Camille Settles   
Math 014- Intro to Data Science   
Dr. Meenakshi Nerolu   
Project 2-Part 2   
April 18, 2025

Title of the Project:   
Understanding School-Level Outcomes in Chicago Public Schools: A Data-Driven Exploration

Student Name and ID:   
Camille Settles, @02986989

Objective:   
This project investigates trends in attendance, academic growth, family involvement, and behavioral data across school types in the Chicago Public Schools (CPS) system. The goal is to uncover key patterns using cleaned public data and visual analytics, identifying differences in educational outcomes across elementary, middle, and high schools.

Introduction:   
This analysis is based on the Chicago Public Schools Progress Report Card dataset from 2011–2012. The dataset includes over 500 schools and 79 variables, offering a comprehensive view of student performance, engagement, and school climate. The objective is to explore how factors such as school type, attendance, academic growth, family involvement, and performance policy levels interact to shape educational experiences and outcomes.

The dataset, publicly available via the City of Chicago Data Portal, provides both quantitative and categorical information. These data points span test performance, disciplinary incidents, attendance rates, and engagement scores across a wide range of school settings.

Dataset Reference:   
City of Chicago Data Portal. (2011). Chicago Public Schools - Progress Report Cards (2011-2012). Retrieved from: https://data.cityofchicago.org

Method:   
The original dataset underwent several key preprocessing steps to ensure the visualizations would be both accurate and meaningful. First, abbreviated school type codes "ES" (elementary), "MS" (middle), and "HS" (high) were mapped to their full descriptive names using the .map() function to improve clarity in plots.

To ensure valid comparisons, rows containing missing values (NaNs) in critical variables such as 'Average Student Attendance', 'ISAT Value Add Read', 'ISAT Value Add Math', 'Family Involvement Score', 'CPS Performance Policy Level', and 'Rate of Misconducts (per 100 students)' were removed using .dropna(). These columns were selected because they directly relate to the study's focus on academic outcomes and engagement. Removing these rows prevented inaccurate averages and misrepresentative visualizations.

Categorical string fields, such as the CPS Performance Policy Level, were cleaned using .str.strip() to eliminate extra spaces that could interfere with grouping and aggregation. Several columns, including the family involvement and ISAT growth scores, were verified or explicitly converted to numeric format using pd.to\_numeric() to prevent issues with plotting functions and to ensure consistent mathematical operations. While columns such as school phone numbers, links, and addresses remained in the dataset, they were excluded from all visual analysis by focusing only on the columns relevant to school performance and student outcomes.

This allowed for meaningful visualizations that accurately illustrated patterns in attendance, academic growth, misconduct, and engagement across different types of schools and performance levels.

Storytelling (Data Visualization & Interpretation):   
The boxplot of student attendance by school type revealed that elementary and middle schools exhibit high and consistent attendance rates, while high schools demonstrate both a lower median attendance and greater variability, including several outliers. This suggests that student presence becomes less reliable at the high school level. High absenteeism may correlate with a range of other academic or social challenges.

A bar chart showing average family involvement scores across showed a result that high schools had the highest average involvement score in the dataset. This challenges the expectation that parental involvement is greatest in early education and may reflect concentrated efforts at the high school level to engage families around graduation readiness, academic planning, or intervention strategies.

A count plot demonstrated that elementary schools account for the vast majority of entries in the dataset, while middle and high schools are sparsely represented. This imbalance is important because it limits the generalizability of findings across all school levels. The overrepresentation of elementary schools must be kept in mind when interpreting trends and drawing conclusions.

In a grouped bar chart comparing average student misconduct across CPS performance policy levels, Level 3 schools had the highest reported rates of misconduct, while Level 1 schools showed the lowest. This inverse relationship highlights how school culture and discipline challenges often correlate with academic underperformance. It reinforces the importance of supportive learning environments and interventions in improving school performance metrics.

A scatter plot comparing ISAT reading and math growth values showed a strong positive correlation between the two metrics, particularly among elementary schools, which were tightly clustered. High schools, by contrast, exhibited more scatter, reflecting inconsistent academic growth across subjects. The strength of the correlation supports the idea that literacy and numeracy development are linked and may benefit from integrated instructional strategies.

Finally, a correlation heatmap summarized the relationships between average student attendance and ISAT value-added metrics. Attendance showed a moderate positive correlation with both reading and math growth, reinforcing the widely accepted idea that students who attend school regularly are more likely to show academic progress. This insight supports policies and programs that prioritize attendance monitoring and engagement efforts.

Conclusion / Summary of Insights:   
The analysis revealed several key insights: schools with higher performance levels report fewer behavioral issues; consistent attendance is more common in elementary and middle schools; and reading and math growth are positively correlated. Surprisingly, family involvement appears highest in high schools in this dataset, possibly due to specific engagement programs or targeted parent outreach. These findings suggest that behavioral and academic outcomes are tightly interwoven, and that early interventions and consistent engagement can improve long-term results. However, it's important to note that the heavy representation of elementary schools introduces bias that must be acknowledged when drawing broader conclusions. Future analyses should consider balancing school type representation and investigating year-over-year data to track progress over time.

References:   
Chicago Public Schools Progress Report Card (2011–2012), City of Chicago Data Portal.   
https://catalog.data.gov/dataset/chicago-public-schools-progress-report-cards-2011-2012

Acknowledgements:   
Special thanks to Dr. Nerolu for providing the project structure and continued guidance, as well as to the City of Chicago for maintaining accessible, high-quality public education data.