Final Project - QMST 3339 Musal

Executive Summary

Our group’s final project utilized a data set regarding student performance(scores) on core subject courses. The analysis consisted of a total combined average of the 3 core subjects. The core subjects chosen were reading, writing, and math seeing that these are some of the most important core subjects that are deterministic in a student’s future. Our group used a regression tree analysis and has shown that reading and writing scores are directly correlated with one another. This is no surprise due to writing and reading having similar characteristics pertaining to one another. Math on the other hand wasn’t directly correlated with reading and writing scores. Math being a more difficult subject for students across all fields was no surprise, as math requires more critical thinking to achieve an answer.

We wanted to dive into the understatement of scores for all students dependent on their backgrounds. To do this we had to understand and categorize the backgrounds of all students. Those categories included the students' gender, their race, whether their parents had a degree if the student ate food at lunch, or not, and if the student had a prep course to help assist him. By understanding these principles our scores will be able to show correlation in each of their respective regions, whether their low scores or high scores.

Furthermore, the analysis went deeper to see the correlation between scores of students of different ethnicity, students whose parents came from various education levels(degrees or not), and if the students had completed a prep course. This analysis also included gender, which came to the conclusion that females across the board on all subjects regardless of parental education, or course prep had done better than males. Among the other conclusions made from the dataset, we observed the scores of the students had completed the prep course. It had shown a positive correlation amongst that particular student's scores. This was of no surprise, that the prep scores better prepared the student for exams for reading, writing and math. Along with that positive correlation on scores, another positive we found shows that if the parent’s have higher education levels the student performed better than if the student's parents had lower education levels.

We also wanted to take a look into the correlation between proper student nutrition and educational success. How was the success rate (exams grades or scores) of students who had been given lunch vs. students who didn’t receive lunch? As one may observe in American society there are lots of low income area schools. Where providing lunch and breakfast to children and students is a major issue. For granted there are some school districts that provide free lunches to their students and these greatly enhance the nature of the student’s education as well as their nutrition and well being. With widespread issues regarding poverty and jobs in America arising to new levels. These factors are a result due to the pandemic and school systems and governments are going to require extra attention to manage student’s well being as it is being greatly affected in ways we cannot fathom right now. There are families who can’t provide food and shelter for their families and that means education becomes even less important. Lower quality of life results in more stress and more worrisome for the students. They have so much more to worry about than education opposed to students who come from wealthy/middle class homes. The more well off students just by observations we have made, perform better than the others. This may be something we can investigate further into the future to attempt to provide better quality education to lower income and students who aren’t as fortunate.

Problem and Data Description

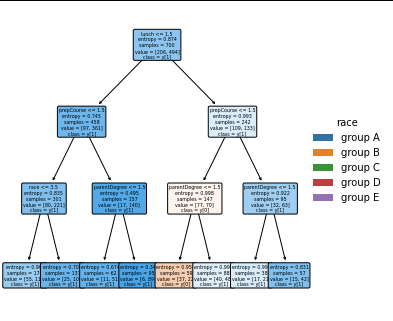
The dataset we provided as stated in the beginning is based on student's performance on reading, writing, and math. These core subjects growing up are very deterministic on how well a student is going to do, not necessarily in life, but in their educational path they choose to go on. The problem our group was trying to achieve was shedding light on important factors that affect a student’s educational future. For example, the education of the student’s parents was a great determinate in how well the student performed. The data model showed a clear representation across different genders and ethnic groups that student’s with parent’s of better education typically scored higher than students whose parents didn't have degrees. Unfortunately this problem is hard for society to address as there is no quick fix for this issue. Each family and student isn’t given the same opportunities as others. Some may argue that there is a same clean slate for everyone, but I believe as one gets older and wiser we realize that this isn’t true. The hand that a person is dealt may not always be as promising as someone else’s hand. Everyone doesn't have the luxury of choosing their parents, surroundings,and situations in life.

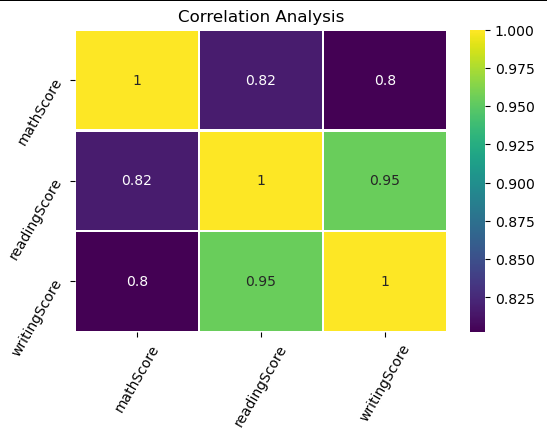
Aside from the stuff we cannot control. We are also seeing the positive effect that preparatory courses have on the overall average of students' future grades. The problem society faces regarding prep courses is they are usually very expensive and not funded by school districts. Parents are having to provide out of pocket costs for these prep courses. In a perfect world it would be awesome for every student to be able to take prep courses, but in the society we live in the amount of lower income students outweighs the higher income students. The lower income has trouble providing students proper lunch much less being able to provide extra funding for schooling. This is more relevant than ever considering the 2020 pandemic we are currently going through. Families are more financially at risk than they have ever been this year. The pandemic has also greatly affected our education across all grade levels from elementary through college students.

