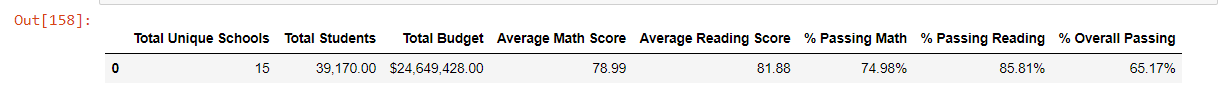
Jack Hayes

Homework 4 Pandas-Challenge

SMU Data Analytics Bootcamp

The Panda’s challenge was two sets of data. There was school data that contained schools in a district or charter with size and budgets of the schools. The other data det was students of the schools. This data set showed: where the student went to school, grade, gender, reading score, and math score. This purpose of the two data sets was to compare reading and math scores of students from each school. In the analysis python was sued to dissect the data. The two data sets were merged together in order to simplify the usage of functions to summarize the data.



(Figure 1.1)

In Figure 1.1 it summarizes the district data. There are 15 different schools with a total budget of $24,649,428. The average mean score per school for math and reading was 78.99 and 81.88 which is passing. On the other hand the percentage of students who passed math was 74.98% and reading was 85.81% which bring an overall 65.17% passing rate. Bailey High school and Hernandez High School and the lowest passing rate.

Next part of this analysis was dissecting per school data. This allowed the data to be further dissected into total school budget and per capita spending per school, average test scores per school, students per school with math/reading scores of 70 or higher, number of students per school that passed both math and reading with scores of 70 or higher and passing rates per school. The data shows that Bailey High school and Johnson High school had the most students, lowest math and reading scores, and lowest math and reading passing rates. Cabrera High School had the highest success in scores and passing rates while their total students were far less than Bailey and Johnson. We can make the conclusion that more budgeting does not create better test scores based on Cabrera high school and Bailey and Johnson High school.