PyCitySchools

The following section will look at and analyze the data frames created using two datasets: schools complete and students complete. The former contains 15 schools, their dependency type (Charter or District), their student population, and their total budget. The latter contains individualized student information, including students’ names, gender, grade, school, and their scores in reading and math—the sum of this latter 39170.

The first two tasks of the homework ask to build upon current information by using a combination of both datasets. The first task, District Summary, summarizes entire columns from the two datasets and, essential for us, finds the average and percentages of their math and reading scores. Of all of this information, the last three columns are probably the most insightful. We see that math scores are significantly lower than reading scores; this, too, appears in the average scores for the subjects, but we don’t see their distribution, so it's hard to get more information from this. The last column shows the percentage of students who passed both. Since this outcome is a lot lower than the sum and average of the former two, this tells us that a fair number of students are passing one but not the other. This is important because it shows that a sizeable proportion of students are capable of one subject but are failing to grasp the other subject.

In the second task, School Summary, the data frame breaks down the former into individualized schools. This allows us to see their sizes, total budget, budget per capita, and the average and percentages of their math and reading scores. From this data frame, we can see that scores and success rates are not equally distributed. Some schools have nearly all their students passing both subjects, while others hover barely over fifty percent. There doesn’t appear to be a middle, and more importantly, since these numbers show such a great contrast to the same metrics as in the first task, it tells us that the distribution of students is more in one type of school than another. Two opinions could be gathered from this contrast. Both opinions point to the type of schools; charter schools are more successful than public. However, this could be attributed to their structure only; charter schools, by their organizational structure, are inherently better than public schools. It could also be argued that population size is the critical factor. The biggest charter school is still eighty percent the size of the smallest public school. It, too, could be a mixture of both opinions. Since public schools’ sizes are bigger and they have more failing students, this skews all results.

To show this further, in the following two tasks of ranking the highest and lowest schools, the highest are all charter schools, and the lowest are all district schools. At the same time, the population of the former is significantly lower than that of the latter. One important thing to point out is that the average scores in both extremes only differ by about seven points in math and four points in reading. This is important because it points to potentially greater diversity in student’s skills at the public than in charter schools. With 90% of students passing both subjects in charter schools, the distribution of scores is potentially more clustered between 70 and 100. Public students need more extreme-achieving students to mitigate failing grades.

The following tasks, distinguishing scores by grade, could hint that the student's score do not change much in the course of high school. Furthermore, this could hint at the capability of the student once nurtured by the quality of the teachers in school, but we lack the information to make such a bold statement.

The task of showing bins of budget and size is important. However, since we do not know the scalability of schools the budget bin can be misleading. Bigger populations require more sizeable and secondary staff that may not be needed in small schools. For example, the ratio of security guard staff in bigger schools could differ from that in smaller schools, but as a whole, bigger schools will need more staff. In Holden High School, a staff of three security guards could suffice, and in Bailey High School, that number could be 25, thus requiring a big budget that does not go to direct educational material.

The last two tasks of grouping by size and type show that math is worst in bigger and district schools. Since this is almost tautological in these datasets, it is hard to know what is genuinely being said here. Nevertheless, it is constant with everything we have highlighted.