

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

GUSTAVO RODRIGUES CARVALHO

**UMA ABORDAGEM DE COMPUTAÇÃO APLICADA PARA OTIMIZAÇÃO
ESPACIAL DE UNIDADES POLICIAIS**

CURITIBA

2026

GUSTAVO RODRIGUES CARVALHO

**UMA ABORDAGEM DE COMPUTAÇÃO APLICADA PARA OTIMIZAÇÃO
ESPACIAL DE UNIDADES POLICIAIS**

An Applied Computing Approach to the Spatial Optimization of Police Units

Dissertação apresentada como requisito para obtenção do título de Mestre em Computação Aplicada do Programa de Pós-Graduação em Computação Aplicada da Universidade Tecnológica Federal do Paraná.
Orientador(a): Profa. Dra. Nadia Puchalski Kozievitch.

CURITIBA

2026



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



**Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Curitiba**



GUSTAVO RODRIGUES CARVALHO

**UMA ABORDAGEM DE COMPUTAÇÃO APLICADA PARA OTIMIZAÇÃO ESPACIAL DE UNIDADES
POLICIAIS**

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Computação Aplicada da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Engenharia De Sistemas Computacionais.

Data de aprovação: 25 de Junho de 2026

Dra. Nadia Puchalski Kozievitch, Doutorado - Universidade Tecnológica Federal do Paraná

Dra. Rita Cristina Galarraga Berardi, Doutorado - Universidade Tecnológica Federal do Paraná

Dra. Veronica Isabela Quandt, Doutorado - Heilbronn University Of Applied Sciences (Hochschule Heilbronn)

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 14/07/2026.

RESUMO

O crescimento da criminalidade em grandes centros urbanos amplia os desafios relacionados ao planejamento territorial e à alocação eficiente de recursos de segurança pública. Nesse contexto, este trabalho propõe um *framework* metodológico para apoio à localização estratégica de unidades policiais, integrando bancos de dados geoespaciais, análise multicritério (MCDA) e algoritmos de agrupamento espacial. A abordagem combina processos de engenharia de dados, ponderação de variáveis urbanas e criminais por meio do *Analytic Hierarchy Process (AHP)* e agrupamento espacial. Como estudo aplicado, *K-means* foi aplicado com 554.832 registros criminais da região metropolitana de Londres, associados a informações espaciais e infraestrutura urbana. A validação foi conduzida com as métricas *Elbow*, *Silhouette* e *Calinski-Harabasz*, permitindo identificar agrupamentos espacialmente coerentes para suporte à tomada de decisão.

Palavras-chave: segurança pública; centros urbanos; análise espacial; criminalidade; mcda.

ABSTRACT

The growth of crime in large urban centers intensifies challenges related to territorial planning and the efficient allocation of public safety resources. In this context, this work proposes a methodological framework to support the strategic location of police facilities by integrating geospatial databases, multicriteria decision analysis (MCDA), and spatial clustering algorithms. The approach combines data engineering processes, weighting of urban and crime-related variables through the Analytic Hierarchy Process (AHP), and spatial clustering techniques. As an applied case study, K-means was applied to 554,832 crime records from the Greater London metropolitan area, associated with spatial and urban infrastructure data. Validation was performed using the Elbow, Silhouette, and Calinski-Harabasz metrics, enabling the identification of spatially coherent clusters to support decision-making.

Keywords: public safety; urban centers; spatial analysis; crime; mcda.

SUMÁRIO

1	INTRODUÇÃO	6
2	METODOLOGIA	9
3	CONCLUSÃO	10
	REFERÊNCIAS	12
	APÊNDICE A – Uma abordagem de computação aplicada para otimiza- ção espacial de unidades policiais	13

1 INTRODUÇÃO

A crescente complexidade dos problemas de segurança pública em grandes centros urbanos tem motivado autoridades e pesquisadores a buscar soluções baseadas em evidências para otimizar a alocação de recursos policiais. Estudos clássicos de *hotspot policing* demonstraram que a concentração de esforços em pequenas áreas de alta incidência criminal geram reduções significativas nos índices de crime da área total analisada, sem aumentar os custos operacionais de forma proporcional segundo Sathiyarayanan, Junejo e Fadahunsi (2019). Pesquisas mostram que 58% de todos os crimes acontecem nos 10% de lugares com os crimes mais graves¹. Nesse contexto, a aplicação de análise espacial e o uso de algoritmos analíticos vêm se consolidando como instrumentos-chave para subsidiar decisões estratégicas em segurança pública, permitindo a identificação de tendências espaciais, temporais e espaço-temporais da criminalidade, tais como a persistência ou deslocamento de *hotspots*, variações sazonais de determinados tipos de crime e padrões de concentração associados a características urbanas específicas, o apoio ao diagnóstico antecipado de crimes e a alocação eficiente de recursos, contribuindo para a inteligência das cidades (Mantovani *et al.*, 2024).

No plano internacional, cidades como Nova York, Los Angeles e Londres têm investido pesadamente em plataformas de inteligência geoespacial. A *Metropolitan Police Service (MPS)* de Londres, por exemplo, disponibiliza bancos de dados abertos que incluem registros de ocorrências criminais georreferenciadas². Essa abordagem integrada, baseada na combinação de dados abertos, ferramentas de Sistemas de Informação Geográfica, técnicas de análise espacial e métodos de visualização da informação, favorece a aplicação de técnicas de Visualização da Informação sobre esses dados, permitindo a criação de painéis e visualizações, como mapas de calor que apresentam a concentração dos atendimentos, auxiliando na tomada de decisão.

No Brasil, observa-se um movimento crescente em direção à adoção de metodologias quantitativas e geoespaciais no planejamento da segurança pública, ainda que sua implementação permaneça desigual, resultando em sobrecarga de algumas delegacias e subutilização de outras. Apesar de avanços pontuais na utilização de Sistemas de Informação Geográfica (SIG) e de ferramentas de *Visual Analytics* pelas forças policiais, a alocação de efetivos e recursos ainda é predominantemente reativa, com reduzido embasamento preditivo ou estratégico.

Estudos recentes indicam, contudo, o potencial dessas abordagens no contexto nacional. Alves *et al.* (2025) demonstram, por meio da aplicação do método de *K-means* aos dados da Secretaria Nacional de Segurança Pública (Senasp), que técnicas de agrupamento podem auxiliar na priorização territorial de políticas públicas ao identificar padrões espaciais de homicídios intencionais entre unidades federativas, considerando variáveis socioeconômicas e demográficas. O algoritmo *K-means* organiza os pontos de ocorrência em K grupos, minimizando a soma das distâncias quadráticas entre cada ponto e o respectivo centroide.

¹ <https://youthendowmentfund.org.uk/toolkit/hot-spots-policing/>

² <https://www.met.police.uk/police-forces/metropolitan-police/areas/stats-and-data/stats-and-data/>

Resultados semelhantes são observados em nível estadual. Couto (2024) utiliza técnicas de aprendizado de máquina e visualização de dados para identificar padrões geográficos e sociodemográficos associados ao roubo de celulares no Rio Grande do Sul, revelando a viabilidade do uso de dados abertos e métodos preditivos na formulação de estratégias locais de prevenção.

No campo da análise espacial e visualização, Neto (2020) propôs o *CrimeWatcher*, um sistema de *Visual Analytics* integrado à Secretaria Nacional de Segurança Pública, capaz de identificar *hotspots* criminais com base em estimativas de densidade *kernel* otimizadas (*MSKDE*). Essa proposta, em linha com o trabalho de Silva e Cestari (2022) sobre a Guarda Municipal de Curitiba, reforça a importância da integração de bases de dados abertos e ferramentas de análise visual para subsidiar decisões em tempo real e aprimorar a transparência das ações de segurança.

De forma complementar, Junior *et al.* (2017) desenvolve a aplicação *ROTA-Analytics*, que integra algoritmos preditivos ao planejamento de patrulhas em cidades inteligentes, exemplificando o potencial da modelagem espaço-temporal no apoio à gestão operacional da polícia. Esses resultados evidenciam que a combinação de modelos preditivos e dados espaciais pode orientar decisões de patrulhamento e instalação de novas unidades policiais de maneira mais eficiente e baseada em evidências.

Assim, a metodologia adotada neste estudo, originalmente aplicada ao contexto britânico segundo Junior *et al.* (2017), mostra-se transferível ao cenário brasileiro. A heterogeneidade dos dados nacionais e a existência de experiências bem-sucedidas em diferentes escalas (municipal, estadual e federal) indicam que o país reúne condições técnicas e institucionais para avançar em práticas de planejamento preditivo e territorial da segurança pública. A persistente limitação identificada por Junior *et al.* (2017) quanto à falta de padronização na coleta e tratamento de dados criminais, embora ainda relevante, vem sendo progressivamente mitigada pela ampliação das iniciativas de dados abertos e integração interinstitucional, que possibilitam análises cada vez mais robustas e comparáveis no tempo e no espaço.

Nesta direção, este estudo apresenta uma abordagem de computação aplicada baseada em *K-means* e análise espacial em banco de dados geográficos para a otimização da localização de unidades policiais. A abordagem *in-database* é adotada para minimizar o *overhead* de latência na transferência de grandes volumes de dados e garantir a integridade topológica das coordenadas espaciais durante o processamento analítico.

Do ponto de vista matemático, o agrupamento busca a partição de n observações em k *clusters*, visando minimizar a variância *intra-cluster*, ou a soma dos quadrados das distâncias euclidianas entre os pontos de ocorrência criminal (x_i) e seus respectivos centroides (μ_j), conforme a Equação 1:

$$J = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \quad (1)$$

Para mitigar a sensibilidade do algoritmo à inicialização dos *seeds* e à suposição de *clusters* aproximadamente esféricos, a robustez dos resultados é assegurada por uma **avaliação comparativa de métricas de validação** segundo Wang *et al.* (2020):

- **Método *Elbow* (Inércia):** Utilizado para identificar o ponto de inflexão onde o aumento do número de *clusters* deixa de proporcionar um ganho significativo de coesão.
- **Coeficiente de *Silhouette*:** Empregado para medir a qualidade da separação e o nível de confiança do pertencimento de cada ponto ao seu respectivo agrupamento.
- **Índice de *Calinski-Harabasz*:** Aplicado para avaliar a razão entre a dispersão *inter-cluster* e *intra-cluster*.

A pesquisa principal associada a esta dissertação está descrita no estudo presente no Apêndice A. O estudo está organizado da seguinte maneira: a Seção 2 apresenta a Fundamentação Teórica; a Seção 3 aborda a Revisão Sistemática de Literatura e a Seção 4 aborda a Metodologia. A Seção 5 apresenta um Caso de Uso, seguida pela Conclusão na Seção 6.

Esta pesquisa também produziu os artigos, listados a seguir, mas que não foram incluídos na íntegra como parte deste documento:

- “Análise de otimização de pontos para abertura de novas unidades policiais” apresenta uma análise de dados geoespaciais da região metropolitana de Londres, verificando a eficácia da distribuição de unidades policiais e sua adequação às necessidades de segurança pública, considerando padrões de criminalidade (Carvalho; Kozievitch; Zavadzki, 2025);
- “Clusterização Espacial In-Database para Planejamento Estratégico de Unidades Policiais” apresenta uma abordagem de clusterização espacial *in-database* para otimização da localização de unidades policiais a partir de dados georreferenciados em larga escala (Carvalho; Kozievitch, 2026);

2 METODOLOGIA

A alocação de recursos públicos de segurança em espaços urbanos de alta complexidade demanda métodos analíticos que superem a intuição administrativa e a simples análise de densidade bruta. Para tanto, esta pesquisa estrutura-se sobre uma abordagem híbrida que integra *MCD*A e algoritmos de Agrupamento Espacial Ponderado (*Weighted Spatial Clustering*).

Para fundamentar a escolha desta modelagem e mapear o estado da arte referente à otimização espacial em segurança pública, foi conduzida uma Revisão Sistemática da Literatura (RSL). O levantamento evidenciou que, embora as pesquisas atuais apresentem ampla maturidade na identificação de *hotspots* criminais, valendo-se majoritariamente de mapas de calor visuais e estimativas de densidade, há uma lacuna no desenvolvimento de *frameworks* capazes de integrar múltiplas variáveis urbanas para apoiar na tomada de decisão estratégica. Essa lacuna que direciona a estruturação da metodologia proposta.

A concepção desta metodologia desdobra-se em duas abordagens complementares. A metodologia genérica (teórica) propõe um *framework* matemático e computacional, aplicável a qualquer problema de locação de facilidades que exija a ponderação de múltiplos atributos contínuos ou discretos no espaço. Trata-se de um modelo que pressupõe dados confiáveis, homogeneidade de fluxo e isotropia espacial. Por outro lado, a metodologia aplicada (simplificada) constitui a transposição desse modelo teórico para a realidade operacional da Região Metropolitana de Londres. Nesta aplicação, lida-se com os limites da granularidade da informação oficial. A abordagem genérica é recomendada para o avanço da teoria da otimização espacial, enquanto a abordagem aplicada justifica-se pelos recursos disponíveis.

3 CONCLUSÃO

A distribuição desigual da criminalidade em grandes centros urbanos impõe desafios relevantes ao planejamento estratégico da segurança pública, especialmente em cenários caracterizados por elevada densidade populacional, limitações operacionais e necessidade de otimização da cobertura territorial. Nesse contexto, este trabalho apresentou uma abordagem aplicada baseada em bancos de dados geoespaciais, análise multicritério e algoritmos de agrupamento espacial para apoiar a localização estratégica de unidades policiais em ambientes urbanos.

Inicialmente, foi realizada uma revisão teórica envolvendo os eixos de Banco de Dados, Algoritmos e Cidades Inteligentes, complementada por uma Revisão Sistemática da Literatura voltada à análise espacial criminal, policiamento orientado por dados e uso de SIG em segurança pública. Os resultados evidenciaram o potencial das técnicas de análise geoespacial para identificação de *hotspots*, apoio à tomada de decisão e otimização da alocação de recursos policiais.

A metodologia proposta integrou processos de engenharia de dados, *MCDA* e agrupamento espacial. A utilização da abordagem *in-database* no caso de uso permitiu o processamento de grandes volumes de dados georreferenciados sem necessidade de exportação para ferramentas externas, preservando integridade espacial e escalabilidade computacional.

Como estudo aplicado, foram utilizados 554.832 registros criminais da região metropolitana de Londres. A aplicação das métricas *Elbow*, *Silhouette* e *Calinski-Harabasz* permitiu validar a definição de $K = 4$, revelando agrupamentos espacialmente coerentes com a distribuição criminal observada. Os centroides dos *clusters* foram interpretados como localizações potenciais para novas unidades policiais, fornecendo subsídios quantitativos para expansão da cobertura territorial e redução de lacunas operacionais.

Os resultados demonstram que a integração entre SIG, bancos de dados espaciais e algoritmos de agrupamento pode contribuir significativamente para o planejamento urbano orientado por evidências, especialmente em iniciativas relacionadas a Cidades Inteligentes e *smart policing*.

