Ellie Culverhouse

1/24/23

Module 4 Analysis

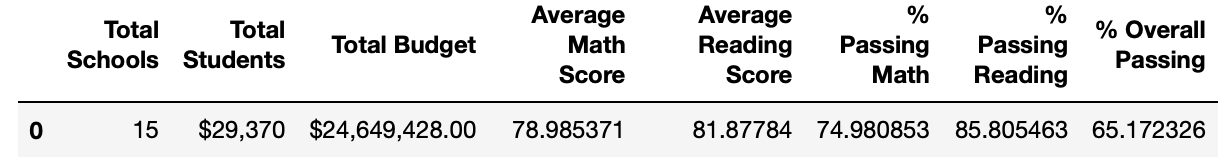
**1) Overview**  
 For this module challenge, I had to act as though I were my city’s, school district’s, new Chief Data Scientist. I assisted the school board and mayor to make strategic decisions regarding future school budgets and priorities.

My task was to analyze the district-wide standardized test results. I had been given access to every student's math and reading scores, as well as various information on the schools they attend. My goal was to aggregate the data to showcase obvious trends in school performance.

**2) Data, models, and results**

The initial data sets I was given were two CSVs. One containing student data which contained the student’s name, the school they attended, grade, gender, and their reading and math scores. The other with the school’s data which was their type of school, size, and budget. To make it easier to manipulate the data we combined the CSV datasets together.

First, I completed a district summary. Retrieving counts of the schools and students, the total budget amongst the schools, averages of the math and reading scores within the district, and what percent of students are passing their reading and math courses. By creating a new data frame of the results of this information I was able to create the following table.



Next, I completed a school summary. Using the necessary calculations, I was able to create a new data frame to summarize key information about each school, I found school name and type, total students for each school, school individual budgets, per student budget, average math and reading scores, and percent passing rates. I was able to display this data frame in the table below.

Table

Description automatically generated

We then look and the top and bottom-performing schools. Creating two new data frames to do this. This helps the district find what is working well and what needs more work within the district. The first table below is of the top-performing schools in the district.

Table

Description automatically generated

This second table is of the bottom-performing schools in the district.

Table

Description automatically generated

Then we will individually look at each grade comparing math and reading scores. This first table, on the left, displays the math scores per grade, per school. The table on the right displays the reading scores per grade, per school. Just by comparing the two, we can assume more students across all grades and most schools do better in reading than in math.

Graphical user interface, text, application

Description automatically generatedTable

Description automatically generated

We also look at the scores comparing them to school size. The following table shows the data frame I used to compare average scores and passing rates to the size of the school.

Table

Description automatically generated

Finally, we looked at school type. I looked at two types of schools, charter, and district. We compared the type of school to the average scores and passing rates.

Table

Description automatically generated

**4) Conclusion**   
 We used this school and student data to make comparisons and try to find patterns between student scores and what school they are going to and how much money that school spends, the type of school, and the size of the school to see if those factors affected the student's math and reading scores.

From our new data frames that made analyzing this data easier, we can make conclusions to answer these questions. First, looking at the school type, District vs. Charter, we can see that Charter schools in this school district have higher averages of both reading and math scores and a higher percentage of their students are passing these courses. When looking at the data frame showing the top and bottom-performing schools we can see that all the top-performing schools, were all Charters and all the bottom-performing were Districts. This can lead us to conclude that students going to Charters will perform better in class than in a District school.

Another conclusion we could draw is that size affects student performance. From our data frame comparing school size and course score we can see that large schools, those ranging from 2000-5000 students have lower reading and math scores and only 56% of the students are passing in the larger schools. But, in the small and medium schools <1000 or 1000-2000, there's a huger increase in overall passing at 90% of the students passing both.