Justin Froehlich

Data Analytics Boot Camp

4/13/23

Pandas Challenge Report

**Summarizing the Analysis**

This challenge was set to look at different high schools within a district and compare student statistics such as populations, budgets, and passing percentages. To start things off, we calculated a few variables and created a new data frame that summarizes statistics for the entire district including total schools, total students, total budget, average math and reading score, and the percentages of students passing math, reading, and overall.

We then move into looking at statistics for each school individually. Through grouping the original data frame by the school’s name, we calculated all the same statistics as the overall district summary. We then created another new data frame that holds summary statistics for each school in its own row, with the school’s name being changed to the index of the data frame. We then sorted the data frame by values in percentage of overall passing students to determine the schools with the highest and lowest percentages.

Next, we split the data frame even further by creating new variables to hold each student based on their grade level. We then used those grade levels to calculate the average math and reading scores by grade at each high school and created another new data frame to reflect those numbers.

Finally, we looked at three other characteristics of the schools which were their size, their overall spending, and their school type (charter or district). We created bins to categorize each school based on these categories. We then calculated the same summary statistics for the three of these incidences and created three new data frames one holding scores by school size, one holding scores by school spending, and one holding scores by school type.

**Drawing Conclusions**

There are many conclusions in this analysis, but I will just focus on a few in this report. The first is comparing the highest performing school and the lowest performing schools. The first thing that shows here is that all the highest performing schools are charter schools, whereas all of the bottom performing schools are district schools. If we look at the data frame that holds score by school type, we can see that charter schools overall passing rate is almost 40% higher than district schools.

Another observation is that when looking at the passing scores by school size, we can see that small and medium schools have a great overall passing rate. However, the large schools with over 2,000 students have a passing rate that is just above 50%. One hypothesis for this is that teachers may be able to spend more time with students at the smaller schools.

One summary statistic that was surprising is that when the spending range per student is less than $585, the overall passing rates are higher than any other category. This seems almost counterintuitive, but when you look closer you can see that the district schools have much larger budgets than the charter schools.