Project Writeup - Best Schools Of Nyc

Objectives

Below is the list of questions for NYC region that we are looking to answer as a final outcome of this project:

- Is Scholastic Aptitude Test (SAT) a fair test in the diverse demographic region like New York City?
- How SAT score vary with race, income, gender, socioeconomic background and other wide array of factors?
- What are the top 10 best schools of NYC based on the average SAT score?
- What are the least expensive NYC neighborhoods with top 10 best schools?
- Which schools consist more than 60% of female students proportion and have average SAT score higher than 1700?
- Which are the schools with more than 95% and less than 10% of the Hispanic students?
- Which schools have low total enrollment and have low average SAT score?

Goal Significance

Why do we need to determine whether SAT is a fair test or not? What advantage we could have by concluding upon the SAT test fairness with evidence? Below are the goals we can enlist:

- This information will give us an idea about:
    - How various factors play role in student's performance in SAT score
    - How diverse the student mix within various NYC neighborhood?
    - How effectively different players of this game, i.e. teachers, parents, students, test system etc., contribute per their roles and responsibility
    - The school regions with strong/weak performances
    - Role of ethnicity, gender, income status and community environment over the test performance
- Based on the results, we could conclude whether:
    - How fair is the SAT system?
    - Which are the factors need to be addressed in case any flaw in the current system is identified?
    - The areas of NYC that should be given more attention in the event of future policy amendment.

Data

Data Source

New York City public data for student SAT score by high school and other relevant demographics is available at:
https://data.cityofnewyork.us/Education/2012-SAT-Results/f9bf-2cp4

The details of the property values in five distinct NYC suburbs is available in the .csv files listed below. This data is available for download at:
http://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page

Data Lists

The data is available in the .csv file format. The detailed file data information is available in the referred links of the file name description.

Project Writeup - Best Schools Of Nyc

`ap_2010.csv`   : Data on [AP Test Results](#))
`class_size.csv`   : Data on [class size](#)
`demographics.csv`   : Data on [demographics](#)
`graduation.csv`   : Data on [graduation outcomes](#)
`hs_directory.csv`   : A directory of [high schools](#)
`sat_results.csv`   : Data on [SAT scores](#)
`survey_all.txt`   : Data on [surveys](#) from all schools
`survey_75.txt`   : Data on [surveys](#) from New York City [district 75](#)

The raw property data for NYC is cleaned based on the most realistic per sq. ft. property rates in each of the five regions. In order to maintain trade-off between the computing and data diversity, the size of each data set is restricted to around 12000 entries.

For more details, see the files listed below:

`BK2017V11.csv`   : Sales Details for Properties in Brooklyn
`BX2017V11.csv`   : Sales Details for Properties in Bronx
`MN2017V11.csv`   : Sales Details for Properties in Manhattan
`QN2017V11.csv`   : Sales Details for Properties in Queen
`SI2017V11.csv`   : Sales Details for Properties in Staten Island

Data Details

SAT scores by school - SAT scores for each high school in New York City

School attendance - Attendance information for each school in New York City

Class size - Information on class size for each school

AP test results - Advanced Placement (AP) exam results for each high school (passing an optional AP exam in a particular subject can earn a student college credit in that subject)

Graduation outcomes - The percentage of students who graduated, and other outcome information

Demographics - Demographic information for each school

School survey - Surveys of parents, teachers, and students at each school

Highlights of the code.

Software packages used:

- Python
- Pandas
- Numpy
- Excel
- re
- random
- Matplotlib.pyplot
- Matplotlib.colors