Entretanto, algumas limitações devem ser consideradas. O modelo adotou distâncias euclidianas simplificadas, desconsiderando restrições reais da malha viária, fluxo urbano e tempo efetivo de deslocamento. Além disso, a granularidade temporal dos dados limitou análises espaço-temporais mais detalhadas, enquanto inconsistências na precisão geográfica podem influenciar a formação dos agrupamentos. A definição experimental de três novas unidades policiais também não considera restrições orçamentárias e institucionais reais.

Outra limitação importante refere-se ao potencial viés algorítmico associado ao uso de registros históricos criminais. Modelos de agrupamento podem reproduzir padrões históricos de policiamento e reforçar desigualdades territoriais, fenômeno conhecido como *feedback loop* de policiamento. Dessa forma, as recomendações produzidas devem ser interpretadas como mecanismos de apoio à decisão, e não como determinismos espaciais automatizados.

Como trabalhos futuros, destacam-se diferentes possibilidades de ampliação metodológica e operacional. Inicialmente, recomenda-se a investigação de algoritmos alternativos de agrupamento espacial, como *DBSCAN*, *HDBSCAN* e *Spectral Clustering*, visando representar padrões criminais complexos, não lineares e espacialmente irregulares. Também se sugere a incorporação de análises baseadas em rede viária (*Network Constrained Analysis*) e geração de isócronas de deslocamento, permitindo estimativas mais realistas de cobertura operacional.

Adicionalmente, futuras pesquisas podem incorporar dados temporais de maior resolução para modelagem espaço-temporal, sazonalidade criminal e patrulhamento preditivo. Modelos clássicos de Pesquisa Operacional, como *P-Median Problem*, *LSCP* e *MCLP*, também podem ser integrados ao *framework* para considerar restrições reais de infraestrutura, orçamento e capacidade operacional.

Sugere-se ainda a realização de análises de sensibilidade relacionadas à qualidade da geocodificação e aos efeitos do *MAUP*, bem como validações quantitativas baseadas em métricas operacionais de cobertura e tempo de resposta. Finalmente, recomenda-se incorporar mecanismos de transparência, equidade algorítmica e governança de dados, buscando garantir que soluções computacionais aplicadas à segurança pública permaneçam alinhadas aos princípios de justiça social e planejamento urbano sustentável.

REFERÊNCIAS

- ALVES, H. J. P. *et al.* An application of the k-means clustering method to national public safety data: The case of intentional homicides. **Brazilian Journal of Biometrics**, v. 43, n. 1, p. 1–9, 2025.
- CARVALHO, G. R.; KOZIEVITCH, N. P. Clusterização espacial In-Database para planejamento estratégico de unidades policiais. **Revista de Sistemas e Computação**, v. 16, n. 1, p. 48–59, mai. 2026.
- CARVALHO, G. R.; KOZIEVITCH, N. P.; ZAVADZKI, L. S. Análise de otimização de pontos para abertura de novas unidades policiais. *In*: ANAIS DA XX ESCOLA REGIONAL DE BANCO DE DADOS. 2025, Porto Alegre, RS, Brasil. **Anais [...]** Porto Alegre, RS, Brasil: SBC, 2025. p. 161–164. ISSN 2595-413X. Disponível em: <https://sol.sbc.org.br/index.php/erbd/article/view/35420>.
- COUTO, L. O. Trabalho de Conclusão de Curso **Identificação de Padrões de Crime: Uma Abordagem Data-Driven para o Roubo de Celulares no RS**. Porto Alegre, 2024. Monografia de Bacharelado em Ciência da Computação.
- JUNIOR, A. A. *et al.* A predictive policing application to support patrol planning in smart cities. *In*: 2017 IEEE INTERNATIONAL SMART CITIES CONFERENCE (ISC2). 2017, Wuxi, China. **Anais [...]** Wuxi, China: IEEE, 2017. p. 1–8.
- MANTOVANI, D. M. N. *et al.* Inteligência em segurança pública municipal: modelos para previsão de crimes e planejamento de zeladoria urbana. *In*: ISLA 2024 PROCEEDINGS LATIN AMERICA (ISLA). 2024. **Anais [...]** [s.n.], 2024. p. 1–10. Disponível em: <https://aisel.aisnet.org/isla2024/14>.
- NETO, J. F. Q. **A Visual Analytics Approach for Geocoded Crime Data**. 2020. Tese (Doutorado) — Universidade Federal do Ceará Fortaleza 2020. Tese de Doutorado em Ciência da Computação.
- SATHIYANARAYANAN, M.; JUNEJO, A. K.; FADAHUNSI, O. Visual analysis of predictive policing to improve crime investigation. *In*: 2019 INTERNATIONAL CONFERENCE ON CONTEMPORARY COMPUTING AND INFORMATICS (IC3I). 2019. **Anais [...]** [S.l.: s.n.], 2019. p. 197–203.
- SILVA, G. M.; CESTARI, J. M. A. P. Visualização da informação com base de dados abertos da guarda municipal de curitiba. *In*: ISLA 2022 PROCEEDINGS LATIN AMERICA (ISLA). 2022. **Anais [...]** [s.n.], 2022. p. 1–10. Disponível em: <https://aisel.aisnet.org/isla2022/1>.
- WANG, J. *et al.* Crime analysis of spatial-temporal distribution based on knn algorithm. *In*: 2020 INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE AND EDUCATION (ICISE-IE). 2020. **Anais [...]** [S.l.: s.n.], 2020. p. 150–154.

APÊNDICE A – Uma abordagem de computação aplicada para otimização espacial de unidades policiais

Uma abordagem de computação aplicada para otimização espacial de unidades policiais

Gustavo R. Carvalho¹

¹PPGCA – Universidade Tecnológica Federal do Paraná (UTFPR)
Avenida Sete de Setembro – 3165 – 80.230-901 – Curitiba – PR – Brasil

`gustavorodriguescarvalho@alunos.utfpr.edu.br`

Abstract. *The growth of crime in large urban centers intensifies challenges related to territorial planning and the efficient allocation of public safety resources. In this context, this work proposes a methodological framework to support the strategic location of police facilities by integrating geospatial databases, multicriteria decision analysis (MCDA), and spatial clustering algorithms. The approach combines data engineering processes, weighting of urban and crime-related variables through the Analytic Hierarchy Process (AHP), and spatial clustering techniques. As an applied case study, K-means was applied to 554,832 crime records from the Greater London metropolitan area, associated with spatial and urban infrastructure data. Validation was performed using the Elbow, Silhouette, and Calinski-Harabasz metrics, enabling the identification of spatially coherent clusters to support decision-making.*

Keywords: *Public safety, urban centers, spatial analysis, crime, MCDA.*

Resumo. *O crescimento da criminalidade em grandes centros urbanos amplia os desafios relacionados ao planejamento territorial e à alocação eficiente de recursos de segurança pública. Nesse contexto, este trabalho propõe um framework metodológico para apoio à localização estratégica de unidades policiais, integrando bancos de dados geoespaciais, análise multicritério (MCDA) e algoritmos de agrupamento espacial. A abordagem combina processos de engenharia de dados, ponderação de variáveis urbanas e criminais por meio do Analytic Hierarchy Process (AHP) e agrupamento espacial. Como estudo aplicado, K-means foi aplicado com 554.832 registros criminais da região metropolitana de Londres, associados a informações espaciais e infraestrutura urbana. A validação foi conduzida com as métricas Elbow, Silhouette e Calinski-Harabasz, permitindo identificar agrupamentos espacialmente coerentes para suporte à tomada de decisão.*

Palavras-chave: *Segurança pública, centros urbanos, análise espacial, criminalidade, MCDA.*

1. Introdução

A crescente complexidade dos problemas de segurança pública em grandes centros urbanos tem motivado autoridades e pesquisadores a buscar soluções baseadas em evidências para otimizar a alocação de recursos policiais. Estudos clássicos de *hotspot policing* demonstraram que a concentração de esforços em pequenas áreas de alta incidência criminal geram reduções significativas nos índices de crime da área total analisada, sem

aumentar os custos operacionais de forma proporcional [Sathiyarayanan et al. 2019]. Pesquisas mostram que 58% de todos os crimes acontecem nos 10% de lugares com os crimes mais graves¹. Nesse contexto, a aplicação de análise espacial e o uso de algoritmos analíticos vêm se consolidando como instrumentos-chave para subsidiar decisões estratégicas em segurança pública, permitindo a identificação de tendências espaciais, temporais e espaço-temporais da criminalidade, tais como a persistência ou deslocamento de *hotspots*, variações sazonais de determinados tipos de crime e padrões de concentração associados a características urbanas específicas, o apoio ao diagnóstico antecipado de crimes e a alocação eficiente de recursos, contribuindo para a inteligência das cidades [Mantovani et al. 2024].

No plano internacional, cidades como Nova York, Los Angeles e Londres têm investido pesadamente em plataformas de inteligência geoespacial. A *Metropolitan Police Service (MPS)* de Londres, por exemplo, disponibiliza bancos de dados abertos que incluem registros de ocorrências criminais georreferenciadas². Essa abordagem integrada, baseada na combinação de dados abertos, ferramentas de Sistemas de Informação Geográfica, técnicas de análise espacial e métodos de visualização da informação, favorece a aplicação de técnicas de Visualização da Informação sobre esses dados, permitindo a criação de painéis e visualizações, como mapas de calor que apresentam a concentração dos atendimentos, auxiliando na tomada de decisão.

No Brasil, observa-se um movimento crescente em direção à adoção de metodologias quantitativas e geoespaciais no planejamento da segurança pública, ainda que sua implementação permaneça desigual, resultando em sobrecarga de algumas delegacias e subutilização de outras. Apesar de avanços pontuais na utilização de Sistemas de Informação Geográfica (SIG) e de ferramentas de *Visual Analytics* pelas forças policiais, a alocação de efetivos e recursos ainda é predominantemente reativa, com reduzido embasamento preditivo ou estratégico.

Estudos recentes indicam, contudo, o potencial dessas abordagens no contexto nacional. [Alves et al. 2025] demonstram, por meio da aplicação do método de *K-means* aos dados da Secretaria Nacional de Segurança Pública (Senasp), que técnicas de agrupamento podem auxiliar na priorização territorial de políticas públicas ao identificar padrões espaciais de homicídios intencionais entre unidades federativas, considerando variáveis socioeconômicas e demográficas. O algoritmo *K-means* organiza os pontos de ocorrência em *K* grupos, minimizando a soma das distâncias quadráticas entre cada ponto e o respectivo centróide.

Resultados semelhantes são observados em nível estadual. [Couto 2024] utiliza técnicas de aprendizado de máquina e visualização de dados para identificar padrões geográficos e sociodemográficos associados ao roubo de celulares no Rio Grande do Sul, revelando a viabilidade do uso de dados abertos e métodos preditivos na formulação de estratégias locais de prevenção.

No campo da análise espacial e visualização, [Neto 2020] propôs o *CrimeWatcher*, um sistema de *Visual Analytics* integrado à Secretaria Nacional de Segurança Pública, capaz de identificar *hotspots* criminais com base em estimativas de densidade *kernel* oti-

¹ <https://youthendowmentfund.org.uk/toolkit/hot-spots-policing/>

² <https://www.met.police.uk/police-forces/metropolitan-police/areas/stats-and-data/stats-and-data/>

mizadas (*MSKDE*). Essa proposta, em linha com o trabalho de [Silva and Cestari 2022] sobre a Guarda Municipal de Curitiba, reforça a importância da integração de bases de dados abertos e ferramentas de análise visual para subsidiar decisões em tempo real e aprimorar a transparência das ações de segurança.

De forma complementar, [Junior et al. 2017] desenvolve a aplicação *ROTA-Analytics*, que integra algoritmos preditivos ao planejamento de patrulhas em cidades inteligentes, exemplificando o potencial da modelagem espaço-temporal no apoio à gestão operacional da polícia. Esses resultados evidenciam que a combinação de modelos preditivos e dados espaciais pode orientar decisões de patrulhamento e instalação de novas unidades policiais de maneira mais eficiente e baseada em evidências.

Assim, a metodologia adotada neste estudo, originalmente aplicada ao contexto britânico [Junior et al. 2017], mostra-se transferível ao cenário brasileiro. A heterogeneidade dos dados nacionais e a existência de experiências bem-sucedidas em diferentes escalas (municipal, estadual e federal) indicam que o país reúne condições técnicas e institucionais para avançar em práticas de planejamento preditivo e territorial da segurança pública. A persistente limitação identificada por [Junior et al. 2017] quanto à falta de padronização na coleta e tratamento de dados criminais, embora ainda relevante, vem sendo progressivamente mitigada pela ampliação das iniciativas de dados abertos e integração interinstitucional, que possibilitam análises cada vez mais robustas e comparáveis no tempo e no espaço.

Nesta direção, este estudo apresenta uma abordagem de computação aplicada baseada em *K-means* e análise espacial em banco de dados geográficos para a otimização da localização de unidades policiais. A abordagem *in-database* é adotada para minimizar o *overhead* de latência na transferência de grandes volumes de dados e garantir a integridade topológica das coordenadas espaciais durante o processamento analítico.

Do ponto de vista matemático, o agrupamento busca a partição de n observações em k *clusters*, visando minimizar a variância *intra-cluster*, ou a soma dos quadrados das distâncias euclidianas entre os pontos de ocorrência criminal (x_i) e seus respectivos centróides (μ_j), conforme a Equação 1:

$$J = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \quad (1)$$

Para mitigar a sensibilidade do algoritmo à inicialização dos *seeds* e à suposição de *clusters* aproximadamente esféricos, a robustez dos resultados é assegurada por uma **avaliação comparativa de métricas de validação** [Wang et al. 2020]:

- **Método *Elbow* (Inércia):** Utilizado para identificar o ponto de inflexão onde o aumento do número de *clusters* deixa de proporcionar um ganho significativo de coesão.
- **Coefficiente de *Silhouette*:** Empregado para medir a qualidade da separação e o nível de confiança do pertencimento de cada ponto ao seu respectivo agrupamento.
- **Índice de *Calinski-Harabasz*:** Aplicado para avaliar a razão entre a dispersão *inter-cluster* e *intra-cluster*.

Este estudo está organizado da seguinte maneira: a Seção 2 apresenta a Fundamentação Teórica; a Seção 3 aborda a Revisão Sistemática de Literatura e a Seção

4 aborda a Metodologia. A Seção 5 apresenta um Caso de Uso, seguida pela Conclusão na Seção 6.

2. Fundamentação teórica

O escopo desta proposta concentra-se em três eixos de pesquisa principais: Cidades Inteligentes, Banco de Dados e Algoritmos.

2.1. Banco de dados

Os Sistemas Gerenciadores de Banco de Dados Espaciais (SGBD-E) constituem uma infraestrutura essencial nas arquiteturas de Cidades Inteligentes, especialmente no domínio da segurança pública. Ao atuarem como repositórios centrais, esses sistemas viabilizam o armazenamento, a indexação espacial e o processamento eficiente de grandes volumes de dados geoespaciais, servindo como camada integradora entre dados abertos, análise espacial e suporte à tomada de decisão estratégica.

