

A CLUSTERING-BASED APPROACH FOR IDENTIFYING GROUPS OF MUNICIPALITIES TO SUPPORT THE DIRECTION OF PUBLIC SECURITY POLICIES

Jefferson Carlos de Oliveira Ribeiro Costa^{1*} and Maísa Mendonça Silva²

Received November 3, 2021 / Accepted July 6, 2022

ABSTRACT. The direction of public policies plays an important role in society as a whole, especially in security, which, in addition to being considered a necessity for every citizen, is constitutionally guaranteed. This study presents the use of an unsupervised learning approach for the establishment of clusters among the municipalities in the State of Pernambuco, Brazil, considering some types of representative crimes, aiming to direct actions to prevent and fight crime in order to support policy makers. The k-means algorithm was used as the main tool in the study, using the software R 3.6.1, and recommendations for actions were directed to each of the obtained clusters. To demonstrate the direction, the grouping with the parameter $k = 26$ was used, referring to the State Security Integration Areas. The results show that the use of a clustering approach for the municipalities provides greater effectiveness in directing actions to combat and prevent crime, given that the municipalities that have the greatest similarities are grouped in the same cluster.

Keywords: cluster, public security, K-means.

1 INTRODUCTION

The complexity of decision-making in public security has been increasing, which is natural concerning the various aspects that influence decisions in this area, ranging among political, economic and socio-cultural issues, among others. In a 30 year-period (from 1980 to 2010), more than 1 million homicides were registered in Brazil (PEREIRA et al., 2016). As stated by the authors (PEREIRA et al., 2016), this number is astonishing for a country that was not experiencing any ethnic, religious and/or territorial conflicts. In addition, this is a scenario that has been getting worse and presents a certain urgency with regard to dedication to interventions that may reduce these numbers. The Institute for Applied Economic Research (IPEA, 2019), in 2012, reported that the homicide rate in Brazil was 29.41 per 100,000 inhabitants, rising to 31.59 homicides per 100,000 inhabitants in 2017. Additionally, there is a significant variation in this rate across the

*Corresponding author

¹Universidade Federal de Pernambuco, Av. Prof. Moraes Rego, 1235, 50670-901 Recife, PE, Brazil – E-mail: jeffersoncrbeiro@gmail.com – <https://orcid.org/0000-0003-3188-4155>

²Universidade Federal de Pernambuco, Av. Prof. Moraes Rego, 1235, 50670-901 Recife, PE, Brazil – E-mail: maisa.ufpe@yahoo.com.br – <https://orcid.org/0000-0003-4223-2769>

country. Some states register a reduction in homicide rates, while others experience a constant increase. Pernambuco, in Brazil, for example, registered a growth of around 40% in homicide rates in 2017, as indicated by the Brazilian Forum on Public Security (MOTA et al, 2020; IPEA, 2019).

The issues related to public security policies, regarding their elaboration and evaluation of efficiency are frequently discussed, since violence directly affects everyone. One of the most studied subjects in the public security area is the geographical analysis of criminal occurrences. Indeed, the research area known as criminal geography, which involves identifying possible countermeasures against violence, uses the microscale approach in several studies (ANDRESEN, 2015). However, these studies have used geographical analysis and have determined crime factors but have not addressed the possibility of grouping similar areas in criminal terms.

As suggested by Mota et al. (2020) the identification of areas with the highest violence rates or with a greater propensity to violence can be considered the first step in the development of effective security policies. In addition, Den Heyer (2014) points out that there is a greater emphasis by police in directing resources to specific geographic areas of high criminality rates or to specific types of crimes. In this context, Weisburd and Amram (2014) claim that it is necessary to investigate chronic places in which the occurrence of a specific type of crime is higher than in other areas. It is worth stating that a region can be chronic in one specific crime, but it can be peaceful in another.

As a consequence of this problem of identifying areas with high criminality, as highlighted in the literature, issues related to the grouping of municipalities are of paramount importance in the analysis of crimes, even if the spatial pattern is not considered. Therefore, the problem addressed in this paper regards the identification of crimes with high rates of occurrence in specific areas, following the allocation of these areas in clusters derived from crime similarity. This is exactly the aim of a cluster analysis, which is one of the standard approaches of data analysis in unsupervised machine learning techniques, in which data that have similarities are grouped in the same cluster. It is worth stating that, in related works, it is possible to find studies using clustering approaches, such as partitioned, hierarchical and density-based approaches (WANG et al, 2017). however, regarding applications in the area of public security, a gap can be observed, as studies tend to be more focused on the search for determinants of crimes and not on grouping regions according to similar crimes (LIMA & MARINHO, 2017).

Furthermore, knowing that these groups can reveal some characteristics that will guide the direction of specific public security policies, this work proposes to group municipalities with similar crimes using the number of occurrences of different crimes of the cities (instances) as value attributes, after these attributes are reduced and assigned to components using a Principal Component Analysis (PCA). Finally, an unsupervised learning approach is used, more precisely, the k-means algorithm, in order to assertively direct actions that focus on combating and preventing crimes with the highest incidence for each of the clusters obtained. The application and validation of the proposal were performed in the 185 municipalities of Pernambuco, Brazil, with real data released by the Secretariat of Social Defense of the state (SDS-PE).

Briefly, compared to similar works in the related literature, this paper innovates in proposing a clustering approach to group municipalities regarding similar types of crimes. It also contributes to the possibility of adopting similar security policies in cities within the same cluster, leading to a more efficient and focused way to combat crimes in terms of resources, police training and cooperation among cities with similarities regarding crimes. This paper is structured as follows: Section 2 is dedicated to the use of data mining techniques in the public security context; Section 3 describes the problem addressed in this paper; Section 4 shows the application of the proposal in a Brazilian State and finally, Section 5 presents the conclusions.

