Education in Danger Incidents: Analyzing Schools and Universities Attacks in War and Conflicted Areas and the Effects on Students, Children and Educators

George Mason University

AIT-580-004 | Prof. Dr. Kazi Kabir

Author: Thy Phan

Abstract

The goal of this research is to analyze schools and universities' attacks in wars and conflicted areas. Previous studies with related topics had been analyzed by qualitative methods, which consisted of case studies, interviews and historical. There is a limitation of research using quantitative methods. This paper will examine these education related incidents over years, months to understand the trends using quantitative methods. The paper also focuses on analyzing number of students attacked, killed and injured and sexual violence cases across different countries and geographic locations. The events descriptions of every incident will be thoroughly analyzed using sentiment analysis to understand the dataset. The dataset has been cleaned, analyzed and visualized or applied natural language processing (NLP) by using Python, R programing languages and SQL queries. The finding reveals that education related violences events increased gradually over time with increasing number of students attacked and distributed across different countries and geographic locations. Many students have been killed, kidnapped and injured in these schools and universities' attacks over year of 2020 to 2024.

Keywords: Education in danger incidents, students attacked, students killed, students injured, SQL, R, Python, NLP

i. INTRODUCTION

Education is among the most important thing to humans, which can help people acquire new knowledge, explore different cultures, develop critical thinking and creative skills and gain access to better opportunities or living standards. Unfortunately, in multiple locations affected by wars and conflicts, children have limited access to education or do not have adequate resources to deepen their knowledge. To explore this topic further, my dataset named 2020-2024 Education-in-danger-incident-data, will focus on education in conflicted and affected areas and its negative effects on students and children.

Previously, there were many research using qualitative methods such as interviews, case study, historical to discuss education in wars and conflicted areas but there is a limitation of study using quantitative method to analyze the problem. Thus, the research is highly important since it focuses on finding patterns of education in danger events. By analyzing different event descriptions, number of attacks, number of people injured, geographic location, we will have more insights to create a protection plan that reduce number of people injured, minimize education barriers for students and children. Also, different humanitarian and advocacy groups can use this research to further their analyses to protect vulnerable population.

This research paper aims to analyze fourth following research questions:

1. How do education related violence events change over time, and which countries or location have the highest frequency of incidents?
2. What is specific month or time of the year that will have the higher number of incidents?

3. Can incidents be classified in different groups based on characteristics of event description?
4. Which countries have the highest number of students attacked, injured and killed?

II. LITERATURE REVIEW

My literature review will synthesize and summarize findings from three different sources that discuss the educational challenges in affected zones and methods have been used to encourage education among students. For instance, Lin [1] discussed barriers to education in armed conflicted areas in Myanmar and the expectations for non-governmental education leaders. Mutu [2] mentioned that millions of children in Nigeria cannot access to education due to various reasons and how to utilize technology to gain access to education for Nigerian children. Akar [3] also discussed about education reform in conflicted areas and ways to empower teachers to promote better education.

A. Educational Challenges in Myanmar

Because of the ongoing civil war in Myanmar since 1949, students still have educational barriers, which included lack of security, adequate academic resources, infrastructure and inclusive policy at national level [1]. Specifically, students have limited access to textbooks or desks and chairs in the classrooms. According to Lin [1], "Myanmar's National Education Strategic Plan was created by the central government without consultation with ethnic minorities within the country", so it failed to provide inclusivity to these students in armed conflicted areas. The author used critical ethnographic methodology to conduct the research regarding possible solutions to reduce the negative effects of civil war to education.

There are two education leaders have been interviewed and an ethical approval has been provided by the Flinders University Social and Behavioral Research Ethics Committee to conduct this study. The educational challenges are included but not limit to lack of qualified teachers due to security reasons and food access and the language barriers for ethnic groups. The study also showed that Burmese was the language barrier to ethnic minorities students to access to education since they did not understand the language and it triggered their trauma of the civil war [1]. The author discussed strategies to resolve human resources challenges, curriculum and language issues, financial issues, high dropout rates as well as the responsibilities of education leaders.

B. Education Challenges in Northen Nigeria and the use of technology to deliver basic education to students

Mutu [2] divided the article in three parts: security situation in Northen Nigeria and its effect on education, government responsibilities regarding the conflicts and the right to access to education for students in affected areas and the utilization of technology to support education in conflict environments. The author mentioned that ongoing conflicts in Northern Nigeria has resulted in the closure of several schools by the government without the alternative methods of accessing to education. Also, Mutu [2] discussed in detail the basic human rights to education even in conflict situation and Nigerian state responsibilities. According to Mutu [2], "even in a conflict situation, children have the right to education which should be respected, protected and fulfilled by the state."

