



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

FERNANDO KELLER

Análise de Locais de Crimes em Matelândia - Paraná:

Análise Estatística de locais de Crimes em Matelândia baseado em Clusterização
Não-supervisionada de Centroides para Suporte à Decisão na Construção de Nova Base da
Polícia Militar

Recife

2025

FERNANDO KELLER

Análise de Locais de Crimes em Matelândia - Paraná:

Análise Estatística de locais de Crimes em Matelândia baseado em Clusterização Não-supervisionada de Centroides para Suporte à Decisão na Construção de Nova Base da Polícia Militar

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Área de Concentração: Estatística Aplicada

Orientador (a): Dr. Thyago Celso Cavalcante Nepomuceno

Coorientador (a): Dr. Victor Diogho Heuer de Carvalho

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Keller, Fernando.

Análise de locais de crimes em Matelândia - Paraná: análise estatística de locais de crimes em Matelândia baseado em Clusterização Não-Supervisionada de Centroides para suporte à decisão na construção de nova base da Polícia Militar / Fernando Keller. - Recife, 2024.

50f.: il.

Dissertação (Mestrado) - Universidade Federal de Pernambuco, Centro de ciências exatas e da natureza, Programa de pós graduação em estatística, 2024.

Orientação: Thyago Celso Cavalcante Nepomuceno.

Inclui referências.

1. Machine learning; 2. K-means; 3. Matelândia. I. Nepomuceno, Thyago Celso Cavalcante. II. Título.

UFPE-Biblioteca Central

RESUMO

Este estudo investiga a criminalidade na cidade de Matelândia, estado do Paraná, utilizando uma técnica de machine learning para agrupar dados referente a locais de crimes ocorridos no município em clusters, usando o método k-means. O objetivo é avaliar a necessidade da construção de uma nova base para a Polícia Militar e também avaliar se a atual base está bem localizada em relação aos locais onde os crimes ocorreram, como também avaliar se a posição atual está satisfatória, usando uma técnica que dará grau de importância para cada natureza criminal. Foram quatro os crimes escolhidos para serem analisados neste trabalho, sendo eles furto, roubo, “Maria da Penha” (violência doméstica) e estupro, referente ao período de outubro de 2022 à outubro de 2023, os quais foram fornecidos pela Polícia Militar do Estado do Paraná. Foi feito um estudo sobre a criminalidade em um contexto nacional, além do impacto dos crimes escolhidos na sociedade. Foi realizado um estudo e análise dos dados criminais, aprofundando em suas características principais e como estão distribuídos na período temporal observado. Os resultados propostos foram alcançados, e após aplicar o método k-means, foi possível afirmar que a localização atual está satisfatória, assim como foi possível concluir que uma segunda base policial militar seria de grande valia no combate ao crime na região. Dessa forma este estudo pode proporcionar uma melhor alocação de recursos da segurança pública do estado do Paraná na cidade de Matelândia.

Palavras-chave: machine learning; k-means; Matelândia; Polícia Militar do Paraná; análise criminal.

ABSTRACT

This study investigates crime in the city of Matelândia, in the state of Paraná, using a machine learning technique to cluster data regarding crime locations in the municipality, applying the k-means method. The aim is to assess the need for building a new base for the Military Police and also to evaluate whether the current base is well-located in relation to where the crimes occurred, as well as to assess if the current position is satisfactory, using a technique that assigns an importance degree to each type of crime. Four types of crimes were selected for analysis in this study: theft, robbery, domestic violence (under the "Maria da Penha" law), and rape, covering the period from October 2022 to October 2023, with data provided by the Military Police of the State of Paraná. A study on national crime trends was conducted, as well as the impact of the selected crimes on society. The criminal data was analyzed, focusing on its key characteristics and distribution within the observed time period. The proposed results were achieved, and after applying the k-means method, it was concluded that the current location is satisfactory, as well as the finding that a second military police base would be highly valuable in combating crime in the region. Thus, this study can contribute to a better allocation of public security resources in the state of Paraná in the city of Matelândia.

Keywords: machine learning; k-means; Matelândia; Polícia Militar do Paraná; análise criminal.

LISTA DE FIGURAS

Figura 1 – Cidade de Matelândia	9
Figura 2 – Base Policial Militar Agro Cafeeira	12
Figura 3 – Diagrama passo a passo da metodologia	23
Figura 4 – Dados formato inicial	25
Figura 5 – Distribuição de ocorrências por natureza criminal	27
Figura 6 – Distribuição de ocorrências criminais por mês do ano	27
Figura 7 – Distribuição de ocorrências criminais por dia do mês	28
Figura 8 – Distribuição de ocorrências por dia da semana	29
Figura 9 – Distribuição de ocorrências por hora do dia	29
Figura 10 – Distribuição de ocorrências por hora do dia crime furto	30
Figura 11 – Distribuição de ocorrências por hora do dia crime “Maria da Penha”	30
Figura 12 – Distribuição de ocorrências por hora do dia crime roubo	31
Figura 13 – Distribuição de ocorrências por hora do dia crime estupro	32
Figura 14 – Top 10 ocorrências distribuídos por bairros	32
Figura 15 – Top 10 ocorrências distribuídos por ruas	33
Figura 16 – Mapa geral dos crimes em Matelândia	34
Figura 17 – Estudo feito com 1 cluster	35
Figura 18 – Estudo feito com dois clusters	36
Figura 19 – Estudo feito com crimes de furto	38
Figura 20 – Estudo feito com crimes de “Maria da Penha”	39
Figura 21 – Estudo feito com crimes de roubo	40
Figura 22 – Estudo feito com crimes de estupro	41
Figura 23 – Todos os centroides dos centroides	44
Figura 24 – Centroide dos centroides	45

LISTA DE TABELAS

Tabela 1 – Dias da semana e o número de ocorrências	24
Tabela 2 – Total de ocorrências por tipo de naturezas criminais	26
Tabela 3 – Distância de todos os crimes com um cluster	35
Tabela 4 – Distância dos dois centroides em comparação a base PM Matelândia .	37
Tabela 5 – Distância do centroide em comparação a base PM Agro Cafeeira	37
Tabela 6 – Distancias da Base PM dos centroides dos crimes	43
Tabela 7 – Distancias da Base PM Agro Cafeeira dos centroides dos crimes	43
Tabela 8 – Distancias da Base PM em relação ao centroide dos centroides	43

SUMÁRIO

1	INTRODUÇÃO	8
1.1	ANÁLISE DE CRIMINALIDADE	10
1.2	OBJETIVOS	11
1.2.1	Objetivos gerais	11
1.2.2	Objetivos específicos	12
2	PRELIMINARES	14
2.1	MACHINE LEARNING	14
2.2	ALGORITIMO K-MEANS	15
2.2.1	Método Centroide dos Centroides (k-means-c²)	18
3	METODOLOGIA	20
4	RESULTADO	24
4.1	ANÁLISE EXPLORATÓRIA DE DADOS	24
5	CONCLUSÃO	46
	REFERÊNCIAS	49

1 INTRODUÇÃO

O estudo realizado neste trabalho, trará informações sobre locais de ocorrências policiais que ocorreram em uma cidade no interior do estado do Paraná, o município de Matelândia, o qual pode ser observado na figura 1, é localizada à aproximadamente 565 quilômetros da capital Curitiba, por via rodoviária, estando situada na região oeste do estado. Segundo o último censo realizado pelo Instituto Brasileiro de Geografia e Estatística (2022), a cidade de Matelândia possui 18.450 habitantes, com uma área territorial total de 639.746 km^2 e um Índice de Desenvolvimento Humano (IDH) de 0,760. Sua economia é majoritariamente advinda do agronegócio, e a maioria de seus empregos formais, com carteira assinada, são provenientes de empresas frigoríficas que se instalaram na região. O município tem como característica ser transeccionado pela principal rodovia federal do estado, a BR-277, e possui dois distritos que ficam afastados do centro do município, sendo eles Agro Cafeeira, que está a uma distância próxima de 8 quilômetros do centro e o distrito de Vila Esmeralda, que está a cerca de 15 quilômetros do centro de Matelândia. A cidade está localizada entre duas das principais cidades do Paraná, a leste pela cidade de Cascavel, cerca de 67 quilômetros e atualmente possui 348.051 habitantes segundo censo realizado pelo Instituto Brasileiro de Geografia e Estatística (2022), e a oeste pela cidade de Foz do Iguaçu, cerca de 72 quilômetros e possui 258.248 habitantes segundo último censo. Matelândia é considerada uma cidade localizada na faixa de fronteira, conforme o § 2º do art. 20 da Constituição Brasil (1988), a faixa de até cento e cinquenta quilômetros de largura, ao longo das fronteiras terrestres, são assim definidas, e essas fronteiras são feitas com o Paraguai e a Argentina.

1.1 ANÁLISE DE CRIMINALIDADE

Segundo a Constituição Federal de 1988, em seu Artigo 5º, é garantido que para todo Brasileiro ou estrangeiro residente no país assegura-se alguns direitos básicos, entre eles o direito a segurança. Brasil (1988). Apesar disso, de acordo com o Instituto de Pesquisa Econômica Aplicada (IPEA) (2018), o Brasil tem níveis alarmantes de violência registrada no país, tendo obtido o recorde de homicídios em 2016, com a quantia de 62.517 pessoas assassinadas em todo território nacional.

