# School District Analysis

Module 4 challenge – deliverable 3.

## Overview of the school district analysis

### Purpose

We are introduced to Maria who is a chief data scientist for a city school district. Maria is tasked with preparing all standardized test data for analysis, reporting, and presentation to provide insights about performance, trends, and patterns. These insights will be used for discissions and for critical decisions at the school district level. This could include which schools should get more funding, less funding, if there is some type of connection between the performance of students and the type of school, etc. We are asked to help Maria analyze data on student funding and standardized test scores. This analysis will be used to aggregate the data and showcase trends in school performance. The Panda’s library along with Jupyter notebook will assist us with the data wrangling. In the beginning parts of this module, we are given several tasks to perform for our full analysis. We are asked to get the district key metrics and each schools key metrics. We also want to get the metrics for:

-The top and bottom 5 performing schools based on overall passing rate.

-The average math score received by students in each grade level at each school.

-The average reading and math score received by students in each grade level at each school.

-The school’s performance based on budget per student, school size, and the type of school but as separate factors.

After completing this analysis, we are then asked to replace the math and reading scores in Thomas High School for the ninth graders, with “NaN” values because of claims of academic dishonesty. We are then to repeat the entire analysis with the ninth-grade values for that specific high school absent. Overall, we want to see how these changes to this specific grade for this specific high school has affected the overall analysis.

## Results

### - How is the district summary affected?

I will admit, for most of the metrics we re-calculated, there was very minimal change from the original data. The difference in the original and the new district summary was in the average math score, % passing math, % passing reading, and % overall passing columns. Because the ninth graders total count was such a microscopic number compared to the total students in the district, the effect is minimal in this summary. All the metrics mentioned here as columns, dropped by 0.1% – 0.3%. The two summaries are shown here:

![original district summary][]

![new district summary][]

### - How is the school summary affected?

This is the main area where there is a huge difference in the values from the original school summary and the new school summary. In the new school summary, the values for % passing math, % passing reading, % overall passing all increased by about 25% - 30%! This is actually a positive impact. It seems that removing the ninth-grade scores put this school in the 90% passing percentage. Using the following images, look for Thomas High School in the third to last row. At the far right of the images, we can see the three columns where the percentage changed from 60’s to 90’s.

![original school summary][]

![new school summary][]

### - How does replacing the ninth graders math and reading scores affect Thomas High Schools performance relative to the other schools?

In the original top 5 schools table, Thomas highschool is in 2nd place. In the new data without the ninth grade data, Thomas highschool is still in 2nd place. In this table more columns are affected. The average math score, % passing math, % passing reading and % overall passing columns decreased by 0.1% - 0.3%. interestingly the average reading score column increased by 0.05%. The changes can be seen here:

![original top schools][]

![new top schools][]

### - How does replacing the ninth-grade scores affect the following:

#### - Math and reading scores by grade

In this section all that changes is that the ninth grade does not have a value in the new data. Because we replaced it with NaN and only the ninth graders, all the other grades remained the same. This is true for both the reading and math scores by grade.

![ original\_math\_scores\_by\_grade.png][]

![ original\_reading\_scores\_by\_grade.png][]

![ new\_math\_scores\_by\_grade.png][]

![ new\_reading\_scores\_by\_grade.png][]

#### - Scores by school spending

In this data, since Thomas High School falls in the 3rd bin ($631 - $645), we would look in this row to see any changes. Here there are no changes. So it did not affect this at all. the images are shown below:

![ original\_scores\_by\_school\_spending.png][]

![ new\_scores\_by\_school\_spending.png][]

#### - Scores by school size

In this data, Thomas High school fell in the medium (1000 – 1999) range. So we look at this row. And we see compared to the original, there is no change. The images are shown below:

![ original\_scores\_by\_school\_type.png][]

![ new\_scores\_by\_school\_type.png][]

#### - Scores by school type

In this data Thomas High School is a charter school so we can look at that row for changes. And again we see no changes. Images are shown below:

![ original\_scores\_by\_school\_type.png][]

![ new\_scores\_by\_school\_type.png][]

## Summary

The first most obvious change we see is the new total student count since we do not count for the students in 9th grade that have been replaced. The second change noticed was in the district summary with mainly the percentages being changed. Here the new data had decreased by a little. The third change noticed was in the school summary which had a massive change ranging in the 30% region. This actually improved the score of Thomas High Schools percentage columns. The fourth change I observed is in the top 5 schools data. Although Thomas high school did not move up or down in the list, it did show slight changes in the percentages columns.