

Identify New York Public High Schools with the Highest SAT Test Scores

Tanya Hall
February 04, 2020

1. Introduction

1.1 Background

The purpose of this project is to identify which public high school's within the five boroughs that make up the New York City area have the highest overall SAT scores, while isolating the public high school's that received the highest score for each section (math, reading, and writing). Then locate the closest vendors to the top ten performing SAT test score public high school for insight on the surrounding area and further planning for families and students.

This report is targeted to people who are looking to relocate or trying to apply to New York City area public high schools with the highest performing SAT scores.

1.2 Problem

The SAT is an entrance exam used by most colleges and universities to make admissions decisions. The SAT is a multiple-choice, pencil-and-paper test created and administered by the College Board and consists of three sections; math, reading, and writing.

The purpose of the SAT is to measure a high school student's readiness for college, and provide colleges with one common data point that can be used to compare all applicants. College admissions officers will review standardized test scores alongside a candidate's high school GPA, classes the candidate took in high school, letters of recommendation from teachers or mentors, extracurricular activities, admissions interviews, and personal essays. Due to the competitive nature of college admission the SAT within the United States, the SAT test scores have become an important data point for all college candidates. Additionally, SAT test scores are utilized for funding college decisions/options. Generally, the higher the SAT score a student achieves the more options for funding college is available through grants, financial aid loans, and work-study programs.

1.3 Interest

Normally, high school students take the SAT test during the spring of their junior year or fall of their senior year. Recognizing that admissions to college for high school's students also start during this time frame, there is not a lot of time to re-take the test if a higher SAT score is needed or wanted. Further warranting that students choose the best public high school for them to learn the necessary skills to achieve a high SAT test score.

2. Data acquisition and cleaning

While there are many different types of data, almost all the data for this study was procured through excel spreadsheets and website scrapping. Considering these data types, pre-processing can be articulated as a set of incremental stages. Beginning with the ingestion of the large data sets into Jupiter Notebooks utilizing the python programming language. The next stage includes cleaning and transforming the data. This is completed by removing invalid records, columns, and translating all the columns to use a normalized set of values. Once the base set of data had been identified for this problem, data acquisition from a website was added. With the new website data extracted, a combination of two different datasets into a single table was achieved.

2.1 Data Sources

The SAT score averages that have been provided by the College Board in each of the five boroughs creating the New York City area during the years 2014-2015. The high school data that has been compiled and published by the New York City Department of Education. The location of venues around the identified public high schools was provided through the use of a Foursquare Application Programming Interface (API)-venues.

2.2 Data Cleaning

Data cleaning was first completed through the preprocessing the average SAT 2014-2015 school year SAT test scores data from Kaggle. This data set was normalized by dropping unused data and fixing

any inconsistencies within the columns, and row values. From the cleaned data a data frame illustrating the average SAT scores for every accredited high school in the New York City, five borough area was crafted. Next data from the Foursquare website was ingested and cleaned. Lastly, a combined data frame was generated with all entries normalized so that the acquire data did not exhibit any inconsistencies through column names or inconsistent location information.

2.3 Feature selection

The data was presented through tables and geographical maps to highlight the location of identified public high schools. The construction of the geographical map of the public high schools with the highest average total SAT test scores and three main sections of the test within the five boroughs area of New York City was done by utilizing the folium library within the python programming language.

3. Exploratory Data Analysis

Identify the closest vendors to the public high schools with the highest SAT test scores with the five borough New York City area, via FourSquare API and geocoding was used as exploratory data points. Though the visualization of the areas of the five borough that possess the highest public high schools with the average total SAT test scores, and closest surrounding vendor's families and students could locate relevant places for transportation or afterschool gathering locations.

3.1 Calculation of target variable

To assist families and students locate public high schools with the highest average SAT test scores and the highest specific sections of the SAT test, the top ten public high schools were identified. The top ten public high schools were displayed in a table format, then visualized in a geographical map. Following the list of the top ten public high schools, a clustered list of the closest vendor to each school was crafted.

3.2 Relationship between average test scores

Stuyvesant High School located in Manhattan was the top performing average SAT test score in all sections of the SAT test. The same ten public high schools were listed for each section of test and subsequently, had the highest total average test scores as well. Those schools are: Stuyvesant High School, Staten Island Technical High School, Bronx High School of Science, High School of American Studies at Lehman College, Townsend Harris High School, Queens High School for the Sciences at York College, Bard High School Early College, Brooklyn Technical High School, Eleanor Roosevelt High School, High School for Mathematics, Science, and Engineering at City College.

4. Conclusions

The aim of this project is to assist families and students identify public high schools within the New York City, five boroughs area with the highest average SAT test scores. SAT test scores are used help colleges interpret students' overall academic performance in relation to the national applicant pool. Through this data families and students are provided with the necessary information to make logical and concise choices when choosing a public high school located within the New York City, five boroughs area. For families and students that are planning for college admission after high school, the public high schools with the total highest test scores are of interest. Additionally, if a student is weak in one section of the test, a public high school that has high average test scores within this academic discipline is of interest as well. Lastly, identifying the closest local vendors to the highest total average test score public high school assist family with planning for after school activities, pick-up/drop-off locations, and an overall general idea of what is within walking distance of the school.

5. Further directions

This project aids families and students alike with a better understanding of public high schools with high academic achievements with respect to the SAT test scores. Venues near the public high schools also allow families and students to find local, walking distance places near the public high schools for planning purposes and neighborhood associations. Recognizing that this project was singularly focused on SAT test scores, future projects could use this data as a platform for more in depth analysis of public high schools located within the New York City, five boroughs area.