Seis anos após a data em que foi realizada esta pesquisa, em 2022, os números de homicídios registrados diminuíram, como pode ser verificado no último relatório divulgado pelo Instituto de Pesquisa Econômica Aplicada (IPEA) (2024). Esta queda no número de homicídios porém, não foi o suficiente para acabar com a sensação de insegurança por parte da população, como mostra pesquisa pública realizada e divulgada em março de 2024 pelo Instituto de Pesquisas Datafolha, onde foram entrevistadas 2.002 pessoas acima de 16 anos, em diversas cidades Brasileiras, e mostrou que a percepção de segurança das pessoas que sentem-se muito inseguros subiu de 34% para 39%, e apenas 16% dos entrevistados disseram sentir-se seguros andando nas ruas após o anoitecer conforme Datafolha (2024). Uma pesquisa realizada pela Federação das Indústrias do Estado do Rio de Janeiro (Firjan), onde é comparado o Brasil em relação ao G20, grupo formado pelas principais economias do mundo, revelou que na área de segurança pública, o país teve índices melhores apenas se comparado com outros dois países integrantes do G20, sendo eles México e África do Sul (Firjan) (2024), mas muito longe do primeiro colocado neste quesito, que é ocupado por Singapura. Ou seja, não basta apenas haver uma diminuição de números brutos, é preciso também que a sensação de segurança seja observada, pois são dois fatores que se complementam.

Uma das formas de aumentar a sensação de segurança, segundo a Teoria das “Janelas Quebradas” (Broken Windows Theory) (WILSON; KELLING, 1982), é com equipes policiais patrulhando as ruas da cidade, entretanto será que essa é a melhor forma de se atacar a violência como um todo, visto que um dos grandes desafios que impedem esse maior fluxo de policiais patrulhando, é a falta de efetivo suficiente para tal. Esta é uma das questões que será levado em conta para ser feita a análise neste estudo.

Analisar crimes é uma das formas de ajudar as forças policiais a otimizarem e conseguirem melhores resultados no controle da violência. Segundo Silva (2015), algumas funções

da análise criminal que podem ser destacadas estão a identificação dos envolvidos no crime ou de existência de padrões de crimes em séries; previsão de ocorrências criminais; análise dos perfis, dos objetivos ou lugares em que poderão ocorrer crimes; oferecer dados de suporte para o policiamento comunitário e programas de prevenção ao crime; fornecer dados para dar suporte ao planejamento de uma unidade e analisar dados operacionais para planejamento de alocação de recursos e orçamentos de uma unidade.

Ainda segundo Silva (2015), se fosse dividir a análise criminal em passos poderia destacar as seguintes etapas:

- Registro: tem-se início da análise criminal, com o boletim de ocorrência;
- Agrupamento dos dados: todas as informações coletadas devem ser analisadas e filtradas conforme o seu conteúdo;
- Classificação: devemos classificar os tipos de delitos, de acordo com suas ações;
- Organizar e qualificar os dados: nesta fase final, as informações adquiridas são transformados em mapeamentos que identificam áreas mais vulneráveis, são confeccionadas tabelas e gráficos com os dados obtidos.

Dados e informações podem ser uma forma relevante no processo de tomada de decisão. Mostrando a importância da análise criminal para os órgãos de segurança pública.

Fernandez (2020) afirma que gestores da área de segurança pública bem instruídos, com informações que lhes possibilitem entender a criminalidade de forma clara e ampla, poderão obter mais sucesso em tomada de decisões e elaboração de políticas públicas eficientes no enfrentamento e na forma de minimizar crimes.

1.2 OBJETIVOS

1.2.1 Objetivos gerais

O presente trabalho, tem como objetivo fornecer informações relevantes que possam vir a ajudar gestores de segurança pública à tomar decisões relacionadas ao tema na cidade de Matelândia no Paraná. Visto que a demanda atual, é o implemento de uma nova base policial no distrito de Agro cafeeira, localizada a cerca de 8 quilômetros, por via rodoviária, do centro de Matelândia. A estrutura física desta base policial já está construída, conforme

pode-se observar na figura 2, e foi custeada pelo município de Matelândia, aguardando apenas o efetivo policial. Esta solicitação por parte dos agentes políticos do município, é fundamentada pela quantidade de ocorrências criminais no distrito.

Com base nestas informações, será realizado uma análise dos dados criminais desta região, e através da técnica de machine learning, usando o método k-means, pretende-se responder a uma demanda política da cidade de Matelândia, onde a gestão pressiona a Policia Militar do Estado do Paraná a aumentar o efetivo em sua cidade, e assim alocar mais uma equipe policial de plantão no distrito de Agro Cafeeira. Será analisado se é mesmo justificável esta demanda, com base no total de ocorrências que acometerem a localidade no período proposto, ou se a situação atual é a mais otimizada dentro da realidade criminal do município.



(a) Base PM Agro Cafeeira



(b) Porta Base PM Agro Cafeeira

Figura 2 – Base Policial Militar Agro Cafeeira

1.2.2 Objetivos específicos

Outro objetivo a ser explorado, será um cenário onde não somente se analisará o numero total de ocorrências que foram registradas, mas também se dará um grau de importância para cada tipificação criminal, tendo como objetivo analisar se o modelo atual de policiamento da cidade de Matelândia está otimizado, levando em consideração que uma eventual falta de efetivo de policiais no estado do Paraná, inviabilize a abertura de uma segunda base policial militar no distrito de Agro Cafeeira.

Para alcançar este objetivo, será feito o a análise de centroides gerados a partir de outros centroides anteriormente obtidos. Como cada tipo de crime, será tratado com um

grau importância, e isto se dará pelo número de clusters atribuído previamente para cada conduta criminal, será extraído um centroide para cada cluster atribuído, por exemplo, um grupo de dados referente a um crime qualquer será atribuído o grau de importância três, sendo então realizado o estudo a partir de três clusters, gerando três centroides distintos, e a partir destes três centroides, será realizado o cálculo do centroide dos centroides, onde será feito o estudo desta localização final. Esta abordagem visa influenciar a área do centroide final e será formalizada e explicitada no decorrer do trabalho.

2 PRELIMINARES

2.1 MACHINE LEARNING

Aprendizado da máquina (machine learning, em inglês) tem se consolidado como uma das áreas mais transformadoras e inovadoras da ciência da computação. Com uma sociedade cada vez mais dependente de dados, o aprendizado da máquina oferece uma abordagem poderosa para resolver problemas complexos e tomar decisões. Esta demanda de dados diários, cria a necessidade de ferramentas que automatizem o uso de dados, com métodos capazes de detectar padrões, aprender com estes padrões e assim fazer algumas previsões, ou executar algum tipo de decisão Murphy (2012).

O aprendizado da máquina concentra-se no desenvolvimento de algoritmos e modelos capazes de aprender padrões a partir de dados. Com uma capacidade de adaptação e melhoria contínua, tornando o aprendizado da máquina tão relevante para uma série de aplicações, como reconhecimento de voz, recomendação de produtos, análise preditiva entre outros.

Em relação ao aprendizado da máquina, existem algumas formas de aprendizado que podem ser aplicadas, são especificamente sete tipos: aprendizado supervisionado; aprendizado Não supervisionado; aprendizado Semi-Supervisionado; aprendizado por reforço; aprendizado por transferência; aprendizado online e aprendizado profundo. Porém pode-se destacar dois deles que são observados com maior frequência, sendo o aprendizado supervisionado e o aprendizado não supervisionado.

No aprendizado supervisionado, conforme Bishop (2006), o modelo é treinado com um conjunto de dados não rotulado, sem saídas explícitas, com o objetivo de identificar estruturas ou padrões subjacentes nos dados. Matematicamente podemos expressar este aprendizado, segundo Murphy (2012), da seguinte forma.

No aprendizado supervisionado, o objetivo é aprender um mapeamento de entradas x para saídas y , dado um conjunto rotulado de pares entrada-saída

$$D = \{(x_i, y_i)\}_{i=1}^N \quad (2.1)$$

Onde:

- D é chamado de conjunto de treinamento;

- N é o número de exemplos de treinamento.

Cada entrada de treinamento x_i é um vetor D -dimensional de números. Esses são chamados de características, atributos ou covariáveis. Sendo assim, x_i pode ser um objeto estruturado complexo, como uma imagem, uma mensagem de e-mail, uma série temporal, um gráfico.

A forma da variável de saída ou resposta pode ser qualquer coisa, mas a maioria dos métodos assume que y_i é uma variável categórica ou nominal de um conjunto finito, $y_i \in \{1, \dots, C\}$ ou que y_i é um escalar de valor real. Quando y_i é categórico, o problema é conhecido como classificação ou reconhecimento de padrões, e quando y_i é real, o problema é conhecido como regressão.

O segundo tipo de aprendizado é a abordagem descritiva ou de aprendizado não supervisionado. São fornecidas apenas as entradas, conforme:

$$D = \{x_i\}_{i=1}^N \quad (2.2)$$

O objetivo é encontrar padrões nos dados, o que pode ser chamado de descoberta do conhecimento. Este é um problema menos definido, pois não é informado quais tipos de padrões procurar, e não há uma métrica de erro a ser usada, onde podemos comparar nossa previsão de y para um dado x com o valor observado.

