Report for Module-4 Challenge

1. **Overview the task.**

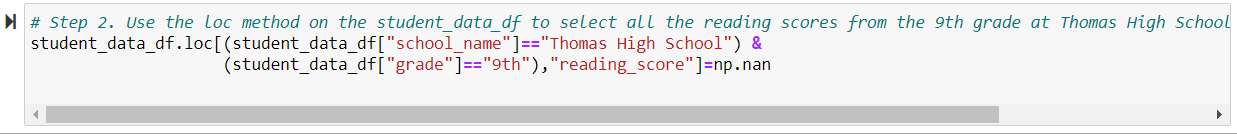
This task is about replacing the 9th grade Thomas High School math and reading scores to missing values, NaN. The reason for the replacement was due to evidence of academic dishonesty. Although the school board does not know the full extent of the academic dishonesty, they want to uphold state-testing standards and have asked me to replace the math and reading scores for Thomas High School with NaNs while keeping the rest of the data intact. Once the replacement is done, we need to repeat the school district analysis that we did in this module and write up a report to describe how these changes affected the overall analysis.

**The purpose of the report is**

1. To write a python code script in jupyter notebook that replace the 9th grade math and reading scores to NaN
2. To compute the percentage of students that pass math, reading and both from Thomas High school based on the number of 10th -12th students from the school
3. Update the school district results and analysis based new number of students

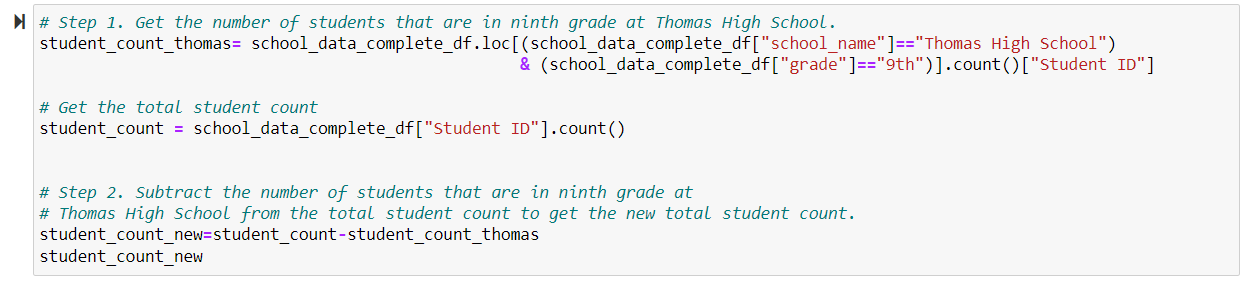
**Code for replacement**

I used the loc method on the student data and inside the loc method,I used the ‘&’ logical operator to select the rows of 9th grade Thomas Highschool students. Outside the loc, I used the equal operator and set it to NaN using ‘numpy nap.nan’ method. The following code replace the reading scores to NaN.

The math results was also replaced to NaN by refactoring the above code(i.e replacing the “reading\_score” with “math\_score”).

To update the school district summary, the newly cleaned student data is merged with the school data using left merge method (student data on the left) and on the school’s name column which is common on both data types.

To update the percentage of students who passed reading, math, and both, we need to reduce the number Thomas Highschool 9th grade students from the total number of students. To get the number of Thomas Highschool 9th grade students, I used the following code.



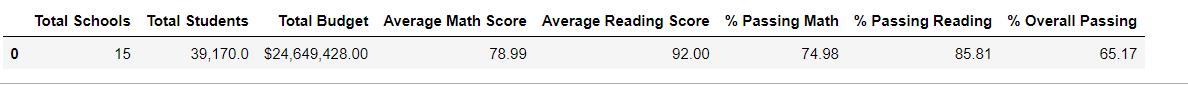
Based on this number, we can get the new total number of students by subtracting the number of 9th grade high school students from the total number of students(student\_count). In what follows, I compare the results with and without replacement of the 9th grade math and reading to NaN.

1. **Results**
   1. **District school summary**

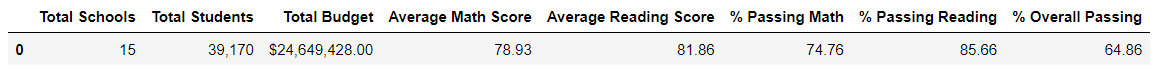
There are some differences in the districts summary results before and after the replacement of 9th grade Thomas Highschool to NaN. This change comes due to the change in total number of students. The Tables-a and b shows that there are some changes in the scores and percentages of pass. For exmaple;

* At a district level, the average reading score was 92 before the replacement, it changed to 81.86 after replacement

