

Team15 (Adhoc4) ASA DataFest 2022 Report:

Members: Hector Beltran, Sean Peck-Collier, Alexandru Folta, Hari Krishnan, Michelle Solorio

Our work as a team was to take the log data provided from this game to try to determine whether or not there was any relationship between selected data collected by the game and the decrease in S5 survey scores. We created a new dataset, which was based on the S5_mean dataset. We took all players who took the S5 survey (which was taken in weeks intervals, 0, 3, 6, etc.), and aggregated the values between the weeks they took the surveys, which included time played during the time period between each survey as well as their in game minigame skill levels. The new data presents a look at what each player's time played and skill levels were each time they took the survey, accompanied by the score they received on the survey.

One odd piece about the log data that we noticed was that there were dates that were set in the distant past, ex. 12-31-1969 and 01-02-1970, in players with ids 6427032, 6427034, 6427035, and 6427037. We removed these players from our dataset. From the 166 players in the log dataset, 64 players took the S5 survey, and our data had 61 players after cleaning. Once the new dataset was created, linear regression models were used to test whether play time, skill level in knowledge, priority, people, refusal, or skill level of oneself in the game had an effect on S5 scores.

The team found that there is practically no correlation between these features, which illustrate time spent and experience in the game, and the S5 scores. These are features you would expect to impact the S5 scores. We created linear regression models that tested a full model as well as an AIC eliminated model. The former performed awful with a p-value of 0.31, and the latter performed better but still very poor, but with a decent p-value of 0.094. The R^2 s of these models were 0.0047 and 0.012 respectively, which shows that even though the AIC model improved upon the p-value, the fit is still very poor. The variables left included in the AIC model were skill_level_priority and skill_level_people. This leads to the conclusion that there is no linear relationship between these features and the S5 scores, as the null hypothesis that the coefficients equal 0 (no impact of including features on the model) is has failed to been rejected, for both linear models we have tried (with $\alpha=0.05$). To further prove that there is no correlatory relationship between these features and the target S5 scores, we also ran some non-linear models, as to cover all bases of possible correlation. We ran both Linear SVM and a Neural Network, using proper training procedures that reflect nonlinearity. Both methods, after training, gave accuracy scores of 0. This shows that there is neither a linear or non-linear relationship between these features and the S5 scores.

After this conclusion was made, different aspects and patterns of the game were explored, and using the AOV function in R the team discovered that someone's hours played and days elapsed had a significant effect on someone's skill level in the knowledge game with both p-values being <0.0001 . Another trend that was discovered was that between different schools, they all displayed similar trends to seconds in play time.

There are some limitations to the broad conclusions we can make from this data: this data as well; in the S5 mean scores dataset, 99% of scores were over 3.5 which shows that 99% of the children surveyed had a low efficacy for drug resistance to begin with this program. In the study *Design and Evaluation of a Computer Game to Promote a Healthy Diet for Young Adults*, Wei Peng shows that video games for health have proven to help youth and adolescent groups in their efficacy to make better decisions in nutrition choices. We believe that a video game will be able to still help these at-risk groups if it were built with more role-playing and uncertainty in the game. Perhaps adding more elements of uncertainty in *PlayForward: Elm City Stories* could help youth better comprehend these risky behaviors and what they will do.