What is the Impact of Economic Need and School Income on the Math and ELA proficiency levels of New York City students as a whole and by Race?

Eco225 Project

Oliver Keyes-Krysakowski

UTORid: krysakow

April 13th, 2022

Introduction

Education is a critical component of a child's early development; as a result, educational authorities must ensure that students can learn in an atmosphere that helps them to achieve academically. The goal of this research is to examine the impact that economic circumstances have on students' academic performance on a school-by-school basis. A student's performance in school can be fairly measured by their ELA and Math proficiency levels. ELA and Math proficiency levels are divided into four levels, each reflecting a series of tasks of increasing difficulty, with Level 4 being the most difficult and Level 1 being the easiest. Students functioning below Level 1 in both ELA and Math are unable to demonstrate the most fundamental types of knowledge and skills consistently. Such students face significant challenges in expanding their knowledge and skills in other areas. It should be noted that standardized tests given to students at each grade level are used to calculate ELA and Math proficiency levels. As a result, by taking the average of the proficiency levels, we may generalize the academic performance of a school's student body.

To better understand the impact of economic issues on students and their performance, this paper will employ the economic need index and school income estimates to quantify the economic health of schools and, consequently, their student population. The economic need index assesses a school's socioeconomic situation and calculates a score based on the number of students eligible for free lunch, public assistance, or who live in temporary housing. A school with a high economic need index, for example, indicates that a large share of the student body is financially unstable. Furthermore, school income estimates are a monetary value that represents the available funding obtained by schools from income sources such as government financing. As a result, by utilizing this economic data, we hope to discover a link between a student's performance and the economic dynamics at work. In addition to the economic need index and school income estimates, other parameters of interest such as enrolment,

attendance, and progress report grades will be investigated to conduct a more detailed investigation of the influence economic conditions have on students' academic achievement on a school-by-school basis.

Furthermore, one aspect of ensuring academic success in a school system is giving all students an equal opportunity to succeed. As a result, this paper examines potential economic disparities among different racial groupings. An investigation will be carried out to determine whether underrepresented groups in the education system, such as Black and Hispanic students, fare better or worse than other groups, particularly White and Asian students. There are 1,094,138 students in the NYC school system. Of those students: 40.8 percent Hispanic, 24.7 percent Black, 16.5 percent Asian, and 14.8 percent White. This investigation will strengthen our understanding not only of probable inequalities in academic success opportunities but also of whether economic issues influence academic performance.

Background

Why is it vital to investigate adolescent education in their formative years? The study, Education and the Subjective Quality of Life, discovered that well-educated people experienced less emotional and physical distress (Ross & Willigen, 1997). Furthermore, education in one's life has been shown to lessen distress through employment and financial opportunities (Ross & Willigen, 1997). Additionally, the findings of the research paper, Neighborhoods and Adolescent Development, reveal that residential area is equally essential for adolescents in terms of educational outcomes (Boardman & Onge, 2005). As a result, because education is critical to a child's progress, studying the impact of out-of-school issues can assist the adolescent in obtaining an education that will help them lead better lives. Therefore, because a student's financial situation is such an important component in their lives, investigating the impact of economic factors in the school environment might lead to discoveries that may aid in a child's academic success.

Relevant Literature

The impact of economic and non-school factors on student achievement is a well-studied topic in educational economics. As a result, our study resembles other research that tackles comparable questions. One relevant literature paper, for example, titled "A New Model for Student Support in High-Poverty Urban Elementary Schools: Effects on Elementary and Middle School Academic Outcomes," investigated the academic achievement of students participating in a student support intervention operating in high-poverty elementary schools (Walsh et al., 2014). As a result, it was determined that students who participated had higher report card scores than their peers and performed better on middle school English language arts and mathematics assessments (Walsh et al., 2014). As a result, this study demonstrates the importance of addressing out-of-school elements that influence student development (Walsh et al., 2014). To continue, another significant literature paper, "The increasing impact of socioeconomics and race on standardized academic test scores across elementary, middle, and high schools," looked into the impact of socioeconomic and racial factors on test scores (White et al., 2016). Socioeconomic and racial characteristics were discovered to account for 52% of the variance in language exam scores and 59% of the variance in math exam scores in high school (White et al., 2016). As a result, these findings have important implications since they show that school-level initiatives to promote academic achievement will be hampered by socioeconomic and ethnic characteristics (White et al., 2016). In addition to the above two quoted publications, "Discrepancies in Advanced Success Outcomes for Texas Students as a Function of Economic Disadvantage" investigated differences in advanced achievement in Texas between economically disadvantaged and non-disadvantaged students (Lee & Slate, 2014). It was discovered that economically disadvantaged students underperformed their classmates statistically significantly in both English language arts and mathematics (Lee & Slate, 2014).

With the inclusion of three relevant studies, the question of how this study is similar to and distinct from the literature may arise. To begin with, this study looks into similar educational system variables. All three studies, including this one, investigated the effect of parameters on student performance. The resulting resemblance is merely the outcome of analyzing educational systems, as student achievement is the most sought-after measure of how well an educational system performs. What is more noteworthy, though, is how this study differs. According to the three publications, each one appears to have concentrated on a different characteristic, such as economic or socioeconomic aspects. We intend to look at many perspectives in this study to conduct a more thorough investigation into the aspects that have the greatest impact on student achievement. Not only will economic and socio-economic aspects be investigated, but so will racial factors. As a result, this study varies due to the development of a more comprehensive profile of the effects that schools and students confront.

Data Source

The dataset provided by PASSNYC will be used in this study. PASSNYC is a non-profit organization that facilitates a collective influence aimed at extending educational opportunities for smart and impoverished children in New York City. PASSNYC identifies students in New York City's underperforming school districts using public data. As a result, this dataset will aid in this research because it provides relevant information on students via the information provided on a school-by-school basis.

Data Source Link: https://www.kaggle.com/passnyc/data-science-for-good

Data Cleaning

Using the PASSNYC dataset, this study performed data cleaning to build datasets that would be utilized to answer the study's inquiry. To begin with, a general dataset labelled "SchoolData" was built from the PASSNYC information. Four dependent datasets were then built from SchoolData, each for a different purpose. EcoData was intended to highlight economic data for all NYC public schools. BhData was designed to showcase economic facts for all of New York City's predominantly Black and Hispanic schools. WaData was established to highlight economic data for all of New York City's predominantly White and Asian schools. MrData was designed to highlight economic data for all mixed-race schools in New York City.

Sample Output:

SchoolData

	School Name	Economic Need Index	School Income Estimate	Percent Black / Hispanic	Average ELA Proficiency	Average Math Proficiency	District	Percent White / Asian	Black/Hispanic Dominant Schools	White/Asian Dominant Schools	Mixed Race Schools
0	P.S. 015 ROBERTO CLEMENTE	0.919	31141.72	92	2.14	2.17	1	6	True	False	False
1	P.S. 019 ASHER LEVY	0.641	56462.88	83	2.63	2.98	1	16	True	False	False
2	P.S. 020 ANNA SILVER	0.744	44342.61	57	2.39	2.54	1	39	False	False	True
3	P.S. 034 FRANKLIN D. ROOSEVELT	0.860	31454.00	92	2.48	2.47	1	9	True	False	False
4	THE STAR ACADEMY - P.S.63	0.730	46435.59	84	2.38	2.54	1	14	True	False	False

EcoData

	Economic Need Index	School Income Estimate	Average ELA Proficiency	Average Math Proficiency	const
0	0.919	31141.72	2.14	2.17	1
1	0.641	56462.88	2.63	2.98	1
2	0.744	44342.61	2.39	2.54	1
3	0.860	31454.00	2.48	2.47	1
4	0.730	46435.59	2.38	2.54	1

BhData

	Average ELA Proficiency	Average Math Proficiency	Economic Need Index	School Income Estimate	District
0	2.14	2.17	0.919	31141.72	1
1	2.63	2.98	0.641	56462.88	1
3	2.48	2.47	0.860	31454.00	1
4	2.38	2.54	0.730	46435.59	1
5	2.29	2.48	0.858	39415.45	1

WaData

	Average ELA Proficiency	Average Math Proficiency	Economic Need Index	School Income Estimate	District
10	3.24	3.63	0.559	40809.90	1
21	3.83	4.03	0.257	76833.96	1
24	2.75	3.24	0.858	26114.78	2
25	3.09	3.41	0.176	103399.19	2
26	3.40	3.71	0.101	144270.13	2

MrData

	Average ELA Proficiency	Average Math Proficiency	Economic Need Index	School Income Estimate	District
2	2.39	2.54	0.744	44342.61	1
6	2.80	3.20	0.499	43706.73	1
14	2.86	3.20	0.362	63760.00	1
16	2.89	2.99	0.451	62519.57	1
17	2.55	2.68	0.430	57504.48	1

Schools with Unavailable Data

Some schools with unavailable data in the parameters of interest were removed to allow for better data analysis. This exclusion of data, however, may result in some bias in the results. In this step, we will examine the data that was eliminated to identify any biases that may have emerged.

