Data 119: Introduction to Data Science II

Analysis: Inequity in the Chicago Public School System

Micayla Roth and DB Christenson

Overview

The Chicago Data Portal[1] provides datasets on the city of Chicago's public operations. Thousands of datasets are available relating to art, public safety, finance, and more. The datasets in this analysis are focused on the annual reports on the Chicago Public School System (CPS) and will use the aggregated data of the CPS Progress Report Card and CPS School Profile datasets from 2021.

Past analyses have used CPS school-level data to analyze discrepancies between schools in the district. Kralis (2022)[2] used demographic, attendance, and college enrollment data from 2015-2020 report cards and created a k-means algorithm to classify schools into three clusters. The first cluster was majority Black or African American (89.20 percent) and low-income (84.46 percent), the second majority Hispanic or Latino (81.30 percent) and low-income (89.21 percent), and the third had no racial majority and about half low-income (54.30 percent). He also found that the correlation between proportion of students that low-income and proportion that enroll in college (-0.644) was nearly as strong as high school graduation rate and proportion enrolled in college (0.771). This analysis suggests that demographic composition metrics are key features that differentiate public schools within the city of Chicago, and that some relationship (whether correlational or causal) exists between school background characteristics and real-life outcomes.

In our analysis, we will be focusing on the characteristics and predictive power of school type. CPS categorizes schools according to 13 different types (Figure 1). Almost 95 percent of schools classified as neighborhood, charter, magnet, citywide-option, small, or selective enrollment; of these, 82 percent of schools are either neighborhood or charter. Neighborhood schools offer automatic admittance to students within their attendance boundaries, but students can apply to such schools from across the city. Charter schools are authorized by the Chicago Board of Education but are operated independently[3].

Our analysis will look at variations in the qualities of these schools in order to determine how the type of school that a student attends can impact their future.

Methodology

The data in this report is made available through the City of Chicago's public data portal from two datasets titled: "Chicago Public Schools - School Progress Reports SY2122"[4] and "Chicago Public Schools - School Profile Information SY2021."[5] These datasets were then merged to contain information from both datasets on schools with IDs present that are present in

---

[1] "City of Chicago: Data Portal: City of Chicago: Data Portal." *Chicago*, data.cityofchicago.org/.

[2] Kralis, M. (2022). Education Disparity in Chicago Public High Schools: A Statistical Analysis. *Socially Responsible Modeling, Computation, and Design*, *2*(1). https://doi.org/10.18409/soremojournal.v2i1.214

[3] "Program Types." *GoCPS*, https://go.cps.edu/program-types.

[4] "CPS - School Progress Reports SY2122" was published by Johnathon Levy with data provided by the City of Chicago and was last updated May 9th, 2022.

[5] "CPS - School Profile Information SY2021" was published by Ted Canji with data provided by the City of Chicago and was last updated October 17th, 2021.

both datasets with duplicate school IDs removed. Because the metric being looked at is student outcomes, only schools that were considered high schools were included in the data.

School_Type	Proportion Low Income	Culture Climate	Student Attendance (%)	4-year Graduation Rate (%)	Student College Enrollment (%)
Career academy	0.889358	Very Well Organized	76.975	78.225	62.725
Charter	0.813863	Organized	79.44	78.360465	44.153659
Citywide-Option	0.770378	Organized	66.315152	23.747059	23.284
Classical	0.213120	Organized	98.066667	Not enough data	Not enough data
Contract	0.733519	Organized	88.025	88.75	70.35
Magnet	0.579245	Organized	92.695349	82.116667	72.228571
Military academy	0.828236	Very Well Organized	85.75	88.3	65.5
Neighborhood	0.716538	Organized	89.713924	74.076744	60.535714
Regional gifted center	0.319499	Organized	95.61	Not enough data	Not enough data
Selective enrollment	0.542590	Organized	89.890909	93.336364	88.609091
Small	0.801610	Organized	83.592	73.115385	61.515385
Special Education	0.586258	Organized	80.685714	3.8	6.1
Virtual	0.772465	Organized	Not enough data	Not enough data	Not enough data

Figure 1: Mean values of proportion low income, culture climate rating, student attendance, graduation rate, and college enrollment for schools of each type.

