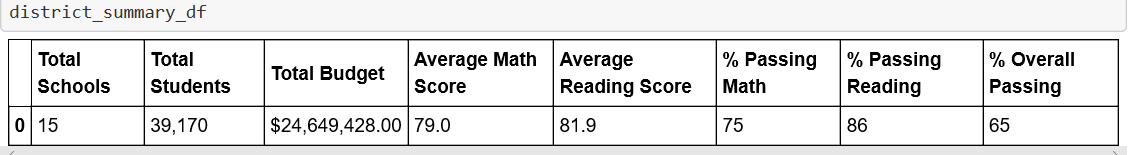
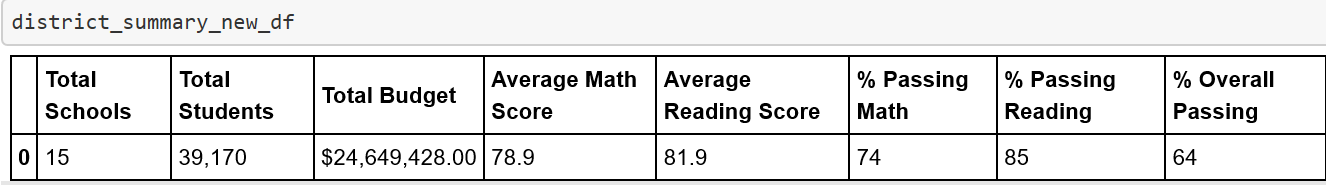
**Module 4 Challenge Analysis**

This module was focused on working with confidential school and student data to perform various functions and analysis. It was found that 9th grade scores from Thomas High School were incorrect, thus we were able to do before and after analysis to see the impact caused from removing the scores, while still performing all the same functions and tests on the data. The scores were replaced by NaN, not a number, so the school data can still be used. The following categories were used to analyze the outputs. The “new” variable was used to indicate the data after the replacement of 9th grade scores. This can be used to tell both datasets apart.

**District Summary**

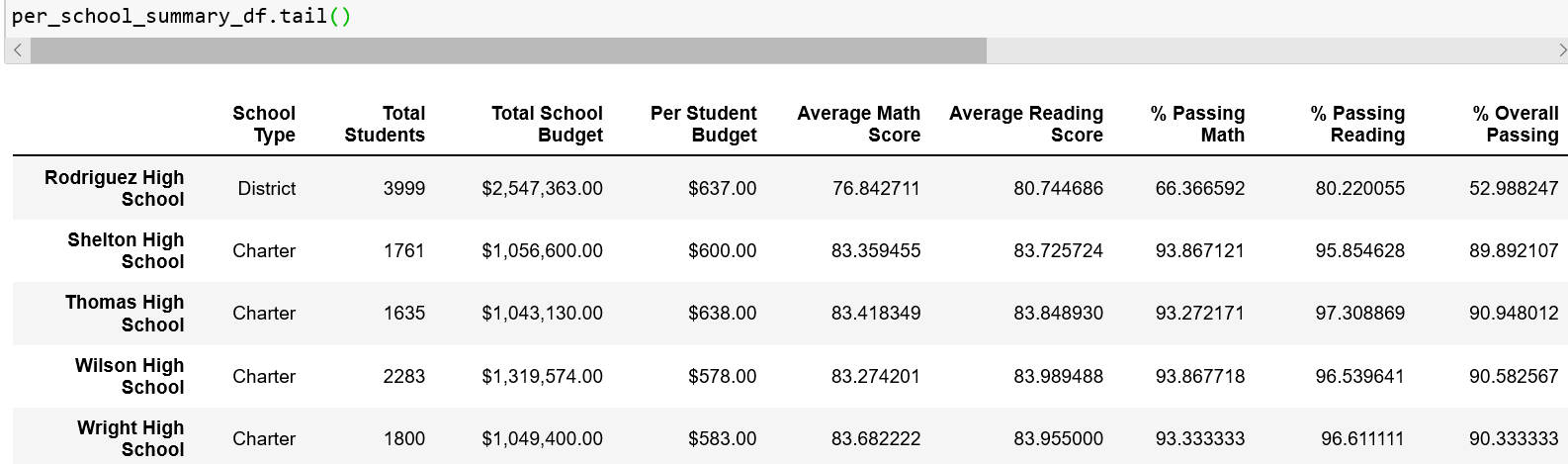
The district summary was a culmination of the average scores, percentages, total budget, total schools, and total students. It can be seen by the tables that it was not impacted much at all. The passing math and reading percentages differ by 1 point similarly to the overall passing percentage. This shows that removing the scores did not affect average reading and math scores as it was not calculated by grades separately. Thus, it can be concluded that the district summary is a big picture snapshot as opposed to a detailed summary.