	ENI	School Income Estimate	Percent Black / Hispanic	Avg ELA Proficiency	Avg Math Proficiency	Percent White / Asian
mean	0.69	44073.06	78.59	2.49	2.52	19.72
std	0.18	19478.92	26.16	0.38	0.47	25.62
min	0.10	20772.06	4.00	1.81	1.83	0.00
25%	0.59	28705.90	66.00	2.23	2.15	2.00
50%	0.73	41049.88	93.00	2.38	2.39	5.00
75%	0.83	49138.86	97.00	2.67	2.79	30.00
max	0.94	118322.46	100.00	3.91	4.19	95.00

Summary statistics were developed to provide an overview of the data, allowing for a better comprehension of the inaccessible data. However, it should be noted that if a school had at least one missing data entry in any of the criteria of interest, it was eliminated; as a result, some data entry points remain. Following data cleansing, only 828 of the initial 1272 schools remain, implying that 444 schools were deleted due to insufficient data. Because of the limited sample size, the findings in the research may not fully represent the entire New York City school system. To continue, the available data for the economic need index, school income estimates, and proficiency levels appear to have a low standard deviation; as a result, because the missing data excludes schools of both high and low levels, it is possible that the results will not be affected. The standard deviation for the percentage of Black/Hispanic and White/Asian students at each school, on the other hand, appears to be large. As a result, when we investigate the covariates, the results may not accurately reflect the genuine variations across race percentiles.

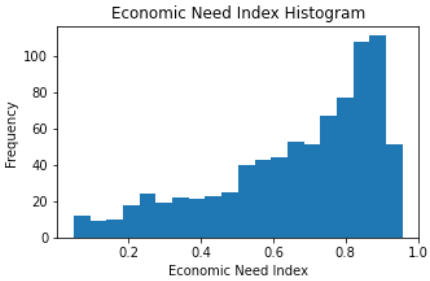
Summary Statistics

Our analytical strategy will begin with a summary of the dataset's economic statistics, followed by a more in-depth examination of the data through the use of graphics. As a result, we will have a better grasp of the impact of economic need and school income estimates on school average proficiency levels.

	ENI	School Income Estimate	Avg ELA Proficiency	Avg Math Proficiency
mean	0.66	48696.47	2.56	2.74
std	0.22	21410.79	0.36	0.46
min	0.05	16901.67	1.97	1.90
25%	0.54	33649.50	2.27	2.36
50%	0.73	43665.94	2.48	2.67
75%	0.85	59073.58	2.79	3.06
max	0.96	181382.06	3.93	4.20

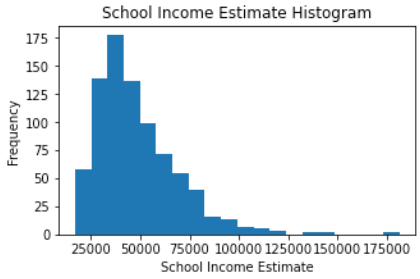
When we interpret the summary statistics, we discover some interesting results. To begin, the means of the economic need index, school income estimates, average ELA proficiency, and average Math proficiency are, respectively, 0.66, $48,696.46, 2.55, and 2.74. We can observe from the standard deviation that each parameter (e.g., economic need index, school income estimates, etc.) appears to have a substantial spread. Looking at the percentile, we can see that the majority (75 percentile) of institutions and students are in financial need. However, the bulk of students (75 percentile) appear to have good proficiency levels. Overall, the dataset appears to contain information about a possible association between the economic need index, school income estimates, and average ELA and Math proficiency. However, after reviewing the dataset's graphics, a more definitive interpretation can be reached.

Histograms

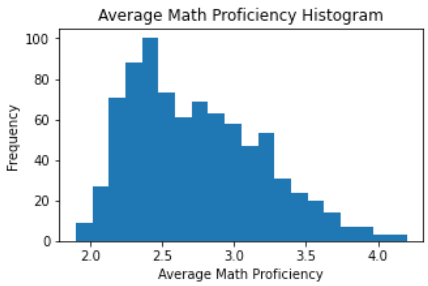


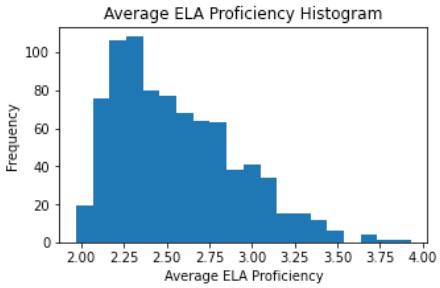
The analysis of the histogram plot of the economic need index shows an important discovery regarding the state of the student body in New York's schools. The left skewness of the histogram indicates that the majority of schools have a substantial number of economically disadvantaged students.

These findings are significant because they show that many of the children pursuing education do not have financial security, which may harm their academic performance.



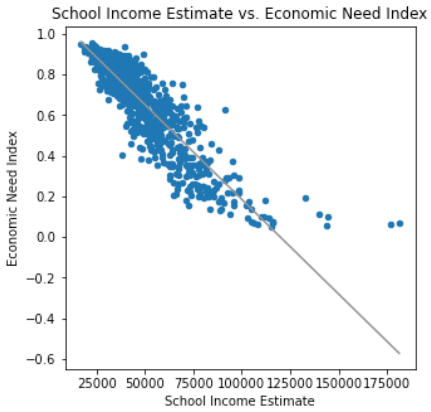
An interesting discovery concerning the situation of school funding is revealed by analyzing the histogram plot of school income estimates. The histogram's right skewness demonstrates that many schools do not receive a lot of state support concerning the small amount of school that does. As a result, this finding may indicate that schools lack the funding to provide resources to support their student population in times of financial distress. Therefore, a student experiencing financial insecurity may be unable to obtain assistance, thereby jeopardizing academic performance.





The average ELA and Math Proficiency histograms shed light on the academic performance of the student population. The small right skewness indicates that the majority of students have ELA and Math proficiency levels ranging from 2 to 3. This observation could be the result of students not receiving the required education to excel academically due to other causes of interest, such as financial need. It should be noted, however, that the proficiency levels appear to be ordinary, and that high levels are typically associated with talented or bright students.

ScatterPlot

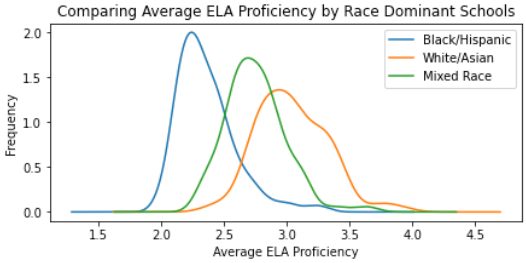


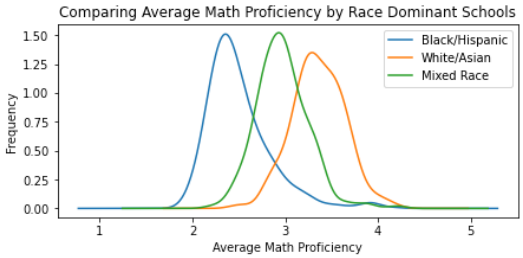
To further investigate the related influence on financial needs and academic performance, a scatterplot was created to discover a potential association between the economic need index and school income estimates. It has been revealed that there is a negative correlation between school funding and the economic need index. As a result, as school income rises, the economic need index falls. This indicates that with proper financing, the financial status of a school's population can be improved. When linked to proficiency levels, an increase in school financing and, consequently, a decrease in economic need may result in higher ELA and Math proficiency levels.