Results

How are Student Outcomes Measured?

The combined data contain nearly 200 metrics including demographic information, outcome measures, and school quality indexes at the school level. Because any measure has limitations in its ability to capture student success, we will use a variety of measures that represent outcomes both while in school and in college. SAT scores are commonly used to compare students from different schools on the same scale[6]. Because many of these schools lack SAT data, we will look at student attendance as a measure of time spent in school, school quality measures created from survey data, and relationships to future college enrollment.

Background Statistics

To understand who has access to schools of each type, we first examined the proportion of each type that is composed of low-income students. For ease of interpretability, we eliminated school categories with ten or fewer schools. Figure 2 shows this distribution.

Most public schools in CPS have a primarily low-income student population. All six school types have at least one school that is at least 60 percent low-income, and most charter schools and neighborhood schools are composed as such. 22 neighborhood schools in the district are 30 percent low-income or less; because neighborhood schools primarily serve students in close geographic proximity to them, these are likely located in more affluent neighborhoods in

[6] Burton, Nancy W., and Leonard Ramist. *Predicting Success in College: SAT® Studies of Classes Graduating Since 1980*, files.eric.ed.gov/fulltext/ED562836.pdf.

the city, although this is beyond the scope of this analysis. Future research may look at affluent neighborhoods and the distribution of students sent to both public and private schools.

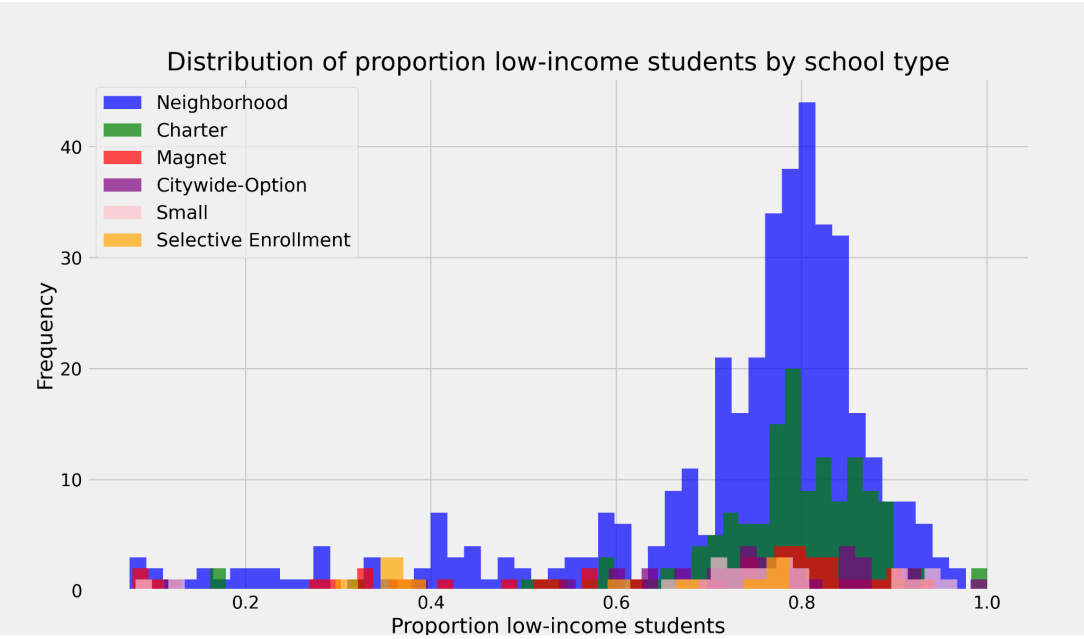

Figure 2: proportion of students that are low-income, by school type. We chose to use proportion rather than pure number of students because large discrepancies in student population between elementary and high schools reduce the comparability of low-income representation between schools.

