PySchools Written Analysis

In this challenge we are presented a large data set which which includes data on student performance in reading reading and math, student names/ID’s, gender, grade, school, school type, and school budget. Throughout the challenge are able to narrow this large data set into smaller data frames to discover trends among different student groups.

Our first task is to understand the metrics across the entire data set, which is represented in the DataFrame “district\_summary”. From the code we are able to count the total number of schools represented (15), total students (39,170), Total Budget, Average Math/Reading Scores, as well as passing rates. This is useful when determining how students are performing at a district level. From the results, we can determine that district-wide, Reading scores are slighty higher on average than math scores and in turn have higher passing rates (85.8% for Reading vs 74.98% for Math). The overall passing rate when combining reading and math passing scores is significantly lower (65.17%), which is a reasonable assumption as this would require the students to pass both courses.

While this may be useful to evaluate scores at a high level, the results may not accurately represent every school in the district. One school’s exceptional performance may be brought down by the average of a lower performing school. In the next section, we dig into the data further and evaluate on a per school level. Continuing the calculations from before and analyzing scores on a per-school level, we are also able to find how budgets may affect score results.

Next we use this more structured data sets to determine which schools are performing the best across the district with Cabrera High School, Thomas High School, and Griffin High School producing the highest overall passing rates. Rodriguez HS, Figueroa HS, and Huange HS produced the bottom 3 overall passing percentage. Once we determine the highest and lowing performing schools, we are able to analyze which grade levels may require more attention by producing the scores per grade.

Several factors may determine how a school is performing and results from the students. From the dataset we are presented with information on each school’s overall spending, size (by student count), and school type- which are outlined in the dataframes “spending\_summary”, “size\_summary”, and “type\_summary”. The most alarming is the discrepancy in overall passing rates between charter schools and district schools, with charter schools producing an overall passing rate of 90.43%, and District type schools severely behind with a passing rate of 53.67%. If I was a parent determining which school I’d like my child to attend, this data tells me charter schools do a much better job of preparing students for success in math and reading.