Covariates

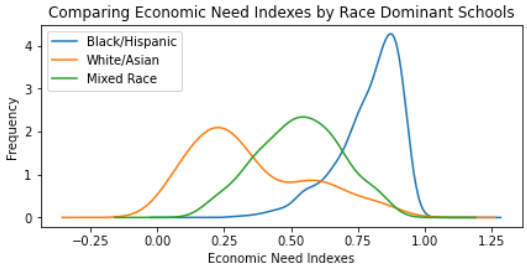
As previously said, education is a vital aspect of a child's success. One of the primary purposes of PASSNYC is to provide kids from all races with equal opportunity for educational excellence. To further the goal of analyzing the impact of schools' and students' economic needs on proficiency levels, a report is prepared for schools with varying race percentiles. Identifying Black/Hispanic dominant schools as those with a Black/Hispanic student population percentage greater than or equal to 70% and White/Asian dominant schools as those with a White/Asian student population percentage greater than or equal to 70%. With this information, we can better assess the benefits and drawbacks of various groups. Furthermore, schools that were classified as neither were recognized as mixed-race schools to

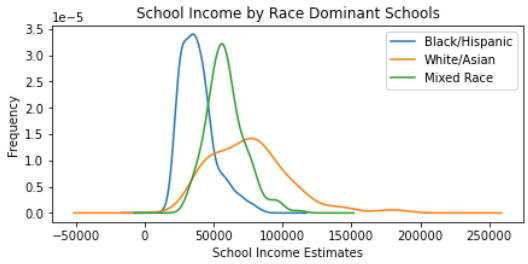
serve as a control. Overall, there are 511 Black/Hispanic dominant schools, 123 White/Asian dominant schools, and 194 mixed-race schools in total. As a result, more Black/Hispanic dominant schools exist than White/Asian dominant and mixed-race schools combined.





When the two density plots above, which compare average ELA and Math proficiency by race dominant schools, are analyzed, it is clear that White/Asian dominant schools have higher ELA and Math proficiency levels than Black/Hispanic dominant schools. Furthermore, an intriguing finding is that the Math and ELA performance levels of mixed-race schools appear to be in the midrange of the dominant race schools. As a result, the findings provide even more impetus to investigate the economic variables affecting each institution.



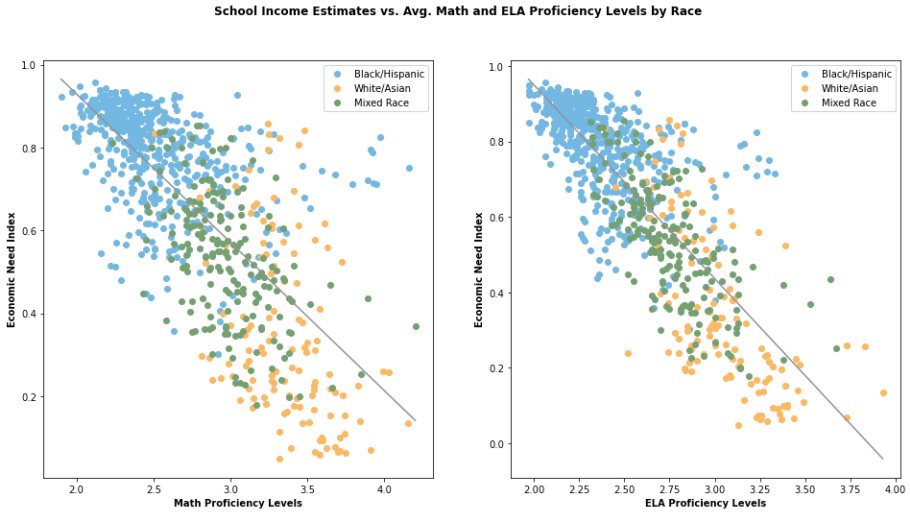


When the two density charts above are compared, it is evident that White/Asian dominant schools have lower Economic Need Indices and greater school income estimates than Black/Hispanic dominant schools. Furthermore, a noteworthy discovery is that mixed-race schools' economic need indexes and school income estimations appear to be at the midpoint of the dominant race schools. As a result, addressing the economic needs of schools with high economic need indexes and low incomes will

help underrepresented groups to flourish academically and offer students from all races equitable chances for educational excellence.

Overall Plots

Many graphs, tables, and interpretations have been shown in the paper; as a result, it is a good practice to reiterate exactly what we are aiming to answer with this data, i.e., the essential message of this study. Two plots are created and understood to supplement the primary idea to better communicate the message. In general, schools with high funding and student bodies with low economic need indices do better in ELA and Math proficiency levels; however, this performance emerges less frequently in schools with a large proportion of underrepresented groups, such as Black and Hispanic students.

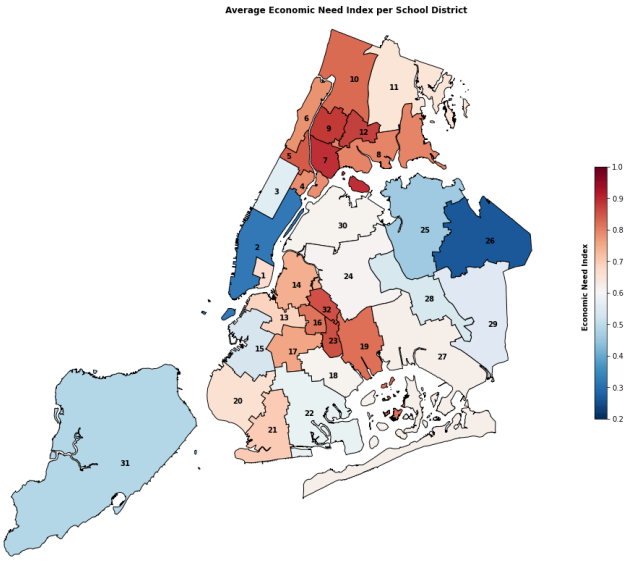


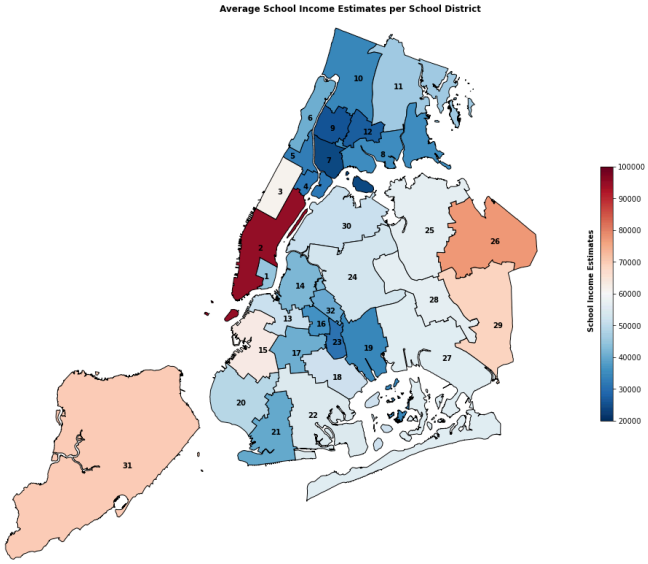
The analysis of the scatterplots above provides the best potential explanation for the message being asked. Overall, it indicates that there is a negative correlation between the economic need index and proficiency levels. On the other hand, there appears to be a positive correlation between school income and proficiency levels. Furthermore, it appears that differences exist between racial groups based on these findings. In all scatterplots, Black/Hispanic dominant schools appear to do worse in all criteria than other racial groups. To continue, White/Asian dominant schools appear to perform better in all parameters than other ethnic groups in all scatterplots. Between the two, mixed-race schools appear to be in the middle of the dominant race schools. Furthermore, White/Asian dominant schools appear to have more variability than other racial groups. As a result, because they try to highlight the impact of economic need and school income on competence levels, these plots are pertinent to the main message. It should be mentioned that multiple plots were utilized to help understand all the factors and their relationships with one another.

Mapping

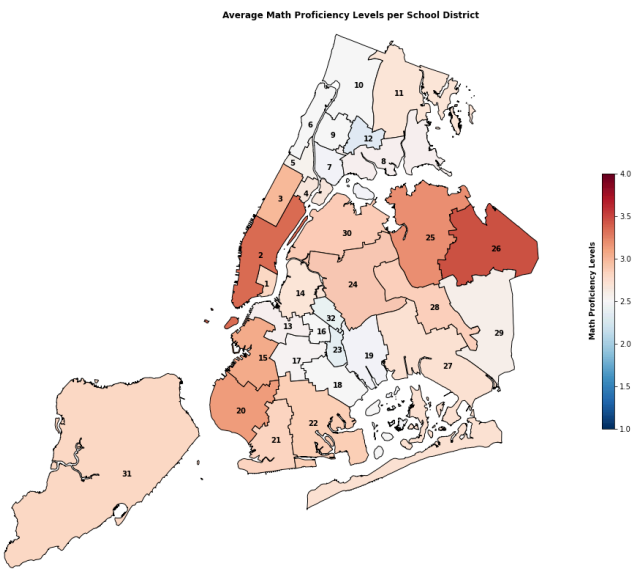
Multiple maps depicting the study's many parameters will be developed in this section of the report (economic need index, school income estimates, Math and ELA proficiency levels). The maps will make use of data from NYC Open Data, a website managed by the City of New York that gives data on various elements of the city. As a result, the map data and appropriate school districts will be retrieved from this site. Furthermore, in this part, we will introduce new statistics on attendance and enrolment in school districts from 2010 to 2011. More details about the new data will be made available later.

