Universidad Tecnológica de Pereira
Università degli Studi di Salerno

Magistrale

Exploring the Historical Behavior of Islamic State Groups: Latent Class Analysis and K-Modes Clustering Approach

Presented by:
Daniel Varona Henao
Mateo Toquica Arango

Supervised by
PhD José Alfredo Jaramillo
PhD Roberto Tagliaferri

January 2021

Abstract

Research on terrorism has always gravitated around qualitative methods and statistical techniques. Technology plays an essential role in terrorism and counter-terrorism analysis by providing collections of large databases in many fields and the computational power to analyze them. Machine learning has shown new methods that could complement standard and well-established methodological approaches. This work contributes to the bridging of machine learning with terrorism studies by analyzing data with a classic statistical method Latent Class Analysis (LCA), and a machine learning method (K-modes). More formally, this work presents a mixed approach to analyze and cluster records from the Global Terrorism Database (GTD) referring to terrorist attacks belonging to the Islamic State. A diverse set of variables are considered, such as the type of weapons, targets, terrorist groups perpetrating the attacks, and geographic location. We identified three analysis periods by relying on a literature review and applied and contrasted LCA and K-Mode models for each period. This project aims to generate a record of how the periods were divided and identify the critical points for using the variables in the GTD database. Finally, we performed a data classification and generated an analysis for whoever requires it for these terrorist groups in the established periods.

Acknowledgments

We want to thank Engineer Jorge Ivan Patiño for his unconditional support during our master's degree in Italy. We also would like to express our sincere thanks to Dr. Roberto Tagliaferri for his help, hospitality, lessons, problem-solving, and accompaniment during the master's degree at the University of Salerno.

We also would like to thank Concetta Nicotra for the love and unconditional support during our stay in the university residences and for making this a warm and pleasant place through her tremendous vocation and approach to working with foreign and local people.

We thank in advance Dr. Jose Alfredo Jaramillo for his remarkable collaboration during the development of this document and for providing the necessary tools to achieve its completion.

Our gratitude also extends to our colleagues for their support in the courses and for generously sharing their knowledge. Also, we would like to express our gratefulness to all the people who were part of this thesis process for contributing their expertise and helping to complete this process.

Finally, we want to acknowledge we would not have been able to complete this quest without the support of our families, who served as pillars and emotional support for the development of our studies in Italy. Thanks to them for being part of this extraordinary experience. Their constant unconditional support and wise advice were crucial to achieving our goals.

Contents

Chapter 1

Introduction

1.1 Problem Description

Terrorism is the use of violent actions to gain social or political control over a given population. The most common target of terrorism is government institutions [1]. According to the Global Terrorism Database (GTD) [2], which has information from 1986 to 2018 on historical phenomena related to terrorist attacks, these events affect society in political, economic, and cultural aspects [2]. Terrorism has a solid connection to numerous social conflicts and is a notable actor in underdeveloped countries. Consequently, there is a need to identify, analyze and prevent potential terrorist attacks. Therefore, research such as *Ritchie et al.* [3] is essential to combat terrorism. That research on terrorism analyzes critical points such as its global distribution and the death rates registered between 2007 and 2017. It stands out that in 2007 there were 12,824 fatalities from terrorism; in 2017, there were 26,445 victims and the highest record in this decade occurred in 2014 when 44,490 people died from actions of this nature.

Likewise, it is worth mentioning that there are studies such as that of Alsaedi, Almobarak, and Alharbi [4], which have the objective of comparing data mining models to predict the success of a terrorist attack and the prediction of the identity of the organizations behind this attack. Our study obtained large volumes of information through data mining techniques to create carefully chosen personalized databases to study terrorist groups belonging to the Islamic State. Smith's study [5] was of great importance for defining the study periods and treating said databases through two statistical methods: Latent Class Analysis (LCA) and K-Modes. Implementation of statistical models like LCA allows knowing the optimal number of clusters within large amounts of behavioral data studies like human behaviors and social perceptions. On the other hand, the K-Modes method efficiently clusters large amounts of categorical data.

This research seeks to answer the following question: Is it possible to identify and compare the evolutionary behavior of terrorist groups based on historical data through two strategies, one being the statistical method latent class analysis and the other being a machine learning strategy of unsupervised models?

1.2 General and Specific Objectives

1.2.1 General Objective

- To cluster organizations and terrorist attacks according to the homogeneity in their behavior and evolution of terrorism related to the Islamic state in the last 30 years.

1.2.2 Specific Objectives

- To identify the periods of analysis that reflect the historical periods and the evolution of Islamic terrorism.
- To identify the behavior of event information within the GTD. for the Islamic State through an Exploratory Data Analysis.
- To implement K-Modes algorithms to identify homogeneous clusters of terrorist organizations.
- To implement Latent Class Analysis algorithms to identify homogeneous clusters of terrorist organizations.
- To compare the results obtained in the K-Modes and Latent Class Analysis.

1.3 Background and Justification

In human geography, terrorism has remained an act of violence searching for the destruction of representative elements or the exposition of messages selected and distributed based on globalization processes, perceived injustices, and other perspectives related to geopolitics, ecological politics, or historical facts *Mustafá et al.* [6].As Mustafá [6] explains, it is crucial to understand that there is no universal term for the definition of terrorism. The term terrorism became popular at the end of the 18th century with the French Revolution, but its apogee began when it expressed the fear applied by anarchy and revolutionary movements in the middle of the 20th century, by Fascist movements in Germany and Italy.

In addition, Sloan [7] explains that in the modern era, new communication systems and technologies have opened the doors to new methods of terrorism, which represent a greater risk of destabilization for society. For example, the use of the internet in some organizations known as the Islamic State has allowed the recruitment of young Muslim women through social networks.

However, new technologies for the use of terrorism also imply new ways to combat terrorism. Nie and Sun's study [8] shows the traceability of terrorist groups according to their actions. Due to the consolidation of information since 1986, global terrorist attacks data has been collected [2]and used in multiple projects and investigations helping the fight against terrorism and allowing the scientific community to contribute to this cause, such as Berkebile and Richard [9]. The study of *Santos et al.* [10] analyzes the excessive use of chemical weapons, finding that 30.5% were unknown gases, 23.3% were corrosive gases, 12.3% were tear gas, 11.6% unspecified gases, 8.2% cyanide, 5.5% pesticides, among others mentioned by the author in of 1970 and 2015. *Yun Woo-suk et al.* [11] carried out a study by exploring the attacks of the last 50 years, concluding that 73.3% of the attacks are of a terrorist type, affecting mainly the Middle East, South Asia, sub-Saharan Africa, and South America.

Hence the importance of applying data mining techniques, with their respective analysis, to define behavior patterns is evident [12]. There are enough events to track terrorist groups and generate traceability with which behavior patterns can be found through data analysis methodologies, information mining, grouping, among others [13], thus becoming an additional tool to combat terrorism.

1.4 Viability and scope

For the execution of the project, we took information belonging to Start. Start is an organization from the University of Maryland where researchers have carried out several studies concerning terrorism, violence, and counter-terrorism. The information that we will analyze within this study is that of GTD. [2]. This database consolidates various information resources such as news, existing data sets, and legal documents. Researchers translated the resources from 80 different languages to English. The variables are continuously updated to classify them under certain relevant groups, such as the type of weaponry used in the different events. The said database contains relevant information dating from 1986 to 2018 on terrorist events worldwide. This database is free to use and will allow the analysis of the behavior of terrorism around the world through statistical techniques. In this way, it offers the possibility of knowing the evolutionary behavior of terrorist groups by obtaining information on their mode of operation, similarities, and weapon preferences. Thus allowing to classify and know-how terrorist groups related to the Islamic state operate.

1.5 Methodology

1.5.1 Hypothesis

- It is possible to obtain equivalent information on clusters of terrorist events focused on periods under the implementation of two unsupervised models.
- It is possible to make clusters based on the variables of the GTD.
- It is possible to observe the cluster changes over time.

1.5.2 Methodological Design

The project's methodology approach studies the data registered in the GTD [2], dating from 1986. The above will allow us to know through a statistical analysis which groups have a similar behavior according to the year in which the event occurred, the country where the attack took place, the type of attack used, type of weapon, target, sub-target, and if it was carried out or not. This analysis allows to establish similarities of operation with the historical data of the Terrorist Groups belonging to Islam (TGBI) registered in this database; that is, the patterns will describe a behavior. Additionally, the exposed behaviors may show anomalies within the database [14], allowing decision-making based on computational algorithms such as the LCA and K-Modes models.

Next, we present the stages we have defined for the execution of the project.

1.5.2.1 Data Mining

An exploratory analysis of the database [12] is necessary for selecting the relevant parameters for the statistical analysis. Different data mining techniques will be implemented, including data cleaning, which aims to obtain reliable information [12]. This process consists of collecting the possible characteristics of the TGBIR from the database; then using data cleaning to arrive at a possible characterization. (Figure 1.1) shows the process of obtaining knowledge through data mining.

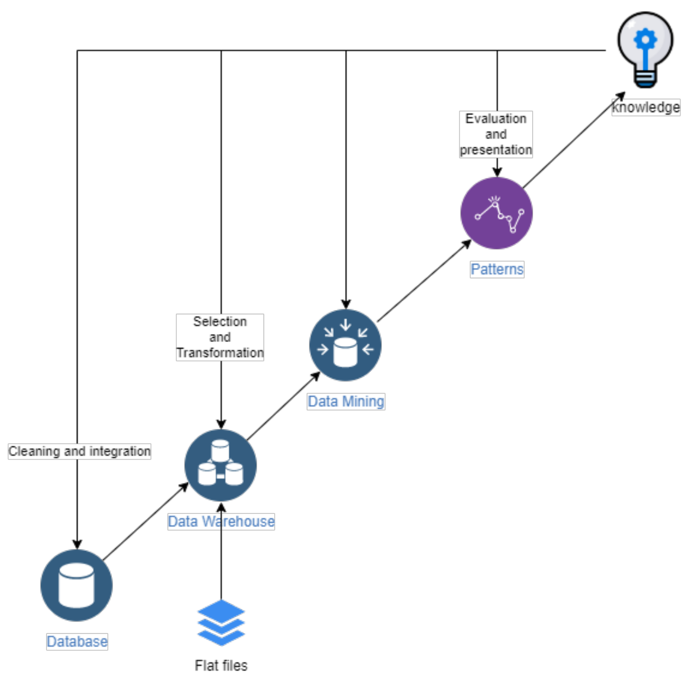

Figure 1.1: Data mining as a stage in the knowledge discovery process. Collected from: Data mining concepts and techniques third edition *Han et al.* [12]© Data mining concepts and techniques third edition 2011 Elsevier

1.5.2.2 Identify Criteria

We will perform a statistical analysis based on crosstables of the results obtained by the LCA and K-Modes models that will provide knowledge about the most relevant historical characteristics of the behavior of terrorist groups belonging to the GTD. The expected results of the analysis will correspond to information related to the years in which the attacks have concentrated, the frequencies with which they occur, the types of attacks most perpetrated, the types of weapons preferred by these groups, what target type classification they are and, finally, the countries where they have been perpetrated, among other criteria.

1.5.2.3 Period Division

Some experts treat the events of September 11 as a turning point for terrorist organizations [15, 16]. However, some researchers suggest a more comprehensive range of essential points to consider with terrorist attacks related to the Islamic State [11, 5], which imposes a discussion point, since based on the information of *Smith et al.*, the segmentation of the problem in different periods. They allow the project to synchronize with a grouping model according to the behavior patterns of each determined period.

1.5.3 Optimal Clusters

Clustering is a data mining technique that differentiates a group A from a group B. Therefore, grouping is the method selected to execute as a research methodology, proposing different types of grouping such as the Latent Class Analysis (LCA) and K-Modes. Researchers use both methodologies to treat categorical data. Consequently, this methodology allows the analysis and results of such grouping. We created nine pairs of models from 2-10 classes or ”clusters” to develop optimal clusters. The purpose of this activity is initially based on verifying the consistency of the pairwise models in terms of results and, consequently, choosing the model with the lowest indicator between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) as the optimal model according to [17].

1.5.4 Implemented Models

Unsupervised algorithms or unsupervised learning consist of learning without a supervisor or a margin of error for each observation. It is generally characterized by its large dimensions for the input values and complexity in terms of properties of interest [18]. These algorithms can identify hidden patterns within the information analyzed. As mentioned above, they have a design that consists of performing such internal learning without a supervisor to evaluate possible responses, according to the study of Sathya and Abraham[19]. The LCA and K-Modes grouping models are in the branch of unsupervised models, responsible for similar grouping information into different sets according to the number of groups and their closeness according to the characteristics they share. As Sathya mentions [19], clustering algorithms are an essential data mining technique where these algorithms receive information randomly, based on characteristics and correlations between the variables that want to be grouped, thus having a lower performance compared to supervised algorithms.

1.5.4.1 Latent Class Analysis

In LCA, we apply statistical methods to analyze human behavior, health sciences, social perceptions, among others [20]. This tool aims to discover variables that underlie the information under scrutiny. This model allows one or many categorical variables to be studied, taking a set of variables as a starting point. Latent classes are defined by conditional independence because the variables are statistically independent of the others within each of these classes. The variables are statistically independent of the others. Denson's study [21] LCA is a clustering approach based on models; its results are not sample-dependent and can be replicated in other samples. Within this model, group membership is unknown because it is assumed that membership in each group is not observed; in other words, it is latent. This technique is usually employed to discover behavioral profiles starting from a large number of data. As Graf mentions [22] , the classification of individuals in a given number of classes is based on similar patterns through the distribution of the study variables. This technique designates a membership probability, representing the value of belonging to each of the classes in the study, which is the likelihood of having certain types of similar characteristics compared to other individuals.

Taking into account the study of *Linzer et al.* [17] which explains how the model works. How the probability that an individual i gives a particular group of answers $\mathbf{J}$ of the manifest variables, assuming local independence, is given by the following formula:

$$(f(Y_i;\pi_r) = \prod_{j=1}^{j}\prod_{k=1}^{K_j}(\pi_{jrk})^{Y_{ijk}} \tag{1.1}$$

- **J** is equal to all the categorical polytomous variables of the study. The polytomous variable is understood to be a variable that has more than two response options.
- K_j they are all possible outcomes.
- **i** are all individuals belonging to the study ranging from 1...N
- (π_{jrk}) is the conditional probability of a class that an element within the study in the class r = 1 ... R will obtain a kth result in the variable jth.
- Y_{ijk} are the observed values of the manifest variables **J**

Suppose we take the job to go through the prior probability in all the classes estimated in the model, the **r** classes with the a priori probability. In that case, we obtain the given probability function:

$$P(Y_i|\pi, p) = \sum_{r=1}^{R} Pr \prod_{j=1}^{J}\prod_{k=1}^{K_j}(\pi_{jrk})^{Y_{ijk}} \tag{1.2}$$

The estimated parameters in this model are p_r and π_{jrk}.

The probability that each individual belongs to each of the classes given the estimates $\widehat{p}_r$,$\widehat{\pi}_{jrk}$ of p_r and π_{jrk}, conditioned by the observed values in the model variables, can be calculated using the Bayes formula:

$$\widehat{P}(r_i|Y_i) = \frac{\widehat{p}_r f(Y_i;\widehat{\pi}_r)}{\sum_{q=1}^{R}\widehat{p}_q f(Y_i;\widehat{\pi}_q)} \tag{1.3}$$

- π_{jrk} are estimates of probabilities of the results conditional on the class **r**.
- The number of independent parameters estimated by the model increases rapidly with *R*, *J* and K_j. Given these values, the number of parameters $R\sum_{r=j}(K_j-1)+(R-1)$ exceeds the total number of observations, or on the contrary one, less than the total number of cells in the cross-classification table of the study variables, the model will not be identified.

1.5.4.2 K-Modes

According to Huang [23], K-Modes is a very efficient way of grouping data by categories. In addition, Huang indicates the problems large amounts of information face based on the K-Means theory to deploy this methodology. Thus, thanks to K-Modes, it is possible to carry out an effective and efficient grouping, a determining factor in the correct development of Data Mining. According to Cao, the natural world describes the information variables through the basic concepts of set theory, using different definitions to classify the information [24]. In the first definition, an information system is a quadruple of data given by the form:

$$IS = (U, A, V, f)$$

U, the set of non-empty objects, called the universe;
A, the non-empty attribute set;
V, the union of all attribute domains that are:
V is the domain of the attribute a and is finite and disordered;
f: U x A $\rightarrow$ V, a mapping called information function such that for any:

$$V = \cup_{a \in A} V_a$$

V is the domain of the attribute a and is finite and disordered;
f: U x A $\rightarrow$ V, a mapping called information function such that for any:

$$a \in U \text{ and } a \in A, f(x, a) \in V_a$$

This definition is adopted in the library belonging to the Python programming language [25] and implemented by de Vos[26] based on the concepts of Huang [27, 28, 23] and Cao [24, 29]. We implemented Cao's results to develop this project as Huang focuses mainly on mixed problems of categorical and non-categorical data [27]. However, Huang's adaptation for categorical values is only an implementation similar to Cao's pseudo-code. Nonetheless, the difference entered is in step 2 where the center is calculated, as shown below:

1 Initialize the variable oldmodes as a k jPj-ary empty array;
2 Randomly choose k distinct objects x1, x2, .. . , xk from U
3 and assign [x1, x2, .. . , xk] to the k jPj-ary array variable newmodes;
4 for l = 1 to k
5 for j = 1 to jPj
6 calculate the similarity Simaj oxl ; xl p according to Definition 2;
7 end;
8 end;
9 while oldmodes¡ ¿newmodes do
10 oldmodes = newmodes;
11 for i = 1 to jUj
12 for l = 1 to k

13 calculate the dissimilarity between the ith object and 14 the lth mode according to Definition 5, and classify the ith 15 object into the cluster whose mode is closest to it; 16 end;
17 end;
18 for l = 1 to k
19 find the mode zl of each cluster and assign to newmodes;
20 for j = 1 to jPj
21 calculate the similarity Simaj ozl ; zl p according to Definition 2; 22 calculate maj of Definition 5; 23 end;
24 end;
25 if oldmodes==newmodes
26 break;
27 end;
28 end.

1.5.4.3 Model Comparison

The last step of the methodology consists of generating a comparative report on the behavior and evolution of terrorist groups through an approach focused on statistical analysis through the LCA and K-Modes techniques, providing a comparison of the results obtained by both. We obtained statistical methods to analyze the similar behaviors of the groups.

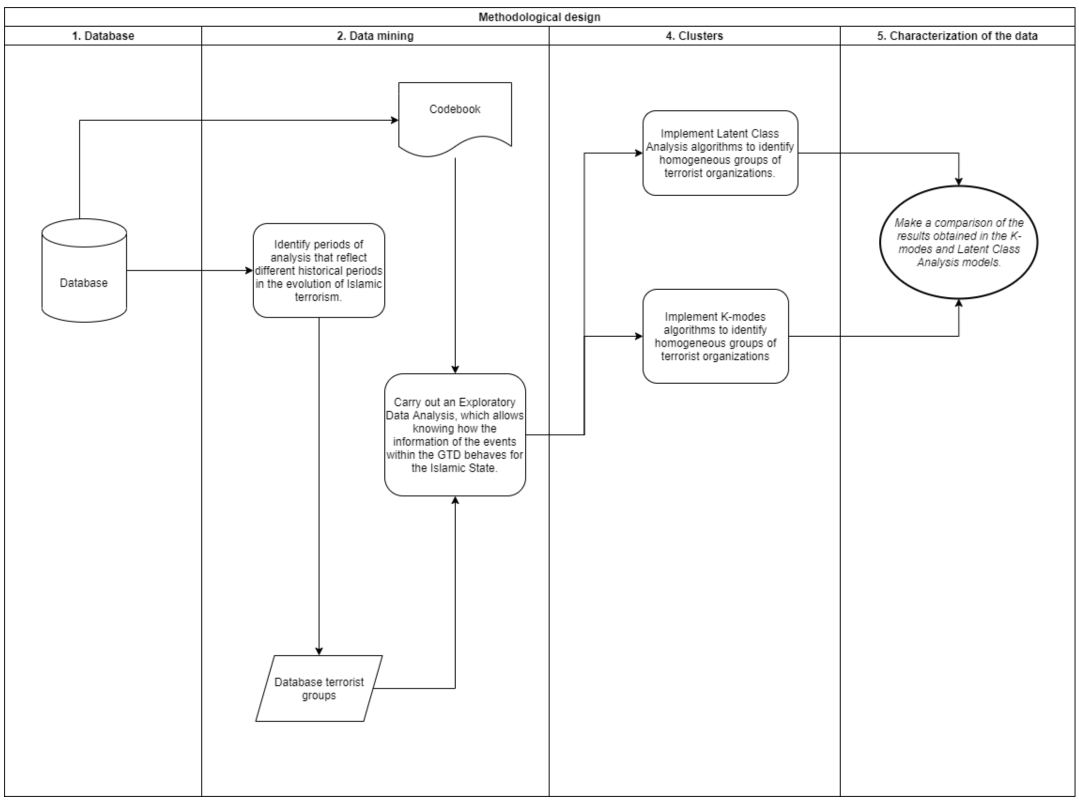

Figure 1.2: Methodology that evidences the steps taken to arrive at the report

1.6 Project sustainability

The project's theme has been studied for several decades and has shown that the effects of terrorism involve not only the victims and injured. In addition, Barker [30] explains that terrorism is something that not only concerns the leaders of the different nations of the world but it also affects the patience of people who have been victims or have been indirectly affected. It is a global issue that addresses different perspectives and impacts depending on the type of terrorism.

On the other hand, the evolution of technology plays a fundamental role within this project because it is an issue that influences the manufacture of weapons, the communication of terrorist organizations with society, and the methods of financing [31], being variables that play an essential role in identifying possible future attacks.

Finally, it is essential to highlight that for the Islamic State, terrorism has left war costs of 88,200 million dollars associated with the reconstruction of the country [32].

In summary, this research seeks to provide knowledge about terrorist actions in the Islamic state using information collected through the GTD database [2] to contribute to the fight against terrorism and the application of data science.

1.7 Administrative Aspects

1.7.1 Necessary Resources: Physical, Logistic, and Human.

- **Physical resources:** computers or virtual tools dedicated to data processing.
- **Logical resources:** Python 3, R-Studio, and dedicated libraries for data treatment and analytics, access to documentation for their implementation.
- **Human Resources:** Experts in Data Analytics, Analytical Models.

1.7.2 Sources of Funding

Universidad Tecnológica de Pereira and Università Degli Studi di Salerno will be providing the necessary equipment, software and human resources in the form of advisors when required.

Table 1.1: Project Costs

	Type	Description	Cost
Physical Resources	Variable	PC renting	$900.000
		Internet	$3.000.000
Logical Resources	Fixed	Global Terrorism Database GTD	$0
		Python3, R-Studio	$0
		Pandas, Numpy, Matplotlib, poLCA,K-Modes	$0
Human Resources	Variable	Development of Advice through an Expert in Data Analytics / Statistics	$0
Unexpected Costs	Variable	10% of total costs	390.000
Total			$4.290.000

- Note: source, self elaboration.

All costs $, are presented in Colombian pesos.

The source of information considered comes from direct collaboration with the Technological University of Pereira, which is a provider of economic resources and educational personnel for support, doubt clarification, and to direct the investigation towards a sharp analysis.

1.7.3 Schedule

Table 1.2: Project Schedule

TASK	MON 1	MON 2	MON 3	MON 4	MON 5	MON 6	MON 7	MON 8	MON 9	MON 10
Research state of the art methods										
Get database										
Preliminary analysis of the data										
Identify periods of analysis that reflect different historical periods in the evolution of Islamic terrorism.										
Implement K-modes algorithms										
Implement Latent Class Analysis algorithms										
Carry out a comparison on the results obtained in the K-modes and Latent Class Analysis models.										

- Note: source, self elaboration. Total project duration: nine months.