Na literatura nacional, diversos trabalhos ilustram o potencial analítico dos bancos de dados espaciais. [Neto et al. 2020] desenvolveram o *CrimeWatcher*, um sistema de *Visual Analytics* integrado à Secretaria Nacional de Segurança Pública para identificação de *hotspots* criminais por meio de estimativas de densidade *kernel* otimizadas. [Carvalho et al. 2018] propuseram o *ROTA-Analytics*, explorando a integração entre banco de dados e algoritmos preditivos para o planejamento de patrulhas em cidades inteligentes. Da mesma forma, [Silva and Cestari 2022] utilizaram bancos de dados geoespaciais para análise da criminalidade junto à Guarda Municipal de Curitiba, enquanto [Alves et al. 2025] aplicaram técnicas de agrupamento sobre dados da Senasp para priorização territorial de políticas públicas e [Couto 2024] investigaram padrões geográficos e sociodemográficos associados ao roubo de celulares no Rio Grande do Sul.

Essa integração entre bancos de dados espaciais e SIG também é amplamente explorada na literatura internacional. [Eck et al. 2005] sintetizam as principais técnicas de SIG aplicadas à segurança pública, detalhando métodos de geração de mapas de densidade (*kernel density estimation*), interpolação e análise de vizinhança, além de enfatizarem a necessidade de cadências temporais para sistemas dinâmicos de mapeamento. [Silva et al. 2019] complementam ao detalhar como a integração entre sistemas de informação e bancos de dados espaciais permite o processamento automatizado de grandes volumes de dados, facilitando a extração de padrões e a sumarização de informações geoespaciais.

Sob a ótica da Teoria da Escolha Racional [Cornish and Clarke 1986], o infrator atua como um agente que avalia racionalmente riscos, recompensas e esforços antes da execução do delito. Nesse contexto, os bancos de dados espaciais vão além da mera função cartográfica: eles permitem identificar ambientes urbanos caracterizados por baixa vigilância natural, elevada acessibilidade e maior potencial de fuga, fatores que influenciam diretamente a percepção subjetiva de risco do ofensor e a atratividade criminal de determinados locais [Lilhare et al. 2024, Sheikh et al. 2017, Kim et al. 2020, Abhay et al. 2023].

Complementarmente, [Ratcliffe 2010] destacam a importância da capacitação de operadores de segurança no uso crítico de ferramentas de *crime mapping*, enquanto [Clarke 1997] evidenciam que a redistribuição espacial de unidades policiais, apoiada

por análise espacial, pode alterar os padrões de comportamento criminal ao modificar a percepção de presença estatal e os cálculos custo-benefício dos infratores.

A origem do conceito de *hotspot policing* remonta aos estudos de [Sherman et al. 1989], que demonstraram empiricamente que uma pequena parcela de locais concentra a maior parte dos crimes predatórios. Utilizando a Teoria das Atividades Rotineiras que concebe o crime como a convergência de um infrator, um alvo e a ausência de um guardião, os autores evidenciaram que o policiamento direcionado a *hotspots* reduz significativamente a criminalidade sem aumento proporcional de recursos. Esse marco impulsionou toda a pesquisa em análise espacial de criminalidade e o desenvolvimento de sistemas de mapeamento.

Complementarmente, a Teoria dos Padrões Criminais de [Brantingham and Brantingham 1993] aprofunda a dimensão espacial do fenômeno, argumentando que a distribuição das ocorrências não é aleatória, mas condicionada pelos trajetos cotidianos dos ofensores e pelas características ambientais urbanas. Ao introduzirem os conceitos de *nodes* (pontos de atividade), *paths* (rotas) e *edges* (fronteiras), os autores permitem compreender como corredores urbanos, centros comerciais, estações de transporte e zonas de transição geram oportunidades delitivas ao diminuírem o esforço e o risco da ação criminosa. Nesse cenário, os Sistemas de Informação Geográfica assumem um papel analítico estratégico, indo além da simples identificação de concentrações para a interpretação das dinâmicas ambientais favoráveis ao crime.

No presente trabalho, o banco de dados espacial atua como núcleo central da arquitetura proposta, integrando os registros georreferenciados da *Metropolitan Police Service*, e as localizações das unidades policiais existentes, viabilizando o pré-processamento, a filtragem espacial e a clusterização necessárias à otimização da localização de novas unidades.

2.2. Algoritmos

Do ponto de vista dos métodos, que incluem técnicas de agrupamento espacial, análise de densidade e identificação de padrões espaciais, o algoritmo *K-means*, originalmente descrito por [MacQueen 1967], surge como uma opção simples e escalável para agrupar grandes volumes de dados georreferenciados. A proposta consistia em particionar um conjunto de observações em K grupos, minimizando a soma das distâncias quadráticas até seus centróides, ajustados iterativamente. Embora assuma *clusters* de forma aproximadamente esférica, o *K-means* permanece amplamente utilizado em análises de criminalidade pela sua eficiência computacional e facilidade de implementação em bancos de dados geoespaciais, como o *PostGIS*.

Quando aplicados à análise criminal, algoritmos de agrupamento espacial também podem contribuir para a identificação de estruturas urbanas recorrentes associadas à dinâmica dos delitos. Sob a perspectiva da criminologia ambiental, agrupamentos detectados por métodos como *K-means* frequentemente correspondem a áreas de convergência de fluxos urbanos, regiões de elevada circulação ou proximidade entre pontos de interesse criminalmente relevantes, características associadas aos chamados *activity nodes* descritos por [Brantingham and Brantingham 1981]. Dessa maneira, o processo de clusterização não se limita à segmentação estatística dos dados, mas auxilia na interpretação de como padrões de mobilidade urbana influenciam a concentração espacial

das ocorrências.

Para validar quantitativamente as soluções de *clustering*, duas métricas clássicas se destacam na literatura. [Rousseeuw 1987] introduziu o coeficiente de *Silhouette*, que compara, para cada ponto, a distância média ao próprio *cluster* com a distância média ao *cluster* mais próximo, gerando um índice que varia de -1 (grupo mal definido) a +1 (grupo bem separado). A *Silhouette* fornece uma medida intuitiva de coesão interna e separação externa, sendo particularmente útil para comparar diferentes valores de K . Adicionalmente, o método *Elbow* é frequentemente empregado para fundamentar a escolha técnica do parâmetro K [Parcianello et al. 2021], identificando o ponto de inflexão onde o aumento do número de *clusters* não reduz significativamente a variância total ou a soma dos quadrados dos agrupamentos. Já [Calinski and Harabasz 1974] propuseram um índice baseado na razão entre a variância *inter-cluster* e a variância *intra-cluster* (também chamado de *Variance Ratio Criterion*), segundo o qual valores mais altos indicam maior distinção entre os grupos formados. Ambos os índices são amplamente empregados em estudos de criminalidade para orientar a escolha do número apropriado de *clusters*.

Do ponto de vista operacional, a definição adequada da quantidade de agrupamentos possui implicações diretas na formulação de estratégias de policiamento orientado por dados. Agrupamentos excessivamente amplos podem mascarar microdinâmicas territoriais relevantes, enquanto segmentações demasiadamente fragmentadas dificultam a identificação de padrões persistentes de oportunidade criminal. Conforme argumenta [Felson 2006], pequenas variações no ambiente urbano podem alterar significativamente a percepção de vulnerabilidade e vigilância, tornando essencial que os modelos computacionais consigam representar escalas espaciais compatíveis com a dinâmica cotidiana dos ofensores e das vítimas.

Estudos recentes destacam a relevância de diferentes técnicas de *clustering* para a análise de dados complexos e de alta dimensionalidade. Como detalhado no tutorial fundamental de [v. Luxburg 2007], o *Spectral Clustering* se baseia na construção de grafos de similaridade para transformar o problema de agrupamento, permitindo a identificação de estruturas não lineares e agrupamentos não convexos. De forma complementar, [Hunt and Reffert 2021] compararam algoritmos como *DBSCAN* e *HDBSCAN*, evidenciando que o *DBSCAN* é eficiente na identificação de *clusters* com formas arbitrárias e na detecção de ruídos, enquanto o *HDBSCAN* aprimora essa abordagem ao ajustar automaticamente parâmetros e oferecer maior flexibilidade na detecção hierárquica de agrupamentos. Juntas, essas análises, baseadas em métodos de *spectral clustering*, técnicas de agrupamento por densidade e em estudos comparativos de algoritmos, reforçam a necessidade de selecionar métodos adequados ao tipo, à densidade e à estrutura dos dados, especialmente em cenários de análise espacial urbana e criminal, onde os padrões podem ser irregulares, multifacetados e de distribuição heterogênea.

Nesse cenário, métodos capazes de identificar agrupamentos não lineares tornam-se particularmente relevantes para estudos de segurança pública, uma vez que a distribuição espacial do crime tende a acompanhar redes viárias, corredores comerciais e trajetos urbanos recorrentes, em vez de padrões geométricos regulares. Conforme discutido por [Brantingham and Brantingham 2008], a configuração espacial das cidades influencia diretamente a formação de oportunidades criminais ao conectar ofensores, alvos e rotas de deslocamento. Assim, algoritmos capazes de detectar densidades variáveis e

estruturas espaciais complexas contribuem para representar de maneira mais fiel a lógica ambiental associada à ocorrência criminal.

Do ponto de vista computacional, além da análise exploratória, outros métodos, como modelos de regressão e séries temporais já foram utilizados para a análise de crime [Leal and Gomes-Jr 2022], por exemplo, para prever a quantidade de crimes em determinadas regiões e períodos do ano, utilizando o modelo *ARIMA*.

2.3. Cidades inteligentes

As iniciativas de Cidades Inteligentes (*Smart Cities*) representam um paradigma contemporâneo de governança urbana que busca integrar tecnologias de informação, análise de dados em larga escala e participação cidadã para resolver desafios complexos, entre os quais se destaca a segurança pública. Nesse contexto, a integração entre SIG, bancos de dados espaciais e abordagens de criminologia ambiental ganha relevância estratégica, permitindo a passagem de um modelo reativo de policiamento para uma atuação preditiva, preventiva e territorialmente otimizada.

A literatura recente evidencia que a aplicação de tecnologias geoespaciais em ambientes de cidades inteligentes tem produzido avanços significativos na prevenção criminal. Estudos demonstram que a combinação de dados abertos, sensoriamento urbano e análise espacial possibilita não apenas a identificação de *hotspots*, mas também a simulação de impactos de intervenções urbanísticas sobre os padrões de criminalidade [Mansourihanis et al. 2025, Kim et al. 2025]. No âmbito internacional, plataformas de *smart policing* têm utilizado SIG integrado a modelos preditivos para otimizar a alocação de recursos policiais, reduzir tempos de resposta e aumentar a percepção de segurança da população.

No Brasil, observa-se um movimento crescente de adoção dessas abordagens. Pesquisas destacam o potencial da integração entre SIG, *IoT* e análise de dados para fortalecer a inteligência de segurança pública em diferentes escalas municipais e estaduais [Palacios et al. 2020, Callegaro et al. 2022]. Tais iniciativas alinham-se ao conceito de *smart governance*, no qual a coprodução de conhecimento entre poder público, academia e sociedade civil fortalece a legitimidade e a efetividade das políticas de segurança.

Entretanto, a implementação de soluções de cidades inteligentes em segurança pública não está isenta de desafios. Questões relacionadas à privacidade de dados, viés algorítmico, desigualdade digital e capacitação de agentes públicos ainda demandam atenção crítica da comunidade científica. Ademais, a efetividade dessas tecnologias depende da qualidade e da integração dos dados, bem como da capacidade institucional de traduzir análises geoespaciais em ações concretas de planejamento territorial.

No presente trabalho, a perspectiva de Cidades Inteligentes orienta a proposta metodológica ao enfatizar a integração nativa entre banco de dados espacial, análise multicritério e algoritmos de otimização para subsidiar a localização estratégica de unidades policiais na região metropolitana de Londres, contribuindo para o campo da governança urbana baseada em evidências.

3. Revisão Sistemática da Literatura

O presente estudo foi conduzido como uma Revisão Sistemática da Literatura (RSL), que

*é um tipo de pesquisa que analisa estudos de um tópico específico para encontrar uma evidência geral.*³

Para direcionar o escopo da revisão, foram utilizadas as seguintes perguntas de pesquisa:

1. Onde ocorrem os crimes com maior frequência?
2. Quais são os padrões temporais na ocorrência de crimes?
3. Como o SIG e dados geoespaciais podem ser usados para análise e previsão de crimes?
4. Quais são os desafios associados à visualização e previsão de crimes?

Para esta RSL, utilizamos a fonte de dados *IEEE Xplore*⁴, em idioma inglês, com a seguinte string de busca:

((“patrolling” OR “policing” OR “law enforcement” OR “crime” OR “delict”) AND (“geoprocessing” OR “geolocation” OR “spatial analysis” OR “geographic data processing” OR “GIS” OR “geospatial analysis” OR “spatial modeling” OR “location tracking” OR “GPS tracking” OR “spatial location” OR “geographic positioning” OR “location detection”)).

Os critérios foram definidos com base no objetivo da revisão, representados na Tabela 1.

Tabela 1. Critérios

Inclusão	
CI1	Estudos que respondem às perguntas de pesquisa
CI2	Estudos publicados a partir de 2014
Exclusão	
CE1	Estudos duplicados
CE2	Indisponibilidade do texto completo
CE3	Livros, editoriais, tutoriais, painéis, sessões de pôsteres, prefácios, opiniões, resumos, cartas, apresentação de slides, relatórios técnicos, trabalho com menos de 5 páginas, ou qualquer trabalho classificado como literatura cinza

A busca na plataforma *IEEE* resultou em 1685 artigos, os quais foram ordenados por relevância. Os dados obtidos foram organizados em uma planilha, na qual foram aplicados filtros baseados em critérios previamente definidos. Para os artigos selecionados, seus resumos foram traduzidos e submetidos à leitura inicial. Após essa triagem, 23 artigos foram considerados adequados para leitura integral, de acordo com os critérios estabelecidos.

3.1. Análises e Resultados

Vinte e três artigos foram selecionados para a análise completa, satisfazendo os critérios Tabela 2.