2.2 ALGORITMO K-MEANS

A clusterização é uma técnica em análise de dados que tem como objetivo o agrupamento de objetos similares em conjuntos chamados clusters. Entre os vários métodos mais conhecidos e usualmente utilizados com esta finalidade, está o método k-means, conforme podemos consultar nos artigos Prabhu e Anbazhagan (2011), Brooks et al. (2018), Capó, Pérez e Lozano (2015), Ali e Kadhum (2015), Mam, G. e Saxena (2017), Martanto et al. (2020).

A primeira vez que o termo “k-means” foi usado, pode ser atribuído a MacQueen (1967), porém a ideia do método possa ser ligada inicialmente a Steinhaus (1957). Stuart Lloyd, em 1957 propôs o algoritmo padrão como técnica para modulação por código de pulso, entretanto, o artigo foi publicado em jornal no ano de 1982, que pode ser consultado em

Lloyd (1982). Contudo é importante citar que Edward W. Forgy publicou essencialmente o mesmo método em 1965, e por esta razão, por vezes o método é referido como o algoritmo Lloyd-Forgy (1965).

O método k-means foi aplicado para encontrar locais candidatos em uma planta industrial, conforme Hartigan (1975) em Clustering Algorithms, e assim descobriram que a abordagem é consistente e confiável, principalmente para modelos que lidam com o problema de cobertura. Como k-means pode ser escalável e eficiente para grandes conjuntos de dados, pois encontra um centroide ideal, que por consequência diminui a soma das distâncias entre os pontos dos dados estudados.

k-means clustering é um método de quantização vetorial, que visa observar “n” partição em “k” clusters, em que cada observação “n” pertence ao cluster chamado de centroide, servindo como um protótipo do cluster. Em outras palavras, o agrupamento K-means é um método para encontrar clusters e centros de clusters em um conjunto de dados não rotulado Hastie, Tibshiran e Friedman (2009). O método muitas vezes acaba encontrando um centroide ideal, que minimiza a soma das distâncias entre os pontos de dados. Podemos abordar o funcionamento do algoritmo em quatro etapas, sendo que o início se dá pela atribuição da quantidade de centroides inicial, logo após se atribui cada ponto do bando de dados ao centroide mais próximo, formando k clusters. Em seguida é calculado novos centroides como os centros dos pontos pertencentes a cada cluster e por fim repete-se os passos um e dois até que a convergência seja alcançada, ou até que o critério pré-estabelecido tenha sido alcançado.

Formalizando matematicamente, o método de k-means como:

$$\text{Min } c = \sum_{i=1}^k \sum_{j=1}^p (x_{ij} - \mu_i)^2 \quad (2.3)$$

Onde:

$$\mu_i = \frac{1}{n_i} \sum_{j=1}^p x_{ij} \quad (2.4)$$

Onde:

- μ_i representa o vetor centroide do cluster $i = 1, 2, 3, \dots, k$;
- x_{ij} são os vetores dos pontos de dados $j = 1, 2, 3, \dots, p$ pertencentes ao cluster i ;

- n_i é o número total de pontos (tamanho do cluster) em i .

O objetivo é dividir x_j pontos totais de dados em k clusters totais, onde cada cluster $i = 1, 2, 3, \dots, k$ é representado pelo seu centroide μ_i . Inicialmente, os primeiros centroides são escolhidos aleatoriamente (naive k-means), convergindo para soluções de mínimos locais a cada iteração. Unindo (2.4) em (2.3), temos a representação formal do processo de otimização:

$$\text{Min } c = \sum_{i=1}^k \sum_{j=1}^p \left(x_{ij} - \frac{1}{n_i} \sum_{j=1}^p x_{ij} \right)^2 \quad (2.5)$$

Ou, de maneira semelhante, minimizando o desvio quadrático pareado dos pontos (x_{ij}, y_{ij}) em um mesmo cluster $\forall (x_{ij} \neq y_{ij})$:

$$\text{Min } c = \sum_{i=1}^k \frac{1}{n_i} \sum_{j=1}^p (x_{ij} - y_{ij})^2 \quad (2.6)$$

A cada iteração, os centroides são recalculados com base na média dos pontos atribuídos a cada cluster. Cada ponto x_j é atribuído ao cluster cujo centroide é o mais próximo, minimizando a distância euclidiana em (2.5). Depois que todos os pontos são atribuídos a um cluster, o centroide de cada cluster é recalculado com base nos pontos atualmente atribuídos. O novo centroide de cada cluster é a média aritmética dos pontos atribuídos a ele:

$$\mu_i^{(t+1)} = \frac{1}{n_i^{(t+1)}} \sum_{j=1}^p x_{ij}^{(t+1)} \quad (2.7)$$

Onde:

- $\mu_i^{(t+1)}$ representa o vetor centroide do cluster $i = 1, 2, 3, \dots, k$ na iteração $(t + 1)$;
- $x_{ij}^{(t+1)}$ são os vetores dos pontos de dados $j = 1, 2, 3, \dots, p$ pertencentes ao cluster $i = 1, 2, 3, \dots, k$ na iteração $(t + 1)$;
- $n_i^{(t+1)}$ é o número total de pontos (tamanho do cluster) em $i = 1, 2, 3, \dots, k$ na iteração $(t + 1)$.

O processo de atualização é repetido até que os centroides não mudem significativamente (ou seja, quando as mudanças nas posições dos centroides forem menores que um limite pré-definido) ou até que um número máximo de iterações seja atingido.

2.2.1 Método Centroide dos Centroides (k-means-c²)

O método de “centroide dos centroides” é uma abordagem útil para determinar um ponto central geral em um conjunto de clusters, especialmente quando se trabalha com problemas de localização de serviços. Em contextos como a alocação de unidades de segurança pública, o objetivo principal desse método é identificar um local estratégico que minimize a distância média para as diversas áreas de atendimento, considerando a distribuição de chamadas de ocorrências de diferentes tipos de crimes entre diferentes regiões geográficas.

Para calcular o centroide dos centroides finais, ou seja, a média dos k -centroides finais $\mu_1, \mu_2, \dots, \mu_k$ dos clusters obtidos ao final do algoritmo, temos:

$$c_{\text{global}} = \frac{1}{c_k} \sum_{i=1}^k \mu_i \quad (2.8)$$

Onde:

- c_{global} representa o centroide dos centroides (centroide global) dos k clusters finais;
- μ_i são os centroides de cada cluster $i = 1, 2, 3, \dots, k$;
- c_k é o número total de centroides finais k (tamanho das amostras de clusters).

Inserindo a equação (2.7) em (2.8), temos que, após $t + 1$ iterações, nossa medida de localização global pode ser formalmente representada por:

$$c_{\text{global}} = \frac{1}{c_k} \sum_{i=1}^k \frac{1}{n_i^{(t+1)}} \sum_{j=1}^p x_{ij}^{(t+1)} \quad (2.9)$$

Por meio da equação (2.9), busca-se identificar um ponto central médio que minimize a distância global entre múltiplos clusters de dados, permitindo uma localização mais equilibrada e estratégica. Esse ponto representa o centro de massa dos centroides de clusters finais, fornecendo uma referência central ideal para posicionamento e alocação de recursos ou serviços em um espaço geográfico.

Na alocação de unidades de segurança pública, o *centroide dos centroides* oferece uma posição estratégica que, em média, equilibra a distância entre várias áreas de alta demanda, melhorando a cobertura geográfica e reduzindo tempos de resposta. A localização

próxima a esse ponto médio permite a uma unidade centralizada responder mais rapidamente a incidentes em diferentes clusters, distribuindo o acesso ao serviço de forma mais uniforme em um território amplo e heterogêneo.

Clusterizações baseadas no centroide global podem dar suporte como referência para o planejamento, sendo adaptáveis conforme as necessidades de segurança mudam, permitindo realocações periódicas e estratégicas que refletem a dinâmica de demandas em cada região.

No contexto de Matelândia, ao definir diferentes números de clusters k para cada tipo de crime com base na gravidade social — como 4 clusters para crimes de maior severidade, como o crime de estupro, 3 clusters para crimes de roubo, 2 clusters para crimes como violência doméstica (Maria da Penha) e 1 cluster para furto — o cálculo do centroide global naturalmente será influenciado pela quantidade e distribuição desses clusters. Crimes mais severos, como estupro, que recebem uma quantidade maior de clusters, terão mais centroides próximos de áreas de alta incidência e severidade, o que acaba trazendo o centroide global mais próximo dessas regiões.

Essa distribuição diferenciada concentra o centroide global em locais onde há maior densidade de clusters de crimes mais graves, uma vez que o cálculo final do centroide dos centroides tende a ponderar áreas com mais centroides. Dessa forma, a localização estratégica sugerida pelo centroide global acaba refletindo, de maneira indireta, uma prioridade geográfica para áreas com alta incidência de crimes graves. Esse ponto final de referência, aqui chamado de centroide dos centroides, permite um posicionamento de unidades de segurança que atenda mais efetivamente às áreas de maior risco social, facilitando uma resposta mais rápida e alinhada à gravidade das ocorrências.