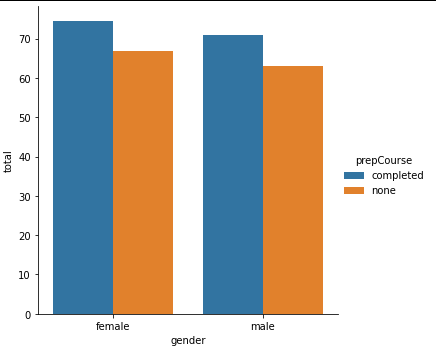
Conclusively, the correlation did show that students male or female who participated in the prep courses were shown to do better overall in math,reading and writing. The prep courses succeeded in better preparing the student’s for success. A solution to this issue would be to make a conscious effort to make prep courses more widely available to students. The prep courses wouldn’t need to be for all courses but mainly focused on the core subjects we tested in our analysis. (Math, Reading,Writing).

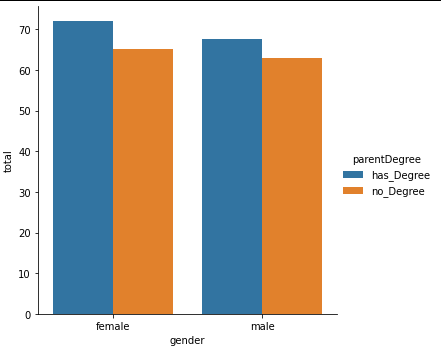
Describing and visualizing Data

In the data we can observe that when a male student has lower exam grades in comparison to a female student, a males exam grades are lower in both experiences. One with eating lunch, and the other, whether their parents had any educational background. While both these events determined some kind of impact on the students' exam score, the absence of one affects the male gender more than the female.









# Analysis

We chose to go with a decision tree (entropy) prediction model and a series of graphs. With us choosing the decision/regression tree route our group was able to predict the scores with an 75% accuracy. The correlation analysis is displaying the correlation between the grades which we were able to determine that reading and writing test scores were closely correlated. Math scores on the other hand weren’t closely correlated as the others. The bar graphs display an easy to read data regarding test prep scores between their respective genders along another bar graph regarding parents' education. The scatter plot analysis reflects the findings amongst the different race/ethnicity groups with each color representing a different race or group of students.

# Summary

The data show the many aspects of a student's nutrition, parents background, gender, race, and whether the student was assisted by taking prep courses for the test. By finding different test points the data proved to show many different data points. From these findings our project used correlation analysis, as well as using decision trees, and a scatterplot. The scatterplot, showed all the data points from the given answers and questions shown from the decision trees. From these data points we were able to gather enough data points to find correlations amongst these answers. Some of those correlations such as performing much better on the exams for the 3 tests to be more positive to when the student has parents with some form of college degree. And a negative correlation when the students’ parents have no educational experience and the student takes no prep classes. These 3 tools helped to vastly understand and prove that students' educational experience has some form of ties with the background they come from, and also including if they are getting their nutrients.

After observing data another one of the many problems we concluded was that a student having lunch showed a positive effect on scores in both males and females. If we as a society in the United States or the world, if we can manage to just provide a better nutrition to students, we can see a positive effect in scores and grades. Students are going to just perform at a higher level when they have the right nutrients. They would worry less about the stresses of hunger and be able to focus entirely on the education at hand.

In conclusion, the project we ran for understanding the different experiences we live as students shows correlative reasons for higher exam performances, and vice versa for lower exam performances. These reasons can be because we are more likely to be links of our parents from morals, education, and characteristics. Thus showing why the exam grades from the students would reflect that of their parents educational levels. Another finding from the whole data set proved or showed correlation that when students eat lunch everyday their scores are likely to improve. Now the data is not proving or disproving that test scores will be better by eating lunch, only showing that when students eat lunch they are more likely to earn higher test grades. To get to these final conclusions, the datasets being analyzed needed to be processed through the 3 different analyses to create a finding of correlation.

# Appendix

: Code. The code you write for your report. This is not going count toward 8 pages.