There are three technologies can provide access to education in conflicted areas, which includes mobile phone-based delivery of educational content, internet-enabled computer lab supporting education, and interactive radio instruction (IRI). For instance, a simple phone (non-Android), which can send and read text messages and a SIM card, can deliver basic education to students in conflicted areas. Furthermore, interactive radio instruction (IRI) approach can be delivered by audio teachers to the students regarding different topics such as life skills, health, conflict prevention, health and mediation. The author mentioned the Somalia Interactive Radio Instruction program, which has delivered basic education to 330,000 children from grade 1 to 5. By providing education through IRI, the program offered knowledge, skillsets and the opportunities to move out of poverty and conflict areas for these people. Regarding the internet-enabled computer lab supporting education, even it is the amazing technology to provide access to education, it can be challenging in Nigeria due to its high set up and maintenance costs. The government also need to shutdown network activities because of terrorist groups, so the internet-enabled approach is not feasible [2].

The author concluded that the interactive radio instruction (IRI) technology is the best approach for Northern Nigeria after exploring three different technologies [2]. According to Mutu [2], the IRI has implemented during COVID-19 lock down and delivered significant results.

C. Education Reform

According to Akar [3], education reform after destructive conflicts is highly sensitive to collective memories of armed conflicts, social injustices, damaged facilities or any potential triggers that can cause new episodes of violence. The educators tend to choose avoid conflict approach in the classrooms, which include preventing discussions regarding controversial issues or any politics and religions topics. Akar [3] mentioned that lack of discussing controversial issues or competing narratives of the wars will limit students' abilities to address social injustices. More importantly, it can make students think that diversity is the roof of destructive conflicts and risks the freedom to express political and religious identities [3].

Akar [3] discussed about high-impact history education teachers in Lebanon, who engaged students in various discussions regarding controversial issues and start a history of education reform. Despite challenging the government's direction of the single narrative of the war, these high-impact teachers have found the confidence and courage to create the safe and collaborative learning environment. Thanks to the innovative work of citizenship education teachers, education has transformed from" an authoritarian-like culture of learning towards a deliberative and empowering one" [3]. Akar [3] concluded that the new approaches from the high-impact teachers are better than any curriculum passed by the government agencies or civil society.

All these three studies have discussed educational challenges in conflicts and affected areas and the strategies can implement to provide access to education. These studies utilized qualitative methods such as interviews, case study, historical to explore these findings. Three authors provided adequate evidence and sources for education access in affected areas. There is a limit of research using quantitative method to explore education in armed conflicted areas. Thus, I will use my project dataset to conduct the study using quantitative method.

III. DATASET

The dataset, *named 2020-2024 Education-in-danger-incident-data*, was taken from the **Humanitarian Data Exchange** website, a platform designed to facilitate sharing, discovery and use of humanitarian data. The dataset was contributed by **Insecurity Insight,** the organization that collects data about violence against civilians, damages and destructions to affected areas. The data will provide supports for the work of aid agencies, healthcare providers, protection services or civil organizations to create protection and intervention plans. The dataset

was published in the *Education in Danger Monthly News Brief*, a complete overview of safety, security and access incidents that impacted education. It contains detailed records on number of schools and universities attacks, the number of students and educators have been injured, kidnapped and killed, the countries and geographic locations where these attacks occurred, the weapons used, the penetrators reported, and event descriptions of these incidents.

The dataset was categorized by country in various locations Afghanistan, Algeria, Angola, Armenia, Australia, Azerbaijan, Bangladesh, Belarus, Brazil, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Chile and many others. The data covers from **January 1, 2020**, to **September 15, 2024,** and has all four data types: **nominal, ordinal, interval and ratio**.

Figure 1

Dataset			
Attribute	**Data Type**	**Attribute**	**Data Type**
Country	Nominal	Location of Event	Nominal
Event Description	Nominal	Reported Penetrator	Nominal
Country ISO	Nominal	Reported Perpetrator Name	Nominal
Admin 1	Nominal	Date	Interval
Geo Precision	Ordinal	Type of education facility	Nominal
Weapon Used/Carried	Ordinal	Latitude	Ratio
Attacks on Schools	Binary Nominal	Longitude	Ratio
Attacks on Universities	Binary Nominal	Forced Entry into education facility	Binary Nominal
Military Occupation of Education facility	Binary Nominal	Arson attack on education facility	Binary Nominal
Damage/Destruction to Ed facility Event	Binary Nominal	Attacks on Students and Teachers	Binary Nominal
Educators Killed	Ordinal	Students Killed	Ordinal
Educators Kidnapped	Ordinal	Students Kidnapped	Ordinal
Educators Injured	Ordinal	Students Injured	Ordinal

Educators Arrested	Ordinal	Students Arrested	Ordinal
Known Educators Kidnap or Arrest Outcome	Nominal	Known Student Kidnap or Arrest Outcome	Nominal
Students Attacked in School	Ratio	Sexual Violence Affecting School Age Children	Ratio
SiND Event ID	Nominal		

IV. TOOLS AND METHODS FOR DATA ANALYSIS

First, I will use Python to load the dataset and perform data cleaning process in Jupyter Notebook. Second, I will do some data analysis in Python, which also includes NLP for Event Description Column. Third, I will do some data visualizations and clustering in R programming Language. Finally, I will create the schema and table of the data in SQL and perform some queries to get total number of students and educators killed, kidnapped, injured or arrested.

