Enhancing Decision-Making for Parents and Authorities

A Comprehensive Analysis and Mapping of School Performance in New York City

MGT 6203 - Summer 2023


Shashikanth Ramakrishnan


Ayush Subedi


Kate Purpura


Syed Mohammad Ali


Tanmay Nag

Georgia Tech

Agenda

1 **Objective and problem statement**

3 **Methodology**

a) Data Collection
b)Research Design & Approach
c)Data Analysis
d)Variables and Measurements

4 **Dashboard**

5 **Results and discussion**

6 **Future Scope and Conclusion**

Georgia Tech

Objective and problem statement

Problem statement

No convenient means for authorities or parents to **evaluate the performance** of NYC schools apart from rankings.

Proposed solution

We aim to **investigate various factors** such as student satisfaction, opportunities for growth, community belongingness, feelings of safety etc. and measure their **impact on the rankings** of public schools in New York City

Key research points

- Identify factors having highest impact
- Measure factors effect on student experience
- Investigate correlation in metrics
- Determine if trend exists b/w factors & rankings

Business justification

Short term

- Promote and market schools on strengths
- Assist parents in informed decisions
- Highlight overlooked factors in rankings

Long Term

- Lead to sustainable improvement over time
- Promote positive competition across schools
- Inform policy and decision making

Our hypothesis is that students' positive experiences correlate positively with schools' ratings.

Methodology | Data Collection

Datasets used

NYC Public Schools Survey

Independent variable based on survey responses

Members of the school community participate in the survey (families, teachers and students

Zip Code data set

_id	zip	borough	post_office	neighborhood	population	density
1	10001	Manhattan	New York, NY	Chelsea and Clinton	21102	33959
2	10002	Manhattan	New York, NY	Lower East Side	81410	92573
3	10003	Manhattan	New York, NY	Lower East Side	56024	97188
4	10004	Manhattan	New York, NY	Lower Manhattan	3089	5519
5	10005	Manhattan	New York, NY	Lower Manhattan	7135	97048
6	10006	Manhattan	New York, NY	Lower Manhattan	3011	32796
7	10007	Manhattan	New York, NY	Lower Manhattan	6988	42751
8	10009	Manhattan	New York, NY	Lower East Side	61347	99492
9	10010	Manhattan	New York, NY	Gramercy Park and Murray Hill	31834	81487
10	10011	Manhattan	New York, NY	Chelsea and Clinton	50984	77436

Great Schools rating

Dependant variable showcasing NYC public schools ranks

Great Schools develops a ‘Summary Rating’ (1-10) in which schools are ranked

We’ll be agile based on our problem statement!

Methodology | Data Collection

Table 1. NYC 2022 School Survey

School Name	Total Student Response Rate	q1. Most students at this school treat each other with respect.				
		Strongly disagree	Disagree	Agree	Strongly agree	I don't know
P.S. 034 Franklin D. Roosevelt	69%	28	27	16	1	4
P.S. 140 Nathan Straus	6%	0	0	4	1	4
P.S. 184m Shuang Wen	98%	17	38	139	39	40
P.S. 188 The Island School	69%	0	0	3	100	1

Table 2. Greatschools.org School Rankings

School	Type	Grades	Total students enrolled	Students per teacher	Reviews	District	Test Scores Rating	Student Progress Rating	College Readiness Rating	Equity Overview Rating
10/10 Above average lennium Brooklyn High Sch 7TH AVE, BROOKLYN, NY, 1 Homes for sale	Public district	9-12	671	10:1	7 Reviews	Nyc Geog District #15 - Brooklyn	10/10	N/A	10/10	10/10
10/10 Above average Peck Slip School (The) Peck Slip, New York, NY, 100 Homes for sale	Public district	PK-5	470	12:1	14 Reviews	New York City Geographic District # 2	10/10	10/10	N/A	9/10
10/10 Above average Manhattan Village Academy t 22nd Street, New York, NY Homes for sale	Public district	9-12	452	15:1	6 Reviews	New York City Geographic District # 2	N/A	N/A	10/10	9/10
10/10 Above average River School (The) st 35th Street, New York, NY Homes for sale	Public district	PK-5	340	12:1	2 Reviews	New York City Geographic District # 2	10/10	10/10	N/A	10/10