2 DATA MINING IN THE PUBLIC SECURITY CONTEXT

Understanding the causes of crimes is a longstanding issue on a researcher's agenda (ALVES et al., 2018). Several techniques are presented in the literature to assist not only in understanding the causes of crimes but also their spatial pattern. Understanding the role of the environment and how the crime spatial pattern occurs is important for its control and investigation (WORTLEY & MAZEROLLE, 2008). Many variables can be considered to make the understanding as robust as possible, such as the fact that critical poverty encourages a rational individual to make irrational decisions, such as committing the crime when the possibility of being caught, exposed and convicted is high or when circumstances are not favorable (NEPOMUCENO et al., 2020). Oatley and Ewart (2003) developed a software that uses mapping and visualization tools. Adeyemi et al. (2021) conducted an exploration of the spatial distribution of crimes in Nigeria, identifying dependency relationships among crimes.

The statistical methods employed and the data mining technologies are used to investigate the impact of the types of evidence available and determine the causality of that domain. Qazi and Wong (2019) proposed a man-centered knowledge discovery for crime text mining that, according to the authors, is capable of extracting plausible associations among crimes, identifying patterns, grouping similar types of crimes, generating a network of co-offenders and a list of suspects based on spatial-temporal and behavioral similarity. In the case of Agarwal, Nagpal and Sehgal (2013), the authors performed a project that focused on the analysis of crimes, implementing a clustering algorithm, the k-means, in the crime data set using a fast-mining tool, and conclusions were reached based on the trend of some types of crimes over the years.

Several other works are found in the literature using data mining and approaching the topic from different perspectives, because there is no field of study as appropriate as data mining to perform crime analysis (AGARWAL et al. 2013). Nonetheless, even with the wide application of mathematical techniques in the context in question, there is still a gap in the literature regarding the grouping of municipalities using a technique for the understanding of an expressive amount of crime data. In this sense, data mining is used in this paper to obtain a reduced data dimension, to group municipalities, and to help the extraction of knowledge to support the effectiveness of policy makers in directing actions.

In the context of the region addressed in this study, the Integrated Security Areas (ISA) are considered, each of them being composed of specific regions (municipalities) with their divisions based on the vicinity/proximity of the regions. The k-means, which is one of the cluster analysis techniques, was selected and a number of k-groups with k equal to different values was determined. It is worth noting that the algorithm is considered practical and well disseminated in the literature; in addition, it has several packages in statistical and data mining tools. However, the main reason for using k-means is that, for a big amount of data, which is the case of this study (185 cities and 11 types of crimes), other types of clustering methods, such as the hierarchical ones, are not recommended, since they are computationally demanding.

To demonstrate this use, a brief literature review was conducted in which papers that used data mining techniques applied to public security were selected. In addition to the works already mentioned, ten papers that demonstrate a line of research more aligned with the content discussed here were selected, as shown in Table 1. As can be seen, the novelty of this paper is mainly the proposal of a clustering approach to group municipalities regarding similar crimes.

Table 1 – Studies on crime analysis by data mining.

Authors, Year	Description
Chapagain et al. (2022)	In this study, the clustering approach based on PCA with k-means is proposed for intrusion detection in a system that detects malicious activities and issues alerts.
Das & Das (2019)	The authors propose a grouping technique based on graphs to discover crime report labels based on paraphrases extracted from large bodies of unidentified crimes.
Das et al. (2019)	The authors propose an incremental adaptive methodology in a crime reporting group, adapted to provide updated cluster sets. In addition, the study makes comparisons with clustering algorithms to express their effectiveness and statistical significance.
Khan et al. (2019)	The authors propose a model for the predictive policing system for street crimes in the Karachi-Pakistan region using the k-means tool and Bayesian methods.
Farias et al. (2018)	This work presents different models related to public security. These models are based on clustering algorithms, on the analysis of formal concept techniques and on the analysis of crime record data collected in Mossoro, Rio Grande do Norte, Brazil.
Wang et al (2018)	They propose a method of tied clues to obtain refined results regarding the phenomenon known in the analysis of crimes as the quasi-repetition effect, adopting a data science perspective, combining correlation coefficient, hierarchical cluster and mining of frequency patterns in a specific order.
Salto & Cocea (2017)	In this paper, models to predict the frequency of several types of crimes by the LSOA code (an administrative system of areas used by the police in the UK) and the frequency of crimes related to anti-social behavior are presented. Three algorithms are used from different categories of approaches: instance-based learning, regression and decision trees.
Borg & Boldt (2016)	The study investigates the use of clustering algorithms to group similar crime reports based on combined characteristics from the structured form.
Curman et al. (2015)	The authors conduct a longitudinal analysis of a 16-year data set using the street segment as the unit of analysis, highlighting the trends in the crimes, and using the group-based trajectory model with the k-means cluster analysis technique.
Tayal et al. (2014)	The authors propose an approach for the design and implementation of crime detection and criminal identification for Indian cities using data mining techniques.

3 PROBLEM DESCRIPTION

This study aims to use an unsupervised learning approach to establish the clusters among the municipalities in the State of Pernambuco, considering several variables, to assist in directing public security policies to combat and prevent crimes. The data used were made available by the National Secretariat of Public Security, through Law No. 12,527/2011 (Law on Access to Information). Data on criminal occurrences were provided for the year 2018, separated according to the 185 municipalities in Pernambuco.

The literature discusses the implementation of crime prevention or reduction programs, assigning this task to the local governments. Nevertheless, in the context adopted in this study, a certain gap is perceived regarding the direction of these programs, which can be supported by specific systems. In addition, the planners' prior knowledge of the problem can be used along with historical data not only to develop new actions but also to critically analyze their direction (FIGUEIREDO & MOTA, 2019; CHATAWAY & HART, 2018).

Brazil has 27 federative units, and Pernambuco is one of them. It is located in the center-east of the Northeastern area, which is the third largest region of the country in territorial extension (estimated at 1,554,000 km²) and the second most populous, with roughly 50 million inhabitants (IBGE, 2010). Despite the socioeconomic improvements in the Northeast, Pernambuco is one of the Brazilian states with the highest crime rates. According to data from the Secretariat of Social Defense of the State of Pernambuco (SDS-PE, 2019), between January and October 2019, there were 2,881 victims of intentional lethal violent crimes and 67,382 violent crimes against property.

