

# Global Terrorism Database: Analysis and Visualization

Sara Khan<sup>1</sup>, Swati Singh<sup>2</sup>, Shraddha Shukla<sup>3</sup>, Utkarsh Dubey<sup>4</sup>,  
Dr. Vibha Oberoi<sup>5</sup>

Greater Noida Institute of Technology

Submitted: 15-04-2022

Revised: 27-04-2022

Accepted: 30-04-2022

**ABSTRACT:** The goal of the project is to interactively visualize the past thirty five years of terrorist attack across the globe using the data in the Global Terrorism Database (GTD). Also, to examine patterns, trends, and distribution of various attack types. The database includes 141,967 terrorist attack records the world, with 137 fields to capture information about, location, attack types, timing, used weapons, casualties, damaged properties, and targets. The dashboard interactively visualizes the concentration of terrorist attacks by region, country, state, city and bar graphs and line graphs to show attacks by year, most common attack types, number of casualties by attack type, and number of attacks by different terrorist group.

**Keywords:** Terrorist Attack, Global Terrorism Database (GTD), Casualties, Terrorism Group, Machine Learning

## I. INTRODUCTION

There has been ever-increasing number and of terrorist attacks worldwide Over the past thirty five years. The Global Terrorism Database (GTD), at <http://www.start.umd.edu/gtd/>, provides useful data on the terrorist attacks from 1970-2017 with characteristics of such terror acts, including the specific tactics, weapons, target types, or fatalities inflicted, also change across time and space [1]. Based on the location of the events, people at large have very limited information about any such event happening in other parts of the world and hence react differently. It is a challenging task to explore and understand the patterns and trends in the dataset. The data may

also be summarized by target types to understand the characteristics of incidents. Artificial Intelligence is used to visualize and predict these types of events.

## II. DATASET DESCRIPTION

The Global Terrorism Database is an open-source database taken from National Consortium for the Study of Terrorism and Responses to Terrorism (START) located at the url <http://www.start.umd.edu/gtd/>. The dataset is comprehensive and contains a lot of terrorism-related information. The dataset includes 1, 56,772 terrorist attacks around the world and 137 features and takes 142.3 MB of disk space. Out of 137 attributes, 23 columns containing both quantitative as well as qualitative information were taken into account thereby taking 26.8 MB of disk space. Some of the columns used are as mentioned:

- i. Date of event
- ii. Location of event
- iii. Type of attack
- iv. Type of weapon used
- v. Extent of property damage
- vi. Target type
- vii. Number of casualties
- viii. Group Name

The dashboard uses only a small subset of these fields, focusing on dates, locations, attack types, responsible party, casualties and extent of property damage etc.

Table1 shows column(s)having numeric values.

Column	Description
Year	Year in which attack occur
Month	Month in which attack occur
Day	Day on which attack occur
Killed	Total Number of Fatalities
Wounded	No of people injured

Latitude	Latitude(of city)
Longitude	Longitude(of city)

**Table 1**

**Table 2**

Column	Description
Region	Region where the incident occur
Country	Country where the incident occur
City	City, Town, or village
Weapon Type	Weapon type used in attack()
Attack Type	General method of attack
Target Type	General type of target/victim
Group Name	Name of the group carrying out the attack

Table 2 shows columns having categorical values.

### III. DATASET PRE-PROCESSING

Global Terrorism Database is inconsistent, incomplete, and contains many errors, missing attributes values. In data pre-processing methodology, the following steps were used in the: Data cleaning is a process of filling missing values, and handle inconsistencies in data. In GTD, numerous fields like motives or responsible organizations which are missing either due to information not available or that field was not relevant for that specific event was not used. Attributes like summary, claim\_mode, and fields like claimmode\_txt, guncertain, nperps etc. were removed since they are not relevant to the analysis of this project. Also, column values including weapsubtype2, and weaptype3\_tx' have more missing values than valid entries. In order to reduce complexity such fields were removed as well.

### IV. TASKS

The tasks to be addressed in visualization includes the following:

- i. During what time period did most terrorist attacks occur?
- ii. In which countries did most terrorist attacks occur?
- iii. What are the most common types of attacks?
- iv. What types of attacks result in the highest number of casualties?
- v. Which terrorist groups are responsible for the highest number of attacks?

### V. RESULTS

#### A. Web-Based Interactive Tool

Figure 1 shows the interactive visualisation dashboard used for customizing the visualizations according to the need of the user.

It presents an exploratory tool which plots all terrorist events mentioned in the dataset on the global map. User can apply multiple filters to get the corresponding statistics. This section also contains the progression graph of terrorism events since 1970. Most affected cities are listed along with the total number of attacks they have faced, based on the filters selected.