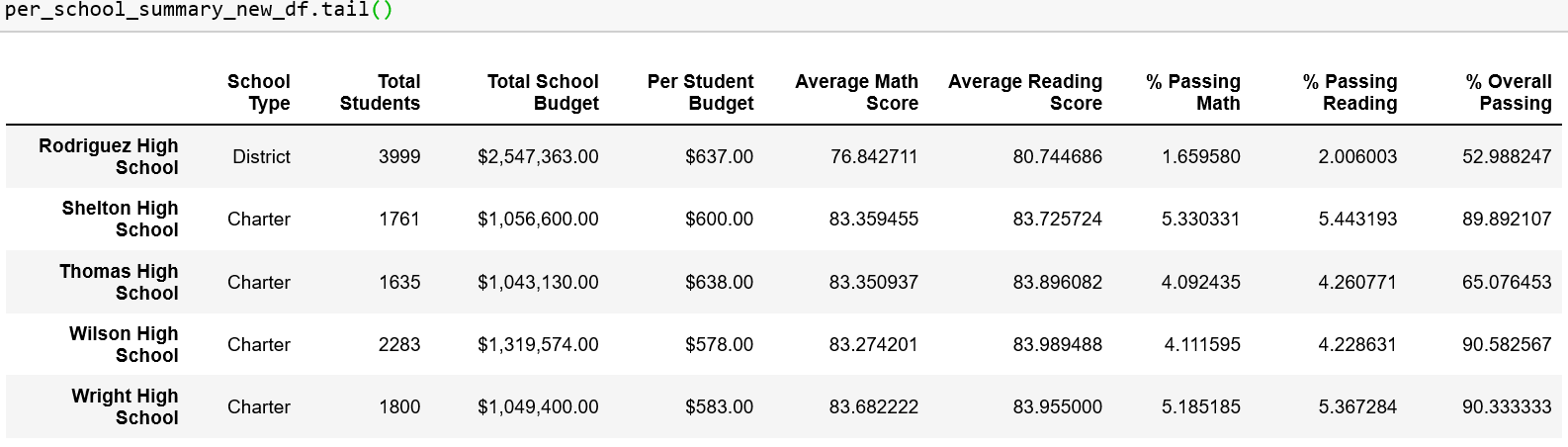




**School Summary**

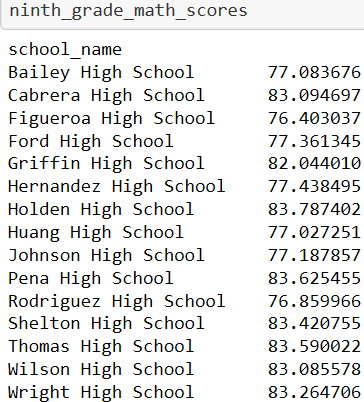
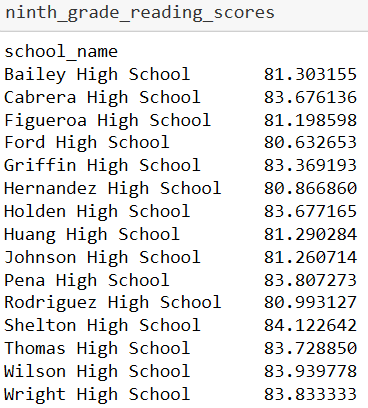
School summary included school type, listed each high school, and per student budget. This summary showed the most differences for Thomas High School. Initially, when all the data was included for 9th graders, the school did well with an overall passing percentage of 93%. However, after removing the data, the school only had a passing of 65%. Although, only the 9th grade data was removed for Thomas High School, it shows that it impacted its standing when compared to other schools in the school summary.