A. Data Cleaning

First, I start the data cleaning process using Python by first few rows of the dataset and basic statistic of numerical columns to understand structure of the dataset. Then, I try to check if there any missing values in these columns to start the cleaning process. I replaced two missing values in categorical column "Admin 1" with "Unknown" value and filled 369 missing values in both two numerical columns "Latitude" and "Longitude" with 0. For categorical column "Location of event", I replace one missing value with "None". After handling the missing values, I also try to verify if any missing values left in these columns.

The "Known Educators Kidnap Or Arrest Outcome" column contains 3248 rows that have "None", "No Information" and "Missing Value" in total of 3472 rows in the Excel file, which will not give any meaningful information for further analysis. Similarly, the "Known Students Kidnap Or Arrest Outcome" column has 3397 rows that

contain "None", "No Information", "Missing" value in total of 3471 rows, which will not deliver any valuable insights if analysis. As a result, I will drop these two columns from my dataset. Figure 2 showed the dataset before and after handling missing value.

Figure 2

```
Missing values:
 Date                                              0
Event Description                                  0
Country                                            0
Country ISO                                        0
Admin 1                                            2
Latitude                                         369
Longitude                                        369
Geo Precision                                      0
Location of event                                  1
Reported Perpetrator                               0
Reported Perpetrator Name                          0
Weapon Carried/Used                                0
Type of education facility                         0
Attacks on Schools                                 0
Attacks on Universities                            0
Military Occupation of Education facility          0
Arson attack on education facility                 0
Forced Entry into education facility               0
Damage/Destruction To Ed facility Event            0
Attacks on Students and Teachers                   0
Educators Killed                                   0
Educators Injured                                  0
Educators Kidnapped                                0
Educators Arrested                                 0
Known Educators Kidnap Or Arrest Outcome        2924
Students Attacked in School                        0
Students Killed                                    0
Students Injured                                   0
Students Kidnapped                                 0
Known Student Kidnap Or Arrest Outcome          3312
Students Arrested                                  0
Sexual Violence Affecting School Age Children      0
SiND Event ID                                      0
dtype: int64
```

```
Date                                              0
Event Description                                 0
Country                                           0
Country ISO                                       0
Admin 1                                           0
Latitude                                          0
Longitude                                         0
Geo Precision                                     0
Location of event                                 0
Reported Perpetrator                              0
Reported Perpetrator Name                         0
Weapon Carried/Used                               0
Type of education facility                        0
Attacks on Schools                                0
Attacks on Universities                           0
Military Occupation of Education facility         0
Arson attack on education facility                0
Forced Entry into education facility              0
Damage/Destruction To Ed facility Event           0
Attacks on Students and Teachers                  0
Educators Killed                                  0
Educators Injured                                 0
Educators Kidnapped                               0
Educators Arrested                                0
Students Attacked in School                       0
Students Killed                                   0
Students Injured                                  0
Students Kidnapped                                0
Students Arrested                                 0
Sexual Violence Affecting School Age Children     0
SiND Event ID                                     0
dtype: int64
```

The dataset has been checked for duplicate rows and correct data types. The data cleaning process
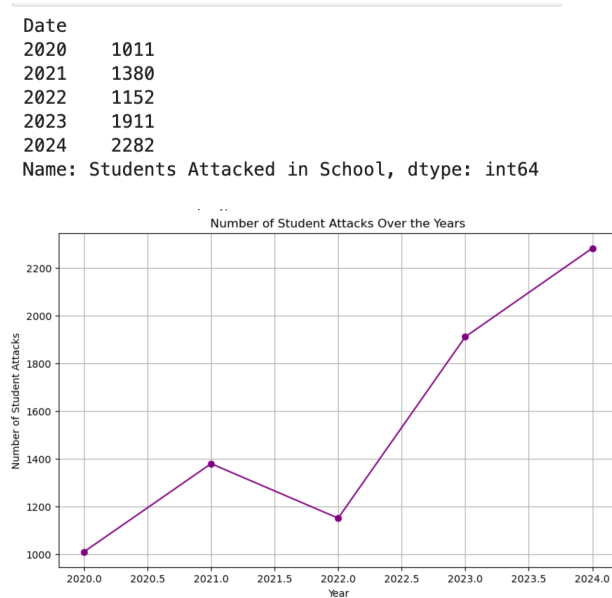
has completed, and the dataset has been cleaned and saved for R, Python and SQL further analyses.