3 METODOLOGIA

Foram selecionados os dados a serem trabalhados de tal forma à diminuir o ruído das informações e focar em crimes com maior gravidade e que causam maior ônus as vítimas. Se fez necessário filtrar quais ocorrências policiais que integrariam o banco de dados a ser estudado. É inegável que crimes contra a vida são os mais graves que possam ocorrer, o homicida, por exemplo, não prejudica apenas sua vítima, mas também seus familiares, como pode ser consultado em Costa, Njaine e Schenker (2017), afinal são crimes irreversíveis e que causam grande prejuízo para todos os envolvidos, sendo assim ocorrências envolvendo homicídio culposo e homicídio doloso, seja em sua forma tentada ou consumada, devem integrar os dados estudados.

São relevantes para o estudo também, crimes de estupro tentado e consumado, e qualquer tipo de ato libidinoso que veio a constranger alguém, conforme as autoras Oliveira e Resende (2020) discutem em sua literatura, este tipo de violência contra a mulher pode ser tão grave quanto qualquer outro tipo de investida, e por esta razão serão dados que estarão presente neste material.

Deve-se incluir crimes de vias de fato (quando há agressão mútua entre os envolvidos) em situação de violência doméstica, ameaça, violência patrimonial, violência verbal e descumprimento de medida protetiva, todas estas modalidades englobam um tipo de crime que neste trabalho será referido com crimes tipo “Maria da Penha”, que são delitos cometidos por parceiros ou ex-parceiros contra mulheres, como também cometido por homens contra mulheres de sua família, e devem ter presença nos dados coletados.

Será incluído crimes contra o patrimônio, entre eles, roubo e furto. Roubo tendo uma prioridade maior, pois envolve violência para subtrair os bens alheios, enquanto furto representa um crime mais “brando”, pois não há violência envolvida no processo de subtração do bem alheio.

Para o estudo aqui proposto, alguns crimes foram deixados de fora da coleta de dados, pois seu grau de periculosidade ou a forma como isso afeta a vida das pessoas seja menos nocivo, pensando em um contexto que gere violência. Entre os crimes, pode-se destacar a contravenção penal de perturbação do sossego ou trabalho alheio, que apesar de ser a ocorrência em que a polícia militar é mais requisitada no estado, representando 32,87% do total de 226.856 requisições feitas no ano de 2020, não gera danos graves aos envolvidos, e

na maioria das vezes, não evolui para um fato que onere o estado com processos judiciais ou investigações por parte da polícia civil, como é possível aprofundar e analisar em Junior (2021). Outro crime que apesar de causar transtornos para sociedade, foi descartado para o presente estudo, é o crime de vandalismo, pois assim como o a perturbação de sossego, tende a não gerar fatos violentos, que necessitem de uma maior investigação ou interferência da máquina pública.

Sendo assim, neste trabalho, foram selecionados os crimes de homicídio, em todas suas modalidades, estupro em todas suas modalidades, violência doméstica em todas as suas formas praticadas e os crime de roubo e furto, em todas suas modalidades.

Para familiarizar os que desconhecem como a legislação Brasileira define estes crimes, será explicitado os artigos criminais escolhidos para o trabalho. Como o estudo não é voltado à área de direito, será realizada apenas a apresentação do artigo, conforme apresentado no Código Penal Brasil de 1940 consultado através do site do Planalto (Brasil (1940)), sem suas majorantes ou quaisquer outras informações que possam estar presente no corpo do artigo.

Respeitando a ordem pela qual foram apresentadas anteriormente:

- Homicídio simples - Art. 121. Matar alguém:

Pena - reclusão, de seis a vinte anos.

- Estupro - Art. 213. Constranger alguém, mediante violência ou grave ameaça, a ter conjunção carnal ou a praticar ou permitir que com ele se pratique outro ato libidinoso (Redação dada pela Lei nº 12.015, de 2009):

Pena - reclusão, de 6 (seis) a 10 (dez) anos. (Redação dada pela Lei nº 12.015, de 2009)

- Lesão corporal - Art. 129. Ofender a integridade corporal ou a saúde de outrem:

Pena - detenção, de três meses a um ano.

- Roubo - Art. 157 - Subtrair coisa móvel alheia, para si ou para outrem, mediante grave ameaça ou violência a pessoa, ou depois de havê-la, por qualquer meio, reduzido à impossibilidade de resistência:

Pena - reclusão, de quatro a dez anos, e multa.

- Furto - Art. 155 - Subtrair, para si ou para outrem, coisa alheia móvel:

Pena - reclusão, de um a quatro anos, e multa.

Cada lei possui muito mais informações quanto a situações específicas, como por exemplo as condições e motivações que levaram ao cometimento do crime, seja na pena aplicada ou na nomenclatura que se dá para casos especiais. Mas para efeito deste estudo, tornam-se irrelevantes estas informações adicionais, exceto quando formos analisar as situações de violência doméstica.

Como forma de esclarecer um fato sobre os dados aqui estudados, devemos observar que apesar de várias tipificações criminais se apresentarem como ameaça, lesão corporal, injúria, desobediência de medida protetiva, para o trabalho aqui proposto, serão todos enquadrados como crimes de “Maria da Penha”, visto que são crimes que foram cometidos contra cônjuges ou mulheres da família.

Conforme observa-se no diagrama explicitado na figura 3, os passos que irão compor o presente trabalho terá início com a análise e escolha dos dados criminais que farão parte do escopo a ser estudado, subsequente a esta escolha destes dados criminais, será realizado o pedido junto ao setor responsável da policia militar do Paraná. Com acesso aos dados fornecidos, atribui-se graus de importância para cada tipificação criminal, sendo feita de acordo com a gravidade e severidade destes crimes. Na etapa seguinte serão listados e organizados as informações obtidas, sendo posteriormente realizado a análise deste banco de dados e o georreferenciamento usando as coordenadas de latitude e longitude para isto. Após será aplicado o banco de dados no método k-means, onde poderão ser realizados os estudos pertinentes, sendo então executado as duas formas de análise, sendo elas definindo graus de importância, e sem definir os graus de importância, onde será possível alcançar os objetivos propostos inicialmente, sendo finalizado, interpretado e discutido os resultados encontrados.

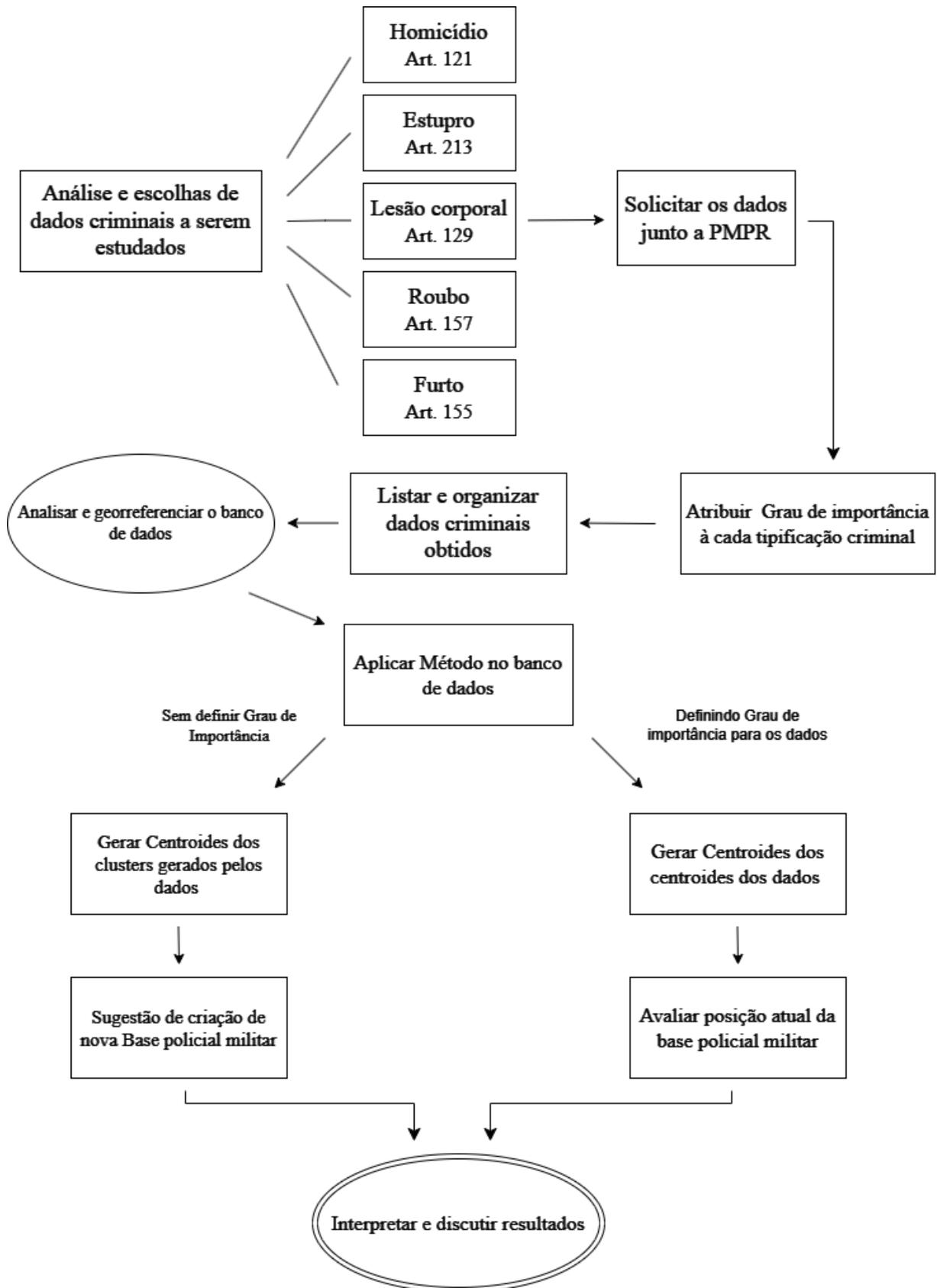


Figura 3 – Diagrama passo a passo da metodologia

4 RESULTADO

4.1 ANÁLISE EXPLORATÓRIA DE DADOS

Os dados criminalísticos foram adquiridos com o setor de comunicação da Polícia Militar do Estado do Paraná, sendo requerido através do site da instituição no setor de ouvidoria Paraná (2024). O período de tempo escolhido foi de outubro de 2022 à outubro de 2023, pois é uma linha temporal que não sofreu influência da Pandemia de Covid-19, que teve início no começo de 2020 e fim em meados de 2022.