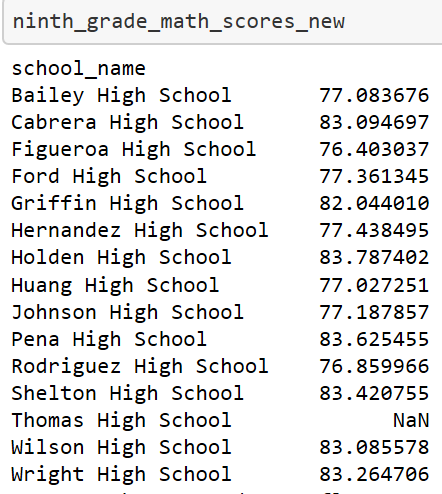
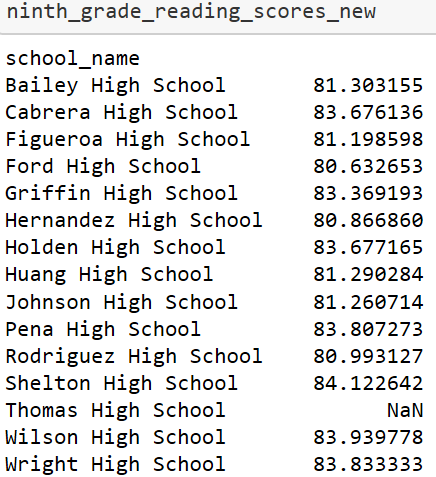




**Ninth Grade Data for All Schools**

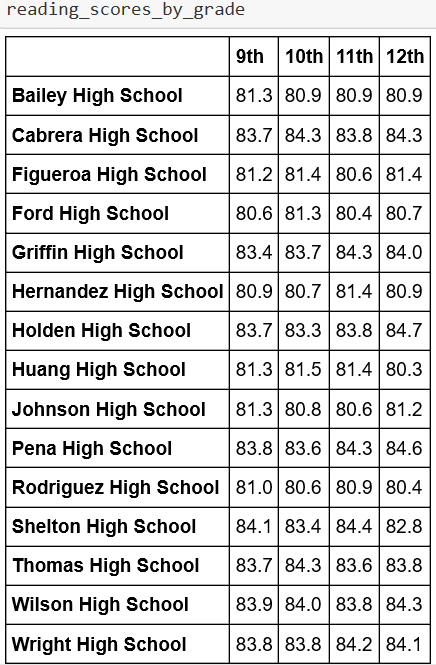
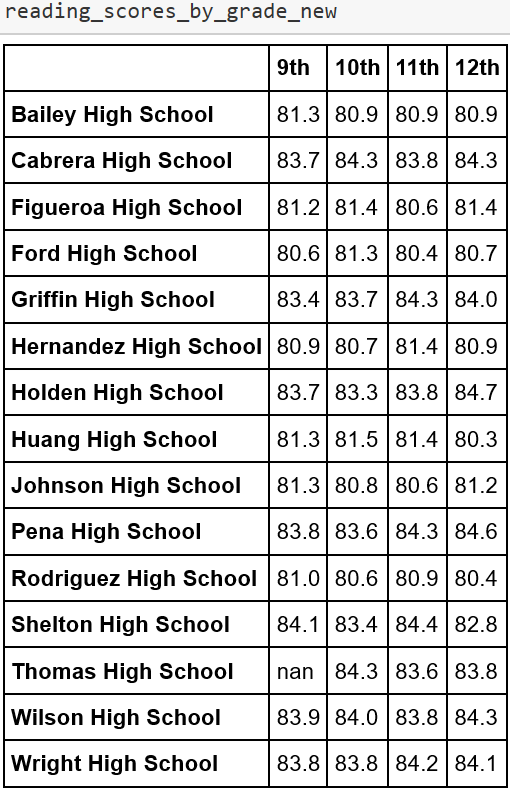
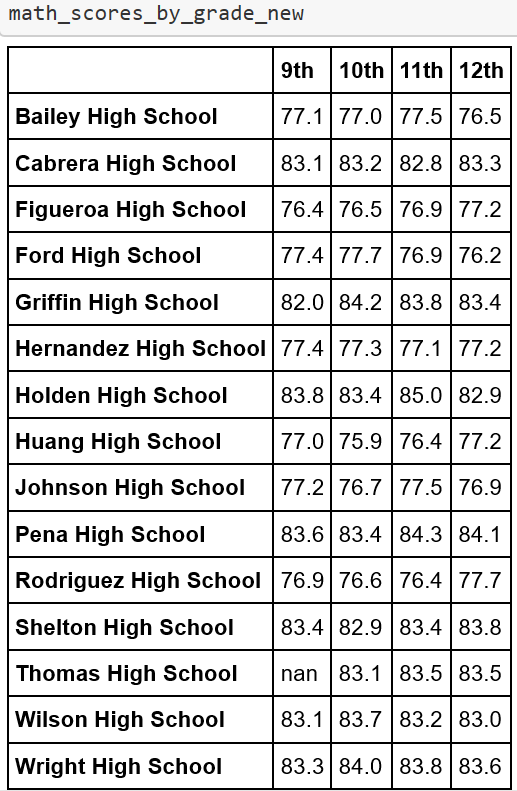
The 4 images below show the before and after of replacing values with NaN. When all schools are listed, Thomas High Schools presents with NaN. This means that for 9th grade data, Thomas High School was not providing any scores. This is a big impact since these scores along with other graders are used to perform further calculations. Thus, it can be said that this step determines the change in the values moving forward.

