

# Analysis Report

Rabeya Siddika Sajia - 23084716

## Math Matters: Examining the New York State Students Performance Over the Time (2006-2023)

Mathematics is a fundamental part of life, influencing every aspect of daily activities and decision-making. From early childhood, math skills lay the foundation for critical thinking and problem-solving. This study analyzes 18 years of student performance data from New York State Math exams to uncover long-term trends and insights. By combining two datasets, 2006 to 2012 and 2013 to 2023, we are investigating similar and dissimilar patterns in test outcomes between grade levels, demographic groups, and geographic areas in New York. Hence, this analysis will provide precise and practical understanding of the question, **“How the students’ performance in New York City Math exams expanded in the past 18 years, and what factors are involved in the changes in different demographics and regions?”**

### I. Data Sources

Data source 1: Math Test Results 2006-2012 [1]

- Data Type: CSV File

Data source 2: Math Test Results 2013-2023 [2]

- Data Type: CSV File

These datasets provide a complete look at New York State Math exam results over a long period, providing a strong foundation for examining trends in student performance. It includes information about different geographic regions, grades, student type, enabling a detailed analysis for answering our main question.

### A. Data Structure

The datasets which we got after cleaning the data, by removing some columns and changing some values, are highly structured and organized. Each column corresponds to specific attributes such as year, grade level (3-8), different regional locations, total number of students tested in the schools, percentage of grading levels, and demographic details.

Report Category	Geographic Subdivision		Grade	Year	Student Category	Number Tested	% of Students Level 1		% of Students Level 2		% of Students Level 3		% of Students Level 4		% of Students Level 3 & 4	
	State	Site					State	Site	State	Site	State	Site	State	Site	State	Site
1	Capekide	Capekide	3	2006	Female	99459	9.2	13.7	51.1	28.5	76.1					
2	Capekide	Capekide	3	2007	Female	10240	5.8	12.3	52.4	30.5	83.1					
3	Capekide	Capekide	3	2008	Female	10850	3.4	9.1	64.1	24.4	88.5					
4	Capekide	Capekide	3	2009	Female	99812	3.3	6.4	66.2	24.3	92.5					
5	Capekide	Capekide	3	2010	Female	94588	11.0	14.4	52.3	22.2	84.6					

Fig. 1. Table of Math Results (2006-2012)

Report Category	Geographic Subdivision		Grade	Year	Student Category	Number Tested	% of Students Level 1		% of Students Level 2		% of Students Level 3		% of Students Level 4		% of Students Level 3 & 4	
	State	Site					State	Site	State	Site	State	Site	State	Site	State	Site
1	Capekide	Capekide	3	2023	Boys (BMD)	41394	11.6	27.3	37.7	23.4	61.1					
2	Capekide	Capekide	3	2023	BMD	10562	32.2	36.7	21.8	7.7	31.1					
3	Capekide	Capekide	4	2023	Boys (BMD)	45756	19.5	21.4	37.4	21.5	58.9					
4	Capekide	Capekide	4	2023	BMD	11435	46.6	25.3	22.4	5.9	28.3					
5	Capekide	Capekide	5	2023	Boys (BMD)	45270	20.9	21.8	34.0	23.3	57.3					

Fig. 2. Table of Math Results (2013-2023)

### B. License

Both datasets are sourced from the United States Government's open data platform about which is also mentioned in the data report. There it is explicitly stated that the datasets are intended for public use. According to their terms and conditions, we are allowed to access, use, and analyze the datasets, but the providers are not responsible for the accuracy or completeness of the information. In adherence to these terms, I will ensure that the data is used responsibly, without engaging in any criminal activity, misuse, or actions that violate the stated guidelines.[3]

## II. Analysis

### A. Methodology

The data integration process combined the two datasets where each contains performance data of students from 2006 to 2023. Both have consistent column names since they both represent similar data for different years. Due to COVID-19, no math tests were conducted in 2020 and 2021, resulting in missing data for these years.

### 1.1 Performance Trends by Year

A stacked bar chart is used to highlight the comparative contributions of each performance level over the years. It shows that Level 1 (Low Performing) students were fewer until 2012 after which their numbers increased, accompanied by a decline in Level 3 & 4 (Proficient and Excelling) students. This indicates an inverse relationship between them. In this analysis we focused on Level 3 & 4 students together because they represent the high performing group, displaying proficiency and excellence compared to the other two levels. After COVID-19 students' performance started to improve significantly with more achieving higher levels.

