A WALK TO SCHOOL

AN ANALYSIS OF NYC HIGH SCHOOL PERFORMANCE FACTORS

TEAM 23

ESSENCE CARSON
JARRELL COOPER
BRYAN GARCIA
NOAH MORTON

Table of Contents

Table of Figures

Project Purpose

Background

New York City is the largest public school district in the United States with 1.1 million students enrolled in K-12 schools across the city[1]. Out of this number, over **290,000 students** go to one of the 486 public high schools across the five boroughs. Though each student goes to high school, their performance varies dramatically due to different factors. One school boasts high graduation rates in one district, while a similar size district struggles to meet state standards. The student achievement gap continues to expand in high school performance.

Our project investigates the factors that influence high performance in New York City high school. All our project data comes from the New York City Department of Education (NYC DOE). We measure high performance based on the data collected for four-year high school graduation rates. Thus, we begin with our question: **What factors influence high performance in New York City public high schools?**

[1] (NYC Department of Education, 2020) - https://www.schools.nyc.gov/about-us/reports/doe-data-at-a-glance

Data Wrangling/Cleaning

Finding an appropriate data source

Thinking about the possible factors that may contribute to high performance (read: graduation rate) in New York City public high school, we searched for a dataset for our education project. Our primary database is the following:

New York City Department of Education Info Hub

The New York City Department of Education (NYC DOE) Info Hub provides exportable data (.xls) on New York City public schools. Based on preliminary analysis, the data from this Info Hub is based on data submitted by the school districts and public schools within New York City. One advantage of this dataset is that it provides rich information on New York City public school enrollment, demographic, graduation rates by the cohort, disciplinary action (suspension, expulsion, etc.), attendance, school funding, and regents (high school standardized exams). One disadvantage of this data set is that it contains different variables across different excel files. Since each school has an unique identification code (DBN and School Code), linking the data sets across excel via Pandas is possible. Cleaning the data set and creating a comprehensive data frame to respond to our project question about the variables that influence high performance factors is essential for the project.

Choosing appropriate datasets

From this database, we focused on two primary datasets to build a data frame that would best be suited for our project. Our first dataset came from the school quality reports which shared enrollment, school name, DBN (school unique identifier), and demographic data of the New York City Public School. NYC DOE only collects racial data in the following categories: White, Asian, Black, and Hispanic.

In addition, this dataset also shared the Economic Needs Index, which is an index created by NYC DOE to measure the economic need based on the social situations of the students and their families. The four criteria:

(1) Be eligible for Human Resources Administration (HRA) benefits (Public Assistance)
(2) Have lived in temporary housing in the last four years
(3) Have entered the NYC DOE for the first time within the last four years, is enrolled in high school
(4) English Language Learner (has a home language other than English)

Thus, a high ENI could be a factor that influences high performance in New York City public high schools due to the social conditions of the students and families.

However, our first dataset did not include school funding, which may provide more context to measure the relationship of how much a school receives compared to their high performance. As a result, we merged the School Transparency Dataset from the NYC DOE via the unique school code identifier (District Borough Number, DBN) to create a more comprehensive data frame for our project analysis.

Merging project data frame

Stage I: *What is public High School? How are we defining public High School in NYC?*

- A public high school in NYC has more than zero enrollments in grades 9-12. We will use the (bottom of the screen "High School" to find the total number of High Schools in NYC. Data is found in the **Summary Tab**

Result #1: Four columns of new data frame: DBN, the name of school and Enrollment of all the NYC public High Schools. Add **2018-19** school Year column. Use Demographic Snapshot data to confirm the consistency of the data with School Quality Report.

DBN	School Name	Year	Enrollment

Stage II: *Who attends NYC public high school and how are they identified by NYC DOE?*

- Using the School Quality Report, **Summary tab**, we pull the Racial Percentages and the Economic Needs Index of each school to build the School View. The data is represented in percentages. NYC DOE defined the following racial categories: **Asian; Black; Hispanic; White**. We hope to continue to interrogate not only the ethics within data collection that allows us to see the multiple ways in which people can embrace a pluralistic racial identity. With this in mind, we cautiously use the NYC DOE data collection on racial categories.

- Using the same dataset, we pulled the Racial Percentages and Economic Need Index

Result #2: Two additional columns: DBN/School Name/Year/Enrollment + Racial Make-up (Asian/Black/Hispanic/White//Economic Need Index

DBN	School Name	Year	Enrollment	Racial Make-up %	Economic Needs Index

Stage III + IV: *Graduation rate across NYC public high school. How are we defining Graduation? What is the graduation rate in NYC public high schools?*

- We recognize that graduation rates differ depending on how many years it takes to graduate NYC public high school as well as the number of students graduating in each cohort. Our graduate rate for NYC public high schools is the **4-year graduation rate in August** measured by cohort, who entered NYC public high school in 9th grade. Though it is common for schools to end in June with regent exams and graduation, there are exceptions where one would graduate in August. We will use the School Quality Report to find the four-year graduation rate across NYC public high schools.

