

Module 4 Challenge

Written Assignment

Summary of the Analysis:

As the new Chief Data Scientist for my city's school district, I'll be helping the school board and mayor make strategic decisions regarding future school budgets and priorities. My first task was to merge school data with student data to create a comprehensive spreadsheet that contained all the relevant data to analyze. I then provided an overview of the district by calculating the number of schools in the district, the total number of students, the total district school budget, the average math and reading scores, the % of students passing math, % of students passing reading, and the % passing both math and reading. This provided me with an overview of the statistics for the main spreadsheet data of the entire district.

I then narrowed this data down in the School Summary. For the School Summary, I filtered the district data to show statistics for each school. In addition to the statistic categories for the District Summary, I also calculated the budget allocated per student for each school by dividing the school's budget by the number of students. I included the school type (district or charter) and the school budget for this datasheet. Looking at the datasheet, the school's name is the index, and the columns (stat categories) are School Type, Total Students, Total School Budget, Per Student Budget, Average Math Score, Average Reading Score, %Passing Math, %Passing Reading, %Overall Passing. The per-school summary allowed me to look at the relevant statistics per school.

I then took the per-school data and arranged it to show the top 5 best-performing schools in the % Overall Passing category and the top 5 worst-performing schools in the % Overall Passing category. I then sorted the data by grade, observing the averages for math and reading scores for each grade (9 through 12) for each school.

The next part of the analysis looked at Scores by School Spending. I calculated each school's spending per student and categorized it within the ranges: ["<\$585", "\$585-630", "\$630-645", "\$645-680"]. Then, I generated another data sheet grouping the collective data for all schools' average math, reading scores, %passing math, reading, %overall passing into the same ranges. This allowed me to see how academic performance may be affected by the amount of money subjected to each student for all schools.

I then looked at Scores by School size. I categorized each school as either Small (<1000 students), Medium (1000-2000 students) or Large (2000-5000 students). I created a data frame that analyzed academic performance based on school size. The same was done for School Types. I categorized each school as either District or Charter and analyzed academic performance based on the school type.

Drawing Conclusions:

- Looking at the Scores by School type, we can see that District schools performed significantly worse academically than Charter schools. District schools had lower scores for math and reading and a lower overall % passing rate. District schools % overall passing percentage (53.67%) vs Charter schools % overall passing percentage (90.43%).

- Additionally, we can see that large schools performed much worse academically than small and medium-sized schools. Based on the Scores by School Size, small and medium-sized schools had similar overall passing percentages of 89.88% and 90.62% respectively. Whereas large schools had an overall passing rate of 58.28%
- It is also interesting to note that the amount of money allocated per student does not correlate to better academic achievement. Looking at Scores by Spending, the %Overall passing decreased as spending ranges per student increased (inverse relationship). This allows us to eliminate the assumption that more spending = better academic performance.