As stated in the Quarterly Bulletin of Criminal Structure in Pernambuco - 1st Quarter of 2018, the disclosure of indicators on public security is guided by scientific criteria for the treatment of information that follow the technical guidelines of the National Statistics System. According to the State Planning and Research Agency of Pernambuco, since 2008 (the beginning of the system for the dissemination of indicators on public security), the state government has been committed to spread information on violence in the state, ensuring the basic principles of data reliability and comparability, along with accessibility to quality information for all citizens of Pernambuco (CONDEPE/FIDEM, 2015). The state is a reference in the collection of crime data, due to significant improvements considering new procedures with high technology in the process. Therefore, the information is considered reliable and accurate (PEREIRA et al., 2016; SAURET, 2012).

In addition, the Brazilian Public Security Forum classifies federative units every two years according to the quality of criminal data. This study considers five axes of evaluation for information quality, as stated in the 12th Brazilian Yearbook of Public Security, which are: (1) the concept of homicides, (2) the information recorded about victim, fact and suspect, (3) the information loss about the victim, fact and suspect, (4) the degree of data convergence from state departments with the official health source and (5) transparency. Based on the defined classifi-

cation, the states are grouped into four groups. Group 1, where Pernambuco is located, has the states with the highest information quality.

Modern conceptions of Public Security Policies involve the coercive/repressive and preventive dimensions. The preventive dimension involves all government strategies aimed at the social prevention of violence and crime (RATTON, 2014). According to Nóbrega (2008), the theme is a complex system that involves at least three subsystems: the police, the judicial and the prison. In the Brazilian model, the integration between investigation and trial procedures is a major obstacle to security. For this reason, the construction of public policies in the area is not a simple act, since such policies imply the articulation between different segments of the three branches of government – federal, states and municipalities.

This reality is what highlights the experience built by the State of Pernambuco in which the implementation of public policies in the areas of Security, Education and Health, in the form of Pacts, has proved to be highly effective. These policies reflect a management model focused on results, which incorporates strategy, alignment of the implementing structure, monitoring and evaluation of the results.

The State of Pernambuco has been suffering from the serious problem of violence, which has intensified from year 2000 onwards and, in the subsequent five years, Recife had the highest homicide rate among all capitals in the country (2000 to 2006). Pernambuco, in turn, had the highest rate of Intentional Lethal Violent Crimes (CVLI) among the Brazilian states, with rates greater than twice the national average (2004 and 2005). During this period, there was a government program with several actions to neutralize the problem of crime and violence in the state. However, such initiatives lacked a single direction from a structured plan of actions. There was no strategy with defined goals or integrated actions involving different actors (outside the police). The reduction of violence became, therefore, one of the priority focuses of the Government in 2007. From a diagnosis performed in the first month of the Government, it became clear that violence in Pernambuco would not be reduced only with the police, requiring the involvement of several actors. Thus, the first movements that would give rise to the Program “Pacto pela vida”– PPV (in English, Pact for life) were initiated, forming a multidisciplinary study group and holding permanent discussion forums on the subject.

The territorialization of the State in 26 Integrated Security Areas (AIS) represented an advance for the management of PPV. Criminal information started to be generated by area, making it possible to understand the different realities of CVLI in Pernambuco. Thus, it was possible to respond more adequately to crimes in these areas. This understanding also allowed an integrated action between the civil and military police, both being held accountable for the result in their areas of crime reduction. To give an idea of this successful management model, the Metropolitan Region of Recife, which was the most violent in Brazil, moved to the 4th place in the Brazilian ranking of violence in 2011 (Annual Management Report, 2011). Nevertheless, over the years, the need to redefine the AIS arose, since the proximity criterion no longer seemed to make sense. The criterion of similarity of crimes was chosen in this article because several authors in the literature indicate advantages of grouping areas using this perspective.

4 APPLICATION AND RESULTS

First, some variables were selected, and among them there are the 185 municipalities of Pernambuco. Following the Law on Access to Information, the 2018 criminal occurrence data were available by the National Secretariat of Public Security, separated according to each municipality. Table 2 shows the description of each type of crime considered in the study.

Table 2 – Description of the variables.

Crime	Description
C1	Trade of illegal substances
C2	Act of intentionally killing another person
C3	An offence to a person's bodily integrity or health
C4	Crime with intent in the previous act and guilt in the subsequent fact
C5	Crime in which there is deceit in killing the victim to subtract something
C6	Appropriation of a vehicle in which the victim is not present
C7	Appropriation of a vehicle through violence or threat
C8	Crime aimed especially at bank branches
C9	Crime that occurs in the transfer of goods
C10	Robbery to commercial establishments, in general
C11	Home invasion and theft

To allow a simplified data interpretation, a Principal Component Analysis (PCA) was performed to reduce the number of variables without significant loss of information about the set of original variables. By reducing the size of a huge amount of data, a representative smaller database can be obtained with reduced noise (PRABAKARAN & MITRA, 2019). More precisely, PCA is applied to reduce the dimensionality of the data, condensing the information contained in its several original variables into a smaller set of variables (components) and with a minimal loss of information. Each main component is defined as a linear combination of the original variables and is represented by a new set of artificial variables that are a linear function of the original ones with maximum variance.

Other studies have followed a procedure similar to the one used in the present work. For example, when trying to monitor network traffic with the detection of malicious activities, the authors²⁸ used PCA to extract resources in reduced dimensions, preceding the clustering through k-means. In the study of theoretical analysis of the k-means algorithm and its applications, it was observed that when combined with PCA, the algorithm achieves higher performance and clustering speed (LI et al., 2019).

In this article, PCA was performed with data on the number of occurrences of the 11 types of crimes (shown in Table 2) in each of the 185 municipalities. The first step of the analysis was to calculate the correlation coefficients among all crimes. The premise of this analysis is to extract non-correlational factors (new variables) that capture the behavior of the original variables (11 types of crimes) with high correlation coefficients.