1. **District summary before replacement**



1. **District summary after replacement**

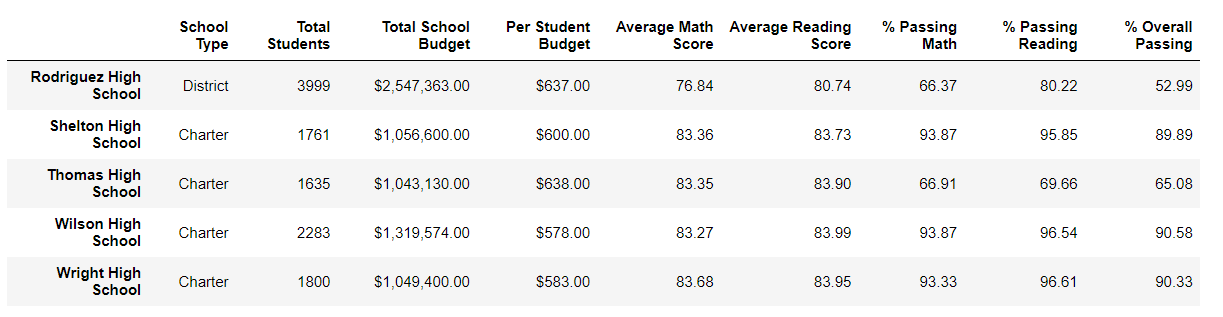
****

* 1. **School Summary**

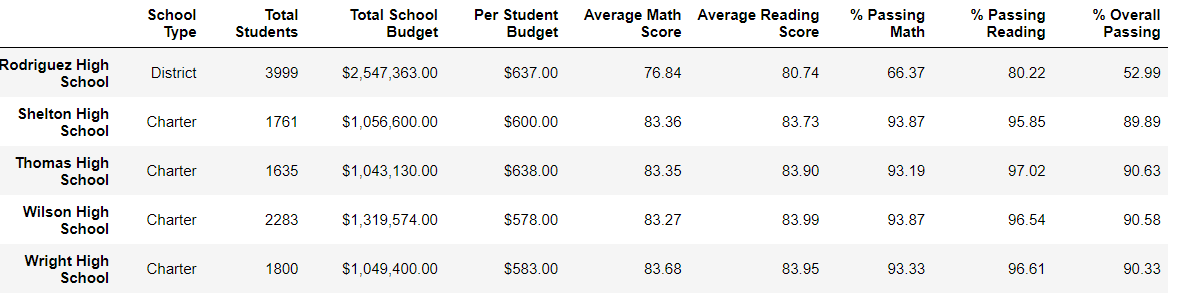
**Tables c and d below shows the school summary for the last five rows of the school summary data frame. As can be seen from these two tables, the results are similar before and after the replacement for all schools except for Thomas High School. Fore Thomas High school**

* **The percentage of students who passed math were 66.91% before the replacement, it became 93.19% after the replacement**
* **The percentage of students who passed reading were 69.66% before the replacement, it became 97.02% after the replacement**
* **The percentage of students who passed both math and reading were 65.08% before the replacement, it became 90.63% after the replacement**

1. **School summary before**

****

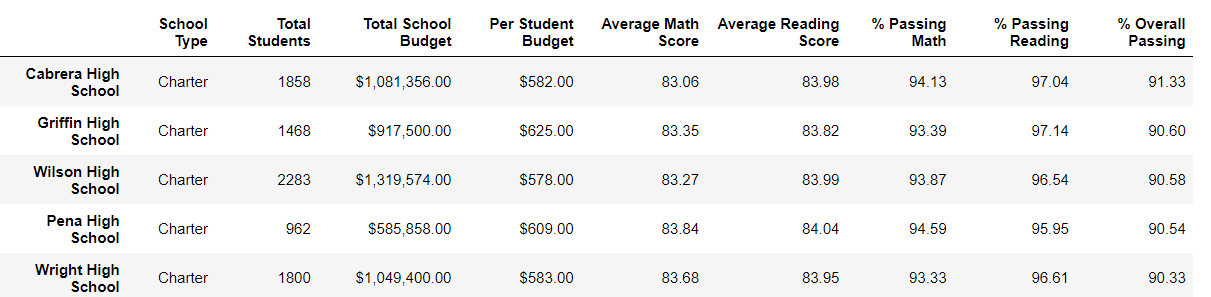
1. **School summary after**

****

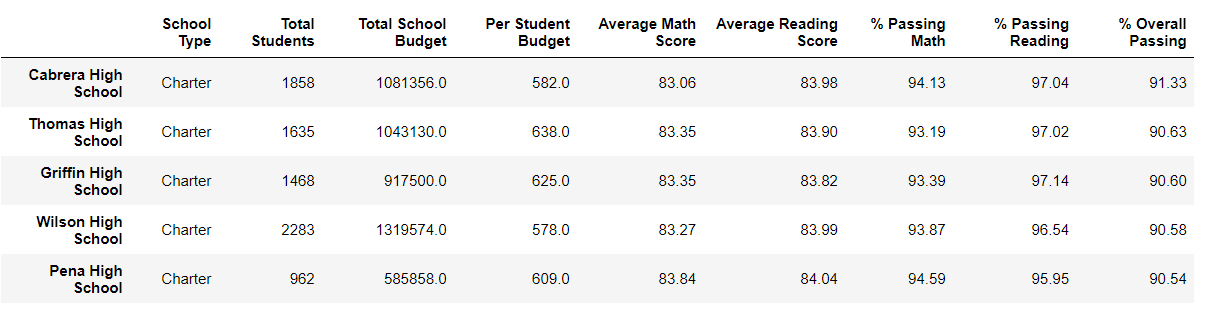
* 1. **Comparing Thomas Highschool before and after the replacement**