Será feito uma análise dos dados que foram repassados pela Polícia Militar do estado do Paraná. Foi fornecido um documento em formato PDF, com tabelas que continham as seguintes informações: município; fato inicial; ano do fato; mês do fato; dia do fato; dia da semana; hora do fato; bairro; logradouro; número; título penal e natureza. Os dados vieram agrupados por ordem cronológica e por tipo de crimes, o que facilita a leitura e entendimento, podendo ser assimilado algumas informações pertinentes, como por exemplo, o crime que mais foi atendido no município de Matelândia, foi de “Maria da Penha”, com 173 casos registrados, seguido pelos crimes furto com 169 registros, roubo com 26 casos e o crime de estupro com 15 casos registrados. Outra informação interessante que pode-se constatar analisando os dados, é que o dia da semana em que ocorre mais crimes nesta cidade, é no sábado, com 64 ocorrências, e em segundo lugar o domingo juntamente com segunda-feira com 57 ocorrências. O dia com menos ocorrências registradas foi na terça-feira, com 48 registros, conforme tabela 1 e figura 8.

Tabela 1 – Dias da semana e o número de ocorrências

Dia da Semana	Segunda-Feira	Terça-Feira	Quarta-Feira	Quinta-Feira	Sexta-Feira	Sábado	Domingo
Ocorrências Atendidas	57	48	52	56	49	64	57

Para o objetivo proposto nesta pesquisa, a informação necessária são os endereços em que esses crimes ocorreram, pois será através deles que será feito o mapa criminal e verificado a necessidade ou não, de uma nova base policial e poderá ser avaliado se a atual base esta em uma localização ideal.

Será iniciado o processo de geolocalização dos endereços mencionados no quadro de dados, e primeiro problema para confecção do estudo. Como estes dados repassados são os boletins de ocorrências registrados por policiais militares que atenderam as situações,

por vezes, tais policiais não se atentaram em dar a devida atenção para um registro fiel e correto dos endereços atendidos, gerando assim muitos erros de português, ou usando escritas diferentes da oficial. Tal fato não permite que a geolocalização seja feita, tendo que num primeiro momento encontrar e corrigir estes equívocos, para então conseguir georeferenciar estes endereços. Um exemplo destes que chamou atenção, por ser um erro dos sites de georeferenciação e não um erro humano, foi o bairro “Vila Pasa”, que apesar de estar escrito de forma correta no quadro de dados, nos sites que realizam a geolocalização, está descrito como "Vila Pazza", o que é um erro no site e não dos policiais que preencheram os dados da ocorrência.

Outra questão que apresentou certa divergência entre a localização real e a geolocalização obtida através do aplicativo Maps, foi no quesito numeração do local onde ocorreu o crime, visto que por algum motivo, provavelmente por falta de atualização no site, em alguns casos acontece uma diferença de alguns metros entre a localização real e a localização virtual. Apesar desta pequena desconformidade, não foi algo que impediu ou inviabilizou o andamento da pesquisa, visto que para os objetivos aqui propostos, não é necessária uma exatidão absoluta.

Para começar a análise do banco de dados objeto de estudo, será apresentado a forma pelo qual foi enviado inicialmente pela Polícia Militar do Paraná (2024), o qual foi feito através de um arquivo formato PDF, conforme figura 4:

DADOS ESTATÍSTICOS CRIMINAIS - FURTO – MUNICÍPIO DE MATELÂNDIA

QUADRO 01: DEMONSTRATIVO DE DADOS DE OCORRÊNCIAS CRIMINAIS – FURTO – MUNICÍPIO DE MATELÂNDIA – PERÍODO: OUTUBRO A DEZEMBRO DO ANO 2022 – CAPE-FRONTIEIRA/SESP.

Município	Fato Inicial	Ano Fato	Mês Fato	Dia Fato	Dia da Semana	Hora Fato	Bairro	Logradouro	Número	Título Penal	Natureza
MATELANDIA	16/10/2022 21:30:00	2022	out	16	dom	21	ZONA RURAL	LINHA ORO	1	CRIMES CONTRA O PATRIMONIO	FURTO SIMPLES
MATELANDIA	20/10/2022 13:30:00	2022	out	20	qui	13	CENTRO	RUA NAPOLEAO LAUREANO	578	CRIMES CONTRA O PATRIMONIO	FURTO SIMPLES
MATELANDIA	27/10/2022 12:00:00	2022	out	27	qui	12	AGROCAFFEEIRA	AVENIDA PRESIDENTE VARGAS	183	CRIMES CONTRA O PATRIMONIO	FURTO SIMPLES

Figura 4 – Dados formato inicial

Estes dados foram extraídos manualmente, e isto foi feito para ser trabalhado na linguagem python, a qual foi escolhida para desenvolver os estudos apresentados nesta dissertação.

As informações repassadas ocupavam doze colunas, sendo elas informando os seguintes dados:

- Município: referente a qual cidade se deu o fato;

- Fato inicial: Mostra o dia e a hora estimada em que o fato ocorreu;
- Ano do fato: Mostra qual ano ocorreu o fato;
- Mês do fato: Em qual mês do ano houve o fato;
- Dia do fato: Repete o dia do fato, mas isoladamente;
- Dia da semana: Informa qual dia da semana aconteceu;
- Hora do fato: Repete a hora estimada que o fato ocorreu, porém só a hora;
- Bairro: Qual bairro do município ocorreu;
- Logradouro: Qual a rua em que aconteceu;
- Número: A numeração do endereço que ocorreu, ou numeração aproximada;
- Título penal: Informa contra o que aquele crime atentou;
- Natureza: Qual crime efetivamente foi cometido.

É possível verificar que há uma coluna que detém as informações de forma aglutinada de outras quatro colunas, mas isso inicialmente não é um problema, visto que suprimir uma coluna em um banco de dados é algo simples.

Alguns apontamentos acerca dos dados estudados, qual é o crime com mais incidência em nosso banco.

Como exposto na figura 5, podemos verificar através dos gráficos, que “Maria da Penha” e furto apresentam número de ocorrência superiores aos outros dois tipo de crime, roubo e estupro. Através destas informações é possível ter uma perspectiva de quanto cada natureza criminal representa em relação ao total dos crimes estudados. “Maria da Penha” por exemplo, representa 45, 17% dos crimes estudados, o crime de furto representa 44, 12%, roubo aparece com 6, 79%, estupro 3, 92% e homicídio com 0% pois não teve ocorrências registradas, conforme explicitado na tabela 2

Tabela 2 – Total de ocorrências por tipo de naturezas criminais

Natureza Criminal	Furto	Roubo	“Maria da Penha”	Estupro
Registros	169	26	173	15

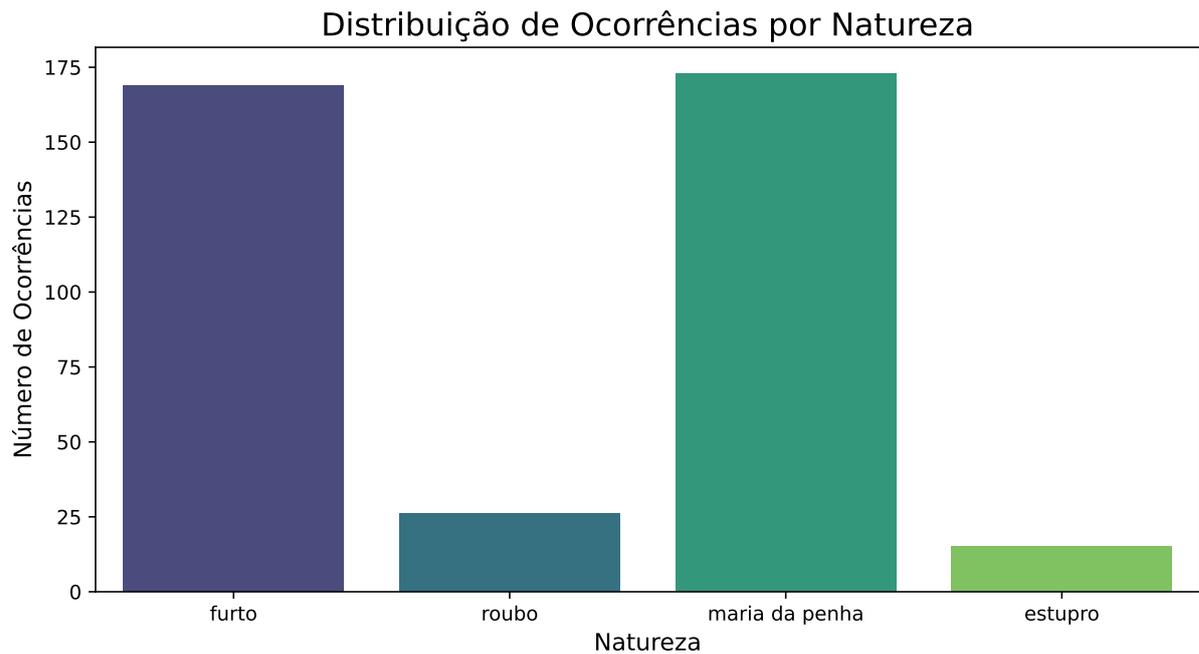


Figura 5 – Distribuição de ocorrências por natureza criminal

É importante ressaltar que na faixa temporal escolhida para este estudo, não foi registrado nenhum homicídio ou tentativa de homicídio, sendo assim ficando omitido de todas as análises futuras, ao invés de ser sempre referida como tendo zero ocorrências.

