A Data-Centric approach to School Discipline

Nicholas Ramroop and Patrick Hosein
Department of Computer Science
The University of the West Indies, St. Augustine, Trinidad
nicholasramroop74@gmail.com, patrick.hosein@uwi.edu

***Abstract*—School violence is a major concern for school administrators and policymakers since it has a significant impact on the academic success of students and their mental well-being. This has led to the increased collection of applicable data for better decision making. Unfortunately, this collected data may not always be utilized fully. We take a "Business Analytics" approach to analyze data collected from schools and develop a suitable dashboard. We first propose a unique performance metric that can be easily tracked and on which predictions can be made so administrators can be proactive rather than reactive. Such predictions can also be used for more long term planning such as the determination of the number of security personnel needed at a particular school. In particular, we analyze data collected from schools about student suspensions (both severity and frequency). We also compare our approach with that presently used and illustrate the benefits of our approach.**

***Index Terms*—School Violence, Weighted Metric, School Bullying, Business Analytics, Data Science**

I. Introduction and Literature Review

The school environment plays a vital role in developing minds, academic performance, and future opportunities for students. Research has shown that school violence can impact students, teachers, and the overall learning environment. In addition to academic failure, school violence has been found to be associated with a myriad of behavioral, social and cognitive problems in students [1]. Violence in schools has also been shown to take a negative toll on the mental health of students, further affecting their academic performance. The authors in [2] found that adolescents, who experience school violence more frequently, exhibited a higher risk of developing depression and suicidal ideation.

This topic remains a pressing issue in the education system. Aggression and violence can manifest itself in several forms of varying severity like physical altercations, verbal abuse, bullying, and other negative behaviors. This problem is complex and stems from a plethora of economic, social and psychological factors that interact in intricate ways [3].

School safety, a key component of school climate, is one of the major concerns of school administrators, teachers, students, and others [4]. Finding a way to measure violence is critical to understanding the nature, magnitude, and prevalence of violence in schools, its impact on students and teachers, and what can be done to address this problem.

Several past studies proposed ways to quantify school violence based on different factors in the school environment. Some examples include the School Violence Intensity Scale (VES) and the school violence index. The VES scale assesses violence in terms of verbal and physical aggression between students of a school environment [5].

In [6], the authors worked to develop a composite metric called the school violence index that encapsulated the violence and peace experienced by students in schools. Their study dissected the term ”violence” into three divisions: direct, structural, and cultural violence. Direct violence was described as the intentional physical and psychological harming of others. Structural and cultural violence were categories that assessed violence in terms of the "social and institutional mechanisms that alienate people from their basic rights" [7]. This study focuses on direct violence as described in [6], however, an alternative method of quantifying violence in schools is explored.

Similar to previous research, the goal of this study was to develop a composite weighted metric that represents school violence. Many prior studies have used questionnaires and self-report surveys as their primary data collection methods and the definitions of violence tends to vary. For example, in Quammie et. al [8], the authors analyzed data collected from surveys taken in several schools. This data was then analyzed to show the correlation between violence in a school and the performance of its students. Note however, that this does not indicate cause since the schools with poor performing students would have attracted those with more violent behavior.

Unlike [5] and [6] where self-report surveys were used to assess violence in schools, the suspensions data from 116 schools was used as a proxy to measure school violence. The self-report surveys on which previous metrics are based, rely heavily on subjective recall and personal perception. Our approach offers a more scalable approach to quantifying violence with reduced bias and improved comparability across different populations.

Studies have shown that schools with high numbers of suspensions are more likely to have a culture and history of violence. Suspensions are directly associated with violent incidents in schools since they only stem from student misbehavior. Additionally, the severity of suspensions can be an indicator of violence in schools. The authors in [9] have found that schools with students who have been suspended for violent incidents, and for a longer duration of time are more likely to have a prevalence of violent behavior.