Chapter 2

State of the Art Review

This chapter will provide information on the topic of terrorism in general. It contains two sections that visit studies that used LCA and K-Modes. The last section will provide knowledge about previous studies on terrorism and the implemented techniques applied for their studies.

2.1 Terrorism

Terrorism covers different areas; there are studies such as terrorism studies, psychological, historical, and criminologist perceptions. In turn, terrorism is divided into different types of analysis: qualitative analysis, quantitative, empirical, historical, comparative analysis, among other types of studies. Literature review studies and perceptions of different researchers when seeking information on terrorism, such as *Schmid et al.* [33], provide extensive knowledge about the focus preferences of terrorism studies. Among its results, through an applied survey, it is evidenced that there is a strong preference for defining terrorism more focused on terms of violence or use of force with 91.1 %, followed by definitions with political elements with 82.2 %.

Terrorism has always captured global attention. Ever since its inception, groups have been formed to combat it. Methods that are of great help to fight it are applied in military and academic approaches. In the academy, large data repositories such as the GTD allow historically studying all kinds of behaviors. In Andary's study [34], they mention that for the United States, the FBI has divided this activity into categories such as domestic and international. Organizations such as Interpol also take actions in the fight against terrorism by raising priority items such as identifying terrorists to find out which are the current terrorist groups and thus identify suspicious groups based on their behavior. The social network analysis and chemical and explosive terrorism analysis concerning the different types of weapons used. The study provides conclusions such as radicalization knowledge to combat Jihadist terrorism as critical points for the fight against terrorism. The effectiveness against terrorism depends mainly on the speed of decision-making to fight or prevent its attacks.

Technological advance is relentless, and terrorism follows its pace. According to studies such as that of Morales [35], the use of new technologies has been of great help to terrorist groups. They use them to gain worldwide recognition and to recruit new members. In this study, there are definitions of terrorism in the educational field: as an action or sequence of violent acts planned and carried out aimed at an objective or target that is already defined. In the legal area, they give it a meaning: a type of violence previously planned against a defined target that is not in combat, developed by clandestine groups, whose purpose is to influence a specific public. It also provides knowledge of different types of terrorism, such as local terrorism, which focuses on attacking particular places with a clear and defined objective. Global or international terrorism aims to be recognized globally and has various objectives to execute its attacks. Finally, it provides conclusions on current terrorism and technology, such as the search for capacities in which they perceive the authorities of different countries, cities, or regions cannot control these events. Therefore, they tend to weaken the image of government figures in high decision-making positions.

Through technological advances, it is possible to obtain the behaviors of a terrorist group and, in turn, know if it may be involved in a new terrorist event. According to Pilley [36], the critical information about an attack in counter-terrorism is to find the name of the terrorist group that committed it. Security is an important issue that has been a priority in global governments and politics to decrease terrorist events. When it is possible to find the name of the perpetrator terrorist group, governments can take action to develop strategies to keep this group on the sidelines. The study developed an application where a user can predict responsible groups based on the information of a terrorist attack due to historical behaviors collected in a repository through data such as the type of attack, the nation, the objective, among others. This type of application allows knowing the names of the possible groups involved in the attack.

Hoolbrook et al. [37] define terrorism as political violence to test government norms. When it comes to religious beliefs, terrorists are not always followers of these; they can be insulted mainly through their ideologies. Virtual content such as documentaries or information on terrorism blogs online is the most significant source of persuasion for these people to lead to attacks. There has been plenty of research on extremist violence for a significant amount of time. Each time it is easier to find reliable information on the subject, but the ideology and violence in extremism continue to provide few indices of understanding. According to this type of study, violent extremism should be investigated more thoroughly and not only as a type of violence to withstand violence.

One approach experts use to explain terrorism is to do it in terms of right-wing, left-wing, and religious terrorism. The study by *Jones et al.* [38] focuses on analyzing the problem of terrorism in the United States according to historical data. They define right-wing terrorism as the use of violent means by racist, misogynist, and anti-abortion groups. In contrast, left-wing terrorism tends to be associated with groups that use violence justified in their nonconformity towards capitalism-imperialism, animal abuse, and support to radical political approaches. Religious terrorism uses violence supporting the faith in all existing religions. The number of terrorist attacks from 1994 to 2020 (893 events in the United States) is relevant information for this problem.

Right-wing terrorism has grown exponentially in the last nine years. Their findings mention a need to control right-wing terrorism as it has expanded since the 2020 presidential elections. Right-wing extremist groups also represent a danger in terms of violence against the federal state. Phenomena such as polarization and misinformation play a crucial role in increasing this issue.

Terrorist groups use violence as a means of intimidation to achieve their objectives. Economic policies related to terrorism regarding security, the increase in the cost against terrorism, and the decrease in the benefits against it play a critical role. These factors reflect in government regulations; these must be solid to prevent activities that benefit terrorist groups. Public and economical financing is essential since, in small countries, it directly affects their governments and foreign investment in them. The study by *Daniel et al.* [39]analyzes the KOF Globalization Index dataset since it has historical information from several countries. Using a regression model, they found that small countries typically have less liberal international policies regarding terrorism due to the loss of money and the lack of stability in terms of public finances.

For tourism, terrorism plays a vital role; tourists plan their trips considering the perceptions of security in the countries they wish to visit. According to *Seabra et al* [40], terrorism directly influences tourism. Coun- tries such as the USA, Italy, Spain, Ireland, among others, have been affected in tourism due to terrorism. Their study mentions that terrorism appears to generate tension with governments and the fight they have carried out against terrorism. On the other hand, Portugal has increased tourism due to the social, economic, and political environment that its government has created, which has increased tourism in this area. Their study [40] analyzed data sets with information on the arrival of tourists in Portugal obtained from the "National Institute of Statistics." Their analysis discovered a problem due to the instability of other countries, which directly affects tourism in the country. They also found that, depending on the countries affected by terrorism, it positively or negatively affects tourists from other regions, as explained in the case of the United Kingdom, Greece, and France, which positively affects the arrival of tourists from Asia.

Experts divide world insurgencies into those that seek to control territory and those that do not, as indicated by *De la Calle et al*[41], also indicates that between the years 1970 and 1997, the patterns of operation of terrorist groups are limited to low-income countries and those that are not. Likewise, the article looks at the insurgencies that appeared in the sixties, where the sources of violence impacted at the continent level. Its emphasis is on the guerrilla commanded by Fidel Castro to overthrow the then president Fulgencio Batista (1959). In this way, the insurgencies between 1970 and 1997 focused either on controlling an area or not.

An insurgency in a country is inversely proportional to the level per capita of its economy. The countries that have the most money are the ones that suffer the least the consequences of an internal problem. Take the armed conflict in Cuba, the civil war in Zambia, and the conflict on the border with Pakistan and Afghanistan.

Some articles[42] collect information from the New York Times dating from 1969 to 2014. Proving that hate speech is not new and results from a complex relationship between the Middle East and the Western hemisphere. Hate speech became standard after the incident of September 11, 2001. They analyze sociolinguistic behavior against Muslims and how public opinion affects this population. Speech of hate is how acts against terrorists negatively involve people from the West and the east—being in the middle of the media controversies opening a gap that Derek treats as "them and us"—making hate speech more prolific to generate even more hatred among the population of the world.

There is a phenomenon called foreign fighters: people of different nationalities that become part of the Islamic armed conflict. According to the article Collective Action, Foreign Fighting, and the Global Struggle for the Islamic State[43], there are cases in universities or local jobs of people who get involved in conflicts abroad. This phenomenon has increased due to globalization. In most cases, as *Schraeder et al* explain, these people come from politically unstable environments where individuals are highly qualified. These highly qualified foreigners are constantly subjected to hate speech, thus creating the need to defend the interests of those affected by the Islamic state by bringing economic capital or human capital into war.

Terrorist groups in the Islamic State have implemented different methods of instilling horror to carry a message. There is data of different types of messages that terrorist groups make, one of them and the strongest, who, according to

The London School of Economics and Political Science, is the videos showing different types of horrors. On the one hand, according to the reports, they make sacrifices as symbolism to the Islamic State [44]. On the one hand, the messages are to demonstrate the capabilities of terrorist groups and the possibility of reaching any corner of the world arguing the acts committed as ”retribution (denouncing the enemy’s way of war, in the grotesque); disgrace (humiliating the West, in the abject) and redemption (glorifying its mythical subjectivities, in the sublime)” [44]. Finally, one of the messages that are an essential part of the logic of these terrorist groups is the ability they have to recruit people and be able to carry out the guidelines of each group.

Blasphemy as a cultural activity in the Islamic religion fuels terrorist attacks, which they see as symbolic acts in defense of the religion. For this, some countries have blasphemy laws; *Saiya et al* [45] indicate that countries with these laws are more likely to suffer from the yoke of terrorism. It also indicates that to end these terrorist groups requires more than weapons since these laws against blasphemy encourage cultures to defend their thinking more reliably. This type of law can create problems of an international nature since there are entities in international instances such as the Organization of the Islamic Cooperation (OIC) that have even created laws against international blasphemy, and many of the orders given apply in predominantly non-Muslim territories.

Since the events of September 11, 2001, a hostile environment has been created for the people of the Middle East, as indicated by *Powell et al* [46]who also indicate there is more media coverage for this type of problem. She also elaborates on the study of 11 terrorist events between 2011 and 2016 to analyze the changes in the media. On the one hand, when attacks with firearms increase, there is a phenomenon of violence in a society, which is the object of study by Powell et al., where she indicates that ”gun violence in the U.S. has become the new normal.” This violent society leads to a mental problem between us in the West and the East, better called ”Us against them.” In today’s society, the news of a Muslim attack carries more weight than another type of terrorist attack, creating a hostile environment for both Muslims and non-Muslims, supporting this statement with the 237 articles of six non-Muslim terrorist incidents versus the 645 articles of five terrorist events.

Hate speech evolved with technology, which is reflected in the study carried out by the publications in the newspaper” Il Fatto Quotidiano”—revealing the different genres of dark humor addressing social problems, political arguments, or even medical issues. *Dynel et al* article [47] devotes itself to analyzing dark humor in response to terrorist attacks, specifically the publications of the Italian newspaper, thus analyzing the attacks that occurred in 2016 in Nice, France—understanding the reasons for the attack and the results in the form of dark humor. They found essential patterns in the posts of hate speech framed as dark humor. It also hints at posting racist humor against people who follow Islam. Different movements show society’s skepticism before political and public entities in a nation. Finally, black humor is shown as an enhancer of messages that terrorists can use in their favor—finding in this type of information support to reach more people with a message to recruit people or instill fear.

The central premise of a terrorist attack is to generate fear in the civilian population [48], where consequently, the regulatory entities have to make every effort to make the population believe that they are safe—starting with a process of counter-terrorism from the need of politicians to respond for their obligations towards the international community.

The need to deal with terrorist attacks has led the great powers or countries with more economic resources to invest in counter-terrorism. As a resource for this project, we find the analysis carried out by *Smith et al*, where they find, i) the averages of intensity invested in security ii) The administration policies have some effect on the security intensities iii) there is some interaction between the policies administration and newspapers during these intensity events.

They [48] were finding that the different intensities in administrations against terrorism are affected proportionally in the newspapers. Understanding how hate speech is reflected in efforts against terrorism. It has been evidenced that during the efforts of the last 24 years, terrorist events have increased as investments in each administration increase. It is also taken into account that during the transitions from Bush (increasing security) to the incoming Obama (decreasing security during the first 2/3 of his government).

Also, in the article [48], they point out the following: "Importantly, we also showed that the power of the administration is not absolute, as demonstrated by differences in the extent to which specific newspapers either increased or decreased in the average intensity of security framing, dependent on ideological alignment with the political administration. As expected, the Wall Street Journal was highly responsive to the increase in securitization driven by the Bush administration and highly resistant to the decrease in securitization driven by the Obama administration", allowing to understand that the speeches that are represented in the continent's most influential newspapers are demarcated by the political contrast they bring.

Some experts [8] indicate that since September 11, 2001, terrorist organizations from Islam were decentralized to become international organizations. Thanks to globalization, organizations have used multiple methodologies to carry the message of hate, allowing them to recruit other followers, alienating convictions, and strengthening criminal structures. At the same time, and with globalization, the authorities have been forced to evolve as these institutions evolve, allowing the use of data science to counter-terrorism. As stated in the *Nie et al* document, the authorities have evolved, allowing them to collect information, pre-process the information obtained, clean the data, and analyze it. Finally, the article sets out the problems faced by the authorities to combat terrorist events and explains that despite all the progress that has been made, privacy policies must still be reinforced; second problems they face when they find that they are not all government organizations do not share this type of information. Finally, the last problem mentioned arises from eliminating redundancy problems in the information, where centralized management of the information required to use Big Data against terrorism is required.

In Great Britain, collective anxiety arose between 2006 and 2011, leading the authorities to implement measures against crime, putting the Muslim communities in the authorities' focus [49], being in some cases victims of the state and harassed by the authorities as well as being the only ones to be stopped at a checkpoint and requisitioned more than any other ethnic group, as *Hargreaves et al* indicates. The limitations in the information have presented problems in the research since they present many questions about how the authorities require individuals. In the document, they pose a series of questions to detail and solve the significant doubts, are the people identified by their name? Does skin color identify them? among other questions. Finally, one enhancer for Muslim communities to commit terrorist crimes in foreign countries is the repressions of Islamophobia.

2.2 Studies Where LCA and K-Modes are Applied

2.2.1 LCA

As mentioned, the LCA model has been implemented several times in studies of social behavior. This model has also been included in medicine, social perceptions, human behavior, and others.

In order to know the behavior of a certain group of people, it is sometimes necessary to divide the study participants into different groups. In the study [50], they mention that the implementation of the LCA model has increased in recent years. On the other hand, their study compares the LCA model against a traditional technique such as the Hierarchical Cluster Analysis (HCA). Perez [50], explains the need to classify people into subgroups or classes through a survey to identify smokers, where the variables belonging to this study are polytomous. On the other hand, the purpose of the study is to provide knowledge about the application of the LCA implemented under different configurations, ensuring consistency in the results by selecting the best model. Said model is chosen using a cross-table analysis to know the results of the classification.

When it comes to perceptions of satisfaction or social behavior, we find studies such as [51] where they analyze the perception of a sample of 2997 students belonging to the Faculty of Economic Sciences and Administration of the UDELAR University of Uruguay. The study is based on knowing the response frequencies of a survey based on eight questions with different responses. This is to know which is the optimal model to carry out the study analysis. In this way, know not only the perception of the services provided by the faculty, also the satisfaction rates provided by the participants concerning these services—all of the above to improve services considering the perceptions of its customers (Students). The results show that the most representative variables according to negative indices have to do with age, dropout, career, and work for this study.

Behavioral studies such as that of [52] aim to identify school profiles through individual variables of students, families, and communities in Chile. This study comprises a population of 486,427 students from 32 different schools, where they studied a sample of 2683. Through a survey, they analyzed the student responses using psychological techniques. Other methods were used to determine the level of respect for teachers. The model with the lowest value of the Bayesian Information Criterion (BIC) has more conservative results than the Akaike Information Criterion (AIC), thus opting for more parsimonious models. Consequently, in the analysis of the information, the classes or clusters of the students were explained, providing a behavioral summary of the groups, finding that 48.8% have a negative perception of the institution. Students who perceive healthy climates and show respectful relationships towards the authorities attend private educational institutions and come from a family nucleus of (father and mother), who live in communities with high levels of security and social control. This represents only a minority of adolescents in Chile since only 8.5% have opportunities to attend private schools.

LCA has also been used for business methods in branches such as marketing. Taking into account the study of *rondan et al*[53], it seeks to make a grouping by some results obtained through a survey of mobile phone users concerning to the quality of service, satisfaction, trust, and intention buyback for participants. To choose the best model and analyze their results, they were based on the model with the minimum BIC. The users belonged to Movistar, Vodafone, Amena (Orange) companies, where 44.7% belonged to class 1, 37.6 % to class 2, and 18% to class 3. As conclusions, the authors contributed to the literature on relationship marketing because the data was efficiently segmented to identify buyback actions.

Studies such as Lapierre [54] seek to identify cyber aggression and victimization status groups. This is an important issue due to the increased use of technology and the media. The study analyzes proactive and reactive cyber aggression. On the other hand seeks to evaluate how the victimization-cyber aggression status groups have some relationship with evolutionarily relevant social advantages and disadvantages in resource control (social domain), reputation (social power). implicit) and reproduction (that is, number of dating and partner sexual relationships, friendship and attachment to the relationship, anxiety, and avoidance). The study comprised a sample of 400 participants in Canada, belonging to community groups such as athletic organizations, churches, extracurricular organizations. The best model to analyze was chosen using the minimum value of the BIC index. Within the results, they discovered that 79.4 % were involved in cyber aggression and victimization. They show that cyber-aggression and victimization are associated with social advantages, including social dominance, dating, and sexual behavior.

Protests have become an essential issue in the United States due to the problems they unleash in terms of public disorder and violence against the authorities. The [55] study aims to identify different groups of political participants. For their analysis, they used the American National Election Studies (ANES) survey to know the participation of voters and Protestants. According to their candidates, Oser explains the importance and need for models that allow knowing behavioral behaviors within society through their political ideologies. LCA plays a vital role within these models because it will empirically enable people to be grouped through similar response behaviors. Within the study, they use the BIC to choose the best model to be analyzed. The importance of the study concludes that LCA is useful for the study of the division of social behaviors based on historical data. Due to technology, models like LCA can be known as participant repertoires and their sociodemographic correlates. They found findings on voter views who tend to be more representative than non-voters because voters' different points of view lead to typical topics of debate.

In terms of violent behavior in young people in the United States, high school students have had numerous violent events. The study [56] aims to employ an LCA to compare the relationship between violent behavior and the use of multiple types of drugs. The study explains the questions they ask in their survey, such as the consumption of cigarettes, alcoholic beverages, drugs, use of weapons, fear, use of alcoholic beverages within school facilities, and feelings of insecurity. They mention that age and drug use have effects on students' behavior modes, according to the types of response and the clusters or classes that were determined. In turn, the behaviors of being a person who carries a weapon are seen in the cases belonging to groups that use drugs and have aggressive behavior, which generates a greater probability that the variables of drug use and violent behavior lead to carrying weapons in schools. It is essential to identify these behaviors to take preventive measures and bring a much more enjoyable school environment.

Many civilians are killed by police officers in the United States every year *wertz et al* [57]. This study is based on data from the National Violent Death Reporting System NVDRS. Their results provide information through a statistical analysis of data such as 616 deaths between 2014 and 2015 of civilians. Different groups were identified in which there was a type of weapon at the time of the operation. Among the weapons were firearms or knives for violent events. In other classes, non-violent events pertained to suicides of civilians involved with weapons and homicides of civilians involved with knives. The study leads to further research of the analysis of police officers' confrontations for the type of variables of a police-involved shooting. The importance of knowing this type of information is high because it can be used to carry out analyzes such as abuse of authority in many cases where the police may be involved.

2.2.2 K-Modes

According to *Cao et al*, information grouping is one of the most critical stages in data mining. Therefore, the analysis through the Kmeans extension is the mode applied in a grouping called KModes [24]. Especially the grouping of values is understood by the difference in the value of the similarity or dissimilarity of two objects, allowing to interpret that the values mainly used to find these values are with quantitative data types. The specific case of KModes allows the same type of analysis to be carried out with qualitative values. One of the most significant benefits of working with KModes is "A distinct characteristic of the new dissimilarity measure takes into account the distribution of attribute values over the whole universe." as stated in the document [24] called A dissimilarity measure for the k-Modes clustering algorithm.

Kmodes is essentially a way to classify categorical data [58], a way that *Chaturvedi et al* says uses a different procedure by using a limit of the L norm approaching zero. As indicated in the Monte Carlo simulations, Kmodes is more efficient than latent class procedures. It is also explained that the way to use KModes is complemented optimally with the latent class analysis (LCA) since finding the optimal groups in LCA is much more varied, unlike KModes. Thus, finally, they strongly recommend using both techniques in market segmentation, as it is indicated that the performance of both techniques is very similar, finding that KModes better handles large amounts of information and LCA allowing great versatility in the way of finding the number of most optimal groups.

Kmodes, despite its performance as a data mining tool, has some improvements in the works. *Nguyen et al* foresees the need to implement a system based on the algorithm created by *Cao et al*. Called private differential or DP [59], it is proposed a task of executing the same functions based on categorical values where the information can be grouped. Design PD schemas for interactive and non-interactive data. They also put as the main objective to identify the critical points in implementing a secure KModes model. However, it is proposed to go further into the subject of the theoretical analysis of the proposal based on algorithms under the private differentiator.

He et al has also proposed that KModes encompass a pattern recognition problem for data analysis. They [60] are exposing as a fundamental pillar for the KModes paradigm: i) to measure a dissimilarity of a simple coincidence for categorical values, ii) Modes as the primary basis of the analysis of the algorithm. According to the document ”Attribute value weighting in k-modes clustering,” they indicate an improvement in the performance of the traditional algorithm since the traditional algorithm, having improvements in response times with algorithms with LCA, requires some performance adjustments with large amounts of information. It also indicates uses of traditional KModes' variants, such as weighting the attribute value in the dissimilarity calculation.

Among the most relevant classical grouping methods, the so-called KModes are created as an alternative to categorize categorical information. By initially selecting a group, KModes defines the selection parameters to categorize the rest of the information. However, a study [61] indicates the precision of the results obtained, finding the deployment of different experiments to determine that the correct deployment of the model depends clearly on the selection of the initial parameter, which is the one that makes the difference in the rest of the group. However, the study does not detail performance or the imprecision that the model may have.

Chapter 3

Theorethical Framework

3.1 Terrorism Study Techniques

There are many techniques for criminal studies. In terrorism, the methods should focus on discovering relationships, behavior patterns, and possible predictions of a terrorism event. There is relevant information for the study and analysis of terrorist behavior, such as the methods of attack they use, the weapons they use, the mortality rates of the events, among other variables.

Taking terrorism as a matter of public interest as a starting point in which this phenomenon can directly or indirectly involve members of society. Large amounts of information are added daily in different repositories. According to the study by *Vajjhala et al* [62], there are nine types of terrorist attacks: Murders, Armed Assault, Hijacking, Barricades, Bombing and Explosion, Kidnapping, Unarmed Assaults, Infrastructure Attacks, Multiple or Unknown attacks. In their study, they use the method of correspondence analysis according to the types of attack by region. This tool is helpful for decision-making since it allows to know where authorities should enforce counter-terrorism policies, according to identified types of behavior. As explained in this paper [62], the correspondence analysis belongs to the multivariate exploratory techniques, which we use due to their ability to provide statistical estimates and graphical diagrams based on relationships between variables. According to this study, the most common variables are dimensional analysis, correspondence analysis, discriminant analysis, Euclidean distance analysis, and vector analysis of contingency tables.

Other studies designed to analyze terrorist events and terrorist organizations. One of them uses game theory. Sandler's study [63] explains the importance of applying this method. Game theory is applied to know the possible answers of an opponent through historical data. Governments frequently use this theory to find ways to predict terrorist attacks, taking into account previous events. Another type of technique used in this study [63] is Intervention analysis since it allows us to know sudden changes in historical behavior through applied politics, which enables us to see the impact on these policies.

According to Tiwari [64], these techniques strongly influence law, crime data analysis, and the discovery of fraudulent banking transactions. In this case, K-means seeks to group information based on training examples of the model to find similar groups. Another K-means method used in this study [64] is the Enhanced K-means Algorithm. This algorithm has as input an amount of data x with a predetermined number of clusters. All this is to calculate the behavioral patterns for the division of groups.

In the study of *Huamaní et al* [65] Machine learning (ML) techniques, such as the decision tree, are a type of model composed of nodes and leaves to perform classification. On the other hand, the Random Forest is also a utility model due to a higher degree of complexity and great accuracy when classifying information. These models help classify behaviors based on historical events to combat terrorism and predict possible terrorist attacks.

