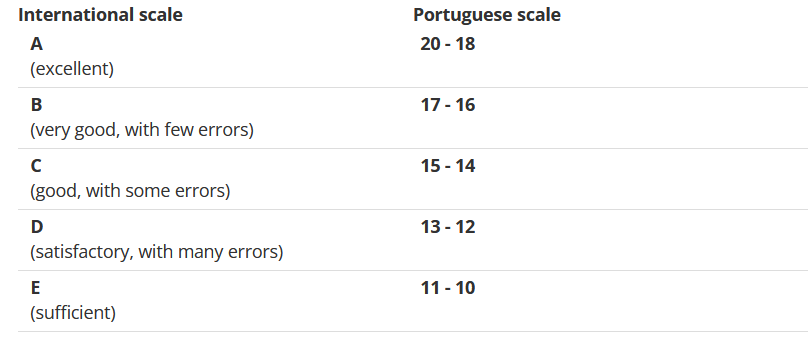
**Project Proposal: Eric Bixby, Chris Rooney, Anthony Shishkin**

1. The dataset we are using for the final project is from students (high school age) on their alcohol consumption and grades. The dataset has over 30 variables, many of which we will not keep for the final project, but the plan is to use such variables like family size or study time to predict a student’s final grade. As mentioned, there is also data on alcohol consumption, which seems like it could be a predictor variable or a response variable.
2. We are planning on using linear regression and multiple linear regression to do some prediction work. After doing some work in linear and multiple linear regression, looking for r-squared values, we are going to do some decision tree stuff in jmp. Some of the results in linear and multiple linear regression should help us find the best variables for decision trees and hypothesis testing. We will also use memory-based reasoning to find cases that are like each other. There seems like there will be many cases that are quite similar. Lastly, we will do some work in clustering to best find outliers in the dataset.
3. sex - student's sex (binary: 0 - female or 1 - male)
4. age - student's age (numeric: from 15 to 22)
5. address - student's home address type (binary: 1 - urban or 0 - rural)
6. Pstatus - parent's cohabitation status (binary: 0 - living together or 1 - apart)
7. Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
8. Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
9. guardian - student's guardian (nominal: 0 - mother, 1 – father, or 2 - other)
10. traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
11. studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
12. failures - number of past class failures (numeric: n if 1<=n<3, else 4)
13. activities - extra-curricular activities (binary: 1 - yes or 0 - no)
14. internet - Internet access at home (binary: 1 - yes or 0 - no)
15. romantic - with a romantic relationship (binary: 1 - yes or 0 - no)
16. famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
17. freetime - free time after school (numeric: from 1 - very low to 5 - very high)
18. goout - going out with friends (numeric: from 1 - very low to 5 - very high)
19. Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
20. Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
21. health - current health status (numeric: from 1 - very bad to 5 - very good)
22. absences - number of school absences (numeric: from 0 to 93)
23. G3 - final grade (numeric: from 0 to 20, output target)

