engineering proceedings

MDPI

Proceedings Paper

Optimizing Police Locations around Football Stadiums Based on Multicriteria Unsupervised Clustering Analysis

Antonio Marcos de Lima * [1], Thyago C. C. Nepomuceno * [1,2], Isaac Pergher [1], Victor D. H. de Carvalho [3] and Thiago Poleto [4]

[1] Federal University of Pernambuco; antonio.marcoslima@ufpe.br; thyago.nepomuceno@ufpe.br; isaac.pergher@ufpe.br;;
[2] Sapienza University of Rome; nepomuceno@diag.uniroma1.it
[3] Federal University of Alagoas; victor.carvalho@delmiro.ufal.br
[4] Federal University of Para; thiagopoleto@ufpa.br
* Correspondence: thyago.nepomuceno@ufpe.br; antonio.marcoslima@ufpe.br
† Presented at The 4th International Electronic Conference on Applied Sciences, 27 Oct–10 Nov 2023.

Abstract: This work proposes a methodology based on Multicriteria Decision Aid (MCDA) and Cluster Analysis to identify ideal locations for the installation of police facilities or vehicle parking and policing around stadiums in Recife, Brazil, during potential violent sports events (criminal occurrences from football supporters or fanbases). K-Means unsupervised clustering algorithm is used to group criminal data into homogeneous clusters based on their characteristics. Each type of criminal occurrence is linked to a single cluster. The optimal location is addressed based on the PROMETHEE method (Preference Ranking Organization Method for Enrichment Evaluation), allowing clusters to be organized into a hierarchy based on the number of facilities (N), average distance (D) from the criminal occurrence to the associated cluster, and the coverage level (C) which is the proportion of crime occurring in a location less than 500m from the associated cluster. Through data analysis on crimes and violence in the region, the study seeks to identify patterns of criminal behaviour and high-risk areas to determine the most strategic location for the police units and enhance the public security decision-making process. The choice for the k parameters ranged from 1 to 30 incorporating all region of analysis, with computational cost of 43 minutes running time using Intel Core i3-3217U (1800GHz and 10 GB RAM). This approach and methodology can be useful to support public security policies in the region and contribute to the reduction of violence around the stadiums. The empirical application can help guide public managers' decisions regarding resource allocation and the implementation of more effective security policies, with the aim of ensuring a safer environment for fans and residents in the areas near the stadiums..

Keywords: Unsupervised Clustering Analysis; Multicriteria Decision Aid (MCDA); K-means; PROMETHEE; Violence; Football; Soccer; Crime; Police Location; Brazil

Citation: To be added by editorial staff during production.

Academic Editor: Alessandro Bruno

Published: date

1. Introduction

Football (soccer) is a social event that brings together people from different social, economic, and cultural backgrounds, representing a synthesis of multiple passions, objective and subjective determinations [1]. Nevertheless, not randomly such events result in episodes of violence and misdemeanors in stadiums and their surroundings, especially in countries with a deep football-rooted culture [2]. Areas close to sports venues are often the scene of conflicts between rival fans, in addition to vandalism, theft and other forms of violence. Some interesting works on the topic include [3,4] studies on the violent behavior of football supporters in Brazil.

Proper location of police units can be an important factor in preventing and reducing violence around stadiums. When there is a police unit close to the stadium, the police can act more quickly in case of conflicts or acts of violence. This can help ensure the safety of fans and residents in areas close to the stadium. Furthermore, police presence can have a deterrent effect on potential offenders, reducing the risk of violence. This work seeks to identify strategic locations for installing police facilities considering their quantity with a hybrid spatial and multicriteria analysis method. In this work, a police vehicle with up to 4 police officers is considered a police facility to provide more efficient and rapid responses to possible demands.

This work proposes a hybrid method using the PROMETHEE and K-means tools to identify possible strategic locations for installing police facilities depending on their quantity, improving public security and mitigating violence in regions around stadiums in Recife, Brazil. Such a hybrid method is has a well-founded statistical base and widely available in languages and software. The justification for using unsupervised clustering is the necessity for an exploratory analysis without predefined labels or categories. In the case of crime data in Recife, defining explicit categories or classes may be challenging, as criminal activities can be multifaceted and dynamic. In addition, unsupervised clustering methods are flexible and adaptable to different data types. In Recife, where several types of crimes are considered, this clustering approach identifies commonalities and differences across different crime categories and locations.