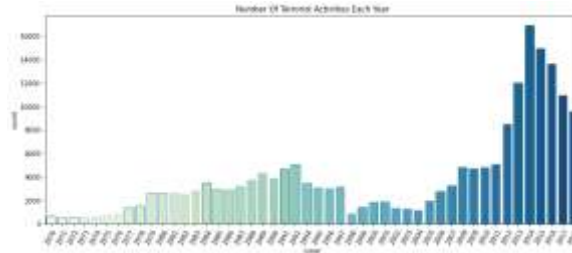


**Figure 1**

**B. Exploratory Analysis**

To better understand the Global Terrorism Dataset various visualizations have been carried out on various attributes that were listed in the Global Terrorism Dataset (GTD). Following are the

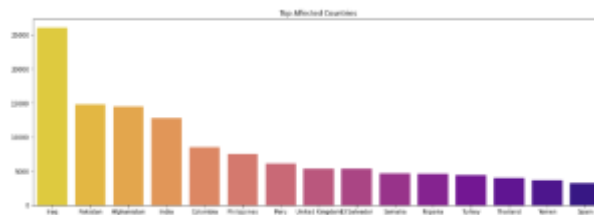
diagrams generated from the GTD after visualization of the contents based on attacks by year, countries by total attack, and attacks by type. The diagrams below have been plotted using plotly.



**Figure 2 Terrorist activities By Year**

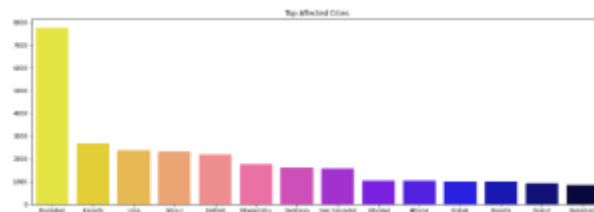
Figure 2 can be used for summarizing the number of terrorist attacks each years. It provide us an idea about how terrorism has evolved and what rate has it impacted the world each year. Terrorist

attacks were quite low in numbers in the decade of 1970 and there was a sudden increase by the year 2014.



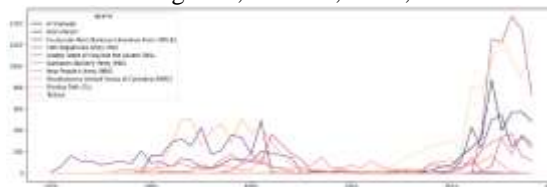
**Figure 3 Top Affected Countries**

Figure 3 shows the highly affected and attacked countries. The most affected countries includes Iraq, Pakistan, Afghanistan, and India based on the total number of attacks.



**Figure 4 Hot Zones(cities affected)**

Figure 4 shows the hot zones that is, the highly affected cities of the most affected countries by terrorist attacks. The hot zones include the cities of Baghdad, Karachi, Lima, Mosul.



**Figure 5 Terrorist Groups**

Figure 5 describes some of the most notorious terrorist groups and their active years. Groups like SL (Shining Path) and FMLN

(Farabundo Marti National Liberation Front) were highly proactive for almost over a decade from the 1980s to 1990s but after that no trace of those



Finding dependencies among the various parameters in the dataset can reveal a key pattern about the nature of terrorism. Out of 137 variables, we have selected sixteen for this map. Some of those parameters are the month, day, year, country, latitude, longitude, type of attack, target type, number of kills, etc. Forming a correlation map can provide us with one-to-one correspondence of each variable with rest.

Figure 8 shows the correlation matrix depicting shows correlation matrix where darker the shade of the block, more the attributes are correlated proportionately. Here we can see that country and latitude are correlated which is expected. Values of neither of those two parameters change and hence they show a strong relation.

## VI. CONCLUSION

Through this research, it is possible to conclude that the use of Machine Learning techniques, it is able to visualize and predict terrorist attacks. The results show that there has been a considerable growth in terrorist attacks since 2014. Regarding the types of attacks, there is still the probability that bombs and explosions are involved, followed by armed assault. The results have been successfully achieved by using the data collected from the GTD. The models that were made through Logistic Regression give results with an efficiency of 0.96. These results demonstrate that Exploratory Data Analysis and Machine Learning are ideal for contributing to research related to world events.

## REFERENCES

- [1]. "Information on More than 140,000 Terrorist Attacks." Global Terrorism Database. Web.
- [2]. Citraro, Dino. "Visualizing Terror: Our Work with the Global Terrorism Database," in Periscope.
- [3]. Lex, A.; Gehlenborg, N.; Strobel, H.; Vuillemot, R.; Pfister, H., "UpSet: Visualization of Intersecting Sets," in Visualization and Computer Graphics, IEEE Transactions on. vol.20, no.12, pp.1983-1992.
- [4]. Ortiz, Santiago. "3D Globe Visualizes the Flows of FoodAid Funding among Countries," in Information Aesthetics.
- [5]. "MetricsGraphics.js." MetricsGraphics.js. Web.
- [6]. "Share Graph." Gapminder. Web.