Urban Noise and Educational Outcomes: Analyzing Environmental Sound Impacts in NYC Schools

Emily Aiken

Springboard Capstone 3

January 2025

1 INTRODUCTION

1.1 BACKGROUND

In educational settings, the challenge of maintaining an effective learning environment amid urban noise has become increasingly complex. Students in urban schools face a diverse array of noise sources, from transportation and construction to community activities and emergency services. These varying noise types may affect cognitive processes differently, potentially impacting academic performance in ways that previous research has not fully explored.

New York City's educational landscape presents a unique opportunity to study these relationships. With over 1,800 schools serving approximately 1.1 million students, NYC schools operate in an environment characterized by intense urban activity and diverse noise sources. The city's dense infrastructure, multiple transportation systems, and vibrant community life create a complex acoustic environment that can significantly impact the learning process. This project will explore the potential affect of these sounds on New York City's academic outcomes.

1.2 RESEARCH QUESTIONS

This study examines the relationship between environmental noise and academic performance through three primary research questions:

1. How do different types of environmental noise correlate with academic achievement in NYC schools?
2. Does the relationship between noise and academic performance differ between lower grades and high schools?
3. Can machine learning models effectively classify schools based on their noise exposure patterns and academic performance?

These questions aim to fill critical gaps in our understanding of how urban noise affects educational outcomes, potentially informing both educational policy and urban planning decisions.

2 DATA AND METHODOLOGY

2.1 DATA SOURCES

This study integrates multiple datasets to analyze the relationship between urban noise and academic achievement in New York City Public Schools. Our analysis draws from three primary data sources:

1. SONYC-UST v2.3 Dataset (Sounds of New York City) - Comprehensive acoustic data from the SONYC sensor network with over 18,000 annotated sound recordings taken from 54 unique sensors. Contains detailed spatiotemporal metadata including sensor coordinates and timestamps for each recording.
2. New York City Public Schools Performance Data - School quality reports from elementary, middle, and high schools in New York City spanning three academic years (2016-17 through 2018-19).
3. NYC Open Data School Point Locations - ESRI shapefile containing geospatial data with precise school locations across all boroughs.

2.2 DATA PROCESSING METHODOLOGY

The data processing pipeline followed several key steps to ensure data quality and compatibility. First, using GeoPandas, we identified schools within a 2km radius of each SONYC sensor, establishing our study's spatial framework. See Figure 1 for a map of these schools. We then synchronized academic years with noise measurements, focusing on the period between August 1, 2016, and June 30, 2019, to align with complete academic years.