B. Analyzing number of student attacks over years

Using Python Programming Language, I will perform different univariate analyses of 'Student Attacked in School' column, which is the ratio data type. First, I will sum the total number of students attacked and grouped them by different years 2020, 2021, 2022, 2023 and 2024. The year of 2020 had the smallest number of students got attacked in school, which is 1011 number of students. The three following years 2021, 2022 and 2023 had 1380, 1152 and 1911 total students attacked in school respectively. In 2024, the total number of students attacked in school had increased to 2282 students. The total number of students got attacks by wars and conflicts have been increased significantly over the years.

Figure 3 showed the total number of students attacked from year 2020 to 2024 and the trend of these incidents throughout the year.

Figure 3

```
Date
2020    1011
2021    1380
2022    1152
2023    1911
2024    2282
Name: Students Attacked in School, dtype: int64
```


Number of Student Attacks Over the Years

Continuing with Python, I will display the summary statistic of 'Student Attacked in School' data. The data has the count of 3471, mean of 2.229, standard deviation of 13.206 max of 347. Then, I want to explore its further by finding the country that has the highest number of students attacks with its latitude, longitude and date of the event. As the result, Nigeria was the country with a highest number of students attacked with Latitude of 11.9313 and longitude of 7.4138 on the date of 12-11-2020.

According to Amnesty International, there were over 500 boarding students have been kidnapped by perpetrators from Government Science Secondary School in the town of Kanara, in Katsina state, Northwest Nigeria in the night of December 11, 2020 [4]. Until December 13, Nigerian authorities still declared that there were still missing 333 students [4].

Figure 4

```
count     3471.000000
mean         2.228753
std         13.206172
min          0.000000
25%          0.000000
50%          0.000000
75%          0.000000
max        347.000000
Name: Students Attacked in School, dtype: float64
```

```
          Date   Country  Admin 1  Latitude  Longitude  \
3045  2020-12-11  Nigeria  Katsina   11.9313     7.4138

         Students Attacked in School
3045                             347
```
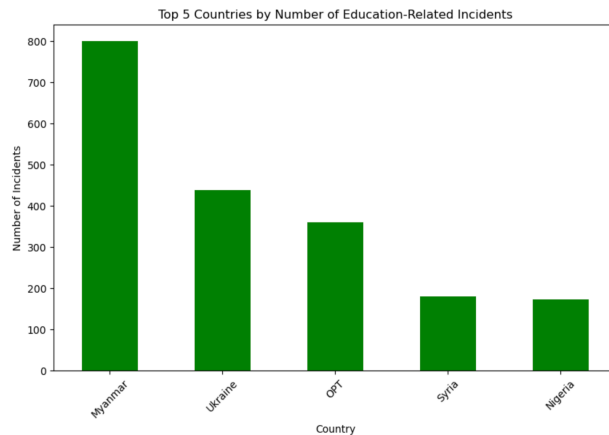
C. Analyze education in danger incidents based on countries and geographic location

Using Python, I try to display five countries that suffered the highest number of schools and universities' attacks from 2020 to 2024, which created education barriers for students and children. These countries are Myanmar, Ukraine, Occupied Palestinian Territory (OPT), Syria and Nigeria. Myanmar had 800 total of attacks from 2020 to 2024, which is the country has the highest number of education in danger incidents. After Myanmar, Ukraine, OPT and Syria had 437, 360 and 179 total number of education-related incidents respectively from 2020 to 2024. Nigeria had 172 total of education related attacks, which is the lowest in top five countries.

Figure 5 showed the frequency of incidents based on five countires and its visualization.

Figure 5

```
Top 5 countries with the highest number of education-related incidents:
Country
Myanmar    800
Ukraine    437
OPT        360
Syria      179
Nigeria    172
Name: count, dtype: int64
```

Top 5 Countries by Number of Education-Related Incidents

```
# Mapping ordinal geo precsion levels to a range of number in order of location exactness
geo_precision_mapping <- c(
  "censored" = 7,
  "censored by HDX" = 7,  # Handle both types of censored entries
  "(2) 25 km Precision" = 2,
  "(3) District, Communicipality or Commune" = 3,
  "(4) Province, State, Governorate" = 4,
  "(5) Linear Feature e.g. Along Road/River" = 5,
  "(6) Country" = 6
)
> print(geo_precision_analysis)
```

	Geo_Precision_Level	Frequency	Percentage
1	2	2047	58.97435897
2	3	904	26.04436762
3	4	122	3.51483722
4	5	3	0.08643042
5	6	26	0.74906367
6	7	369	10.63094209