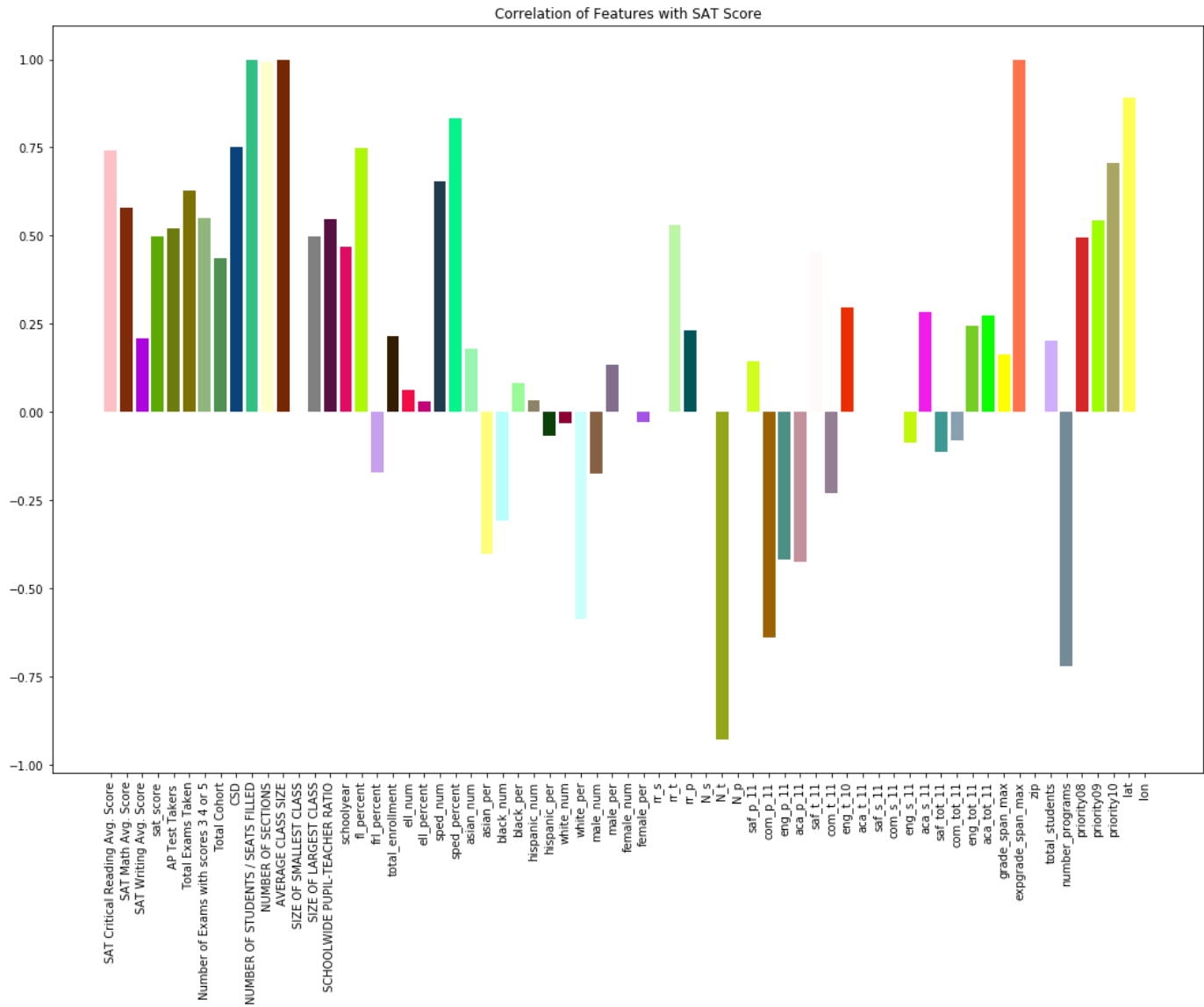# Project Writeup - Best Schools Of Nyc

Overview:

- Import files of different formats (.csv, .txt) and combine them to consolidate the data frame
- Understand the data and its relevance to the ultimate project goal
- Identify list of features and target feature for data analysis
- Combine various fields in the scattered data files and generate primary keys for database merge
- Format alternation for data frame column info
- Extract latitude and longitude of NYC schools and map them in basemap
- Condense various datasets to focus onto NYC High School Students SAT score info
- Clean the dataset for unwanted info and NaN values
- Review features correlation with the target column and visualize it through chart
- Various features verified for correlation with SAT score are:
    - Safety and Respect score based on student responses
    - Racial percentage
    - Hispanic proportion
    - Gender
    - AP Test Takers
    - Class size
- List schools with:
    - More than 95% Hispanic students
    - Less than 10% Hispanic Students
    - Low total enrollment and low SAT score
    - Higher SAT score and high female percentage
    - SAT score more than 1700 & female percent higher than 60%
- Correlation between Parent, Teacher and Student Responses to survey
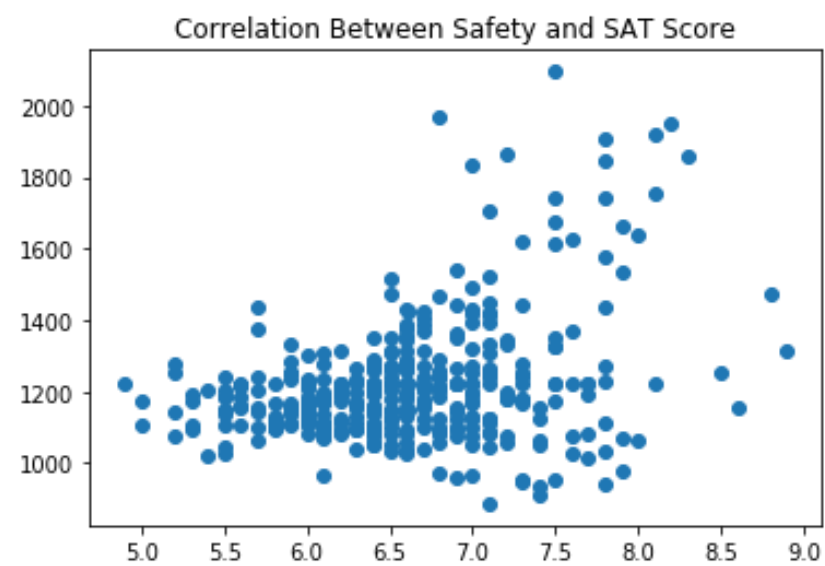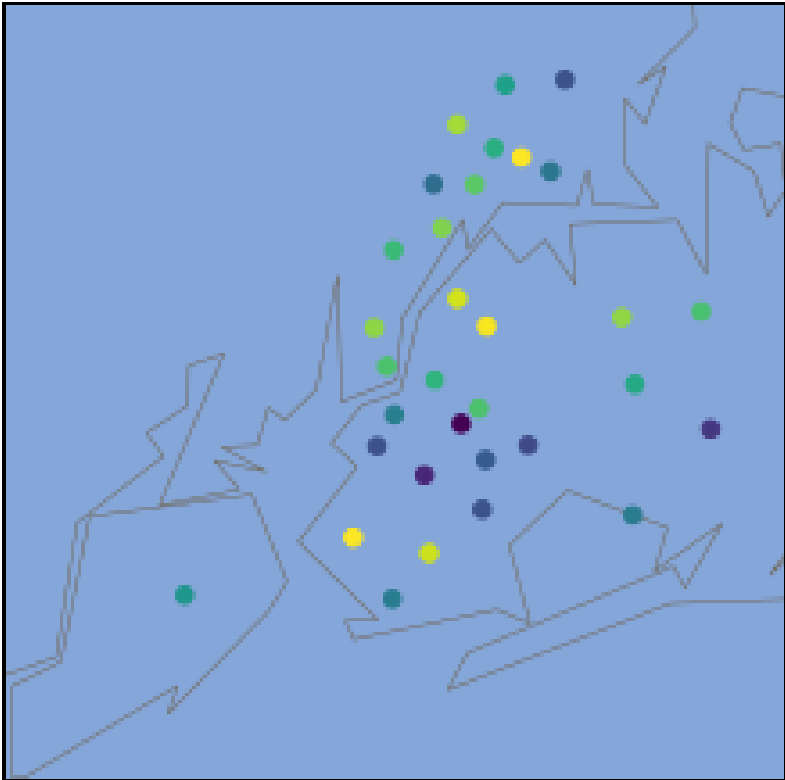- Find out the least expensive NYC neighborhoods for top 10 schools