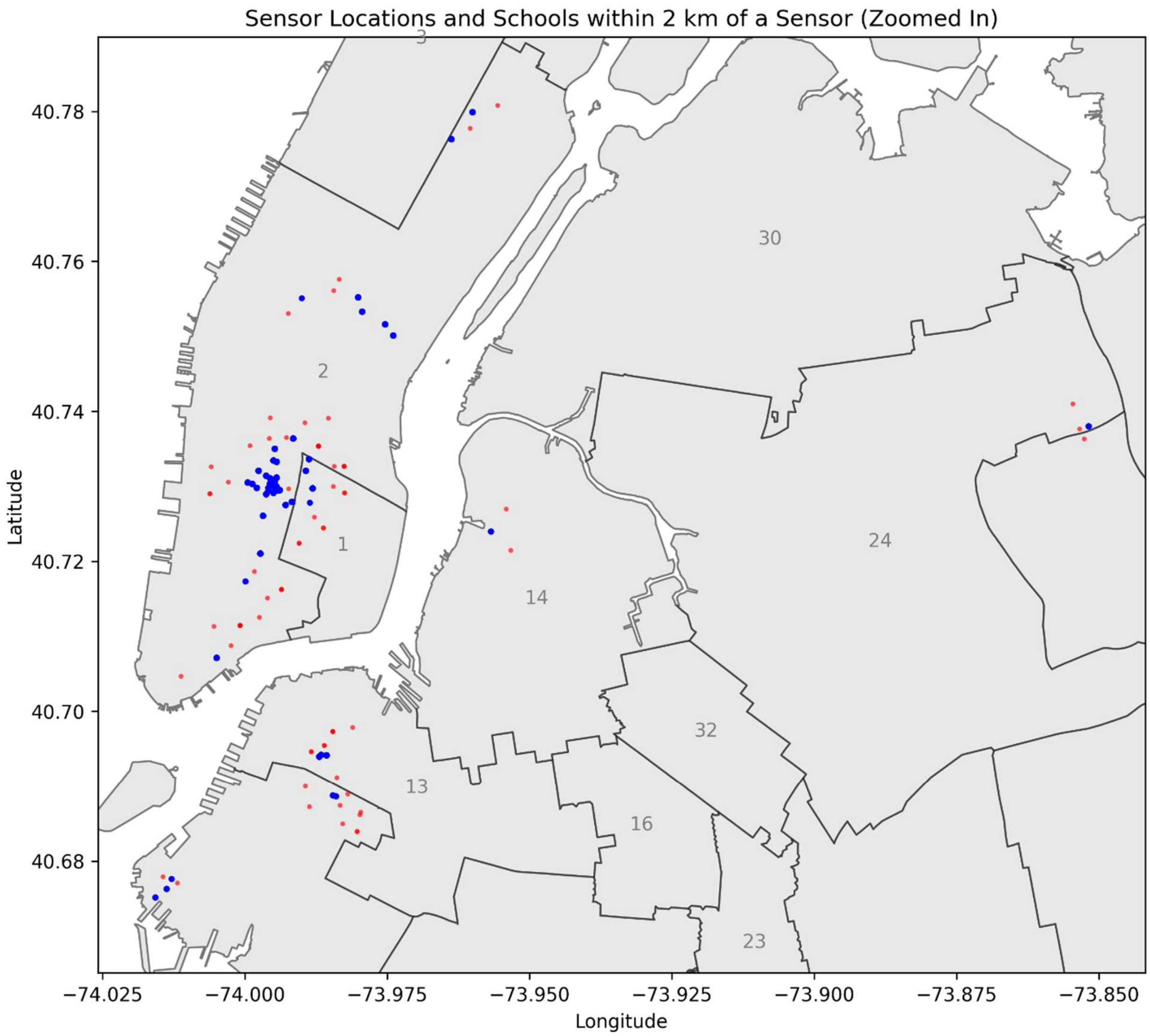

Figure 1. Final map of New York City after removing school locations that were outside the 2km range of the SONYC sensors. Sensor locations are plotted in blue, while school locations are plotted in red.

The SONYC dataset's 23 distinct sound classes were aggregated into broader categories including alert signals (emergency vehicles, alarms), engine noise (cars, trucks, aircraft), music, and ambient urban sounds. From the School Quality Reports, we extracted and normalized performance metrics, particularly the lower grades' ELA proficiency (mean: 0.433, std: 0.205) and Math proficiency (mean: 0.414, std: 0.237) scores. These datasets were merged into a unified analysis framework, with careful attention to temporal alignment and data quality validation.