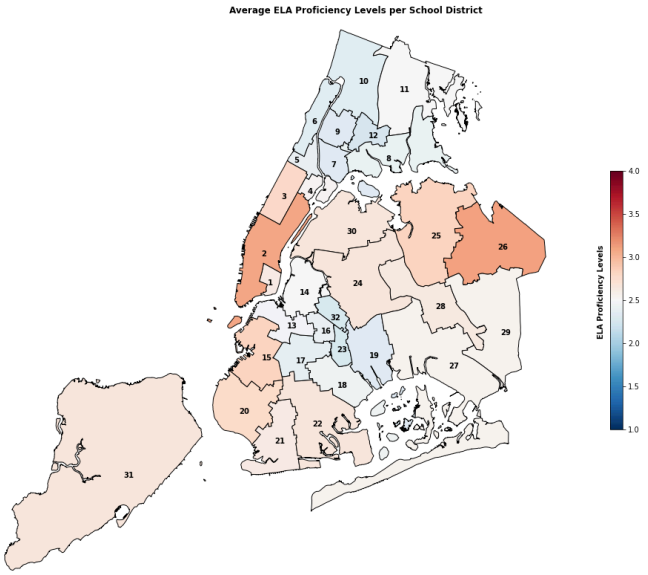
Data Link: https://data.cityofnewyork.us/City-Government/Community-Districts/yfnk-k7r4





According to the analysis of the two maps above on average economic need index per school district and average school income estimates per school district, large disparities exist between school districts. Districts such as 2, 26, 29, and 31 appear to perform better in both the economic need index and School Income when compared to poorer districts such as 7, 9, 10, 12, and 23, which perform worse in both the economic need index and school income estimates. As a result, these maps not only depict potential links between the economic need index and School Income, but also the economic and funding inequalities that exist in New York City.





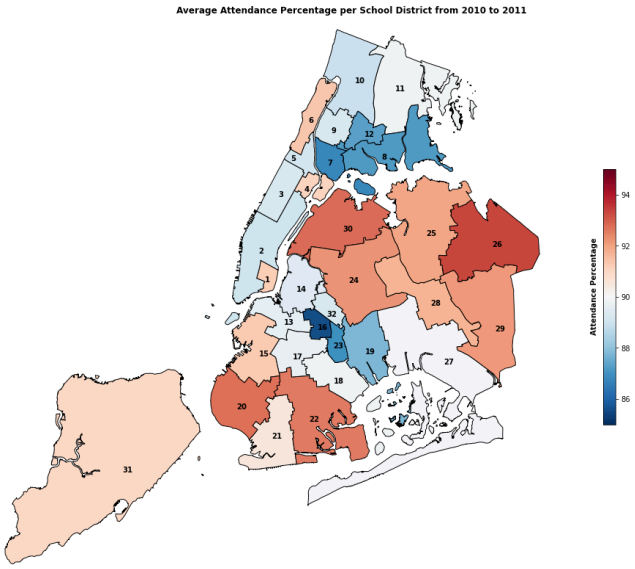
According to the analysis of the two maps above on average Math and ELA proficiency levels by the school district, considerable discrepancies occur between school districts. School districts such as 2, 25, and 26 appear to have the highest levels of proficiency, while districts such as 32, 23, and 12 appear to have the lowest levels of proficiency. In general, ELA and Math competence levels appear to be comparable across school districts; however, Math proficiency levels appear to be higher overall than ELA proficiency levels. To continue, a review of the maps of the economic need index and School Income shows that there may be a link between the economic state of school districts and proficiency levels.

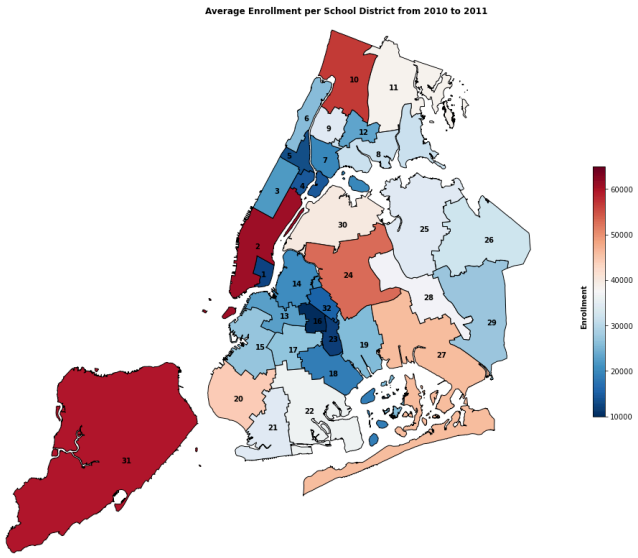
New Data

A new dataset will be added to supplement the data in our analysis, which will offer information on attendance and enrolment statistics by the school district from 2010 to 2011. Each school district's relevant average attendance percentage and average enrolment are included in the dataset. Using this new information, a map of the variables will be produced and examined for a possible correlation with the existing data. As previously noted, this data was obtained from NYC OpenData.

The addition of attendance and enrolment data by the school district will provide a unique perspective on how economic issues influence Math and ELA proficiency levels. A student's and their family's financial situation may have an impact on their attendance at school. As a result, this data may identify more opportunities for improvement in the education system. For example, if a kid must miss school to support their family's financial responsibilities then these circumstances may have an impact on the student's academic progress. Therefore, this data will not only present a new area of interest but will also improve the investigation in the paper.

Data Link: https://data.cityofnewyork.us/Education/2010-2011-School-Attendance-and-Enrollment-Statist/7z8d-msnt





The two maps of average attendance and enrolment by the school district from 2010 to 2011 reveal no overall association or pattern between the two. It is worth noting, however, that the districts that fared best and worst in all categories in the preceding maps did not have significant variations in enrolment and attendance percentages. It's debatable whether there's a link between these two factors and the ones mentioned previously. It may be argued, however, that low attendance and enrolment in impoverished districts like 16 and 23 may show a link between low proficiency levels, economic necessity, and school income.

Web-Scraping

Creating a standardized grading system for each school to allow for comparison and ranking is a critical component of guaranteeing a well-functioning educational system. The New York City Department of Education created the Progress Report Grades to demonstrate how each school is performing. The reports' goal is to hold schools accountable for student accomplishment and to assist schools in charting a plan of action to increase student learning. Schools are graded with letters ranging from A to F based on their contribution to student learning in the areas of the school environment,

student performance, and student progress. Schools can also earn extra credit for assisting their high-needs students in making exceptional progress.

The School Environment score indicates how well schools have prepared students for learning. It covers the attendance rate as well as the results of the Learning Environment Survey. The Student Performance score comprises the proportion of students who scored proficient or above on the New York State Assessment English Language Arts and Math tests, as well as the median student proficiency. It also contains the proportion of students who graduated in four years and the proportion of students who finished in six years. Schools are also evaluated based on the sort of diplomas pupils get after four and six years, with more credit given to diplomas indicating a greater level of competency and college readiness. The Student Progress Score compares students' current year performance levels to their prior-year performance levels to determine how much schools assist students' progress during the school year. As a result, to have a better knowledge of the health of each school within the NYC education system, this study will compare the Overall Grade and Overall Score criteria to the original parameters. The Overall Score is the sum of the four scores (School Environment, Student Performance, Student Progress, and Additional Credit), and the Overall Grade is the letter grade assigned to the overall score. This data will be scraped from the New York City open data API (https://opendata.cityofnewyork.us), and specifically for this section of the paper, the school progress report grades will be scraped from the relevant API serial number for the years 2010 and 2011 (https://dev.socrata.com/foundry/data.cityofnewyork.us/yig9-9zum).

This data will improve the paper by providing a different viewpoint on the state of each school and the student body. This will enable new analyses of the impact of economic factors on New York City schools, notably the economic need index and school income estimates. Furthermore, because

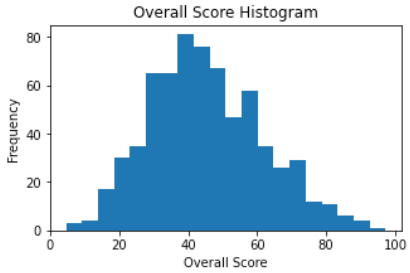
proficiency levels may not accurately reflect the state of the student body, integrating a new measure of success will allow for more in-depth comparison as well as highlight areas for development. Because Progress Report Grades are assigned to each school, we can easily integrate this new information with our original dataset because both contain data on every school in New York City. Therefore, the merging process will simply entail associating the new parameters with each of the appropriate schools. We next use this merged dataset to generate visualizations like histograms, scatterplots, density plots, and maps to allow for a better understanding of the data and the possibility of discovering new areas of interest and correlations with the original parameters. In addition to looking at the entire school, we will continue our investigation of covariates, namely race percentiles, to study possible discrepancies among different racial groups.