Magnet schools and selective enrollment schools have more variation in low-income populations; however, their smaller sample sizes make it more difficult to make generalizations. Five out of eleven selective enrollment schools are 40 percent low-income or less, while four are 70 percent low-income or more. While some of these schools provide similar representation to that of more widely available neighborhood and charter schools, half do not reflect the public school population, thus limiting opportunities for lower-income students.

School Quality Measures

CPS progress reports quantify school quality via the 5Essentials Student Survey, the 5Essentials Teacher Survey, and the My Voice My School Parent Survey[7]. The progress report data categorizes schools on a five-point scale ranging from "Very weak" to "Very strong" on five measures: effective leaders, collaborative teachers, supportive environment, involved families, and ambitious instruction. We created a composite school quality variable for each school that represents the average of these five metrics. Figure 3, below, presents the proportion of schools in each type that fall into each classification.

[7] "My Voice, My School Surveys." *CPS*, https://www.cps.edu/about/district-data/metrics/surveys/.

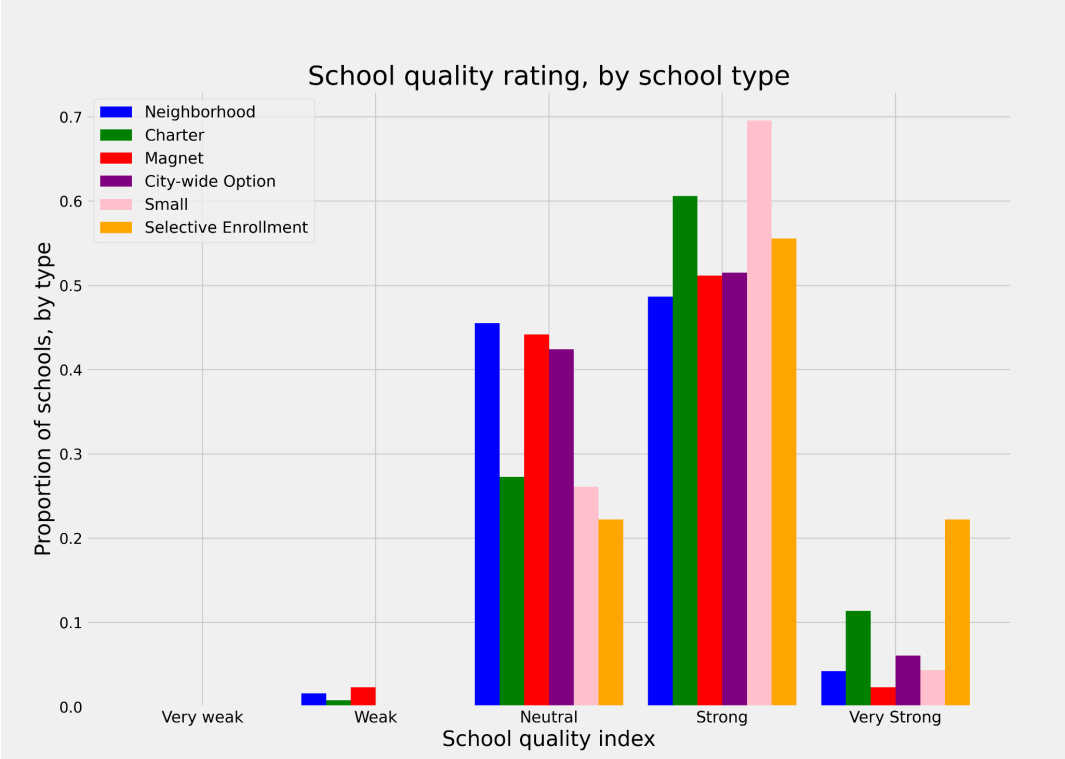

Figure 3: school quality rating, by school type.