After PCA, clusters were obtained using the k-means algorithm. The algorithm was executed using the software R 3.6.1 and recommendations for actions were directed to each cluster. The clustering package had an important role in the application and implementation of the clustering technique, and the R package named *factoextra* facilitated both the extraction and visualization of the output of the exploratory analysis of multivariate data conducted herein. Additionally, *factoextra* also allows the implementation of the PCA.

To provide a two-dimensional visualization throughout the study and to simplify the analysis, only the first two main components were used. Table 3 contains the proportions of variation explained by each component and the accumulated proportions, indicating that components 1 and 2 alone already explain 80.57% of the entire data variation, with 70.46% corresponding to the first main component and 10.11%, to the second.

Table 3 – Proportion of variation explained by the components.

Principal Component	Proportion (%)	Accumulated Proportion (%)
PC1	70.46	70.46
PC2	10.11	80.57
PC3	9.36	89.93
PC4	5.70	95.63
PC5	2.58	98.21
PC6	0.67	98.88
PC7	0.39	99.27
PC8	0.33	99.60
PC9	0.20	99.79
PC10	0.12	99.91
PC11	0.08	100.00

The structural relationship between the variables and the main components can be analyzed from the variable graph (Figure 1), which presents a view of the projections of the observed variables projected on the plane measured by the first two components. To understand the graph, it is important to mention that its deduction is focused on the ends (top, bottom, left and right).

Thus, the first principal component mainly reflects violent acts of any kind, with lesser contribution from crimes 4 (Crime with intent in the previous act and guilt in the subsequent fact), 5 (Crime in which there is an intent on killing the victim in order to subtract something from them) and 8 (Crime modality aimed especially at bank branches). Regarding the second principal component, crimes 4, 5 and 8 have the greatest weight.

Data size was duly reduced, and, considering the application of the Principal Component Analysis, clusters were obtained using the k-means clustering technique. Initially, it was necessary to establish the number of groups/clusters. An integer must be assigned to parameter k in advance,

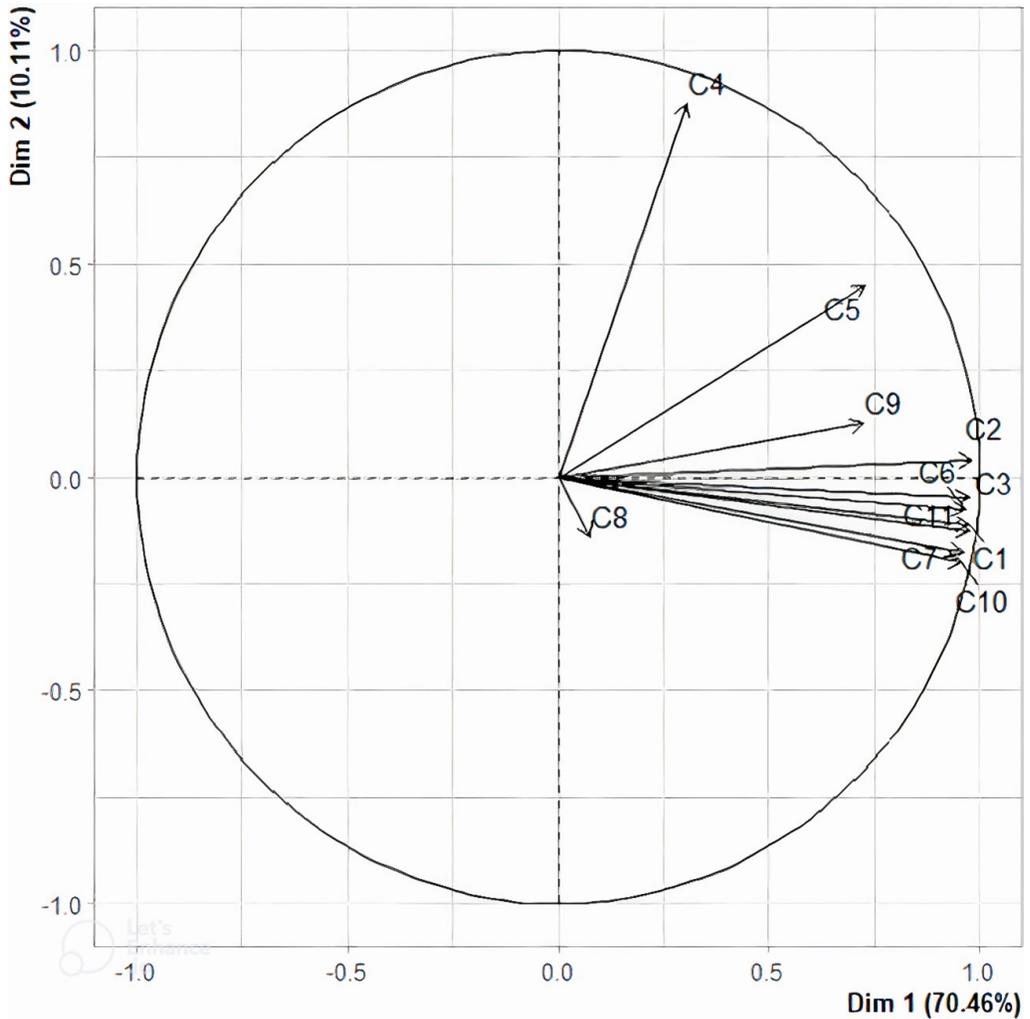


Figure 1 – Relationship between the variables and the main components.

which will precisely be the cluster amount. This parameter is usually defined as an *ad hoc* basis by the user. Typically, the value of k is chosen based on a priori knowledge of the problem, requiring an expert's critical view (WITTEN et al. 2011; SILVA & RIBEIRO, 2018). In this article, experts from the SDS-PE participated in this decision.

The algorithm was firstly applied to the following k values: 2, 3, 4 and 5 (Figure 2). In all cases, a certain disparity was observed in one of the points: Recife, the state's capital. This is related to the high crime rate compared to other municipalities. Crime 2, for instance, had the highest number of occurrences in Jaboatão dos Guararapes, in 2018. Except for the state's capital, still considering this type of crime, Jaboatão dos Guararapes appeared at the top ranking in the state. Nonetheless, the number of occurrences of crime 2 in Recife was approximately 3.2 times higher.