Normally exploratory analyses of *Bhatia et al* [66] are produced on variables of interest to see symmetrical plots. This study helps summarize the information and group it, allowing experts to make decisions and identify risk areas. Their results show that private citizens are the most affected due to terrorism issues.

Hate speech is a problem in terrorism [67], and with hate speech, disseminating information has evolved to have a global impact [68]. Consequently, Halloran finds a way to interpret the text of images related to acts of violence in hate speech through data analysis methods [69]The study's objective is to legitimize the information spread by terrorist groups, allowing the identification of both used materials and propaganda and determining the reused material on different platforms. Finally, *Halloran et al* propose interdisciplinary strategies for studying hate speech. Different articles point to a new form of counterterrorism; the case is for the article ”Hypotheses Analysis and Assessment in Counterterrorism Activities: A Method Based on OWA and Fuzzy Probabilistic Rough Sets” [70]. The authors develop the article using a combination of probability, approximate, and fuzzy theories. They focus the information analysis on already known terrorist attacks, allowing the model to parameterize the information with decision outputs. Providing research through the three-way decision theory (3WDT), access the universe of information and dividing it into three positive, negative, and border parts. Finally, the information is classified thanks to the tool designed for decision-making.

In conjunction with counter-terrorism, the data science revolution has significantly impacted security and intelligence. According to REF, data science has divided opinions into those who agree with the use of technologies and those who do not. However, it is highlighted how effective emerging technologies have been in making decisions against terrorism. Artificial intelligence and counter-terrorism have been used together to implement various solutions worldwide, such as determining terrorist structures, identifying disputes, recognizing the increase in terrorism online, and so on [71]. In summary, as indicated by *Ganor et al*, as time goes by, more structured information is required on a large scale to carry out more implementations of the type of counter-terrorism. Finally, the future of artificial intelligence must be based on respect and good handling of data from cautious and respectful scientists of the population, returning people the hope of having an accessible and equitable democracy.

The more data is recovered on the internet, the more the dilemma framed in the article [72] is questioned, who consider a problem of conflicts between national security and people who may be victims of privacy. Considering that fingerprints can be used both for convenience for citizens, they can also be used for purposes against citizens. *Verhelst et al* It aims to lead the reader to highlight the importance of three kinds of problems: class imbalance, the curse of dimensionality, and spurious correlations. Machine learning is also one of the fastest-growing areas in computer science, allowing implementations in many fields, an interdisciplinary nature. Feeding the appropriate algorithms, we can find as many advances as possible, allowing us to find behavior patterns, data classification, and other outputs that can be obtained. However, in the case of predicting terrorist attacks, the files have to pass through susceptible areas for the civilian population, bordering between what is morally correct and what is incorrect. According to security, according to *Verhelst et al*, it is essential to determine the border of the third parties that can intervene in this type of information. Who considers and concludes that mass surveillance algorithms have a significant challenge in the future to deal with the significant errors that may be faced.

Terrorism has excellent repercussions on tourism, as shown by *Cheng et al* [73] who indicates that tourism in Xinjinag has grown. Expressing that the social situation and violence have improved, strengthening the commerce and tourism industry. It is finding a direct relationship in the opportunities that can find in a region with terrorism. It is also indicated that it is essential to increase efforts against terrorism and investments in strengthening investment in science and technology. The document argued the integration of all police, military and maritime forces.

3.2 Terrorism Issues

As explained in the study by [63] terrorism impacts are measured in specific measures such as 1) Direct Costs, which are calculated based on damage to property, equipment, structures, among others. 2) Indirect Costs: which include trade losses, security costs, among other costs. According to this study for the United States, in one of the most critical terrorist events such as 9/11, the estimated Direct and Indirect costs were around 90-100 million dollars.

The consequences of violence, such as corruption and illegal groups, are factors that trigger terrorism. As stated in [74] variables such as 1) Physical Consequences: such as bodily injuries such as burns, fractures, disability, among others. 2) Mental health and behavior problems: which are concerning, alcohol, drugs, depression, anxiety, suicidal thoughts, post-traumatic stress disorders, among others. 3) Chronic diseases: cancer, cardiovascular disorders, kidney problems, strokes, among others, are factors that make it difficult for people to interact due to their ethnicities, races, or religious beliefs due to the exclusion of people. All the aforementioned fits into the difficulties generated for the interrelation of people due to their ethnicities, races, or religious beliefs, providing a conflict observed in the exclusion of people.

The tourism industry is one of the sectors most affected by the phenomenon of terrorism. According to Martín [75] countries that count on tourism as one of their primary sources of income means that security and the fight against terrorism is an objective of great relevance for their economies. As mentioned in the study, world-renowned events such as fairs, Olympic games and business meetings are affected by this problem, thus deteriorating the income of their economies due to the low perception of security.

Studies in terrorism are essential due to the behaviors that arise from the different study approaches in this field. These analyzes contribute to directing the notion of terrorism in other areas, leading to important findings on political violence. Social media plays a fundamental role in the problem of terrorism. A clear example is seen in the study of [76] the concern of the Islamic State to fight rebel groups, the Syrian government, and get involved in economic activities.

Terrorism in Syria in conjunction with social media has revolutionized the methodology of spreading hate speech. Through various platforms, terrorism has evolved with more than 6 billion hours of reproduction only on "YouTube" [77]. That way, social media is found in terrorism its various forms to escalate the regional problem globally. *Awan et al* intervenes with more than 100 Facebook pages of the Islamic State, and 50 different user accounts on Twitter to determine what impact the material spread on the internet has, what type of strategies they use in recruiting personnel, and how they are seen on the platforms social networks. Determining the speed with which terrorist groups adapt as the media advances, social networks being powerful media to get a message across.

The dissemination of information plays a fundamental role in the problem of terrorism. Therefore, and as mentioned in the study by *Silva et al* [78], social media has an important responsibility when it comes to sharing virtual content. The perception of security is linked to what people see daily in their networks or news sources. On the other hand, and as *Silva et al* mentions, social networks are responsible for sharing myths regarding the information on terrorism. These myths have left negative images of Muslim migrants who cross the border from Mexico to the United States to perpetuate only terrorist attacks. According to this study, one of the most relevant myths is that Jihadists inspire extremism, where events like Jihadist-inspired represent 18.3%, while Far-right represents 49.6% and Left-wing 32.% Which leads to knowing the importance of the information shared on extremism and leading to perfect studies on national discourse, political creation, and laws to combat violence.

Knowledge of terrorism such as violent actions for political or population control purposes, the objective most affected by terrorism being government institutions [1] plays an important role. But there is a need to be in the capacity and know the causes that trigger terrorism and the events that take place. Taking into account Mawloo's study [79] it is evident that this phenomenon does not differ between ethnicity, status, or religion. Simply attacks are carried out with a common objective depending on the organization that carries out said event and the affected victim. On the other hand *Mawlood* mentions that problems such as social, economic, and political instability in terms of security are variables of high impact on the issue of terrorism at a global level.

According to [80] since the beginning of the '80s, globalization has touched tourism, allowing large masses of tourists to access tourism in areas such as Europe, Asia, and the Pacific. However, there is a significant decline in tourism from numerous causes, such as critical stability problems related to terrorism, migrants, and political problems as indicated *Brondoni et al*. It is also indicated that the most significant movement of tourists is given between June and September, except for countries with more stable geography that can be extended over time. It is essential to highlight the negative impact left by terrorist attacks, leaving a negative impact both globally and locally, demonstrating the volatility of the tourism business. *Brondoni et al* indicates that tourism in an area can take from 13 months to 24 months, depending on the impact, to retake the natural course of business.

According to [38] the United States of America will go through one of the worst terrorist crises during 2022 as they explain that some decisions that were taken in 2020 will have substantial repercussions shortly. They explain in 3 parts the breaking points: 1) the 2020 votes for the presidency were a matter of polarization for many North Americans, increasing the sources of hatred exploited by the extremists in the United States of America. 2) The events of 2020 that left during the pandemic as unemployed or some officials threatening to close non-essential businesses in response to the second and third waves of Covid-19 have increased hate speech promoting violence and the need to regain freedom. 3) Racist events or school shootings and mass protests. These three stages are presented, which according to *Jones et al* indicates the need for all parts of the United States of America to have an essential role in the framework of counter-terrorism, thus reinforcing the rules of anti-hate speech on the various network platforms we can tackle this global problem.

Since the events of September 11, 2001, counter-terrorism efforts have increased globally [81]. Also increasing the level of research and, in turn, the level of internet searches. Finding that Al-Qaeda is one of the most sought-after items when it comes to terrorism and extremism from 2007 to 2016, a period in which counter-terrorism/war on terror and Afghanistan are also searched. As also *schuurman et al* indicates, the study of this phenomenon began to increase only in Africa until 2010 and 2014. In Asia, the percentage of articles decreased from 2007 with 21.7% to 10.7% in 2016. While in Asia Northern and Western Europe, the percentages in the same years increased from 13% to 22.8%. The literature analysis was carried out with 3442 articles published from 2007 to 2016 in nine academic journals. They obtained Terrorism and Political Violence (TPV) and Studies in Conflict & Terrorism (SCT) as the primary source in journals.

There has been a change in the concept of terrorism since what happened on September 11, 2001, where violence is the only common factor. However *Gaibulloev et al* classifies terrorist attacks in 2 parts, domestic terrorism whom those affected and the perpetrator are citizens of the same country. Transnational terrorism involves two citizens, one from the country where the events occur and the other from a different citizenry; in this way, the two forms of terrorism in the world are detailed according to *Gaibulloev et al*. The document [82] also indicates the evolution of domestic terrorist attacks and transnational terrorist attacks. The study concludes that since 1990 transnational terrorist attacks have drastically decreased to 40%; however, they indicate a more significant number of victims involved.

On the other hand, internal terrorism increased until 1992; since then, the pattern of fall was similar until 2008, increasing terrorist attacks. A change of headquarters is highlighted after the attack of September 11, 200 to the Middle East, Africa, and South Asia, this fact in response to the industrial countries for adopting defensive measures against terrorism. The post 9/11 era (REF PERSONA) expresses the use of hostage-taking and armed attacks. It is essential to highlight the use of two different databases for the analysis in the article [82] as well as the GTD database [2] and the database ”International terrorism: Attributes of terrorist events” [83].

3.3 Islamic Terrorism

Various groups are derived from the invasion of the United States of America that was born in 2003. There is a terrorist group that was born in Iraq through Abu Musab al-Zarqawi commanding as a volunteer in Herat, Afghanistan. This non-al-Qaeda group is a precursor to AQI [84], a pioneer group after the events of invasion, taking power in 2013 through fear and horror thanks to *Laub et al*, who explains the form and methods—using car bombs and suicide attacks, making this one of Iraq's bloodiest years. After the expansion of ISIS in the territory, establishing control zones on the borders with Iraq and Syria, transferring the police forces and terrorist groups from the area, ISIS remains in power as of January 2014. In June, they take control of the rest of the area. Thanks to desertions and making a withdrawal at night, ISIS takes control of the entire city. Thus demonstrating that the areas in this country were divided by terrorist groups and determining the breaking point to show since when the terrorist group ISIS was at the fore in this area.

ISIS has a systemic organization [85] where there is a territorial vision divided into four categories where issues related to the state of the terrorist group ISIS, take into account the geography where the terrorist groups are present, highlight the territorial concept and ideology of ISIS, and categorize the group as an insurgency or revolutionary state. The ISIS terrorist groups in Syria and Iraq contradict the historical practices of Islam. The terrorist group ISIS does not act only because of its geographical position, but they behave as a single cell worldwide. Finally *Kadercan et al* indicates the need to create synergies around this problem, since scientists, analysts from different fields in international relations, and especially political geographers; make their own decisions, leaving aside the main reason for the problem, which is the civilian population and all the problems that revolve around it.

The transformation over the years for the terrorist groups of the Islamic State is imminent; there is evidence that shows a decrease in productivity part. They began to produce half of the propaganda from 2015 to 2017. Through *Winter et al* [86] the study is carried out focused on who has directed the propaganda of these terrorist groups because, until 2017, it is shown that the propaganda has had a digital transformation. Finally, it is shown how the Islamic State has used the media to distort digital reality, putting all its efforts into telling a different reality to its supporters.

There is a recent problem with Indonesian women, where they have been included as an essential part of terrorist attacks. As indicated [87], they include women in suicide attacks taking advantage of the vulnerabilities of migrants or women who want to be part of a change. It also indicates *Nuraniyah et al* that this inclusion of these women for these purposes can be avoided by helping women who are victims of these terrorist groups and by not allowing Islamic

State groups to brainwashing these women. However, the study also shows how it is intended to change one indoctrination for another through tutoring and community aid, leading him to the same terrorism problem but with a different name.

The media have an important position in world terrorism. The content they share is mainly given by terrorist events of the Islamic State or Muslims.

This content is sensitive and critical because, by not being differentiated, these events can generate phobias towards Muslims in the United States. Concerning this, the study by *Von Sikorski et al* [88] mentions that the media are mainly responsible for spreading fear due to the content they share. Through a survey, the researchers provide results on the adverse treatment of Muslims due to shared news where it does not differ if it came from IS terrorism or was a Muslim attack. Not differentiating the groups involved in a terrorist event leads to social problems regarding the attitudes that others take towards Muslims. Considering the above, it is clear that if the media does not do a proper job to differentiate the type of groups involved, it leads to the rejection of Muslim communities for their stereotypes. In another study *Von Sikorski et al* [89] mentions that since the attacks attributed to IS in 2014, observers continue to analyze which acts should be considered terrorism. Islamic terrorism is generally associated with ”Radical Islamic Fundamentalist calling for Jihad” and with people directly involved in Islam and Muslims. Countries such as Barcelona, Brussels, Berlin, London, among others, have been victims of terrorist events in recent years [89]. It is there where the media play a fundamental role because 1) news shared in great quantity achieves the main objective for terrorist groups, which is worldwide recognition. 2) The feeling of insecurity is generated worldwide in the areas where these attacks occur most, thus negatively affecting Muslims in the face of the negative perception that other members of society obtain from them.

There are a large number of Muslims who are in economic distress. The study by *Sirgy et al* [90] mentions that this large number have beliefs that terrorism is a viable means of solving problems such as corruption, crime, social conflicts, and politics. This is a subject of great attention because most Muslims with strong religious beliefs blame national issues, thus accepting terrorism as a national solution. They conclude that there is a need to create a type of education about modern society and democracy through a survey. This is for understanding the new existing methods to combat corruption, crime, and other essential variables.

Gomes [91] explains in detail how the groups emerged: 1) Al-Qaeda: which arose due to the Soviet invasion in Afghanistan in 1979, turbulent relations between the West and the Muslims. This group has three objectives: serve as a terrorist group, to recruit Muslim militants for the fight beyond Afghanistan, and finally to be a resistance to lead and unify the Jihadist movement. 2) ISIS:

which gets its origin due to the intervention of the US military in Iraq in 2003 and the civil war in 2011. These differences found by [91] conclude that ISIS is considered as a state while Al-Qaeda is viewed as an organization.

In the Islamic State, women have played a fundamental role in pro-terrorism. In the study by *Patel et al* [92] they explain that women have contributed to the implementation of clear objectives in a certain way, have carried out terrorist attacks such as explosions, or have even immolated themselves. They have also taken up tasks like recruiting people, transforming people to their religion, and all the above without leaving radicalization aside.

Chapter 4

Thesis Development

4.1 Global Terrorism Database

GTD is a database belonging to the National Consortium for the Study of Terrorism and Responses (START) [2], an organization of the University of Maryland. The GTD is a classified database that has collected historical information regarding terrorist events worldwide and has approximately 190,000 data. Technology has played a fundamental role in collecting information because it has made it possible to access large amounts of material to be studied and stored. This database has been updated by experts taking into account specific criteria such as:

- The act must be aimed at achieving a political, economic, religious, or social objective.
- There must be evidence of an intention to coerce, intimidate or transmit some other type of message to a broader public than the immediate victims affected by the event.
- The action must be outside the context of legitimate war activities.

The compilation of data based on these criteria allowed the experts to consolidate said database, ensuring consistency in the information added to this repository. On the other hand, the collection of information was carried out in an automated way and, in some cases, manually, thus allowing accurate information on terrorist attacks. It was necessary to translate more than 80 languages into English to unify the information.

In addition, this organization and database contain a guideline, which allows understanding how it works, and the updates that this repository has had. On the other hand, it provides a detailed explanation of the variables that comprise it. We used the following variables and definitions, according to the Codebook [93] of the GTD, to carry out the analysis for this study:

- ID: unique classification of each attack, containing data such as year, month, and day in which the event occurred, followed by the sequence of the event development.
- Year: the year in which the event occurred.
- Country: each country is classified with a number; there is also a variable that contains country names.
- Region: the regions have their code variable and the name of each region.
- Attack type: this variable contains the code and the name of the different types of attack classification, divided as 1) Murder which is geared more towards perpetrating attacks on specific and prominent people. 2) Armed Assault is intended to cause physical harm or death directly to human beings through the use of firearms or sharp instruments, which do not include rocks, fists, or other hand weapons. 3) Bombardment / Blast attack is caused by a hazardous material and causes physical damage to a surrounding environment. 4) Hijacking, the objective is to take control of a vehicle such as an airplane, boat, or bus and divert it from its programmed destination to force the release of prisoners or political targets. 5) Hostage tacking (barricade incident) the objective is to take control of hostages to achieve a political objective, be it political agreements or concessions. 6) Hostage tacking (kidnapping) the objective is to take control of hostages to achieve political agreements or other types of agreements; it differs from the previous one because it involves moving the hostages to another place. 7) Facility / Infrastructure Attack the objective is to cause damage to a non-human target, excluding the use of an explosive. 8) Melee the objective is to cause physical harm or death to human beings by any means without the use of explosives, firearms, sharp instruments, or incendiary instruments, except for attacks with chemical, biological or radiological weapons since they are considered as unarmed assaults. 9) Unknown
- Weapon Type: this variable counts as the others with its code variable, and the name that corresponds to each type of weapon among the types of weapon are: 1) Biological. 2) Chemical. 3) Radiological. 4) Nuclear 5)

Firearms. 6) Explosives. 7) Fake Weapons. 8) Incendiary. 9) Melee. 10) Vehicle. 11) Sabotage Equipment. 13) Unknown.

- Target / Victim Type: this variable has the numerical and textual description of the type of victim or target to which the terrorists aimed: 1) Business: such as Restaurants, Banks, Commerce, Industrial sector buildings and Pharmaceutical sector buildings, Entertainment buildings such as Casinos Stadium and Cultural buildings, Construction sector and Farms or Ranches. 2) Government (General): government entities and figures like Government Buildings excluding Embassy buildings and Diplomatic personnel. 3) Police: Includes all types of buildings and people that belong to the police except the military. 4) Military: Military personnel buildings including maritime and Military aircraft, weaponry, and convoys. 5) Abortion Related. 6) Airports & Airports & Aircraft. 7) Government (Diplomatic): Diplomatic buildings such as Em- bassy's, Consulates, Organizations, and Diplomatic figures. 8) Educational Institution: Including teachers. 9) Food or Water Supply. 10) Journalist & Media. 11) Maritime: Including civilian and commercial maritime ships and Oil Tankers, it also includes ports and maritime facilities. 12) NGO. 13) Other. 14) Private Citizens & Property. 15) Religious Figure / Institutions. 16) Telecommunication. 17) Terrorists / Non-State Militias. 18) Tourists. 19) Transportation (other than aviation). 20) Utilities. 22) Violent Political Parties.

4.2 Period Division

The periods identified in this project are determined by lapses of 10 years where multiple events seem to have been able to modify the data variations as can be understood in the document [9]. Some studies show that the stage of multiple presidents in the United States of America (USA); has acted as a determining factor in the increment of hate speech present in TGBI. Therefore, determining these periods is of high importance for the project. According to the different US presidential periods, an essential factor in this period division was the perception of hate speech and the security of American citizens. Under these time lapses, said studies showed an increased concern by the United States citizens regarding terrorism and becoming possible victims.

4.2.1 Period 1

The first period began in 1986 when the database began the constitution of the information until 2000. As explained by Smith et al., it was after the terrorist attack on the United States in 2001 that hate speech increased; For this reason, it has been chosen in this study to define the first analysis period between 1986 and 2000, collecting information since the creation of the database.

This period includes Bill Clinton's mandate, as *Benjamin et al* [48]. mentioned. In this lapsus, the perception of insecurity by a terrorist attack had relatively low values on American citizens considering Oklahoma's City Bombing event occurred in 1995, which consisted of a series of explosions in this location.

4.2.2 Period 2

The second period dates from the event that occurred with the twin towers in the USA in 2001 and when George W. Bush began as president. This period, Smith explains, exhibits a more pronounced trend line in the record of counter-terrorism and terrorist actions [5]. As mentioned in the study, since the 9/11 attacks, a problem has been carried out between Islam and terrorism because such attacks and the continuous discourses have created hate speech, translating into religious and ethnic tension in the United States. Considering this, it is evident that there was an increase in insecurity from US citizens for this presidential term.

4.2.3 Period 3

Finally, the last period is given between the lapse of change of presidency of George W. Bush and Barack Obama because it is there, according to Smith, where the national security investment took a downward trend. In this project, the differences in the timelines use Smith's study to delimit the three periods of activity of TGBI mentioned. This period represents a high increase in the perception of insecurity in front of US citizens, reflected by the worldwide proclamation of the Caliphate by ISIS.

4.3 Data Processing

The project will be developed based on the pillars of the techniques compiled by *Han et al* [12]who indicate the need to eliminate null values, exposing six possible solutions to this need. For this project, we use the option to ignore values first, which is indicated as the best solution for value classification problems. Second, to eliminate noisy data, it is necessary to use graphic tools to filter information. Tools such as boxplots are suggested to show how to proceed with the information. Thus, determining the Outlier Analysis technique, we defined two data groups, the first group with more than 50 terrorist attacks and the second group with less than 50 terrorist attacks. This last technique was extended, allowing adopting this solution to the thesis requirements. Finally, through this adaptation of the technique, three previously described periods were determined, thus concluding the adaptation carried out where we suggest using a previous grouping of the data. It concludes with two types of grouping prior to the solution described for this project. The number of terrorist attacks determines the first grouping. We used the second grouping with periods. At the end of this section, it will be possible to determine to use these selected groupings thanks to the use of the description provided by Python with its data analysis tool called ”Description,” which shows the standard deviation, mean, minimum, maximum, and the quartiles of the analyzed information.

In order to do what was described above, we analyzed the information to determine the null values. In this way, as shown in table 4.1, a difference is evident between the minimum, first quartile, second quartile, and third quartile (see figure 4.1). We can find a maximum value with a notable difference inconsistent with concluding anything from the data collected.

Table 4.1: General description

Count	3617.000
Mean	52.934476
Std	1455.936869
Min	1.000000
25%	1.000000
50%	2.000000
75%	5.000000
Max	86261.000000

- source: self elaboration.

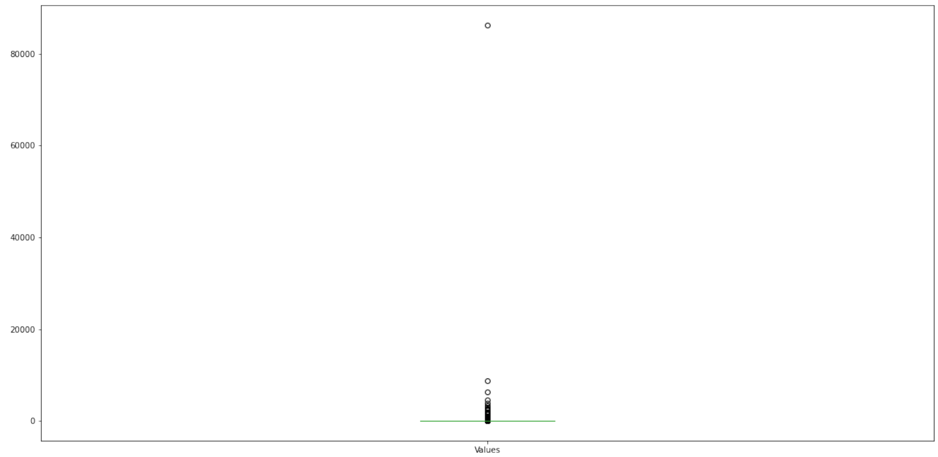

Figure 4.1: Plot of the General report

In this way, the first step to clean the data is to discard the missing values according to *Han et al*; therefore, after cleaning the missing values on all the columns, the result can be seen in the table 4.2.