Using the reduced dataset from 3, Run code to pull columns of "Metric Value 4-year graduation rate -All students" (tab "Student Achievement, column BG). Link via DBN code

Result #3/4: One additional column to data frame: DBN/School Name/Year/Enrollment/Racial Make-up/Economic Needs Index + Grad%

DBN	**School Name**	**Year**	**Enrollment**	**Racial Make-up**	**Economic Needs Index**	**Grad %**

Stage V: *What is the NYC public High School Funding Situation? How much funds are given to each public high school? Where is the funding (Local, State, Federal) coming from?*

- We will use the Financial data Transparency Data Report from the NYC DOE to pull data regarding School Funding Sources as well as (potentially) funding per pupil.

- Run code to narrow **Financial Transparency** dataset to only Total Funding. Link to DBN of NYC public high schools, **pull funding data**)

Result #5: Three additional columns: DBN/School Name/Year/Enrollment/Racial Make-up/Economic Needs Index/ Grad % + Total Funding/ Total Funding per Pupil

DBN	**School Name**	**Year**	**Enrollment**	**Racial Make-up**	**Economic Needs Index**	**Grad %**	**Total Funding**	**Total Funding per pupil**

Imputing Values for Total School Funding & Total School Funding per Pupil

After merging the datasets to create a comprehensive data frame, we noticed that there were some Null values present in our data frame. Out of the 486 New York City public high schools, only 65 school's Financial Funding Data was missing. Since only 13% of the financial data was missing, we decided to impute Total Funding per Pupil (K) with the median since that difference between the mean and median were only a $400 difference. For the other 3 funding, we imputed with the average since their mean was greater than the medium by at least $100M. These imputing methods provide better accuracy for our project than just using an arbitrary value. It is important to note that ENI and demographic data did not have any Nan values.

Table 1: Funding simple statistics

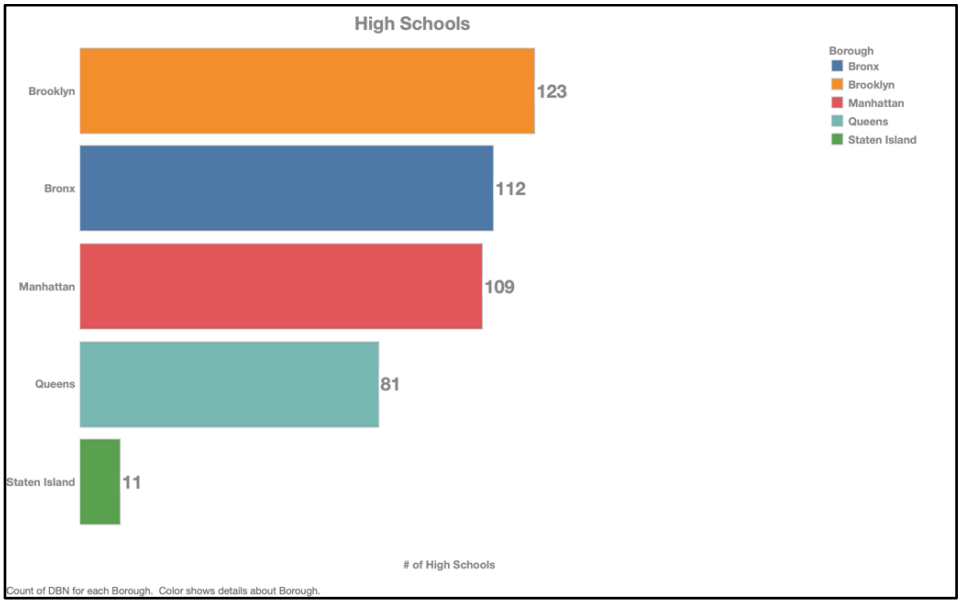
	State & Local Funding (M)	Federal Funding (M)	Total Funding (M)	Total Funding per Pupil (K)
Mean	9.625	0.537	10.162	22.732
Median	6.772	0.405	7.194	22.314