Furthermore, the grouping of municipalities was analyzed for the parameter $k = 26$; this number refers to a territorial division currently used in Pernambuco, comprising eight areas (Integrated Security Areas): Capital, Metropolitan Region, North Forest Zone, South Forest Zone, Agreste 1, Agreste 2, Sertão 1 and Sertão 2. Each of these ISA is composed of specific circumscriptions, being neighborhoods and/or municipalities, and has its divisions based on the proximity of the regions.

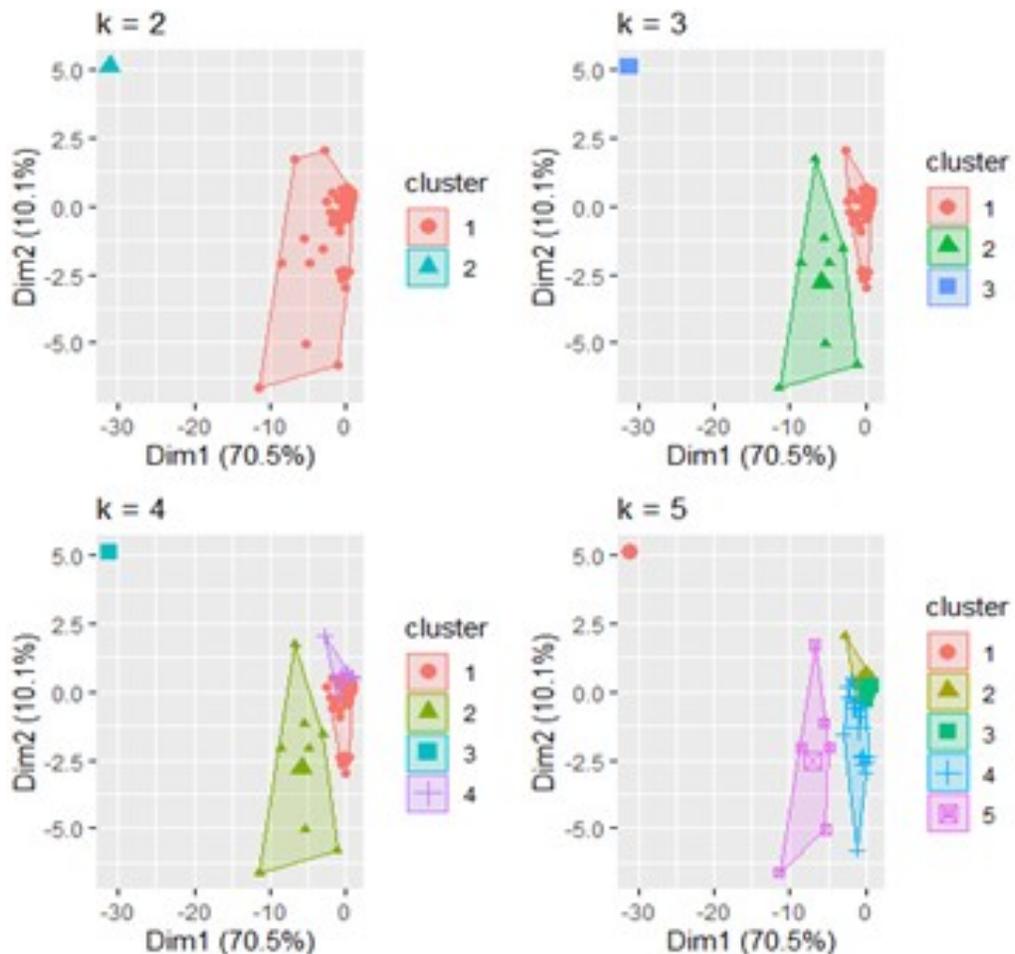


Figure 2 – Clusters 2018 ($k=2, 3, 4$ and 5).

As stated in the Decree No. 1197, June 11, 2010, the boundary between territories, areas and circumscriptions considers the technical-cartographic criteria provided by institutions such as the Brazilian Institute of Geography and Statistics, the State Planning and Research Agency of Pernambuco, among others. These areas can be seen in Table 4.

These territorial divisions were established in order to distribute the role of the military and civil police in an isometric manner, stimulating the integrated action of both institutions and allowing them to be used as a way to target public security policies (LOPES, 2016).

Figure 3 represents the state map divided by the clusters obtained using the parameter $k = 26$. There is not necessarily a relationship in proximity, regarding the location of the municipalities, for the grouping of the regions. A clear example is the cluster presented in green (number 21), which is composed of Petrolina and Paulista. Petrolina is 712 km west of the state capital, while Paulista is 18 km away from Recife.

Table 4 – Pernambuco’s Integrated Security Areas.

Territory	ISA	ISA Nomenclature
Capital	1	Santo Amaro
	2	Espinheiro
	3	Boa Viagem
	4	Várzea
	5	Apipucos
Metropolitan Region	6	Jaboatão dos Guararapes
	7	Olinda
	8	Paulista
	9	São Lourenço da Mata
	10	Cabo de Santo Agostinho
North Forest Zone	11	Nazaré da Mata
	16	Limoeiro
South Forest Zone	12	Vitória de Santo Antão
	13	Palmares
Agreste 01	14	Caruaru
	17	Santa Cruz do Capibaribe
Agreste 02	15	Belo Jardim
	18	Garanhuns
Sertão 01	19	Arcoverde
	20	Afogados da Ingazeira
	21	Serra Talhada
	22	Floresta
Sertão 02	23	Salgueiro
	24	Ouricuri
	25	Cabrobó
	26	Petrolina

The fact that these two municipalities are in the same group is directly related to the Euclidean distance, calculated based on the number of occurrences of each type of crime. Therefore, they