Results

Visualize the results.

Correlation of Features with SAT Score

Project Writeup - Best Schools Of Nyc

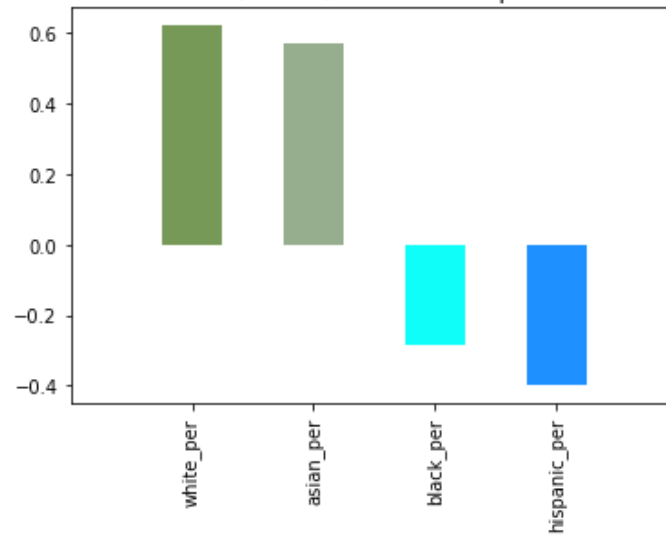### Correlation Between Safety and SAT Score
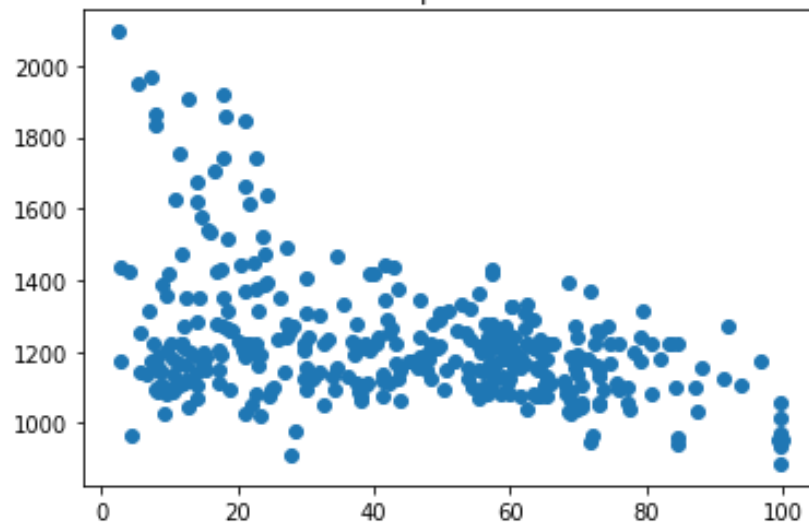

Map Safety Score by Districts
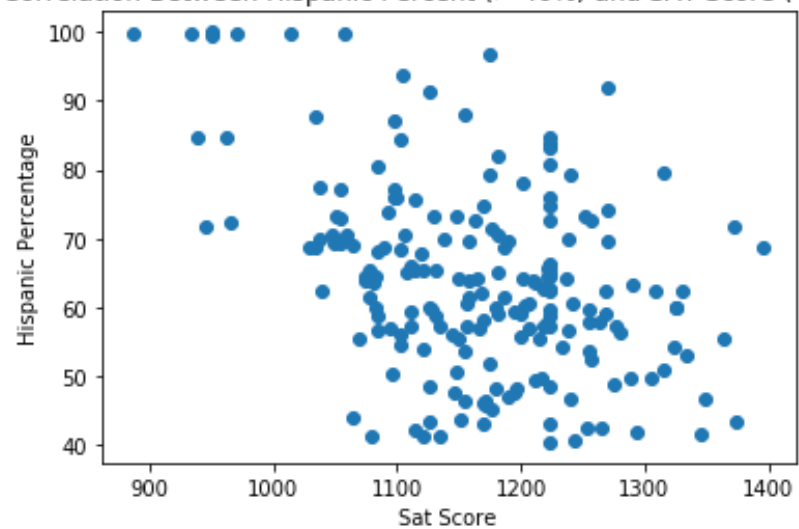
Project Writeup - Best Schools Of Nyc