Sample Output:

Web-Scraping DataSet

	school	Overall Grade	Overall Score
0	p.s. 015 roberto clemente	C	27.0
1	p.s. 019 asher levy	B	48.7
2	p.s. 020 anna silver	B	48.2
3	p.s. 034 franklin d. roosevelt	C	39.8
4	p.s. 063 william mckinley	B	49.0

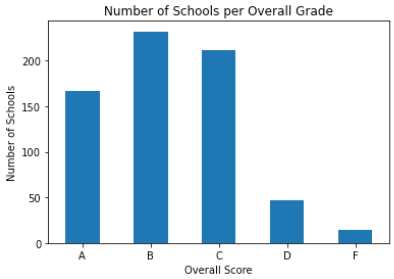
Web-Scraping Histograms



The investigation of the histogram plot of the overall score reveals an important discovery about the state of New York City's schools. The histogram's minor right skewness suggests that the majority of institutions have an overall score of about 40. These findings are crucial because they reveal that many

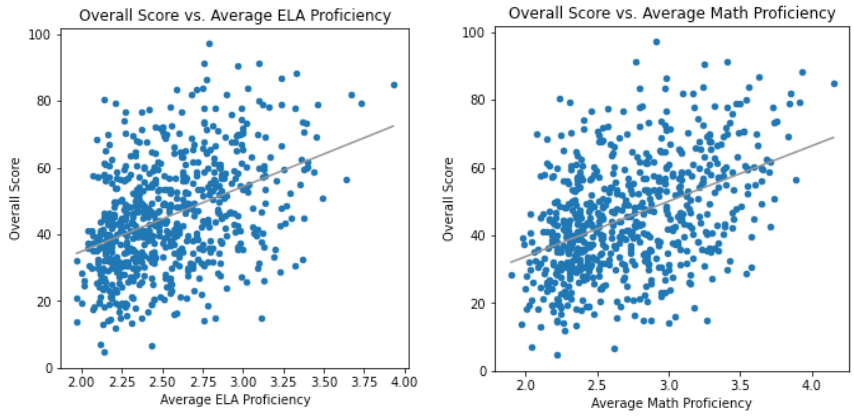
children pursuing an education do not attend a school that might help them improve in many aspects, which may damage their academic achievement.

Web-Scraping BarPlot



To get a better picture of the state of the schools in New York City, we classified the number of schools by their overall grade. As a result, the Bar Plot provides a different view of the state of education in New York City than the histogram. Because of the way letter grades are assigned, it looks like the majority of schools receive B or C letter grades. These statistics indicate that the majority of schools have areas to improve student growth in academics.

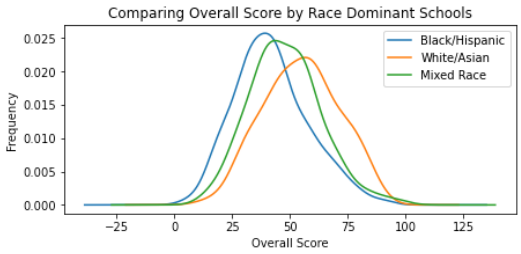
Web-Scraping ScatterPlots



The scatterplots of Overall Score versus average ELA and Math proficiency levels demonstrate a minor positive correlation between Overall Score and student academic achievement. However, it appears that there is a lot of variation in both plots. To continue, the positive association shows that

higher-scoring schools have higher average ELA and Math proficiency levels. This statement emphasizes the fact that the majority of schools have significant areas for improvement after analyzing the histogram and bar plot. As a result, students' ability to achieve academically may be limited.

Web-Scraping Covariates



When the density plots comparing overall scores by race dominant schools are evaluated, it is obvious that White/Asian dominant schools outperform Black/Hispanic dominant schools. Intriguingly, the overall score of mixed-race schools appears to be in the middle of the dominant race schools. As a result, the findings provide even more pressure to explore the economic aspects influencing each school, as well as the discrepancies in opportunities between students of different races.

OLS Regression

In this part, a regression analysis will be performed on the study's initial parameters. To be more specific, the independent variables (X) will be the economic need index and school income estimates, while the dependent variables (Y) will be average ELA and Math proficiency levels. With the regression analysis, we hope to expand our inquiry into potential relationships between the two categories of variables. Due to the scatterplots of the parameters exhibiting linearity, we hypothesize that the economic relationship between Y and X is linear. Additionally, a regression line of degree one was fitted to the scatterplots, demonstrating that the data appears to follow a linear trend.

The appropriate variables to analyze using regression are the economic need index and school income estimates as the X variables and average ELA and Math proficiency levels as the Y variables. Because the goal of the study is to investigate the impact of economic factors such as the economic need index and school income estimates on proficiency levels, a regression analysis of these variables is thought to be the most useful. The existence of a possible relationship between the X and Y variables has been proven throughout the investigation, for example, through plots and maps. As a result, since a link has been established, the economic need index and school income estimates appear to be the best variables for explaining variations in average ELA and Math proficiency levels. Therefore, using regression, we can discover a more precise correlation, which will strengthen our examination of the study's question.

In addition to the factors listed above, a regression analysis will be performed to investigate the impact of school income estimations on economic need indices. Previous scatterplots of the two factors revealed a possible linear link; as a result, further analysis is warranted. Because one of the major economic policies at the disposal of educational systems is school funding, determining whether such efforts might help to reduce economic need indices can bring to light an avenue that will ideally allow for a multiplier effect on boosting proficiency levels. As a result, the independent variable (X) in the OLS regression model will be school income estimates, while the dependent variable (Y) will be economic need index.

It should be noted that the covariates will not be subjected to a regression analysis because it is assumed that they will reveal the same relationship as the total dataset. Furthermore, other variables such as attendance, enrolment, and overall score did not appear to have a meaningful link with the major variables of interest, average ELA and Math proficiency levels.

OLS Regression Results

Dep. Variable:	Average ELA Proficiency	R-squared:	0.665
Model:	OLS	Adj. R-squared:	0.665
Method:	Least Squares	F-statistic:	1643.
Date:	Wed, 13 Apr 2022	Prob (F-statistic):	1.36e-198
Time:	01:23:12	Log-Likelihood:	134.46
No. Observations:	828	AIC:	-264.9
Df Residuals:	826	BIC:	-255.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.4181	0.022	152.425	0.000	3.374	3.462
Economic Need Index	-1.2971	0.032	-40.534	0.000	-1.360	-1.234

Omnibus:	157.336	Durbin-Watson:	1.488
Prob(Omnibus):	0.000	Jarque-Bera (JB):	342.068
Skew:	1.048	Prob(JB):	5.26e-75
Kurtosis:	5.350	Cond. No.	6.51

An OLS regression model was used to analyze the impact of Economic Need Indices on average ELA proficiency levels in this result table. The slope is $\hat{\beta}_1 = -1.31$ and the intercept is $\hat{\beta}_0 = 3.42$. The negative $\hat{\beta}_1$ parameter estimate implies that the economic need index has a negative effect on average ELA proficiency levels, as evidenced by the scatterplots. The p-value of 0.000 indicates that the influence of Economic Need Indices on average ELA proficiency levels is statistically significant (using $p < 0.05$ as a rejection rule). The R-squared value of 0.66 indicates that economic need indices account for approximately 66% of the variation in average ELA proficiency levels.

OLS Regression Results

Dep. Variable:	Average ELA Proficiency	R-squared:	0.455
Model:	OLS	Adj. R-squared:	0.454
Method:	Least Squares	F-statistic:	688.9
Date:	Wed, 13 Apr 2022	Prob (F-statistic):	6.80e-111
Time:	01:23:12	Log-Likelihood:	-67.772
No. Observations:	828	AIC:	139.5
Df Residuals:	826	BIC:	149.0
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.0109	0.023	88.533	0.000	1.966	2.055
School Income Estimate	1.121e-05	4.27e-07	26.246	0.000	1.04e-05	1.2e-05

Omnibus:	96.271	Durbin-Watson:	1.403
Prob(Omnibus):	0.000	Jarque-Bera (JB):	148.636
Skew:	0.800	Prob(JB):	5.30e-33
Kurtosis:	4.323	Cond. No.	1.32e+05

An OLS regression model was used to analyze the impact of school income estimates on average ELA proficiency levels in this result table. The slope is $\hat{\beta}_1 = 0.00001134$ and the intercept is $\hat{\beta}_0 = 2$. The positive $\hat{\beta}_1$ parameter estimate implies that the School Income Estimate has a positive effect on average ELA proficiency levels, as evidenced by the scatterplots. The p-value of 0.000 indicates that the

influence of school income estimates on average ELA proficiency levels is statistically significant (using $p < 0.05$ as a rejection rule). The R-squared value of 0.458 indicates that school income estimates account for approximately 45.8% of the variation in average ELA proficiency levels.