³<https://library.leeds.ac.uk/info/1404/literature-searching/15/systematic-reviews>

⁴<https://ieeexplore.ieee.org/Xplore/home.jsp>

Tabela 2. Artigos Selecionados

Nº	Título do Artigo	Autor	Ano
1	<i>Crime Analysis of spatial-temporal distribution based on KNN Algorithm</i>	[Wang et al. 2020]	2020
2	<i>IST: Role of GIS in crime mapping and analysis</i>	[Sheikh et al. 2017]	2017
3	<i>Crime Hotspots Mapping and FIR Data Interface</i>	[Lilhare et al. 2024]	2024
4	<i>A predictive policing application to support patrol planning in smart cities</i>	[Junior et al. 2017]	2017
5	<i>Fuzzy Based Geo-Spatial Crime Category Prediction for Crime Mapping and Safe Route Travel</i>	[Sharma et al. 2022]	2022
6	<i>Comparative Analysis Of Crime Hotspot Detection And Prediction Using Convolutional Neural Network Over Support Vector Machine with Engineered Spatial Features Towards Increase in Classifier Accuracy</i>	[Sravani and Suguna 2022]	2022
7	<i>Impact of Spatial Correlation on Crime Prediction in Communities with Different Crime Densities</i>	[Dong et al. 2022]	2022
8	<i>A Comprehensive Review on Crime Patterns and Trends Analysis using Machine Learning</i>	[Ratra et al. 2023]	2023
9	<i>Micro-Level Incident Analysis using Spatial Association Rule Mining</i>	[Yoo et al. 2019]	2019
10	<i>Prediction of Crime in Neighbourhoods of New York City using Spatial Data Analysis</i>	[Almuhanna et al. 2021]	2021
11	<i>Crime Detection Using Data Mining Techniques</i>	[Sayal et al. 2024]	2024
12	<i>Smart Policing: Using Geospatial Crime Data to Plan Patrol Routes</i>	[Abhay et al. 2023]	2023
13	<i>Space-Time Variation of Property Crime in Beijing with ESDA Method</i>	[Li and Tang 2019]	2019
14	<i>Crime sequencing: Fighting crime with mathematics and technology</i>	[Steier et al. 2017]	2017
15	<i>Time series analysis for crime forecasting</i>	[Borowik et al. 2018]	2018
16	<i>Crime Prediction Using Auto Regression Techniques for Time Series Data</i>	[Yadav and Sheoran 2018]	2018
17	<i>Visual Analysis of Predictive Policing to Improve Crime Investigation</i>	[Sathiyarayanan et al. 2019]	2019
18	<i>Bandwidth Selection of Kernel Density Estimation for GIS-based Crime Occurrence Map Visualization</i>	[Kim et al. 2020]	2020
19	<i>Predicting High-Risk Areas for Crime Hotspot Using Hybrid KNN Machine Learning Framework</i>	[Kim et al. 2025]	2025
20	<i>The Role of Machine Learning in Crime Analysis and Prediction</i>	[Vadav et al. 2024]	2024
21	<i>An Approach on Cyber Crime Prediction Using Prophet Time Series</i>	[Nag et al. 2022]	2022
22	<i>Geographical crime rate prediction</i>	[Jain and Riyank 2023]	2023
23	<i>Geographical Crime Rate Prediction System</i>	[Tarlekar et al. 2021]	2021

Considerando (**Pergunta 1**) o artigo de [Sathiyarayanan et al. 2019] que identifica *hotspots* criminosos por meio de análises espaciais, revelando áreas com alta concentração de crimes. Esse método é corroborado pelos artigos de [Wang et al. 2020] e [Steier et al. 2017], que destacam como a análise de densidade e agrupamento podem evidenciar essas áreas de alta incidência. Além disso, o artigo de [Tarlekar et al. 2021] reforça a ideia de que a análise espacial é crucial para detectar e entender a frequência

dos crimes em diferentes regiões. Já para identificar e mapear áreas de alta incidência de crimes, os artigos de [Wang et al. 2020], [Steier et al. 2017] e [Tarlekar et al. 2021] discutem técnicas para mapear *hotspots*, como a estimativa de densidade de *Kernel (KDE)* e análise de agrupamento. O artigo de [Wang et al. 2020] foca na visualização dos dados criminais, enquanto o artigo de [Steier et al. 2017] detalha como o *KDE* pode destacar áreas críticas. O artigo de [Tarlekar et al. 2021] amplia essa perspectiva ao combinar diversas técnicas analíticas para uma visualização completa das áreas com alta incidência de crimes. O artigo de [Steier et al. 2017] observa que criminosos tendem a seguir um padrão que se estabelece entre 1 a 5 milhas de suas residências, com o último local de crime frequentemente situado próximo de sua residência. Esse padrão é apoiado pelos artigos de [Sravani and Suguna 2022], [Ratra et al. 2023] e [Li and Tang 2019], que discutem como a análise de movimento dos criminosos e padrões espaciais ajuda a prever futuras ocorrências baseadas na proximidade dos locais de crime. E quanto às áreas que são mais propensas a crimes, o artigo de [Steier et al. 2017] identifica áreas próximas a lojas de bebidas, transportes públicos e lojas de conveniência como mais propensas a crimes. O artigo de [Sathiyarayanan et al. 2019] complementa essa visão ao mostrar que a concentração de crimes também está associada à presença desses estabelecimentos em *hotspots* identificados. O artigo de [Tarlekar et al. 2021] reforça essa relação ao mapear a incidência de crimes em áreas com alta atividade comercial. O artigo de [Steier et al. 2017] sugere que criminosos provavelmente utilizam transporte público para facilitar a fuga. Essa ideia é corroborada pelo artigo de [Shenoy and Auradkar 2023], que analisa como o acesso a transportes públicos pode influenciar a mobilidade dos criminosos e suas escolhas de locais para cometer crimes. O artigo de [Steier et al. 2017] recomenda aumentar as patrulhas em áreas de alta criminalidade e utilizar câmeras para monitorar suspeitos. Essa recomendação é apoiada pelo artigo [Tarlekar et al. 2021], que sugere o uso de dados para orientar a alocação de recursos policiais em pontos críticos e melhorar a eficácia das operações de prevenção.

Ao se tratar de padrões temporais na ocorrência de crimes (**Pergunta 2**), os artigos de [Wang et al. 2020], [Dong et al. 2022] e [Nag et al. 2022] identificam padrões de sazonalidade e variações diárias na ocorrência de crimes. O artigo [Dong et al. 2022] usa dados geográficos para mostrar como esses padrões se manifestam ao longo do tempo, enquanto o artigo de [Nag et al. 2022] detalha como esses padrões podem variar com base em fatores sazonais e horários específicos. Os artigos de [Sravani and Suguna 2022], [Yoo et al. 2019] e [Nag et al. 2022] indicam que crimes tendem a ocorrer em locais de alta atividade durante horários específicos, com picos noturnos e finais de semana. O artigo de [Yoo et al. 2019] oferece uma análise detalhada sobre como esses horários influenciam a localização dos crimes, alinhando-se com os padrões temporais identificados nos artigos de [Sravani and Suguna 2022] e [Nag et al. 2022]. Já com relação a previsão de crimes futuros, os artigos de [Sheikh et al. 2017], [Dong et al. 2022], [Li and Tang 2019], [Nag et al. 2022] e [Jain and Riyank 2023] exploram como combinar dados espaciais e temporais para criar modelos preditivos. O artigo de [Li and Tang 2019] destaca a eficácia dessa combinação na previsão de áreas e horários de alto risco, enquanto o artigo de [Jain and Riyank 2023] analisa a influência desses padrões na criação dos modelos. Considerando a remoção de *outliers* e a inclusão de feriados nas previsões, o artigo de [Nag et al. 2022] avalia como há *outliers* e feriados no modelo, considerando esta abordagem crucial para ajustar modelos e garantir que as previsões considerem todas as variáveis

relevantes, como os efeitos de eventos sazonais e datas comemorativas.

Os artigos de [Sheikh et al. 2017], [Lilhare et al. 2024], [Shenoy and Auradkar 2023] e [Kim et al. 2020] mostram a importância de SIG e análise espacial na identificação e análise de *hotspots* de crimes (**Pergunta 3**). Os artigos de [Sheikh et al. 2017] e [Shenoy and Auradkar 2023] detalham como essas ferramentas ajudam a mapear a distribuição dos crimes, enquanto o artigo de [Kim et al. 2020] demonstra como a visualização de dados geoespaciais pode auxiliar nas políticas de segurança pública. Os artigos de [Junior et al. 2017] e [Dong et al. 2022] mostram como a análise espacial é essencial para direcionar recursos de forma eficaz e melhorar a resposta a áreas de maior risco. E [Kim et al. 2020] sugere que a visualização de dados, como mapas de densidade e pontos, é crucial para a formulação de políticas e estratégias de segurança pública. Essas visualizações ajudam a identificar padrões e alocar recursos de forma mais eficaz, conforme detalhado pelos artigos de [Sheikh et al. 2017] e [Shenoy and Auradkar 2023].

Ao considerar desafios (**Pergunta 4**), itens como a qualidade dos dados, a diversidade de fontes e questões éticas na análise de dados criminais é mencionada [Sharma et al. 2022][Sayal et al. 2024] [Borowik et al. 2018]. Nesta direção, modelos preditivos são essenciais para identificar áreas de alto risco e planejar patrulhas mais eficazes [Junior et al. 2017][Almuhanna et al. 2021] [Shenoy and Auradkar 2023] [Steier et al. 2017]. O artigo de [Almuhanna et al. 2021] detalha como essas previsões podem direcionar a alocação de recursos, enquanto o artigo de [Steier et al. 2017] recomenda estratégias para aumentar a eficácia do policiamento em áreas identificadas como de alta criminalidade. Algoritmos de *machine learning*, como *Random Forest*, *KNN*, *ARIMA* e modelos de aprendizado profundo podem ser utilizados [Sravani and Suguna 2022] [Almuhanna et al. 2021] [Sayal et al. 2024] [Vadav et al. 2024] [Jain and Riyank 2023] [Tarlekar et al. 2021]. O artigo de [Sayal et al. 2024] compara a eficácia desses algoritmos, mostrando que *Random Forest* e métodos de aprendizado profundo frequentemente oferecem maior precisão em comparação com outros métodos. Porém, estes algoritmos têm limitações, como a qualidade dos dados, ajuste de hiperparâmetros e utilização de técnicas avançadas [Sravani and Suguna 2022] [Dong et al. 2022] [Borowik et al. 2018], essas abordagens visam aumentar a precisão e a eficácia dos modelos preditivos de crimes. E quanto a alocação estratégica de recursos policiais, [Almuhanna et al. 2021] [Shenoy and Auradkar 2023] [Vadav et al. 2024] [Vadav et al. 2024], a previsão é um método mais utilizado. O artigo de [Vadav et al. 2024] oferece diretrizes práticas para a alocação de recursos, enquanto o artigo de [Vadav et al. 2024] analisa como utilizar dados preditivos para otimizar a resposta policial e a prevenção de crimes.

Os 23 artigos foram classificados entre cinco temas de aplicação, como ilustra a Tabela 3. Os artigos também apresentam estudos empíricos e teóricos Tabela 4, divididos em três abordagens metodológicas Tabela 5: Análise espacial e Visualização de Dados, *Machine Learning* e Algoritmos de Previsão e Análise Temporal e Sazonalidade. As áreas de aplicação encontradas foram Segurança Pública e Policiamento e Tecnologias e Ferramentas de Análise Tabela 6.

As quatro tabelas apresentadas a seguir nos auxiliam a elucidar, como se classificam os artigos com base nas características em comum identificadas, sendo elas: **temas**

e áreas de aplicação representadas pela Tabela 3, tipos de estudos contidos na Tabela 4, abordagens metodológicas listadas na Tabela 5 e áreas de aplicação na Tabela 6.

Tabela 3. Temas e áreas de aplicação

Temas e Áreas de Aplicação	Artigos	Descrição
Identificação e Mapeamento de <i>Hotspots</i> de Crimes	1, 3, 4, 6, 8, 12, 13, 14, 17, 19, 23	Estes artigos concentram-se na identificação de áreas com alta incidência de crimes e na aplicação de técnicas para mapear <i>hotspots</i> . Eles abordam o uso de técnicas de análise espacial, visualização de dados, e métodos de <i>clustering</i> para identificar e mapear áreas críticas.
Padrões Temporais e Espaciais dos Crimes	1, 2, 4, 6, 7, 9, 10, 13, 14, 15, 21, 22	Esta categoria abrange estudos que investigam como os crimes ocorrem em relação ao tempo e ao espaço. Inclui a análise de sazonalidade, variações diárias, e padrões de crime em diferentes horários e épocas do ano.
Métodos e Algoritmos de Previsão	1, 4, 6, 7, 10, 11, 15, 16, 19, 20, 21, 22	Artigos que discutem diferentes algoritmos e métodos de <i>machine learning</i> aplicados à previsão de crimes. Incluem comparações entre algoritmos e análises sobre como a engenharia de características e o pré-processamento afetam o desempenho dos modelos.
Utilização de Dados Geoespaciais e SIG	2, 3, 4, 6, 7, 9, 12, 18	Foca na aplicação de SIG e dados geoespaciais na análise e previsão de crimes. Inclui a utilização de SIG para mapear e analisar a distribuição de crimes e a importância das características geográficas na previsão.
Desafios e Limitações na Previsão de Crimes	5, 6, 7, 11, 15, 16, 22	Aborda os desafios e limitações encontradas na visualização e previsão de crimes. Inclui questões relacionadas à qualidade dos dados, diversidade de fontes, e problemas éticos na análise de dados.
Planejamento e Alocação de Recursos Policiais	4, 10, 12, 14, 16, 20	Explora como modelos preditivos podem ser usados para otimizar o planejamento de patrulhamento e a alocação de recursos policiais. Inclui estratégias para melhorar a eficiência e eficácia das operações policiais.

Foram identificadas algumas categorias e subcategorias recorrentes nos artigos conforme representa a Tabela 7.

3.2. Discussão

Os artigos revisados oferecem uma visão abrangente sobre a identificação de *hotspots* de crimes, destacando a importância das técnicas de análise espacial e agrupamento para mapear áreas com alta incidência de criminalidade. O artigo de [Steier et al. 2017] destaca

Tabela 4. Tipos de estudos

Tipos de Estudos	Artigos	Descrição
Estudos Empíricos	1, 2, 3, 6, 7, 8, 12, 13, 14, 15, 19, 21, 22	Artigos que baseiam suas conclusões em dados reais e experimentos. Incluem estudos de caso, análises de dados e experimentos práticos para validar teorias e métodos.
Estudos Teóricos	4, 5, 9, 10, 11, 16, 20	Focam na discussão de teorias, modelos e conceitos relacionados à previsão e análise de crimes. Incluem revisões de literatura, propostas de novos modelos e discussões conceituais sobre métodos e algoritmos.

Tabela 5. Abordagens metodológicas

Abordagens Metodológicas	Artigos	Descrição
Análise Espacial e Visualização de Dados	1, 3, 4, 6, 12, 13, 14, 17, 18	Utilizam técnicas de análise espacial e visualização para mapear hotspots e identificar padrões de crimes. Incluem métodos como <i>KDE</i> e análise de <i>clustering</i> .
<i>Machine Learning</i> e Algoritmos de Previsão	1, 4, 6, 7, 10, 11, 15, 16, 19, 20, 21, 22	Aplicam algoritmos de <i>machine learning</i> para prever a ocorrência de crimes. Incluem técnicas como <i>Random Forest</i> , <i>KNN</i> , <i>ARIMA</i> , e modelos de aprendizado profundo.
Análise Temporal e Sazonalidade	1, 2, 7, 9, 10, 13, 14, 15, 21	Investigam como os crimes variam ao longo do tempo, considerando padrões sazonais e horários específicos. Incluem a análise de variações diárias e sazonais.