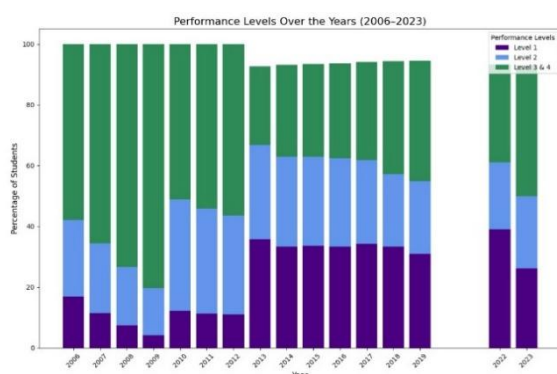


Fig. 3. Performance Trends Over the Years

### 1.2 Performance Trends by Demographic Group

The three grouped bar charts highlight the performance trends across demographic groups (Asian, Black, Hispanic, and White). We can observe that Hispanic and Black students are higher in percentage among Level 1 (Low-Performing) and

Level 2 (Below Proficient) students, while Asian and White students dominate Levels 3 & 4 (Proficient & Excelling). There is a decline in performance for all groups after 2012, followed by a little recovery in recent years.

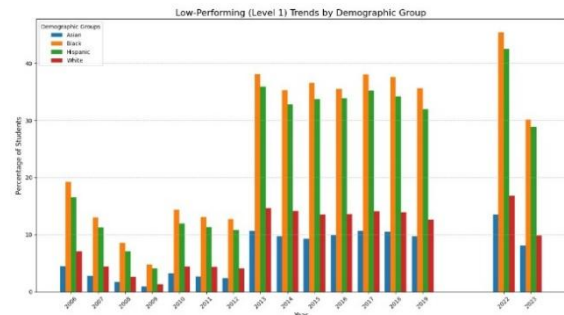


Fig. 4.1. Level 1 Students

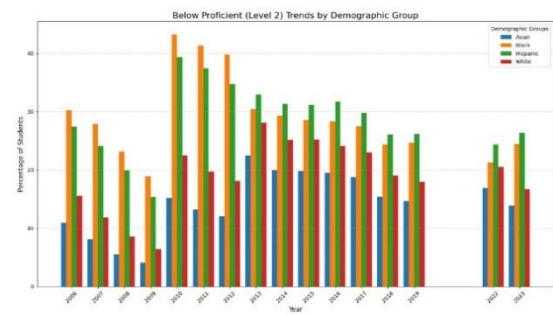


Fig. 4.2. Level 2 Students

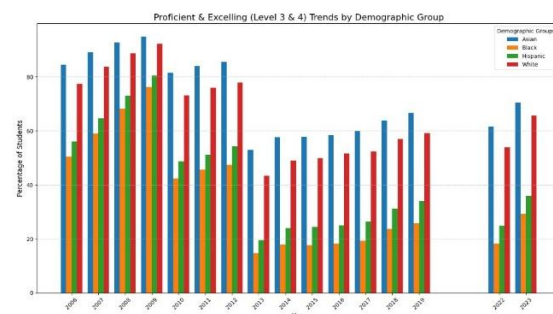


Fig. 4.3. Level 3 & 4 Students

### 1.3 Performance Trends by Geographic Region

Due to our huge dataset, we have selected some data for this analysis. The three scatter plots highlight the performance trends across the selected subdivisions (Bronx, Brooklyn, Citywide, Manhattan, Queens, Staten Island) from 2006 to 2023. Subdivisions like the Bronx consistently show higher percentages of Level 1 and Level 2 students, indicating areas where

students' performance is not promising. On the other hand, Manhattan and Queens consistently perform better, with a high percentage in Levels 3 & 4 students.