Distribution of Geo Precision Levels

Then , I start using R programming language to examine the Geo Precision ordinal data. The Geo Precison is the level of accuracy of the location that the events occurred , which ranges from highly precise to highly general. For this dataset, "(2) 25 km Precison" is the highest level of location exactness, followed by "(3) District, Communicipality or Commune", "(4) Province, State, Governorate" and "(5) Linear Feature e.g. Along Road/River" and "(6) Country" and the last one is "Censored" or Censored by HDX" means the location has not been provided. I mapped the ordinal data into the range of 2 to 7 in which 2 is the highest level of location exactness, and 7 is the lowest level of location exactness. Then, I cleaned the data to remove any leading and trailing spaces and standardlized "censored by HDX" to "censored" to remove data redundancy. I converted to numeric data and tried to display frequency and percentage of each Geo Precison. Geo Precision level 2 has the highest frequency 2047 incidents and a percentage 58.97, followed by Geo Precsion level 3 with 904 frequency and 26.045 percentage. Geo level 5 has the lowest frequency of 3 and the percentage of 0.0864.
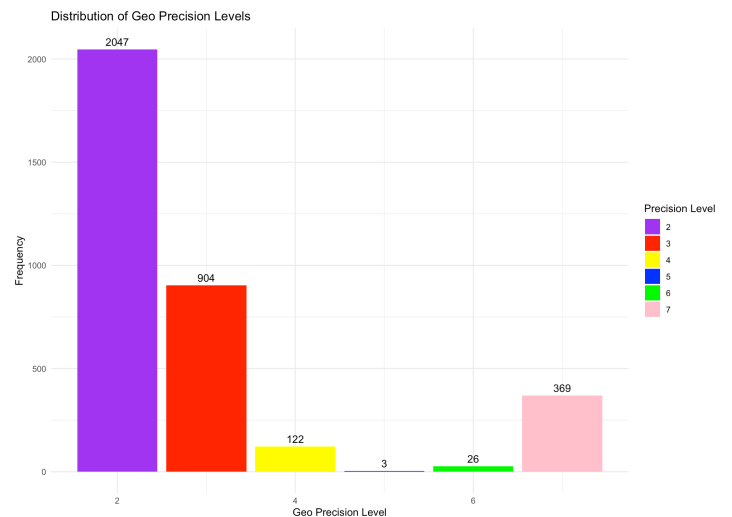
In general, this dataset has a high level of location accuracy for all events that happened with the high frequency for Geo Precison level 2 and 3. Figure 6 showed a plot of the Frequency of different Geo Precison Levels with various colors.

Figure 6

Then, I will use R programming language to apply clustering for the dataset. I chose eight numerical colums, which include 'Latitude', 'Longtitude',' Students Attacked in School', 'Student Killed', 'Student Injured', 'Students Kidnapped', 'Student Arrested' and 'Sexual Violence Affecting School Age Children' to do the clustering. I do some basic data cleaning such as replace 0 in column Latitude and Longtitude with their respective colum mean, so the missing data in Latitude and Longtitude will not affect clustering. I standarlized the data, so all data variables can contribute equally to the clustering.

I divided eight colums into four clusters and tried to visualize them geographically using Latitude and Longtitude. The data points that grouped in the same cluster has similar characteristics. The first picture in Firgure 7 displayed the Count, Average and Max of numerical colums in each cluster 1,2,3,4. Specifically, Cluster 1 has the count of 1099, average latitude of 7.40, average longitude 1.44, the maximum of 53 students attacked, 15 killed, 53 students injured, 106 students kidnapped and 300 arrested and maximum of two sexual violence incidents. Cluster 2 has the count of 1478, average latitude of 19.7, average longitude of 76.8, the maximum of 131 students

attacked, 13 killed, 40 injured, 40 kidnapped and 173 arrested and maximum of 17 sexual violence incidents. Cluster 3 has the count of the count of 878, average latitude of 42.4 and longtitude of 39.4, the maximum of 69 students attacked, 11 killed, 40 injured, 4 kidnapped, 40 arrested and maximum of two sexual violence. Cluster 4 had 16 counts, 18.2 and 39.7 average latitude and longtitude, maximum of 347 students attacked, 51 killed, 344 injured, 346 kidnapped, 0 arrested and 0 sexual violence.

Cluster 2 has the highest number of data points, and Cluster 4 has the smallest number of data points. However, data points in Cluster 4 have the highest number of students got attacked, killed, injured and kidnapped. Also, Cluster 2 has the highest number sexual violence incidents occurred, 17, and Cluster 1 has the maximum number of students got arrested, 300. In the visualization, data points in cluster 1 with the color of orange have distributed evenly across different geographic locations. The green-colored data points representing Cluster 2 appeared mostly within the latitude range of 10 to 30 and around a longitude of 100. Cluster 3 in the color of blue located in latitude of 30 to 55 and around longitude of 50. Cluster 4 in the color of purple occupied small areas in the dataset around latitude 10.

Figure 7 showed the summary statistic of eight numeric colums using for clustering; Count, Average of Latitude and Longtitude and Max of students related incidents of colums in each cluster 1,2,3,4; and the geographic clustering visualization.