Table 4.2: Cleaned missing values

count	3286.000000
mean	27.658551
std	219.442613
min	1.000000
25%	1.000000
50%	2.000000
75%	5.000000
max	6601.000000

- source: self elaboration.

Finding an adjustment in the standard deviation, the mean dispersion fell from 1455.936 to 219.442, a notorious reduction in the maximum values, going from 86261 to 6601. However, the data are still inconclusive since the percentiles cannot yet be identified to conclude on them. Where graphically it is concluding in the figure 4.2.

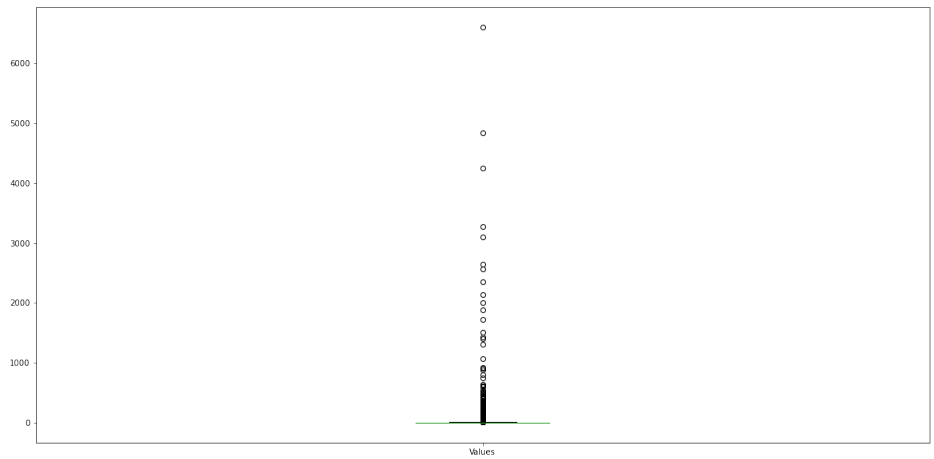

Figure 4.2: Plot after dropping missing values

Although the outliers are a bit cleaner, the data is inconclusive. Thus, in this way, in a continuous search for the development of the analysis, we find a critical item to analyze: the number of terrorist groups with non-relevant data levels. Finding that of the 3286 data previously analyzed, 3087 belong to the data group of terrorist groups with less than 49 records of terrorist attacks, as shown in the following table 4.3.

Table 4.3: Groups with less than 50 attacks

count	3087.000000
mean	4.426628
std	7.444788
min	1.000000
25%	1.000000
50%	1.000000
75%	4.000000
max	49.000000

- source: self elaboration.

On the other hand, through this filter, we selected the terrorist groups that are of importance and that provide valuable information for the development of the analysis. Obtaining 199 terrorist groups worldwide that have at least 50 terrorist attacks, as is shown in the following table 4.4.

Table 4.4: Groups with more than 50 attacks

count	199.000000
mean	388.045226
std	811.863975
min	50.000000
25%	72.500000
50%	128.000000
75%	274.000000
max	6601.000000

- source: self elaboration.

Employing the last filter, it is possible to interpret the information, where we graphically 4.2 find some quartiles that begin to have a form for the research. However, the outliers misrepresent the information.

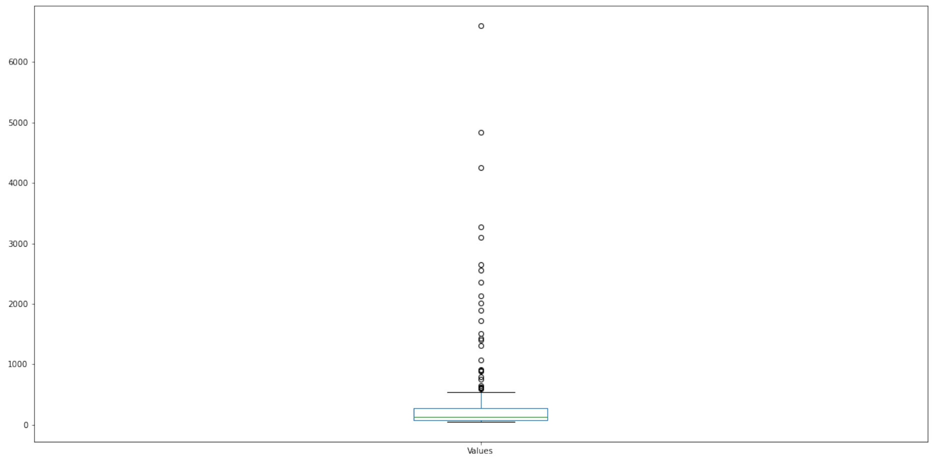

Figure 4.3: Plotting groups with more than 50 attacks

Through an analysis of the 199 groups found, we separated a list of all the terrorist groups associated with the Islamic state, allowing us to filter and reach the goal of cleaning the data to start treating them. Below is the list of groups obtained.

1. Islamic State of Iraq and the Levant (ISIL)
2. Al-Shabaab
3. Taliban
4. Al-Qaida in the Arabian Peninsula (AQAP)
5. Al-Qaida in Iraq
6. Sinai Province of the Islamic State
7. Khorasan Chapter of the Islamic State
8. Algerian Islamic Extremists
9. Moro Islamic Liberation Front (MILF)
10. Al-Gama'at al-Islamiyya (IG)
11. Al-Qaida in the Islamic Maghreb (AQIM)
12. Armed Islamic Group (GIA)
13. Tripoli Province of the Islamic State
14. Palestinian Islamic Jihad (PIJ)
15. Islamist extremists
16. Islamic Salvation Front (FIS)
17. Islamic State of Iraq (ISI)
18. Lashkar-e-Islam (Pakistan)
19. Jemaah Islamiya (JI)
20. Al-Qaida
21. Thai Islamic Militants

By filtering the 21 terrorist groups, it was possible to find values that fit the project's needs. In this way, and through the following table, it can be observed how the quartiles take a form that allows an adequate reading of the values. However, it is essential to highlight that the standard deviation that we obtain is still very dispersed since the division of project periods is missing up to this point.

Table 4.5: Groups related to Islamic groups

count	21.000000
mean	922.333333
std	1744.283129
min	58.000000
25%	140.000000
50%	236.000000
75%	426.000000
max	6601.000000

- source: self elaboration.

Graphically 4.4, we can identify the critical points of these values. So then we observe four outliers.As mentioned above, we can start using these values to draw conclusions.

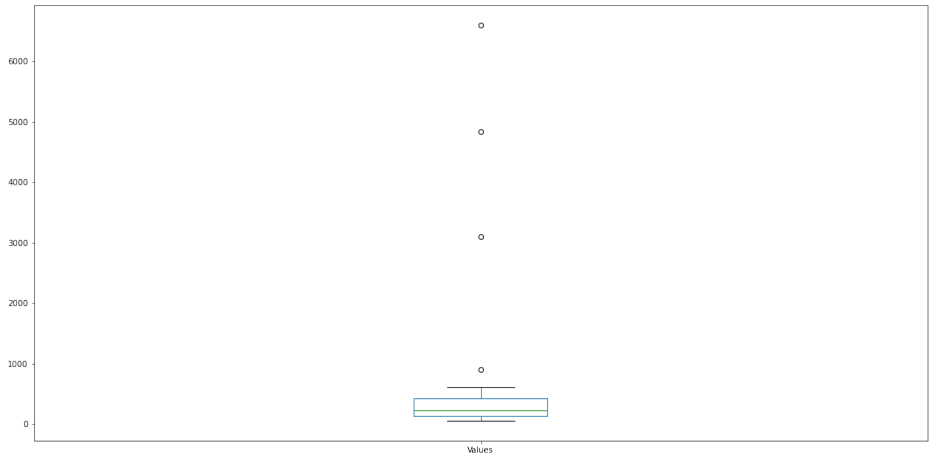

Figure 4.4: Plotting 21 groups related to Islamic groups

Finally, we established the period between 1985 and 2000 as period one, between 2000 and 2008 as period two. For the last period, established between 2008 and 2018, the values ended up settling, allowing the research development to interpret the values obtained as shown below.

For period one 4.6, we find a standard deviation of 75,433. Unlike the previous values, it is a dispersion of information according to what is required by the minimum standards. We find the information of the quartiles much fairer with the information obtained.

Table 4.6: First period

count	10.000000
mean	90.800000
std	75.433267
min	3.000000
25%	29.750000
50%	86.000000
75%	139.000000
max	236.000000

- source: self elaboration.

We can see how the data is adapted to our research through the graphics 4.5.

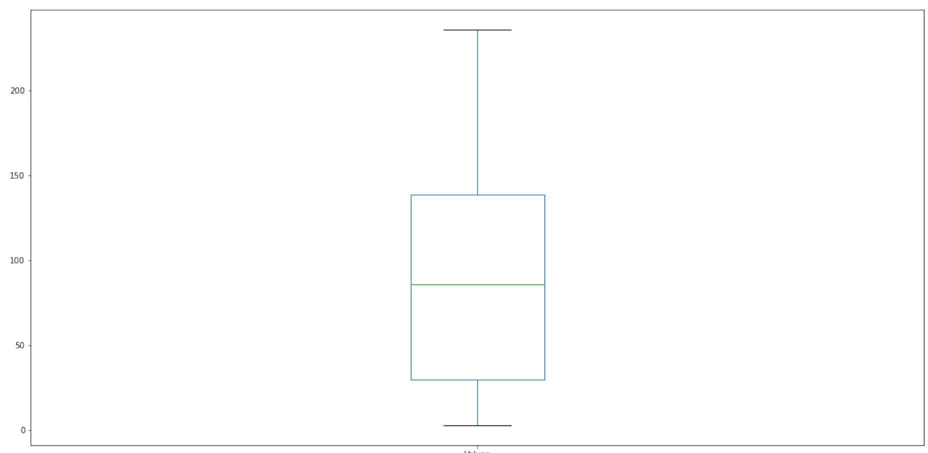

Figure 4.5: Plotting the first period of time

We obtained 14 terrorist groups from the Islamic state with a higher standard deviation than the previous one for period two 4.7. However, it is more data, and it is a fair value that we find in the development of the research.

Table 4.7: Second period

count	14.000000
mean	131.857143
std	196.804416
min	3.000000
25%	30.750000
50%	63.000000
75%	141.000000
max	766.000000

- source: self elaboration.

Graphically 4.6 we observe an outlier value outside the standards. However, we observe the rest of the values with a normalized behavior around the mean. Below is the graph.

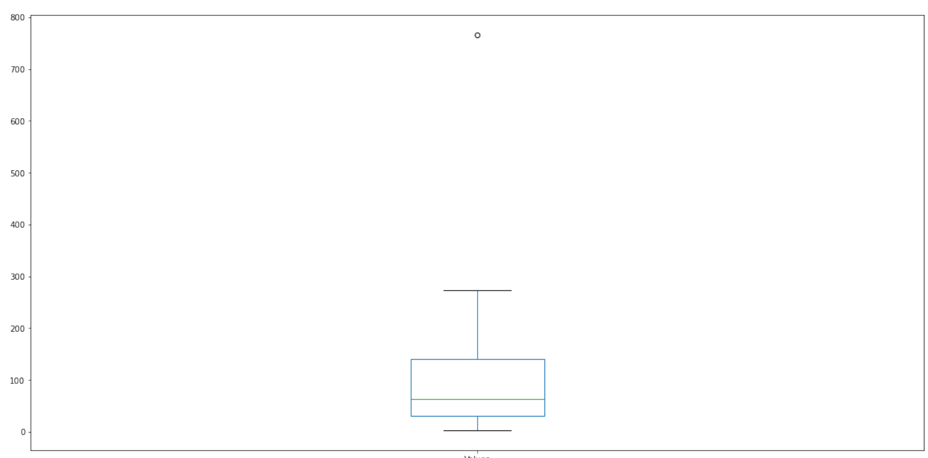

Figure 4.6: Plotting the second period of time

After period II, by eliminating the atypical values as evidenced in Table 4.8, and comparing with the previous value obtained from the same period, without atypical values, we find similar behavior in the standard deviation, to a lesser extent than the parent with the outliers but essentially the same shape.

We corroborate this information with the distances between the quartiles, the minimum and the maximum value of terrorist attacks. Graphically we find a significant distance between the quartiles of the minimum and the maximum value of terrorist attacks.

Table 4.8: Second period

count	13.000000
mean	83.076923
std	76.618603
min	3.000000
25%	30.000000
50%	60.000000
75%	132.000000
max	273.000000

- source: self elaboration.

Graphically 4.7, the highlighted central values are detailed; a distributed value is seen with an upward approach. We observe the distribution of the most focused quartiles up to the fourth quartile, where a rise in these values is denoted—concluding with a volume of terrorist attacks between 3 and 132 terrorist attacks.

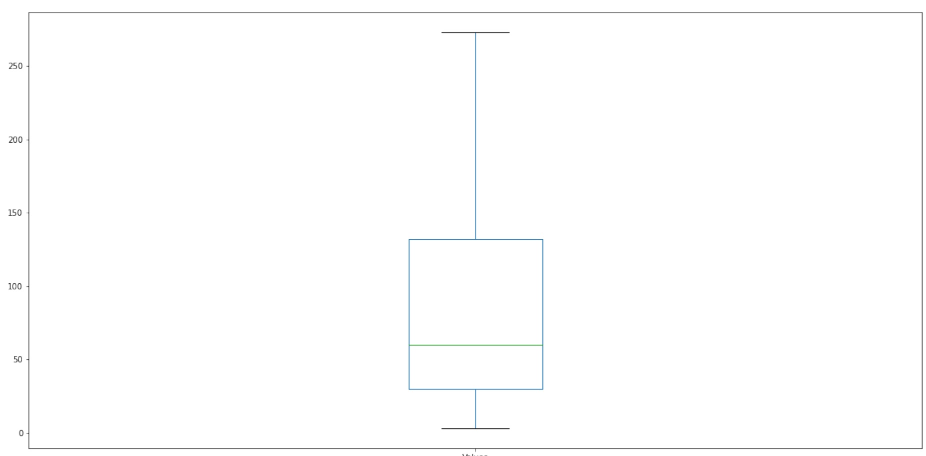

Figure 4.7: Plotting the second period of time

Finally, we obtain the values of the third and last period 4.9. We find more chaos within the standards found previously, the explanation for this event can be found in unknown facts for the research, with which it is suggested as a scope for this project to carry out the researchs only in the first two periods since we find a standard deviation for the last period a value that exceeds the margins of 200 as shown below.

Table 4.9: Third period

count	16.000000
mean	1038.437500
std	1845.686899
min	1.000000
25%	61.250000
50%	165.500000
75%	571.000000
max	5832.000000

- source: self elaboration.

Graphically 4.8, we find more than one outlier, leaving the continuity of the research at risk if this significant amount of data is erased. Below is the graph.

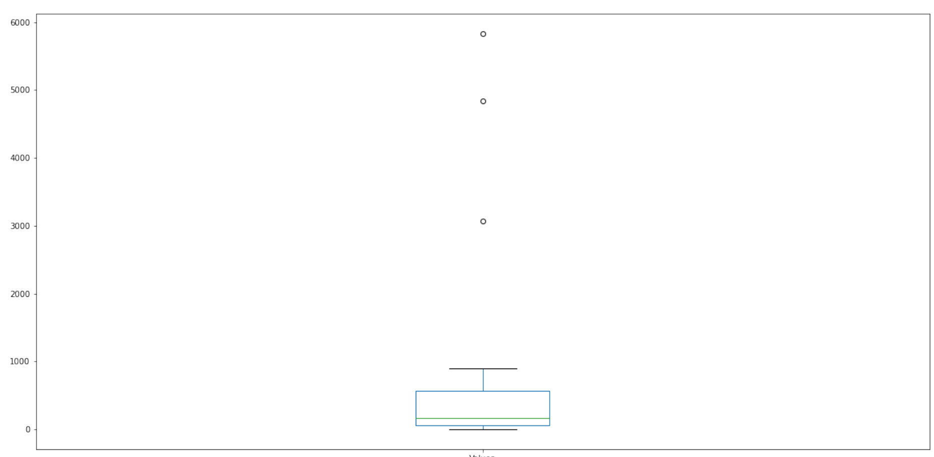

Figure 4.8: Plotting the third period of time

After period III, by eliminating the atypical values as evidenced in Table 4.10, and comparison with the previous value obtained from the same period but with atypical values, we find similar behavior in the standard deviation, to a lesser extent than the parent with the outliers but essentially the same shape. We corroborate this information with the distances between the quartiles, the minimum and the maximum value of terrorist attacks. Graphically 4.9, we find a significant distance between the quartiles of the minimum and the maximum value of terrorist attacks.

Table 4.10: Second period

count	13.000000
mean	220.769231
std	264.146473
min	1.000000
25%	26.000000
50%	106.000000
75%	424.000000
max	901.000000

- source: self elaboration.

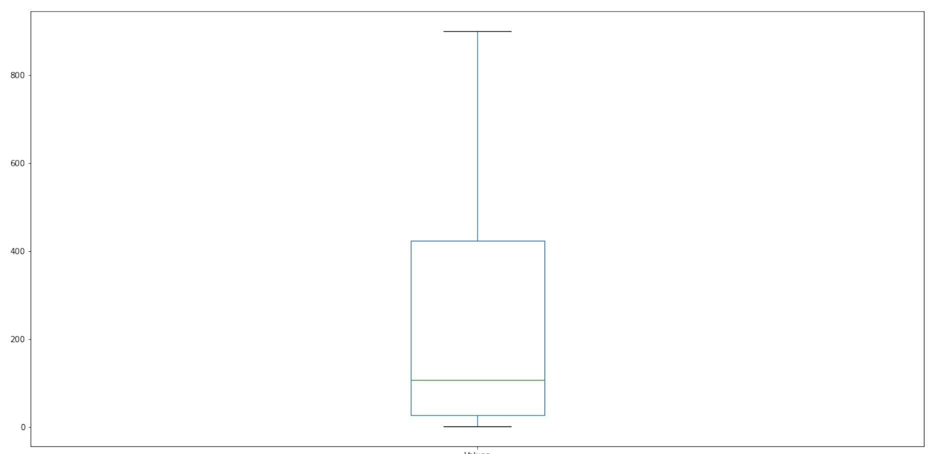

Figure 4.9: Plotting the second period of time

4.4 Exploratory data analysis

4.4.1 Introduction

Studying scientific articles is critical to determine the relationship between the data extracted from the GTD database. It allows through crucial factors such as year of events, way of operating at the beginning, and their last appearance. In this way, contribute to validating the information in the database in conjunction with scientific articles. We used Python programming language with libraries such as pandas for practical methods. Finally, this section of exploratory data analysis will end with a sample of five scientific articles that will verify the information.

4.4.2 Lashkar-e-Islam

Formed in 2004, but according to the data found in the GTD, the terrorist attacks of this group began in 2009. We consider a time gap not too far away to determine that the data is incorrect. With significant growth in 2012, see figure 4.10. The first attack, the last attack, and the article is supported by the information related to GTD. Thus, establish a relationship during the three periods, which takes as a standard factor bomb attack, as shown in figure 4.11. Also determining through the war that the terrorist groups Lashkar-e-Islam and Ansar-ul-Islam occupy [94]. that the armed assault was the second common factor in this terrorist group in period III.

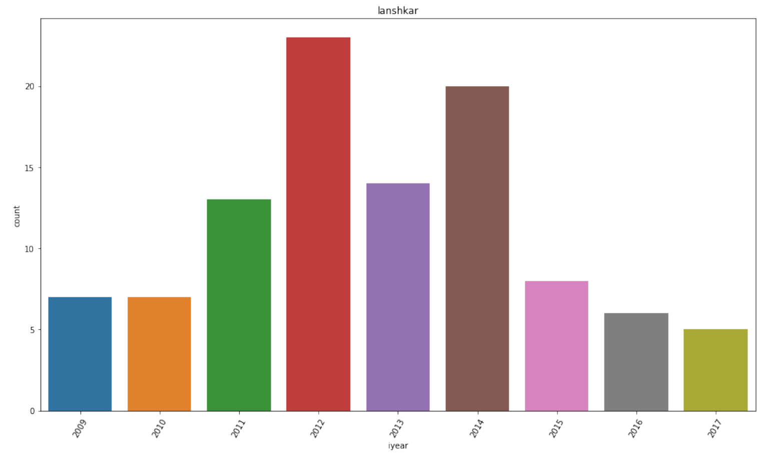

Figure 4.10: Year graph Lashkar-e-Islam period III GTD

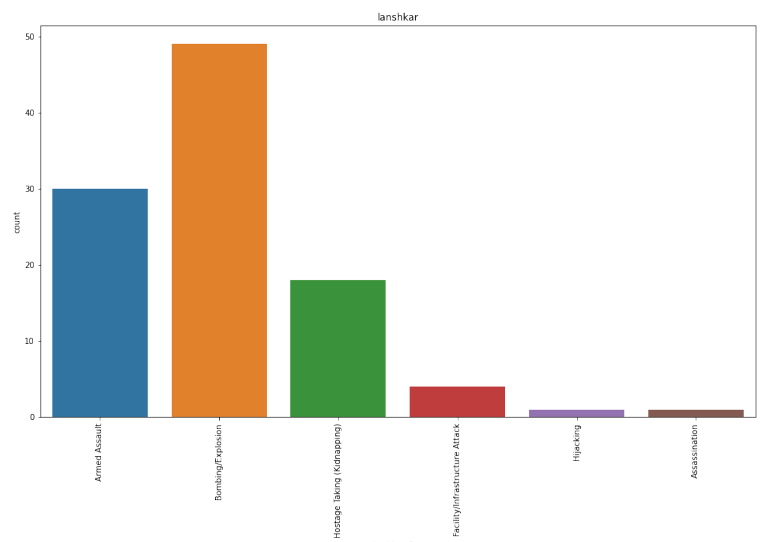

Figure 4.11: Attack type graph Lashkar-e-Islam period III GTD

4.4.3 Al-Qaida in the Islamic Maghreb (AQIM)

AQIM was formed in 1998 with the first attack in 2003; however, the first attack registered in the GTD dates from 2007 (figure 4.13). As indicated in the report [95], AQIM dates from suicide bomb attacks where it is registered in the second period, which It indicates mainly with suicide bombs, see figure 4.12. Finally, for the record, it is indicated that the terrorist group AQIM increased its attacks by engaging in extortion, data that is also registered in the GTD database—determining a similar behavior pattern for both the report and the data provided by the GTD.