Georgia Tech

Methodology | Research Design/Approach

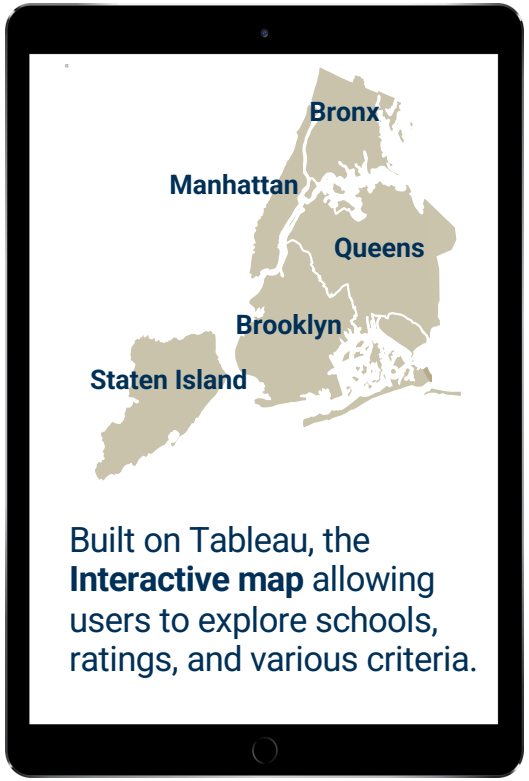
The research followed a two-phase approach: I) Exploratory data analysis and reporting and II) Statistical modeling.

Phase I: Exploratory data analysis and reporting → Phase II: Statistical modeling

In the first phase we developed tableau visualizations on the datasets to understand the responses to different questions in the survey and study the commonality in patterns.

Additionally, for schools that were common in the two datasets, we created a map-based visualization of New York schools and showcased the rankings alongside the responses to different survey questions.

Part 1 - Dashboarding



Findings on EDA, and reporting of insights from statistical modelling.

Georgia Tech

Methodology | Research Design/Approach

In this phase, we investigated analysis relevant to survey data (ANOVA, correlation analysis, etc.), as well as used logistic regression to get a ranking of student satisfaction and experiences from dataset A.

Phase I: Exploratory data analysis and reporting

Phase II: Statistical modeling

Split data

Into training and test data.

Plot data points

To understand existing relationships between independent and dependent variables. & check for multicollinearity

Run initial regression.

