

Objective:

For this project, I am interested in exploring what factors cause students to achieve higher grades with machine learning. However, I am most interested in understanding what the data is telling me through visualizations. To measure students' grades, I will use the column **SEGRADES**. This column has the values “Mostly A’s”, “Mostly B’s”, “Mostly C’s”, and “Mostly D’s”.

The Dataset:

For this project, I used the curated 2019 National Household Education Surveys Program: Parent and Family Involvement Dataset made available by the After the Bell Fall 2022 Data challenge. This data was collected by survey from the [US Census Bureau](#) for the Department of Education. The PFI survey collects data about students who are enrolled in grades K-12 and asks parents questions about school choice and various aspects of parent involvement in education.

The Dataset consists of 75 variables/columns, including the **BASPID**, index/unique identifier, along with 15,500 responses/rows.

Setup:

First, I will start off the project with EDA by exploring some of the characteristics of the sample, trends, and relationships in the data, and checking for any anomalies. Next, because this dataframe is extremely large, I will use machine learning to find the ten most relevant columns that most contribute to predicting students' grades. After finding the four most relevant columns, I will explore them more through visualizations to try and find out what exactly the data is telling me. Lastly, I will analyze what the visualizations tell me and conduct research to figure out more about why these relationships may exist.

EDA:

The gathered data contains 8055 males and 7445 females. 48.86% are White, 13.58% are Black, 25.24% are Hispanic, 5.99% are Asian or Pacific Islander, and 6.33% all other races. 13782 students attend public school and 1718 students attend private school. In addition to student characteristics, 64.99% of students come from families with two parents and siblings, and 70.79% of parents are married. The parent marital status is shown in figure 3.

Some relationships that I explored first are the relationships that sex, public or private school, disability, and race have with each other. I didn't find any meaningful relationships other than Race vs Public or Private schools. I found that there are over eight percent more white students in private schools compared to public schools, as shown by figure 1. This relationship could be present because the average white family has higher income than other families, therefore can afford private school. [5]