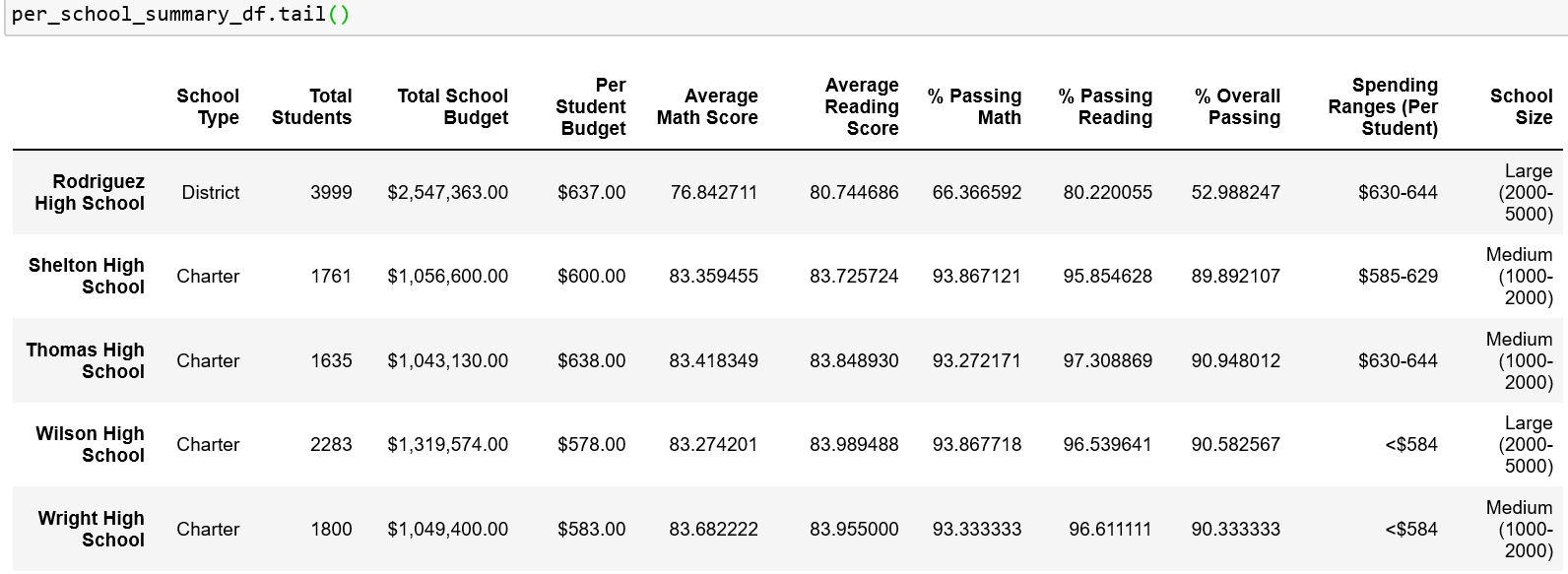
**Math and Reading Scores by Grade**

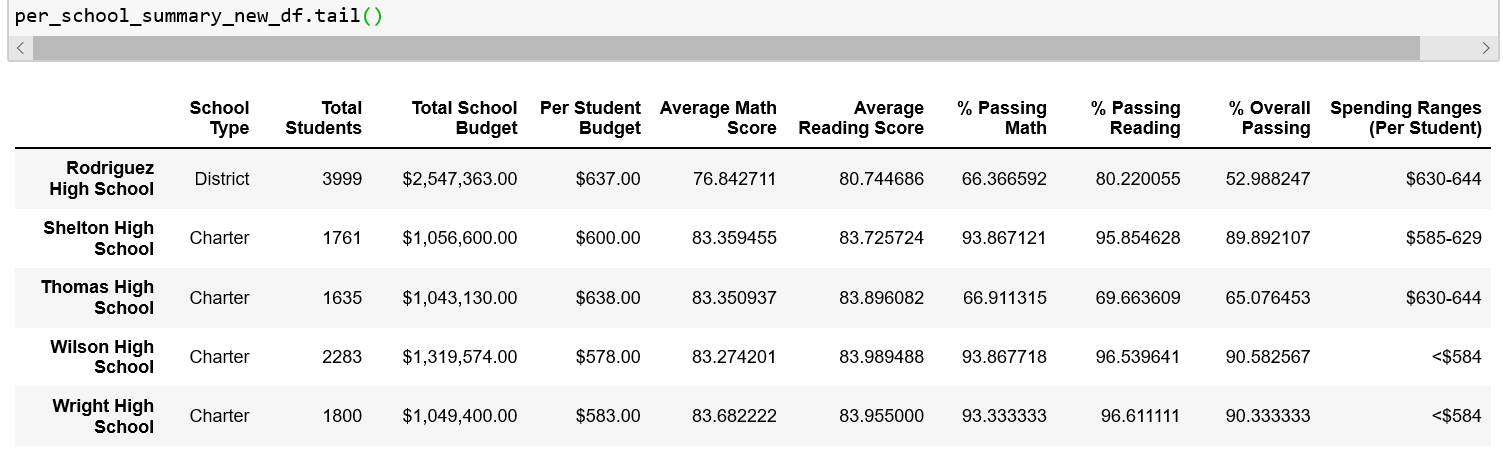
Similarly, to the above examples, the math and reading scores by grade continued to show how removing the 9th scores, eliminated any analysis to be performed for 9th graders in Thomas High School. Although, 10-12th graders data is still present, it is incomplete given an entire class is missing. This shows that Thomas High School’s analysis after was always incomplete and not accurate.

**Per School Spending Summary**

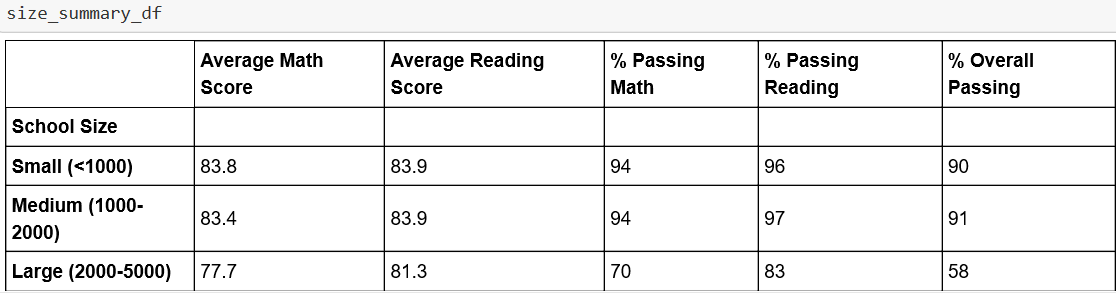
The spending summary contained previous values and included spending ranges for student which was calculated by spending bins. It can be seen from the images that the average scores for math, reading, percent passing math, reading and overall were all changed. The average scores, passing math and reading percentage show a few changes, but the most impacted was overall passing percentage. It was 90% and it is now, 65%. That large change indicated overall academic success for school and Thomas High School is unable compete now, as it was able to before. That is important to note because if this data is used to monitor academic success, Thomas High School would be up for scrutiny.