Tabela 6. Áreas de aplicação

Áreas de Aplicação	Artigos	Descrição
Segurança Pública e Policiamento	4, 10, 12, 14, 16, 20	Focam na aplicação dos resultados da previsão de crimes para melhorar o policiamento e alocação de recursos. Incluem estratégias para otimização de patrulhamento e resposta policial.
Tecnologias e Ferramentas de Análise	2, 3, 4, 6, 7, 12, 18	Discussão sobre o uso de tecnologias como SIG e ferramentas de visualização para análise de crimes. Incluem a aplicação de tecnologias emergentes na análise espacial.

que a análise de padrões cronológicos e a proximidade com a residência dos criminosos pode revelar informações valiosas para a prevenção de crimes. Além disso, a descoberta

Tabela 7. Categorias

Categoria	Subcategoria	Quantidade de Artigos
Métodos e Algoritmos	<i>Machine Learning</i>	8
	Análise Estatística	5
Análise Espacial	SIG e Dados Geoespaciais	7
	Visualização de Dados	4
Desafios e Limitações	Qualidade dos Dados	6
	Questões Éticas	3

de que áreas comerciais tendem a ter mais crimes do que áreas residenciais sugere a necessidade de intervenções específicas. É importante notar que a literatura ainda carece de estudos que integrem completamente dados espaciais e temporais para prever e entender melhor a dinâmica dos *hotspots* criminosos.

Os padrões temporais e espaciais de crimes, abordados em artigos como [Wang et al. 2020], [Dong et al. 2022], e [Nag et al. 2022], revelam variações significativas ao longo do tempo e em diferentes horários. No entanto, ainda há uma lacuna significativa na integração desses padrões para construir modelos preditivos mais robustos. Artigos como [Sheikh et al. 2017] e [Li and Tang 2019] indicam que a combinação de dados espaciais e temporais pode melhorar a previsão de futuros crimes, mas a aplicação prática dessa combinação ainda é um desafio. A literatura sugere que a inclusão de variáveis temporais e a consideração de eventos sazonais são áreas emergentes que precisam de mais desenvolvimento para aprimorar a eficácia das previsões.

A eficácia dos métodos e algoritmos de previsão de crimes é amplamente discutida em artigos como [Sravani and Suguna 2022], [Almuhanna et al. 2021], e [Jain and Riyank 2023]. A comparação entre diferentes algoritmos, como *Random Forest* e *KNN*, revela que métodos mais avançados, como aprendizado profundo, frequentemente superam outros em termos de precisão conforme [Sayal et al. 2024]. No entanto, a literatura ainda precisa de uma comparação mais abrangente e sistemática dos diferentes algoritmos. A seleção do valor de K no *KNN*, conforme sugerido pelos artigos de [Wang et al. 2020] e [Junior et al. 2017], e a engenharia de características conforme [Jain and Riyank 2023] são cruciais para o desempenho dos modelos preditivos, mas muitas vezes são negligenciadas.

O uso de SIG e dados geoespaciais é fundamental para a análise e previsão de crimes, conforme discutido nos artigos de [Sheikh et al. 2017], [Lilhare et al. 2024], [Shenoy and Auradkar 2023], e [Kim et al. 2020]. No entanto, a literatura revela desafios na integração de dados geoespaciais com outras variáveis e na visualização eficaz desses dados para decisões estratégicas. Artigos como [Lilhare et al. 2024] e [Kim et al. 2020] destacam a importância da visualização avançada de dados, como mapas dinâmicos, para a formulação de políticas de segurança pública.

Os modelos preditivos desempenham um papel crucial no planejamento e na alocação de recursos policiais, como mostrado nos artigos de [Junior et al. 2017], [Almuhanna et al. 2021], [Shenoy and Auradkar 2023], e [Steier et al. 2017]. No entanto, a literatura ainda não fornece uma abordagem detalhada sobre como alocar recursos com base nas previsões de forma eficiente. A integração de diferentes fontes de dados e

a otimização da alocação de recursos são áreas emergentes que necessitam de mais pesquisa para aprimorar as estratégias de policiamento conforme [Almuhanna et al. 2021], [Yadav and Sheoran 2018] e [Vadav et al. 2024].

Os desafios associados à visualização e previsão de crimes, conforme discutido em artigos como [Sharma et al. 2022], [Sayal et al. 2024], e [Borowik et al. 2018], incluem questões relacionadas à qualidade dos dados, diversidade de fontes e considerações éticas. Embora haja avanços significativos, a literatura ainda carece de estudos que abordem como superar limitações nos modelos existentes. Artigos de [Sravani and Suguna 2022], [Dong et al. 2022], e [Borowik et al. 2018] sugerem que melhorias na qualidade dos dados e ajustes nos hiperparâmetros são essenciais para superar essas limitações. Além disso, questões éticas relacionadas à privacidade e ao uso responsável dos dados, como destacado pelo artigo de [Sayal et al. 2024], são uma preocupação crescente que precisa ser abordada com mais rigor.

4. Metodologia

A alocação de recursos públicos de segurança em espaços urbanos de alta complexidade demanda métodos analíticos que superem a intuição administrativa e a simples análise de densidade bruta. Para tanto, esta pesquisa estrutura-se sobre uma abordagem híbrida que integra MCDA e algoritmos de Agrupamento Espacial Ponderado (*Weighted Spatial Clustering*).

A concepção desta metodologia desdobra-se em duas abordagens complementares. A metodologia genérica (teórica) propõe um *framework* matemático e computacional representado pela Figura 1, aplicável a qualquer problema de locação de facilidades que exija a ponderação de múltiplos atributos contínuos ou discretos no espaço. Trata-se de um modelo que pressupõe dados confiáveis, homogeneidade de fluxo e isotropia espacial. Por outro lado, a metodologia aplicada (simplificada) constitui a transposição desse modelo teórico para a realidade operacional da Região Metropolitana de Londres (caso de uso aplicado na Seção 5). Nesta aplicação, lida-se com os limites da granularidade da informação oficial. A abordagem genérica é recomendada para o avanço da teoria da otimização espacial, enquanto a abordagem aplicada justifica-se pelos recursos disponíveis.

O *framework* proposto Figura 1 possui cinco etapas: 1) Dados de Entrada; 2) Engenharia de Dados; 3) Modelo multicritério; 4) Validação e Avaliação e 5) Análise de Cobertura da Rede Viária.

4.1. Dados de entrada

A assertividade de um modelo de otimização espacial é fundamentalmente atrelada à qualidade, resolução e natureza dos dados de entrada (*inputs*). Para capturar a multidimensionalidade do fenômeno criminal e do risco urbano, vários tipos de dados podem ser usados, como eventos criminais, severidades de crimes, densidade populacional, privação socioeconômica, infraestrutura de transporte, infraestrutura sensível (escola, hospital) e unidades policiais existentes.

Variáveis como densidade populacional podem ser combinadas ao *Index of Multiple Deprivation (IMD)*, que quantifica a privação relativa nas áreas de habitação, emprego e saúde, atuando como uma *proxy* para a vulnerabilidade social.

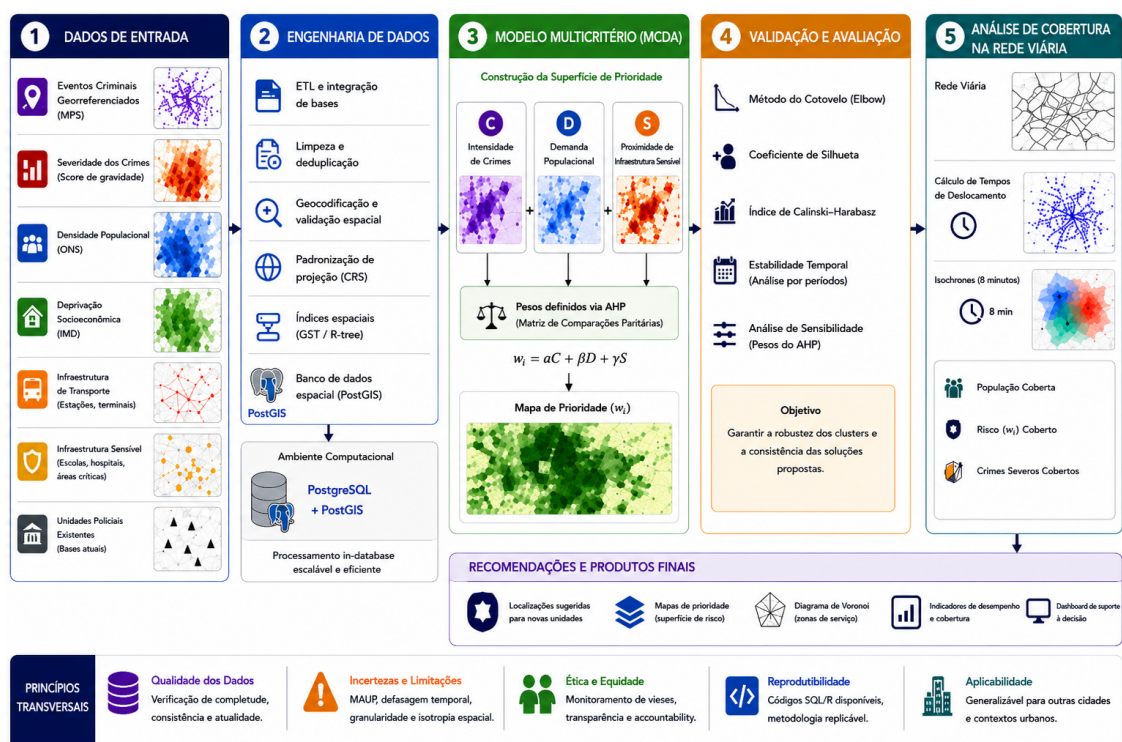


Figura 1. Framework Metodológico para otimização espacial de unidades policiais.

Também podem ser incorporados dados de infraestrutura urbana e segurança nacional, contemplando a localização de *hubs* de transporte, bases militares e edifícios críticos, que representam tanto nós de alta circulação quanto ativos de alto risco que demandam vigilância constante.

É importante reconhecer o impacto da qualidade e da resolução geométrica desses dados no modelo. A agregação espacial por *Lower Layer Super Output Areas (LSOA)*, por exemplo, introduz o Problema da Unidade de Área Modificável (*MAUP - Modifiable Areal Unit Problem*), onde os limites administrativos arbitrários podem mascarar variações intrassetoriais de criminalidade. Além disso, a utilização de dados censitários podem possuir defasagem temporal, visto que a dinâmica criminal sofre mutações mais rápidas do que os ciclos de atualização demográfica.

4.2. Engenharia de Dados

A etapa de Engenharia de Dados é responsável pela preparação e organização das informações utilizadas no modelo analítico. Nessa fase, são realizados processos de *ETL (Extract, Transform and Load)*, integração de diferentes bases de dados, limpeza e deduplicação de registros, além da padronização espacial por meio de sistemas de referência cartográfica (*CRS*). Também são executadas etapas de geocodificação, validação espacial e criação de índices geoespaciais para otimizar consultas e análises. Eventualmente os dados podem ser armazenados em um banco de dados (como *PostgreSQL/PostGIS*).

4.3. Modelagem multicritério

A MCDA, neste contexto, não é apenas um método de ranqueamento, mas uma ferramenta de tradução espacial: ela converte variáveis díspares em um único índice numérico contínuo, denotando a “gravidade” ou a “necessidade” de cada unidade espacial.

Para estruturar este modelo, cada célula espacial i recebe um peso final w_i , calculado a partir de variáveis independentes selecionadas pelo modelo. Por exemplo, poderia-se escolher intensidade criminal (C_i), densidade populacional agregada (D_i) e um fator de proximidade a infraestruturas sensíveis (S_i). Todos os valores de entrada podem sofrer padronização estatística (como o método *Z-Score*), permitindo a comparabilidade entre escalas dissimilares (por exemplo, contagem de crimes versus metros de distância). A normalização segue a formulação:

$$z_i = \frac{x_i - \mu}{\sigma} \quad (2)$$

Onde x_i representa o valor observado na área i , μ denota a média da variável em todo o território de estudo, e σ corresponde ao desvio padrão.

Para a atribuição dos pesos relativos a cada variável, métodos como Processo Hierárquico Analítico (*AHP - Analytic Hierarchy Process*), Entropia (*Entropy Weight Method*) ou Ponderação Igualitária (*Equal Weighting*) poderiam ser utilizados. Embora métodos alternativos como a ponderação por Entropia reduzam a subjetividade do tomador de decisão, ou a ponderação igualitária simplifique o modelo, o *AHP* justifica-se por permitir a calibração do modelo com base no conhecimento de especialistas em segurança pública, equilibrando a importância relativa entre densidade populacional e manchas criminais estruturais.

Considerando as três variáveis citadas anteriormente, o índice espacial resultante, ou peso da observação (w_i), seria dado pela equação:

$$w_i = \alpha \cdot C_i + \beta \cdot D_i + \gamma \cdot S_i \quad (3)$$

Onde α , β e γ representam os vetores de prioridade derivados da matriz de comparação paritária do *AHP* (atendendo ao limiar de Razão de Consistência aceitável), e C_i representa intensidade criminal, D_i representa densidade populacional agregada e S_i um fator de proximidade a infraestruturas sensíveis.

4.3.1. K-means

Com os dados espaciais modelados e ponderados, a etapa de localização das unidades policiais pode ser resolvida como um problema de partição não supervisionada. O algoritmo tradicional *K-means* busca minimizar as distâncias euclidianas geométricas entre os pontos de um conjunto e os centros dos grupos formados. Entretanto, para responder à topologia de risco calculada, aplicou-se a extensão geométrica do *Weighted K-means*.

Neste cenário, cada observação espacial (x_i) atrai o centro de gravidade do *cluster* não apenas pela sua posição, mas de acordo com o peso de demanda (w_i) construído

pela MCDA. O objetivo do algoritmo passa a ser a minimização da seguinte função custo global (J):

$$J = \sum_{j=1}^k \sum_{i \in S_j} w_i \cdot ||x_i - \mu_j||^2 \quad (4)$$

Nesta formulação, μ_j é o centróide ponderado do *cluster* j , que atuará como a coordenada sugerida para a alocação de uma nova unidade operacional. O problema da localização espacial de facilidades possui uma extensa literatura na pesquisa operacional, exigindo a justificativa da adoção do agrupamento ponderado em detrimento de outros algoritmos clássicos.

4.3.2. Outras Opções na Literatura

Algoritmos de densidade não paramétricos, como *DBSCAN* e *HDBSCAN*, são reportados na literatura forense para a extração de áreas de mancha criminal (*hotspots* de formatos irregulares). No entanto, eles são limitados para alocação de recursos físicos, visto que seu resultado é uma região morfológica, e não um ponto de coordenação exato, além de não permitirem fixar previamente o número de unidades (o parâmetro K).

Por outro lado, abordagens puramente oriundas da Pesquisa Operacional, como o Problema das P-Medianas (*P-Median Problem*), o *Location Set Covering Problem (LSCP)* ou o *Maximal Covering Location Problem (MCLP)*, oferecem soluções ótimas globais de alocação de redes. Contudo, em problemas com milhões de pontos de demanda contínuos, os problemas combinatórios clássicos tendem à explosão computacional (NP-Difícil). O *Weighted K-means*, operando como uma meta-heurística eficiente, permite processar todo o tecido da região metropolitana de Londres em tempo exequível, gerando soluções sub-ótimas que podem convergir matematicamente de forma estável.