Most schools are either "Neutral" or "Strong;" none are very weak on average, but only a small proportion are considered "Very Strong." Selective schools by far are most likely to be very strong, which suggests that students who are admitted to these schools have access to a higher quality education. Neighborhood schools are almost evenly split between neutral and strong, and charter schools seem to be more likely to provide a higher-quality education. Magnet schools do not seem to provide a significantly better education than charter schools or even than neighborhood schools; however, small schools seem to overwhelmingly foster a positive learning environment. The graph suggests that small and selective schools have the most supportive environment for students, but these schools are inherently less accessible than other options throughout the city.

Analyzing General Trends in Student Outcomes

Figure 4, below, shows the relationship between college enrollment rate and student attendance; a good predictor of outcomes and a fair statistic to measure it against. Students who pursue higher education are more engaged and more likely to attend school. Alongside this data another metric, school type, is established to display the different distributions between school types.

Each school type has its own unique stratification across the graph. Almost every citywide option is far below the mean rate, but selective enrollment schools clear it quite comfortably. Neighborhood schools show a strong correlation between student attendance and college enrollment rates. Charter schools seem to have most of their weight at the bottom or at

the top with less in the middle of the other schools. There seem to be very clear differences in the behavior of the students and in the student outcomes of each school type.

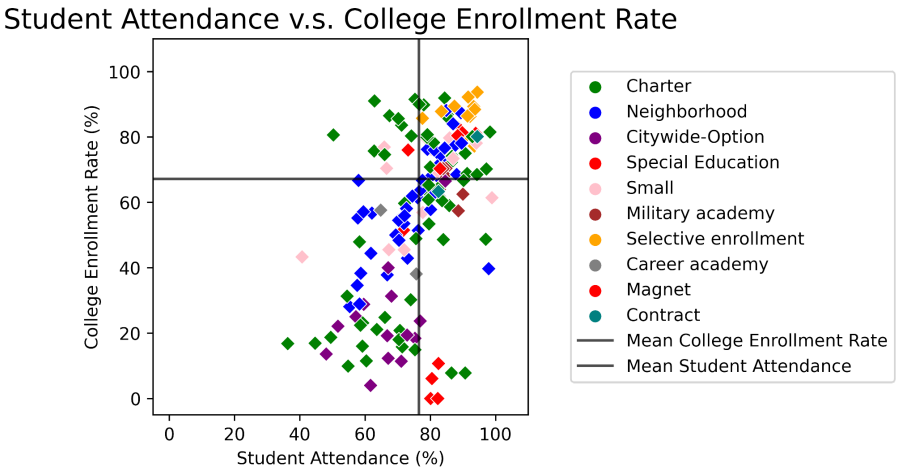

Figure 4

Understanding the Importance of School Type using Neural Networks

Using different machine learning models and training them using different data allows us to see the impact that the type of school that a student attends has on correctly classifying certain metrics. One example of this is using classification models to identify whether a school's college enrollment rate is higher than the CPS average college enrollment rate. A neural network model was fed two sets of data, one with school type included and one without. Then it was made to go through a set of test data and predict, for each school, whether its college enrollment rate was above the CPS average or not.

Figures 2A and 2B, shown below, are confusion matrices describing the accuracy of each model. The value 0 represents a school's college enrollment rate being below the mean, meanwhile a value of 1 represents a school's college enrollment rate being above the mean.