Correlation of Whites, Asians, Blacks and Hispanics with SAT Score
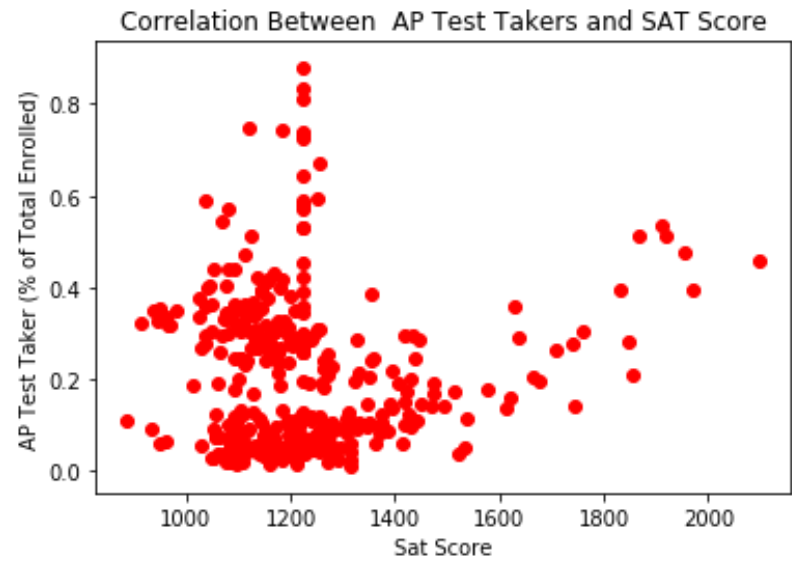

Correlation Between Hispanic Percent and SAT Score

Correlation Between Hispanic Percent (> 40%) and SAT Score (< 1400)

Project Writeup - Best Schools Of Nyc
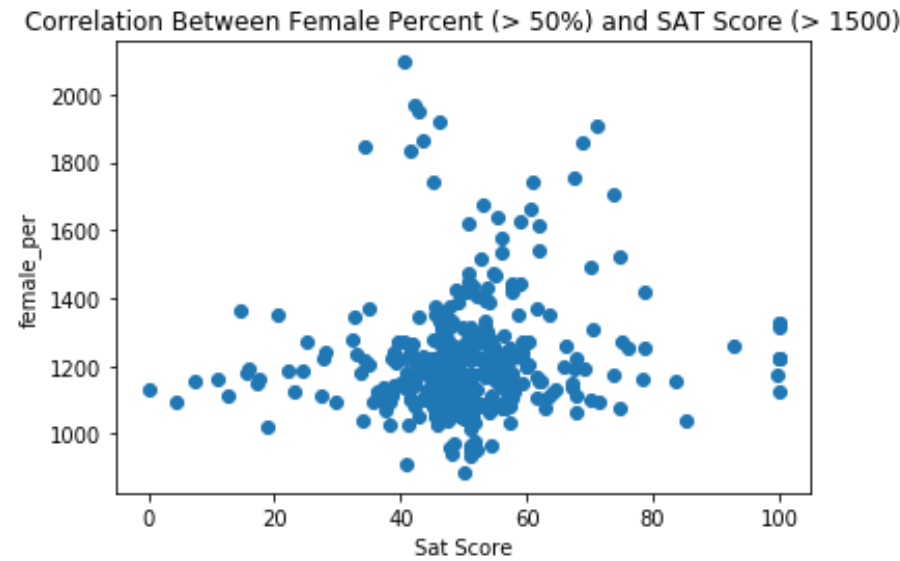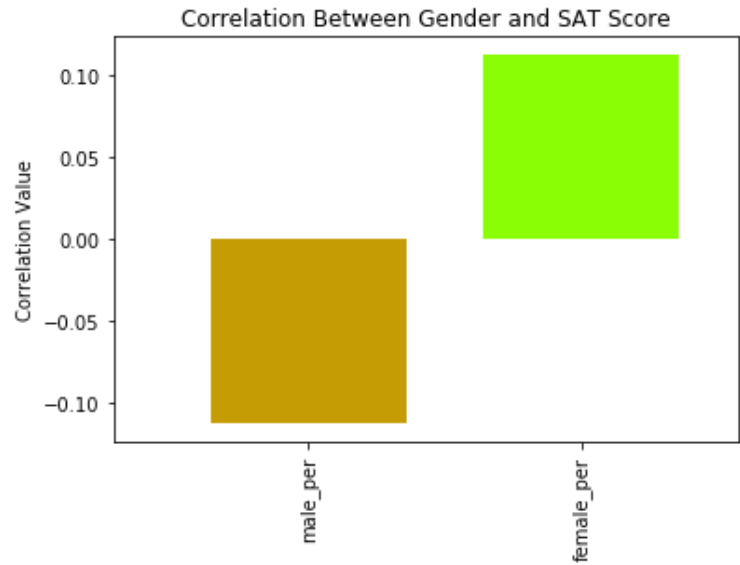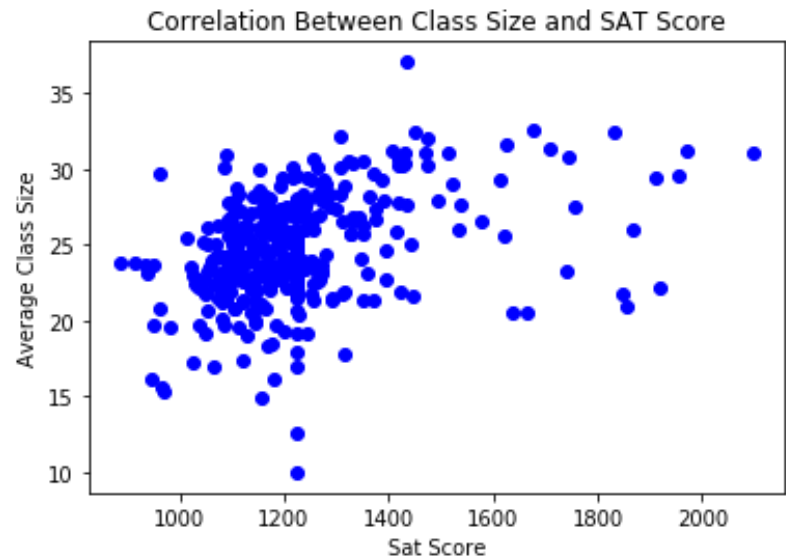
Correlation Between Gender and SAT Score