Check linearity, constant variance, and normality assumptions

Perform transformation

Validate various models on validation data

Select the most suitable model to test on test data

Positive answers increase student happiness probability

Negative answers decrease student happiness probability

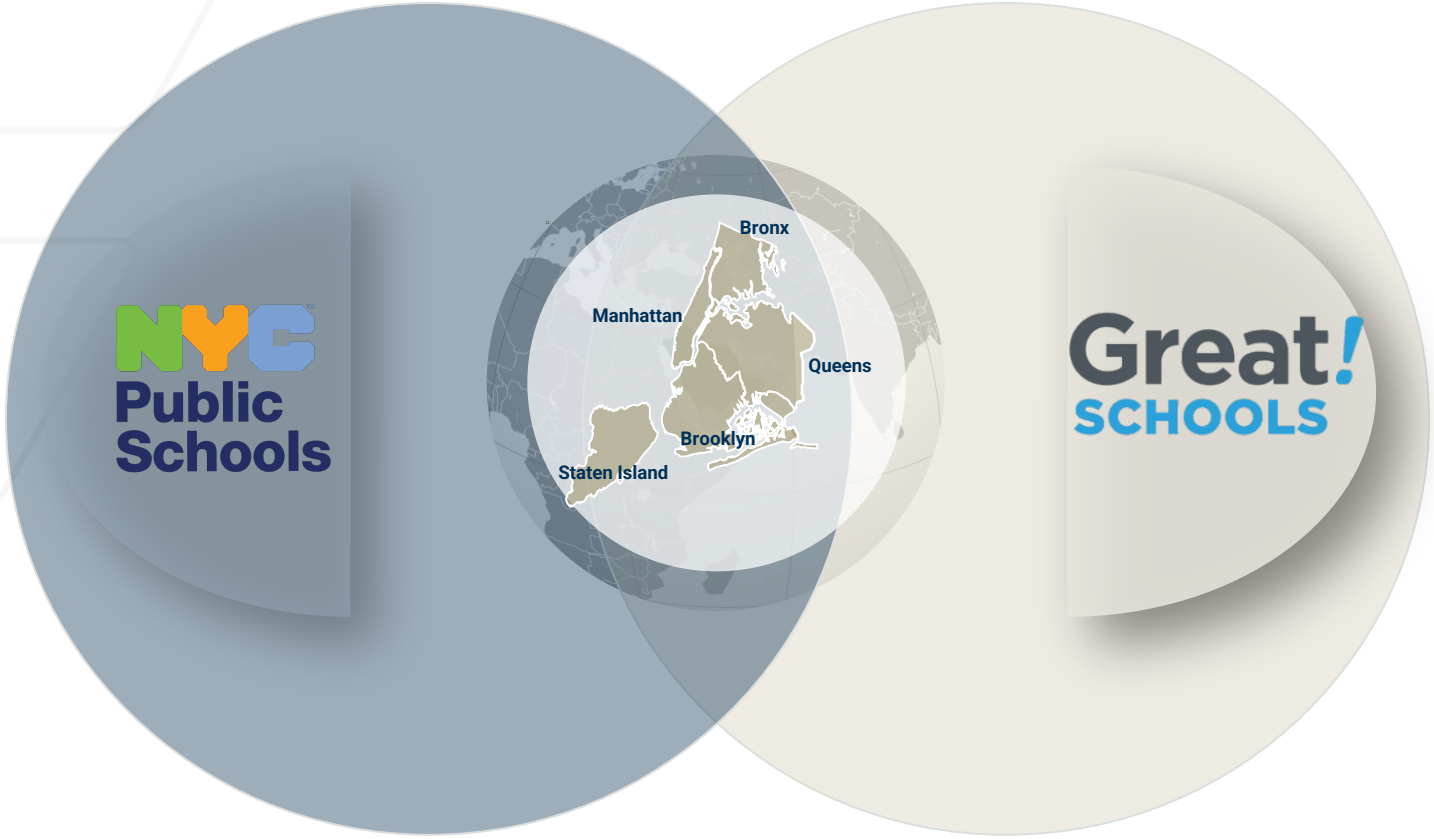
Analyze

Identify schools with great rankings and student happiness ratings as the top schools to promote to parents

Georgia Tech

Methodology | Data Analysis

The two datasets were combined based on common schools. 296 schools matched exactly and misspellings were corrected. 296 schools provided sufficient data to analyze



We further used a 3rd supporting data set which included the zipcodes…

_id	zip	borough	post_office	neighborhood	population	density
1	10001	Manhattan	New York, NY	Chelsea and Clinton	21102	33959
2	10002	Manhattan	New York, NY	Lower East Side	81410	92573
3	10003	Manhattan	New York, NY	Lower East Side	56024	97188
4	10004	Manhattan	New York, NY	Lower Manhattan	3089	5519
5	10005	Manhattan	New York, NY	Lower Manhattan	7135	97048
6	10006	Manhattan	New York, NY	Lower Manhattan	3011	32796
7	10007	Manhattan	New York, NY	Lower Manhattan	6988	42751
8	10009	Manhattan	New York, NY	Lower East Side	61347	99492
9	10010	Manhattan	New York, NY	Gramercy Park and Murray Hill	31834	81487
10	10011	Manhattan	New York, NY	Chelsea and Clinton	50984	77436

0	NaN	NaN	NaN	Strongly disagree	Disagree	Agree	Strongly agree	I don't know	Strongly disagree	Disagree	...	Strongly disagree	Disagree	Agree	Strongly agree
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
2	01M034	P.S. 034 Franklin D. Roosevelt	69%	28	27	16	1	4	0	2	...	NaN	NaN	NaN	NaN
3	01M140	P.S. 140 Nathan Straus	6%	0	0	4	1	4	0	2	...	NaN	NaN	NaN	NaN
		P.S. 184m													

Additionally, we converted the joined dataset to long format for Tableau dashboards and statistical modeling.