We introduce a metric that considers both the frequency and severity of infractions that resulted in suspensions. The length of suspensions and the number of students present in each school were also taken into consideration when developing this metric. The impact of security personnel and teacher attendance on this weighted metric was examined. Of 116 schools used in the study, 26 were designated as “Schools of Focus” by the school governing body. These Schools of Focus were identified as requiring additional support from the government in improving their student outcomes, particularly in literacy and numeracy. The violence in these schools was also examined using the weighted metric and compared to schools which were not given the aforementioned designation.

II. Methodology

Violence in schools is of increasing concern to school administrators. Additional emphasis is being placed on the monitoring of serious violent offenses in schools such as bullying, assaults, fighting, extortion, etc. The objective of this research was to provide a universal metric that measures violence in schools, taking into account the varying severity of different suspension categories and the number of students present in each school, and then using this metric to derive insights into items that may positively or negatively impact school violence.

The methodology of this study involves a comparative analysis of the methods used in [10], the National School Discipline Matrix of Trinidad and Tobago using a dataset of suspension data for Terms 1, 2 and 3 of the 2023-2024 academic year. The National School Discipline Matrix uses purely the numbers of student suspensions to compare the level of indiscipline for each educational district and class level in schools. Analysis on a per-school basis was not presented in the School Discipline Matrix, however, the same drawbacks with this method exist when examining violence on a district and class level. These are as follows:

- **Population Size:** Using the number of suspensions does not consider the differences in student populations when ranking schools and education districts.
- **Severity of Infractions:** The Government [10] includes a predefined list of suspension categories, their relative severity, and specifies the disciplinary actions that school administrators are required to take when dealing with students who commit an infraction from each category. The aforementioned analysis method using numbers of student suspensions, however, does not consider the severity of suspensions in its analysis.

Data on the number of suspensions for each school in the sample country is submitted daily to the school administrative body. For each suspension, detailed information including the offense/suspension category, duration, number of repeat offenses committed by the student, etc. are recorded. A sample of this dataset was used which contained the suspensions reported by 116 schools for the 2023-2024 academic year. This data was separated into three school terms; Term 1 - September to December, Term 2 - January to March and Term 3 - April to July. Each row in the sample dataset represented a single student suspension. The duration of each suspension was used as a weight which reflected the severity of the infraction committed by a student. The data was aggregated by school and time period, ensuring that student privacy was not a concern. The Ministry [10] specifies the disciplinary actions that must be taken when a student commits an infraction. Infractions are classified as belonging to one of three levels; Minor, Major and Severe.

In addition to the level of each offense, the number of times a student repeated an offense is also considered when specifying the duration of a suspension. As a result, developing a metric that uses the length or duration of a suspension to quantify violence encapsulates both the severity, as well as the prevalence of repeat offenses. That is, if many students repeat minor offenses, the length of their suspension increases, which leads to their school having a higher violence value using our weighted metric, thus indicating that some increased attention should be given to the school. Table II contains the disciplinary actions for different levels of infractions; Minor, Major and Severe.

If we denote, for some school with population P, the set of suspensions over some period T by S_T and denote the number of suspension days for suspension i by v_i then we define the violence rate of the school by:

$$R = \frac{1}{PT} \sum_{i \in S_T} v_i \tag{1}$$

We also compute the violence rate of all schools using the above formula but taking the summation over all schools and denote this by $\bar{R}$. Our violence metric for a school is then given by $M = R/\bar{R}$ and so value above 1 indicates the school is more violent than average.

This metric was used to evaluate the violence in schools over different time periods. It should be noted that the names of the schools have been replaced with generic names, however the names remain consistent for all comparisons and the data used is from a real-life dataset.

After the overall ranking of schools was generated using the entire dataset, the metric was calculated on a weekly basis for each school. This weekly data was used to look for trends and correlations between the weighted metric and the teacher to student ratio in a school, the number of security personnel (for schools with additional security personnel) and the total enrollment in the schools. Comparisons were also drawn between the schools with additional security and those without using the weighted metric.