Correlation Between Female Percent (> 50%) and SAT Score (> 1500)

Correlation Between AP Test Takers and SAT Score

Project Writeup - Best Schools Of Nyc

Correlation Between Class Size and SAT Score

Correlation for Difference between Parent-Student Survey Response and SAT Score
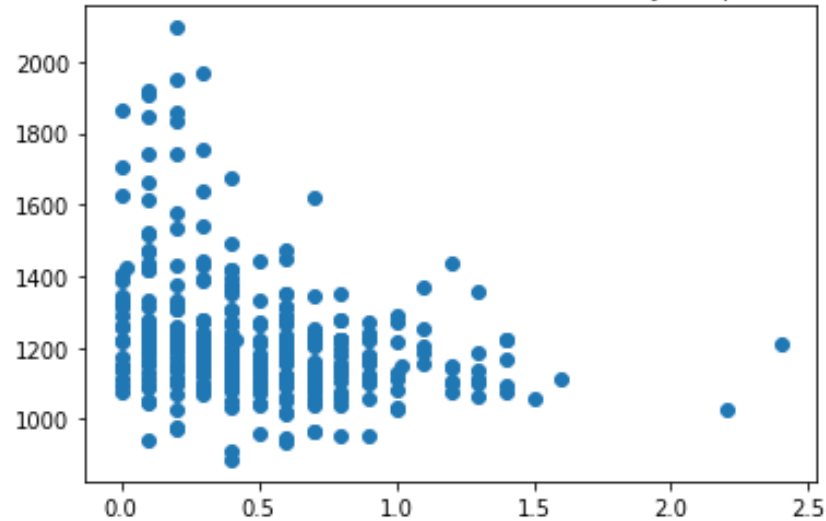
Project Writeup - Best Schools Of Nyc