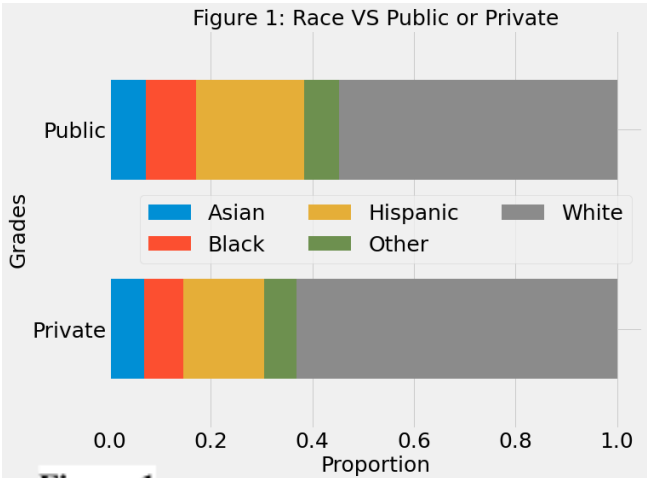


Figure 1

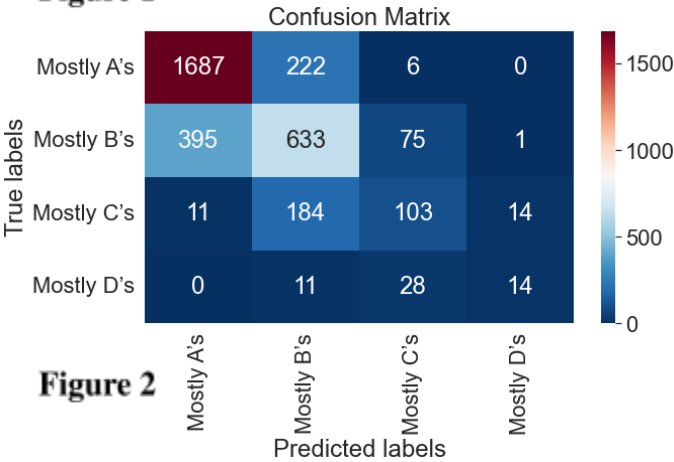


Figure 2

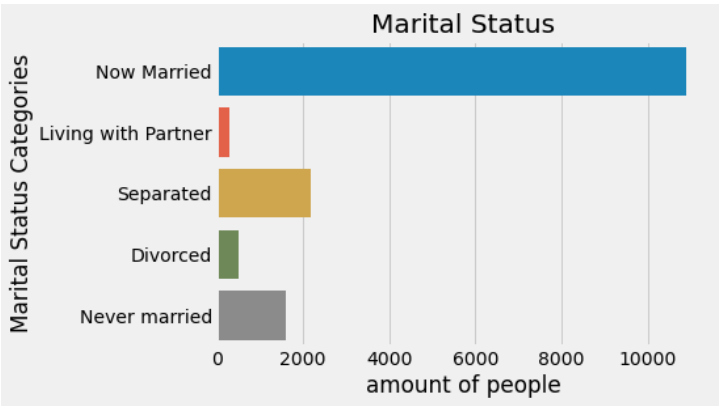


Figure 3

Machine Learning:

First, to figure out what columns are more relevant for predicting the child's grades, **SEGRADES**, I will use machine learning algorithms to find the 4 most relevant columns. I will do this by isolating **SEGRADES** as the output variable and using the rest of the columns, excluding **BASMID**, as input variables.

I will test multiple algorithms to see which model performs best, including Logistic Regression, Socratic Gradient Descent (SGD), Decision Tree, Random Forest, K-Nearest Neighbor, and a neural network model.

After doing this, I determined that the Logistic Regression Model performed best with an accuracy of .72. Figure 2 shows a confusion matrix with how the model predicted the child's grades vs what the true grades were. The diagonal going from top left to bottom right represents the correctly predicted labels. Randomly guessing the grade would result in an accuracy of .25, so an accuracy of .72 is decent. I was also impressed by the models ability to never classify someone who got a D with an A or an A with a D as shown by the top right and bottom left corners of Figure 2. Further cleaning the features in this dataset would result in a higher accuracy. It is also important to consider that the weights of the model are dependent on all the features in the dataset, however, I only analyzed the top 4.

Using this model I generated a table with the 4 most important features and the absolute value of their weight, a higher weight should somewhat correspond to a greater significance. This is shown in Figure 4.

Feature	Weight
SEGRADEQ	2.682816
SEENJOY	0.516907
CSEX	0.439534
FSCOUNSLR	0.396326

Parents View on Child's Work:

As shown by the first row of figure 2, with a weight of 2.68, the feature that best predicts the grades a child receives, is

Figure 4

SEGRADEQ. This column lists the responses to the question “How would you describe his or her work at school?”. Responses are as follows: "Excellent", "Above Average", "Average", "Below Average", and "Failing". There is also another option “Valid Skip”, however, I ignored this column.

As shown by Figure 5, the better a parent thinks their child is doing, correlates to a higher grade. When parents say that their child does excellent work, it is very likely that their child will get “Mostly A’s”, while “Above Average” will get “Mostly A’s” or “Mostly B’s”, “Average” will get “Mostly B’s”, or “Mostly C’s”, “Below Average” will get “Mostly C’s”, or “Mostly D’s”, and Failing will most likely get “Mostly D’s”. Something I found interesting is how parents had very critical views on how they described their child's work. For example, over 2500 parents said their child was doing “Above Average” instead of “Excellent”, even though their child placed in the highest category, “Mostly A’s”.

Similarly, over 2000 parents said their child’s work was “Average” even though their child received “Mostly B’s”. This could be because parents who have high expectations on their child's work, will achieve higher grades. [1] However, it is also more important to consider that if a child’s work is lower like “Failing”, then it is less likely they would get grades as high as “Mostly A’s”.