Georgia Tech

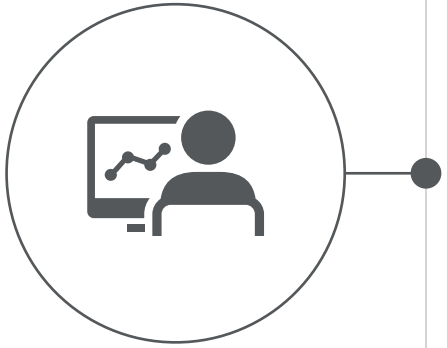
Methodology | Variables and Measurements

Our research is based on the four key points:

Identify
factors having the highest impact on the schools' ranking.

Measure
factors that influence the student satisfaction and experiences.

Investigate
correlation in metrics.



Determine
if trend exists between survey factors & rankings.

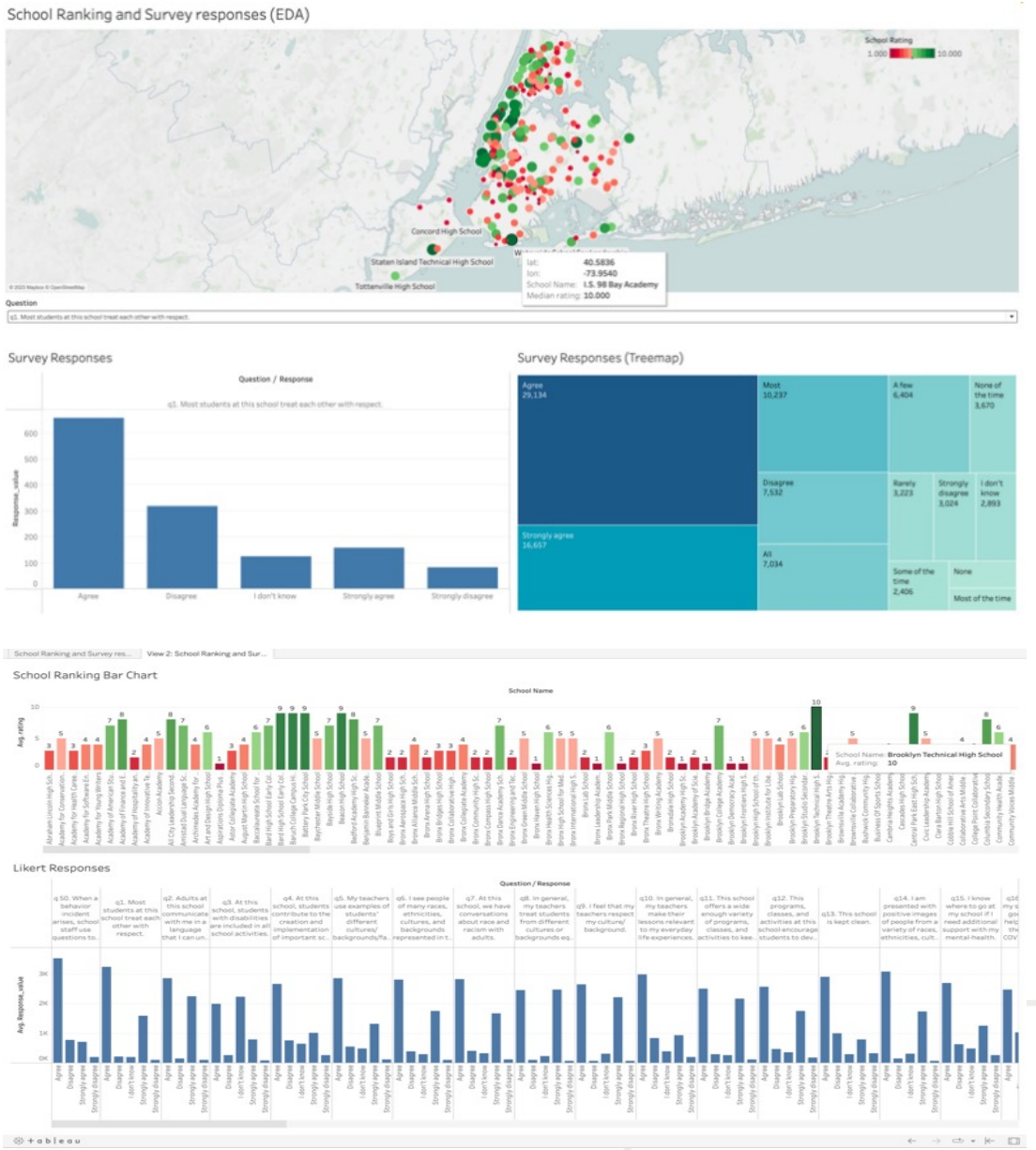
We used the school names to join the two datasets. We also experimented with feature engineering Likert scale to binary responses. The independent variables are survey responses, whereas the dependent variable is the Great School ranking.