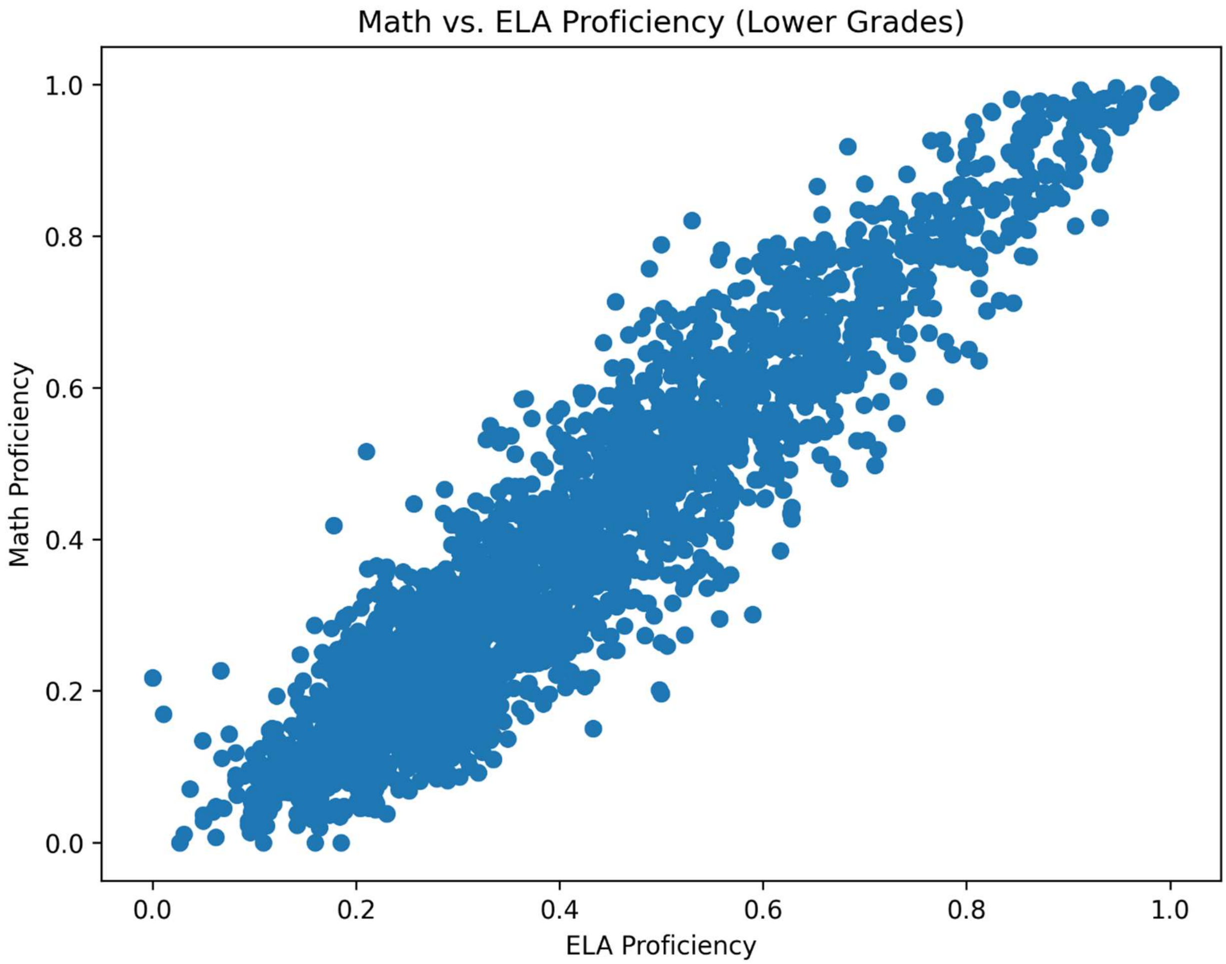

Figure 2. The correlation between Math and ELA proficiency for lower grades was 1.00:0.94. This nearly 1:1 correlation enabled us to use a combined metric for modeling purposes without losing much information.

2.3 Analysis Framework

After standardizing numerical features and handling missing values, we employed K-means clustering to identify patterns in noise exposure across schools. The classification analysis, focused on predicting academic performance from noise patterns, utilized several models, with Logistic Regression proving most effective (CV Score: 0.892 ±0.063, F1 Score: 0.916, Accuracy: 0.91).

To ensure robust results, we addressed class imbalance using SMOTE technique, achieving a balanced training set. Feature importance analysis through the Logistic Regression model revealed varying impacts of different noise types, with alert signals (1.530), engine noise (1.067), music (-2.080), and dog presence (-1.581) emerging as key factors. We validated our findings through 5-fold cross-validation and sensitivity analyses across different school subsets, complemented by geographic analysis of noise impact patterns.

3 Results

3.1 Model Performance

We trained and evaluated three different classification models – Logistic Regression, Random Forest Classifier, and Support Vector Machine – to understand how different types of environmental noise relate to student proficiency scores. For the sake of simplicity, and since prior analysis revealed that stronger correlations between noise levels and academic performance in lower grades than in higher grades, we used the combined ELA and Math proficiency metric for lower grades discussed in section 2.2.