A próxima figura 6 irá apresentar um panorama de como estes dados estão distribuídos pelos meses do ano.

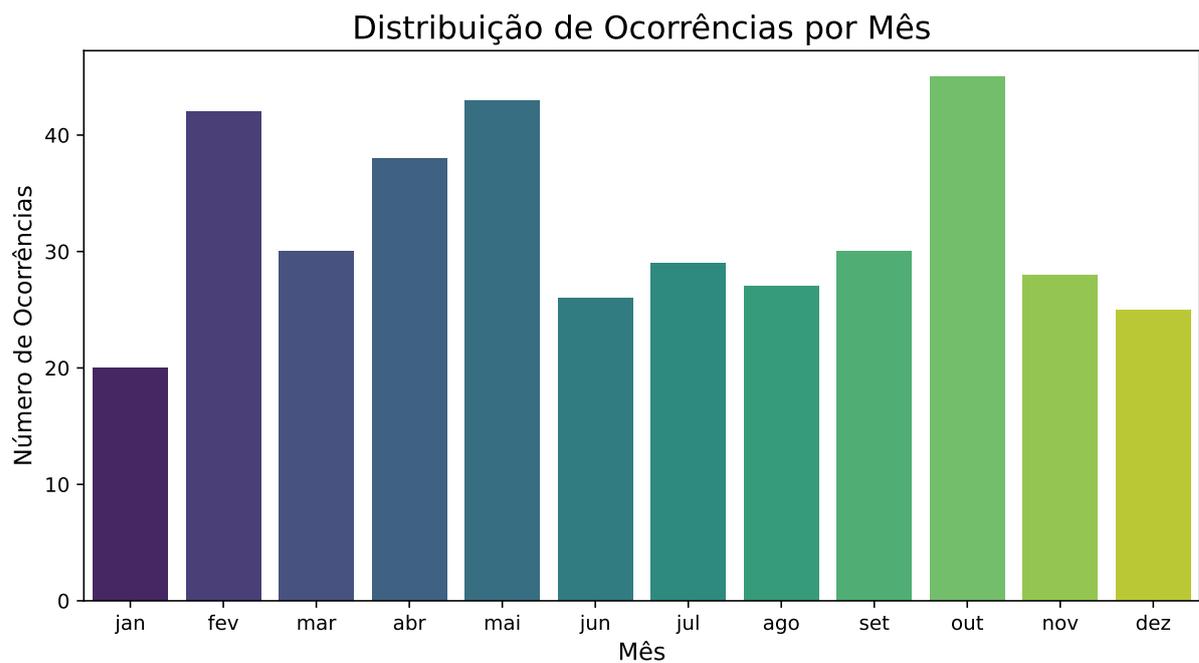


Figura 6 – Distribuição de ocorrências criminais por mês do ano

Neste gráfico representado na imagem 6 é possível visualizar que o mês com mais crimes atendidos, foi o de outubro com 45 ocorrências, seguido por maio com 43 e fevereiro com 42 fatos. Já o mês com menos ocorrência é o mês de janeiro, com 20 ocorrências.

Quando analisamos as informações disponíveis no banco de dados, é feito um estudo da distribuição dos crimes de acordo com os dias do mês, vide figura 7, e é possível analisar que o dia com mais fatos é o dia 5, com um total de 22 ocorrências. Já o dia em que menos ocorrem crimes na cidade de Matelândia, é o dia 31 com 5 ocorrências. Porém como o dia 31 não recorre todos os meses do ano, é interessante ressaltar que o segundo dia com menos crimes no mês, é o dia 22, com 6 ocorrências.

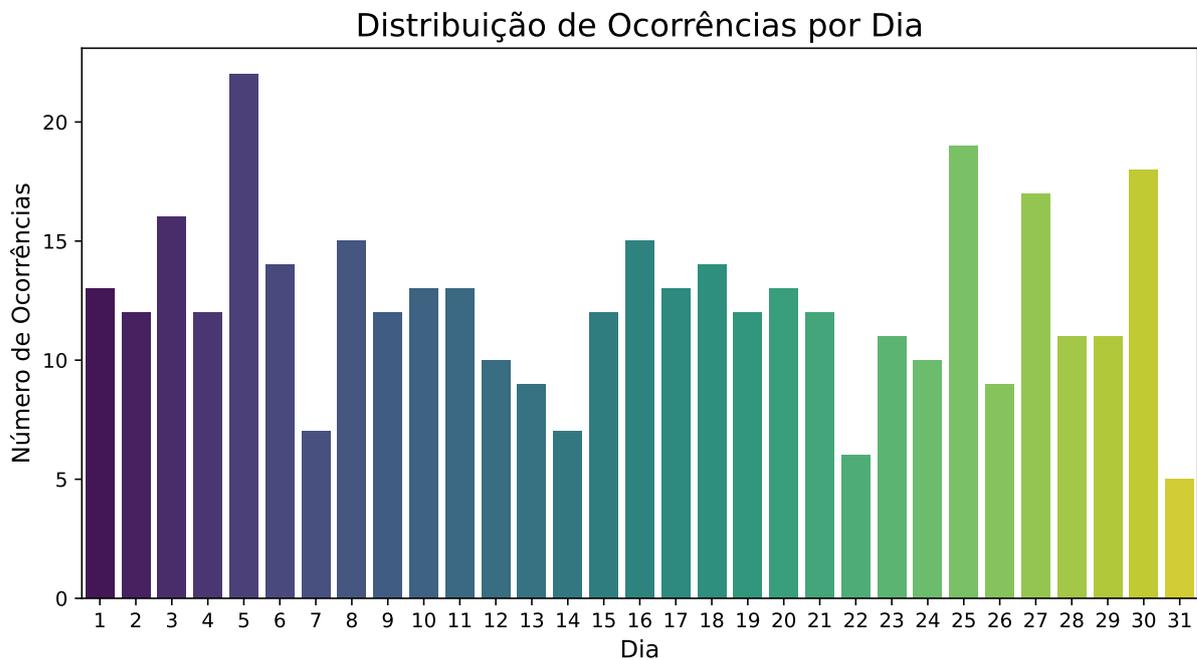


Figura 7 – Distribuição de ocorrências criminais por dia do mês

Pode-se verificar ainda, qual a incidência de acordo com os dias da semana. E ao fazê-lo é possível afirmar que o dia com mais ocorrências é o sábado, com um total de 64 registros, e o dia com menos crimes relatados é na terça-feira, com 48 chamados, este resultado está representado na figura 8.

Na figura 9 pode-se analisar qual a dinâmica dessas ocorrências hora a hora no decorrer do dia, e ao fazê-lo verificasse que a hora do dia com mais incidência de crimes é as 14 : 00 horas com 33 registros e o momento do dia com menos crimes é as 3 : 00, 4 : 00 e 6 : 00 da manhã, ambos os horários com estatísticas iguais, com 6 registros.

Para melhor visualizar esta dinâmica da distribuição criminal em relação as horas do dia, será feito a mesma análise anterior, porém discriminando cada natureza criminal.

