Brendan Golden

Module 4 Analysis

This module was a great introduction to the Panda’s notebook, allowing us to explore the possible capabilities. The data set introduces high schools and their students in a specific district, their respective grades, their individual math and reading scores, the number of students per school, and the total school budget for each subsequent high school. After initially merging the data, we calculated a plethora of relevant information, allowing for quantitative analysis and comparison for the district and each school. The preliminary calculations concerned the district wholistically as I calculated basic information such as the total number of students, the total budget, the average math and reading scores, and the percentage of students who individually passed the math and reading curriculum, and then the percentage of students who passed both simultaneously. Moving into the next part of the analysis, the assignment asked us to generate the same values for each school in the district individually, rather than for the entire district at once. The bulk of the calculations were the same except for using the group by function to display all statistical values at once using the number unique, size, and count functions. After the calculations, we were tasked with creating a complex data frame, which summarizes the statistics for each school at once; located in cell thirty of the notebook. The next requirement was straightforward as we needed to sort the data frame by the top and bottom performing schools based on the percentage of students who passed both math and reading. Continuing, we needed to filter the data frame, with the code provided, to examine the data for each grade and their average passing math and reading percentages. The final part of the module required cutting and binning the data to manipulate the school summary data frame into individualized, descriptive ones. The final data frames illustrate the spending ranges of each student at each school, and the relationship between the related percentage of students who passed math and reading, and the distinction between private and public-school pass rates.

The first conclusion, or distinction, that I found relevant in the data was the correlation between how the math and reading scores tended to decrease across the board as the spending ranges per student increased. By looking at the school spending data frame you can see that the schools having a lower spending range per student, either district or charter, exhibit higher scores across the board. While carefully examining each high school you tend to notice that this trend holds, transitioning into the spending range data frame. The spending range data frame validates the first conclusion as you can depict that the students with the lowest spending range have higher average scores and passing rates.

The second conclusion that I found present in the data is the correlation, and causation, that the size of the school's student body affects the average test scores and passing rates. As in the first conclusion, when cutting the per-school summary data frame by school size, the data depicts that smaller schools produce higher scores and passing percentages. This conclusion is summarized in the size summary data frame, as the concise data frame illustrates that schools with less than 1000 students have higher overall results compared to schools with between 1000 and 2000 students and even more so than schools with over 2000 students.