Pandas Challenge Analysis

Analyzing the Highest-Performing Schools, and Bottom Performing Schools dataframes, we can see that the schools with the highest percent (%) overall passing all belong to the "charter" school type, while those with the lowest passing rate all belong to the "district" school type. Performing a hypothesis test to compare the % overall passing between district and charter schools, we obtain a p-value of 6.188e-21. The extremely small value allows us to reject the null hypothesis that the means of the two groups are equal, concluding that the difference between the passing rates is unlikely to have occurred by chance alone, and is therefore statistically significant. Despite the low p-score, the correlation between district schools to low overall passing, and charter schools to high overall passing, does not necessarily mean that school type causes the difference in passing rate.

When analyzing the Scores by School Size dataframe, we can also see that, in general, larger schools have a lower passing rate than smaller schools. A hypothesis test between small and large schools gives us a small p-value of 0.011, indicating that the difference between % overall passing, according to school size, is also statistically significant. Taking a further look at these factors, we observe that all of the district schools are large, while all but one of the charter schools are either small or medium sized. The charter school with the lowest student count (Holden High School - 427 students) had a % overall passing of 89.227%, while the one with the highest student count (Wilson High School - 2283 students) had a higher % overall passing (90.583%). Although the total number of students in the district school with the lowest student count (Ford High School - 2739 students) was not significantly higher than that of Wilson High School, the % overall passing was significantly lower (54.290%). This indicates % overall passing is likely more influenced by school type than school size.

A third factor that we can assume to have an impact on overall school performance is the Spending Range (per Student). However, looking at the Scores by School Spending dataframe, we notice that a higher budget is not indicative of higher passing rates. In fact, schools with the lowest spending range per student (less than $585) have a higher % overall passing rate (all above 89%), than those with the highest spending range ($645 - $680), all of which have a

passing rate of approximately 53.5%. Overall, district schools have a higher Total School Budget than Charter Schools, which is consistent with the previous observations.

Looking at the Math/Reading Score by Grade dataframes, we notice that there is not a lot of variability in the mean math and reading scores across different grades at a given school. Moreover, comparing specific schools' math and reading scores, we notice a consistency in the mean scores in charter schools, while math mean scores are significantly lower than reading scores across all district schools.