Our Logistic Regression model demonstrated strong predictive capability in analyzing the relationship between urban noise and academic achievement. While Random Forest achieved a slightly higher cross-validation score of 0.919 (±0.053), Logistic Regression provided superior F1 Score (0.916) and accuracy (0.91) compared to both Random Forest (F1: 0.818, accuracy: 0.82) and SVM (F1: 0.758, accuracy: 0.73). The Logistic Regression model's balanced performance, with a cross-validation score of 0.892 (±0.063), indicates robust and consistent performance across different subsets of the data, making it the most reliable choice in capturing the noise-achievement relationship. See Figure 3 for a bar chart offering a visual comparison of each model's metrics.

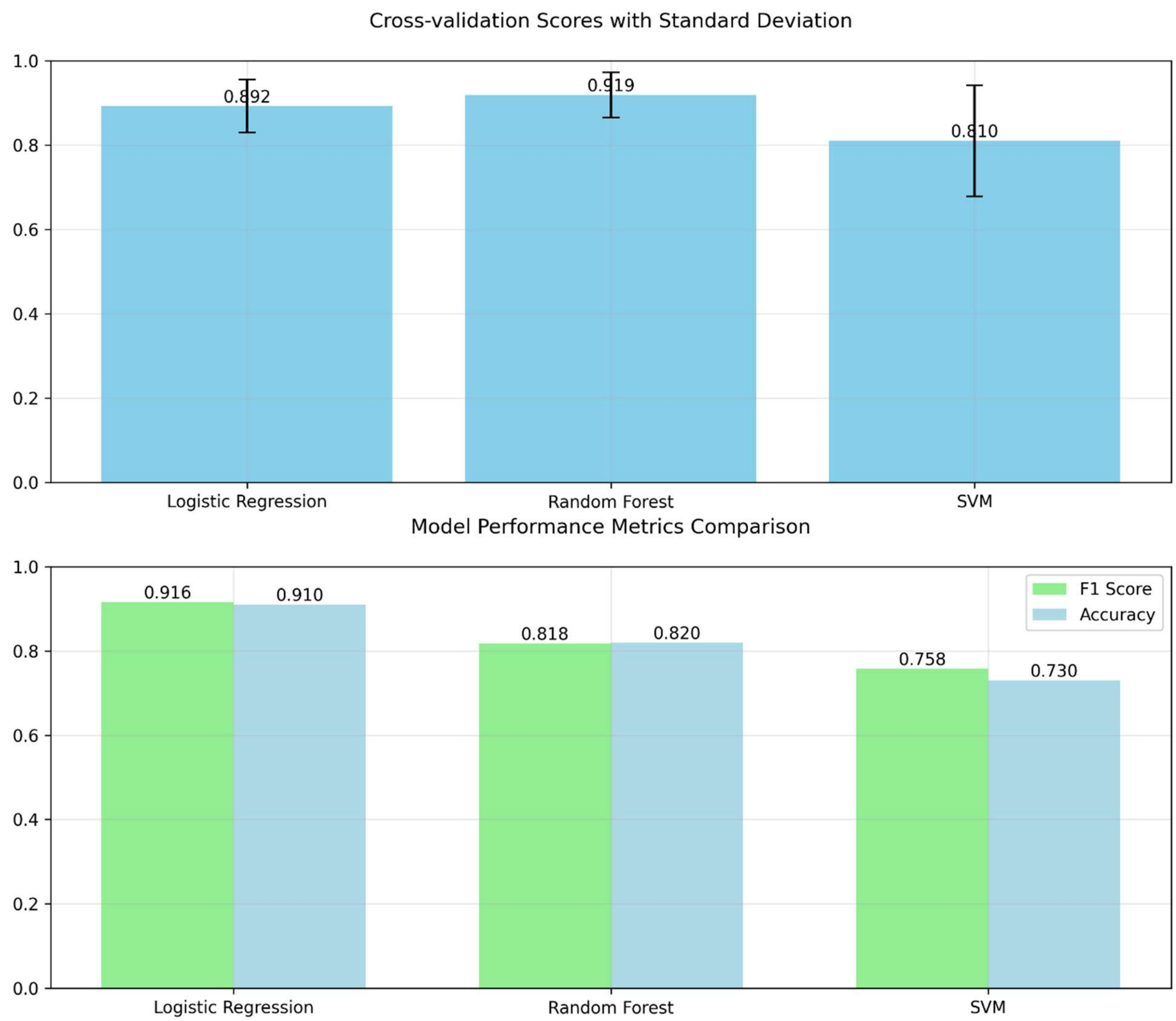

Figure 3. A comparison of cross validation scores (top) and model performance metrics, including F1 and Accuracy Scores (bottom).