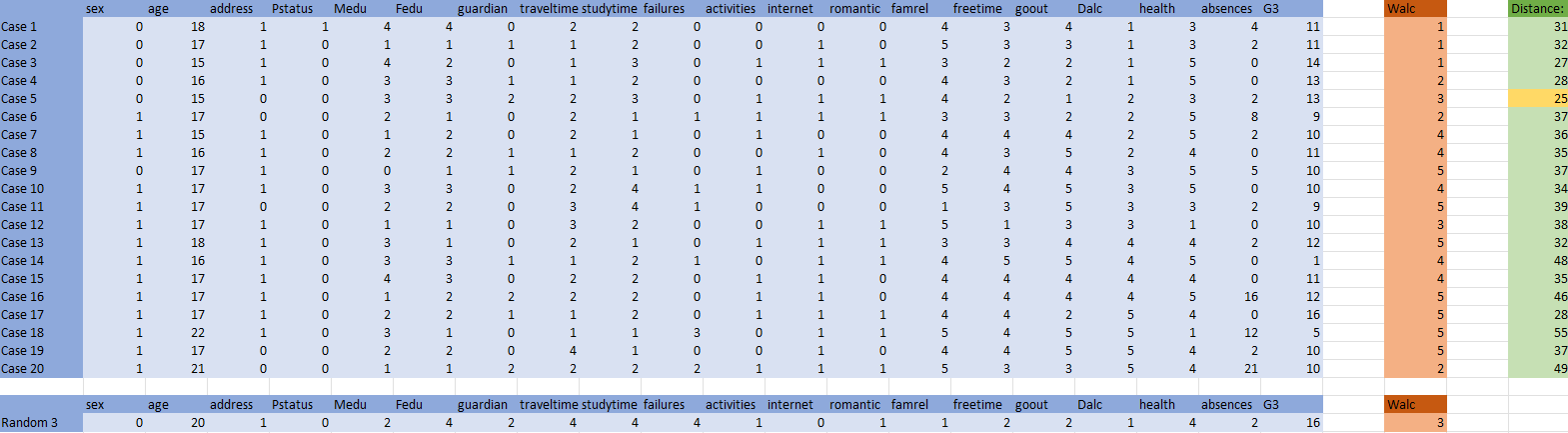
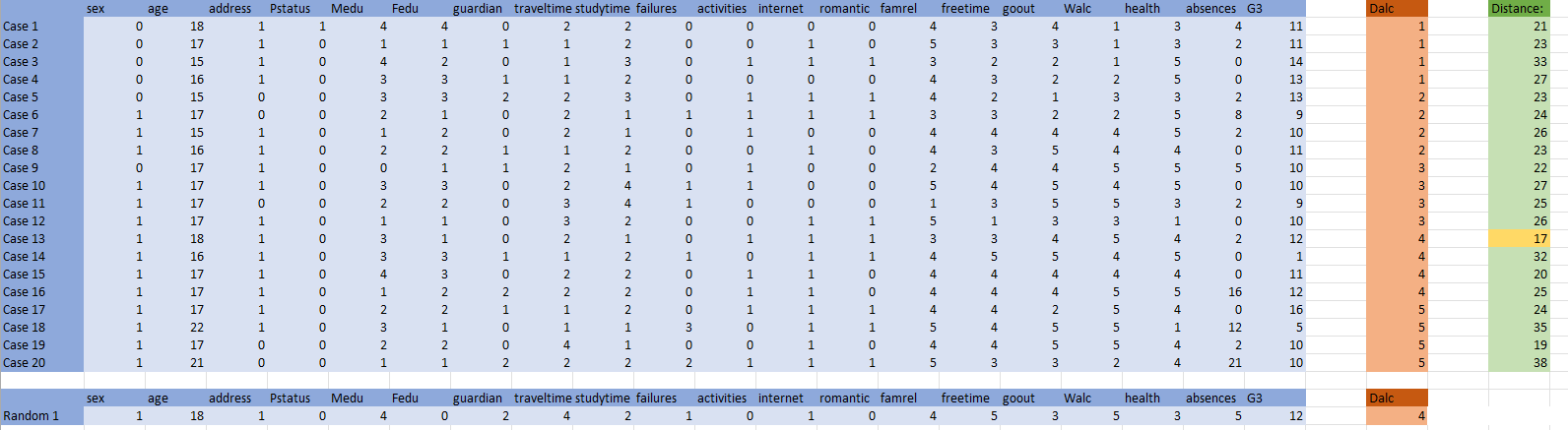