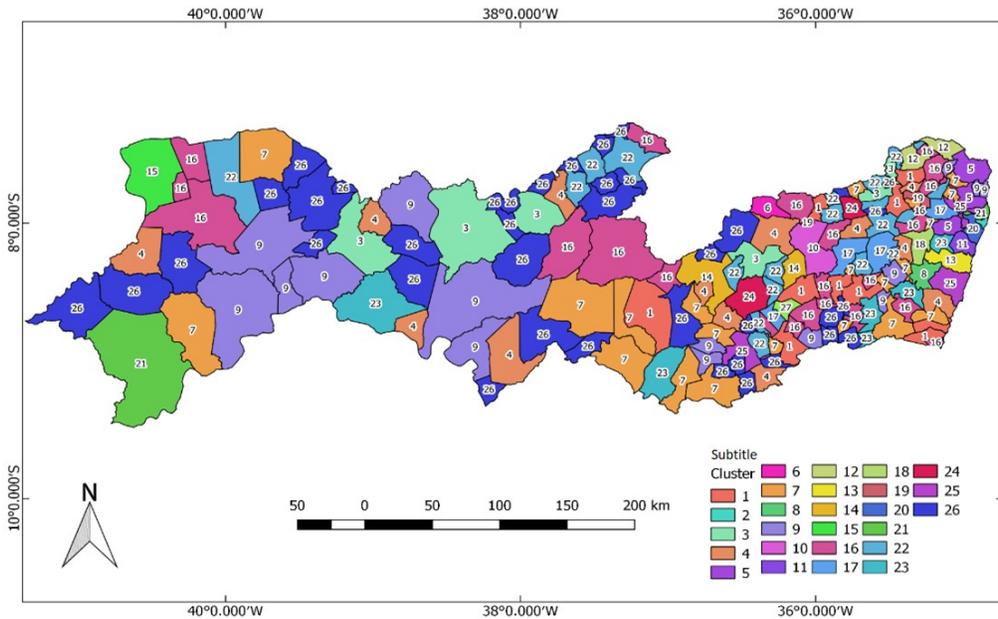


Figure 3 – Pernambuco's map segregated by clusters.

were grouped focusing on similarity. This observation allows the understanding of the need to use clustering approaches to direct actions against crime, and not only the use of regions established according to the proximity among municipalities.

Similar crime practices among cities are a common sense, and, due to the grouping into clusters, it was possible to reach groups with similar occurrences. As a theoretical contribution, to the extent of our knowledge and based on the literature review performed in Section 2, it can be concluded that this is the first paper that proposes a clustering approach to group municipalities considering similar crimes. Furthermore, in a practical perspective, this paper contributes to the redefinition of the ISAs in the State of Pernambuco and the possibility of adopting specific security policies in cities within the same cluster.

To demonstrate the direction of appropriate policies for each of these groups, a survey of some national and state policies and programs was conducted for each type of crime, such as the National Drug Policy, the documents and programs defined by Secretariat for Violence Prevention and Drugs Policy of the State of Pernambuco, the Social Defense Secretariat of Pernambuco and the National Public Security Policy. Therefore, the actions proposed in this study were defined based on information from different areas, and together with experts from the SDS-PE. Several meetings were held with SDS-PE to elaborate the actions to be implemented to combat each type of crime.

As a result, it was possible to list specific actions to combat and prevent crimes and, thus, direct categories of actions that best suit each of the clusters. Both the categories of crimes and their respective actions are described in Table 5.

Table 5 – Actions according to the type of crime.

Crime	Actions
C1	Investing in educational programs Promoting continuous action to dismantle criminal organizations Investing in technologies to monitor drug plantations Establishing a state intelligence division to combat drug trade Deploying police crackdowns
C2	Implementing a plan to combat crimes against life Deploying monitoring cameras in specific locations Requalifying structures or expanding protection programs to groups that are highly vulnerable to violence Strengthening investigative skills of police officers
C3	Implementing a plan to combat crimes against life Creating social projects Increasing the number of police rounds Requalifying structures or expanding protection programs to groups that are highly vulnerable to violence
C4	Implementing a plan to combat crimes against life Gradually increasing the number of police officers Capacity building through courses on conflict mediation and humanized care Support and qualification courses for socially vulnerable groups
C5	Conducting police operations where violence incidence is high Reducing bureaucracy and strengthening investigative skills of police officers Training professionals on investigation, interview and interrogation techniques Deploying monitoring cameras on strategic routes
C6	Expanding information exchange and system sharing among states Deploying/expanding the video surveillance and license plate tracking system Expanding the use of information technology in vehicle monitoring Defining vehicle verification operations in strategic locations
C7	Establishing a task force in vehicle theft investigation Deploying (or investing) in police stations to combat vehicle theft Expanding information exchange and system sharing among states Deploying/expanding the video surveillance and license plate tracking system
C8	Investing in risk management together with the Brazilian Federation of Banks Determining intelligence actions with a multidisciplinary team Investing in image sharing systems from bank security cameras with law enforcement authorities
C9	Promoting continuous actions to dismantle criminal organizations responsible for cargo theft Expanding inspection on highways Investing in specialized personnel, such as tax auditors
C10	Investigating and dismantling gangs practicing thefts Periodically increasing the coordination of patrols in strategic areas Promoting advisory and partnership actions with traders from the regions
C11	Expanding camera monitoring systems Identifying the regions with the highest incidence and conduct patrols Performing actions in partnership with city halls to identify the lighting need in strategic points

To exemplify the direction of the actions, the grouping with 26 clusters was considered, since this number refers to the number of the Integrated Security Areas in Pernambuco currently in use. In all clusters a prevalence of crime C3 and crime C7 was identified, which are crimes related to bodily integrity or someone's health and vehicle appropriation through violence or threat, respectively.

When considering the actions proposed in Table 5, some indications for combating crime C3 could be associated with the national/state plan, such as: the creation of social projects, the increase in police patrol, as well as the requalification, creation or structuring programs to protect social groups in situations of high vulnerability to violence.

The second type of crime with the highest number of incidences is vehicle theft (C7). In this case, it is important that there be investments in devices and programs to combat this crime, and some actions to be adopted are: the establishment of a task force to investigate vehicle theft gangs, the implementation or investment in police stations aimed at combating theft, in addition to expanding the exchange of information and sharing of information systems among states and the implementation/expansion of video surveillance and tracking systems.