4.4. Validação e Avaliação

Esta etapa tem como objetivo verificar a robustez e a consistência das soluções geradas pelo modelo multicritério. Para isso, podem ser utilizadas métricas de análise de agrupamentos, incluindo o Método do Cotovelo (*Elbow Method*), o Coeficiente de *Silhouette* e o Índice de *Calinski-Harabasz*, permitindo avaliar a qualidade e a separação dos *clusters* espaciais identificados.

Para a validação, também pode ser integrado o modelo de Cobertura Máxima (*MCLP*). Proposto por [Church and ReVelle 1974], é um modelo de otimização espacial cujo objetivo consiste em selecionar um conjunto limitado de instalações de forma a maximizar a cobertura de áreas de demanda dentro de uma determinada distância ou tempo de atendimento. O modelo é amplamente utilizado em problemas de localização-alocação, apoio à decisão e planejamento urbano, especialmente em cenários que envolvem restrições operacionais e recursos limitados.

Adicionalmente, pode ser realizada uma análise temporal para verificar a estabilidade das regiões prioritárias ao longo do tempo, bem como uma análise de sensibilidade dos pesos definidos pelo método *AHP*, visando avaliar o impacto de variações nos critérios

sobre os resultados obtidos. Essas etapas contribuem para aumentar a confiabilidade, a interpretabilidade e a reprodutibilidade da metodologia proposta.

4.5. Análise de Cobertura da Rede

A etapa de análise da cobertura viária tem como objetivo avaliar a acessibilidade e o alcance operacional das áreas prioritárias identificadas pelo modelo multicritério. Para isso, são utilizadas métricas espaciais e técnicas de análise em redes viárias, considerando distâncias, conectividade urbana e tempo estimado de deslocamento. Adicionalmente, o modelo de Cobertura Máxima (*MCLP*) pode ser empregado para identificar localizações estratégicas capazes de maximizar a cobertura das regiões prioritárias, contribuindo para o apoio à tomada de decisão em ambientes urbanos.

5. Caso de uso - Dados da região metropolitana de Londres

Esta Seção apresenta um caso de uso prático, utilizando dados abertos da cidade de Londres, com o objetivo de identificar os três locais mais adequados para a instalação de novas unidades policiais. As etapas utilizadas são apresentadas na Seção 4. O processo envolve a coleta e o tratamento de dados abertos, a aplicação de filtros espaciais, técnicas de clusterização e o cálculo de centróides para sugerir pontos estratégicos de cobertura.

Embora a metodologia proposta esteja estruturada em cinco etapas complementares, a aplicação prática apresentada nesta pesquisa constitui uma adaptação simplificada condicionada à disponibilidade, granularidade e limitações dos dados públicos oficiais da *Metropolitan Police Service (MPS)*. Dessa forma, determinadas etapas foram aplicadas parcialmente, enquanto outras foram apenas discutidas conceitualmente, preservando-se, contudo, a coerência metodológica do modelo proposto.

5.1. Dados de Entrada

A base de dados utilizada para a pesquisa refere-se a dados de crimes individuais e incidentes de comportamentos antissociais (*ASB Anti-social Behaviour*) da região metropolitana de Londres, atendida pela *Metropolitan Police Service (MPS)*⁵, incluindo informações de localização em nível de rua e os resultados subsequentes de ações policiais e judiciais associados ao crime.

O segundo tipo de dado, correspondente às unidades policiais, foi obtido a partir da base de *amenities*⁶, especificamente dos registros relacionados a estabelecimentos classificados como unidades de segurança pública, totalizando 921 unidades. Esses dados foram importados para o banco de dados por meio do *QGIS 3.22.4 64 bits*⁷, considerando a delimitação espacial definida para a área de estudo.

Embora a metodologia proposta contemple a utilização de múltiplas variáveis urbanas, socioeconômicas e territoriais, nesta aplicação prática foram utilizados predominantemente os dados criminais e as localizações das unidades policiais existentes, em razão da maior disponibilidade, padronização e consistência estatística dessas bases públicas.

⁵<https://data.police.uk/data/fetch/0a319eae-59b1-47d6-adcf-b485883123b3>

⁶<https://overpass-api.de/>

⁷<https://www.qgis.org/project/visual-changelogs/visualchangelog322>

5.2. Engenharia de Dados

Na etapa de importação e armazenamento, foram utilizados o *PostgreSQL 17.2*⁸, o *PostGIS 3.5*⁹, e o *QGIS*.

Para dados de crimes foi utilizado o período de Agosto-2023 a Agosto-2024, totalizando 554.832 registros. A Tabela 8 apresenta cada coluna contida na tabela referente aos crimes. Os dados de unidades policiais totalizaram 921 registros. Os dados fornecidos são anonimizados, motivo pelo qual as ocorrências são agrupadas em meses e o tipo de crime é enquadrado em categorias. A localização passa por um processo de aproximação geométrica ao nível de rua (*street-level*): as coordenadas reais são intencionalmente deslocadas (em uma escala de dezenas a centenas de metros) e vinculadas a pontos de ancoragem (*snap points*) predefinidos em vias públicas. Esse procedimento de *geomasking* garante que nenhuma ocorrência seja associada a um endereço específico, exigindo uma vizinhança mínima de oito propriedades postais para que o ponto seja validado.

Tabela 8. Colunas da tabela

Coluna	Descrição
<i>id</i>	Identificador incremental único de cada registro, gerado via planilha
<i>crime_id</i>	Identificador do crime fornecido pela base de dados (nem sempre há valor)
<i>month</i>	Mês e ano em que a ocorrência foi registrada
<i>longitude</i>	<i>Longitude</i> aproximada de onde o crime ocorreu
<i>latitude</i>	<i>Latitude</i> aproximada de onde o crime ocorreu
<i>location</i>	Nome da localização aproximada em que o crime ocorreu
<i>lsoa_code</i>	Código de referência à <i>LSOA</i> na qual o ponto anonimizado se enquadra
<i>lsoa_name</i>	Nome de referência à <i>LSOA</i> na qual o ponto anonimizado se enquadra
<i>crime_type</i>	Um dos tipos de crimes listados nas Perguntas Frequentes do <i>Police.UK</i> .
<i>last_outcome_category</i>	Uma referência ao resultado associado ao crime ocorrido mais recentemente
<i>geom</i>	Geocodificação gerada a partir da <i>latitude</i> e <i>longitude</i>

Os registros de dados de crimes foram inseridos, indexados, geocodificados e atualizados para a projeção cartográfica *EPSG:3395 (WGS 84/Pseudo Mercator)* na base de dados, utilizando o *SQL* do Anexo A. Considerando a limitação de gastos e a otimização de recursos em cidades, o objetivo era responder à seguinte pergunta: **Quais são as três localizações mais adequadas para a instalação de uma nova unidade policial, considerando a distribuição e a quantidade de crimes na região?**

As etapas relacionadas à integração avançada de múltiplas fontes heterogêneas, automação completa de *ETL* e modelagem contínua de atualização de dados foram apli-

⁸<https://www.postgresql.org/docs/17/index.html>

⁹<https://postgis.net/2024/09/PostGIS-3.5.0>

cadadas parcialmente, principalmente porque as bases utilizadas já se encontravam previamente estruturadas em formatos compatíveis com SIG.

5.3. Modelagem Multicritério

O sistema de pesos do Reino Unido é reconhecido como um dos mais sofisticados do mundo: o *Crime Severity Score (CSS)* ¹⁰. A polícia de Londres organiza os incidentes criminais em dois níveis hierárquicos: *Major Crime Types* (Categorias Principal) e *Minor Crime Types* (Subcategorias). Apresentamos essa estrutura na Tabela 9.

Tabela 9. Categorização Oficial de Crimes da *Metropolitan Police Service (MPS)*

Categoria Principal (<i>Major</i>)	Subcategorias (<i>Minor</i>)
Violência Contra a Pessoa	Homicídio, Violência com Lesão Corporal, Violência sem Lesão Corporal.
Furtos	Furto de Pessoa, Furto em Lojas, Furto de Bicicletas, Outros Furtos.
Crimes Relacionados a Veículos	Furto ou Roubo de Veículo, Furto de Objetos em Veículo, Tomada Agravada de Veículo, Interferência em Veículo.
Invasão e Arrombamento	Residencial, Comercial e Comunitário.
Roubo	Roubo de Bens Pessoais, Roubo de Propriedade Comercial.
Crimes Sexuais	Estupro, Outros Crimes Sexuais.
Incêndio Criminoso e Dano	Incêndio Criminoso, Dano ao Patrimônio.
Infrações Relacionadas a Drogas	Tráfico de Drogas, Posse de Drogas.
Perturbação da Ordem Pública	Desordem Violenta, Atos que Causam Medo, Alarme ou Angústia.
Posse de Armas	Posse de Arma de Fogo, Posse de Armas Brancas ou Objetos Ofensivos.
Crimes Diversos Contra a Sociedade	Direção Perigosa, Violação de Fiança, Outras Infrações Sociais.

Na análise multicritério aqui desenvolvida, não contabilizamos apenas a frequência absoluta de crimes, mas priorizamos sua severidade intrínseca. Adotamos os pesos atribuídos pelo *Office for National Statistics (ONS)* ¹¹, baseados no tempo de sentença aplicado pelos tribunais. Utilizamos esses valores para compor o parâmetro S_i (*score* de adequabilidade), conforme exemplificado na Tabela 10.

Verificamos que, em termos de volume, os dados mais recentes indicam uma concentração de crimes de oportunidade em áreas centrais e *hubs* de transporte. Destacamos os seguintes pontos:

- **Furto de Pessoa:** Identificamos um aumento significativo impulsionado pelo furto de dispositivos móveis em zonas turísticas (ex: *Westminster* e *Camden*).

¹⁰<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/crimeseverityscoreexperimentalstatistics>

¹¹<https://www.ons.gov.uk>

Tabela 10. Pesos Oficiais do *Crime Severity Score* (CSS/ONS) Aplicados às Categorias da MPS

Categoria Principal (Major)	Subcategoria de Exemplo (Minor)	Peso Oficial	Impacto Metodológico / Territorial
Violência Contra a Pessoa	Homicídio	7.979	Gravidade Crítica (Forte atração de centróides)
Crimes Sexuais	Estupro / Abuso Grave	3.656	Impacto Altíssimo (Foco em vulnerabilidade)
Invasão e Arrombamento	Invasão Residencial Agravada	2.127	Impacto Elevado (Invasão com violência)
Incêndio Criminoso e Dano	Incêndio com Risco à Vida	902	Alta Periculosidade e risco social
Roubo	Roubo de Bens Pessoais com Força	746	Alta Prioridade (Violência patrimonial)
Infrações Relacionadas a Drogas	Tráfico de Drogas	667	Médio-Alto (Indicador de redes criminais)
Invasão e Arrombamento	Invasão Residencial Comum	438	Impacto Moderado-Alto (Alarme social)
Posse de Armas	Posse de Arma de Fogo / Branca	350	Risco de Escalação (Prevenção de violência)
Crimes Relacionados a Veículos	Furto ou Roubo de Veículo	200	Impacto Médio (Dinâmica de mobilidade)
Violência Contra a Pessoa	Lesão Corporal Grave	127	Impacto Médio (Agressões urbanas)
Perturbação da Ordem Pública	Desordem Violenta / Ameaças	80	Impacto Moderado (Convivência civil)
Furtos	Furto de Pessoa	40 – 60	Alta Frequência, Baixo Peso Individual
Crimes Relacionados a Veículos	Furto de Objetos em Veículo	34	Baixo Impacto (Crime de oportunidade)
Crimes Diversos Contra a Sociedade	Infrações Sociais / Administrativas	10 – 20	Baixa Prioridade (Baixo potencial ofensivo)
Infrações Relacionadas a Drogas	Posse de Cannabis (Uso Próprio)	3	Impacto Mínimo (Foco estrito em volume)

- **Furto em Lojas:** Constatamos que esta categoria atingiu níveis recordes em 2025, representando um dos principais desafios operacionais.
- **Violência sem Lesão Corporal:** Classificamos como a categoria de maior volume dentro dos crimes contra a pessoa, englobando assédios e agressões verbais.
- **Crimes Relacionados a Veículos:** Notamos que o furto de componentes automotivos permanece constante em distritos periféricos, exigindo estratégias de vigilância específicas.

Embora esta etapa tenha sido conceitualmente levantada e estruturada no trabalho, a modelagem multicritério completa proposta na metodologia genérica não foi efetiva-

mente aplicada na execução prática do caso de uso. Dessa forma, variáveis adicionais, métodos formais de ponderação multicritério e modelos de priorização espacial foram discutidos apenas em nível teórico, não sendo operacionalizados integralmente devido às limitações de granularidade, padronização e disponibilidade das bases públicas utilizadas. Em vista disso, o principal entregável desta etapa, originalmente concebido sob a forma de um mapa de prioridade, reconfigura-se na delimitação de *clusters* espaciais e na extração de seus respectivos centróides.

Em uma primeira etapa, foram selecionadas as ocorrências de crimes em um raio limite de 2 km das delegacias existentes, por meio da função `ST_DWithin`, a qual permite verificar se duas geometrias estão separadas por uma distância inferior a um determinado limiar. Essa função realiza a avaliação de proximidade espacial diretamente sobre as geometrias, considerando o sistema de referência espacial adotado, e retorna verdadeiro para os casos em que a distância entre os pontos representativos das ocorrências e as localizações das delegacias é menor ou igual ao valor especificado. Dessa forma, foi possível identificar e filtrar, de maneira objetiva, os registros situados dentro da área de influência definida para cada unidade policial, estabelecendo um critério espacial consistente para a análise de cobertura territorial.

Em seguida, foi utilizado o algoritmo *K-means*, por meio da função `ST_ClusterKMeans` do *PostGIS*, com o objetivo de agrupar as ocorrências criminais georreferenciadas em *clusters* espaciais e identificar regiões com maior concentração de eventos, subsidiando a análise de padrões espaciais e a proposição de áreas candidatas para a instalação de novas unidades policiais. Essa função implementa o método de particionamento *K-means* diretamente sobre conjuntos de geometrias, atribuindo cada ocorrência ao *cluster* cujo centróide se encontra à menor distância, de modo a minimizar a variabilidade interna de cada grupo e maximizar a separação entre os diferentes agrupamentos. Como resultado, obtêm-se *clusters* espacialmente coerentes, representados por seus respectivos centróides, que sintetizam a distribuição espacial das ocorrências e facilitam a interpretação dos padrões de concentração criminal no território analisado.

5.4. Validação e Avaliação

Para determinar o número ideal de *clusters* K no algoritmo *K-means*, este trabalho utilizou três métricas complementares:

- **Método *Elbow*:** analisa a inércia (soma das distâncias quadráticas entre cada ponto e o centro do seu *cluster*).
- **Coefficiente de *Silhouette*:** mede a similaridade entre os pontos do mesmo agrupamento e dos agrupamentos vizinhos.
- **Índice de *Calinski–Harabasz*:** avalia a razão entre dispersão inter-*cluster* e dispersão intra-*cluster*.