Correlation for Difference between Parent-Teacher Survey Response and SAT Score

Correlation for Difference between Teacher-Student Survey Response and SAT Score

Project Writeup - Best Schools Of Nyc

Tabular Results

Top 10 Best Schools in New York City

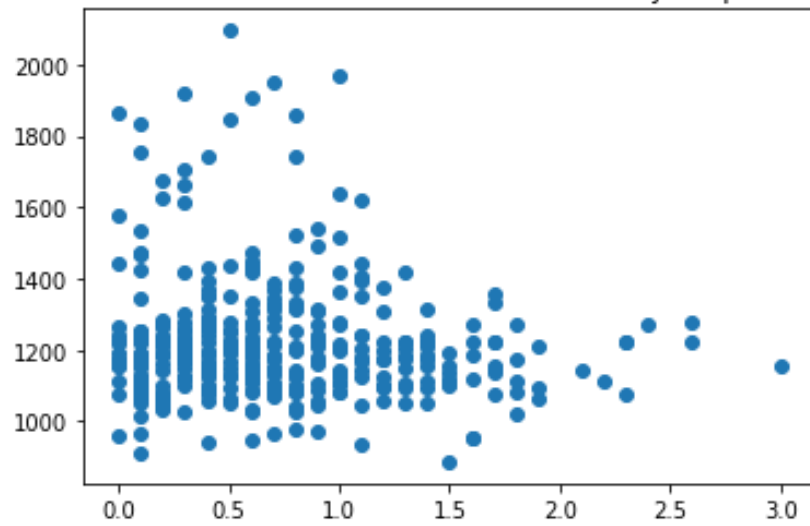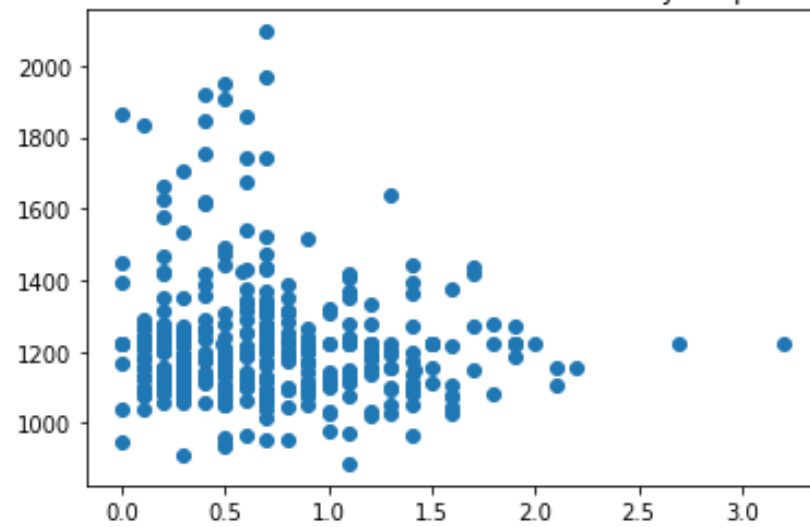| boro	sat_score	SCHOOL NAME	CSD	zip
Manhattan	2096.0	STUYVESANT HIGH SCHOOL	2	10282
Bronx	1969.0	BRONX HIGH SCHOOL OF SCIENCE	10	10468
Staten Island	1953.0	STATEN ISLAND TECHNICAL HIGH SCHOOL	31	10306
Bronx	1920.0	HIGH SCHOOL OF AMERICAN STUDIES AT LEHMAN COLLEGE	10	10468
Queens	1910.0	TOWNSEND HARRIS HIGH SCHOOL	25	11367
Queens	1868.0	QUEENS HIGH SCHOOL FOR THE SCIENCES AT YORK CO...	28	11433
Manhattan	1856.0	BARD HIGH SCHOOL EARLY COLLEGE	1	10002
Manhattan	1847.0	HIGH SCHOOL FOR MATHEMATICS, SCIENCE AND ENGIN...	5	10031
Brooklyn	1833.0	BROOKLYN TECHNICAL HIGH SCHOOL	13	11217
Manhattan	1758.0	ELEANOR ROOSEVELT HIGH SCHOOL	2	10021

The Least Expensive NYC Neighborhoods with Top 10 Best Schools

Borough	Property Address	Owner	Price/Sq Ft	Neighborhood School
Manhattan	201 WARREN STREET	TRIBECA NORTH END LLC	112.479	STUYVESANT HIGH SCHOOL
Manhattan	399 CHAMBERS STREET	TRIBECA POINTE LLC	114.277	STUYVESANT HIGH SCHOOL
Bronx	2450 DAVIDSON AVENUE	2460 DAVIDSON REALTY	51.4895	BRONX HIGH SCHOOL OF SCIENCE
Bronx	2776 SEDGWICK AVENUE	TINEO, JOSE RAMON	51.8865	BRONX HIGH SCHOOL OF SCIENCE
Staten Island	10 PEEL PLACE	ANN SOLAZZO	24.215	STATEN ISLAND TECHNICAL HIGH SCHOOL
Staten Island	41 PETER AVENUE	SCAROLA, GREGG M	24.2163	STATEN ISLAND TECHNICAL HIGH SCHOOL
Bronx	2450 DAVIDSON AVENUE	2460 DAVIDSON REALTY	51.4895	HIGH SCHOOL OF AMERICAN STUDIES AT LEHMAN COLLEGE
Bronx	2776 SEDGWICK AVENUE	TINEO, JOSE RAMON	51.8865	HIGH SCHOOL OF AMERICAN STUDIES AT LEHMAN COLLEGE
Queens	152-11 UNION TURNPIKE	ST. JOHN'S UNIVERSITY	50.8769	TOWNSEND HARRIS HIGH SCHOOL
Queens	154-27 64 AVENUE	WISCHHUSEN RICHARD	66.6693	TOWNSEND HARRIS HIGH SCHOOL
Queens	162-25 112 ROAD	CALVARY GRANDPARENT R	59.1462	QUEENS HIGH SCHOOL FOR THE SCIENCES AT YORK CO...
Manhattan	85 STANTON STREET	85 OWNERS CORP	97.2278	BARD HIGH SCHOOL EARLY COLLEGE
Manhattan	138 LUDLOW STREET	TRIPUKA REALTY CORP	97.401	BARD HIGH SCHOOL EARLY COLLEGE
Manhattan	No Resi. Property Found	N/A	N/A	HIGH SCHOOL FOR MATHEMATICS, SCIENCE AND ENGIN....
Brooklyn	220 LINCOLN PLACE	220 LINCOLN LLC	50.0592	BROOKLYN TECHNICAL HIGH SCHOOL
Brooklyn	233 BERKELEY PLACE	233 BERKELEY PLACE LL	50.2336	BROOKLYN TECHNICAL HIGH SCHOOL
Manhattan	230 EAST 71 STREET	230 OWNERS CORP	97.0392	ELEANOR ROOSEVELT HIGH SCHOOL
Manhattan	305 EAST 75 STREET	GRUMA REALTY CORP	97.0882	ELEANOR ROOSEVELT HIGH SCHOOL

Schools with More Than 60% Female & Greater Than 1700 SAT Score

```
Schools with SAT > 1700 & Female > 60% =
 ['BARD HIGH SCHOOL EARLY COLLEGE' 'ELEANOR ROOSEVELT HIGH SCHOOL'
 'BEACON HIGH SCHOOL'
 'FIORELLO H. LAGUARDIA HIGH SCHOOL OF MUSIC & ART AND PERFORMING ARTS'
 'TOWNSEND HARRIS HIGH SCHOOL']
```

