

Student Data Analysis Project

This is just an overview of the project and will only contain descriptions of the analysis and pictures of the graphs/charts produced by the code itself. For a full look at the descriptions, graph/charts and code please click the link for the [Google Colab Notebook](#).

Data Analysis Overview

In this project I will be analyzing the achievement and demographic information of 1000 students to determine if any or all of their demographic information can help to predict their scores on the different parts of the exam.

Data Overview

In the dataset there are 5 different non-numeric values which represent the student's information. They include **gender**, **race/ethnicity**, **parental level of education**, **lunch** and **test preparation course**. These values will be used to make predictions about a student's scores. Specifically, the scores are **math**, **reading** and **writing**.

Project Overview

The objective of this project is to be able to draw conclusions about the dataset as a whole. It is important to note that this is not a machine learning project designed to be able to take new input of students and make predictions about any new information. A future extension to this could be to run this data through a neural network to be able to make predictions on new data.

Project Code

The project data can be found [here](#). The initial exploration is just data shape and the mean/range of the different test scores (math / reading / writing).

Data Import/Exploration

Below is just a small sample of the data. In the code, the whole data set is imported at this point.

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
2	female	group B	bachelor's degree	standard	none	72	72	74
3	female	group C	some college	standard	completed	69	90	88
4	female	group B	master's degree	standard	none	90	95	93
5	male	group A	associate's degree	free/reduced	none	47	57	44

Below is the output from an initial data exploration.

```
Rows = 1000
Columns = 8
Header Names:
Index(['gender', 'race/ethnicity', 'parental level of education', 'lunch',
      'test preparation course', 'math score', 'reading score',
      'writing score'],
      dtype='object')
Math Information
Average Math Score = 66.1
Range = 100
Reading Information
Average Reading Score = 69.2
Range = 83
Writing Information
Average Writing Score = 68.1
Range = 90
```

Data Analysis

Project Imports

This project uses several open-source python libraries in order to analyze the data and plot the results of the analysis. Specifically, the Pandas, Seaborn and Matplotlib libraries are used in this project.

Exploratory Analysis and Visualization

In the sections below, the data will be analyzed in different ways and graphs will be displayed to show the analysis. For a look at the code which generated the graphs, make sure to check out the [Google Colab Notebook](#).

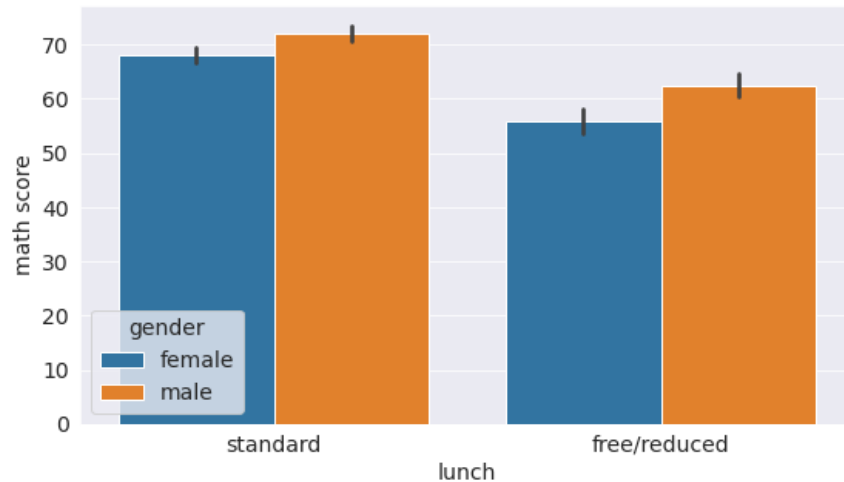
Average Performance By Gender

A breakdown of the average test scores based on a students gender.

	math score	reading score	writing score
gender			
female	63.6	72.6	72.5
male	68.7	65.5	63.3

Average Math Score By Lunch Option and Gender

The graph below shows that overall, if you have free and reduced lunch, you are likely to score low on the math test. Further than that, if you are a female and on free and reduced lunch, then your chances of scoring high are even less.



