**Rutgers Data Science Boot Camp**

**Pandas**

**Homework Assignment 04**

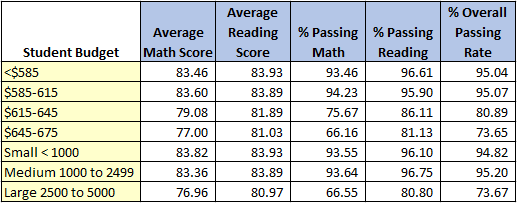
**Mark Visco**

**Assignment**

This assignment is to analyze district-wide standardized test results. The information is stored on two different csv files.

**Observations**

There were two significant observations which stood out once the entire analysis had been completed. For the first observation I combined two similar types of analyses together; the cost per student and the school size. One would think that spending more money per student would produce greater results, but the opposite was true in this case. Schools with a lower cost per student budget performed better than schools with larger cost per student basis. Another assumption would be that larger schools with larger budgets would have the resources to ensure their students would perform better over smaller schools with smaller budgets but that was also not the case. The larger the school the lower the overall scores in math and reading. This could be because smaller schools may have a better student to teacher ratio but because this data was not provided I am not able to draw that conclusion.



The other observation involved the reading and math scores by high school grade and the type of school (Charter or District). In this project, the data output was in the form of tables, but I thought putting the tables into a bar graph would be easier to view and would immediately point out any differences. As you can see from the two graphs, the first shows the reading scores and the second shows the math scores, across the board all Charter schools outperformed the District schools in both reading and math. Reading scores for both the Charter and District schools were slightly better than math scores.

The data available didn’t have any other pieces of information that would help explain why this was the case where every charter school performed better in both subjects. The other observation is that all of the charter schools performed at the same level and all of the district schools, while lower also performed at the same level. There weren’t any significant outliers.

Additional information such as gender, ethnic breakdown or age of the students in each of the classes might help provide additional insight into why these numbers landed where they did in the graph at this level. Slicing the data further might uncover these reasons or raise additional questions.