Next section offers a theoretical foundation for the employed methods. Section 3 is dedicated to results and discussion of the empirical application. Section 4 concludes the work with a summary of results and discussions.

2. Methods

Facility Location Problem (FLP) is a topic of great interest in areas such as logistics, transport, public service management and other areas where there is a need to decide where to install new service units or equipment. The objective is to decide in the best possible way where to locate a certain number of services to minimize costs and maximize the efficiency of the service provided [5].

FLP is a complex problem involving several factors, such as the distance between facilities and customers, customer characteristics, facility capabilities, and demand for services. The specialized literature presents several approaches to the solution of the problem, from mathematical models that seek the exact solution of the problem to heuristic algorithms that seek approximate solutions in a reasonable time. Many studies have considered methods to address the facility selection problem through the use of mathematical analysis and computer simulations [6].

The FLP solution can bring many benefits to society, such as improving the quality of services provided, reducing costs, optimizing resource distribution and improving users' well-being. Ideally, undesirable events (such as robberies, assaults, vandalism, etc.) should be as far away from society as possible, but we have no control over them. Therefore, bringing desirable services closer to society is of interest. Many quantitative tools and information technologies are adequate for addressing socioeconomic issues and the problem of public security [7–10]. For the present study, the facility location problem is addressed by combining K-means clustering with Multicriteria Decision Aid (PROMETHEE model).

2.1. K-means

K-means is a widely used unsupervised clustering algorithm for grouping data into homogeneous groups based on their characteristics. It is one of the most popular cluster classification algorithms due to its simplicity and efficiency [11].

This algorithm is iterative, starting with a random choice of k centroids representing each cluster's center. Then, each data point is assigned to the cluster whose centroid is closest. Once all points have been assigned to a group, the centroids are recalculated based on the data points assigned to each group. This process is repeated until there are no more changes in assigning data points to groups or until the maximum number of iterations is reached.

Although K-means is a simple and fast algorithm, it has some limitations, such as sensitivity to choosing the number of k groups and random initialization of centroids. However, there are techniques to deal with these problems, such as multiple random initialization and hierarchical sampling initialization. We combined this algorithm with the PROMEHTEE methodology for a more robust classification of vulnerable urban spaces.

2.1. PROMETHEE

The PROMETHEE method (Preference Ranking Organization Method for Enrichment Evaluation) is an approach that guides the decision-making process, allowing the options to be organized in a hierarchy based on the preferences of the decision-maker [12]. Developed by Brans et al. [13], this technique helps classify actions according to the established criteria, allowing a more precise and structured analysis of the available possibilities. Non-compensatory decision-making models such as PROMETHEE have been well used in the literature to evaluate job satisfaction and resource allocation [14,15].

The PROMETHEE method can be defined in two steps. In the first step, the comparison is made using preference functions of different types, such as difference functions, similarity functions and others. These comparisons generate a pre-order matrix that indicates the preference relationship between the alternatives in relation to each criterion individually. In a second step, the person responsible for decision-making assigns weights to the criteria according to priorities so that the most relevant criteria receive more significant weight.

It is necessary to assign weights to each criterion where each criterion i must have a weight p_i. Thus, the degree of overclassification given by the equation can be calculated:

$$\pi(a,b) = \sum_{i=1}^{n} p_i F_i(a,b) \tag{1}$$

Where:

n= decision criteria

$\pi(a,b)$ = outranking of alternative a over alternative b

F_i= preference function for criteria i

p_i= weight for criteria i

Based on the outranking degree of all alternatives, flows of alternatives can be calculated. There are three types of flow, positive, negative and net flow. The positive flow $\emptyset^{+}(a)$ measures how much an action is preferred over the other n-1 alternatives. Analogously, the negative flow $\emptyset^{-}(a)$ tells us how preferable the n-1 alternatives are. The net flow $\emptyset(a)$ subtracts the negative flow from the positive flow. This is basically how the alternative is performing compared to the others. The greater the net flow, the more preferable the alternative. The equations calculate the flows are:

$$\emptyset^{+}(a) = \frac{1}{n-1}\sum_{b\neq a} \pi(a,b) \tag{2}$$

$$\emptyset^{-}(a) = \frac{1}{n-1}\sum_{b\neq a} \pi(b,a) \tag{3}$$

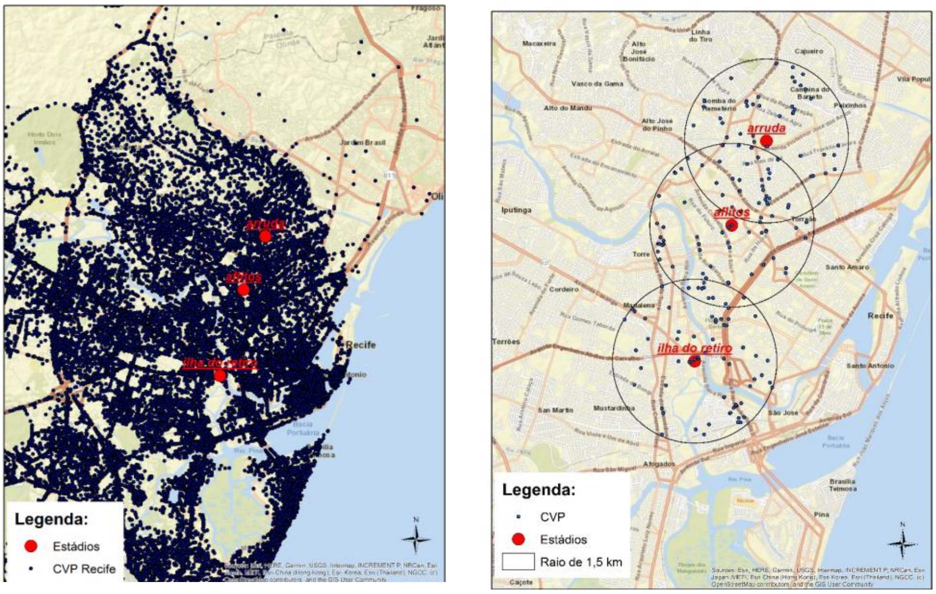
$$\emptyset(a) = \emptyset^{+}(a) - \emptyset^{-}(a) \tag{4}$$

3. Results and Discussion

The results for applying data-driven analytics and virtual learning technologies are crucial in optimizing resource provision and allocation with valuable insights into the socioeconomic landscape [7,9,16]. By analyzing diverse datasets, including crime patterns, demographic information, and economic indicators, policymakers can make informed decisions to address specific challenges and allocate resources effectively.

The SEPLAG (State Planning and Management Secretariat) made available a database with CVP cases (Property Crimes) and their respective geolocations throughout Pernambuco. This base covers cases from 2018 to 2021 across the state. Among the crimes present in this data are: thefts, theft, theft occurring in bank institutions, theft in public transport, mugging and robberies, theft of cargo, burglary, and commercial burglary. See Nepomuceno & Costa [17], De Carvalho and Costa [8] and Borba et al. [18] for a better description of those geographic-referenced crimes and the context where they operate.

Even not having other types of crimes such as physical, sexual, psychological or moral, the existence of robbery data in general can be a useful indicator of occurrences of violence in urban areas. This is because robberies are often associated with acts of violence, such as physical aggression, threats and intimidation. Figure 1 illustrates the plot of crimes in Recife through the ArcGis software using OpenStreetMap as a base map.



(**a**) (**b**)

Figure 1. CVP (property crimes) Occurrences: (**a**) Total occurrences in Recife; (**b**) Occurences around stadiums (up to 1.5 km).

The following table 1 reports the proportion of occurrences for each stadium based on the type of crime.

Table 1. Proportion of CVP (property crimes).

Types of CVP	Arruda	Ilha Do Retiro	Aflitos
Robberies	13,5%	2,9%	6,7%
Mugging	76,9%	88,4%	81,3%
Burglary	3,8%	1,4%	0,0%
Theft and Robberies in Bus	3,8%	0,0%	1,3%
Commercial Burglary	1,9%	5,9%	10,7%
Theft and Robberies in Public Transport	0%	1,4%	0%

In order to minimize the problem of the initial choice of centroids, the multiple random initialization technique was adopted. This technique is the default in the *Scikit* library. The algorithm is executed with seeds from different centroids according to the specified number of times. The seed that generated the clusters with the smallest sum of the squared distances of the points to the cluster centroid is chosen. For this study, each clustering ran with 5,000 different seeds. It was using a notebook equipped with an Intel(R) Core(TM) i3-3217U processor at a frequency of 1.80GHz and 10GB of RAM memory, taking one hour and 43 minutes to complete.

Scenarios were created to identify the best allocation of facilities based on the number of facilities (N), average distance (D) and coverage level (C). The weights for each of them were given by the ROC technique [19] considering the order of preference (under the criteria). The following Table 2 reports the number of facilities (locations) for the 5 top-ranked alternatives for each stadium considering all six scenarios: first scenario: (C, D, N), second scenario: (C, N, D), third scenario: (D, C, N), fourth scenario: (N, C, D), fith scenario: (D, N, C), sixth scenario: (N, D, C).