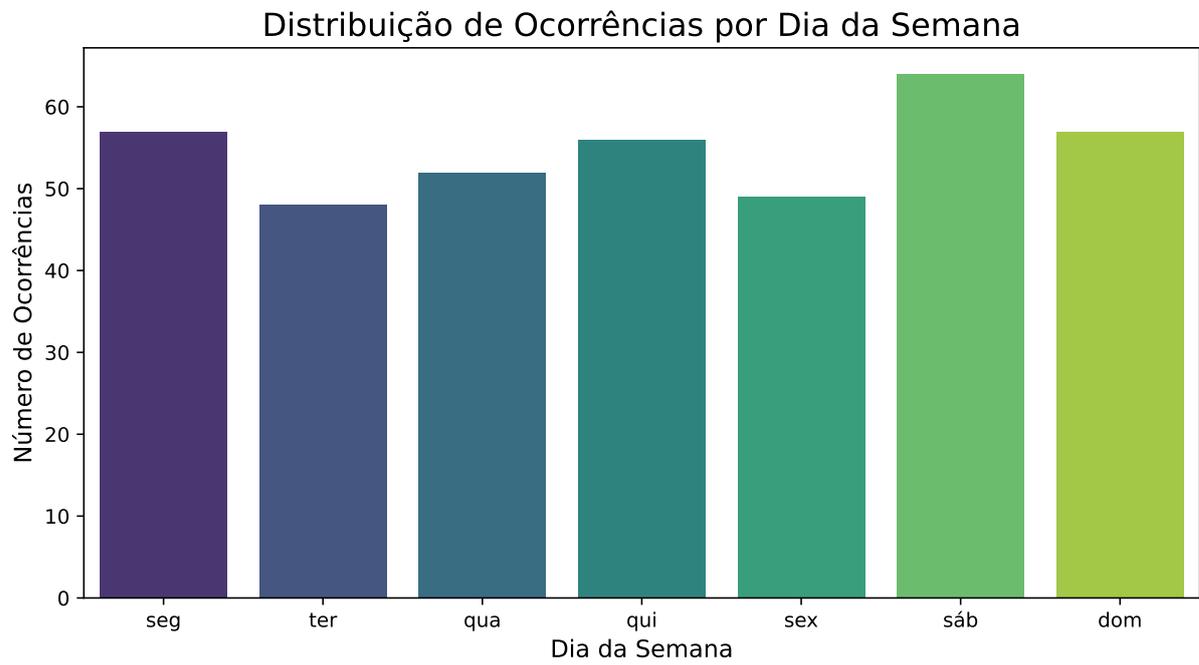


Figura 8 – Distribuição de ocorrências por dia da semana

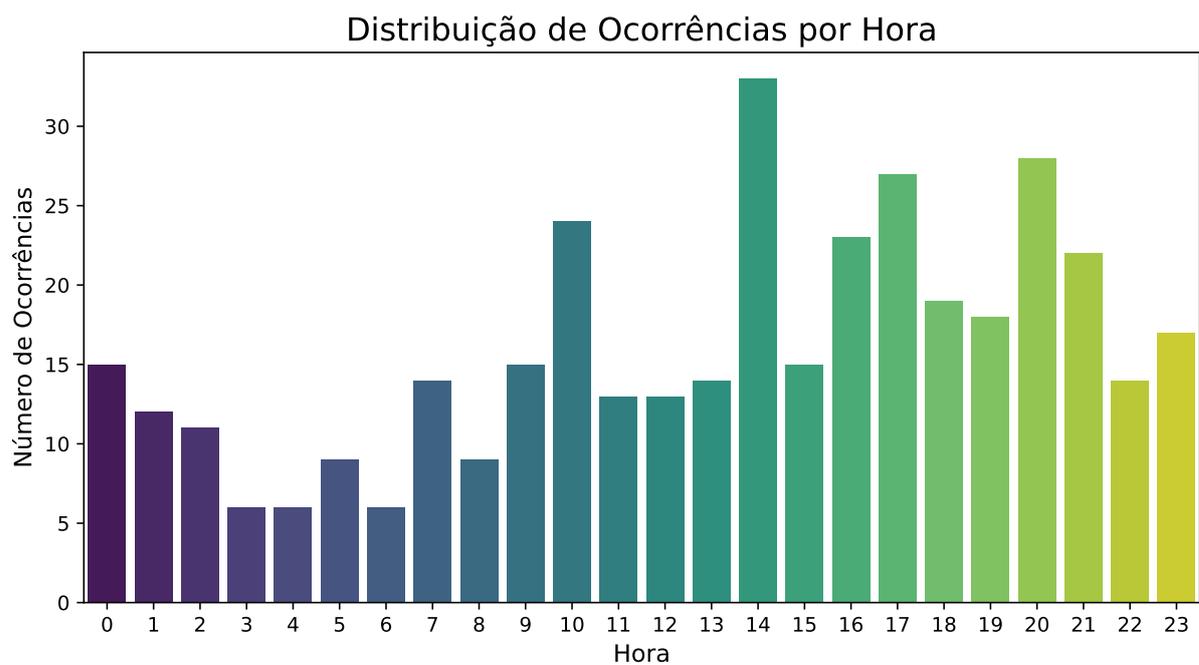


Figura 9 – Distribuição de ocorrências por hora do dia

Inicialmente o crime de furto, conforme figura 10:

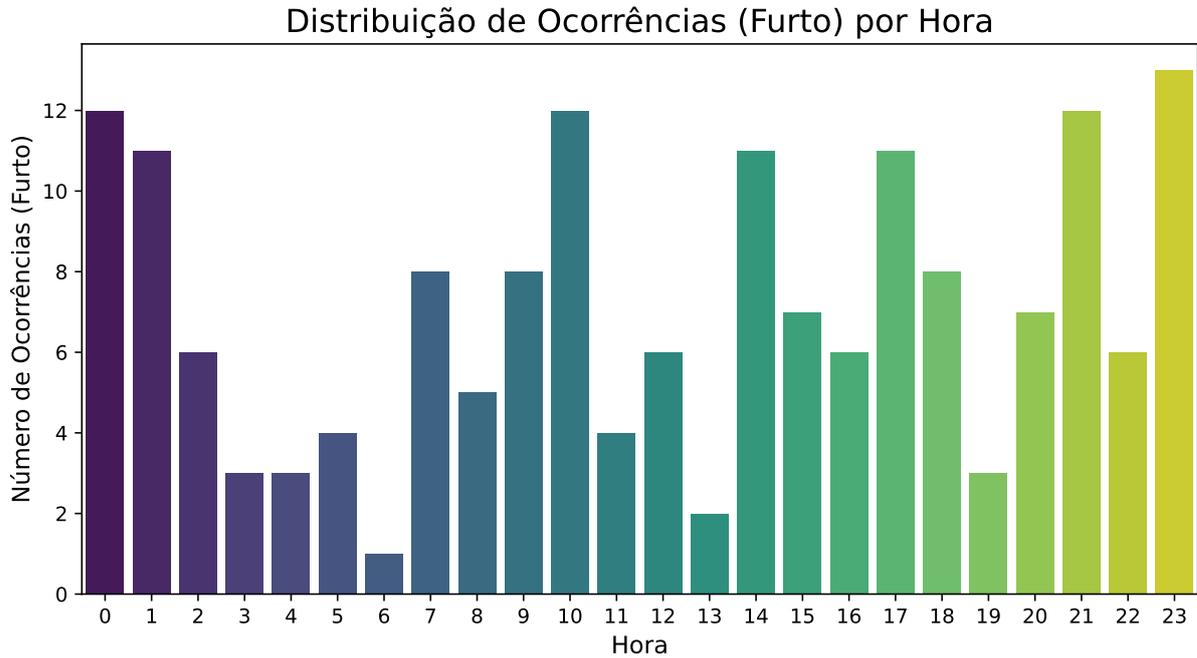


Figura 10 – Distribuição de ocorrências por hora do dia crime furto

Os dados mostram que a maior ocorrência de furtos, ocorre às 23 : 00 horas, seguido pela segunda maior ocorrência que ocorre as 00 : 00. O horário com menos ocorrências são as 06 : 00, com apenas 1 dado registrado.

Na figura 11, os dados referentes ao crime de “Maria da Penha”.

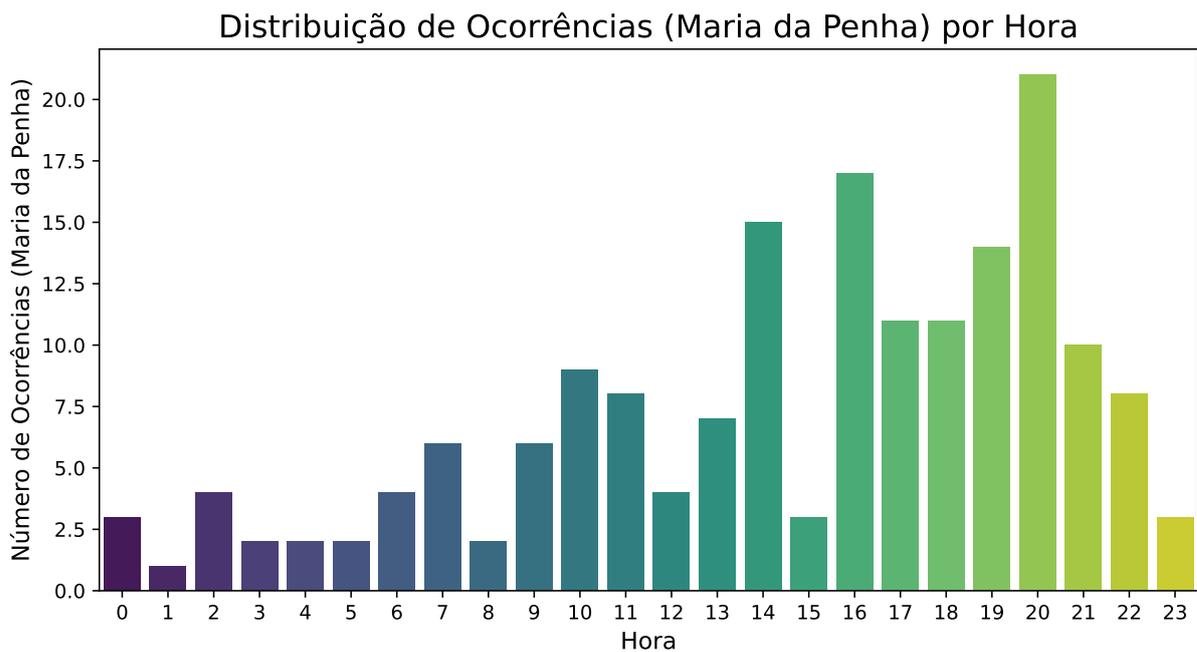


Figura 11 – Distribuição de ocorrências por hora do dia crime “Maria da Penha”

Nesta situação os dados mostram que o pico de ocorrências é às 20 : 00 horas, com 22 registros, e o horário com menos incidência é às 01 : 00 com 1 registro.

