RJ Krimtez, Yoonsung Park

Approach:

After looking at the provided dataset, the team recognized that the only variable we could reasonably use to evaluate the effectiveness of the data was the **Self-efficacy for Drug Use Resistance Survey**. Looking at the data it was quickly apparent that many students scored 4's on all testing occasions, implying that they were highly susceptible to future drug use and survey results were not impacted by participation in the Play2Prevent video game. Accordingly, the team looked to classify students into two groups, those whose survey scores improved and those whose did not. The team looked at the average time it took students to complete challenges across sessions as a proxy for a student's genuine participation in the game. Using this 'focus indicator' as the independent variable, the team looked to see if there was any correlation between a student's active participation in the game and an improvement in the Drug Resistance Survey score.

Data Cleaning:

The largest obstacle with respect to data cleaning was relabeling the Session attribute in the student activity log data. It was crucial for us to be able to compare the average time a student took to complete challenges as they progressed through the sessions, however there were disparities in spelling, capitalization, and numeric representation of the session number.

Feature Generation:

Our primary feature of interest was the average time it took an individual student to complete a 'challenge' in the game. We used this feature as a proxy for active participation in the game and expected that if a student was focusing on the material presented and the instructions of the game, their average challenge completion time would 1) generally be lower than an student who was not focusing on the game and 2) would decrease as they progressed through the sessions.

To build this feature, the team first calculated the time it took to complete each challenge, found the average time for completion at each session, and compressed the multiple sessions each student completed to find the average completion time for each student. The students were then assigned to one of four groups with group Q1 reflecting the 25% of students who completed challenges the fastest and Q4 representing the 25% students who took the most time to complete challenges.

The response feature we were interested in analyzing was stored as a boolean that reflected if a student's Drug Resistance Survey scores improved or not (True meaning improvement). To generate this, the team calculated the change in scores between each period to calculate the net change from the beginning of the experiment to the end. Both data sets were then merged on the player_id variable.

Statistical Tests:

In order to determine if the time spent per challenge was correlated with improvement in the Drug Resistance Survey scores, the team conducted a Chi-Squared ratio test over 55 samples. The Null hypothesis of the test was that the two variables are independent. The results are below...

$$X\text{-squared} = 8.5079, \text{df} = 3, \text{p-value} = 0.0366$$

According to the above results, we can conclude at a 95% confidence level that the average time a student spends completing a challenge and the probability of their results from the drug resistance scores improving are not independent. In other words, there is a statistically significant observed correlation between a student focusing on the materials presented in the Play2Prevent video game and their assumed ability to resist illegal substances.

Visualization:

The average time to complete a task in the videogame for each session the student participated in is displayed as a line graph (one line for each student), color coordinated by whether that student's Drug Resistance Survey scores improved over the trial.

Conclusion:

Looking at the results from above the team is able to conclude that when students actively participate in the game play provided through the Play2Prevent platform, their scores on the Drug Resistance Survey improve. This implies that the platform is able to prevent at risk adolescents from participating in self-destructive behavior.