Nonetheless, excluding these two types of crimes that are common in all clusters, Table 6 presents the most representative types of crime that need special attention for each of the clusters. Therefore, it is possible to associate Table 6 and Table 5 to obtain the recommendations of public security actions that best suit the municipalities from a given cluster. For instance, the municipalities that are part of cluster 1 must focus the recommendations for crime 2, while the municipalities assigned to cluster 2 must focus the recommendations for crime 1, and so on. To give some examples, clusters 8 and 14 stand out for crimes C9 (cargo robbery) and C11 (home robbery), respectively. For the municipalities included in cluster 8, it is suggested to promote continuous actions to dismantle criminal organizations responsible for cargo theft, to expand inspections on highways and invest in specialized personnel. With regard to cluster 14, it is important to expand the camera monitoring systems, as well as conduct actions to improve public lighting at strategic points.

As can be seen in Table 6, different clusters present the same representative type of crime; this can be justified from the particularities related to each cluster. For instance, there is a disparity in the proportion per inhabitant of each of the crimes. For example, clusters 1, 4, 7, 13 and 26 have the same most common type of crime, which is crime C2. In addition to crimes C3 and C7, which are presented in all clusters, one may think that these clusters are the same. However, the average number of each type of crime in each cluster can be critical for differentiating them. Another important point is that this paper, for simplification reasons, only presents the three more common types of crimes for each cluster. Nevertheless, each cluster has other types of crimes that also make them diverse from each other.

Finally, a positive consequence of using the proposal presented here is that the possibility of adopting similar security policies in cities within the same cluster leads to a more efficient way to combat crimes. For instance, cities within the same cluster, that is, with similar types of crimes,

Table 6 – Representative types of crime per cluster.

Cluster	Crime	Cluster	Crime
1	C2	14	C11
2	C1	15	C1
3	C1	16	C6
4	C2	17	C6
5	C6	18	C1
6	C6	19	C6
7	C2	20	C10
8	C9	21	C1
9	C2	22	C6
10	C6	23	C2
11	C1	24	C6
12	C1	25	C1
13	C2	26	C2

can perform joint police training. There is also a clear opportunity to better assign all types of resources – human, financial and technical – since the operations would be more focused. Finally, in a general way, the adoption of the crime clustering perspective proposed here may promote a greater cooperation among cities with similar types of crimes.

5 CONCLUSION

The extremely high incidence of violence, in general, leads to the need to investigate the subject from different perspectives to reduce and control its rates. By understanding the context of Brazilian violence and aiming to achieve positive results, it is crucial that decisions be structured efficiently and that public resources have an effective direction.

Therefore, this work proposed to use an unsupervised learning approach to establish clusters among the municipalities of Pernambuco, considering some representative crime variables in the State, to direct actions to prevent and fight crimes and support public security policy makers.

This decision is characterized by requiring the ability of the decision-maker to better evaluate the regions regarding their crimes and the use of the referred technique is a strong support in this sense, since the aim of clusters is homogeneity within the groups. As a result, the components (in this case, the municipalities) share common criminal characteristics that are different from the municipalities in other clusters. Therefore, clustering as a tool to aid the decisions that direct public security actions provides a better view of the regions that present a greater need for attention in Pernambuco.

By obtaining clusters based on the number of criminal occurrences, experts can analyze clusters of municipalities simultaneously, prioritize resources and direct policies more assertively, so that

the clusters can provide a source for an interdisciplinary study. In this sense, a simple way of directing policies was demonstrated based on the incidence of the types of crimes. Actions to combat and prevent specific crimes were segregated, and the most registered crimes in the cluster were analyzed.

As a limitation of this work, since crimes are dynamic by nature and consequently, the proposal must be applied from time to time or even in real time, the practical results may not reflect the actual situation of the cities of Pernambuco, since the data employed referred to the year 2018. For future works, the development of a computational tool integrated with real data is suggested to allow a dynamic analysis of the crimes. An MCDA approach, such as CPP-TRI (SANT'ANNA ET AL., 2015), which deals with large amounts of data, can also be applied after the understanding of the peculiarities of the clusters, since predetermined classes could be defined.

Acknowledgements

The authors would like to thank the Secretariat of Social Defense of Pernambuco (SDS-PE) for the information provided. We are also grateful for the reviewers' comments which helped us to improve the quality of the paper.

References

- ADEYEMI RA, MAYAKI J, ZEWOTIR TT & RAMROOP S. 2021. Demography and Crime: A Spatial analysis of geographical patterns and risk factors of Crimes in Nigeria. *Spatial Statistics*, **41**.
- AGARWAL J, NAGPAL R & SEHGAL R. 2013. Crime Analysis Using K-means Clustering. *International Journal of Computer Applications*,**83**(4): 1-4.
- ALVES LGA, RIBEIRO HV & RODRIGUES FA. 2018. Crime prediction through urban metrics and statistical learning. *Physica A: Statistical Mechanics and Its Applications*, **505**:435-443.
- ANDRESEN MA. 2015. Predicting local crime clusters using (multinomial) logistic regression. *Cityscape: A Journal of Policy Development and Research*, **17**(3):327-339.
- BOLT & BOLDT M. 2016. Clustering Residential Burglaries Using Modus Operandi and Spatiotemporal Information. *International Journal of Information Technology & Decision Making*, **15**(1): 23-42.
- BRAZILIAN PUBLIC SECURITY FORUM. 2019.*Brazilian Yearbook of Public Security*, Year 13.
- CHAPAGAIN P, TIMALSINA A, BHANDARI M & CHITRAKAR R. 2022. Intrusion Detection Based on PCA with Improved K-means. *ICEEE International Conference on Electrical and Electronics Engineering*, **894**: 13-27.

CHATAWAY ML & HART TC. 2018. Crime prevention and reduction programs: How does knowing about community initiatives moderate attitudes towards criminal victimization? *Australian New Zealand Journal of Criminology*, **51**: 239–257.

CONDEPE/FIDEM - STATE PLANNING AND RESEARCH AGENCY OF PERNAMBUCO. 2015. *Statistics of Violent Crime in Pernambuco*.