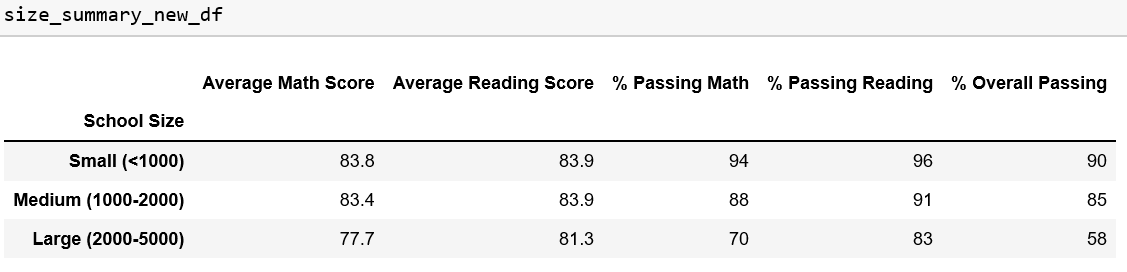




**School Size Summary**

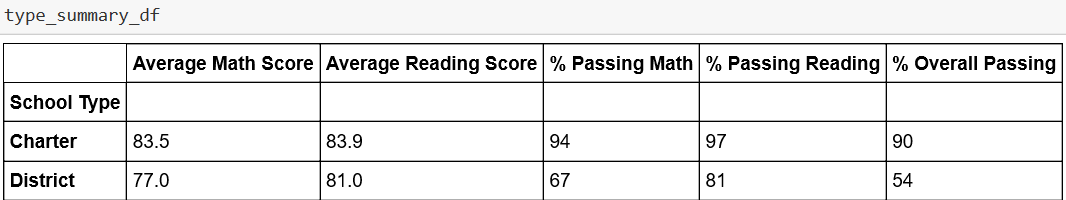
In the school size summary, Thomas High School falls into the medium range with 1000-2000 students. Thus, the values in those columns were changed. Specifically, the percentage of math, reading, and overall. There are larger differences now in comparison to the District Summary. This is because more data was used in culmination to calculate the School Size Summary. Also, after removing the 9th graders from Thomas High School, the number of data in the frame was also decreased which would cause a decrease in the percentages.

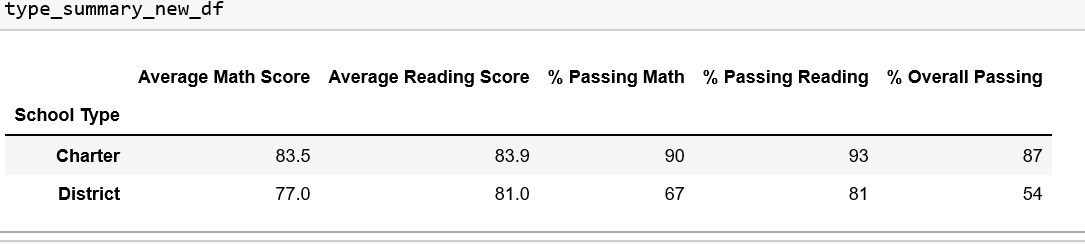




**School Type Summary**

Thomas High School belongs to the charter type. Those rows for passing reading and math percentage and overall percentage were affected by a few points. Like the size summary above, the removal of 9th grade data of Thomas High School impacted all the functions that were performed when the final school type summary was calculated. Although, it is minor differences, it still shows that within the category of charter and district types, there were significant changes to observe.





**Conclusion**

The premise of this challenge was to repeat all necessary calculations and formatting after filtering by Thomas High School and then replacing the 9th grade scores with NaN. This was done as the scores were deemed inaccurate and should not be included in the data analysis. After performing all the steps, we were able to do before and after data analysis to see the impact of removing the data from Thomas High School. The replacing of data affected the school’s academic standing in comparison to other schools. If this analysis was being used to dictate school performance, Thomas High School would be questioned. Lastly, the analysis provided big and small picture impacts to the data before and after. In the future, if multiple school’s grades needed to be replaced, the large impact can be seen again if all the steps are performed. This is a good way to analyze the data repetitively even if new information was added.