Figure 7

```
> print(summary_by_cluster)
# A tibble: 4 × 10
  Cluster Count Avg_Latitude Avg_Longitude Max_Students_Attacked Max_Students_Killed
  <fct>   <int>        <dbl>         <dbl>                 <dbl>               <dbl>
1 1        1099         7.40          1.44                    53                  15
2 2        1478        19.7          76.8                    131                  13
3 3         878        42.4          39.4                     69                  11
4 4          16        18.2          39.7                    347                  51
> summary(numerical_data_with_clusters)

 Max_Students_Injured Max_Students_Kidnapped Max_Students_Arrested Max_Sexual_Violence
                <dbl>                  <dbl>                 <dbl>               <dbl>
                   53                    106                   300                   2
                   40                     40                   163                  17
                   40                      4                    40                   2
                  344                    346                     0                   0

> summary(numerical_data_with_clusters)
     Latitude         Longitude       Students Attacked in School  Students Killed   Students Injured
 Min.   :-38.63   Min.   :-120.80   Min.   :  0.000             Min.   : 0.0000   Min.   :  0.000
 1st Qu.: 13.07   1st Qu.:  28.86   1st Qu.:  0.000             1st Qu.: 0.0000   1st Qu.:  0.000
 Median : 19.48   Median :  39.13   Median :  0.000             Median : 0.0000   Median :  0.000
 Mean   : 21.55   Mean   :  43.29   Mean   :  2.229             Mean   : 0.1608   Mean   :  0.677
 3rd Qu.: 32.19   3rd Qu.:  91.92   3rd Qu.:  0.000             3rd Qu.: 0.0000   3rd Qu.:  0.000
 Max.   : 56.87   Max.   : 144.15   Max.   :347.000             Max.   :51.0000   Max.   :344.000


 Students Kidnapped Students Arrested  Sexual Violence Affecting School Age Children Cluster
 Min.   :  0.0000   Min.   :  0.0000   Min.   : 0.000000                            1:1099
 1st Qu.:  0.0000   1st Qu.:  0.0000   1st Qu.: 0.000000                            2:1478
 Median :  0.0000   Median :  0.0000   Median : 0.000000                            3: 878
 Mean   :  0.5782   Mean   :  0.2858   Mean   : 0.009795                            4:  16
 3rd Qu.:  0.0000   3rd Qu.:  0.0000   3rd Qu.: 0.000000
 Max.   :346.0000   Max.   :300.0000   Max.   :17.000000
```


Geographical Distribution of Clusters
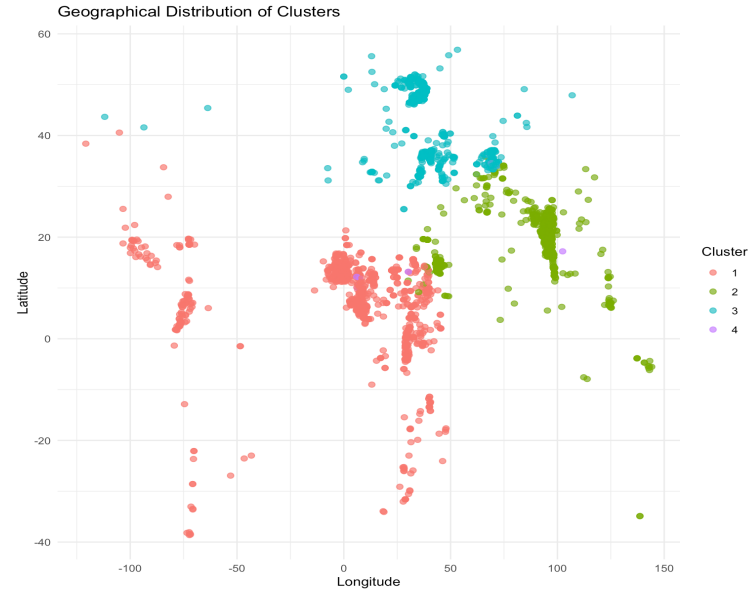
Then, I started defining schema in SQL and use it to display some queries for top five countries have highest number of students injured and killed in these schools' attacks over years or which country had the highest sexual violence events occurred in these attacked.

The top five countries had the highest number of students killed over the years were Afghanistan with 114 students killed, Somalia with 66 students killed, Occupied Palestine Territory with 53 students killed, Uganda with 51 students killed, and Myanmar with 40 students killed. Also, the five countries with the highest number of students injured over the years were Nigeria with 690 students, Afghanistan with 222 students, Myanmar with 202 students, Sudan with 191 students and Pakistan with 174 students.

Based on the results in Figure 8, Afghanistan was an outlier among the top five countries with the highest total students killed, which is around two times higher than Somalia and the remaining countries. Also, Nigeria appeared as an outlier among the top five countries with the highest students injured because the Nigeria's number of students injured was three time higher than Afghanistan and the remaining countries. Thus, further analyses should be made for Afghanistan and Nigeria. Myanmar also needs further analysis, since it appeared both in Figure 8 as top five countries for total students injured and killed and Figure 5 as top country had highest education in danger incidents over time.