Dashboard

We developed a Tableau dashboard to compare lower-rated and higher-rated schools in NYC. The dashboard allows users to visualize and explore the data interactively, promoting a better understanding of the observed trends.

The dashboard showed no clear difference between how lower-rated and higher-rated schools responded to the survey questions. This challenges our initial hypothesis and shows that school ratings are influenced by a complex set of factors. The dashboard is available on Tableau Public. [here]

The Tableau dashboard is a valuable tool for presenting and interpreting our analysis. The absence of distinct trends between lower-rated and higher-rated schools encourages further exploration into the determinants of school ratings. We are optimistic that this research will contribute to improving educational equity.



Results and Discussion

The survey questions q21 & q39. had the highest positive and negative correlation with school ranking, respectively

Response	Question	rating
26	q21. During this school year, I have felt stressed out when learning.	0.337309
28	q22. During this school year, I have felt worried when learning.	0.319305

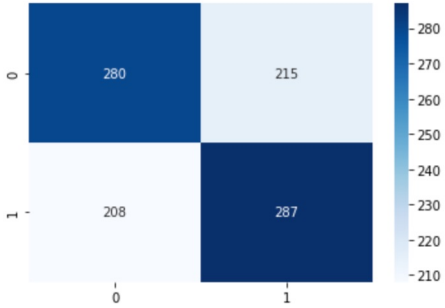
None of the independent variables demonstrated a correlation close to 0.60, which means that the initial hypothesis was not supported.

The random forest algorithm and logistic regression model had low accuracy scores, indicating that student responses were not able to predict school rankings.

The ordinality of the data was not as big of an issue as the data mostly being noise and not a signal to represent school ranking.

A binary classification model with a threshold of 7 was slightly better than a 10-point scale model, but the accuracy scores were still poor.

The XGBoost model had a slightly better accuracy score than the random forest algorithm, but it was still poor.



Confusion Matrix for XGBoost model.

	Predicted									
	0	1	2	3	4	5	6	7	8	9
Actual 0	128	0	203	74	10	9	0	0	0	0
Actual 1	21	0	76	21	1	2	0	0	0	0
Actual 2	70	0	438	250	3	19	0	0	0	0
Actual 3	29	0	444	234	10	23	0	0	0	0
Actual 4	10	0	289	217	11	13	0	0	0	0
Actual 5	15	0	341	228	12	25	0	0	0	0
Actual 6	15	0	286	134	7	20	0	0	0	0
Actual 7	10	0	141	146	8	14	0	0	0	0
Actual 8	14	0	98	41	0	7	0	0	0	0
Actual 9	5	0	139	55	3	11	0	0	0	0