OLS Regression Results

Dep. Variable:	Average Math Proficiency	R-squared:	0.532
Model:	OLS	Adj. R-squared:	0.532
Method:	Least Squares	F-statistic:	940.3
Date:	Wed, 13 Apr 2022	Prob (F-statistic):	1.80e-138
Time:	01:23:12	Log-Likelihood:	-209.44
No. Observations:	828	AIC:	422.9
Df Residuals:	826	BIC:	432.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.7273	0.034	109.721	0.000	3.661	3.794
Economic Need Index	-1.4866	0.048	-30.664	0.000	-1.582	-1.391

Omnibus:	189.076	Durbin-Watson:	1.248
Prob(Omnibus):	0.000	Jarque-Bera (JB):	488.713
Skew:	1.173	Prob(JB):	7.54e-107
Kurtosis:	5.944	Cond. No.	6.51

An OLS regression model was used to analyze the impact of Economic Need Indices on average Math proficiency levels in this result table. The slope is $\hat{\beta}_1 = -1.50$ and the intercept is $\hat{\beta}_0 = 3.73$. The negative $\hat{\beta}_1$ parameter estimate implies that the economic need index has a negative effect on average Math proficiency levels, as evidenced by the scatterplots. The p-value of 0.000 indicates that the influence of Economic Need Indices on average Math proficiency levels is statistically significant (using $p < 0.05$ as a rejection rule). The R-squared value of 0.532 indicates that economic need indices account for approximately 53.2% of the variation in average Math proficiency levels.

OLS Regression Results

Dep. Variable:	Average Math Proficiency	R-squared:	0.343
Model:	OLS	Adj. R-squared:	0.342
Method:	Least Squares	F-statistic:	430.5
Date:	Wed, 13 Apr 2022	Prob (F-statistic):	2.74e-77
Time:	01:23:12	Log-Likelihood:	-350.45
No. Observations:	828	AIC:	704.9
Df Residuals:	826	BIC:	714.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.1331	0.032	66.753	0.000	2.070	2.196
School Income Estimate	1.246e-05	6.01e-07	20.748	0.000	1.13e-05	1.36e-05

Omnibus:	101.641	Durbin-Watson:	1.178
Prob(Omnibus):	0.000	Jarque-Bera (JB):	155.258
Skew:	0.843	Prob(JB):	1.93e-34
Kurtosis:	4.288	Cond. No.	1.32e+05

An OLS regression model was used to analyze the impact of school income estimates on average Math proficiency levels in this result table. The slope is $\hat{\beta}_0 = 0.00001261$ and the intercept is $\hat{\beta}_1 = 2.12$.

The positive $\hat{\beta}_1$ parameter estimate implies that the School Income Estimate has a positive effect on average Math proficiency levels, as evidenced by the scatterplots. The p-value of 0.000 indicates that the influence of school income estimates on average Math proficiency levels is statistically significant (using $p < 0.05$ as a rejection rule). The R-squared value of 0.347 indicates that school income estimates account for approximately 34.7%of the variation in average Math proficiency levels.

OLS Regression Results

Dep. Variable:	Economic Need Index	R-squared:	0.796
Model:	OLS	Adj. R-squared:	0.796
Method:	Least Squares	F-statistic:	3219.
Date:	Wed, 13 Apr 2022	Prob (F-statistic):	3.54e-287
Time:	01:23:12	Log-Likelihood:	722.87
No. Observations:	828	AIC:	-1442.
Df Residuals:	826	BIC:	-1432.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.1181	0.009	127.913	0.000	1.101	1.135
School Income Estimate	-9.324e-06	1.64e-07	-56.736	0.000	-9.65e-06	-9e-06

Omnibus:	84.571	Durbin-Watson:	1.402
Prob(Omnibus):	0.000	Jarque-Bera (JB):	447.668
Skew:	0.274	Prob(JB):	6.17e-98
Kurtosis:	6.560	Cond. No.	1.32e+05

An OLS regression model was used to analyze the impact of school income estimates on Economic Need Indices in this result table. The slope is $\hat{\beta}_0 = -0.000009309$ and the intercept is $\hat{\beta}_1 = 1.12$. The negative $\hat{\beta}_1$ parameter estimate implies that the School Income Estimate has a negative effect on Economic Need Indices, as evidenced by the scatterplots. The p-value of 0.000 indicates that the influence of school income estimates on Economic Need Indices levels is statistically significant (using $p < 0.05$ as a rejection rule). The R-squared value of 0.796 indicates that school income estimates account for approximately 79.6% of the variation in Economic Need Indices.

OLS Regression Summary

The results of the OLS Regression Models revealed some intriguing discoveries about the effects of economic factors like economic need indices and school income estimates on ELA and Math proficiency levels. To begin with, the most significant conclusion was that the OLS regression revealed a negative association between economic need indices and proficiency levels. As previously noted, a one-

point increase in ENI results in a 1.31 fall in average ELA proficiency levels and a 1.50 fall in average Math proficiency levels. Furthermore, a high R-squared value of 66 percent for ELA proficiency and 53.2 percent for Math proficiency indicated that economic need indices accounted for a significant portion of the changes in proficiency levels. To continue, the influence of school income estimates was found to have a positive relationship with proficiency levels. It was discovered that a one-point rise in school income results in a 0.00001134 increase in average ELA proficiency levels and a 0.00001261 increase in average Math proficiency levels. Furthermore, a high R-squared value of 45.8 percent for ELA proficiency and 34.7 percent for Math proficiency indicated that estimates of school income accounted for a significant portion of the variations in proficiency levels. In addition to these regressions, one was built to examine the impact of school income estimates on economic need indices, which was shown to be negative. A one-point increase in school income estimates, for example, results in a 0.000009309 decrease in the economic need index. School income estimates accounted for a significant portion of the fluctuations in the economic need index, owing to the high R-squared value of 79.6 percent in this regression. It should be emphasized that schools are typically supported with thousands of dollars; hence, even with a tiny slope, considerable funding can have a significant impact on proficiency levels and ENI.

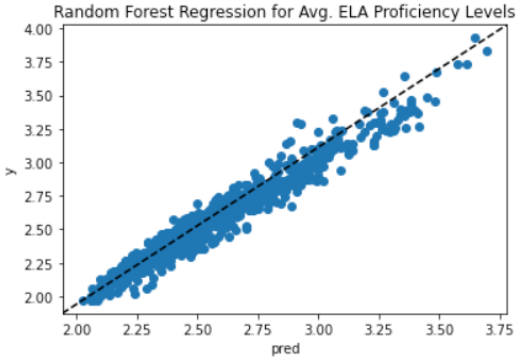
The economic theory underpinning these findings is that they show educational systems how to enhance the average competency levels of their schools. The findings indicate that favourable economic policies, such as raising school financing and assisting students and their families in achieving more financial stability, will benefit the school's student body's proficiency levels. As a result, these findings add weight to the study's answer to the topic of how economic circumstances influence proficiency levels. They provide a rough but quantifiable estimate of how the effects of various economic factors can enhance schools.

Machine Learning

Machine Learning will be employed in this phase of the study to improve forecasts of average ELA and Math proficiency levels. To determine the best and worst outcomes of the proficiency levels, the parameters of the economic need index, school income estimates, and racial percentiles will be applied. The use of regression trees will provide a better understanding of the impact that economic circumstances and ethnic groupings have on Math and ELA competency, allowing for better solutions on how to improve these levels. Because there are two y variables of relevance, two tree regressions will be performed: one for Math proficiency and one for ELA proficiency. The decision to employ the study's primary factors was chosen because the OLS regressions and scatterplots from the study revealed that these parameters have the greatest impact on proficiency levels and contain the greatest number of samples, allowing for more accurate predictions. Additionally, because race percentiles are being analyzed in the study, it is best to practice incorporating these into the machine learning analysis.