Figure 8

	country character varying (100) 🔒	total_students_killed 🔒 bigint
1	Afghanistan	114
2	Somalia	66
3	OPT	53
4	Uganda	51
5	Myanmar	40

	country character varying (100) 🔒	total_students_injured 🔒 bigint
1	Nigeria	690
2	Afghanistan	222
3	Myanmar	202
4	Sudan	191
5	Pakistan	174

Then, I will analyze countries that have the highest sexual violence incidents and display the exact date and countries that the sexual violence incidents happened. I try to display both top five and top ten countries with the highest sexual violence events to compare the difference. So, the country named, Papua New Guinea (PNG) had the highest number of sexual violence incidents related to schools' attacks, followed by Occupied Palestine Territory (OPT) with 4 incidents and Rwanda and Greece with 2 sexual violence incidents. The remaining countries in top 10 such as UK, Kenya and Pakistan had a record of 1 incident of sexual violence.

Figure 9

	date date 🔒	country character varying (100) 🔒	total_sexual_violence 🔒 bigint
1	2023-06-07	PNG	17
2	2023-12-07	OPT	4
3	2020-02-22	Greece	2
4	2020-04-03	Rwanda	2
5	2022-10-05	UK	1
6	2024-03-30	Bangladesh	1
7	2024-06-01	Pakistan	1
8	2022-09-09	UK	1
9	2023-06-29	Kenya	1
10	2021-04-18	Burundi	1

	date date 🔒	country character varying (100) 🔒	total_sexual_violence 🔒 bigint
1	2023-06-07	PNG	17
2	2023-12-07	OPT	4
3	2020-04-03	Rwanda	2
4	2020-02-22	Greece	2
5	2024-06-01	Pakistan	1

D. Analyze Education In Danger Incidents By Specific Date and Month

Date is an interval data type in this dataset, which will analyze further in this section using SQL. I analyzed the frequency of date in this dataset and displayed top ten date that have highest incidents. In 2021- 03-07, there were 14 total education incidents happened, which was the date with the most incidents. Followed by 2024-05-17 with 13 incidents across different locations, 2024-01-08 and 2020-02-24 with 12 incidents. There were five dates with 11 incidents based on the result. Then, I tried to extract month from Date data type and calculate the total incidents per month. January recorded the highest number of education-in-danger incidents, totaling 386, followed by May and June with 378 each. February reported 339 incidents, March 323, September 267, October 261, December 247, November 240, August 235, and April 213, July 204 over the years of 2020 to 2024. July was the month with the lowest number of incidents over years.

Also, I able to use Python to visualize the Number of Education Incidents per Month over the years of 2020 to 2024. Figure 10 contained the results analyzing in both SQL and Python.
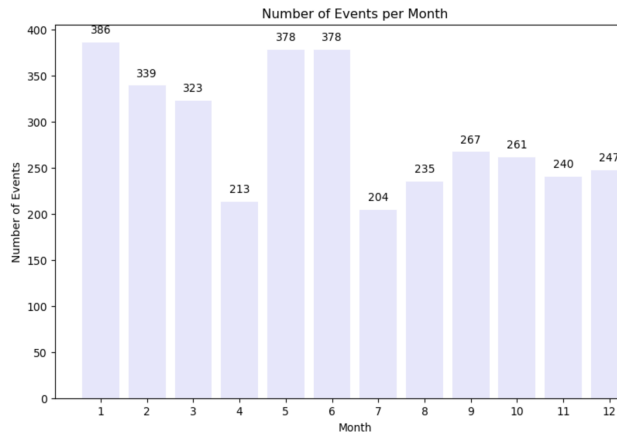
Figure 10

	date date 🔒	total_incidents 🔒 bigint
1	2021-03-07	14
2	2024-05-17	13
3	2024-01-08	12
4	2020-02-25	12
5	2024-04-22	11
6	2023-12-13	11
7	2022-03-01	11
8	2021-05-27	11
9	2024-07-03	11
10	2021-05-17	10

	month numeric 🔒	total_incidents 🔒 bigint
1	1	386
2	5	378
3	6	378
4	2	339
5	3	323
6	9	267
7	10	261
8	12	247
9	11	240
10	8	235
11	4	213
12	7	204

Number of Events per Month

E. Analyze Different Event Descriptions Using NLP

I analyzed 'Event Description' column using Python to perform the Setiment Analysis. I applied the polarity scores to the dataset in order to classify them as Negative, Neutral or Positve for further analysis. The summary of sentiment analysis showed that the Event Description have 2826 negative responses, 531 neutral reponses and 114 positve reponses. Beside the sentiment analysis summary, Firgure 11 also showed the sentiment scores of event description for few rows in the dataset and the visualization of the sentiment analysis.

The purple chart was the highest chart in the plot with 2826 negative description, followed by the Neutral green chart, and Pink chart was the smallest chart with 114 positive descriptions. In general, most of these event descriptions were negative, which indicated the negative experiences for all students and educators suffered in the incidents.

