#School District Analysis Using Pandas

##Overview

We’re using Python/Pandas libraries to analyze a dataset of student information within a school district. It includes reading/math scores, school name/district, test scores, etc. for about 40,000 students. This analysis involves refactoring Python code to accommodate to potential tampering within the dataset, and how it may affect the final results. The original analysis is kept/accounted for, and the code is once again ran through with suspicious aspects counted out to see if they had any affect.

##Suspected Altering of Data:

The reading/math scores for the 9th grade students at Thomas High School are suspected to have been changed. The reasons are unknown and conclusions have yet to be drawn, so for this analysis this part of the dataset has been counted out, so we can determine the effects relative to the other grades in that school.

#####Results:

The isolation/removal of the potentially tampered data is best looked at when we gather math/reading scores by grade including the “tampered” data, as well as after.

* Initially we find that Thomas high school has a passing percentage of ≈ 65%, where once refined that number changes to ≈ 90%. This is seen in numerous figures throughout the code, but is best displayed here:

\*\*Before 9th Grade Removal:\*\*

| **School Type** | **Total Students** | **Total School Budget** | **Per Student Budget** | | **Average Math Score** | | **Average Reading Score** | | **% Passing Math** | | **% Passing Reading** | **% Overall Passing** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Thomas High School** | Charter | 1635 | $1,043,130.00 | $638.00 | | 83.350937 | | 83.896082 | | 66.911315 | 69.663609 | 65.076453 |

| **School Type** | **Total Students** | **Total School Budget** | **Per Student Budget** | **Average Math Score** | **Average Reading Score** | **% Passing Math** | **% Passing Reading** | **% Overall Passing** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Thomas High School** | Charter | 1635 | $1,043,130.00 | $638.00 | 83.350937 | 83.896082 | 93.185690 | 97.018739 | 90.630324 |

Initial filtering of information was done by sorting by number of students, budget per student, and even type of school, but this information is not currently of use to us. As it’s beyond the point of the initial analysis. Even comparing the initial overall district passing percent of 64.9% to the performance of other schools excluding Thomas high, we can tell something is not right.

##Summary

What we find is that the removal of this small aspect of the dataset has a significant affect on the placing of the school, and the main thing that can be taken away is that whether this is a result of a mistake, corruption, immoral action, etc. this suspicion warrants further investigation. By sorting school’s performances by grade, budget, district type, etc. we now have started to gather information to have at our disposal, filtered out when needed.