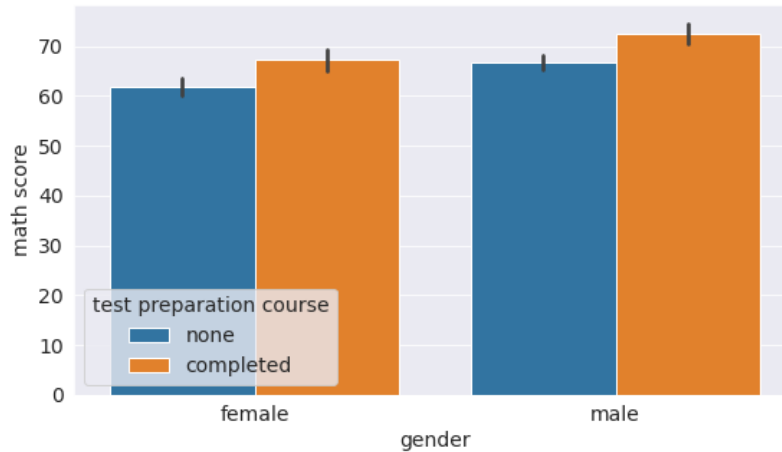
Average Math Score By Lunch Option and Test Prep

The graph below shows the students who take free and reduced lunch are still not able to get to the level of the standard students. This being said, if you are a student who has standard lunch, taking the course does not increase your score that much compared to the free / reduced lunch.



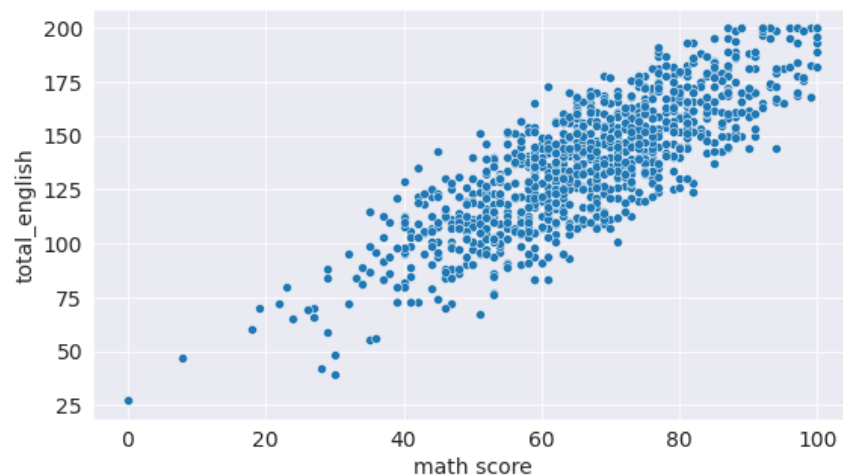
Average Math Score By Gender and Test Prep

This graph shows that on average, the male students are better than the female students on math. This also shows the average female student should be taking a test prep course if she wants to do as well as the male students on math.



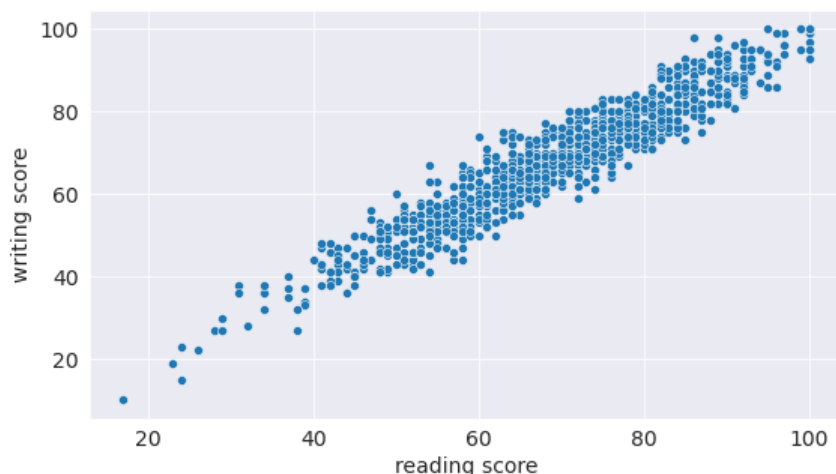
Math vs "Total English"

In the code below, a new 'total_english' score is created by adding both the reading and writing scores together. This is then plotted against the math score to compare the two. This shows first that there is a trend between the two sets of data and secondly that a student who does well in one area is likely to do well in another area.