Confusion matrix output from logistic regression

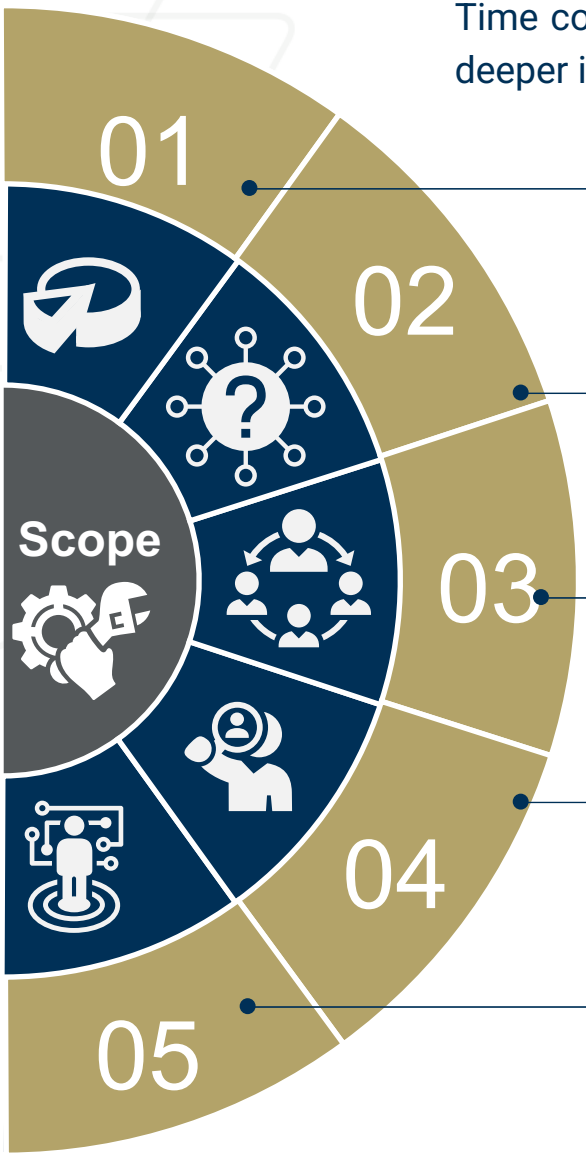
All of the models were trained on balanced data and there was no data leakage.

ANOVA is not ideal for Likert scales and an ordinal logistic regression would be more appropriate.

Overall, the study found that there is no clear relationship between student responses and school rankings. The results suggest that other factors, such as school resources, teacher quality, and student demographics, may be more important in determining school rankings.

Future Scope

Time constraints limited our exploration, but with more time, we would have pursued alternative approaches to delve deeper into the data and uncover further insights.



Study size
The study was limited by the size of the datasets.

Hypothesis
The initial hypothesis was rejected, but further research could uncover meaningful correlations between school ratings and other relevant factors.

Correlation
For example, examining the connection between school ratings and average income, average class size, or the size of the school district could offer valuable insights.

Findings
The study presents valuable findings, but the authors recognize the potential for further research to expand upon our existing knowledge.

Evidence
The pursuit of new datasets and exploration of additional correlations will undoubtedly strengthen the foundation for evidence-based decision-making in the realm of education

Georgia Tech

Conclusion

Our analysis found that student satisfaction and experiences are not strongly correlated with school rankings in NYC

The survey questions "During this school year, I have felt stressed" and "My teachers check-in with me frequently" had positive and negative correlations with rankings, respectively.

The initial hypothesis that student satisfaction and experiences are correlated with school rankings was not supported by our analysis.

We also identified several limitations, such as challenges with handling ordinal data and the relatively small cardinality in the intersection set.

Data analytics may not always align with our intuition or common sense.

Our analysis suggests that other factors, such as the school's location or students' income levels, may be more strongly correlated with school rankings.

Our project has provided valuable insights into the factors that influence school rankings. It also highlights the need for continuous improvement in research methodologies to better understand educational systems.

This analysis aims to assist parents, authorities, and students in making more informed decisions.

Georgia Tech

Georgia Tech