Exploratory Data Analysis

Painting the Picture

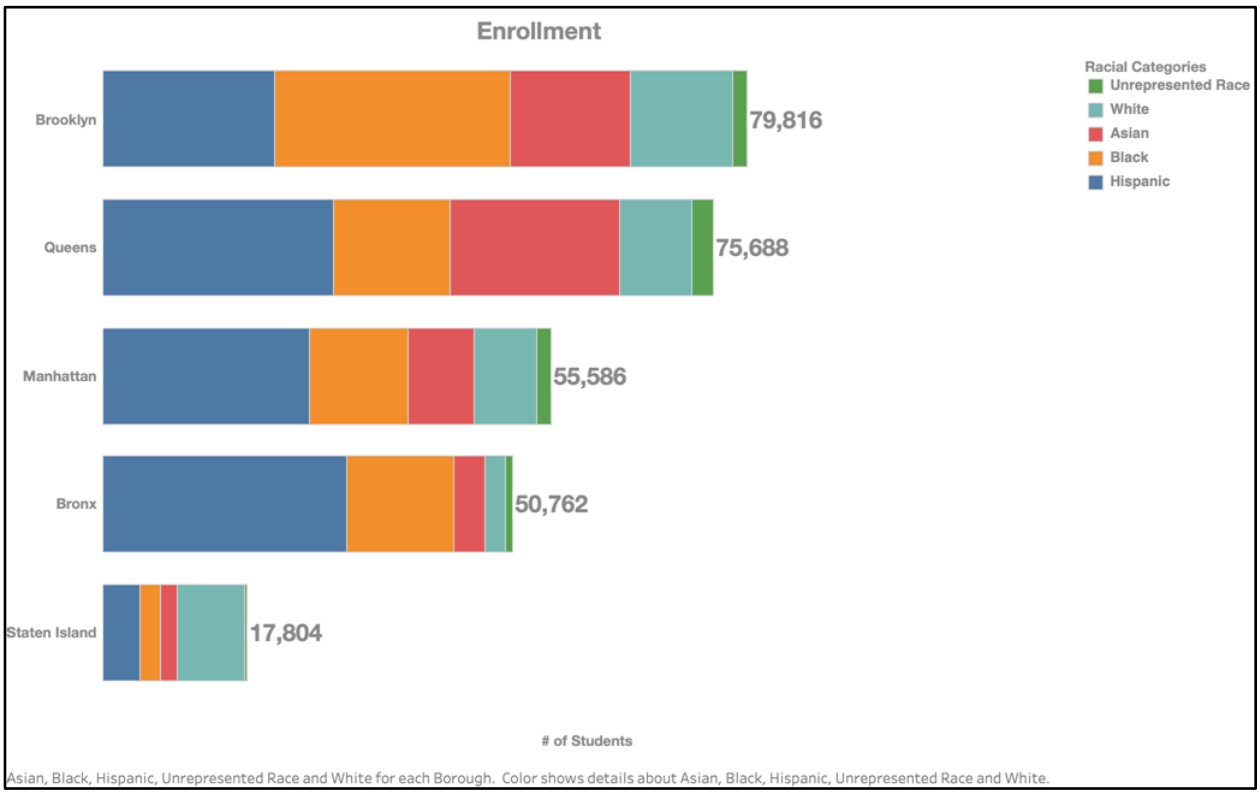
From our thoroughly cleaned data frame, we continue with the exploratory data analysis of our project to respond to the question: what factors influence high performance in NYC high school? As mentioned before, there are 486 public high schools across the five boroughs in New York City. The distribution of the public high school is represented in the chart below:

Table 2: NYC High School Distribution Table



The number of high schools are not evenly distributed among the boroughs. Brooklyn (123) has the highest schools. After Brooklyn, Manhattan and Bronx are tied with 112 schools. Investigating student graduation rate encourages us to ask questions to learn more about the 290,000+ students that are enrolled within NYC public high school. This is seen in our next table:

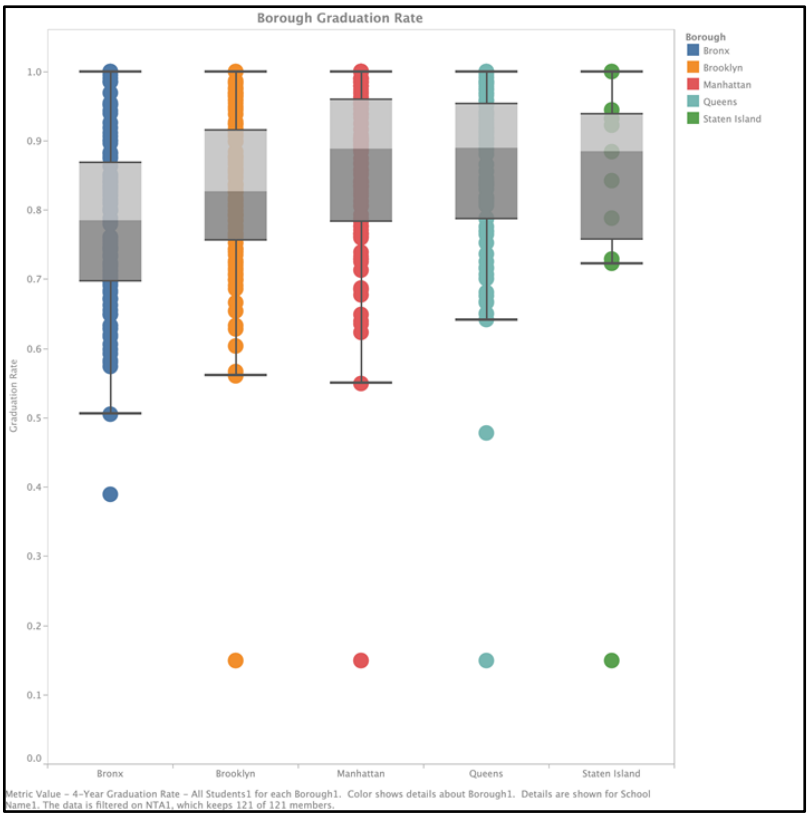
Table 2: NYC High School Enrollments



Graduation Rate Boxplots

Using a boxplot, we analyzed graduation rates across boroughs. The first clear insight from the plot is that Bronx and Brooklyn have the lowest medium graduation rates. These two boroughs also happen to have the highest number of high schools in NYC. Further analysis shows that 20% of students do not graduate.