Finally, an attempt was made to investigate the trends in overall school violence using the weighted metric. The weekly metric for all schools ($\bar{R}$) was used for this trending. Geometric or exponential smoothing was applied to the dataset. This technique is used in time series analysis to improve the smoothness of the data and reduce the impact of noise. It uses a geometrically weighted moving average of the past observations to estimate the current value. The authors of [11] also noted that this technique can assign exponentially decaying weights over the time series so earlier values in the series have a decreasing impact on the current smoothed value. One drawback is that geometric smoothing can be sensitive

TABLE I
SCHOOL DISCIPLINE MATRIX ACTIONS BY OFFENSE LEVEL

Level of Offense	First Offense	Second Offense	Third Offense
Minor	Teacher Conference	Parent Conference and/or detention or 1–2 days of in-school suspension	Parent Conference and 3–5 days of out-of-school suspension, escalation to Major level if continued.
Major	Parent Conference and possible in-school detention	Parent Conference and/or detention or 1–2 days of in-school suspension or 3–5 days out-of-school suspension	7 days out-of-school suspension, referral to external social services, escalation to Severe level if continued.
Severe	Parent Conference, 3–5 days out-of-school suspension, referral to external social services, and/or Police	Parent Conference, 7 days out-of-school suspension, possible extended suspension, referral to external social services, and/or Police	Expulsion and referral to alternative educational institution, referral to external social services, and/or Police.

to outliers in the data. The formula used to carry out this smoothing on the dataset is given by the formula:

$$S_t = \alpha X_t + (1 - \alpha)S_{t-1} \tag{2}$$

where S_t is the smoothed value for time t, α is the variable that specifies the degree of smoothing, i.e., the smoothing factor, and X_t is the actual value at time t. The optimal value for α was selected by varying the value of alpha from 0.1 to 0.9 and selecting the value that had the lowest mean square error when predicting future values. We found this value to be $\alpha = 0.5$.

III. RESULTS AND DISCUSSION

The first item explored using the weighted metric was the ranking of schools. Schools were ranked by metric value for the entire academic year. The schools were then sorted in order of largest metric value to smallest metric value. Schools higher on the list are said to have the highest national violence. Schools were also ranked using the number of suspensions (without accounting for severity) to illustrate the differences in the rankings. In this case we divide number of suspensions of a school by the average suspension over all schools.

Figure 1 contains a ranking of the 25 schools from the sample dataset with the highest values of the weighted metric. The school with the highest metric value, Crestwood Secondary (shown in green), is six times more violent than average. However it is not considered a very violent school when looking at the Schools of Focus.

Figure 2 contains a ranking of the 25 schools taken from the sample dataset with the highest number of suspensions, as presently done. By ranking schools based on their number of suspensions, the sample school, Crestwood Secondary, is now the 19th school in the list, instead of the first. The same comparison is done for the top school in Figure 2. The most violent school was found to be Silver Creek Secondary. As seen in Figure 1, when using the weighted violence metric, this school is ranked as the 17th most violent school.

The reason for this difference is highlighted when the attendance and populations of the schools are considered. The student population of Crestwood Secondary was 283, with an average attendance of 135 students for the entire sample period. Silver Creek Secondary however, had a student population of 1026 and an average attendance of 554 students for the same period. Using the weighted metric, as illustrated in Figure 3 and 4 we see that Crestwood Secondary has a notably higher violence per student, even though its population is significantly lower than the larger school.

Using these results, it can be seen that it is crucial to consider the relative populations of schools when comparing their violence, since schools with higher populations are more likely to have a larger number of suspensions.

The violence of each school is also an important consideration. Taking two of the schools with similar average attendance and weighted metric values, Meadowbrook Secondary and Mountain Ridge High School, we can illustrate the importance of using a weighted metric to assess the violence in schools. Both schools had an average student attendance of approximately 255 students per day over the sample time period. However, despite the similarity in weighted scores and attendances, Meadowbrook Secondary had 34% fewer suspensions. This is a result of the differing proportions of suspensions lengths issued to students for their various infractions. Figure 5 illustrates the frequency of suspension durations for both schools. Although Meadowbrook Secondary had significantly fewer suspensions, the severity of their suspensions was significantly higher, hence increasing their score.

In addition to ranking on a school level, the districts were ranked using the same sample data for the academic year 2023-2024 using both the presently used ranking and the weighted metric. As illustrated in Figure 6 and Figure 7 when calculating the violence and the difference in attendance, there is a significant difference in the order in which districts are ranked.