3.2 IMPACT OF NOISE TYPES

The analysis revealed varying impacts of different noise types on academic achievement. Alert signals showed the strongest positive correlation with achievement (coefficient: 1.530), followed by engine noise (coefficient: 1.067). This unexpected finding suggests potential underlying factors, such as proximity to emergency services or main thoroughfares, which may correlate with other beneficial neighborhood characteristics.

Conversely, certain noise types demonstrated negative associations with academic performance. Music emerged as the strongest negative predictor (coefficient: -2.080), followed by the presence of dog-related noise (coefficient: -1.581). These findings suggest that certain types of ambient noise may be more disruptive to the learning environment than others.

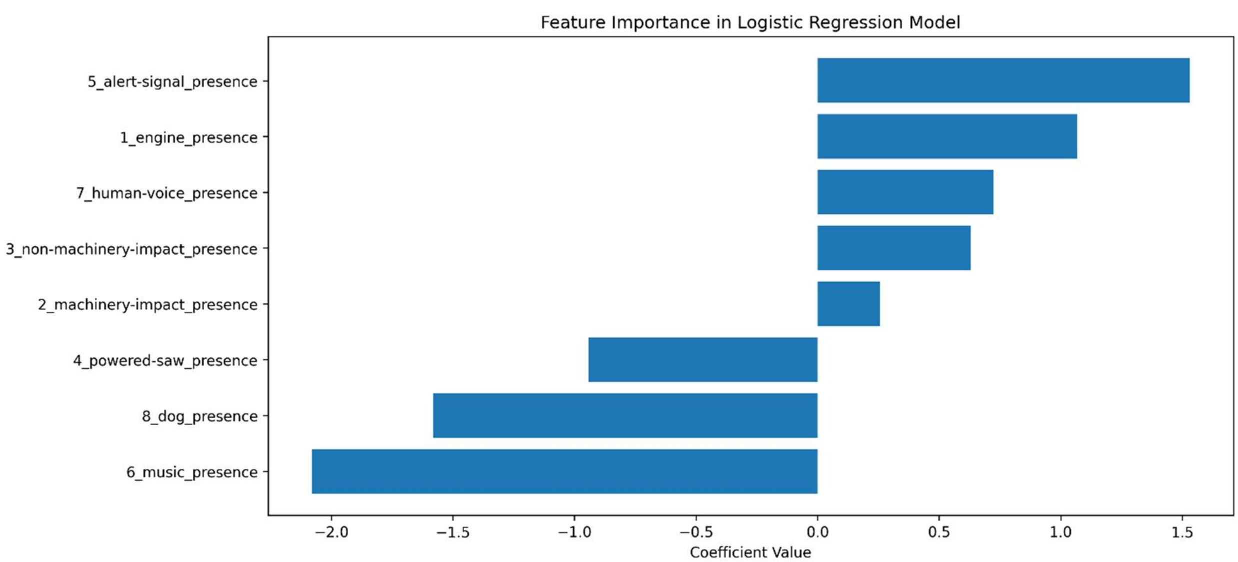

Figure 4. Coefficient values of feature importance in our logistic regression model.

3.3 GRADE-LEVEL ANALYSIS

Our results indicate stronger correlations between noise exposure and academic performance in lower grades compared to higher grades. Elementary school students showed particular sensitivity to noise impacts, with variations in both ELA (mean: 0.433, std: 0.205) and Math proficiency (mean: 0.414, std: 0.237) correlating more strongly with noise levels than in higher grades.