Regression Trees

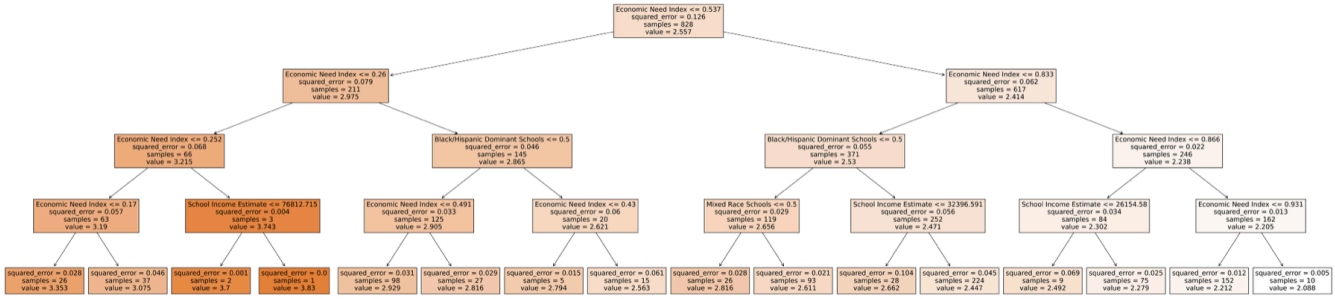
Average ELA Proficiency



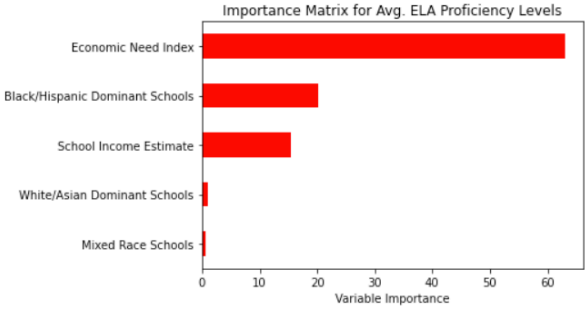
The scatterplot above illustrates the results of a Random Forest Regression with the previously mentioned X variables. When the scatterplot and the regression line are examined, it appears that the data follows a strong positive linear correlation. As a result, the X variables are expected to be strong

predictors of average ELA proficiency levels, which reinforces the study's question and previous findings. Furthermore, it demonstrates to school systems a viable way to increase ELA competence levels.

The mean squared error of the Random Forest Regression for average ELA Proficiency Levels was calculated and found to be around 0.005627. This mean squared error looks to be small enough for the study's query while still having a strong predictive power. Minimizing prediction error, even more, would result in an extremely convoluted tree and an increased likelihood of over-fitting.

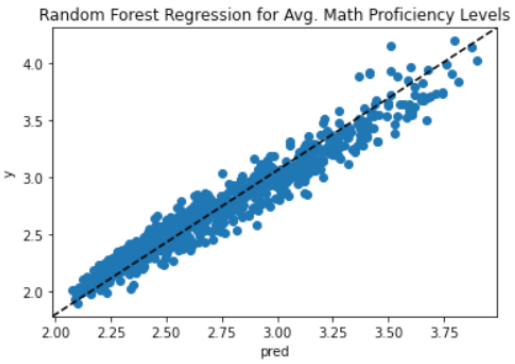


The decision tree regression analysis yielded an interesting discovery for average ELA proficiency levels. To begin with, it appears that 0.537 is a significant economic need index value that affects whether average ELA proficiency levels will be greater than 2.563 if less than 0.537 or lower than 2.816 if larger than 0.547. Overall, it indicates that the best outcome (3.83) for school ELA proficiency levels is if the student body has an ENI of less than 0.252 and a school income estimate greater than $76812.715. The worst outcome (2.088) for school ELA proficiency levels is if the student body has an ENI greater than 0.931. Throughout the tree, it appears that ENI and school income estimates are not the only factors influencing ELA proficiency levels, but also whether the school is Black/Hispanic dominant or mixed race.



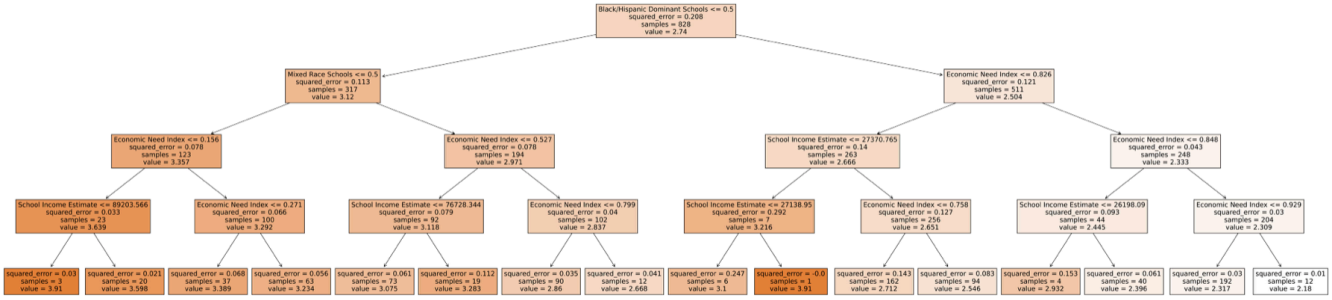
An importance matrix bar-plot was created to better understand which parameters have the greatest impact on the average ELA proficiency levels of schools. According to the plot, the largest factors in ELA proficiency are the economic need index of a school's student population and whether the school is Black or Hispanic dominant or not. These findings are critical for gaining a better grasp of how to enhance ELA competence levels. According to the plot, schools that focus on resolving students' financial problems and diversifying their student population are more likely to enhance average ELA proficiency levels.

Average Math Proficiency

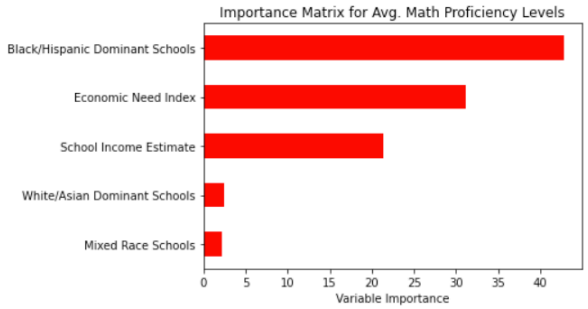


The scatterplot above illustrates the results of a Random Forest Regression with the previously mentioned X variables. When the scatterplot and the regression line are examined, it appears that the data follows a strong positive linear correlation. As a result, the X variables are expected to be strong predictors of average Math proficiency levels, which reinforces the study's question and previous findings. Furthermore, it demonstrates to school systems a viable way to increase Math competence levels.

The mean squared error of the Random Forest Regression for average Math Proficiency Levels was determined to be around 0.01216. Though not as low as that reported for average Math Proficiency Levels, the mean squared error appears to be low enough for the study's inquiry while still having a good predictive ability. Even further minimizing prediction error would result in an exceedingly complicated tree and an increased likelihood of over-fitting.



The decision tree regression study produced an unusual finding for average Math competency levels. To begin, whether the school is predominantly Black/Hispanic or not appears to be a significant divide; however, both findings result in the best outcome for average Math proficiency levels. Overall, it appears that whether the school is Black/Hispanic dominant or not, the best outcome (3.91) for school Math proficiency levels is if the student body has an ENI of less than 0.156 or 0.836 and a school income estimate of less than $892903.566 or greater than 27138.95. If the ENI of the student population is greater than 0.929, the worst outcome (2.18) for school Math proficiency levels is obtained. Throughout the tree, it appears that ENI and school income estimates are not the only factors impacting Math proficiency levels, but also whether the school is predominantly Black/Hispanic or White/Asian.



An importance matrix bar-plot was produced to better understand which criteria have the most impact on the average Math proficiency levels of schools. According to the plot, the most important elements in Math proficiency are whether the school is Black or Hispanic dominant or not, as well as the student population's economic need index. Furthermore, albeit not as relevant as the other two factors mentioned, school income appears to have a rather large impact on Math proficiency, which is not apparent in ELA performance. These findings are essential for understanding how to improve Math proficiency levels. According to the plot, schools that focus on resolving financial issues for students, increasing funding, and diversifying their student population are more likely to improve their average Math proficiency levels.