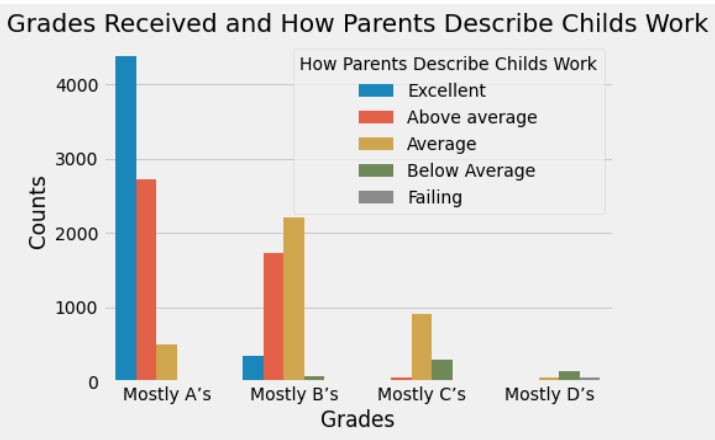


Figure 5

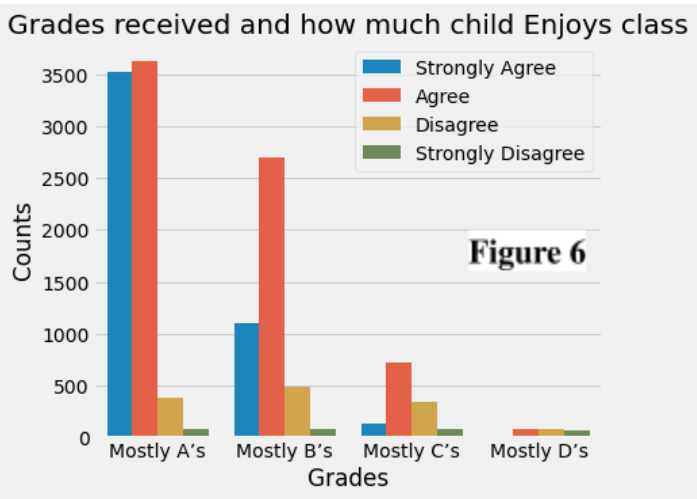


Figure 6

Child's Enjoyment of school:

The next most important category for predicting a child's grade is SEENJOY, with a weight of .52. SEENJOY asks parents how much they agree with the statement: “This child enjoys school.” Responses to this question are either “Strongly Agree”, “Agree”, “Disagree”, or “Strongly Disagree”. Once again I ignored the column “Valid Skip”. Figure 6 shows that if a child enjoys school more, they will receive higher grades. When parents “Strongly Agree” that their child enjoys school, it is most likely that their child will receive “Mostly A’s”. Similarly, when parents “Agree” that their child enjoys school it is most likely that their child will receive “Mostly A’s” or “Mostly B’s”. This trend could be because when students are happier, they will naturally devote more time to their work and put forth a better effort. When students at St. Andrew’s Episcopal School were asked about class, they often responded that happiness and positive feelings promoted quality schoolwork. [2] In addition, one student said, “I always feel pushed to do my best when I have a project that I find to be really interesting and fun” [2]. These perspectives clearly show why increased happiness might promote more quality effort which promotes higher grades.

Childs Sex:

The next most important category for predicting a child's grade is CSEX, with a weight of .44. CSEX asks if the child's sex is “Male” or “Female”. The relationship sex has on a child's grade is interesting. Females are more likely to get “Mostly A’s”, however, males are more likely to get “Mostly B’s”, “Mostly C’s”, and “Mostly D’s”. This is shown by figure 7. I was super interested in this trend of females performing better than males and wanted to explore it more. To start off, I wanted to find out the number of males vs females in each grade as illustrated by figure 8. After taking a closer look at this graph, I realized that the number of males vs females in the younger grades were approximately the same, however, I saw that as the kids got older, there were more males than females, showing that males possibly would go on to high school, but females would drop out.