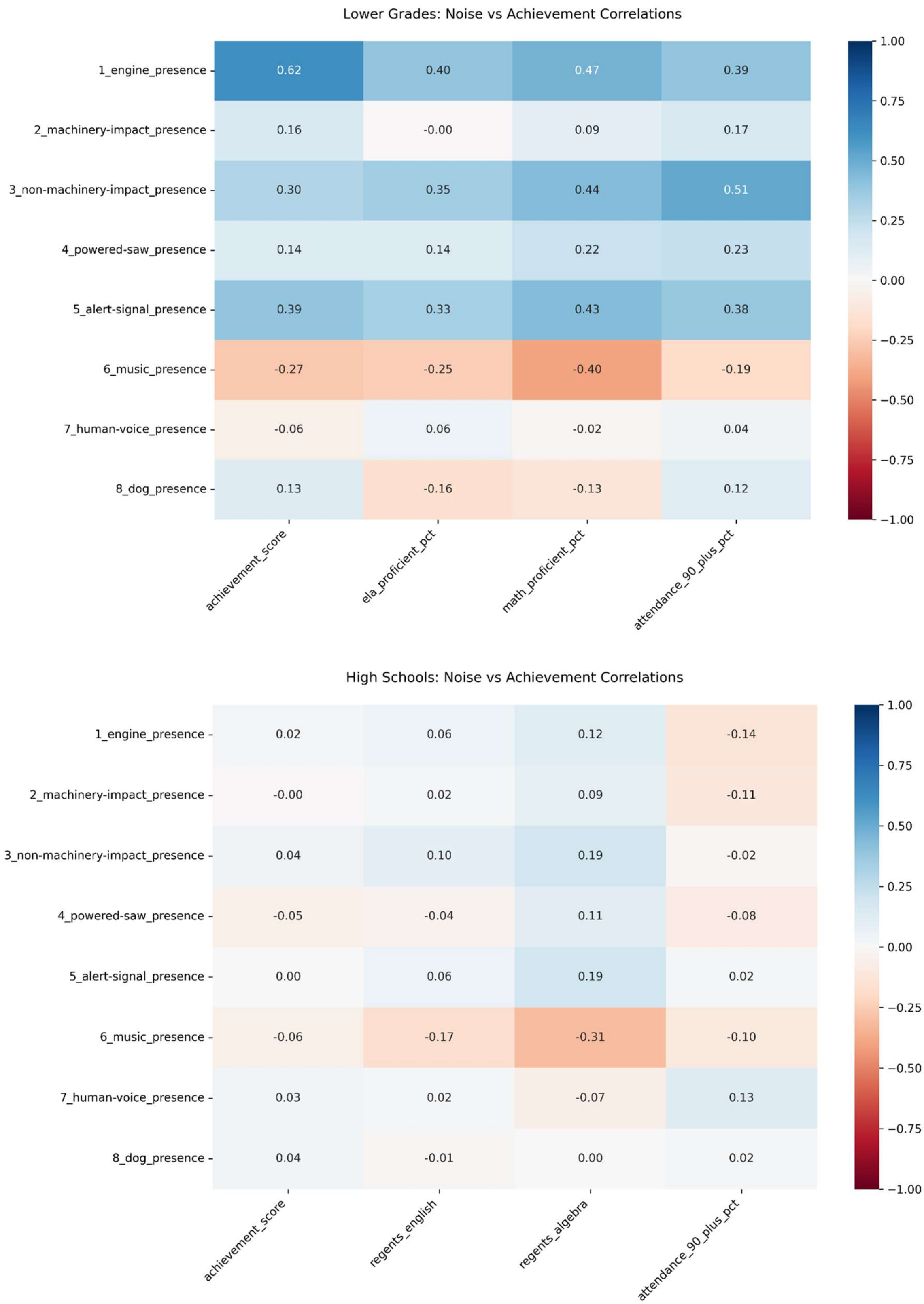

Figure 5. Correlation heatmaps between sound types and various achievement metrics. Note that for lower grades (top), correlations are stronger across the board than they are for high school grades (bottom).

3.4 GEOGRAPHIC PATTERNS

Spatial analysis revealed distinct patterns in noise exposure across New York City schools. Schools within our 2km sensor radius showed varying levels of noise impact, with some geographic clusters demonstrating higher resilience to noise exposure than others. This suggests that local factors, such as building design, surrounding infrastructure, or neighborhood characteristics, may moderate the relationship between noise and academic performance. See Figure 5 for a map of schools categorized by cluster.