Para a determinação do número adequado de *clusters* K no algoritmo *K-means*, foram realizados experimentos variando-se sistematicamente o valor de K no intervalo de 2 a 10. Para cada valor considerado, o algoritmo foi aplicado ao conjunto de dados e, em seguida, foram calculadas três métricas de avaliação: a inércia, utilizada no Método *Elbow*; o Coeficiente de *Silhouette*; e o Índice de *Calinski–Harabasz*. Os resultados obtidos foram analisados de forma comparativa por meio de gráficos, permitindo avaliar o comportamento de cada métrica em função de K .

A análise da curva de inércia evidenciou um ponto de inflexão em $K = 3$, caracterizando o “cotovelo” apresentado na Figura 2. O Coeficiente de *Silhouette* apresentou seu maior valor médio em $K = 2$, conforme mostrado na Figura 3, enquanto o Índice de *Calinski–Harabasz* atingiu seu pico em $K = 7$, como ilustrado na Figura 4. Esses resultados indicam que diferentes métricas sugerem valores distintos para K , reforçando a necessidade de uma decisão baseada em análise conjunta e não em um único critério isolado.

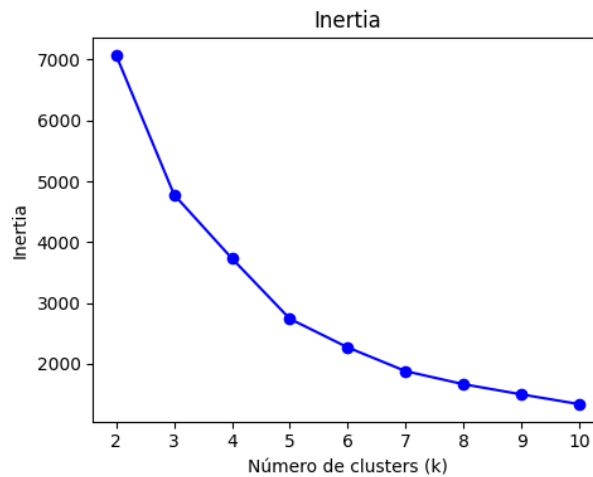


Figura 2. Número de *clusters* para *Inertia*

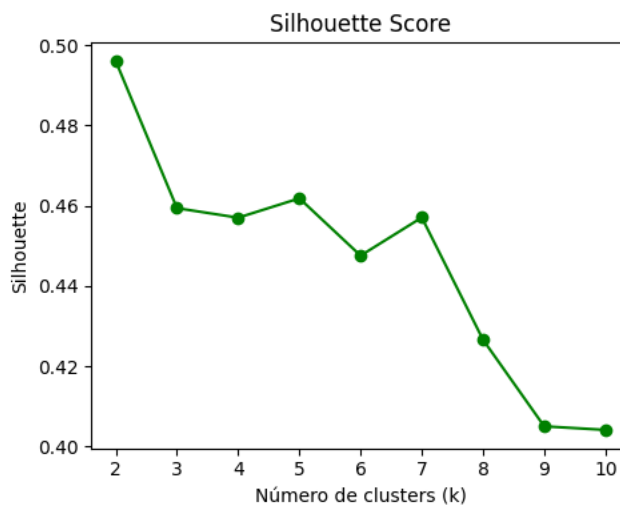


Figura 3. Número de *clusters* para *Silhouette*

A definição do número de *clusters* foi, portanto, realizada com base em experimentação orientada por métricas de avaliação, na qual foram efetuados testes com K variando unitariamente, em ordem crescente, de 2 a 10, com tempo total de execução de 3h e 52min. A escolha final considerou $K = 4$, que apresentou um equilíbrio entre as métricas, com *Silhouette Score* de 0,4570, Índice de *Calinski–Harabasz* de 332.760,86 e Inércia de $3,73 \times 10^3$, indicando uma maior separação entre os *clusters*, conforme representado na Figura 5.

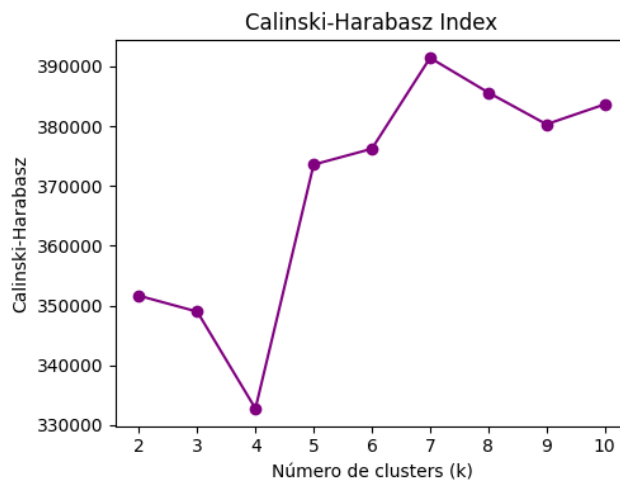


Figura 4. Número de *clusters* para *Calinski Harabasz*

Do ponto de vista conceitual, a escolha de $K = 4$ mostrou-se mais adequada por representar um compromisso entre qualidade estatística dos agrupamentos e interpretabilidade espacial dos resultados. Embora valores menores de K tenham apresentado melhores escores em métricas específicas, tais configurações resultaram em agrupamentos excessivamente amplos, enquanto valores mais elevados implicaram fragmentação excessiva do espaço urbano. Nesse sentido, $K = 4$ apresentou uma solução intermediária, capaz de preservar padrões espaciais significativos de concentração de ocorrências, ao mesmo tempo em que gera regiões territorialmente contínuas e operacionalmente mais coerentes.

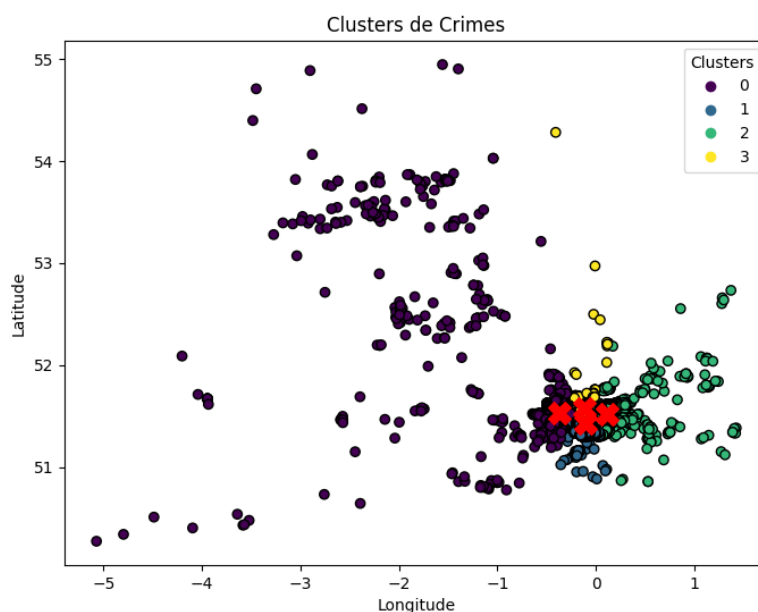


Figura 5. *Clusters* de crimes e seus centróides

5.5. Análise de Cobertura da Rede

A segunda etapa teve como objetivo identificar as melhores localizações para as novas unidades policiais, com a função `ST_Centroid`. Na Figura 6 é apresentado o resultado gerado após a obtenção dos pontos através do *SQL* do Anexo B:

- A cor vermelha representa as ocorrências de crimes individuais e incidentes de comportamentos antissociais.
- A distribuição das unidades de delegacias já existentes está representada pelas estrelas rosa.
- As localizações sugeridas para as novas unidades estão representadas pelos símbolos verdes.

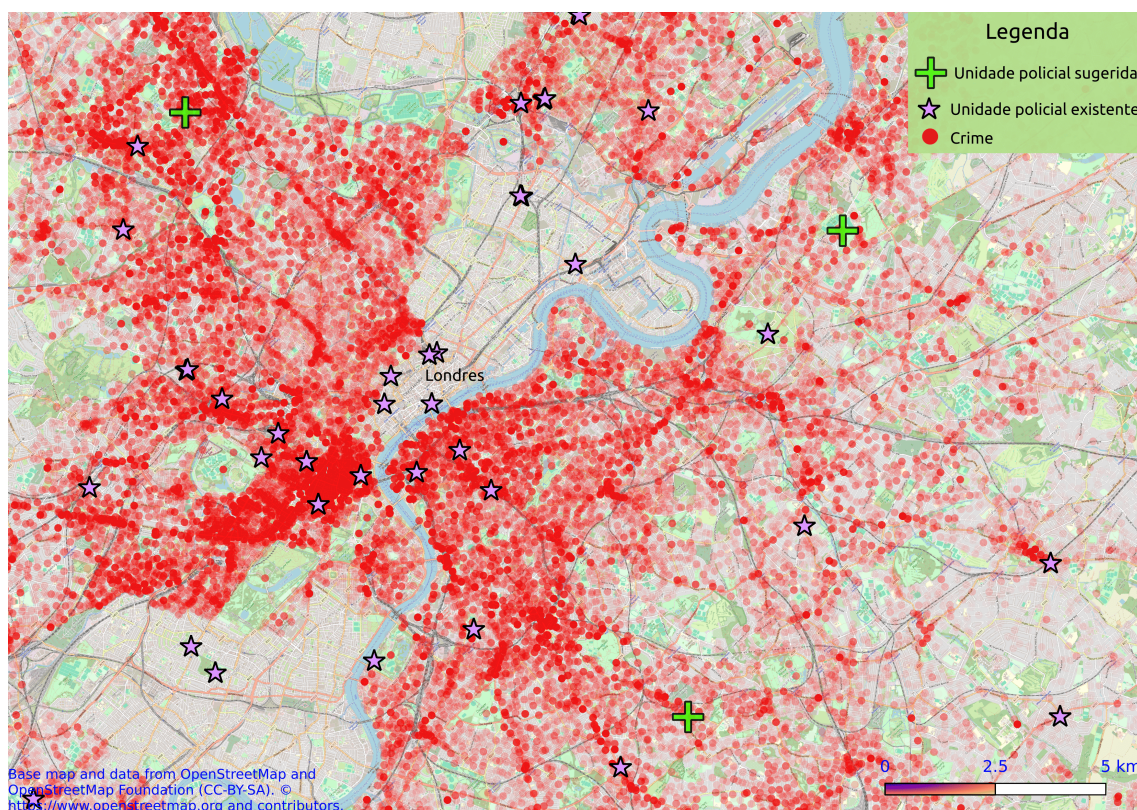


Figura 6. Crimes na região metropolitana de Londres e unidades policiais

A escolha pelo *K-means* justifica-se por sua eficiência e escalabilidade, características fundamentais para lidar com grandes volumes de dados criminais georreferenciados. A implementação no *PostGIS*, por meio da função `ST_ClusterKMeans`, permite agrupar geometrias com base na proximidade espacial sem exportar dados para ferramentas externas, garantindo processamento interno e integração nativa com consultas *SQL*. Estudos indicam que o *K-means* apresenta desempenho superior em tempo de execução em relação a métodos mais complexos, como *DBSCAN*, especialmente em bases dinâmicas e de grande escala conforme [Chakraborty et al. 2014]. Além disso, avanços recentes demonstram que o *K-means* continua sendo uma solução eficiente e previsível para vetores espaciais em larga escala, mesmo diante de desafios de memória e custo computacional [Ji et al. 2025]. Assim, a simplicidade, a rapidez e a integração direta no

banco de dados justificam sua adoção em detrimento de abordagens como *DBSCAN* e *Spectral Clustering*, que embora sofisticadas, exigem maior custo de processamento.

A análise de cobertura da rede viária foi aplicada parcialmente. Embora a metodologia proposta contemple análises baseadas em rede viária, conectividade urbana, isócronas e tempo estimado de deslocamento, nesta aplicação prática foram utilizadas predominantemente distâncias euclidianas e critérios de proximidade espacial direta, em razão da indisponibilidade de dados operacionais detalhados sobre mobilidade, tráfego e tempo real de resposta.

5.6. Limitações do Estudo

A abordagem proposta apresenta limitações metodológicas e operacionais que devem ser consideradas na interpretação dos resultados. Essas restrições derivam principalmente das características dos dados e das escolhas experimentais adotadas.

O raio de abrangência das delegacias foi fixado em 2 km, utilizando distância euclidiana. Essa simplificação desconsidera a malha viária, barreiras urbanas, fluxo de tráfego e tempo real de deslocamento, podendo gerar distorções na estimativa da cobertura efetiva. Estudos futuros devem incorporar análises baseadas em rede viária e isócronas de tempo de resposta.

No processo de clusterização, adotou-se $K = 4$, definido empiricamente após testes exploratórios e avaliação de métricas de qualidade. No entanto, a definição do número ótimo de *clusters* constitui limitação inerente ao *K-means*, que tende a produzir agrupamentos esféricos e depende de distâncias euclidianas. Métodos alternativos, como *DBSCAN* e *Spectral Clustering*, podem ser explorados para capturar padrões espaciais mais complexos.

Os dados apresentam granularidade temporal restrita a mês e ano, impedindo análises espaço-temporais detalhadas sobre horários, dias da semana e sazonalidades. Consequentemente, os resultados refletem uma distribuição espacial predominantemente estática. Pesquisas futuras devem utilizar dados de maior resolução temporal para modelos dinâmicos e preditivos.

A definição de três novas localizações para unidades policiais teve caráter experimental, com o objetivo de demonstrar a aplicabilidade da metodologia, e não reflete restrições orçamentárias, operacionais ou institucionais reais. Trabalhos futuros podem empregar modelos de otimização mais robustos, como o Problema da *P-Median*, incorporando restrições reais de infraestrutura, orçamento e capacidade operacional.

Por fim, inconsistências na precisão espacial dos dados de geocodificação podem introduzir vieses nos agrupamentos e centróides gerados. Recomenda-se a realização de análises de sensibilidade e a discussão dos efeitos do *MAUP*. Adicionalmente, não foi realizada validação operacional quantitativa das novas localizações, considerando métricas como tempo médio de resposta e percentual de cobertura.

6. Conclusão

A distribuição desigual da criminalidade em grandes centros urbanos impõe desafios relevantes ao planejamento estratégico da segurança pública, especialmente em cenários caracterizados por elevada densidade populacional, limitações operacionais e necessidade

de otimização da cobertura territorial. Nesse contexto, este trabalho apresentou uma abordagem aplicada baseada em bancos de dados geoespaciais, análise multicritério e algoritmos de agrupamento espacial para apoiar a localização estratégica de unidades policiais em ambientes urbanos.

Inicialmente, foi realizada uma revisão teórica envolvendo os eixos de Banco de Dados, Algoritmos e Cidades Inteligentes, complementada por uma Revisão Sistemática da Literatura voltada à análise espacial criminal, policiamento orientado por dados e uso de SIG em segurança pública. Os resultados evidenciaram o potencial das técnicas de análise geoespacial para identificação de *hotspots*, apoio à tomada de decisão e otimização da alocação de recursos policiais.