Reading vs Writing

The graph below shows there is a very strong correlation between a student's reading and writing score.



Questions

In this section, questions will be posed and answered based on the data in the dataset. Each question will have a written response and some sort of graph or table to explain the written answer given.

What factors most affect the Math score?

Based on the data, there is not much a student or family can do to influence math ability. Even factors such as the level of education of the parents has little effect on the average math score. The biggest factor on a student's math score is the lunch option. Students who have standard lunch (meaning they pay full price) scored an average of 70.0 points on the math test. Students who have free/reduced lunch (meaning they can't afford to pay full price) scored an average of 58.9 on the math test.

The reason for this could be as simple as some students not having the same resources as others. It is not known if a calculator can be used on this test, but if it can the data makes sense. Parents who can afford to pay full price for their student's lunch are more likely to be able to afford a calculator for their student. Parents who are unable to afford full price lunch may not have the means to provide their student with a calculator for the test.

parental level of education		test preparation course	
associate's degree	67.9	completed	69.7
bachelor's degree	69.4	none	64.1
high school	62.1	lunch	
master's degree	69.7	free/reduced	58.9
some college	67.1	standard	70.0
some high school	63.5	gender	
		female	63.6
		male	68.7

Are male or female students likely to score well on the tests?

As you can see from the table below, male students are more likely to score high on the math test while female students are more likely to score well on the reading/writing portion of the test.

	math score	reading score	writing score	total_english
gender				
female	63.6	72.6	72.5	145.1
male	68.7	65.5	63.3	128.8

What factors most affect the English score?

Looking at the data, it is clear that 2 major aspects of a student's life are important in determining their result:

1. **Their parents education:** Parents who have more education may be more likely to read more and therefore read more to their children.
2. **Taking a prep test:** While taking a prep course for math has shown to not do much, for English, the students may get practice with the specific types of reading and writing assignments found on the test.

Looking at the data the lowest total_english score comes from **high school** for education level and **none** for test prep (123.7). In contrast, the highest score comes from **master's degree** for education level and **completed** for test prep (158.4).

		reading score	writing score	total_english
parental level of education	test preparation course			
associate's degree	completed	76.2	76.8	153.0
	none	67.9	65.8	133.7
bachelor's degree	completed	76.7	78.7	155.4
	none	70.6	70.0	140.6
high school	completed	67.8	68.1	135.9
	none	63.4	60.2	123.7
master's degree	completed	78.2	80.1	158.4
	none	73.9	73.4	147.3
some college	completed	76.0	76.5	152.5
	none	66.1	64.9	131.0
some high school	completed	70.9	70.4	141.3
	none	63.9	60.8	124.7

Can a student's race/ethnicity predict their score?

While these are just averages, group A did the worst on the exams across the board while group E did the best overall on all exams. In conclusion, this data shows race/ethnicity can be used as a means to help predict how a student might perform on the test.

	math score	reading score	writing score	total_english
race/ethnicity				
group A	61.6	64.7	62.7	127.3
group B	63.5	67.4	65.6	133.0
group C	64.5	69.1	67.8	136.9
group D	67.4	70.0	70.1	140.2
group E	73.8	73.0	71.4	144.4

Overall Conclusions

This data set of 1000 students shows what most educators already know. If your parents were wealthy enough to go to college and read lots of books, then they will read to you. If you are exposed to lots of reading material and also have the money to take a prep course, you will be fine on the test. Math on the other hand, there is no straight path for success. Some students are born with more of a math brain and others need to work hard to develop the problem solving skills to do well in math.