**We can compare schools based on the overall (both math and reading) passing percentage. Before the replacement, Thomas High School was not among top five performing schools (Table-e). But after the replacement, Thomas High School became the second top performing school (Table-f)**

1. **Top five performing schools before**

****

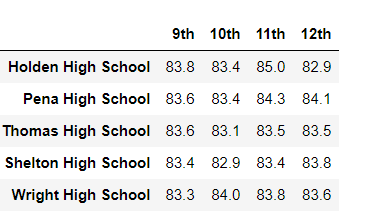
1. **Top -5 performing schools after**

****

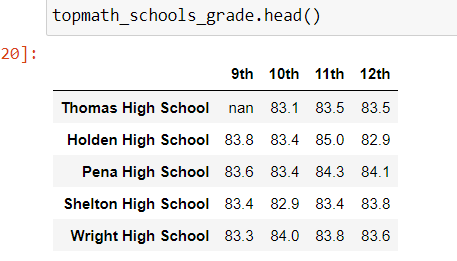
* 1. **Math and reading by grade**

**The math result for grade 9th for Thomas High School before the replacement was 83.6(Table-g). After the replacement, the result is changed to NaN(table-h). For all other schools, there is no change.**

1. **Top five math result by grade, before the replacement**

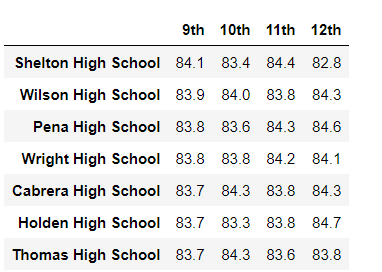
****

1. **Top five math result by grade, after the replacement**

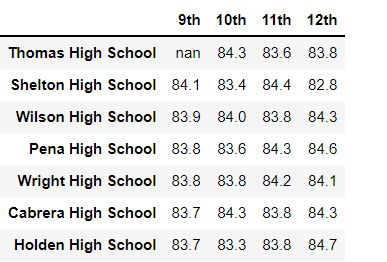
****

**The result for reading is similar to the math. Before the replacement, the reading score for 9th grade Thomas Highschool was 83.7(table-i), after the replacement it became “nan”(Table-j).**

1. **Top seven reading by grade, before replacement**

****

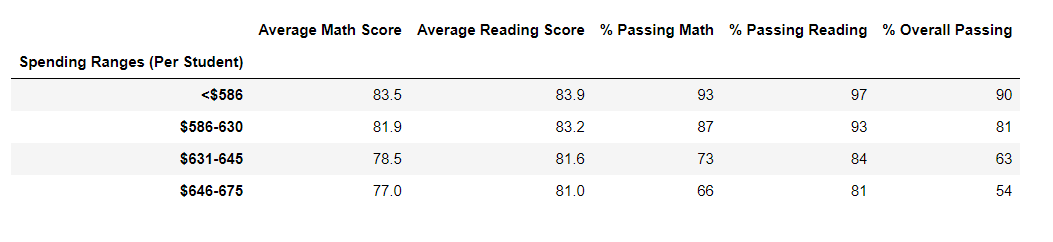
1. **Top seven reading by grade, after replacement**

****

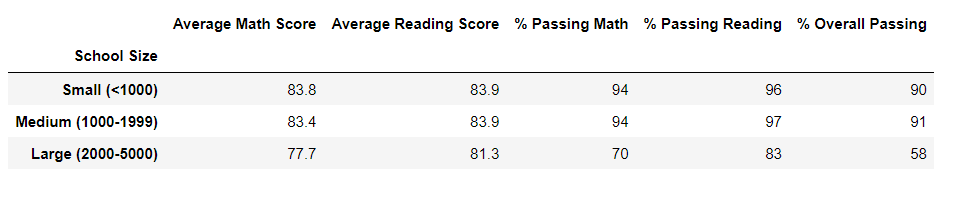
* 1. **Scores by spending, size and type**

**The codes result in similar results before and after the replacement.**

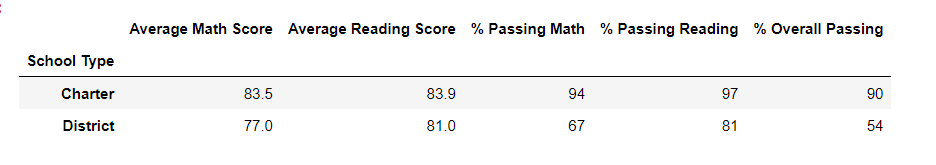
1. **Scores by school spending**

****

1. **Scores by size**

****

1. **Scores by type**

****

1. **Summary.**

**The Math and reading results were replaced for 9th grade students of** **Thomas High school, but this does not bring change to the magnitude of scores and percentages of other schools. The replacement; however, improves Thomas High school’s performance relative to other schools. The replacement also slightly changes the districts school level performances. The results show that scores by size, spending and school type were unchanged by the replacement.**