Project Writeup - Best Schools Of Nyc

Schools with More Than 95% Hispanic Students

```
['MANHATTAN BRIDGES HIGH SCHOOL'
 'WASHINGTON HEIGHTS EXPEDITIONARY LEARNING SCHOOL'
 'GREGORIO LUPERON HIGH SCHOOL FOR SCIENCE AND MATHEMATICS'
 'ACADEMY FOR LANGUAGE AND TECHNOLOGY'
 'INTERNATIONAL SCHOOL FOR LIBERAL ARTS'
 'PAN AMERICAN INTERNATIONAL HIGH SCHOOL AT MONROE'
 'MULTICULTURAL HIGH SCHOOL' 'PAN AMERICAN INTERNATIONAL HIGH SCHOOL']
```

Schools with Less Than 10% Hispanic Students

```
['STUYVESANT HIGH SCHOOL' 'BRONX HIGH SCHOOL OF SCIENCE'
 'BROOKLYN TECHNICAL HIGH SCHOOL'
 'QUEENS HIGH SCHOOL FOR THE SCIENCES AT YORK COLLEGE'
 'STATEN ISLAND TECHNICAL HIGH SCHOOL']
```

Schools with Low Total Enrollment for SAT & have Low Avg. SAT Score

```
        INTERNATIONAL COMMUNITY HIGH SCHOOL
             BRONX INTERNATIONAL HIGH SCHOOL
       KINGSBRIDGE INTERNATIONAL HIGH SCHOOL
       INTERNATIONAL SCHOOL FOR LIBERAL ARTS
             HIGH SCHOOL OF WORLD CULTURES
          BROOKLYN INTERNATIONAL HIGH SCHOOL
    INTERNATIONAL HIGH SCHOOL AT PROSPECT
                  IT TAKES A VILLAGE ACADEMY
                    MULTICULTURAL HIGH SCHOOL
    PAN AMERICAN INTERNATIONAL HIGH SCHOO
```

Explain the results in a simple, concise and easy way. (non-technical audience)

The analysis and results give very useful info to shape our future planning based on:

- Student mix based on certain gender and/or ethnicity in various neighborhood
- Significance of various factors such as safety, racial mix, gender, female proportion, Advance Placement Testing system, classroom size, parents-student-teachers expectations for academia etc. over the SAT score
- School regions with strong/moderate/weak SAT performance

Explain the results in the most technical way. (technical, data-scientist audience)

- We observe high correlations between SAT score and following factors:
    - Number of Students / Seats Filled
    - Average Class Size
    - The highest grade the school expects to serve eventually
    - Number of Teacher Respondents (-ve)
    - SAT Critical Reading Average Score
    - Free Lunch Percent
    - Number of distinct programs available at the school (-ve)

  It's obvious to see how the 'class-size', 'number of students' and the 'highest grade the school expects' creates a competitive but cohesive social environment for the students and encourage them working hard to perform better in SAT.

Project Writeup - Best Schools Of Nyc

The strong positive relationship of 'SAT Critical Reading Average Score' with the SAT score shows the impact of effective reading skills on the overall SAT performance.

The positive correlation of 'Free Lunch' with SAT score tells the success story of government's welfare program in the area of economically developing neighborhood and low-income immigrant regions.

It is interesting to note that higher the number of teachers' participation in survey for expectation about student's academia negatively influence the overall student performance in SAT.

The negative correlation of 'Number of distinct programs available at the school' with SAT score goes in reverse parity with the positive correlation of 'class-size' and 'number of students' factors above as more program diversity tends to reduce the class-size and number of students per course.

- Correlation: Safety & SAT Score:

There appears to be a positive correlation between safety and SAT score though not so strong. There are some schools that have high safety standard and achieve high SAT score and some schools with low safety standard got low SAT score. Majority of the school from the sample falls between the safety score of 6.0 to 7.5 with an average SAT score below 1500.

- Safety Map

Brooklyn seems to have better safety score compared to the same for the parts of Manhattan, Queens and Bronx region.

- Racial Correlation:

The whites and the Asians are found having strong positive correlation while the Blacks and the Hispanic are found having moderate negative correlation with the SAT score.

This is due to the fact that the Hispanic and the Blacks might be coming from the immigrant families. The socio-economic factors, family support and surrounding environment play significant role in overall performance of the students.

- Correlation between Hispanic Proportion and SAT score:

There seems to be a moderately strong downward trend with the increase of Hispanic Population. Very small portion of population attains a SAT score of 1800 or more while majority falls under the score of 1500. Majority of the schools with Hispanic population in the range of 40% - 80% get average SAT score of around 1150. The results become clearly evident when plotted with 40% or more Hispanic Proportion for SAT score of 1400 or less.

Project Writeup - Best Schools Of Nyc

This is mainly due to the immigrant population along with the socio-economic and demographic factors.

- Gender Correlation:

Female show very weak positive and male show very weak negative correlation for SAT score. This shows the relative performance of female and male students – which is more or less the same.

- Relationship with Female Percentage:

There exists moderately positive correlation between the female proportion and SAT score. While tested for pool of students having more than 50% female proportion and SAT score of greater than 1500, it is observed that around 40% - 65% of the total students attain SAT score in the range of 1000 – 1350.