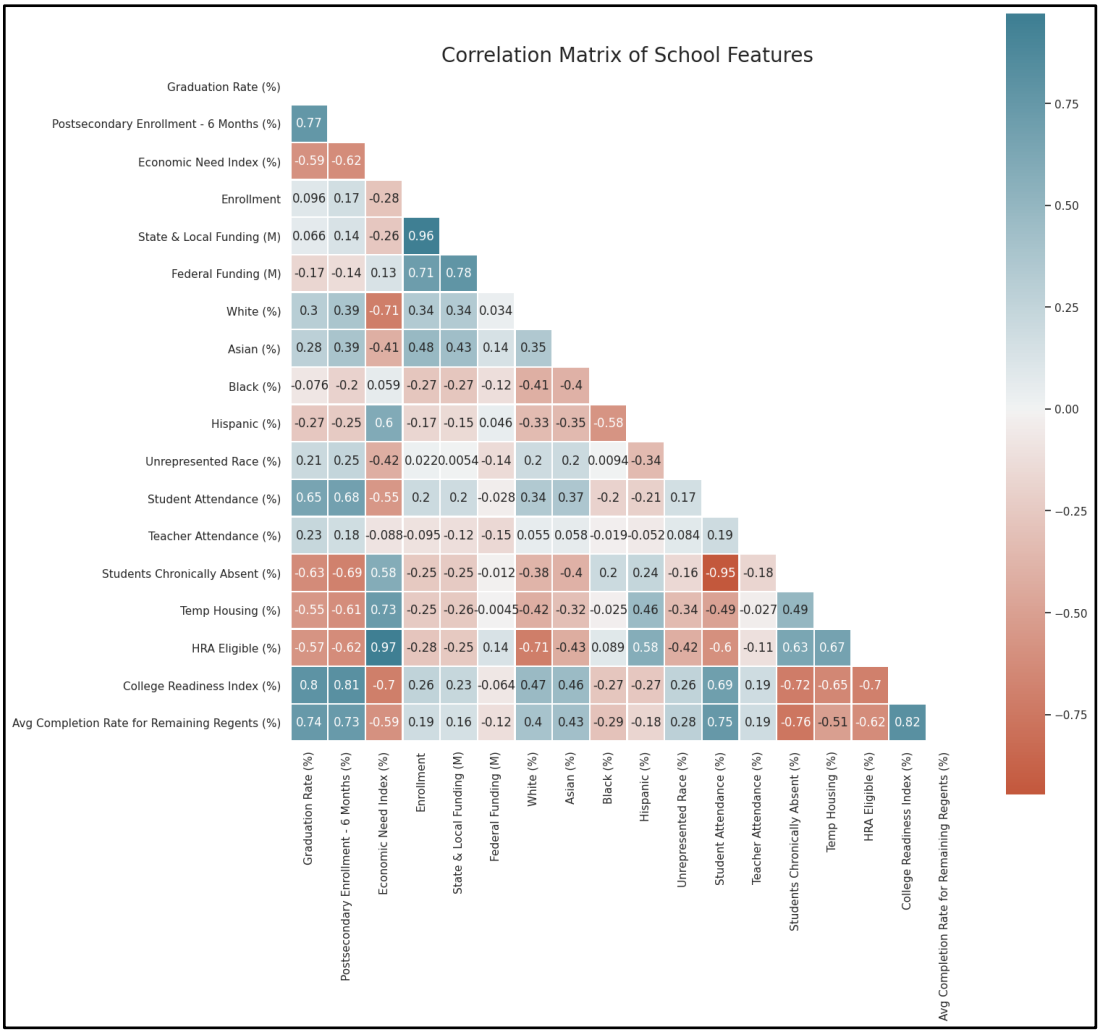
Table 3: Boxplot Graduation Rates Across Boroughs



Correlation Matrix

With the knowledge of the school, student, and graduation rate overview of our education project, we return to our question: what factors influence high performance (read: graduation rate) across New York public high schools. The following correlation matrix provides further insight to this response:

Table 4: Graduation Correlation Matrix

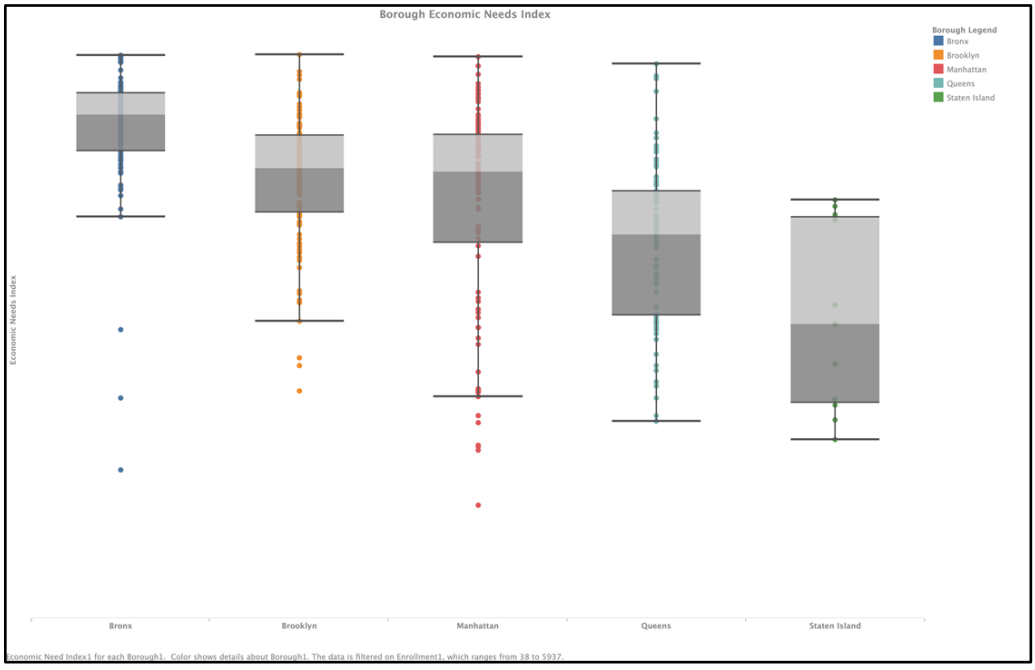


We notice how the Economic Needs Index is highly negatively correlated to the four-year graduation rate, which is the indicator we are using for our project on the factors that influence high performance in New York City public high schools. In addition, a high ENI is negatively correlated to College Readiness Index, Student Achievement as well as the Percentage of Whites and Asians in a school. However, ENI is positively correlated to the percentage of students in temporary housing as well as the percentage of Hispanics in a school.

Recognizing how the ENI refers to economic disparity, the amount of support on public assistance, multilingual families, and new as well as describes the social context of the student, it

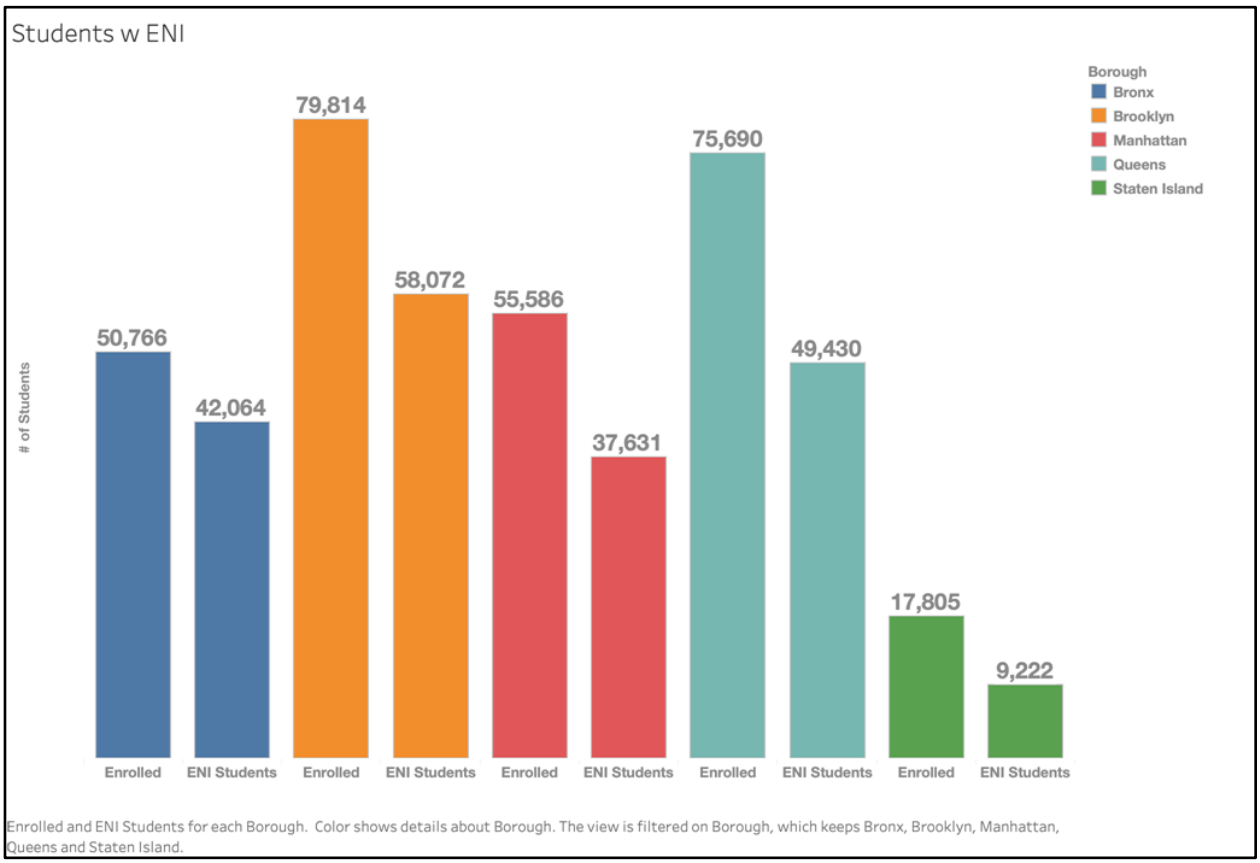
is interesting to witness these correlations. We question: if ENI indicates economic need, in what ways is Total School Funding distributed in relation to ENI? What would this correlation look like across boroughs? Though ENI and Total Funding is negatively correlated, we are curious to investigate different data visualizations to provide a more interactive approach to our project.