Table 2. Best Alternatives on each Scenario.

	Scenarios						
Stadiums	**1**	**2**	**3**	**4**	**5**	**6**	**Ranking**
Aflitos	30 Locations	11 Locations	30 Locations	7 Locations	26 Locations	1 Local	1º
	26 Locations	10 Locations	29 Locations	8 Locations	28 Locations	2 Locations	2º
	29 Locations	12 Locations	28 Locations	9 Locations	27 Locations	7 Locations	3º
	28 Locations	13 Locations	27 Locations	10 Locations	29 Locations	3 Locations	4º
	27 Locations	14 Locations	26 Locations	1 Locations	30 Locations	6 Locations	5º
Ilha do retiro	30 Locations	10 Locations	30 Locations	7 Locations	30 Locations	1 Local	1º
	29 Locations	11 Locations	29 Locations	6 Locations	29 Locations	2 Locations	2º
	28 Locations	12 Locations	28 Locations	8 Locations	26 Locations	6 Locations	3º
	27 Locations	9 Locations	27 Locations	9 Locations	28 Locations	7 Locations	4º
	26 Locations	14 Locations	26 Locations	5 Locations	27 Locations	3 Locations	5º
arruda	30 Locations	8 Locations	30 Locations	6 Locations	30 Locations	4 Locations	1º
	28 Locations	10 Locations	29 Locations	7 Locations	28 Locations	1 Local	2º
	29 Locations	11 Locations	28 Locations	5 Locations	29 Locations	5 Locations	3º
	27 Locations	7 Locations	27 Locations	4 Locations	27 Locations	6 Locations	4º
	26 Locations	9 Locations	26 Locations	8 Locations	26 Locations	7 Locations	5º

4. Conclusion

The results report interesting prospects considering three criteria and six options in order of preference (scenarios). Among the scenarios, only two obtained a balanced amount of facilities. The remaining scenarios are not recommended options for a decision-maker in any of the stages since these scenarios imply a high or insufficient number of facilities. Among the recommended scenarios at the Ilha do Retiro stadium, the second best alternative solved the problem of intersections with rivers.

The proposed method has the advantage of simultaneous consideration of multiple criteria, such as socioeconomic factors, demographic data, and crime rates. This integration enhances identifying patterns and relationships within crime data, aiding in comprehensive decision-making for law enforcement and policymakers. Combining unsupervised clustering with multicriteria decision aid addresses subjectivity by balancing subjective criteria with data-driven insights. It also assists in identifying trade-offs in crime prevention strategies and uncovers complex interactions within the data, leading to more effective crime analysis. This integration ensures a consistent and informed decision-making process in addressing crime challenges.

Applying this methodology in other Brazilian stadiums and considering other criteria not addressed in this work can be an interesting extension of the current analysis in addition to using data standardization to verify the impact on the generation of clusters. Furthermore, the possibility of applying a multicriteria system that serves multiple decision-makers can also add value for the current approach.

Author Contributions: Conceptualization, A.M.L. and T.C.C.N.; methodology, A.M.L. and T.C.C.N.; software, A.M.L. and T.C.C.N.; validation, A.M.L. and T.C.C.N.; formal analysis, A.M.L. and T.C.C.N.; investigation, A.M.L. and T.C.C.N.; resources A.M.L. and T.C.C.N.; data curation A.M.L. and T.C.C.N.; writing—original draft preparation, A.M.L. and T.C.C.N.; writing—review and editing A.M.L. and T.C.C.N.; visualization, A.M.L. and T.C.C.N.; supervision, T.C.C.N, I.P, V.D.H.C and T.P.; project administration, A.M.L. and T.C.C.N.; funding acquisition, T.C.C.N.; All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: This research received support from the Brazilian National Council for Scientific and Technological Development (CNPq).

Data Availability Statement: Data can be provided by request to the corresponding author.