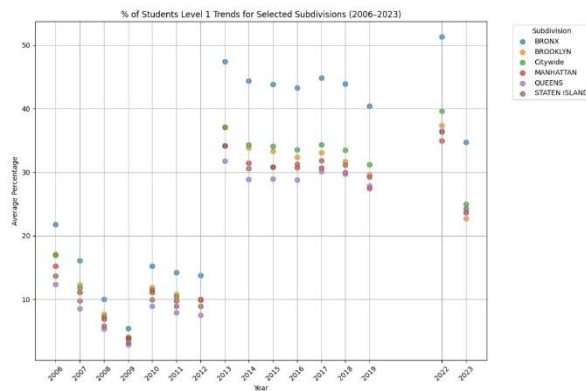


Fig. 5.1 Level 1 Students

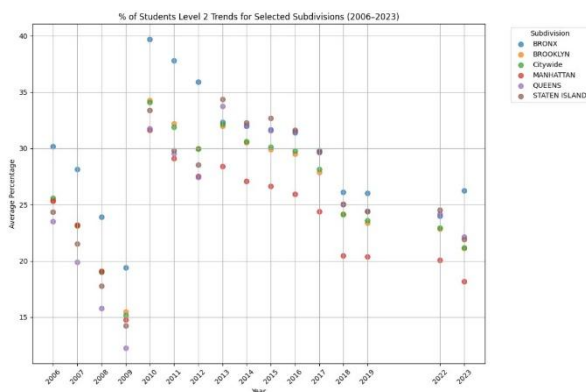


Fig. 5.2. Level 2 Students

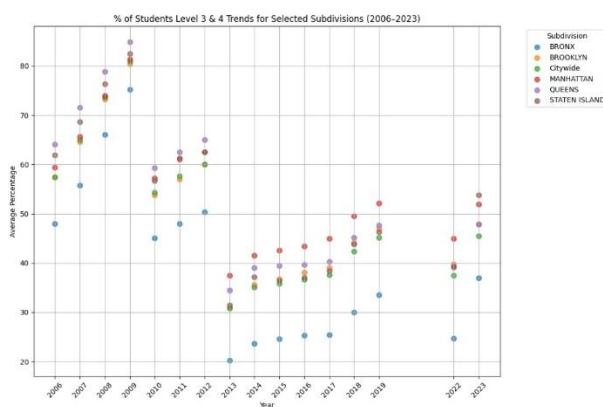


Fig. 5.3. Level 3 & 4 Students

## B. Correlation Analysis

The heatmap shows the correlation between performance metrics (percentage of students in Levels 1, 2, and 3 & 4) and the year. Level 1 students have a positive

correlation with year (0.37), indicating an increase in Level 1 students over the years. In contrast, Level 3 & 4 students have a strong negative correlation with Level 1 (-0.74) and a slight negative correlation with year (-0.37), indicating there exists an inverse relationship between them. Level 2 shows very weak correlations with year (-0.02) and other performance levels, indicating little variability.

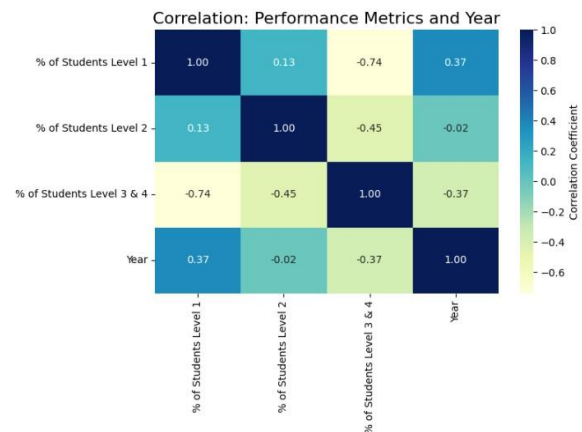


Fig. 6. Correlation Analysis between Year and Percentage of Levels of Students

## C. Results

- The level of high-performing students seen before 2012 has not yet been achieved in the post-pandemic years. It highlights that the students need greater focus and attention on improving their education.
- All groups of students show some variations, especially after post-pandemic, which could be linked to external factors like education policies, pandemic related disruptions or mental situations. The Black and Hispanic students require greater support and attention to help them reach proficient and excelling performance levels.
- The trends over time show declining high performance and increasing low performance, indicating the need for focused involvement to

reverse the situation. Besides, Level 2 students have remained relatively stable, acting as a middle category between low and high performance.

### III. Limitations

- Due to the dataset's large size, we could not analyze all geographic subdivisions, which limits the ability to fully represent trends across all areas.
- Due to some missing values and also missing data of 2020 and 2021 (because of COVID-19) creates a gap, by making it harder to observe the continuous trends.
- The correlation analysis reveals only relationships but doesn't explain why certain trends, like declining high performance are happening.
- Influences like socio-economic conditions, policy changes, or differences in school resources which aren't included, could affect performance trends.

### IV. Conclusion

This report studies how students' performance in New York City math exams has evolved over the past 18 years and tries to explore the points contributing to changes across different demographics and regions. The analysis recognizes key trends, such as an inverse relationship between the high-performing (Levels 3 & 4) and low-performing students (Level 1) after 2012, which is still not yet recovered. Geographic and demographic imbalance are clear, with Manhattan and Queens consistently outperforming areas like the Bronx, and Asian students achieving better results compared to other students. To address our question, we analyzed performance trends over the years. However, to understand the elements driving changes across different demographics and regions, more information is needed such as, socio-

economic data, school policies, home environments, and other contributing factors that have direct influence on students' performance.

### References

- [1] Math Test Results 2006-2012. <https://catalog.data.gov/dataset/math-test-results-2006-2012>
- [2] Math Test Results 2013-2023. <https://catalog.data.gov/dataset/math-test-results-2013-2023>
- [3] NYC.gov, Terms of Use. <https://www.nyc.gov/home/terms-of-use.page>