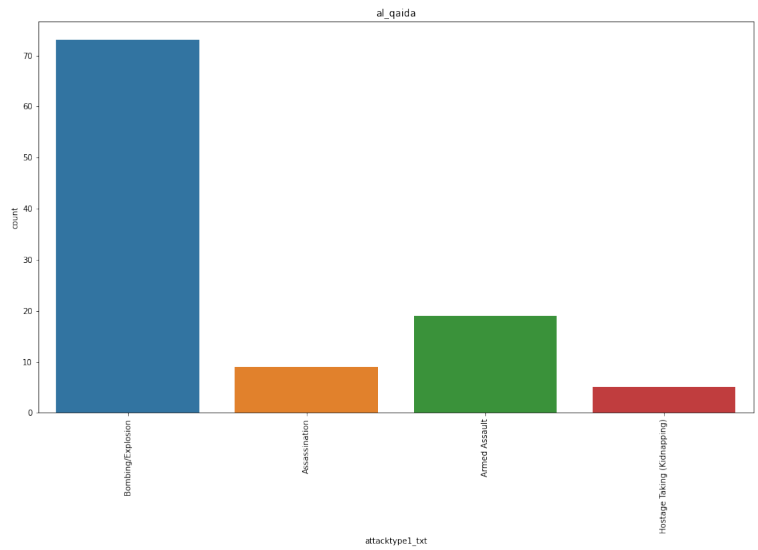

Figure 4.12: Attack type Al-Qaida in the Islamic Maghreb period II GTD

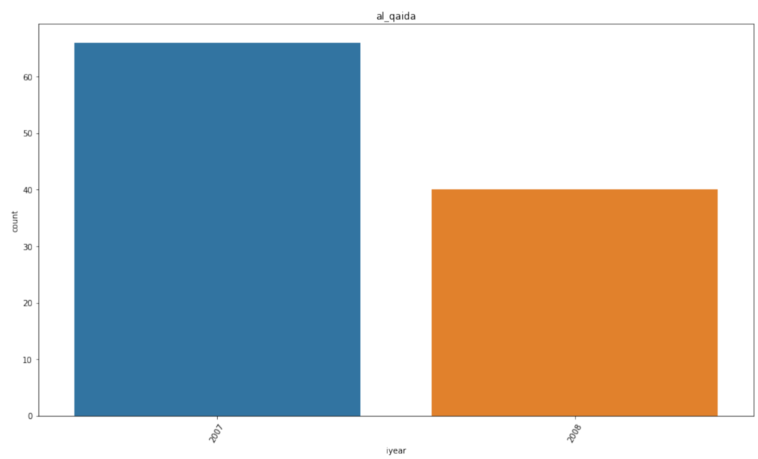

Figure 4.13: Year graph Al-Qaida in the Islamic Maghreb period II GTD

4.4.4 Al-Qaida

According to University of Stanford [96], Al-Qaida, founded in 1988, appeared for the first time on the world's radar in 2003. However, by analyzing the information made to the GTD, it was possible to locate the first attack in 1992. However, it can be noted in figure 4.14 that it is an isolated attack. As well as the attacks of 1998 (figure 4.14) and 2000(figure 4.15), since having two attacks each one, it is impossible to draw any conclusion. Until the second period, a curve is shown in the volume of attacks perpetrated by Al-Qaida as shown in figure 4.12 and figure 4.16. Going to 5 attacks in 2001, 14 attacks in a consolidated Al-Qaida in 2003. In this way, that is why we consider this point consistent for the research.

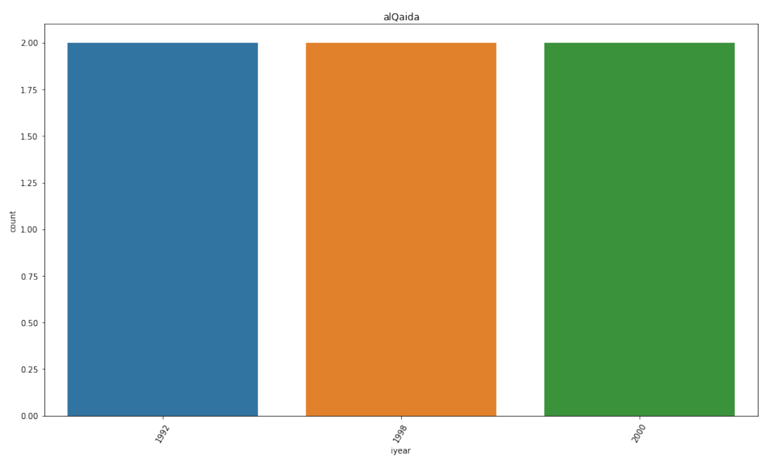

Figure 4.14: Year Al-Qaida graph period I GTD

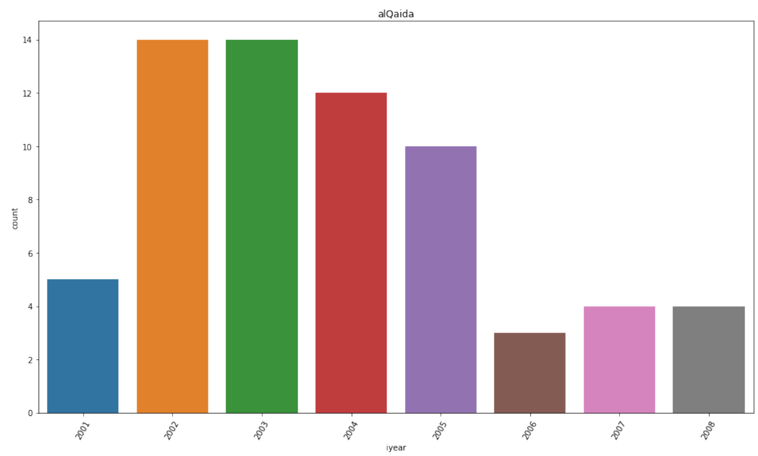

Figure 4.15: Year Al-Qaida graph period II GTD

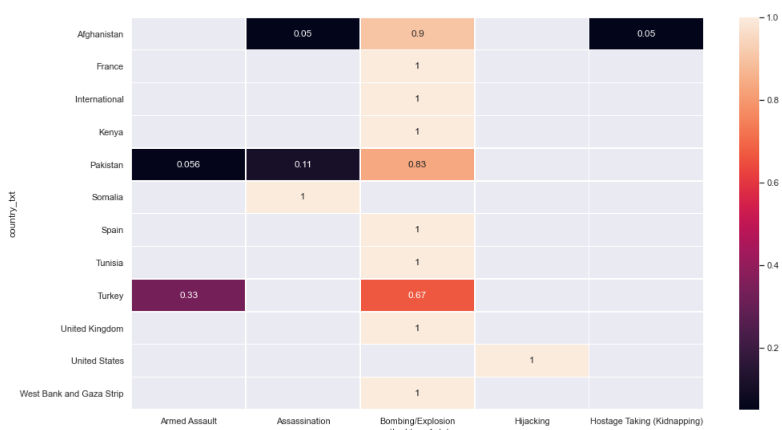

Figure 4.16: Heatmap Al-Qaida graph period II GTD

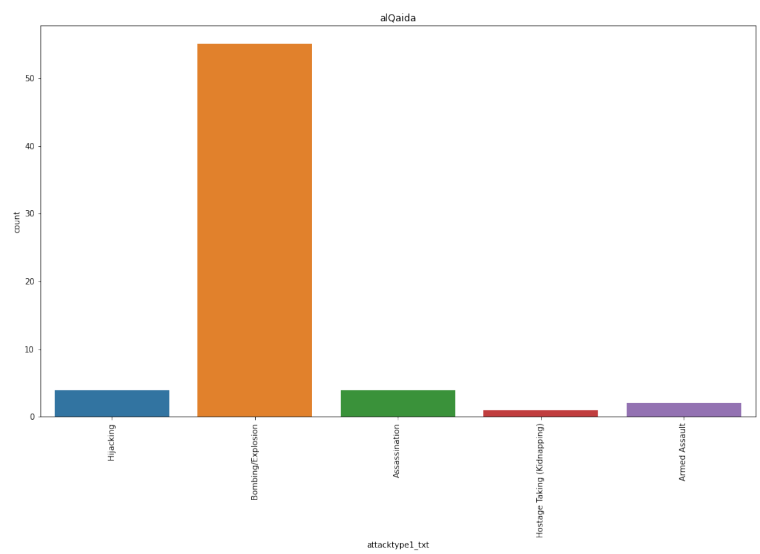

Figure 4.17: Attack type Al-Qaida graph period II GTD

4.4.5 Jemaah Islamiyah

In the article [97], we have the terrorist group Jemaah Islamiyah formed in the early 90s; however, the data shows that the first appearance of this terrorist group dates from the early 2000s. It is important to note that for this first period of this terrorist group and supported by the article, the main objectives of this terrorist group were the churches, supported by figure 4.16 that corresponds to period I. Finally, in the three periods, it is seen graphically (Figure 4.19, Figure 4.20 and Figure 4.21) as the terrorist group Jemaah Islamiyah uses bombs as its primary method, supported in the article that has the same argument.

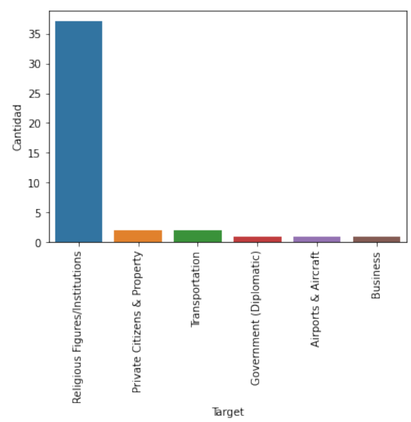

Figure 4.18: Target Jemaah Islamiyah graph period I GTD

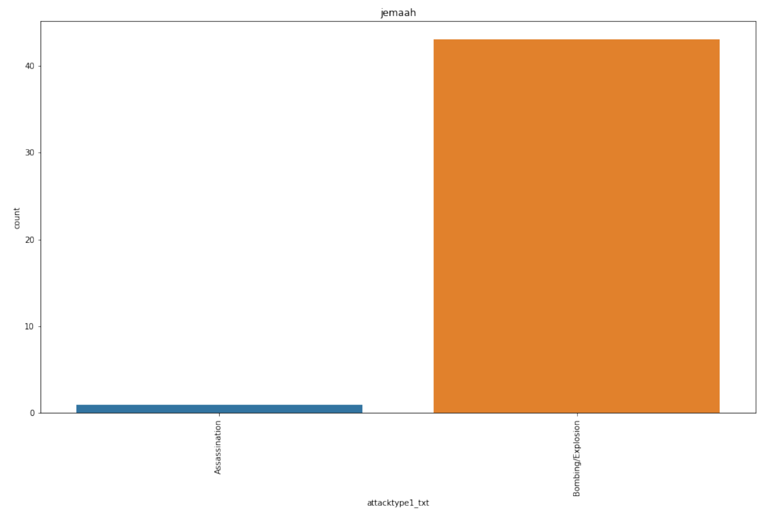

Figure 4.19: Attack type Jemaah Islamiyah graph period I GTD

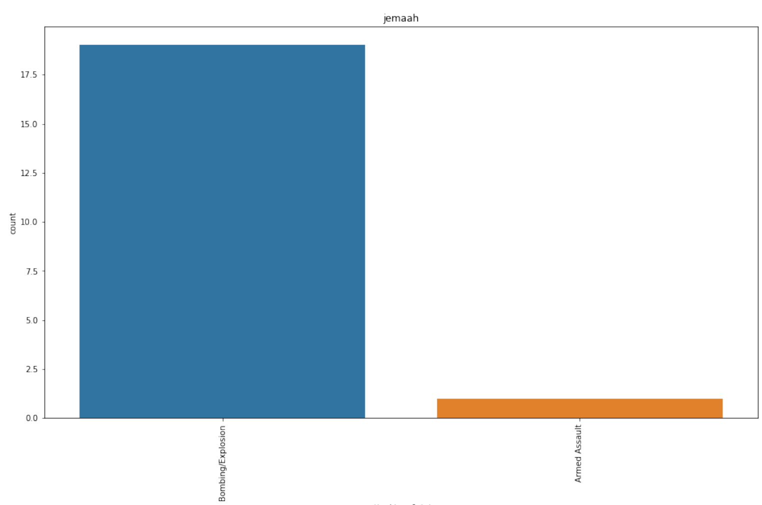

Figure 4.20: Attack type Jemaah Islamiyah graph period II GTD

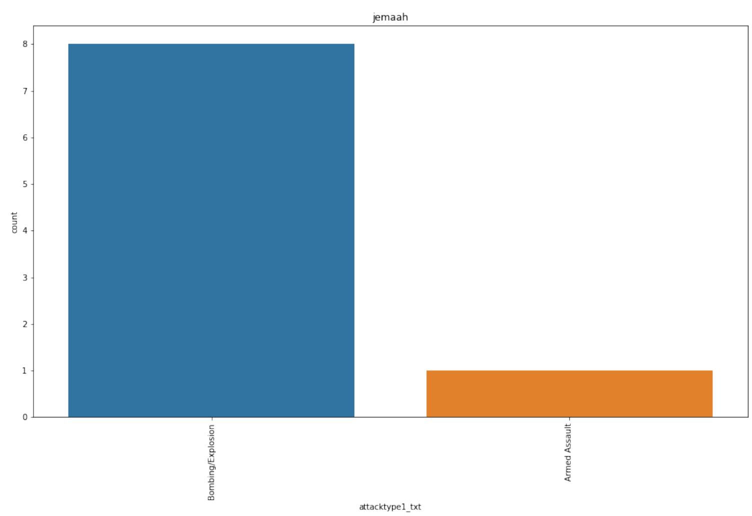

Figure 4.21: Attack type Jemaah Islamiyah graph period III GTD

4.5 Latent Class Analysis

R's poLCA library allows the calculation of different latent class models using a single command line. Next, the variables entered in the model to carry out the study based on terrorist events belonging to the GTD and the Islamic State will be explained.

4.5.1 Data Input

The variables are entered in the **poLCA** function. In turn, the response variables were coded as integer values. Each variable started with the value of one to indicate the first value of the response result and increased until reaching the maximum number of response variables. Given that the response values for each of the variables in the study started at zero or had a decimal value, the **poLCA** library would finish the calculation without estimating the model, and once it would return an error value.

4.5.2 Command Options

To specify the latent class models, the library uses the formula expression of the standard symbolic R model. The response variables will be the manifest variables of the model since latent class models can have multiple manifest variables. These variables must be linked or bound as **cbind**( ) , the names of the response variables that will be included within the model will be entered. For this study, the variables entered in this command line were entered as follows:

- **cbind**(country_txt, attacktype1_txt, targtype1_txt, gname, weaptype1_txt, iyear)

For this study, the basic model of latent classes without covariates was implemented. Therefore the command line for the response variables would look as follows:

- **cbind**(country_txt, attacktype1_txt, targtype1_txt, gname, weaptype1_txt, iyear)

To estimate the Latent Class Analysis model, it is necessary to use the command predetermined by the **poLCA** library, which contains the following characteristics: It is necessary first to assign the name you want to give to the model as follows:

- **Model1**$< -$ poLCA(Model$_Variables$, data=Info, nclass=2, graphs = FALSE, na.rm = TRUE, probs.start =NULL, nrep = 150)

where:

- **Model1**: Is the designated name for the first model.
- **Model_Variables**: Are the response variables bounded in which the model will be focused.
- **data = Info**: The dataset that contains the Islamic State events.
- **nclass=2**:This command will be the number of latent classes or clusters that the model will assume. The default value is two. For this study, nine different models were calculated, consisting of 2 to 10 classes.
- **graphs = FALSE**: Its default value is FALSE; if it is assigned as TRUE, the **poLCA** library graphically shows the division of the classes in a predetermined graph. It is designated as FALSE because otherwise, the estimation of the process will be slower.
- **na.rm = TRUE**: This command is for handling missing values in manifest response variables. If the value is TRUE, the missing values are deleted before estimating the model. Otherwise, if it is FALSE, the missing values are retained within the model. Typically in **poLCA** you always work under full value parameters, so the default value is TRUE.
- **probs.start =NULL**: This command provides the ability to add a list of class conditional-response probability matrices. The model uses these values as initial values for the EM estimation algorithm. The default value is NULL, which means the initial values are generated randomly.
- **nrep = 150**: The number of times the model will be repeated to obtain the model with the maximum log-likelihood. Usually, this command is predetermined as one, but for best results and consistency in the data for this study, each model was replicated 150 times.

Table 4.11: Period 1 Best Model Selection

Model	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)
Model 1	16307.75	17390.28
Model 2	15395.93	17022.13
Model 3	**14712.28**	**16882.15**
Model 4	14366.02	17079.57
Model 5	14209.26	17466.48
Model 6	14109.24	17910.12
Model 7	14042.56	18387.12
Model 8	14090.16	18978.38
Model 9	14002.11	19434

- source: self elaboration.

Table 4.12: Period 2 Best Model Selection

Model	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)
Model 1	33691.49	34613.46
Model 2	32176.95	33562.66
Model 3	31065.6	32915.06
Model 4	30046.73	32359.94
Model 5	29570.03	32346.98
Model 6	**29073.33**	**32314.02**
Model 7	28769.53	32473.97
Model 8	28554.29	32722.47
Model 9	28372.39	33004.32

- source: self elaboration.

The selection of the best model obtained in LCA through the poLCA library was based on the results obtained from Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Where:

$$AIC \ = \ -2\Lambda + 2\Phi \tag{4.1}$$

$$BIC \ = \ -2\Lambda + \Phi \ln N \tag{4.2}$$

- Λ represents the maximum log-likelihood of the model
- Φ represents the total number of estimated parameters.
- **N** represents the sample size.

The selection of the optimal models is given by the minimum values of the AIC and the BIC. This study considers the minimum values obtained in the BIC results because it provides more parsimonious values according to [17].

4.6 K-Modes

Fuzzy K-modes is a transaction model for classifying unsupervised categorical data. Information within the framework of this research is of two types; first, data of categorical types using six variables for the development of the research, in which we can find data such as the name of terrorist groups, types of attacks, objectives, etc. Second, we found the need to use an unsupervised method for classifying the database records and determining the need to use LCA and K-modes as grouping methods, complementing the use of LCA for the project by defining the optimal clusters as concluded in the previous section.

4.6.1 Input

Determining that for K-Modes, it is necessary to use the length of the data in the table, finding in each column the utility found in the categorical data. Using a sequence of steps, first, the algorithm extracts observation data or pivot data; for the first period, four pivots are extracted, which is determined by the number of clusters. The clusters, in turn, are determined in the previous section, finding four clusters for the first period and seven clusters for the second period. They are finding four pivots for the first period and seven for the second period. In the same way, the pivots are made up of each of the variables that database has; thus, for the first period, there are four samples with six variables each, and for the case of the second period, there are seven samples with six variables each. Then the algorithm needs to do a tour and find what values have similarities, to say comparison is made with the entire database in conjunction with the pivot data already mentioned above. Finding a similarity adds a value of one, and if they are not equal, they do not add values. Finally, with the values of similarity found, the values closest to the centroids must be determined, looking in each record for the minimum value of the resulting values and in this way classifying the information through the selected clusters.

4.6.2 Deploy

Through python[25] and its K-Modes library [26] that uses the K-Modes cauterization method, it is possible to implement it in conjunction with pandas. Thus through the data already mentioned. For correct execution of this library, the number of clusters and the method to be cauterized must be selected.

4.7 Results

4.7.1 Period 1

For the development of the results and consequently with Bayesian Information Criterion, through table 4.11, which yields results of 16882.15, model 3 is selected, that is, 3 clusters for the deployment of period one. Composed of values dating from 1986 to 2000, it is essential to highlight the similar behavior denoted in conglomerates one and three. Concluding with a notorious activity in models two and four, however, it should be clarified that under the standards set out in section 4.5.2, the model to be selected is selected thanks to the lowest value that we can determine.

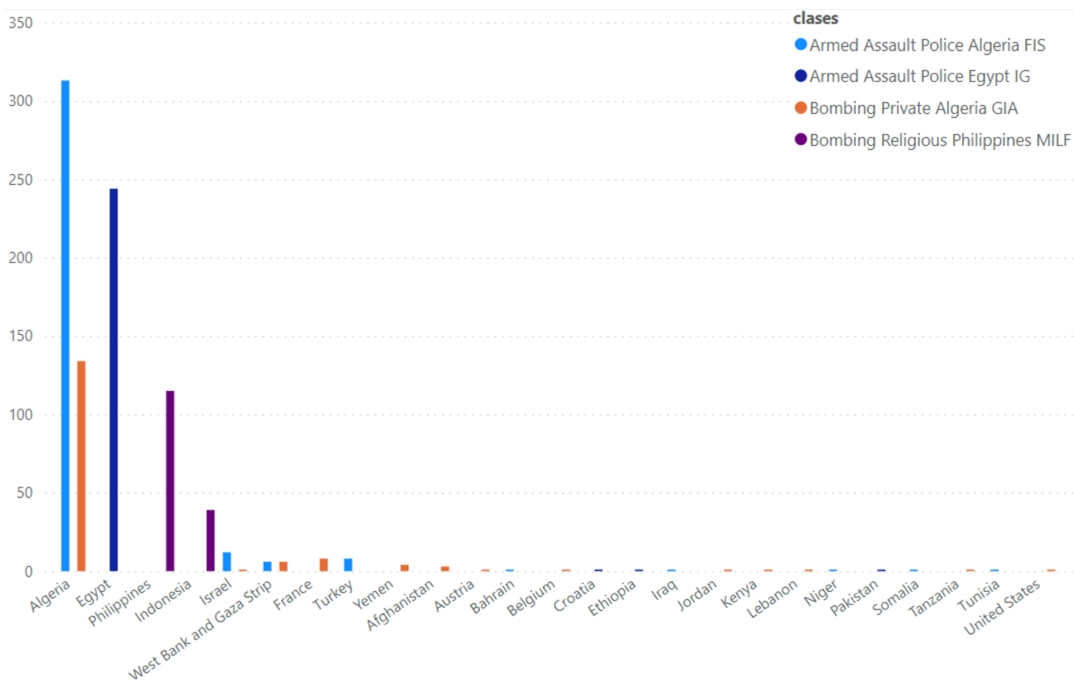

Figure 4.22: Countries LCA Period 1

The most affected countries by terrorist attacks during the first period were: Algeria, Egypt, the Philippines, and Indonesia,according to the definitions presented in Section 4.1, with: 447, 244, 115, and 39 attacks, respectively.

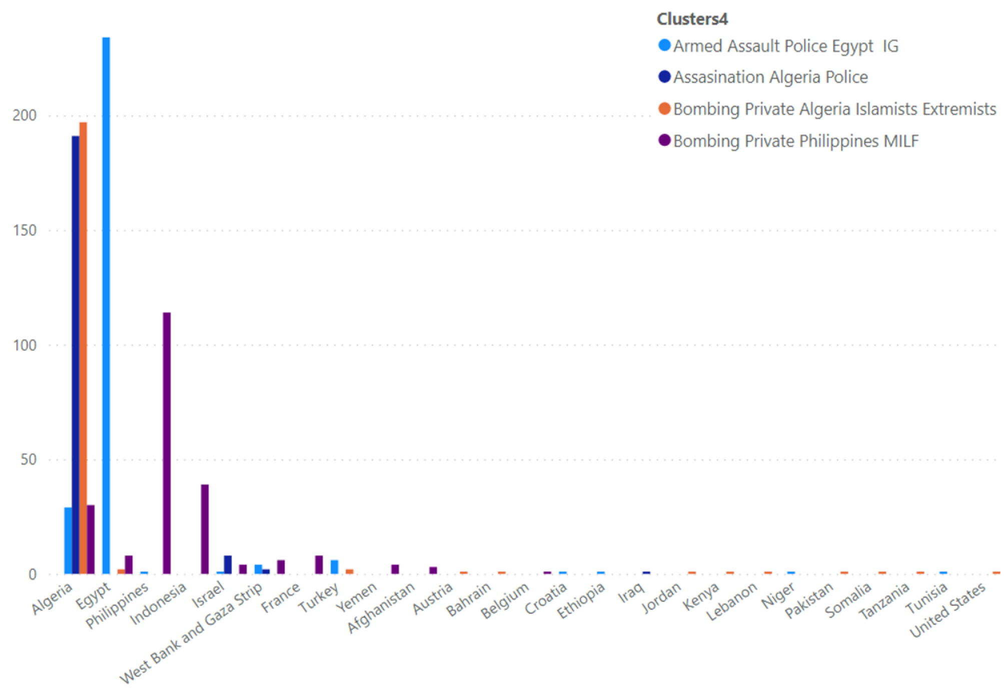

Figure 4.23: Countries K-modes Period 1

For period one and the cluster "Armed Assault Police Egypt IG," terrorist attack behaviors are detected, as shown in figure 4.23. There is a considerable number, and the largest of all is located in Egypt. There is also an agglomeration of terrorist attacks in Algeria. Finally, there is evidence of many terrorist attacks in different countries, finding that there has been a presence of terrorist attacks in 25 countries of the world. The same behavior determined by terrorist attacks in Algeria, Egypt, the Philippines, Indonesia, and Israel is evidenced with the LCA results.