As mentioned above, using a weighted metric is also useful to draw a more accurate comparison between schools of focus and others. Next we investigate the temporal aspects of this metric and look at the weekly variation. We also investigate

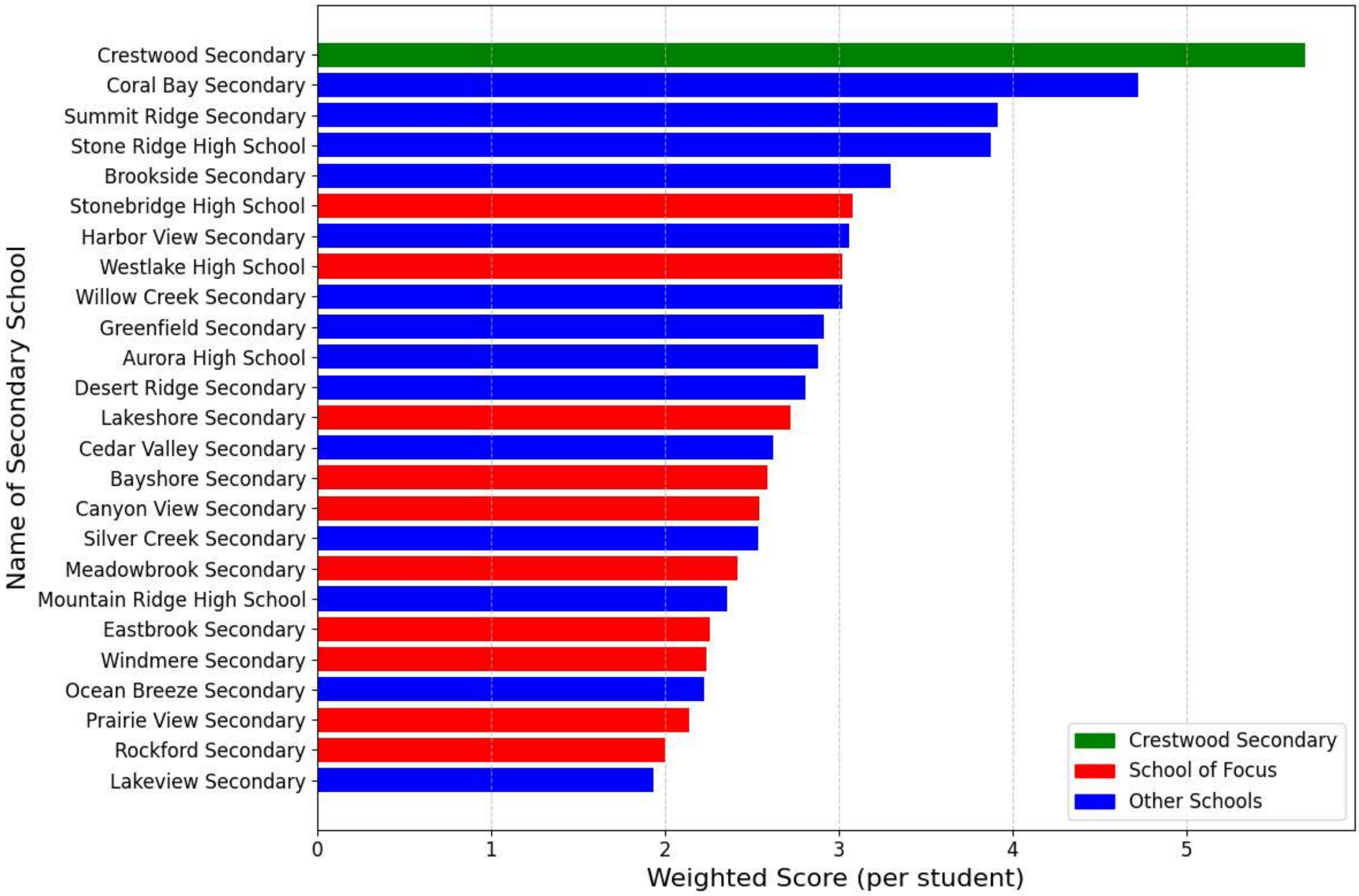

Fig. 1. Top 25 schools by Proposed Metric M

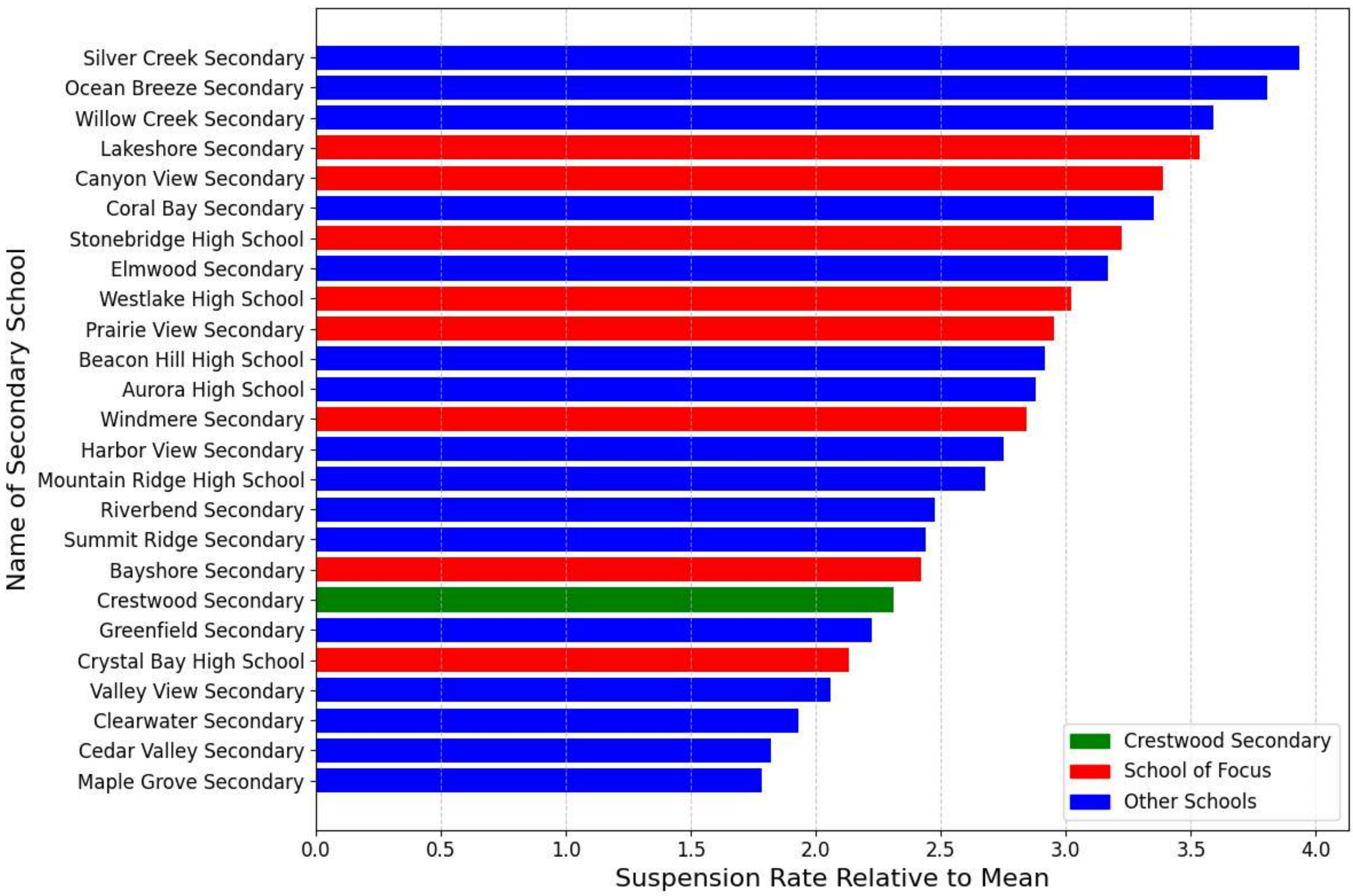

Fig. 2. Top 25 schools by Number of Suspensions normalized by the average suspensions per school