**Highest R-squared**

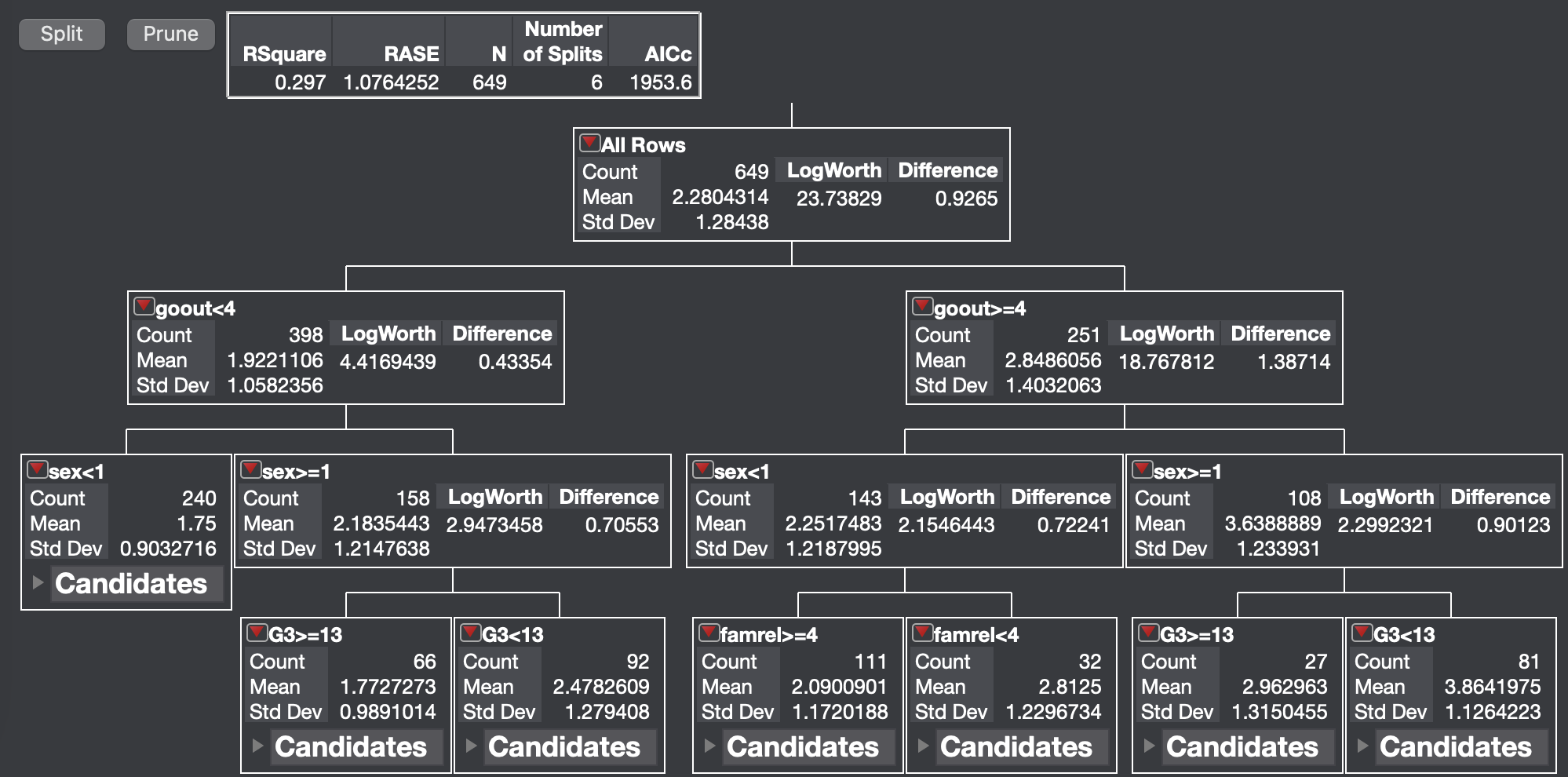
Weekend

1. Sex .32
2. Study time -.21
3. Going out .39
4. Health .11
5. Absences .16
6. G3 -.18

Memory Based Reasoning

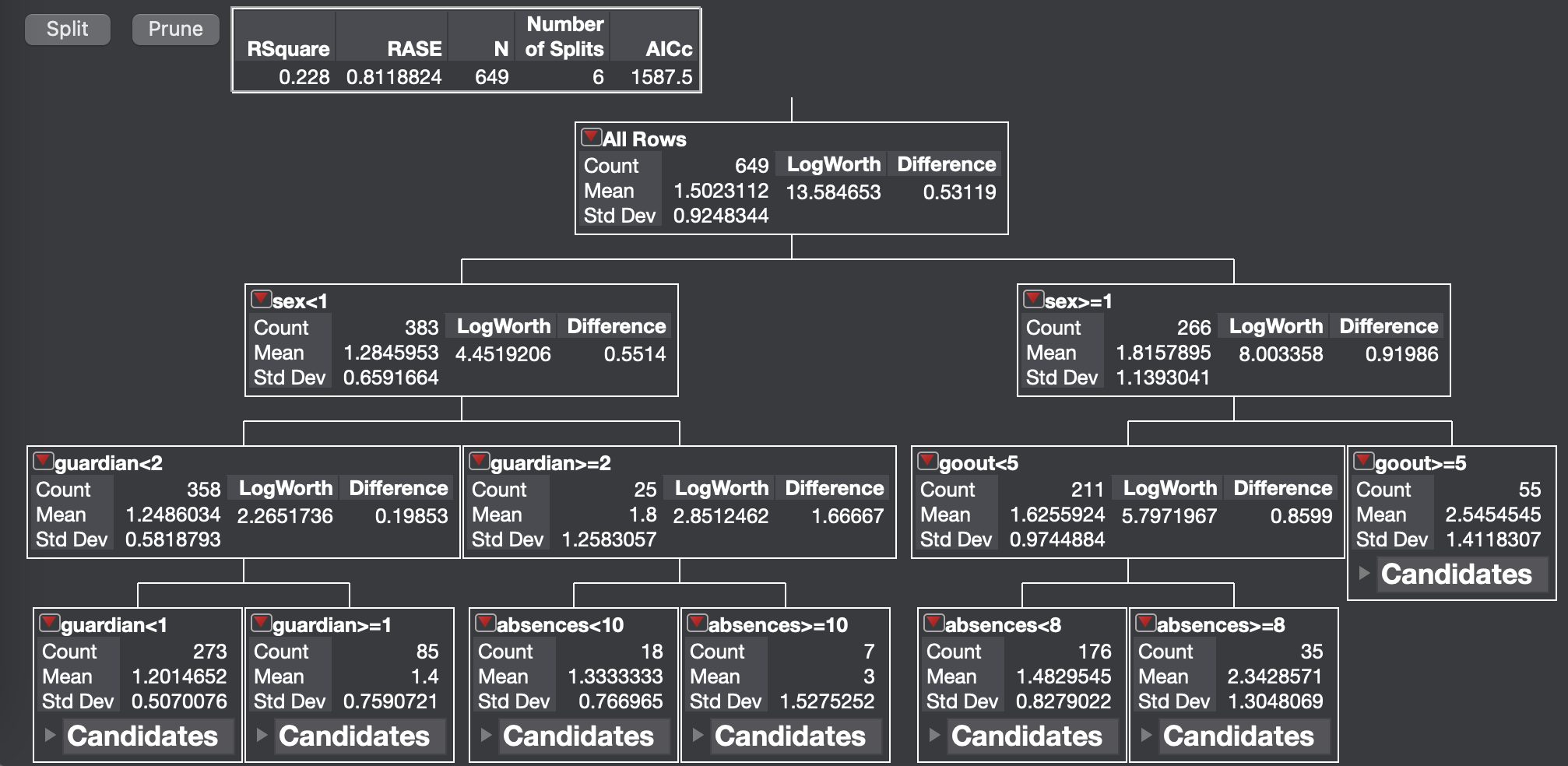


Weekend Alcohol Consumption Decision Tree



From this decision tree we can see that for both those who go out frequently and those who do not, their level of weekend alcohol consumption can be best predicted by their sex. For males who go out frequently, their alcohol consumption is best predicted by their final grade in school. For females their alcohol consumption is best predicted by the quality of their family relationships.

Workday Alcohol Consumption Decision Tree



This decision tree shows that sex is a primary predictor of workday alcohol consumption. For males, the frequency of going out is the next predictor while for females, whether their guardian is their mother or father compared to someone else is more important.

Neural Network