Comparison

When the two forms of regression analysis are compared, it appears that both procedures arrive at the same conclusion, but each reveals different information along the process. To begin, the OLS regression was successful in finding the presence of linear relationships between the parameters of interest. On the other hand, it was unable to discern other factors of influence, as demonstrated by the regression tree. The regression tree analysis found that whether or not the school was predominantly Black or Hispanic, the amount of school funding, and the economic need index all played a part in determining what would result in changes in average proficiency levels. These findings demonstrate the significant advantage achieved through decision tree regression since it identifies which parameters and the values of those parameters lead to distinct outcomes. The OLS regression, on the other hand, does not represent a specific value in the data that leads to a different outcome, but rather the overall trend of the relationship between parameters. As a result, the economic intuition underlying this discovery is that, while a parameter may be identified as having an impact on an outcome, that parameter may be

influenced by other factors. The conclusion for educational systems is that investing in one parameter may not be as beneficial as investing in multiple parameters that influence each other.

Conclusion

After analyzing the ensuing statistics and graphs of the information present in the PASSNYC dataset, we are better qualified to answer the issue of whether the impact of a school's economic needs affects the Math and ELA proficiency levels of students from all backgrounds. To begin, looking especially at the economic need index, school income estimates, ELA and Math proficiency levels, and the correlation between all parameters, it is reasonable to conclude that economic circumstances do influence the academic performances of all students. According to the findings, a considerable proportion of pupils in New York City school districts appear to have high economic need indices. Furthermore, many of the schools have been identified as having insufficient resources to meet the needs of their student populations. To continue, our research into racial groupings yielded a remarkable revelation. Schools dominated by underrepresented groups, such as Blacks and Hispanics, were found to have higher financial insecurity, lower school funding, and inferior academic performance than schools dominated by Whites and Asians.

To continue, the findings from the maps developed to demonstrate the differences between school districts in New York City revealed several significant conclusions. Overall, the existence of inequality between school districts was detected on all maps. There were significant inequalities in economic need, school income, and academic performance. In addition, maps were created using a new data source that identified attendance and enrolment in each school district from 2010 to 2011. Despite the lack of a clear correlation or pattern, it was discovered that low attendance and enrolment in disadvantaged areas may indicate a link between low proficiency levels, high economic needs, and low

school income. Furthermore, a new data source containing the overall score and overall grade of each school in the New York City school system was web-scraped from the NYC Department of Education. After analyzing the new data and our original dataset, several conclusions were reached. To begin with, it was discovered that many children pursuing an education do not attend a school with a high score, and this was proved to affect their proficiency levels due to the positive correlation that was observed between overall score and proficiency levels. However, no apparent link was established between the overall score, economic need index, and school income estimates. As a result, this research reveals that the student body's economic characteristics and school budget do not limit the school's ability to provide an adequate education to its students. When covariates were considered, it was discovered that White/ Asian dominant schools outperformed Black/Hispanic dominant schools in overall score, demonstrating disparities in opportunity amongst students of different races. To continue, while there may be a correlation between the initial parameters and overall score at the district level, the study of the map on the average overall score by school district demonstrated inequalities in the type of education received by children in each district.

Furthermore, an OLS and tree regression analysis was performed to better evaluate any correlations between economic and ethnic characteristics and average Math and ELA proficiency levels. To begin with, the OLS regression was successful in discovering linear relationships between parameters, all of which were supported by large r-squared values. A negative linear association was discovered between economic need indices and average proficiency levels, as well as school income estimates and economic need indices. A one-point increase in ENI, for example, results in a 1.31 drop in average ELA proficiency levels and a 1.50 drop in average Math proficiency levels. Furthermore, a one-point rise in school income estimates resulted in a 0.000009309 fall in the economic need index. Estimates of school income, on the other hand, were shown to have a positive linear relationship with

average proficiency levels. A one-point increase in school income is associated with a 0.00001134 increase in average ELA proficiency levels and a 0.00001261 increase in average Math proficiency levels. It should be noted that schools are often funded with thousands of dollars; hence, even if the impact is minimal, significant funding can have a large impact on proficiency levels and ENI. To continue, the regression tree analysis found that whether the school was predominantly Black and Hispanic, as well as the amount of school funding, had an important influence in determining what degree of economic need index would result in changes in average proficiency levels. It was discovered that schools with a high ENI that were predominantly Black and Hispanic outperformed schools with a low ENI that were not predominantly Black and Hispanic in terms of average proficiency levels. The economic theory underlying these findings is that they demonstrate how educational systems can improve the average proficiency levels of their schools. The findings indicate that favourable economic policies, such as increasing school funding and assisting students with financial insecurity, and diversity policies, such as creating a more diverse school environment and providing additional assistance to schools dominated by Blacks and Hispanics, will positively benefit average proficiency levels.

The significance of these findings is that they show that students' academic success is linked not only to their performance in school but also to other circumstances beyond their control. These findings are critical for educational leaders who want to see a more successful and diverse student body. For example, based on the findings of this study, an increase in school funding or addressing the financial hardships faced by students is likely to help students perform better academically. Furthermore, providing funding and financial security not only helps all students perform better but also gives underrepresented groups a higher chance of academic success. Additionally, because mixed-race schools fared better in all metrics tested than Black and Hispanic dominated schools, one probable inference is that schools with more diversity will do better in all aspects. In conclusion, economic and racial issues

likely influence all students' academic performance. However, it is not only up to students to perform better in school, but it is also up to educational authorities to support them in doing so.

Distinguishability

It is a valuable practice to state the primary findings that separate this research from other publications in similar literature at the end of the study. To begin with, a thorough examination of the impact of the economic need index and school income estimates on average Math and ELA proficiency levels provided a more credible answer. The study's question was studied from many angles using graphs, maps, and regression analysis. This is a significant distinction from other publications since it allows for a clearer understanding of the relationship between a student's performance and the economic dynamics at work. To continue, with the inclusion of economic and racial aspects in the study, a more in-depth investigation of what genuinely influences student performance could be conducted. The majority of the relevant literature studied the influence using a single component. However, because educational systems are tremendously complex, a single-factor analysis may not be sufficient to explain changes. In addition, new criteria of interest were included to enable a deeper investigation into what other influences can impact proficiency levels. Exploration of other parameters will result in more trustworthy responses to the study's questions because it allows fewer possibilities for third-party impacts.

Future Works

Even though this paper produced substantial findings that helped answer the question of what effect economic need and school income have on the average Math and ELA proficiency levels of New York City students as a whole and by Race? Based on this study, it is expected that there will be more to investigate in future work. To begin with, one key concern that this paper may raise is the lack of a large sample size. Because this study was focused primarily on New York City schools, which were then

filtered to include those with the most relevant data to the study, the inclusion of more school systems would provide a more robust answer to the study's inquiry. For example, following data cleansing, only 828 of the initial 1272 schools remain, implying that 444 schools were deleted due to insufficient data. Because of the limited sample size, the findings in the research may not fully represent the entire New York City school system. To continue, it was found that the standard deviation for the percentage of Black/Hispanic and White/Asian students at each school appeared to be large. As a result, when the covariates were investigated, the results may not have accurately reflected the genuine variations across race percentiles. As a result, the inclusion of more schools will allow for a more thorough examination of the parameters of interest. Additionally, with larger sample size, a stronger regression analysis might be performed to further quantify the effect of economic and racial characteristics on average Math and ELA proficiency levels, potentially leading to other regression studies such as IV and difference-in-difference. To continue, the paper has limitations, such as relying solely on public data, which may exclude characteristics that might have a greater influence on proficiency levels. School income estimates, for example, may not fully reflect the number of government resources provided to schools. These constraints can be addressed in future work by collecting data from additional educational systems as well as directly from schools and government institutions. This direct data can include, for example, the previously mentioned government resources as well as other essential factors such as teacher and student relations.

References

Boardman, J. D., & Saint Onge, J. M. (2005). Neighborhoods and Adolescent Development. *Children, youth and environments*, *15*(1), 138–164.

Lee, K. M., & Slate, J. R. (2014). Discrepancies in Advanced Success Outcomes for Texas Students as a Function of Economic Disadvantage. *Journal of Education Research*, *8*(3), 137–149.

Ross, C. E., & Willigen, M. V. (1997). Education and the subjective quality of life. *Journal of Health and Social Behavior*, *38*(3), 275. https://doi.org/10.2307/2955371

Walsh, M. E., Madaus, G. F., Raczek, A. E., Dearing, E., Foley, C., An, C., Lee-St. John, T. J., & Beaton, A. (2014). A new model for student support in high-poverty urban elementary schools. *American Educational Research Journal*, *51*(4), 704–737. https://doi.org/10.3102/0002831214541669

White, G. W., Stepney, C. T., Hatchimonji, D. R., Moceri, D. C., Linsky, A. V., Reyes-Portillo, J. A., & Elias, M. J. (2016). The increasing impact of socioeconomics and race on standardized academic test scores across elementary, middle, and high school. *American Journal of Orthopsychiatry*, *86*(1), 10–23. https://doi.org/10.1037/ort0000122