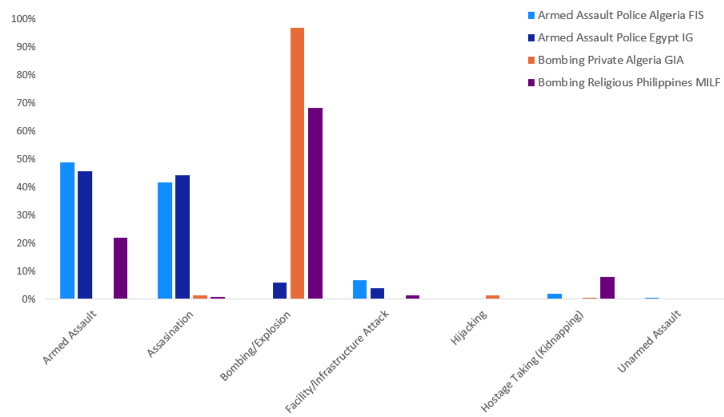

Figure 4.24: Attack Type LCA Period 1

The most frequent terrorist attack types during the first period were: Armed Assault, which represented a 48.84 % for the ”Armed Assault Police FIS” and 45.75% for ”Armed Assault Police Egypt IG” Bombing/Explosion represented the 96.93% for ”Bombing Private Algeria GIA”, and 68.18% for ”Bombing Religious Philippines MILF” respectively.

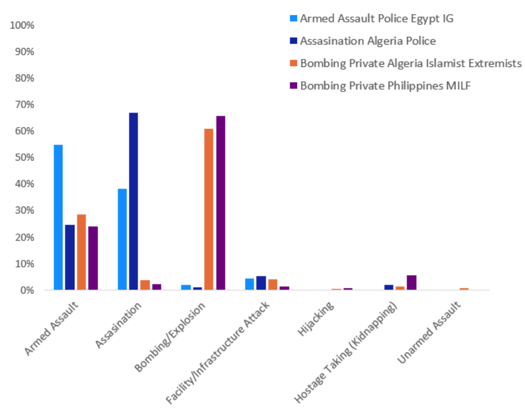

Figure 4.25: Attack Type K-modes Period 1

The cluster ”Armed Assault Police Egypt IG,” as evidenced in figure 4.25, there are two essential peaks for the cluster ”Armed Assault Police Egypt IG,” in armed assault and murder. The cluster ”Assassination's Algeria Police” has greater participation of attacks in murders. Finally, for the cluster ”Bombing private Algeria Islamist Extremist” and ”Bombing Private Philippines MILF” there is evidence of much more participation in handling explosive bombs. It is also possible to demonstrate the use of different techniques in terrorist attacks, such as hostage-taking, which is a common factor in all clusters but to a lesser extent, as is also the case of attacks on infrastructure where a similar amount in terrorist attacks. Finally, there are unique attacks, such as the case of extortion for the ”Bombing Private Philippines MILF” cluster.

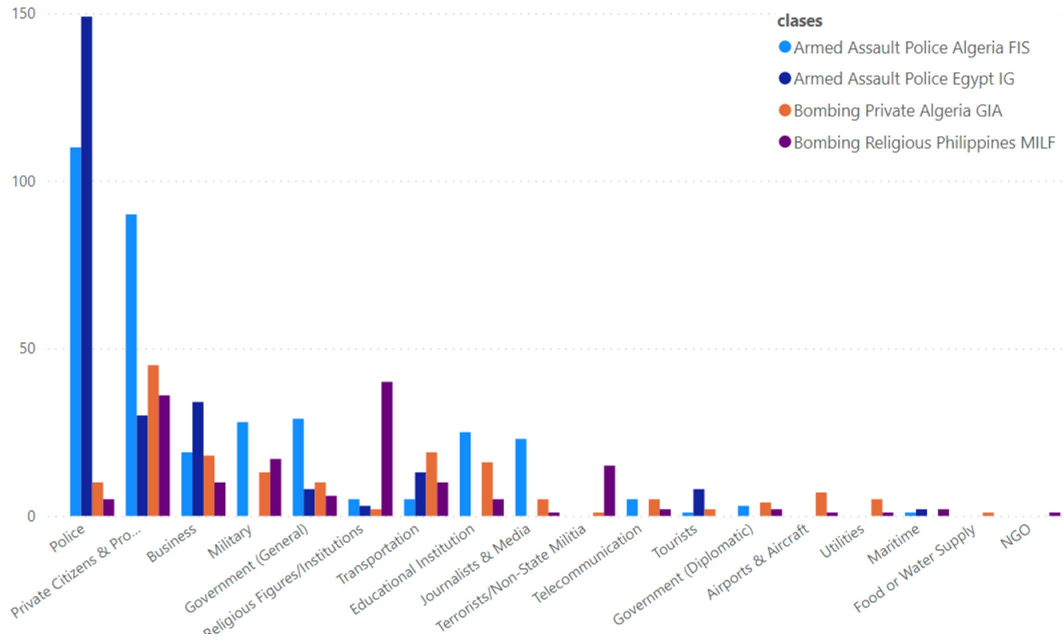

Figure 4.26: Targets LCA Period 1

The most affected institutions by terrorist attacks during the first period were: Police, Private Citizens & Property, and Business, according to the definitions presented in Section 4.1, with a total of attacks of:274, 201, and 81, respectively.

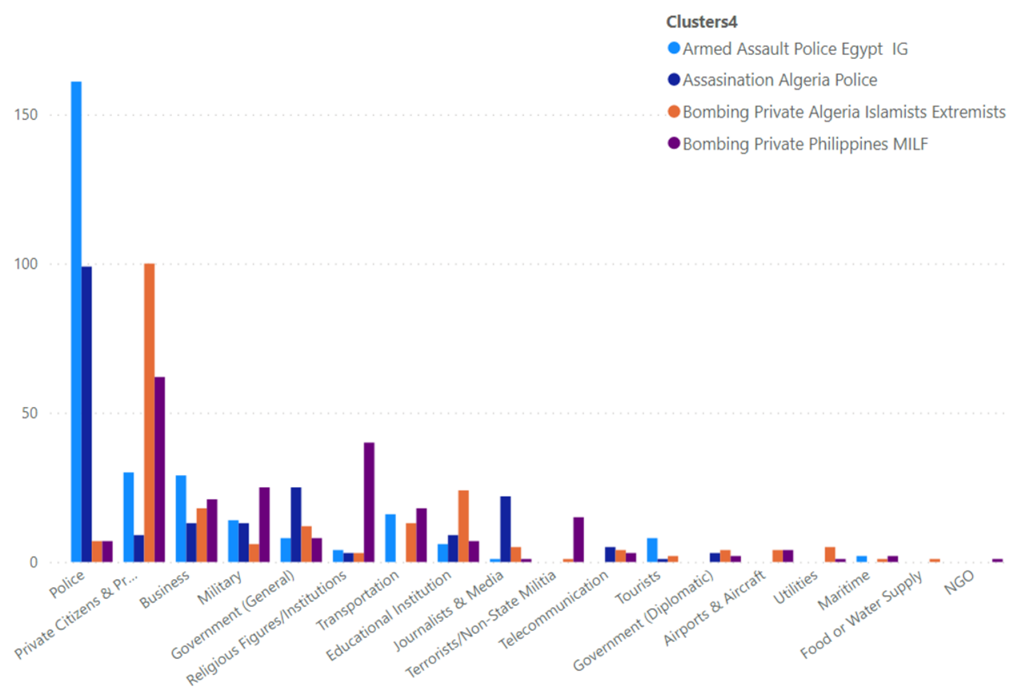

Figure 4.27: Targets K-modes Period 1

For the objectives, a greater quantity is evidenced than the others in attacks on policemen for the cluster ”Armed Assault Police Egypt IG,” as seen in figure 4.27. Also, for the cluster ”Assassination's Algeria Police,” the main objective of this cluster is the police, finding a large number of terrorist attacks focused on the police. However, the cluster ”Bombing private Algeria Islamist Extremist” and ”Bombing Private Philippines MILF”, focus more on attacks on private citizens and property, finding a close relationship between attacks with an armed assault on police and bombs on private citizens and property.

Table 4.13: Period 1 Weapon Types LCA

	Armed Assault Police Algeria FIS	Armed Assault Police Egypt IG	Bombing Private Algeria GIA	Bombing Religious Philippines MILF
Explosives	0.29%	6.88%	98.77%	68.18%
Firearms	70.64%	91.90%	0.61%	31.17%
Incendiary	6.69%	0%	0.61%	0.65%
Melee	21.80%	1.21%	0%	0%
Other	0.29%	0%	0%	0%
Sabotage Equipment	0%	0%	0%	0%

- source: self elaboration.

The most used weapon types according to the definitions presented in Section 4.1, in terrorist events during the first period were: Fire Arms representing the 70.64% for the weapons used by the cluster ”Armed Assault Police Egypt” , Explosives representing the 98.77% ”Bombing Private Algeria GIA”, and 68.18% for the cluster ”Bombing Religious Philippines MILF for the weapons used in the attacks. Melee and the other weapon types are not so relevant for the events in the first period.

Table 4.14: Period 1 Weapon Types K-Modes

	Armed Assault Police Egypt IG	Assasination Algeria Police	Bombing Private Algeria Islamist Extremists	Bombing Private Philippines MILF
Explosives	2.87%	0.50%	61.90%	66.82%
Firearms	93.55%	85.15%	14.76%	25.35%
Incendiary	0%	5.45%	4.76%	1.84%
Melee	3.23%	8.24%	18.57	5.99%
Other	0.36%	0%	0%	0%
Sabotage Equipment	0%	0.50%	0%	0%

- source: self elaboration.

Through the previous images (figure 4.27 and table 4.14), the behavior of the types of weapons is expected to be firearms in the case of the clusters "Armed Assault Police Egypt IG" and "Assassination's Algeria Police." The case of " Bombing private Algeria Islamist Extremist "and" Bombing Private Philippines MILF "use explosive-type weapons, corroborating the information in figure **??**.

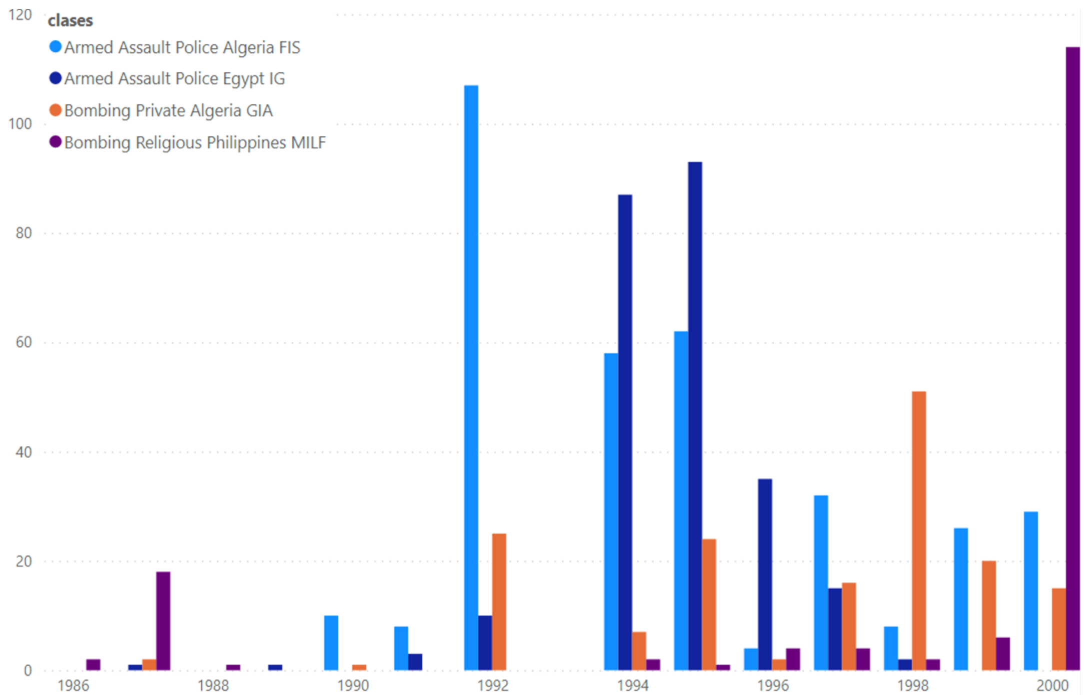

Figure 4.28: Year Events LCA Period 1

The most affected years by terrorist attacks during the first period were: 1995, 2000, 1994, and 1992, with a total number of attacks of: 180, 158, 154, and 142, respectively.

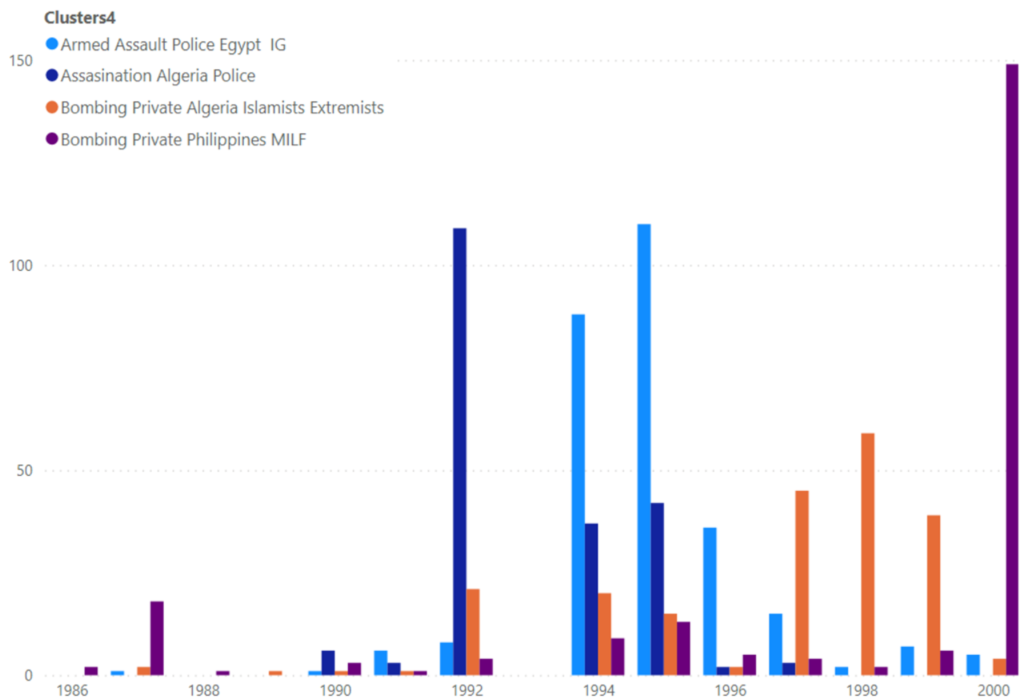

Figure 4.29: Year Events K-modes Period 1

An increase in terrorist attacks has been observed over the years, as shown in figure 4.25. Determining the year with the most attacks in 2000, which is part of the "Bombing Private Philippines MILF" cluster. However, it found a second peak for the cluster "Assassination's Algeria Police" in 1992. Already for the clusters "Armed Assault Police Egypt IG" and "Bombing private Algeria Islamist Extremist" are clusters deployed in the years since 1994 to 1999.

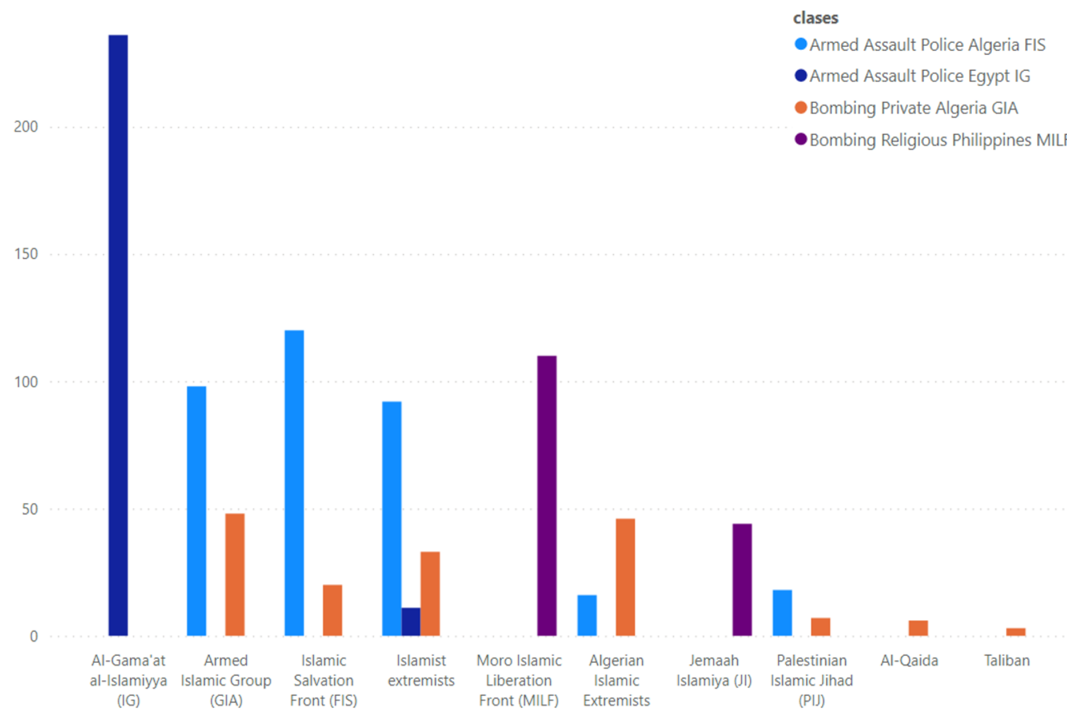

Figure 4.30: Terrorist Groups LCA Period 1

The terrorist groups with the highest rate of terrorist attacks during the first period were: Al-Gama'at al-Islamiyya (IG), Armed Islamic Group (GIA), Islamic Salvation Front (FIS), Islamist extremists, and Moro Islamic Liberation Front (MILF), with a total of attacks of: 236, 146, 140, 136 and 110 respectively.

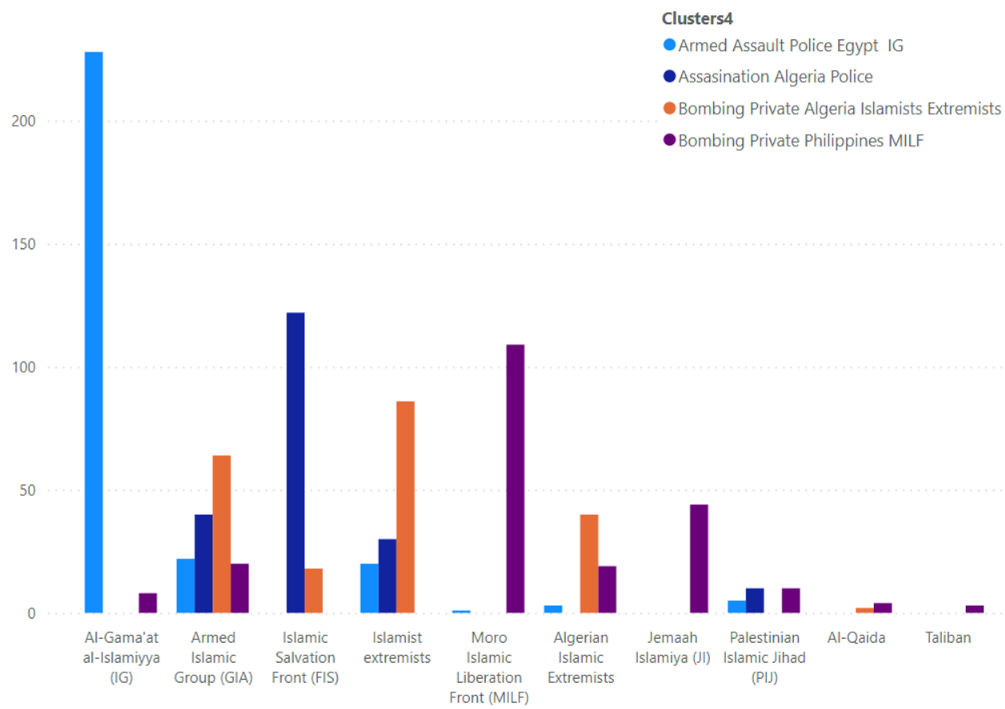

Figure 4.31: Terrorist Groups K-modes Period 1

In the terrorist groups, as seen in figure 4.31, it is observed that Al-gama'at al-Islamiyya (IG) is in almost the entire "Armed Assault Police Egypt IG" cluster. It is also seen that the groups Armed Islamic Group, Islamist extremist and Algerian Islamic Extremist that are located in the cluster "Bombing private Algeria Islamist Extremist." Islamic Salvation Front is in the cluster "Assassination's Algeria Police," with almost all its extension in this cluster.

4.7.2 Period 2

According to the BIC criteria, Model 6 was selected from table 4.12. The following results were obtained through a cross-table analysis for the second period, which comprises the years 2001-2008, according to the parameters defined in Section 4.5.2. According to the behaviors, it is evident that clusters 1 and 6 are the most active, followed by clusters 2 and 3 and finally clusters 4, 5, and 7.

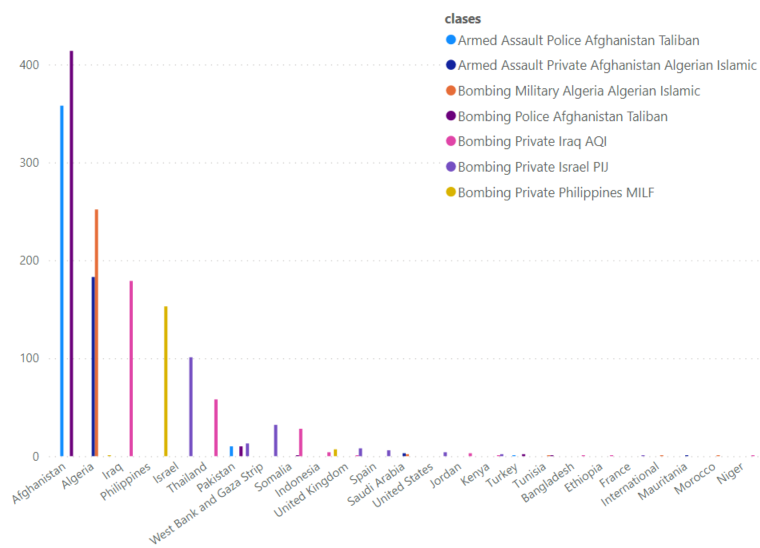

Figure 4.32: Countries LCA Period 2

The most affected countries by terrorist attacks during the second period were: Afghanistan, Algeria, Iraq, the Philippines, and Israel, according to the definitions presented in Section 4.1, with a total number of attacks of 772, 436, 179, 153, and 101, respectively.

