The data is from NYC Open Data . <https://data.cityofnewyork.us/api/views/f9bf-2cp4/rows.csv?accessType=DOWNLOAD>

The data is about SAT scores from graduating seniors of 2012 at NYC schools. Records contain 2012 College-bound seniors mean SAT scores taken during SY 2012. In the data set, you will see school names with unique code , DBN and number of student takers and average of maths, reading and writing scores at each school. I use python to parse and clean up data such as removing headers, clean up the missing rows and change the strings in school names such as “,” to “-” so that importing to sql from CSV format will not be an issue.

Looking at data, Brooklyn Technical High School ranks with highest number of students who took SAT (1277) with the highest sum of average total score of 1905 in Brooklyn district. Although number of student at Stuyvesant High School is lower (832 students) than those from Brooklyn school, the total average score is higher with 2149.Stuyvesant ranks the highest in terms of average total SAT scores in the whole New York City, followed by Bronx high school for medical science.

Looking at stats, we can see that school like Multicultural high school has only 29 students of exam takers and lowest total average score with 923. We can also see that majority of lowest number of SAT takers with lowest total average SAT scores are not from schools at Manhattan area. This also points out that number of SAT takers somehow relate to total average score of students. School like Stuyvesant and Brooklyn high schools have above 800 test takers and average score of 1905.

Although the data is reliable since it is from credible New York City Open data source, it would be much better to make assumptions if more information related to different factors such as lowest and highest score in each school (so that we can see the outliers) or their family’s income or may be some type of quality evaluator from individual school.

I think that the data follows my assumption in terms of population and highest scores relation to the geographical location.