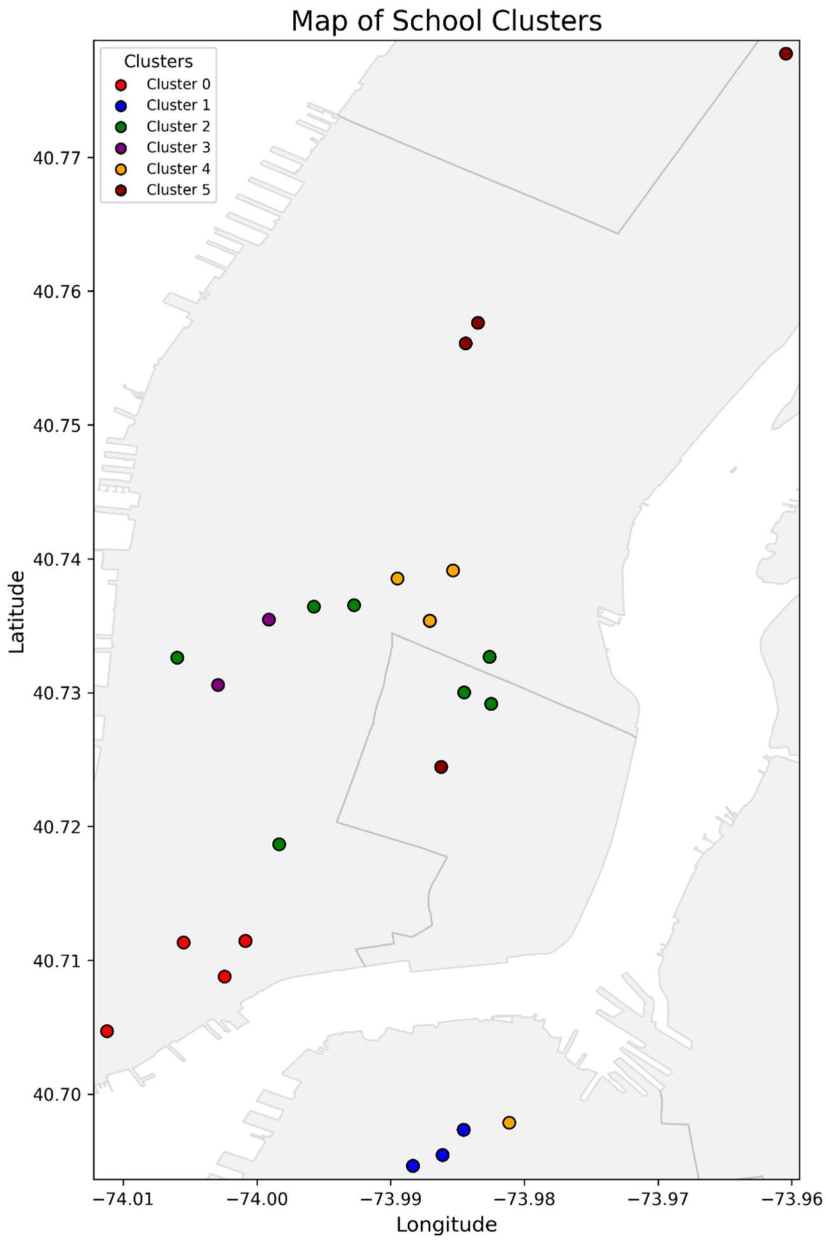

Figure 6. Schools categorized by cluster visualized on a map.

4 Discussion

4.1 Interpretation of Findings

The relationship between urban noise and academic achievement in New York City schools reveals several unexpected patterns. The positive correlation between alert signals (coefficient: 1.530) and academic achievement appears counterintuitive at first glance. However, this finding may reflect underlying neighborhood characteristics rather than direct causation. Areas with frequent emergency vehicle activity often coincide with well-resourced neighborhoods or proximity to hospitals and fire stations, which might correlate with other beneficial community factors.

The strong negative association between music (-2.080) and academic performance presents an interesting challenge for urban planning. This could indicate issues with sound isolation in school buildings, particularly in areas with high density of entertainment venues or outdoor community spaces. The negative impact of dog-related noise (-1.581) similarly suggests that certain types of ambient neighborhood sounds may be more disruptive to learning than previously considered.

4.2 Policy Implications

These findings have significant implications for educational policy and urban planning. The stronger correlation between noise exposure and academic performance in lower grades suggests that noise mitigation efforts should prioritize elementary schools. Traditional noise reduction approaches often focus on traffic and construction noise, but our results indicate that a more nuanced approach considering various types of urban sounds may be more effective.