A metodologia proposta integrou processos de engenharia de dados, MCDA e agrupamento espacial. A utilização da abordagem *in-database* no caso de uso permitiu o processamento de grandes volumes de dados georreferenciados sem necessidade de exportação para ferramentas externas, preservando integridade espacial e escalabilidade computacional.

Como estudo aplicado, foram utilizados 554.832 registros criminais da região metropolitana de Londres. A aplicação das métricas *Elbow*, *Silhouette* e *Calinski-Harabasz* permitiu validar a definição de $K = 4$, revelando agrupamentos espacialmente coerentes com a distribuição criminal observada. Os centróides dos *clusters* foram interpretados como localizações potenciais para novas unidades policiais, fornecendo subsídios quantitativos para expansão da cobertura territorial e redução de lacunas operacionais.

Os resultados demonstram que a integração entre SIG, bancos de dados espaciais e algoritmos de agrupamento pode contribuir significativamente para o planejamento urbano orientado por evidências, especialmente em iniciativas relacionadas a Cidades Inteligentes e *smart policing*.

Entretanto, algumas limitações devem ser consideradas. O modelo adotou distâncias euclidianas simplificadas, desconsiderando restrições reais da malha viária, fluxo urbano e tempo efetivo de deslocamento. Além disso, a granularidade temporal dos dados limitou análises espaço-temporais mais detalhadas, enquanto inconsistências na precisão geográfica podem influenciar a formação dos agrupamentos. A definição experimental de três novas unidades policiais também não considera restrições orçamentárias e institucionais reais.

Outra limitação importante refere-se ao potencial viés algorítmico associado ao uso de registros históricos criminais. Modelos de agrupamento podem reproduzir padrões históricos de policiamento e reforçar desigualdades territoriais, fenômeno conhecido como *feedback loop* de policiamento. Dessa forma, as recomendações produzidas devem ser interpretadas como mecanismos de apoio à decisão, e não como determinismos espaciais automatizados.

Como trabalhos futuros, destacam-se diferentes possibilidades de ampliação metodológica e operacional. Inicialmente, recomenda-se a investigação de algoritmos alternativos de agrupamento espacial, como *DBSCAN*, *HDBSCAN* e *Spectral Clustering*, visando representar padrões criminais complexos, não lineares e espacialmente irregulares. Também se sugere a incorporação de análises baseadas em rede viária (*Network Constrained Analysis*) e geração de isócronas de deslocamento, permitindo estimativas

mais realistas de cobertura operacional.

Adicionalmente, futuras pesquisas podem incorporar dados temporais de maior resolução para modelagem espaço-temporal, sazonalidade criminal e patrulhamento preditivo. Modelos clássicos de Pesquisa Operacional, como *P-Median Problem*, *LSCP* e *MCLP*, também podem ser integrados ao *framework* para considerar restrições reais de infraestrutura, orçamento e capacidade operacional.

Sugere-se ainda a realização de análises de sensibilidade relacionadas à qualidade da geocodificação e aos efeitos do *MAUP*, bem como validações quantitativas baseadas em métricas operacionais de cobertura e tempo de resposta. Finalmente, recomenda-se incorporar mecanismos de transparência, equidade algorítmica e governança de dados, buscando garantir que soluções computacionais aplicadas à segurança pública permaneçam alinhadas aos princípios de justiça social e planejamento urbano sustentável.

Referências

- Abhay, D. A., Akash, S., Ashwin, K., Shenoy, A. G., and Auradkar, P. K. (2023). Smart policing: Using geospatial crime data to plan patrol routes. *2023 4th International Conference for Emerging Technology (INCET)*, pages 1–7.
- Almuhanna, A. A., Alrehili, M. M., Alsubhi, S. H., and Syed, L. (2021). Prediction of crime in neighbourhoods of new york city using spatial data analysis. *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*.
- Alves, H. J. P., Fernandes, F. A., Silva, E. M., and Batista, B. D. O. (2025). An application of the k-means clustering method to national public safety data: The case of intentional homicides. *Brazilian Journal of Biometrics*, 43(1):1–9.
- Borowik, G., Wawrzyniak, Z. M., and Cichosz, P. (2018). Time series analysis for crime forecasting. *2018 26th International Conference on Systems Engineering (ICSEng)*.
- Brantingham, P. J. and Brantingham, P. L. (1981). Notes on the geometry of crime. *Environmental Criminology*, pages 27–54.
- Brantingham, P. J. and Brantingham, P. L. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, 13(1):3–28.
- Brantingham, P. J. and Brantingham, P. L. (2008). Crime pattern theory. *Environmental Criminology and Crime Analysis*, pages 78–93.
- Calinski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), pages 1–27.
- Callegaro, S., Palacios, R. A., and Luciano, E. M. (2022). Discutindo o uso da pesquisa-ação e suas variações em estudos sobre cidades inteligentes. *ISLA 2022 Proceedings Latin America (ISLA)*, pages 1–10.
- Carvalho, G. A., Neto, P. B., Cacho, N., Adachi, E., and Lopes, F. (2018). Plataforma rota: Histórico, desafios e soluções para segurança pública em cidades inteligentes. In *Anais do I Workshop Brasileiro de Cidades Inteligentes*, pages 61–69, Porto Alegre, RS, Brasil. SBC.

- Chakraborty, S., Nagwani, N. K., and Dey, L. (2014). Performance comparison of incremental k-means and incremental dbSCAN algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*.
- Church, R. and ReVelle, C. (1974). The maximal covering location problem. *Papers in Regional Science*, 32(1):101–118.
- Clarke, R. V., editor (1997). *Situational Crime Prevention: Successful Case Studies*. Harrow and Heston, Albany, NY, second edition. 2nd edition.
- Cornish, D. B. and Clarke, R. V. (1986). *The Reasoning Criminal: Rational Choice Perspectives on Offending*. Springer-Verlag, New York.
- Couto, L. O. (2024). Identificação de Padrões de Crime: Uma Abordagem Data-Driven para o Roubo de Celulares no RS. Monografia de Bacharelado em Ciência da Computação.
- Dong, Q., Ye, R., and Fan, Y. (2022). Impact of spatial correlation on crime prediction in communities with different crime densities. *2022 IEEE 8th International Conference on Computer and Communications (ICCC)*.
- Eck, J. E., Chainey, S., Cameron, J. G., Leitner, M., and Wilson, R. E. (2005). Mapping crime: Understanding hotspots. *U.S. Department of Justice, National Institute of Justice*, pages 1–71.
- Felson, M. (2006). *Crime and Nature*. Sage Publications, Thousand Oaks.
- Hunt, E. L. and Reffert, S. (2021). Improving the open cluster census: I. comparison of clustering algorithms applied to Gaia DR2 data. *Astronomy & Astrophysics*, 646:A104.
- Jain, S. and Riyank (2023). Geographical crime rate prediction. In *2023 4th International Conference on Intelligent Engineering and Management (ICIEM)*, pages 1–6.
- Ji, Y., Liu, Z., Wang, S., Sun, Y., and Peng, Z. (2025). On simplifying large-scale spatial vectors: Fast, memory-efficient, and cost-predictable k-means. *Proceedings of the 41st IEEE International Conference on Data Engineering (ICDE)*.
- Junior, A. A., Cacho, N., Thome, A. C., Medeiros, A., and Borges, J. (2017). A predictive policing application to support patrol planning in smart cities. In *2017 IEEE International Smart Cities Conference (ISC2)*, pages 1–8, Wuxi, China. IEEE.
- Kim, G. et al. (2025). Crime mapping in urban environments using explainable AI. *Sustainable Cities and Society*.
- Kim, Y., Kim, G., Lee, Y., and Jang, K. (2020). Bandwidth selection of kernel density estimation for GIS-based crime occurrence map visualization. *ICTC 2020*, pages 1705–1708.
- Leal, M. and Gomes-Jr, L. (2022). Impacto da pandemia da COVID-19 nos padrões de crimes no município de Curitiba. In *Anais da XVII Escola Regional de Banco de Dados*, pages 101–108, Porto Alegre, RS, Brasil. SBC.
- Li, S. and Tang, W. (2019). Space-time variation of property crime in Beijing with ESDA method. *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*.

- Lilhare, S. G., Kumavat, Y., Banait, G., and Kurkelli, A. (2024). Crime hotspots mapping and fir data interface. *ICICET 2024*, pages 1–7.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1)*. University of California Press, pages 281–297.
- Mansourihanis, O. et al. (2025). Evolution and prospects of geographic information systems (gis) applications in urban crime analysis: A review of literature. *Sustainable Cities and Society*.
- Mantovani, D. M. N., Santos, L. T., Machado, C. J., and Leal, G. A. (2024). Inteligência em segurança pública municipal: modelos para previsão de crimes e planejamento de zeladoria urbana. *ISLA 2024 Proceedings Latin America (ISLA)*, pages 1–10.
- Nag, A., Ranjan, R., and Kumar, C. N. S. V. (2022). An approach on cyber crime prediction using prophet time series. *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*.
- Neto, J. F. Q. (2020). *A Visual Analytics Approach for Geocoded Crime Data*. PhD thesis, Universidade Federal do Ceará, Fortaleza. Tese de Doutorado em Ciência da Computação.
- Neto, M. P. et al. (2020). Crimewatcher: Um sistema de visual analytics integrado à secretaria nacional de segurança pública. *Revista Brasileira de Computação Aplicada*, 12(2):45–62.
- Palacios, R. A., Callegaro, S., and Luciano, E. M. (2020). Collaborative governance for decision making on smart cities initiatives. *ISLA 2020 Proceedings Latin America (ISLA)*, pages 1–10.
- Parcianello, Y., Kozievitch, N. P., Fonseca, K. V. O., and Rosa, M. (2021). Origin-destination data: a prototype and related scenarios. *Revista Brasileira de Computação Aplicada*, 13(2):16–27.
- Ratcliffe, J. H. (2010). Crime mapping and the training needs of law enforcement. *Police Practice and Research*, 11(1), pages 3–15.
- Ratra, A., Agarwal, A., Vats, S., Sharma, V., Kukreja, V., and Yadav, S. P. (2023). A comprehensive review on crime patterns and trends analysis using machine learning. *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, pages 53–65.
- Sathiyanarayanan, M., Junejo, A. K., and Fadahunsi, O. (2019). Visual analysis of predictive policing to improve crime investigation. *IC3I 2019*, pages 197–203.
- Sayal, A., Gupta, A., Vasundhara, C., M, Y. B., Gupta, V., and Maheshwari, H. (2024). Crime detection using data mining techniques. *2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT)*.

- Sharma, S., Uniyal, A., Srinivasan, P., and Chaudhari, S. (2022). Fuzzy based geo-spatial crime category prediction for crime mapping and safe route travel. *2022 IEEE Region 10 Symposium (TENSYP)*.
- Sheikh, J., Shafique, I., Sharif, M., Zahra, S. A., and Farid, T. (2017). Role of gis in crime mapping and analysis. *ComTech 2017*, pages 126–131.
- Shenoy, A. G. and Auradkar, P. K. (2023). Smart policing: Using geospatial crime data to plan patrol routes. *2023 4th International Conference for Emerging Technology (INCET)*.
- Sherman, L. W., Gartin, P. R., and Buerger, M. E. (1989). Hotspots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27(1), pages 27–56.
- Silva, G. M. and Cestari, J. M. A. P. (2022). Visualização da informação com base de dados abertos da guarda municipal de curitiba. *ISLA 2022 Proceedings Latin America (ISLA)*, pages 1–10.
- Silva, R. S., Silva, R. R., Kuribayashi, H. P., Cunha, C. V., Francês, C. R. L., and Sousa, K. N. S. (2019). Clusterização de dados mistos para análise da atividade pesqueira artesanal na bacia araguaia-tocantins. *Revista Brasileira de Computação Aplicada*, 11(3):155–164.
- Sravani, T. and Suguna, M. R. (2022). Comparative analysis of crime hotspot detection and prediction using convolutional neural network over support vector machine with engineered spatial features towards increase in classifier accuracy. *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*.
- Steier, J., Zigarelli, A., Giannini, E., and Minimair, M. (2017). Crime sequencing: Fighting crime with mathematics and technology. *2017 IEEE MIT Undergraduate Research Technology Conference (URTC)*.
- Tarlekar, S., Bhosle, R., D'souza, E., and Sheikh, S. (2021). Geographical crime rate prediction system. *2021 IEEE India Council International Subsections Conference (INDISCON)*.
- v. Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*.
- Vadav, M. G., N, R., Reddy, E. S., Vishal, M. S., and Vishal, G. (2024). The role of machine learning in crime analysis and prediction. *2024 International Conference on Expert Clouds and Applications (ICOECA)*.
- Wang, J., Zhang, S., Lan, Y., Wu, C., Xia, Y., and Chen, L. (2020). Crime analysis of spatial-temporal distribution based on knn algorithm. *ICISE-IE 2020*, pages 150–154.
- Yadav, R. and Sheoran, S. K. (2018). Crime prediction using auto regression techniques for time series data. *2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*.
- Yoo, J. S., Park, S. J., and Raman, A. (2019). Micro-level incident analysis using spatial association rule mining. *2019 IEEE International Conference on Big Knowledge (ICBK)*.

A. SQL para atualizar a projeção cartográfica e criar índice

```
ALTER TABLEISTRADO_gustavo_rodrigues_metropolitan
    ADD COLUMN geom_3395 geometry(Point, 3395);
UPDATEISTRADO_gustavo_rodrigues_metropolitan
    SET geom_3395 = ST_Transform(geom, 3395)
    WHERE geom IS NOT NULL;
CREATE INDEXISTRADO_gustavo_rodrigues_metropolitan_geom_3395_idx
    ONISTRADO_gustavo_rodrigues_metropolitan
    USING GIST (geom_3395);

ALTER TABLEISTRADO_gustavo_rodrigues_police
    ADD COLUMN wkb_geometry_3395 geometry(Point, 3395);
UPDATEISTRADO_gustavo_rodrigues_police
    SET wkb_geometry_3395 = ST_Transform(wkb_geometry, 3395)
    WHERE wkb_geometry IS NOT NULL;
CREATE INDEXISTRADO_gustavo_rodrigues_police_wkb_geometry_3395_idx
    ONISTRADO_gustavo_rodrigues_police
    USING GIST (wkb_geometry_3395);
```

B. SQL para aplicar o algoritmo K-means

```
WITH crimes_outside_coverage AS (
    SELECT c.geom_3395
    FROMISTRADO_gustavo_rodrigues_metropolitan AS c
    LEFT JOINISTRADO_gustavo_rodrigues_police AS p
        ON ST_DWithin(c.geom_3395, p.wkb_geometry_3395, 2000)
    WHERE p.ogc_fid IS NULL
),
crimes_clustered AS (
    SELECT
        geom_3395,
        ST_ClusterKMeans(geom_3395, 4) OVER () AS cluster_id
    FROM crimes_outside_coverage
),
cluster_centroids AS (
    SELECT
        cluster_id,
        COUNT(*) AS crime_count,
        ST_Centroid(ST_Collect(geom_3395)) AS centroid
    FROM crimes_clustered
    GROUP BY cluster_id
)
SELECT
    cluster_id, crime_count, centroid
FROM cluster_centroids
ORDER BY crime_count DESC;
```