Acknowledgments: We acknowledge the support from the Brazilian National Council for Scientific and Technological Development (CNPq).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Murad, M. (2007). A violência e o futebol: dos estudos clássicos aos dias de hoje. FGV Editora.
2. Palhares, M. F. S., Schwartz, G. M., Teruel, A. P., Santiago, D. R. P., & Trevisan, P. R. T. D. C. (2012). Lazer, agressividade e violência: considerações sobre o comportamento das torcidas organizadas. Motriz: Revista de Educação Física, 18, 186-199.
3. Nepomuceno, T. C. C., de Moura, J. A., e Silva, L. C., & Costa, A. P. C. S. (2017). Alcohol and violent behavior among football spectators: An empirical assessment of Brazilian's criminalization. International journal of law, crime and justice, 51, 34-44.
4. Nepomuceno, T. C. C., de Carvalho, V. D. H., Silva, L. C. E., de Moura, J. A., & Costa, A. P. C. S. (2022). Exploring the Bedouin syndrome in the football fan culture: Addressing the hooliganism phenomena through networks of violent behavior. International journal of environmental research and public health, 19(15), 9711.
5. Fernandes, S., Captivo, M. E., & Clımaco, J. (2005). Um sistema de apoioa decisao para análise de problemas de localizaçao bicritério. Working paper 2/05, Centro de Investigaçao Operacional, Faculdade de Ciências, Universidade de Lisboa.(in Portuguese).
6. Adeleke, O. J., & Olukanni, D. O. (2020). Facility location problems: models, techniques, and applications in waste management. Recycling, 5(2), 10.
7. Nepomuceno, T. C. C. (2023). Data-driven Analytics for Socioeconomic Challenges in a Contemporary World, Socioeconomic Analytics, Vol. 1, Issue 1, Pages 1 - 4. https://doi.org/10.51359/2965-4661.2023.259300.
8. de Carvalho, V. D. H., Costa, A. P. C. S. (2023). Exploring Text Mining and Analytics for Applications in Public Security: an in-depth dive into a Systematic Literature Review., *Socioeconomic Analytics*, Vol. 1, Issue 1, Pages 5 – 55. https://doi.org/10.51359/2965-4661.2023.259008.
9. Poleto, T., de Carvalho, V. D. H, de Oliveira, R. C. P. (2023). Applying Spatial Decision Support for Maternal Mortality Analysis in a Brazilian State. *Socioeconomic Analytics*, 1(1), 92-102. https://doi.org/10.51359/2965-4661.2023.259360.
10. Boezi, G., Cottarelli, A, Manimendra, I., Pagmozzi, D. (2023) The Efficiency of European Universities and the Role of Exogenous Factors. *Socioeconomic Analytics*, 1(1), 103-125.
11. Cabrera-Barona, P. F., Bayón, M., Durán, G., Bonilla, A., & Mejía, V. (2020). Generating and mapping Amazonian urban regions using a geospatial approach. ISPRS International Journal of Geo-Information, 9(7), 453.
12. Mareschal, B., & De Smet, Y. (2009, December). Visual PROMETHEE: Developments of the PROMETHEE & GAIA multicriteria decision aid methods. In 2009 IEEE International conference on industrial engineering and engineering management (pp. 1646-1649). IEEE.

13. Brans, J. P., Vincke, P., & Mareschal, B. (1986). How to select and how to rank projects: The PROMETHEE method. *European journal of operational research*, 24(2), 228-238.
14. Nepomuceno, T. C., Daraio, C., & Costa, A. P. C. (2021). Multicriteria ranking for the efficient and effective assessment of police departments. *Sustainability*, 13(8), 4251.
15. Nepomuceno, T. C. C., Daraio, C., & Costa, A. P. C. S. (2020). Combining multicriteria and directional distances to decompose non-compensatory measures of sustainable banking efficiency. Applied Economics Letters, 27(4), 329-334.
16. Ji, S., Dai, M., Wen, H., Zhang, H., Zhang, Z., Xia, Z., & Zhu, J. (2022). An improved virtual metrology method in chemical vapor deposition systems via multitask gaussian processes and adaptive active learning. *The International Journal of Advanced Manufacturing Technology*, 122(7-8), 3149-3159. https://doi.org/10.1007/s00170-022-10115-4
17. Nepomuceno, T. C. C., & Costa, A. P. C. S. (2019). Spatial visualization on patterns of disaggregate robberies. Operational Research, 19, 857-886.
18. Borba, B. F. D. C., de Gusmão, A. P. H., Clemente, T. R. N., & Nepomuceno, T. C. C. (2022). Optimizing Police Facility Locations Based on Cluster Analysis and the Maximal Covering Location Problem. Applied System Innovation, 5(4), 74.
19. Barron, F. H. (1992). Selecting a best multiattribute alternative with partial information about attribute weights. Acta psychologica, 80(1-3), 91-103.