The geographic patterns identified in our analysis suggest that local factors may moderate noise impact. This points to the potential effectiveness of targeted interventions based on specific neighborhood characteristics rather than one-size-fits-all solutions. Schools demonstrating higher resilience to noise exposure may offer valuable insights for building design and noise mitigation strategies.

4.3 Practical Considerations

The varying impacts of different noise types suggest the need for tailored approach to noise management in educational settings. While some noise sources may be difficult to control, understanding their relative impact can help prioritize mitigation efforts. For instance, the strong negative correlation with music suggests that attention to sound isolation from entertainment venues should be a priority in school building design and renovation.

Our findings also highlight the importance of considering noise impact in school site selection and urban development planning. The interaction between school location, noise exposure, and academic performance suggests that environmental impact assessments for new school construction or renovation should include detailed analysis of the local acoustic environment.

5 Recommendations

5.1 Immediate Actions

Urban schools should prioritize targeted noise reduction strategies based on our findings about specific noise impacts. For elementary schools, which showed higher sensitivity to noise, implementing sound-absorbing materials and providing noise reduction equipment for students during independent work periods could offer immediate benefits. Schools should also evaluate their alert systems, potentially adopting visual alternatives or graduated volume approaches for non-emergency notifications.

5.2 Infrastructure Improvements

Building modifications should focus on addressing the specific noise types identified as most impactful. Sound isolation improvements should target music-related noise (-2.080 coefficient), particularly in schools near entertainment venues or community spaces. For schools exposed to frequent alert signals, while positively correlated with performance, building designs should still incorporate noise reduction features to minimize disruption while maintaining community connectivity.

5.3 Policy Development

Educational administrators should consider noise exposure when making decisions about school scheduling and resource allocation. The stronger correlation between noise and academic performance in lower grades suggests that younger students should be prioritized for quiet classrooms and noise-mitigated spaces. Additionally, urban planning policies should incorporate our findings about the varying impacts of different noise types when considering new school locations or renovations.

5.4 Community Engagement

Schools should develop partnerships with local communities to address noise concerns collaboratively. This includes working with entertainment venues to manage music-related noise, coordinating with emergency services about alert signal patterns, and engaging with neighborhood groups about community noise management. These partnerships can help develop solutions that balance educational needs with community activities.

5.5 Monitoring and Assessment

Implement ongoing noise monitoring systems to track the effectiveness of interventions. Schools should establish baseline measurements and regular assessment protocols to evaluate the impact of noise reduction strategies. This data can inform future improvements and help identify emerging noise concerns before they significantly impact academic performance.

6 FUTURE RESEARCH DIRECTIONS

This study provides foundational insights into the relationship between urban noise and academic achievement in New York City schools, while also highlighting opportunities for future research.

Several promising directions for future research emerge from our findings. First, longitudinal studies spanning longer time periods could reveal how sustained exposure to different noise types affects academic progress over multiple years. The unexpected positive correlation between alert signals and academic performance warrants deeper investigation into potential confounding variables and community characteristics.

The grade-level variations in noise sensitivity we observed suggest the need for age-specific studies. Future research could explore the developmental aspects of noise tolerance and its impact on specific cognitive functions across different age groups. Additionally, the complex relationship between music-related noise and academic performance calls for more detailed analysis of frequency patterns, timing, and duration of exposure.

Integration with emerging smart city technologies could provide richer contextual data. Combining noise measurements with other environmental factors such as air quality, traffic patterns, and urban development could offer a more complete picture of environmental impacts on education. The geographic patterns identified in our study also suggest the value of comparative analyses across different urban environments.

Finally, the evolution of acoustic monitoring technology, particularly through projects like SONYC (Sounds of New York City), promises to enhance our understanding of these relationships. SONYC aims to deploy sensors that will "listen continuously to ambient outdoor sounds, measure the total volume, and transmit a statistical description of the audio it hears." The maturation of this sensor network would enable more nuanced analysis of temporal patterns, including seasonal variations and time-of-day effects. Such granular data could help identify optimal times for noise-sensitive academic activities and inform more effective intervention strategies.