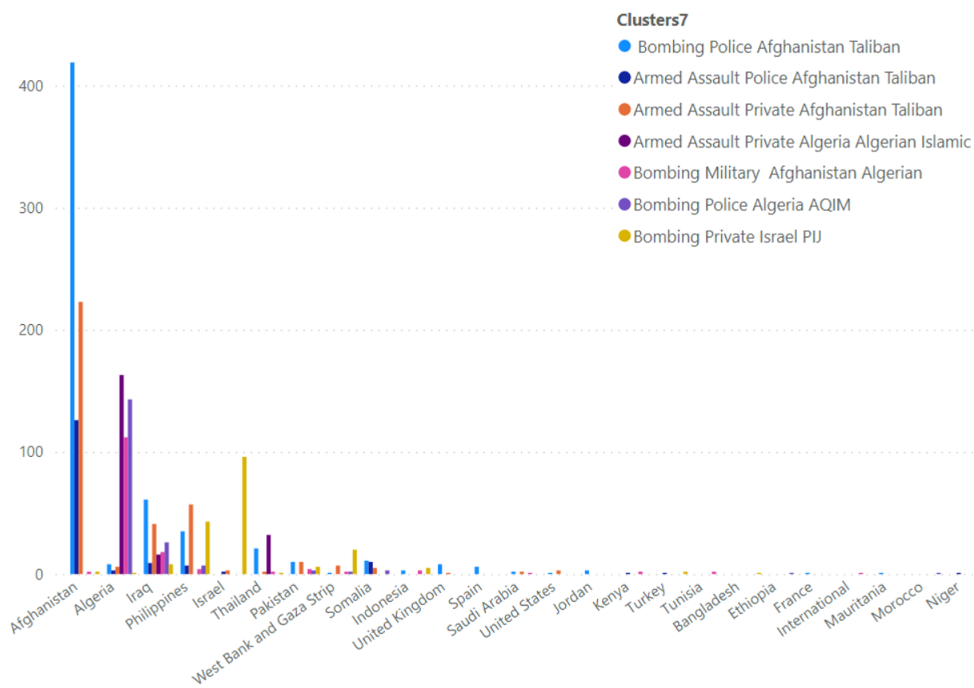

Figure 4.33: Countries K-modes Period 2

For period two, among the 7 clusters determined for the research, there is an equal number of terrorist attacks in different countries compared to period one. For this period, figure 4.33 shows how the cluster “Bombing police Afghanistan Taliban” has the majority of terrorist attacks exceeding 400 attacks. It is also possible to verify how, following this cluster, we can determine that the second cluster with the most terrorist attacks is ”Armed assault private Afghanistan Taliban.” It is also essential to refer to the image to determine that the countries most affected by terrorist attacks are Afghanistan, Algeria, Iraq, the Philippines, and Israel.

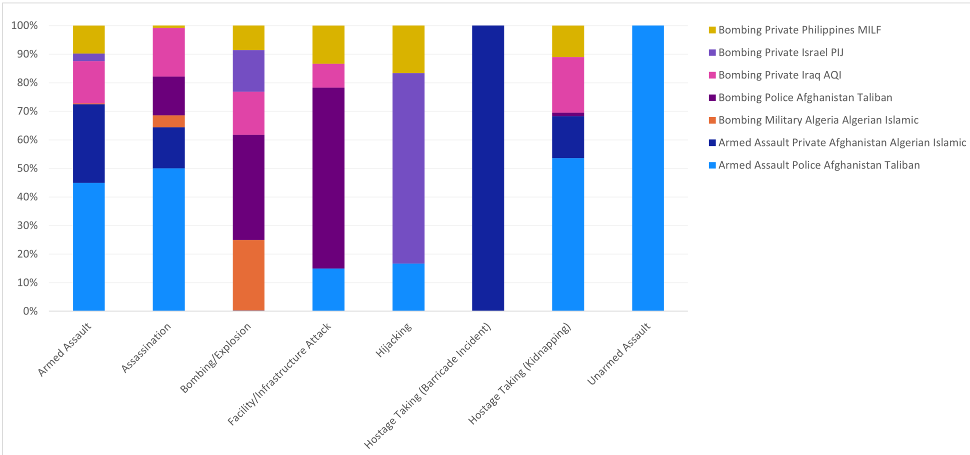

Figure 4.34: Attack Type LCA Period 2

The most frequent terrorist attack types during the second period were: Bombs / Explosives attack type, events increased from 278 to 1014, for this Period 2 the 36.79% of the events belong to ”Bombing Police Afghanistan Taliban”. Armed Assault attack types increased from 315 to 563, by comparing it with Period 1, for this Period 2 the 44.94% belong to ”Armed Assault Police Afghanistan Taliban. Assassinations decreased from 255 to 118 for the second Period, ”Armed Assault Police Afghanistan Taliban” has the 50% of the total amount of this attack type events.

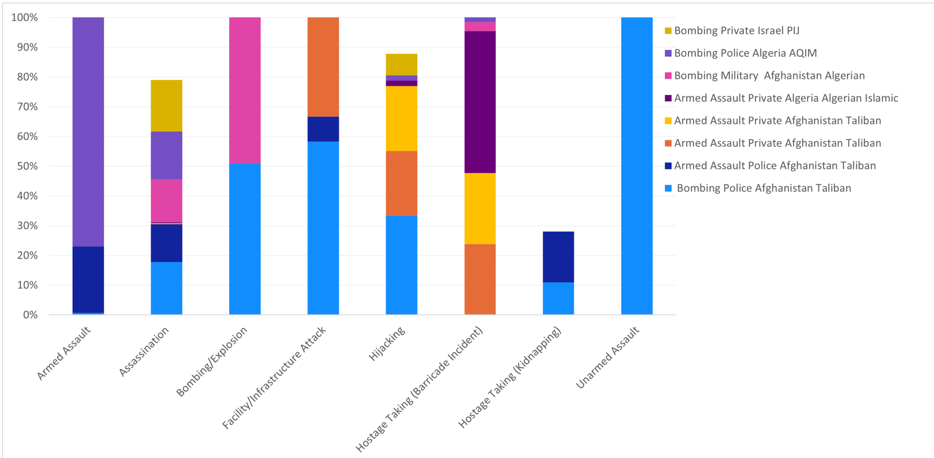

Figure 4.35: Attack Type K-modes Period 2

For the types of attack, it is observed in figure 4.35, a greater dominance in the attacks with explosives for the cluster of "Bombing police Afghanistan Taliban" with more than 500 terrorist attacks, as seen in the figure. It is also determined that the second most significant use of an attack-type is armed assault, especially for the cluster "Armed assault private Afghanistan Taliban." They also have isolated cases of some terrorist attacks, such as unarmed attacks and hostage-taking, observed in the clusters "Bombing police Afghanistan Taliban" and "Armed assault private Afghanistan Taliban," respectively. Finally, it is crucial to clarify how the use of explosives and armed assault is the predominant type of attack in almost all the research clusters.

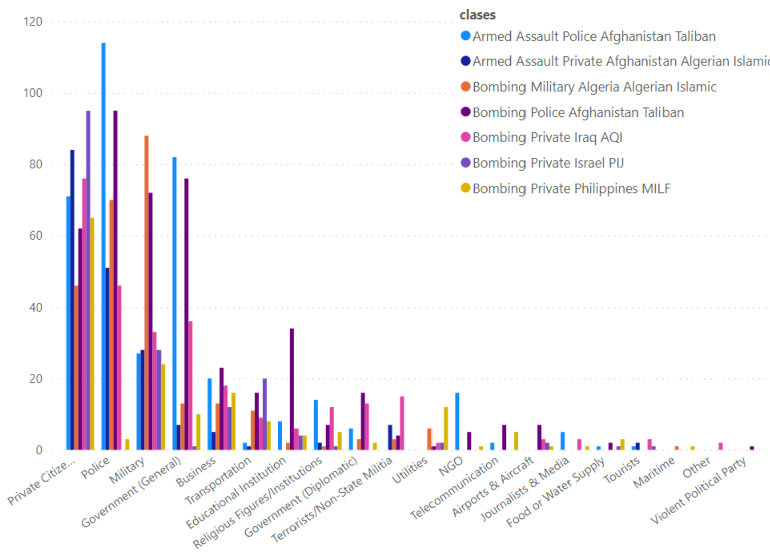

Figure 4.36: Targets LCA Period 2

The most affected institutions by terrorist attacks during the first period were: Private Citizens & Property, Police, Military, Government (General) and Business, according to the definitions presented in Section 4.1, with a total of attacks of: 499, 379, 300, 225 and 107 respectively.

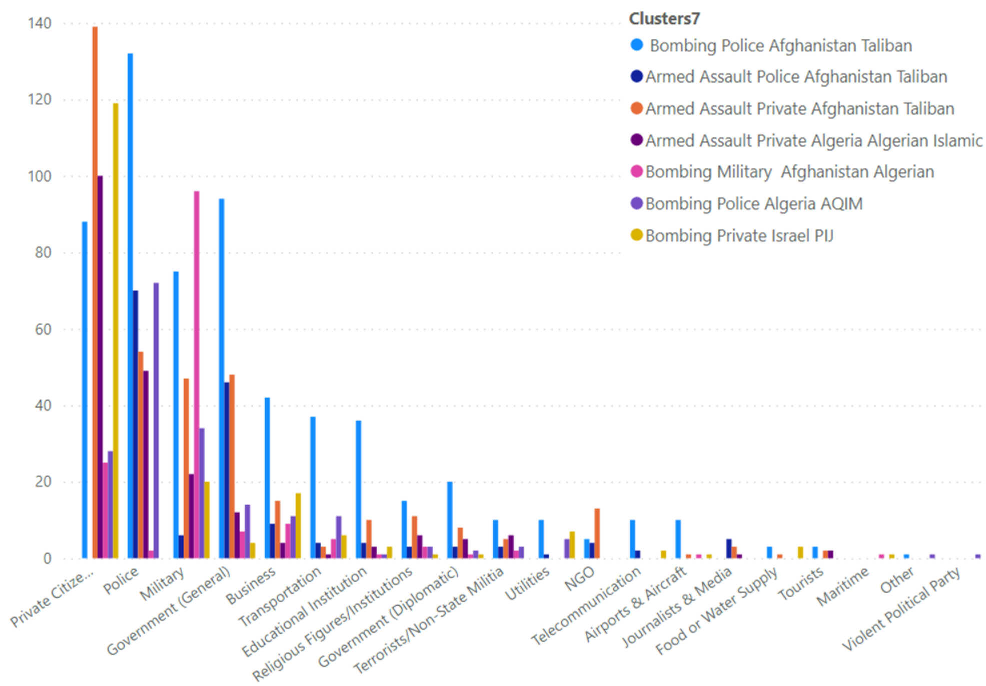

Figure 4.37: Targets K-modes Period 2

Compared with period one, it is observed in figure 4.37 how attacks on private citizens and property are the factor by which most clusters decide to perpetrate terrorist attacks. However, a greater use of various attacks is seen compared to period one. They are implementing large amounts of attacks for private citizens and property, Police, Military, and government in general.

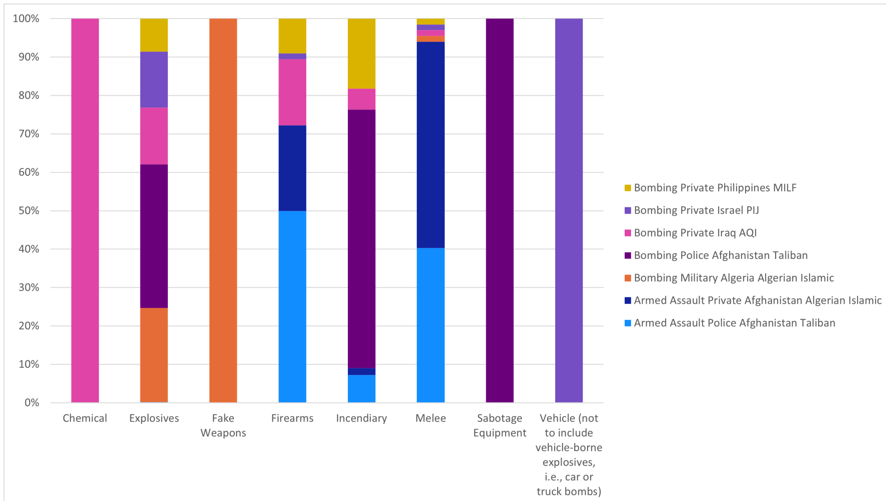

Figure 4.38: Weapons LCA Period 2

The most used weapon types in terrorist attacks during the first period were: Explosives with a total of 1042 events, where the 37.43% belongs to ”Bombing Police Afghanistan Taliban” and 24.47% belongs to ”Bombing Military Algerian Islamic. Fire Arms with a total of 673 events where the 49.93% belongs to ”Armed Assault Police Afghanistan Taliban” and the 22.29% belongs to ” Armed Assault Algerian Islamic. Another Relevant results for ”Bombing Military Afghanistan Algerian Islamic is that the 100% of the weapon type Fake Weapons. The group ”Bombing Police Afghanistan Taliban has also the 100% in using Sabotage Equipment.

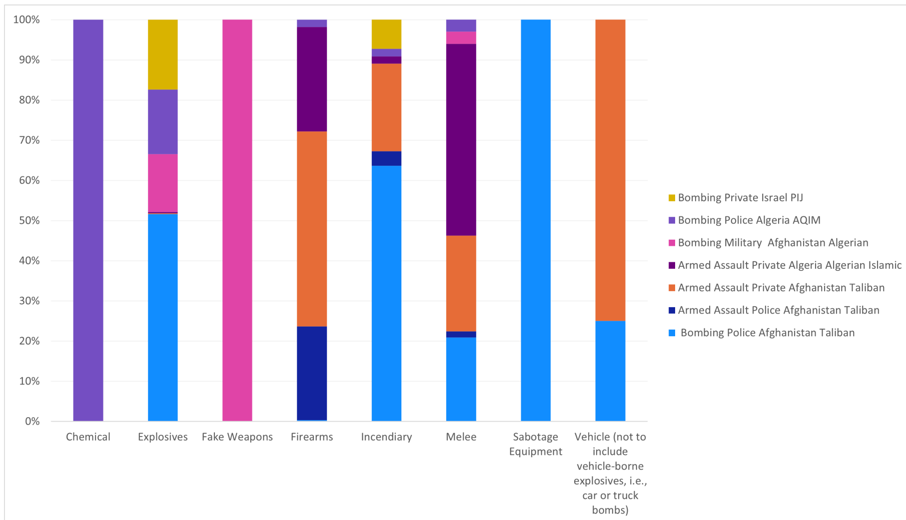

Figure 4.39: Weapons K-modes Period 2

Implementing various attack targets also increases the use of weapons, as can be seen in figure 4.39, where the use of different types of weapons such as chemical weapons is observed in the case of the ”Bombing police Algeria AQIM” cluster or even the use of false weapons in the case of the cluster ”Bombing military Afghanistan Algerian.”

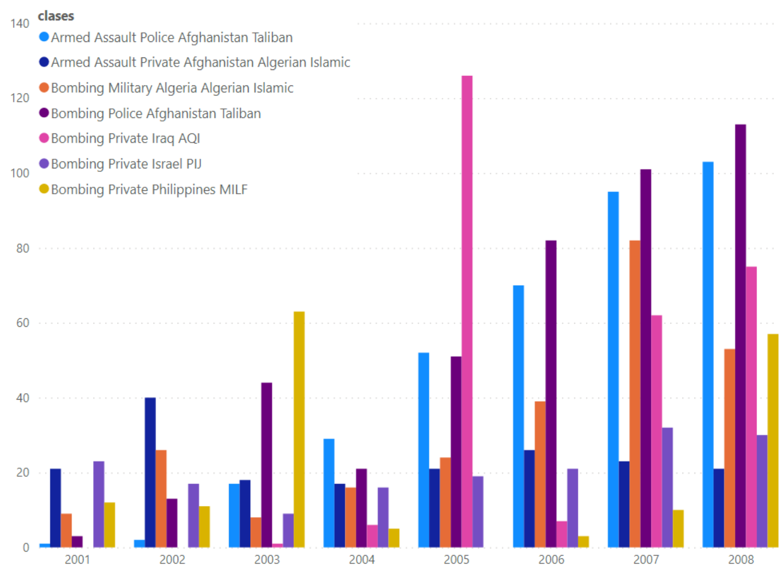

Figure 4.40: Year Events LCA Period 2

During the second period, the most affected years by terrorist attacks were: 2008, 2007, 2005, 2006, and 2003, with a total number of attacks of: 452, 405, 293, 248, and 160 respectively.

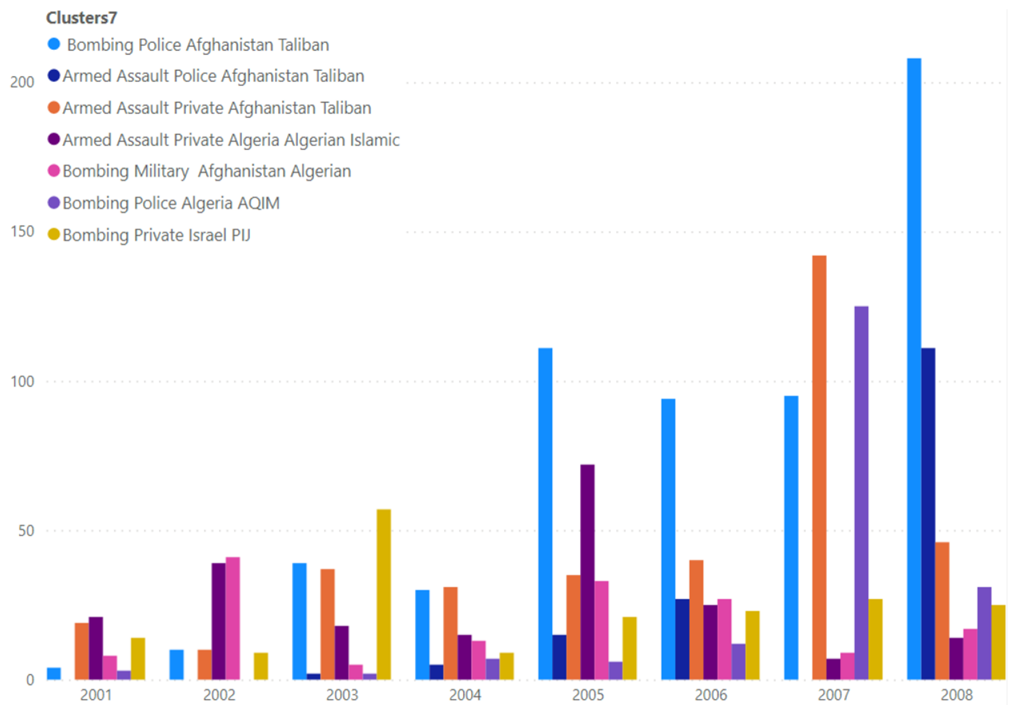

Figure 4.41: Year Events K-modes Period 2

For the years, figure 4.41 shows an increase in terrorist attacks. In the case of "Bombing Police Afghanistan Taliban," for 2001 it had less than 25 terrorist attacks, and for 2008 it had more than 200 terrorist attacks. This case has a similar pattern for the rest of the years. It is observing a higher concentration of terrorist attacks for 2008.

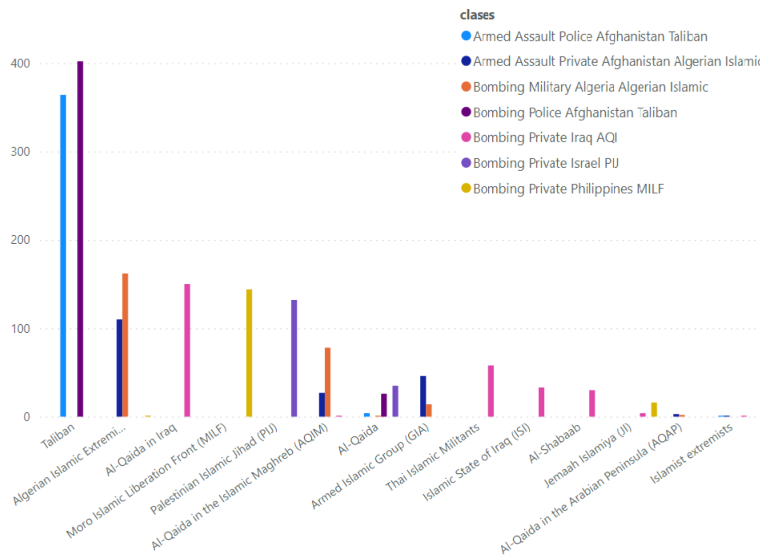

Figure 4.42: Terrorist Groups LCA Period 2

The terrorist groups with the highest rate of terrorist attacks during the second period were: Taliban, Algerian Islamic Extremists, Al-Qaida in Iraq, Moro Islamic Liberation Front (MILF), and Palestinian Islamic Jihad (PIJ), with a total number of attacks of: 766, 273, 150, 144 and 132 respectively. As it is shown groups such as Moro Islamic Liberation Front MILF changed their Target Types from Religious shown in the figure 4.26 to Private Citizens as it is shown in the figure 4.36

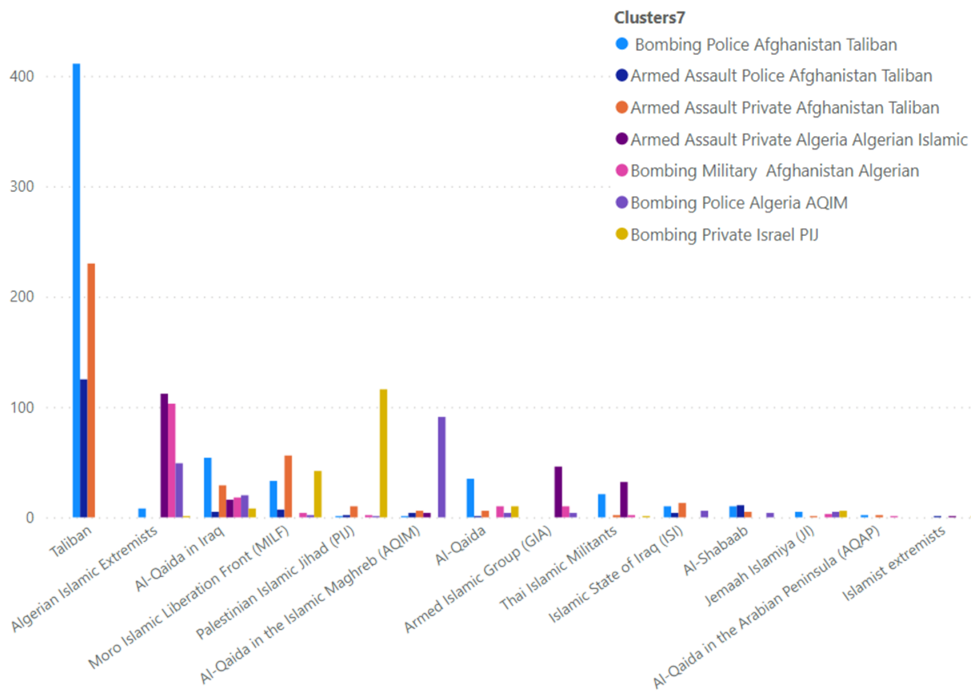

Figure 4.43: Terrorist Groups K-modes Period 2

Chapter 5

Conclusions and Future Works

5.1 Conclusions

5.1.1 Data processing

The division of periods presented in the report can be supported by the data processed in conjunction with articles related to relevant dates. It is also noteworthy to highlight how, through computing tools, the dispersion of the values assigned to each of the three periods analyzed can be measured. Finally, it is essential to analyze the evolution of a data mining process and its management, from starting with a null data cleaning process to going through a grouping process by analyzed values (number of attacks), followed by extraction of relevant terrorist groups for research, ending with the linking of the analysis of the documents related to the terrorist attacks over time together with the result of the data mining, resulting in the consolidated database for each of the periods.

5.1.2 Third period

The exploratory data analysis lets the research know how to endorse the established periods and determine the need to delve further into the third period. Thus limit the research only to period one and period two. We are specifying the problem of chaos that this period suffers through the standard deviations resulting from the exploratory data analysis. For the first period, we show it as the dispersion of the data concerning the mean with values of 75.433267 and 90.8, respectively, unlike the other two periods, where adjustment for outliers was necessary. The analysis also allowed us to understand two things. First, period two with its adjustment allowed us to understand that it was possible to work with it, demonstrating the relationship between the database with the information sources. Second, period three, after showing chaos is its standard deviation compared to the other two periods, allowed the project to determine the need to open a possible second thesis around this period.

5.1.3 Groups evolution

The predominant terrorist group for three of the clusters that monopolize a large part of the terrorist attacks, as shown in figure 4.43, there are also more terrorist groups than in period one.