CURMAN ASN, ANDRESEN MA & BRANTINGHAM PJ. 2015. Crime and Place: A Longitudinal Examination of Street Segment Patterns in Vancouver. *BC Journal of Quantitative Criminology*, **31**: 127-147.

DAS P & DAS AK. 2019. Graph-based clustering of extracted paraphrases for labelling crime reports. *Knowledge-Based Systems*, **179**: 55-76.

DAS P, DAS AK, NAYAK J, PELUSI D & DING W. 2019. Group incremental adaptive clustering based on neural network and rough set theory or crime report categorization. *Neurocomputing*, doi: <https://doi.org/10.1016/j.neucom.2019.10.109>.

DEN HEYER G. 2014. Examining police strategic resource allocation in a time of austerity. *Salus*, **2**(1).

FARIAS AMG, CINTRA ME & FELIX AC. 2018. Definition of Strategies for Crime Prevention and Combat Using Fuzzy Clustering and Formal Concept Analysis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **26**(3): 429-452.

FIGUEIREDO C & MOTA C. 2019. Learning Preferences in a Spatial Multiple Criteria Decision Approach: An Application in Public Security Planning. *International Journal of Information Technology & Decision Making*, **31**(4): 1403-1432.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. 2010. *Censo Demográfico: Brasil/Pernambuco*, Brazil.

IPEA - INSTITUTO DE PESQUISA ECONÔMICA APLICADA. 2019. *Atlas da Violência: Retrato dos Municípios Brasileiros*, Rio de Janeiro.

KHAN JR, SAEED M, SIDDIQUI FA, MAHMOOD N & UL ARIFEEN Q. 2019. Predictive Policing: A Machine Learning Approach to Predict and Control Crimes in Metropolitan Cities, *Journal of Information in Communication Technology*, **3**(1): 17-26.

DE LIMA FS & MARINHO E. 2017. Public security in Brazil: Efficiency and technological gaps. *Economia*, **18**(1): 129-145.

LOPES JMA. 2016. *Políticas de Segurança Pública nos Estados de Minas Gerais e Pernambuco em Perspectiva Comparada*. Dissertation Master in Sociology - UFPE, Recife.

MOTA CMM, FIGUEIREDO CJJ & PEREIRA DVS. 2020. Identifying areas vulnerable to homicide using multiple criteria analysis and spatial analysis. *Omega*, **100**.

- NEPOMUCENO TCC, SANTIAGO KTM, DARAIO C & COSTA APCS. 2020. Exogenous crimes and the assessment of public safety efficiency and effectiveness. *Annals of Operations Research*, 1-34.
- NÓBREGA JM. 2008. Barômetro da Violência e da Segurança na cidade do Recife. *Política Hoje*, **17**(1).
- OATLEY GC & EWART BW. 2003. Crimes analysis software: “pins in maps”, clustering and Bayes net prediction. *Expert Systems with Applications*, **25**(4): 569-588.
- PEREIRA DVS, MOTA CMM & ANDRESEN MA. 2016. The Homicide Drop in Recife, Brazil. *Homicide Studies*, **21**(1): 21-38.
- PRABAKARAN S & MITRA S. 2019. Design and development of machine learning algorithm for forecasting crime rate. *International Journal of Innovative Technology and Exploring Engineering*, **8**(11): 1217-1222.
- QAZI N & WONG BLW. 2019. An interactive human centered data science approach towards crime pattern analysis. *Information Processing & Management*, **56**(6).
- QI H, LI J, DI X, REN W & ZHANG F. 2019. Improved k-means clustering algorithm and its applications. *Recent Patents on Engineering*, **13**(4): 403-409.
- RATTON JL. 2014. *Governança da Segurança Pública e Redução dos Homicídios em Pernambuco: o caso do Pacto pela Vida*. Mimeo.
- SECRETARIAT OF SOCIAL DEFENSE OF THE STATE OF PERNAMBUCO (SDS-PE). 2019. *Crimes violentos contra o patrimônio: número de ocorrências de CVP em Pernambuco por município*, Jan-Out.
- SILVA C & RIBEIRO B. 2018. *Aprendizagem Computacional em Engenharia*. Coimbra: Coimbra University Press.
- SALTOS G & COCEA M. 2017. An Exploration of Crime Prediction Using Data Mining on Open Data. *International Journal of Information Technology & Decision Making*, **16**(5): 1155-1181.
- SANT’ANNA AP, COSTA HG & PEREIRA V. 2015. CPP-TRI: a sorting method based on the probabilistic composition of preferences. *Int. J. Information and Decision Sciences*, **7**(3).
- SAURET G. 2012. *Estatísticas pela Vida*. Recife: Editora Bagaço.
- TAYAL DK, JAIN A, ARORA S, AGARWAL S, GUPTA T & TYAGI N. 2014. Crime detection and criminal identification in India using data mining techniques. *AI & Society*, **30**(1): 117-127.
- WANG H, ZHAO Z, GUO Z, WANG Z & XU G. 2017. An improved clustering method for detection system of public security events based on genetic algorithm and semi-supervised learning. *Complexity*.

WANG S, WANG X, YE P, YUAN Y, LIU S & WANG F. 2018. Parallel Crime Scene Analysis Based on ACP Approach. *IEEE Transactions on Computational Social Systems*, **5**(1): 244-255.

WEISBURD D & AMRAM S. 2014. The law of concentrations of crime at place: The case of Tel Aviv-Jaffa. *Police Practice and Research*, **15**: 101-114.

WITTEN IH, FRANK E & HALL MA. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Amsterdam: Morgan Kaufmann Publishers, pp. 1-15.

WORTLEY R & MAZEROLLE L. 2008. *Environmental Criminology and Crime Analysis: Situation the theory, analytic approach and application*. Environmental criminology and crime analysis. London: Willian Publishers.

How to cite

COSTA JCOR & SILVA MM. 2022. A clustering-based approach for identifying groups of municipalities To support the direction of public security policies. *Pesquisa Operacional*, **42**: e257930. doi: 10.1590/0101-7438.2022.042.00257930.