Figure 11

```
Event Description and Sentiment Score:
                                   Event Description  Sentiment_Score
0      Reported in September 2024: An unspecified num...         0.0772
1      August 2024: A teacher coordinator of an indig...        -0.8860
2      28 August 2024: Nine Palestinians, including t...        -0.7845
3      August 2024: A soldier was killed and three pe...        -0.8779
4      August 2024: A high school was damaged in an a...        -0.7184
...                                              ...              ...
3466   January 2020: A school was damaged by missiles...        -0.7579
3467   January 2020: A university student and CNL sup...        -0.5267
3468   January 2020: A school was damaged and childre...        -0.6808
3469   January 2020: A school director was assassinat...         0.0000
3470   January 2020: A teacher and six students were ...        -0.6705

[3471 rows x 2 columns]
```

```
Sentiment Analysis Summary:
Sentiment_Category
Negative    2826
Neutral      531
Positive     114
Name: count, dtype: int64
```


Sentiment Analysis Summary
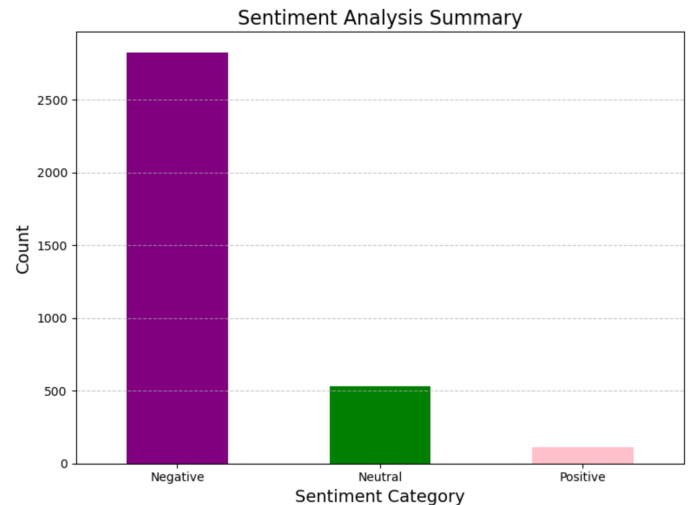
v. CONCLUSION

The purpose of this research paper is to analyze education in danger incidents over years, across different countries or geographic locations. Additionally, the research aims to categorize event descriptions of these incidents using Natural Language Processing (NLP). The number of students attacked have been increased significantly over the years, which has analyzed in Figure 3. Using eight numerical colums, the dataset had been divided in four different clusters 1, 2,3,4 and used latitude and longtitude to visualize geographically in Figure 7. Cluster 2 has the highest number of data points, and Cluster 4 have the highest number of students attacked, killed and kidnapped. The top five countries had the highest schools' attacks, highest number of students got injured, killed over years and total sexual violence cases have been analyzed in Figure 5,8 and 9. Myanmar had the highest schools' attacks over years, Afghanistan had the highest number of student killed over years and Nigeria had the highest number of students injured over years. Also, January was the month with the highest incidents occurred thoughout the years, followed by May and June, which have been analyzed in Figure 10. Using NLP, the sentiment analysis had been performed for Event Description, which resulted in a high level of negative descriptions of the events.

Further analyses are needed to create an action plan aim to reduce number of students got attacked in these education in danger incidents. Further analyses needed for geographic locations in each cluster to understand their characteristics and why these locations grouped in the same clusters. These three countries Myanmar, Afghanistan and Nigeria will need further in-depth analysis to figure out why these numbers of students killed, injured and

attacked were high. Addition qualitative analysis can be performed to examine what happened in the months with high education in danger incidents such as January, May and June.

REFERENCES

[1] H. K. Lin, "Educational challenges and expectancy of non-governmental education leaders in armed conflict areas in Myanmar," International Journal of Educational Management and Development Studies, vol. 4, no. 4, pp. 61–80, 2023. [Online]. Available: https://doi.org/10.53378/353022.

[2] P. Mutu, "Leveraging technology to deliver basic education to children in conflict areas of Northern Nigeria," *African Human Rights Law Journal*, vol. 23, no. 1, pp. 182–204, Jul. 2023. Available: https://doi.org/10.17159/1996-2096/2023/v23n1a8.

[3] B. Akar, "Citizenship education in conflict-affected areas and nation-states: empowering teachers for sustainable reform," Intercultural Education, vol. 31, no. 5, pp. 519–532, 2020. [Online]. Available: https://doi.org/10.1080/14675986.2020.1794123.

[4] Amnesty International, "Nigeria: Education under attack in the north," *Amnesty International*, Aug. 08, 2022. https://www.amnesty.org/en/latest/press-release/2020/12/nigeria-education-under-attack-in-the-north/

[5] Humanitarian Data Exchange, "Education in Danger Monthly News Briefs Dataset," Humanitarian Data Exchange, [Online]. Available: https://data.humdata.org/dataset/sind-education-in-danger-monthly-news-briefs-dataset.