This made me question if females were receiving higher grades in only the lower grades. To see if this hypothesis was true, I made another graph, Figure 9 that only looks at the students who receive “Mostly A’s”. After analyzing this graph, I saw that females still achieved higher grades than males in all grades except 4th grade. This is interesting because even though there are more males than females, there are still more females that receive higher grades, proving my hypothesis incorrect.

This really fascinated me and prompted me to do more research about why females are achieving higher grades than males. I came to realize that females achieve higher grades than males possibly because of their higher conscientiousness as shown by “a host of cross-cultural studies”[3] with one being from Columbia University that showed “female students far more likely than males to jot down detailed notes in class, transcribe what professors say more accurately, and remember lecture content better”[3]. This would naturally lead to females earning better grades than males.

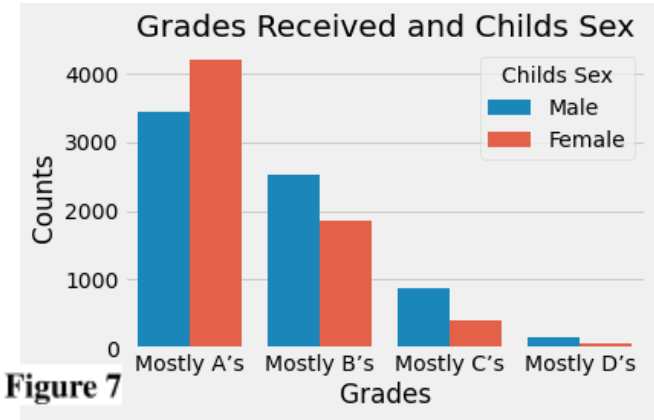


Figure 7

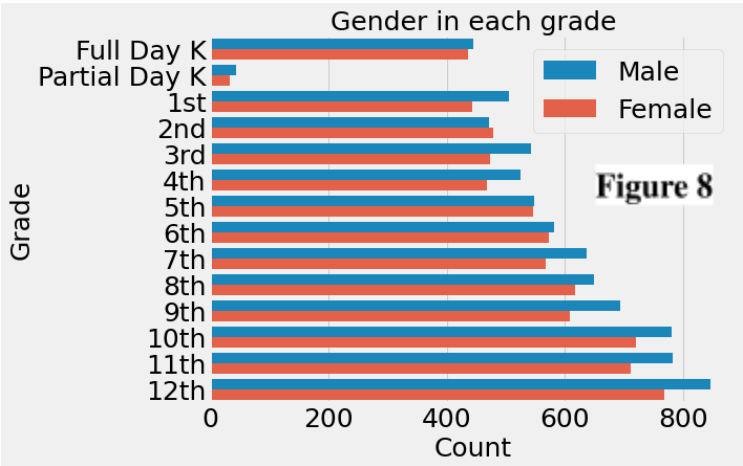
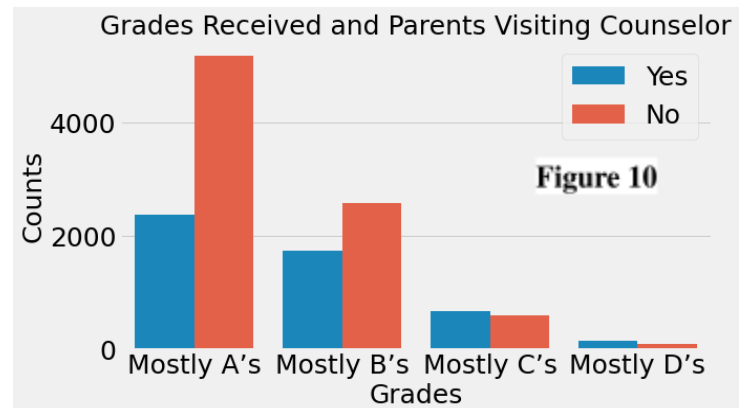
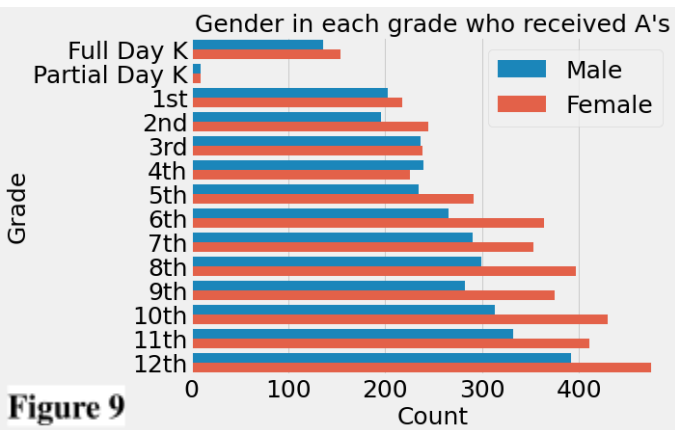