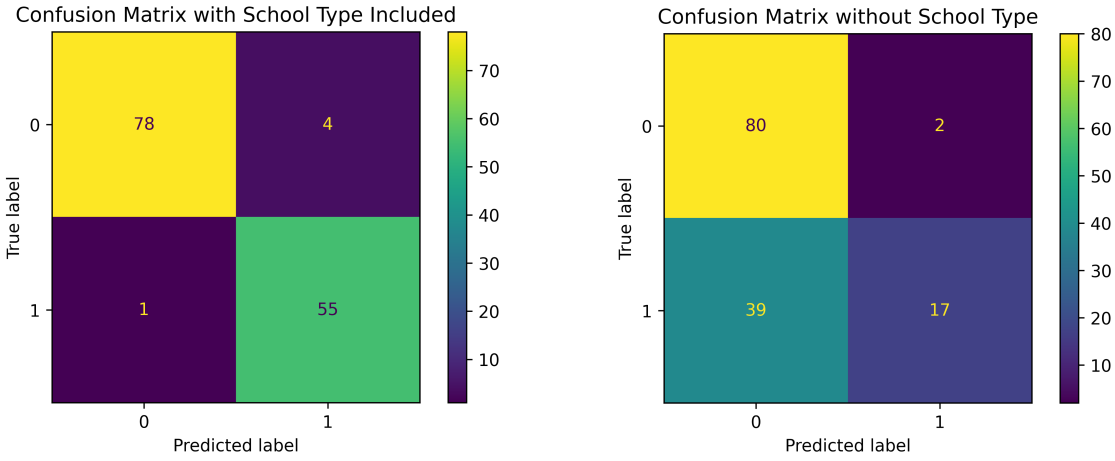

Figure 5A, 5B

Including school type in the training data allowed the model to make significantly more accurate predictions about schools with a college enrollment rate that is above the CPS mean. Without including the school type in the training data, the model was accurate just 70.2% of the time, while including the school type meant it was accurate 96.3% of the time. Because the model performed better when given the school type parameter, that feature seems to hold significant importance when it comes to how the model decides whether the given school was above the CPS mean or not.

Factors that may explain why school type is an important feature to consider are the funding given to the school, the accreditation and passion of the faculty, the types of families that surround the students, and even more confounding variables. All these elements severely affect how a student views education and its importance. Therefore, the model interprets school type as a useful metric for gauging the average student's ambitions at the school. At a selective enrollment school, many students have the goal of attending post-secondary education, while at a city-wide school, students are less likely to see college as an option because it is either unattainable or because a weaker educational background does not stimulate the same educational goals.

Conclusion

Our analysis shows that there are tangible differences between schools of different types. Neural network algorithms more accurately predict college enrollment rates when school type is included, providing evidence that attending these schools opens up access to future opportunities. These schools score highest on all five quality measures, and students attending them spend more time attending school. They're also less likely to have larger proportions of low-income students. Future research should investigate whether discrepancies in outcomes are directly determined by qualities of the schools, or whether they are a product of self-selection of the students in accordance with motivation, family educational background, and access to supplementary educational resources and opportunities. This can be done by focusing on schools with a wide distribution of socioeconomic backgrounds; this would require a dataset extending beyond Chicago to provide a large enough sample size. Finally, looking at funding distribution across the city could provide insights on whether resources are equitably distributed to neighborhoods and students who need them most.

While the city provides options for higher quality schooling, these opportunities are not available to all; understanding how resources are distributed and who can access them will contribute to a better understanding of how education and child development support can become more equitable across Chicago and the rest of the country.

Works Cited

Burton, Nancy W., and Leonard Ramist. *Predicting Success in College: SAT® Studies of Classes Graduating Since 1980*, files.eric.ed.gov/fulltext/ED562836.pdf

"City of Chicago: Data Portal: City of Chicago: Data Portal." *Chicago*, data.cityofchicago.org/

"My Voice, My School Surveys." *CPS*, https://www.cps.edu/about/district-data/metrics/surveys/

Kralis, M. (2022). Education Disparity in Chicago Public High Schools: A Statistical Analysis. *Socially Responsible Modeling, Computation, and Design*, *2*(1). https://doi.org/10.18409/soremojournal.v2i1.214