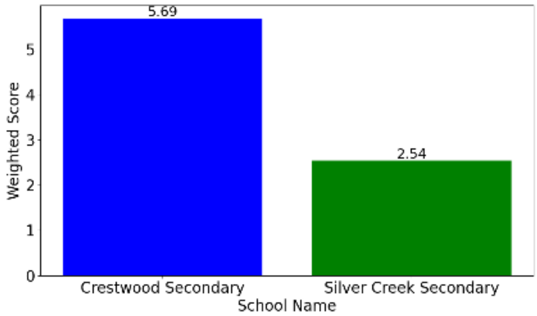

Fig. 3. Comparison in Weighted Violence metric between Schools

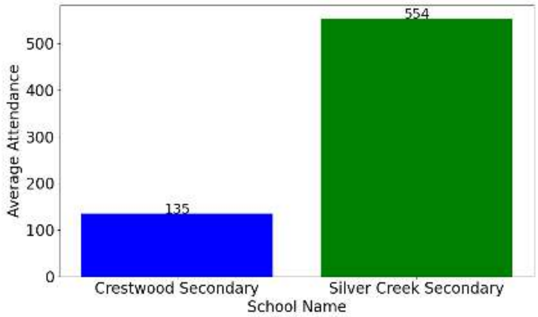

Fig. 4. Relative Attendance of Highlighted Schools

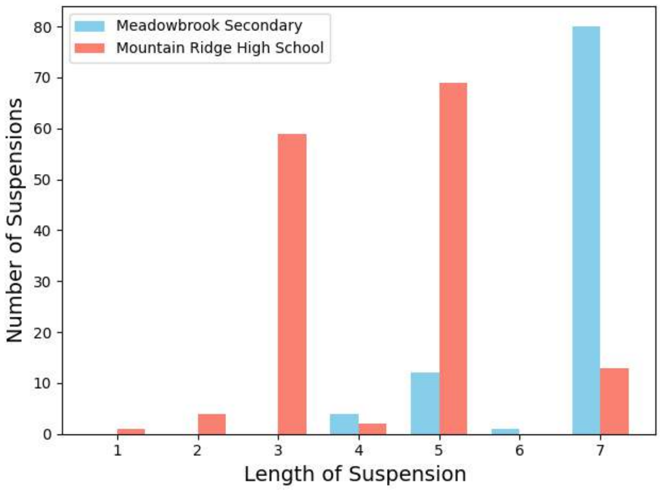

Fig. 5. Distribution of suspensions for Schools with Similar metrics

which factors are correlated with violence such as Teacher to Student Ratio, security attendance, health and safety officers, and the total enrollment of students in a school. To do this, the weighted metric was calculated on a weekly basis for the entire sample period.

Spearman correlation coefficient was chosen as the metric of choice due to its ability to capture both linear and non-linear relationships. It is also less sensitive to outliers and scale but can be affected by tied values. Conducting correlation analysis on the dataset, no notable correlation was found between the weighted metric and teacher attendance rate, school enrollment or the number of security personnel. It should be noted

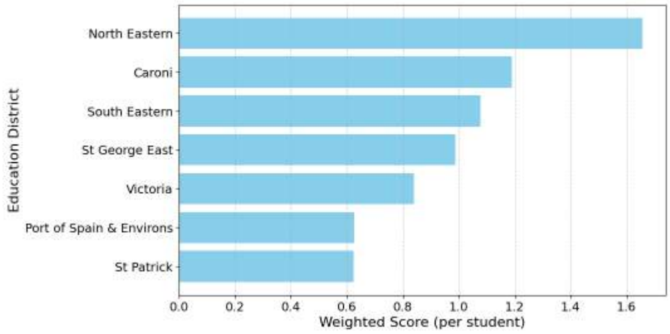

Fig. 6. Districts Ranked by Weighted Violence Metric.