- Relationship with AP Test Takers:

No correlation is found between the AP Test Takers and the SAT score. In fact, up to 40% of the total enrollment who took AP Test scored less than 1400 in SAT. There are few percent of students who took AP Test and secured SAT score of 1800 – 2000.

The data about students' performance in AP Test Vs. SAT could give us more insight to further comment about the AP Test results predictive relevance for SAT score.

- Correlation with the Class Size:

Moderately strong positive correlation has been observed between the class size and SAT score. Students' SAT score slowly improves as the class size increases from 20 to 30 students. Majority of the students score between 1000 – 1400 irrespective of class size. However, there are few exceptional cases where the students score beyond 1800.

This clearly shows that performance of students in SAT is strongly impacted by his/her class size, higher student-teacher ratio and consequently the dense social environment - which collectively proves to be encouraging for the students.

- Correlation for Difference between Parent-Student Survey Response:

It appears that there is a moderately strong negative correlation between the difference in Parent-Student Survey Response and SAT Score. Hence, more similarity of parents and student's opinion about the latter's academic expectation, better would it be reflected in to the higher SAT score. In other words, more the disparity between parents and student's responses for students' academic expectation score, lower would be the SAT score.

This clearly indicates that parental involvement in student's education results into a better SAT performance and vice versa.

Project Writeup - Best Schools Of Nyc

- Correlation for Difference between Parent-Teacher Survey Response:

  There exists weak negative relationship between the difference in parents and teachers' survey response and SAT score. This indicates that difference of opinion between the parents and the teachers over the expectation of student's academia has moderately reverse impact over student's performance in SAT. In other words, less the differences between the parental and teacher's opinion about student's academic expectation, moderately better would be student's performance in SAT.

- Correlation for Difference between Teacher-Student Survey Response:

  No correlation is found between the difference in teachers' and students' survey response and SAT score. This indicates that difference of opinion between the teachers and students over the academic expectation for the students have no impact over student's performance in SAT.

- Following information is deduced based on the analysis to answer our certain open-ended questions:
  - List of Schools with Hispanic More than 95%
  - List of Schools with Hispanic Less than 10%
  - List of Schools with Low Total Enrollment and Low Sat Score
  - List of Schools with High SAT score & High Female Percentage
  - List of School with more than 60% Female Percentage and SAT score higher than 1700
  - Top 10 Best Schools NYC based on SAT Score
  - List of Great Neighborhood (Least Expensive) in NYC for Top 10 Best Schools

Conclusion

What we learn from this outcome. (non-technical audience)

- We understand the impact of various demographic, socio-economic, and current credit and grading system related factors over the performance of students in SAT examination. This helps us to determine and verify how this test maintains balance amongst these factors for overall fairness of test within the widely diverse community.
- We are able to determine top 10 best schools based on their average SAT score.
- We shortlist the least expensive but situated with the top 10 best school neighborhood regions.
- We could identify the regions of NYC neighborhood where schools do not perform well and understand whether or not there exists any ethnic diversity [to consider] for policy improvement.
- We learn the impact of similarity and disparity of students/teachers/parents' opinion towards student's academia over the SAT performance. This highlights the roles and responsibilities of various players of the game and its relative significance for the final outcome.

Project Writeup - Best Schools Of Nyc

Technical significance of the results. (technical, data-science audience)

- We are able to know various critical factors that do/do not affect students' performance in SAT examination. Based on this info, we can verify and validate how fare the current SAT system is for the diverse demographics like NYC. In case we found any flaws in the SAT system for unfair testing performances, we can get better idea about the causes to focus on and fix the system.

- We found the factor significance as summarized below:

Factor with Positive Impact on SAT Score (Severity)	Factors with Negative Impact on SAT Score (Severity)	Factors with No Impact on SAT Score
Class Size (Moderately Strong)	Difference Between Parent-Teacher Survey Response (Weak)	Difference Between Teacher-Student Survey Response
Race: The Whites & The Asians (Strong)	Race: The Blacks and The Hispanic (Moderately Strong)	AP Test Takers
Female Proportion (Moderately Strong)	Difference Between Parent-Student Survey Response (Moderately Strong)	
Gender – Female (Weak)	Gender – Male (Weak)	
Safety (Weak)	Hispanic Percentage (Moderate)	

- Based on the correlations and its severity on students' SAT score, we can mention that SAT is reasonably fair for the diverse NYC demographics.
- We got answers to all of our open-ended questions enumerated in the objectives of the project. (See Results above for details)

Suggestion for Further Development

How differently you would have done this project, if get a chance to do it again. Why?

In case of gotten second chance to do this project again, I would have gathered and used some more features such as:

- Parents employment status
- Average annual family income
- Present residence status: Own/Rented, House/Apt.
- Family size
- Number of siblings (elder/younger)
- Details of student's extracurricular activities
- Participated in free / free and reduced lunch program
- Student's total years of enrollment with the same school
- Past scholarship record

Project Writeup - Best Schools Of Nyc

to determine each of these factors correlation with SAT score to further deepen our understanding about the fairness of the SAT system.

Your suggestions for someone if s/he wants to further continue this work.

Someone could pick one or more of the untouched data fields mentioned above, explore and gather additional features and continue this journey further to cherish more flavors in the analysis.