Table 5: Boxplot ENI Across Boroughs



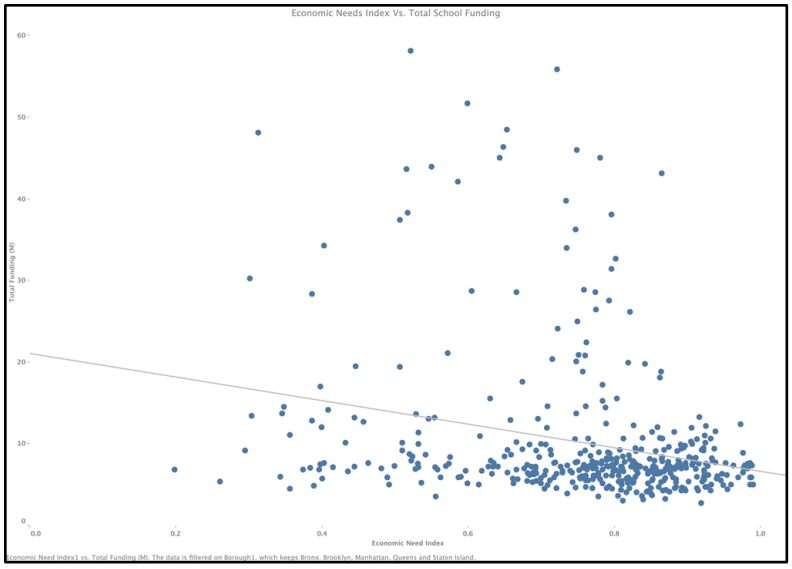
From the ENI Boxplot, it is clear the ENI for schools in the Bronx (~.85) is the highest on average than the four other boroughs. Brooklyn (~.79) and Manhattan (~.78) show similar ENI results. Though Queens (~.65) is lower than the first three boroughs, there exist a significant gap in ENI compared with Staten Island (.55). Bronx, Brooklyn, Manhattan and Queens all have schools with an ENI close to 1, which further reveals the Economic Need of the Schools in that area. We created another data visualization to illustrate how many students have economic need across the five boroughs.