According to the previous graphs, we can evidence the evolutionary behaviors, considering historical aspects for the second period, groups such as Islamic Salvation Front (FIS) and Armed Islamic Group (GIA). A large part of the members of the FIS group became part of the GIA group by the time of the second war in Algeria called (the years of Lead.)

On the other hand, for the second period, events in which bombing and explosive events occupy 55% compared to those for this period. Furthermore, the critical countries are Algeria, Afghanistan, Iraq, Egypt, and the Philippines in these two periods. These results can be corroborated with studies such as *Win et al* [98] where the authors mention a rapid increase in terrorism over the years in countries such as Iraq, Afghanistan, and Pakistan. According to *Laite et al* [**?**], there are some similarities in specific results concerning the countries most affected by terrorism. There are high mortality rates in countries like Iraq, Afghanistan, and Pakistan due to terrorism, according to the GTD data. There are also results regarding the Taliban terrorist group and its preference for bomb attacks.

According to the terrorist organizations' evolutionary behavior, we find that regarding terrorist organizations such as the Taliban, there is a notable increase concerning the second period. This behavior is reflected in the Taliban's struggle to force the withdrawal of US troops from Afghanistan [99].

According to the different regions, the emergence of new terrorist groups, such as Al-Qaeda, is evident. Also, the emergence of new clusters is apparent. For the second period, the only group that maintains its presence during both periods is the MILF group; in this case, it has an evolutionary change by changing its attack objective from religious targets to private citizen targets [100].

The data and graphs previously exposed show conclusive data on possible terrorist group behaviors through the LCA and K-modes clustering techniques. These techniques converge towards similar results, grouping four types of behavior for the first period and seven for the second. Thus, we can confirm the robustness of the data obtained through the crossover of the methods.

5.1.4 LCA and K-Modes

LCA is commonly used to study social behaviors, medical studies, and social perceptions. Our research broadens the spectrum using this technique towards recognizing terrorist patterns. For the development of the study, we used mixed data extraction, transformation, and loading techniques in conjunction with information classification and grouping techniques through various tools such as the K-Modes and LCA classification. We used verification of the information by going to scientific articles to specify the integrity of the records through study techniques such as exploratory data. We also used programming languages such as Python and R for the research development. In this way, we can conclude with an analysis of the groups obtained and suggest patterns that reveal the behavior of terrorist groups over time. In conclusion, with the development of this research, we obtained the necessary support to understand the nature of terrorist groups of the Islamic State from 1986 to 2008, offering a classification of the data and simplifying its understanding.

5.1.5 Method comparison

Allowing to carry out an analysis supported by the tables obtained. For the comparative analysis of results between the two selected methods, we highlight valuable similarities for the project due to their coincidence with some values of their respective cluster. From the graphs, we could identify and highlight the most critical points of each cluster and their maximum values. For the first period, the names for the LCA results are determined based on the most relevant values of each period. Thus, the resulting names for LCA are ”Armed Assault Police Algerian FIS,” ”Armed Assault Police Egypt IG,” ”Bombing Private Algeria GA,” and ”Bombing Religious Philippines MILF”. On the other hand, the names of the result obtained in K-Modes are ”Armed Assault Police Egypt IG,” ”Assassination Algerian Police,” ”Bombing Private Algeria Islamists Extremists,” ”Bombing Private Philippines MILF”. Finding in the two periods values in their maximum similarities with at least up to one variable, that is, in the example of the first names, we find similarities Armed, assault, Police changing only the country and the perpetrator group.

Finally, to understand the maximum points and their comparisons between LCA and K-Modes, as for period one, we determined the names of the groups using their maximums. They allow the project to converge to a name, according to each cluster, that transmits the value corresponding to that group. So then we can validate the list of names extracted as follows—finding some similar maximum peaks determined with up to at least one value on the graph in one period.

5.1.5.1 LCA List

- Armed Assault Police Afghanistan Taliban
- Armed Assault Private Afghanistan Algerian Islamic
- Bombing Military Algerian Algerian Islamic
- Bombing Police Afghanistan Taliban
- Bombing Private Iraq AQI
- Bombing Private Israel PIJ
- Bombing Private Philippines MILF

5.1.5.2 K-Modes list

- Bombing Police Afghanistan Taliban
- Armed Assault Police Afghanistan Taliban
- Armed Assault Private Afghanistan Taliban
- Armed Assault Private Algerian Algerian Islamic
- Bombing Military Afghanistan Algerian
- Bombing Police Algeria AQIM
- Bombing Private Israel PIJ

5.2 Future works

- The possibility of analyzing in-depth the third period is left open.
- Test this same method used within the framework of this research with different terrorist groups
- An in-depth evaluation of the models based on the information of the third period taking into account the abrupt evolutionary behaviors in terms of the increase in terrorism between 2009-2018 to find new behavioral patterns for this phenomenon of terrorism.
- Implementing models in other regions with other problems different from the Islamic State and its conflicts.

Bibliography

[1] O. E. Dictionary. ”terrorism, n.”., tipo @ONLINE. [Online]. Available: https://www.oed.com/view/Entry/199608?redirectedFrom=Terrorism

[2] L. G. Miller E, Dugan L. ”university of maryland, n.”., tipo @ONLINE. [Online]. Available: https://start.umd.edu/gtd/

[3] C. A. Hannah Ritchie, Joe Hasell and M. Roser, “Terrorism,” *Our World in Data*, 2013, https://ourworldindata.org/terrorism.

[4] A. S. Alsaedi, A. S. Almobarak, and S. T. Alharbi, “Mining the global terrorism dataset using machine learning algorithms,” in *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2019, pp. 1–7.

[5] B. K. Smith, M. Stohl, and M. Al-Gharbi, “Discourses on countering violent extremism: the strategic interplay between fear and security after 9/11,” *Critical Studies on Terrorism*, vol. 12, no. 1, pp. 151–168, 2019.

[6] R. Kitchin and N. Thrift, *International encyclopedia of human geography*. Elsevier, 2009.

[7] S. Sloan, “Technology and terrorism: Privatizing public violence,” *IEEE Technology and Society Magazine*, vol. 10, no. 2, pp. 8–14, 1991.

[8] S. Nie and D. Sun, “Research on counter-terrorism based on big data,” in *2016 IEEE International Conference on Big Data Analysis (ICBDA)*. IEEE, 2016, pp. 1–5.

[9] R. E. Berkebile, “What is domestic terrorism? a method for classifying events from the global terrorism database,” *Terrorism and political violence*, vol. 29, no. 1, pp. 1–26, 2017.

[10] C. Santos, T. El Zahran, J. Weiland, M. Anwar, and J. Schier, “Characterizing chemical terrorism incidents collected by the global terrorism

database, 1970–2015," *Prehospital and disaster medicine*, vol. 34, no. 4, p. 385, 2019.

[11] Y. Woo-suk, "Periodical and spatial differences of terrorism examining global terrorism database from 1970˜ 2018."

[12] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[13] C. C. Aggarwal and P. S. Yu, "Data mining techniques for associations, clustering and classification," in *Methodologies for Knowledge Discovery and Data Mining*, N. Zhong and L. Zhou, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 13–23.

[14] D. Talreja, J. Nagaraj, N. Varsha, and K. Mahesh, "Terrorism analytics: Learning to predict the perpetrator," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2017, pp. 1723–1726.

[15] A. Keivan Hosseiny, "The evolution of european anti-americanism in the post-september era," *POLITICAL QUARTERLY*, vol. 51, no. 2, pp. 563–537, 2021.

[16] A. K. Hosseiny, "The evolution of european anti-americanism in the post-september 11 era."

[17] D. A. Linzer, J. B. Lewis *et al.*, "polca: An r package for polytomous variable latent class analysis," *Journal of statistical software*, vol. 42, no. 10, pp. 1–29, 2011.

[18] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2017.

[19] R. Sathya and A. Abraham, "Comparison of supervised and unsupervised learning algorithms for pattern classification," *International Journal of Advanced Research in Artificial Intelligence*, vol. 2, no. 2, pp. 34–38, 2013.

[20] C. R. C. López, A. M. Rivas, and L. O. Zarate, "Modelos de clases latentes para definir perfiles conductuales en niños de 4 y 5 años." *Revista Electrónica de Psicología Iztacala*, vol. 14, no. 1, p. 354, 2011.

[21] N. Denson and M. Ing, "Latent class analysis in higher education: An illustrative example of pluralistic orientation," *Research in Higher Education*, vol. 55, no. 5, pp. 508–526, 2014.

[22] S. Graf and M. Cecchini, "Identifying patterns of unhealthy diet and physical activity in four countries of the americas: a latent class analysis," *Revista Panamericana de Salud Pública*, vol. 42, p. e56, 2018.

[23] Z. Huang and M. K. Ng, "A fuzzy k-modes algorithm for clustering categorical data," *IEEE transactions on Fuzzy Systems*, vol. 7, no. 4, pp. 446–452, 1999.

[24] F. Cao, J. Liang, D. Li, L. Bai, and C. Dang, "A dissimilarity measure for the k-Modes clustering algorithm," *Knowledge-Based Systems*, vol. 26, pp. 120–127, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.knosys.2011.07.011

[25] Y. B. G. Challenger Pérez, Ivet Díaz Ricardo and R. Antonio, "El lenguaje de programación python," *Ciencias Holguín*, 2014. [Online]. Available: https://www.redalyc.org/articulo.oa?id=181531232001

[26] N. J. de Vos, "kmodes categorical clustering library," https://github.com/nicodv/kmodes, 2015-2021.

[27] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data mining and knowledge discovery*, vol. 2, no. 3, pp. 283–304, 1998.

[28] ——, "Clustering large data sets with mixed numeric and categorical values," in *In The First Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 1997, pp. 21–34.

[29] F. Cao, J. Liang, and L. Bai, "A new initialization method for categorical data clustering," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10 223–10 228, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417409001043

[30] J. Barker, *El sinsentido del terrorismo*. Intermón Oxfam Editorial, 2004, vol. 22.

[31] F. C. Navarro Cardoso, "Criptomonedas (en especial, bitcóin) y blanqueo de dinero," *Revista electrónica de ciencia penal y criminología*, 2019.

[32] R. Ares García, "Desarrollo de algoritmos no supervisados para información sensorial en robótica autónoma," 2019.

[33] A. P. Schmid, J. J. Forest, and T. Lowe, "Terrorism studies," *Perspectives on Terrorism*, vol. 15, no. 3, pp. 142–152, 2021.

[34] A. Andary, "Terrorism and counter-terrorism," vol. 1, p. 14, 01 2021.

[35] T. G. R. Morales, "El terrorismo y nuevas formas de terrorismo," *Espacios Públicos*, vol. 15, no. 33, pp. 72–95, 2012.

[36] P. H. Pilley and S. Sikchi, "Model for predicting terrorist group by clope algorithm," 2014.

[37] D. Holbrook and J. Horgan, "Terrorism and ideology," *Perspectives on Terrorism*, vol. 13, no. 6, pp. 2–15, 2019.

[38] S. G. Jones, C. Doxsee, and N. Harrington, "The escalating terrorism problem in the united states," 2020.

[39] D. Meierrieks and F. Schneider, "Terrorism and international economic policy," *European Journal of Political Economy*, vol. 69, p. 102011, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0176268021000124

[40] C. Seabra, P. Reis, and J. L. Abrantes, "The influence of terrorism in tourism arrivals: A longitudinal approach in a mediterranean country," *Annals of Tourism Research*, vol. 80, p. 102811, 2020.

[41] L. de la Calle and I. Sánchez-Cuenca, "Rebels without a Territory," *Journal of Conflict Resolution*, vol. 56, no. 4, pp. 580–603, aug 2012. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0022002711431800

[42] D. M. D. Silva, "The Othering of Muslims: Discourses of Radicalization in the New York Times , 1969-2014," *Sociological Forum*, vol. 32, no. 1, pp. 138–161, mar 2017. [Online]. Available: http://doi.wiley.com/10.1111/socf.12321

[43] P. J. Schraeder and M. J. Schumacher, "Collective Action, Foreign Fighting, and the Global Struggle for the Islamic State," *Democracy and Security*, vol. 16, no. 3, pp. 234–259, jul 2020. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/17419166.2020.1802690

[44] J. Baudrillard, *The spirit of terrorism and other essays*. Verso Trade, 2013.

[45] N. Saiya, "Blasphemy and terrorism in the Muslim world," *Terrorism and Political Violence*, vol. 29, no. 6, pp. 1087–1105, nov 2017. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/09546553.2015.1115759

[46] K. A. Powell, "Framing Islam/Creating Fear: An Analysis of US Media Coverage of Terrorism from 2011-2016," *RELIGIONS*, vol. 9, no. 9, sep 2018.

[47] M. Dynel and F. I. M. Poppi, "In tragoedia risus: Analysis of dark humour in post-terrorist attack discourse," *DISCOURSE & COMMUNICATION*, vol. 12, no. 4, pp. 382–400, 2018.

[48] B. K. Smith, M. Stohl, and M. Al-Gharbi, "Discourses on countering violent extremism: the strategic interplay between fear and security after 9/11," *Critical Studies on Terrorism*, vol. 12, no. 1, pp. 151–168, jan 2019. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/17539153.2018.1494793

[49] J. Hargreaves, "Police stop and search within british muslim communities: Evidence from the crime survey 2006–11," *The British Journal of Criminology*, vol. 58, no. 6, pp. 1281–1302, 2018.

[50] D. O. Pérez and J. M. A. Izquierdo, "Análisis de clases latentes como técnica de identificación de tipologías," *Revista INFAD de Psicología. International Journal of Developmental and Educational Psychology.*, vol. 5, no. 1, pp. 251–260, 2019.

[51] E. Vernazza, M. Lezcano, and R. Álvarez-Vaz, "Aplicacion del análisis de clases latentes a la construccion de la satisfaccion estudiantil en la facultad de ciencias economicas y de administracion, udelar (uruguay)." *XLVIII Coloquio Argentino de Estadística.*, 2020.

[52] J. L. Galvez-Nieto, J. A. Garcia, D. Vera-Bachmann, I. Trizano-Hermosilla, and K. Polanco, "Multilevel latent class cluster analysis of school climate: individual, family and community factors," *Revista de Psicodidáctica (English ed.)*, vol. 25, no. 2, pp. 85–92, 2020.

[53] F. J. Rondán Cataluña, M. J. Sánchez Franco, and Á. F. Villarejo Ramos, "Análisis de clases latentes en la relación entre la calidad de servicio, satisfacción y confianza con la intención de recompra," *Conocimiento, innovación y emprendedores: Camino al futuro*, 2007.

[54] K. R. Lapierre and A. V. Dane, "Social advantages and disadvantages associated with cyber aggression-victimization: A latent class analysis," *Computers in human behavior*, vol. 113, p. 106497, 2020.

[55] J. Oser, "Protest as one political act in individuals' participation repertoires: Latent class analysis and political participant types," *American Behavioral Scientist*, p. 00027642211021633, 2021.

[56] J. W. Lee and H. Chung, "Latent class analysis with multiple latent group variables," *Communications for Statistical Applications and Methods*, vol. 24, no. 2, pp. 173–191, 2017.

[57] J. Wertz, D. Azrael, J. Berrigan, C. Barber, E. Nelson, D. Hemenway, C. Salhi, and M. Miller, "A typology of civilians shot and killed by us police: a latent class analysis of firearm legal intervention homicide in the 2014–2015 national violent death reporting system," *Journal of urban health: bulletin of the New York Academy of Medicine*, vol. 97, no. 3, p. 317, 2020.

[58] A. Chaturvedi, K. Foods, P. E. Green, and J. D. Carroll, "K-modes clustering," pp. 35–55, 2001.

[59] H. H. Nguyen, "Privacy-preserving mechanisms for k-modes clustering," *Computers & Security*, vol. 78, pp. 60–75, 2018.

[60] Z. He, X. Xu, and S. Deng, "Attribute value weighting in k-modes clustering," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15 365–15 369, 2011.

[61] N. Sharma and N. Gaud, "K-modes clustering algorithm for categorical data," *International Journal of Computer Applications*, vol. 127, no. 1, p. 46, 2015.

[62] N. R. Vajjhala, K. D. Strang, and Z. Sun, "Statistical modeling and visualizing open big data using a terrorism case study," in *2015 3rd International Conference on Future Internet of Things and Cloud*. IEEE, 2015, pp. 489–496.

[63] T. Sandler and W. Enders, "Applying analytical methods to study terrorism," *International Studies Perspectives*, vol. 8, no. 3, pp. 287–302, 2007.

[64] M. S. Tiwari and M. A. Kalam, "Comparative study of clustering algorithms used in counter terrorism."

[65] E. L. Huamaní, A. M. Alicia, and A. Roman-Gonzalez, "Machine learning techniques to visualize and predict terrorist attacks worldwide using the global terrorism database," *Machine Learning*, vol. 11, no. 4, 2020.

[66] K. Bhatia, B. Chhabra, and M. Kumar, "Data analysis of various terrorism activities using big data approaches on global terrorism database," in *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. IEEE, 2020, pp. 137–140.

[67] J. Bernal del Castillo, "El enaltecimiento del terrorismo y la humillación a sus víctimas como formas del discurso del odio," *Revista de derecho penal y criminología*, vol. 16, pp. 13–44, 2016.

[68] L. Bustos Martínez, P. P. De Santiago Ortega, M. Martínez Miró, and M. S. Rengifo Hidalgo, "Discursos de odio: una epidemia que se propaga en la red. estado de la cuestión sobre el racismo y la xenofobia en las redes sociales," *Revista Mediaciones Sociales,(18)*, pp. 25–42, 2019.

[69] K. L. O'Halloran, S. Tan, P. Wignell, J. A. Bateman, D.-S. Pham, M. Grossman, and A. V. Moere, "Interpreting text and image relations in violent extremist discourse: A mixed methods approach for big data analytics," *Terrorism and Political Violence*, vol. 31, no. 3, pp. 454–474, 2019.

[70] H. Fujita, A. Gaeta, V. Loia, and F. Orciuoli, "Hypotheses analysis and assessment in counterterrorism activities: A method based on owa and fuzzy probabilistic rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 831–845, 2019.

[71] B. Ganor, "Artificial or human: A new era of counterterrorism intelligence?" *Studies in Conflict & Terrorism*, vol. 44, no. 7, pp. 605–624, 2021.

[72] H. M. Verhelst, A. Stannat, and G. Mecacci, "Machine learning against terrorism: how big data collection and analysis influences the privacy-security dilemma," *Science and engineering ethics*, vol. 26, no. 6, pp. 2975–2984, 2020.

[73] C. Cheng and B. Zhao, "Intelligent counter-terrorism equipment status and development," in *International Conference on Applications and Techniques in Cyber Security and Intelligence*. Springer, 2018, pp. 318–324.

[74] O. OPS and U. PNUD, "Informe sobre la situación mundial de la prevención de la violencia 2014 [internet]. new york: Oms; 2016 [citado 2021 oct 17]."

[75] P. M. Martín, "El impacto del terrorismo internacional en la industria del turismo: balance y perspectivas en el mediterráneo," *bie3: Boletín IEEE*, no. 3, pp. 349–364, 2016.

[76] M. Youngman, "Building "terrorism studies" as an interdisciplinary space: Addressing recurring issues in the study of terrorism," *Terrorism and political violence*, vol. 32, no. 5, pp. 1091–1105, 2020.

[77] I. Awan, "Cyber-extremism: Isis and the power of social media," *Society*, vol. 54, no. 2, pp. 138–149, 2017.

[78] J. R. Silva, C. Duran, J. D. Freilich, and S. M. Chermak, "Addressing the myths of terrorism in america," *International Criminal Justice Review*, vol. 30, no. 3, pp. 302–324, 2020.

[79] S. Mawlood, "Global view of the main reasons terrorism emergence," *Public Administration and Law Review*, no. 3, pp. 95–108, 2020.

[80] S. M. Brondoni, "Global tourism and terrorism. safety and security management," *Symphonya. Emerging Issues in Management*, no. 2, pp. 7–16, 2017.

[81] B. Schuurman, "Topics in terrorism research: reviewing trends and gaps, 2007-2016," *Critical Studies on Terrorism*, vol. 12, no. 3, pp. 463–480, 2019.

[82] K. Gaibulloev and T. Sandler, "What we have learned about terrorism since 9/11," *Journal of Economic Literature*, vol. 57, no. 2, pp. 275–328, 2019.

[83] E. F. Mickolus, T. Sandler, J. M. Murdock, and P. A. Flemming, *International Terrorism: Attributes of Terrorist Events: ITERATE...* Inter-university Consortium for Political and Social Research, 1976.

[84] Z. Laub and J. Masters, "Islamic state in iraq and greater syria," *The Council on Foreign Relations. June*, vol. 12, 2014.

[85] B. Kadercan, “Territorial logic of the islamic state: an interdisciplinary approach,” *Territory, Politics, Governance*, vol. 9, no. 1, pp. 94–110, 2021.

[86] C. Winter, “Apocalypse, later: a longitudinal study of the islamic state brand,” *Critical Studies in Media Communication*, vol. 35, no. 1, pp. 103–121, 2018.

[87] N. Nuraniyah, “Not just brainwashed: understanding the radicalization of indonesian female supporters of the islamic state,” *Terrorism and political violence*, vol. 30, no. 6, pp. 890–910, 2018.

[88] C. Von Sikorski, D. Schmuck, J. Matthes, and A. Binder, "“muslims are not terrorists”: Islamic state coverage, journalistic differentiation between terrorism and islam, fear reactions, and attitudes toward muslims,” *Mass Communication and Society*, vol. 20, no. 6, pp. 825–848, 2017.

[89] C. Von Sikorski, J. Matthes, and D. Schmuck, “The islamic state in the news: Journalistic differentiation of islamist terrorism from islam, terror news proximity, and islamophobic attitudes,” *Communication Research*, vol. 48, no. 2, pp. 203–232, 2021.

[90] M. J. Sirgy, M. Joshanloo, and R. J. Estes, “The global challenge of jihadist terrorism: A quality-of-life model,” *Social Indicators Research*, vol. 141, no. 1, pp. 191–215, 2019.

[91] A. d. T. Gomes and M. M. Mikhael, “Terror or terrorism? al-qaeda and the islamic state in comparative perspective,” *Brazilian Political Science Review*, vol. 12, 2018.

[92] S. Patel and J. Westermann, “Women and islamic-state terrorism: An assessment of how gender perspectives are integrated in countering violent extremism policy and practices,” *Security Challenges*, vol. 14, no. 2, pp. 53–83, 2018.

[93] U. O. MARYLAND, “Codebook: Methodology, inclusion criteria, and variables,” https://www.start.umd.edu/gtd/downloads/Codebook.pdf, august 2021.

[94] “Mmp: Lashkar-e-islam.” [Online]. Available: https://cisac.fsi.stanford.edu/mappingmilitants/profiles/lashkar-e-islam

[95] “Mmp: Al qaeda in the islamic maghreb.” [Online]. Available: https://cisac.fsi.stanford.edu/mappingmilitants/profiles/aqim

[96] “Mmp: Al qaeda.” [Online]. Available: https://cisac.fsi.stanford.edu/mappingmilitants/profiles/al-qaeda

[97] “Mmp: Jemaah islamiyah.” [Online]. Available: https://cisac.fsi.stanford.edu/mappingmilitants/profiles/jemaah-islamiyah

[98] K. N. Win, J. Chen, G. Xiao, Y. Chen, and P. F. Viger, "A parallel crime activity clustering algorithm based on apache spark cloud computing platform," in *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 2019, pp. 68–74.

[99] "Mapping militant organizations. "afghan taliban." stanford university. last modified june 2018." https://cisac.fsi.stanford.edu/mappingmilitants/profiles/afghan-taliban, accessed: 2021-12-7.

[100] "Mapping militant organizations. "al qaeda." stanford university. last modified january 2019." https://cisac.fsi.stanford.edu/mappingmilitants/profiles/al-qaeda, accessed: 2021-12-7.