time on the refusal stack and their S5-scores were still higher. For RQ2, we hypothesized that there would be a significant difference between the two groups. But the p-value of 0.8789 for the knowledge stack and the p-value of 0.1839 for the refusal stacks were non-significant. These results signified that it would be hard to predict the real-life behaviours of the students by just looking at their game-play.

Limitation

We found that there could be potential biases as there was no information provided about the users of the game. This information would be helpful for our analysis so that we can determine if the amount of time spent on the knowledge and refusal minigames was due to the users' age, education, and racial background. This information would also be useful to make conclusions about the S5 scores.