Table 6: Number of Students Economic Need



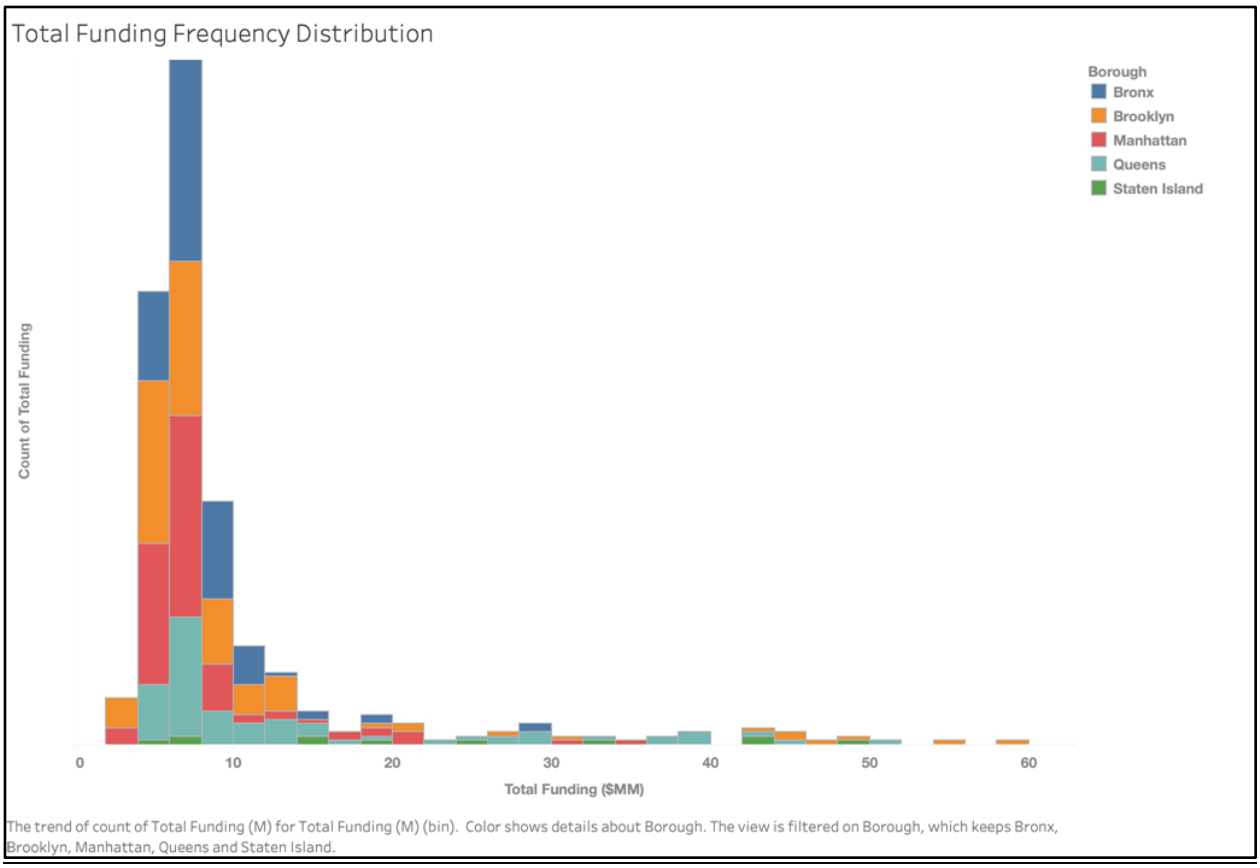
Recognizing how the Total Funding for Schools is higher in Queens and Staten Island despite their relatively "low" ENI, it is important to question the relationship between funding and ENI to observe the different ways Schools with high Economic Need can receive Economic Support.

Graph 1: Total Funding and ENI Linear Regression



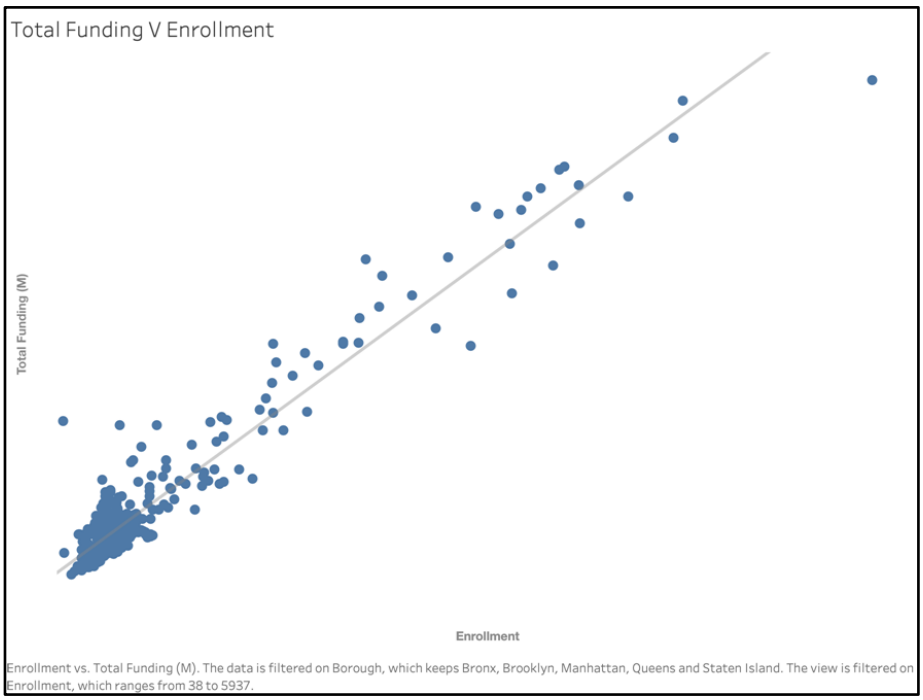
Total Funding and ENI are negatively correlated, which means that funding is not distributed based on economic need. If economic need is not supported economically through funding, then what determines the distribution of funding. Thus, we analyze the total distribution per pupil:

Table 7: Histogram of Total Funding Across Boroughs



At first glance of the distribution of Total Funding per Pupil, it seems that since the funding is equally distributed then NYC public high schools have achieved an equal, balanced education system. However, ENI is not equally distributed as we have witnessed in the Boxplot earlier. Thus, each pupil/family/school may come from a social context and experience structure in a way that increases their Economic need. As demonstrated below Total Funding is linked to Student Enrollment:

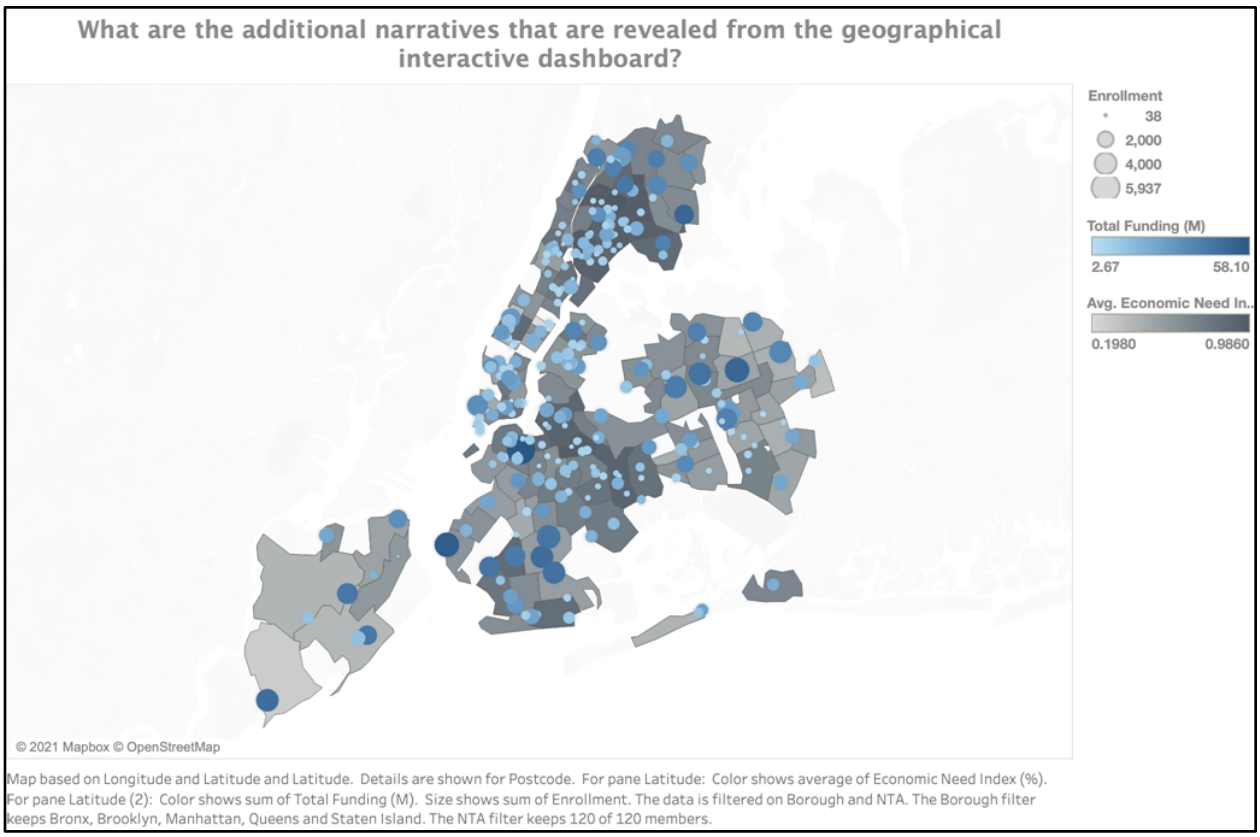
Graph 2: Total Funding and School Enrollment Linear Regression



Though a pupil may go to a school with a high economic need, it receives the same support as another child who attends a school with a lower economic need. As a result, equal funding across pupils in NYC further exacerbates the social inequality that may influence a high ENI.

Recognizing the interlocking nature of education, which is woven into different aspects of social inequality, we realize how certain factors may simultaneously influence each other. Thus, a high ENI rate may be correlated to a high percentage of students in a school community in temporary housing, which may influence school's attendance rating, which is further correlated with high performance (read: graduation rate). Witnessing the interconnectedness of the factors, we hope to illustrate alternative ways of conceptualizing NYC high school education throughout our interactive dashboard:

Image 1: Interactive Tableau Dashboard



SCAN ME

Recommendations and Considerations

From our data analysis of the factors that influence NYC public high schools, we observe that distributing funding solely based on Total Enrollment disregards the various economic needs of the schools and students. We recommend that a distribution of funding that centers ENI can encourage and increase graduation rate across NYC public high schools. In addition, we hope that our analysis can promote more consistency regarding the funding policies according to ENI across both State & Local as well as Federal Funding.

We realize that graduation rate factors are not singular: A high ENI rate is correlated to a high percentage of students in a school in temporary housing, which may influence a school's attendance rating, which is another factor that is negatively correlated with the graduation rate. Thus, a collective approach is essential to understanding the interlocking confounding factors that influence high performance in NYC public high schools.

References

New York Department of Education. (2020). *DOE Data at a Glance*. Retrieved from: https://www.schools.nyc.gov/about-us/reports/doe-data-at-a-glance