

Data Fest report 2022
Team: Nihongokurasu

Data insight:

Although there are about 131 columns and over 2,000,000 observations in this data, there are an extremely large number of null values which poses enormous obstacles in analyzing the data. To a short data insight, there were 212 players registered for this game, 166 players succeeded in playing forwards, and 82 participants filled in the questionnaire.

Assumption:

Our hypothesis is that after playing the game, the player will receive significant educational effects which will help them build a healthy life. This can be reflected by the S5 score, the higher the S5 score, the more likely they are able to resist drug use by themselves. We assume that the longer time they played the game, which can be implied by how many sessions they participated in, the larger the positive impact would be on themselves. We are also interested in whether their minority background, age, and gender would influence their outcome S5 score.

Analysis:

Because we assumed that the educational function of the game can affect students' ability to refuse. Firstly, we explore how people keep moving on in the process. The distribution of sessions shows that most people stopped at the session around 10, while only one person reach session 16. And then we want to explore if the capability to keep moving on is different between gender and ethics. And we found that there is no obvious difference between genders and ethics, but the most persistent people are all female African Americans. And then, we want to explore the assumption that experience and knowledge gained from the game can help people to improve their steadfast ability to refuse to go astray. From the histogram we made, we can see that the point differences between the last time and the first time that the participants fill the questionnaire are symmetric by the y-axis, but there is one outlier who got improvement a lot. And we explore the effect of playing the game at a higher level. From the graph we made, we found that the score has a rising intention before week 10 and decreasing trend after week 10. And in total there is a rising trend, which means there is a slight improvement after playing the game. But for people who is more persistent in the game, initially have higher grade compared to the whole questionnaire participants, and their improvement and negative improvement is not that obvious. And then we want to explore if the difference of improvement is different in gender or ethics, finding that male African American is more affected by the game both positively and negatively. And Caucasians are negatively improved by the game of their ability to refuse the things that will lead them astray. After that, we are thinking about maybe the negative influence is due to students initially having a 4-point grade. After our exploration, we find that it is true that students with higher scores initially will be more easily affected by both games. And the male is more easily to be affected. But there is one thing that we think should be mentioned is that there are some students who took the game 2-3 times with different identities, so there is bias in our analysis.

Conclusion:

After exploring and analyzing the data, we found that our hypothesis is untenable. The majority of the participants maintained similar scores before and after playing the game and there was a slight difference between distinct gender and ethnicity. This shows that this game may not have those important educational effects and might have a large improvement space.