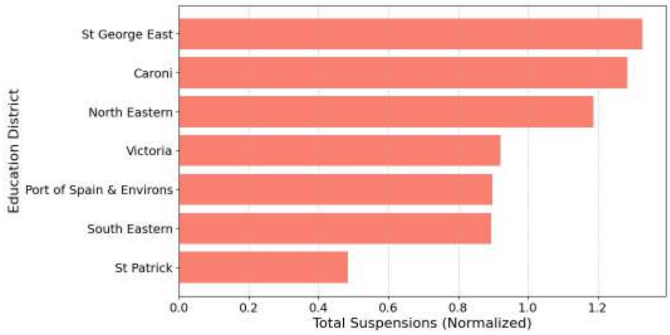

Fig. 7. Districts Ranked by number of Suspensions.

however, that schools with additional security, and health and safety personnel had an average weighted metric that was 86% higher than that of schools without. An independent t-test with a p-value of less than 0.05 confirmed the statistical significance of this comparison. This is illustrated in figure 8 and clearly indicates that adding health and safety, and security personnel to the violent schools did not seem to address problems and that a different approach was needed. This may be due to external factors such as school location, as those in high crime areas can be influenced by broader social dynamics that can limit the effectiveness of additional health and safety, and security personnel.

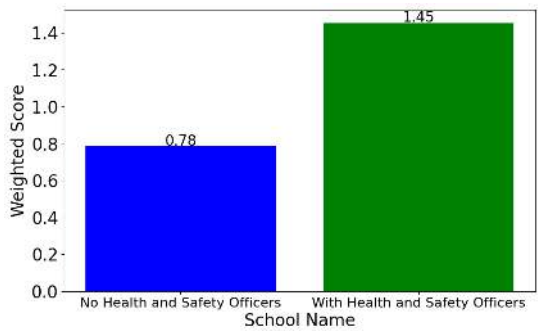

Fig. 8. Weighted Metric Score for schools with/without Health and Safety Officers

Figure 9 shows a weekly plot of the weighted metric for all schools across the entire period covered by the sample data. The raw values were plotted and then smoothed using geometric smoothing with an alpha value of 0.5 as it was a good balance of reducing noise and maintaining the overall trend in the data. The goal of this portion of the analysis was to identify trends in school violence and correlate irregularities with real world events as well as use trend and seasonality to make a prediction of the metric score for an upcoming term. The X-Axis is labelled in value pairs where the first value represents the term and the second value represents the week number in each term (e.g. A value of 2-4 is Term 2, week 4).

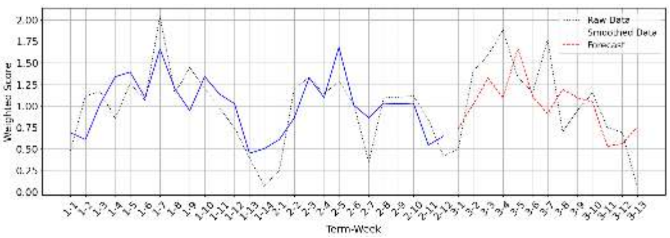

Fig. 9. Termly Trend in Weighted Scores by Week (Exponential Smoothing)

As illustrated in the plot, the graph for each term follows a similar general trend where violence is lowest at the beginning and ends of each term with a peak near the center of the term. Two notable dips in violence were observed in Term 2 and Term 3. In Term 2, a drop in violence across all schools was observed in week 7. This is directly aligned with the week of Carnival 2024. Furthermore, there was a decrease in violence in week 6 of the third school term. A possible explanation for this is that many schools hosted Caribbean Secondary Math and English Examinations and were not carrying out usual school operations hence, exhibiting a decrease in violence.

Using geometric smoothing while taking into account trend and seasonality to predict the metric scores for Term 3 of the academic year yielded a mean squared error (MSE) of 0.20 and a root mean squared error (RMSE) of 0.45. Additionally, the ARIMA (Autoregressive Moving Average) model was used to forecast the weighted score for the same time period. ARIMA is another widely used method in time series forecasting. This model combines autoregression, differencing and moving averages to gain insight into the trends in the time series data. A grid search was conducted to find the optimal hyperparameters for the model, which resulted in a forecast that also had a mean squared error of 0.20. However, as seen in 10, much of the variation that was preserved with exponential smoothing, was not present in this forecast.