Figure 8

Parent visits to the guidance counselors office:

The next column I explored is FSCOUNSLR, parent visits to the guidance counselor's office. FSCOUNSLR asks the

question “has any adult in this child’s household met with a guidance counselor in person?”. The responses are either “Yes” or “No”. Figure 10 shows that parent visits to the guidance counselor's office correlates to lower grades. Students who receive “Mostly A’s” have over two times less parents that have visited the guidance counselor's office, while as grades decrease, the proportion of parents that have visited the guidance office increases. In fact, students who receive “Mostly D’s” have about twice as many parents who have visited the guidance counselor's office, inversely related to “Mostly A’s”. This relationship might exist because when parents visit the guidance counselor's office, it is usually for a disciplinary measure. Therefore, if a child is misbehaving at school, this can affect their learning and ability and performance on tests. [4]



Conclusion:

After analyzing several factors that influence grade, I came to realize that how parents rank their child's work, a child's enjoyment in school, sex, and parent visits to the guidance office all influence the grades a child will receive. Although I only analyzed these four variables, there are countless other variables that influence a child's grade. One other variable I think would be interesting to analyze but is not in my current dataset is the difficulty of the school the child attends, because some schools may be more difficult to earn higher grades at. For example, it might be more difficult to earn a high grade at a top ranked high school compared to an average elementary school. Therefore, this information would be important in predicting the grade a child would get. This data can be generated by asking each school for their grade distribution, the proportion of students who receive A's, B's, C's, and D's. I am also interested in learning how to better optimize my machine learning models to create better predictions in DATA 11900

Works Cited

- [5] Brown, Emma. “The Overwhelming Whiteness of U.S. Private Schools, in Six Maps and Charts.” *The Washington Post*, WP Company, 28 Oct. 2021, <https://www.washingtonpost.com/news/education/wp/2016/03/29/the-overwhelming-whiteness-of-u-s-private-schools-in-six-maps-and-charts>
- [4] Gnauti, Enrico. “Why Girls Tend to Get Better Grades than Boys Do.” *The Atlantic*, Atlantic Media Company, 23 Sept. 2014, <https://www.theatlantic.com/education/archive/2014/09/why-girls-get-better-grades-than-boys-do/380318/>.
- [3] *How Can Behavior Affect Academics for Students?* - Jano.us. http://jano.us/ngms/support/How%20Can%20Behavior%20Affect%20Academics%20for%20Students_%20_%20Everyday%20Life%20-%20Global%20Post.pdf.
- [2] Lauren Schiller PhD Student, and Christina Hinton Adjunct Lecturer on Education. “It's True: Happier Students Get Higher Grades.” *The Conversation*, 13 Sept. 2022, <https://theconversation.com/its-true-happier-students-get-higher-grades-41488>.
- [1] Yamamoto, Yoko, and Susan D. Holloway. “Parental Expectations and Children's Academic Performance in Sociocultural Context - Educational Psychology Review.” *SpringerLink*, Springer US, 4 Mar. 2010, <https://link.springer.com/article/10.1007/s10648-010-9121-z#:~:text=Students%20whose%20parents%20hold%20high,2007>