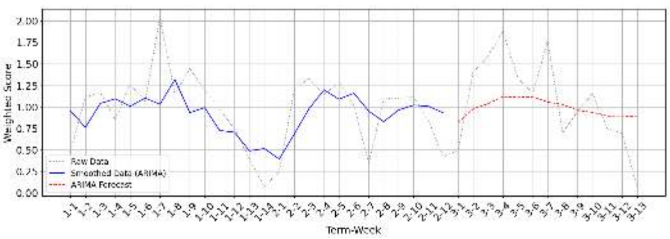

Fig. 10. Termly Trend in Weighted Scores by Week (ARIMA)

IV. Conclusion

The primary objective of this study was to develop an improved method of quantifying violence in schools. By using school suspension data as the key metric the aim was to take into consideration the relative severity of different infractions that lead to suspensions and provide a more accurate representation of school violence. Our metric was able to successfully capture this while accounting for the different populations and attendance rates in schools.

This weighted metric has shown that suspension data can be a valuable tool for measuring the intensity of violence in schools. Due to the confidentiality of the data, individual incident reports could not be used, which limited the granularity of the analysis. However, further refinement of this model is necessary. Incorporating environmental factors such as student demographics, socioeconomic status, and academic performance data could enhance the comprehensiveness of the analysis. Incorporating these additional factors could be instrumental in applying machine learning techniques in future research to improve the model's predictive power and provide insights into even deeper patterns in school violence. These additions would also offer a deeper understanding of how school violence relates to academic outcomes and how these patterns are influenced by various environmental conditions.

References

[1] T. Shim and C. Ye, "Teachers' roles in coping with school violence from the perspectives of prospective teachers: A q methodological approach," *Behavioral Sciences*, vol. 14, p. 1099, 11 2024.

[2] E. LEE, "Effects of school violence on depression and suicidal ideation in multicultural adolescents," *Iranian Journal of Public Health*, 05 2019.

[3] M. P. Wijeratne, R. Seneviratne, N. Gunawardena, T. Østbye, and C. Lynch, "Descriptive study of the psychosocial and physical environments of school in relation to violence among adolescents in the gampaha district of sri lanka," *Sri Lanka Journal of Child Health*, vol. 44, p. 138, 09 2015.

[4] C. Crawford and R. Burns, "Support or punishment practices: What works to reduce school violence," *Social Sciences*, vol. 9, p. 220, 11 2020.

[5] M. Garcés-Prettel, Y. Santoya-Montes, and J. Jiménez-Osorio, "Influence of family and pedagogical communication on school violence," *Comunicar*, vol. 28, pp. 77–86, 04 2020.

[6] D. Capistrano, S. Sloan, J. Symonds, E. Samonova, C. Sugrue, and D. Devine, "Assessing experiences with violence and peace in primary schools in sierra leone," *Journal of Aggression, Conflict and Peace Research*, vol. 14, pp. 231–243, 01 2022.

[7] J. Galtung, "Violence, peace, and peace research," *Journal of Peace Research*, vol. 6, p. 167–191, 1969. [Online]. Available: https://www.jstor.org/stable/422690

[8] M. Quammie and P. Hosein, "School climate factors as predictors of school performance: A machine learning approach," in *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, 2024, pp. 1312–1316.

[9] K. Kirkwood and W. Richardson, "Profiling expelled students," *Brock Education Journal*, vol. 15, 01 2006.

[10] "Government of the republic of trinidad and tobago national school discipline matrix," 2022. [Online]. Available: https://storage.moe.gov.tt/corporate/2022/09/School-Discipline-Matrix-Approved.pdf

[11] Q. Zhou, "Portfolio optimization with robust covariance and conditional value-at-risk constraints," 2024. [Online]. Available: https://www.semanticscholar.org/paper/Portfolio-Optimization-with-Robust-Covariance-and-Zhou/92d79bcc98f29ee472a3d01f09dc7c0efffbb64f