Quando analisado hora a hora dos dados referente ao crime de roubo, observa-se que sua maior incidência se dá no período das 14 : 00 horas, sendo que às 13 : 00 e 15 : 00 ambas com 4 registros, ficam em segundo lugar com horário com mais incidências, conforme figura 12.

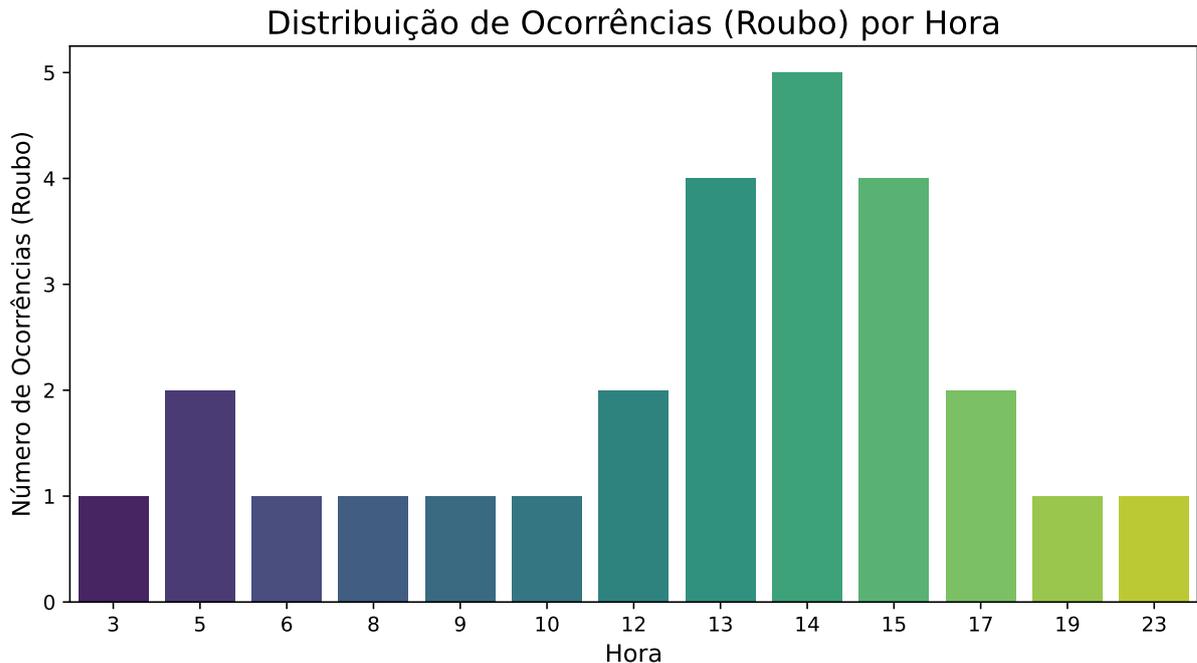


Figura 12 – Distribuição de ocorrências por hora do dia crime roubo

Ao analisar esta distribuição por hora com relação aos crimes de estupro, é possível observar:

Na figura 13 é possível observar que a maior incidência destes crimes é relatada às 17 : 00, com 3 registros.

Ao analisar o banco de dados focando nos endereços em que ocorreram os crimes, pode-se visualizar no gráfico a seguir, onde foi feito um ranking com os 10 bairros em que mais foram registrados fatos, conforme figura 14.

É possível concluir que o bairro com mais incidência de crimes, é o centro da cidade de Matelândia, com um total de 107 registros, seguido por Agro cafeeira que conta com 88 registros e em terceiro lugar é o bairro da Vila Nova, que foi acometido por 50 crimes.

Ao analisar estes crimes visando observar em quais ruas eles ocorreram, o banco de dados mostra um fato curioso, conforme o gráfico é possível notar que inverteu-se as posições quando comparado aos bairros.

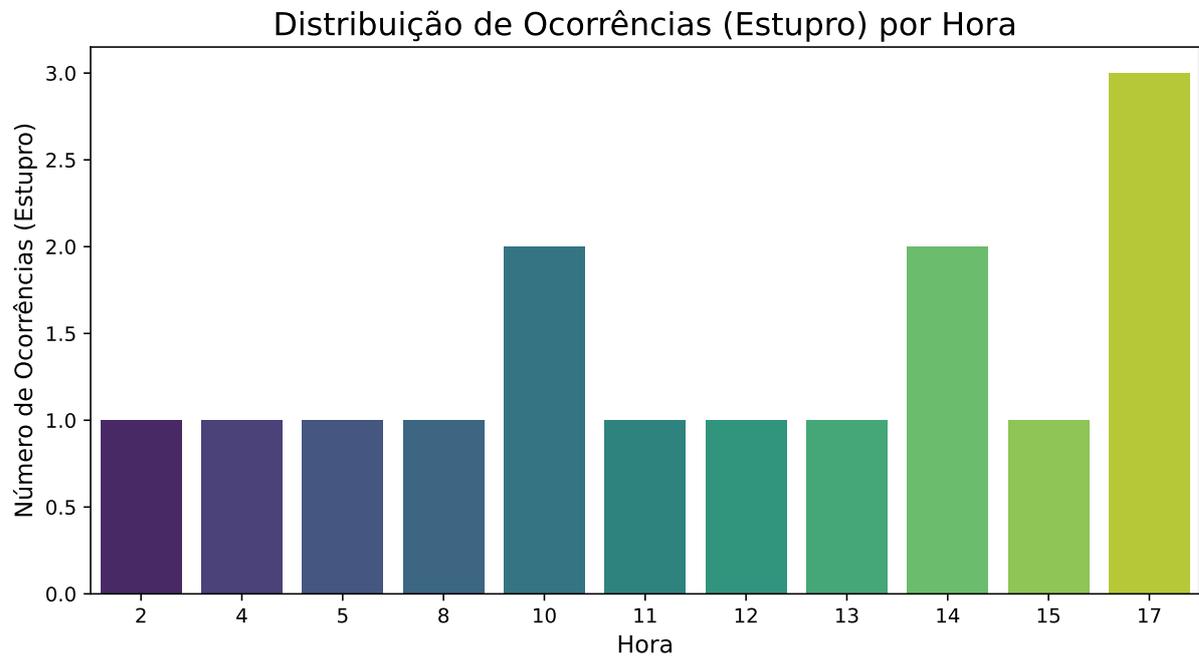


Figura 13 – Distribuição de ocorrências por hora do dia crime estupro

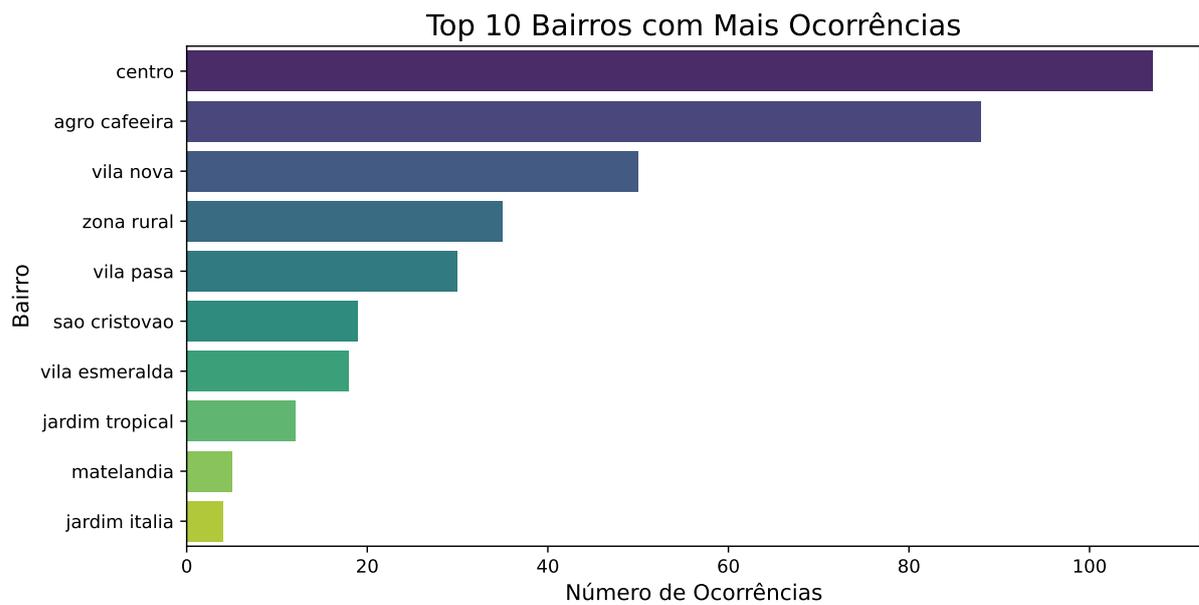


Figura 14 – Top 10 ocorrências distribuídos por bairros

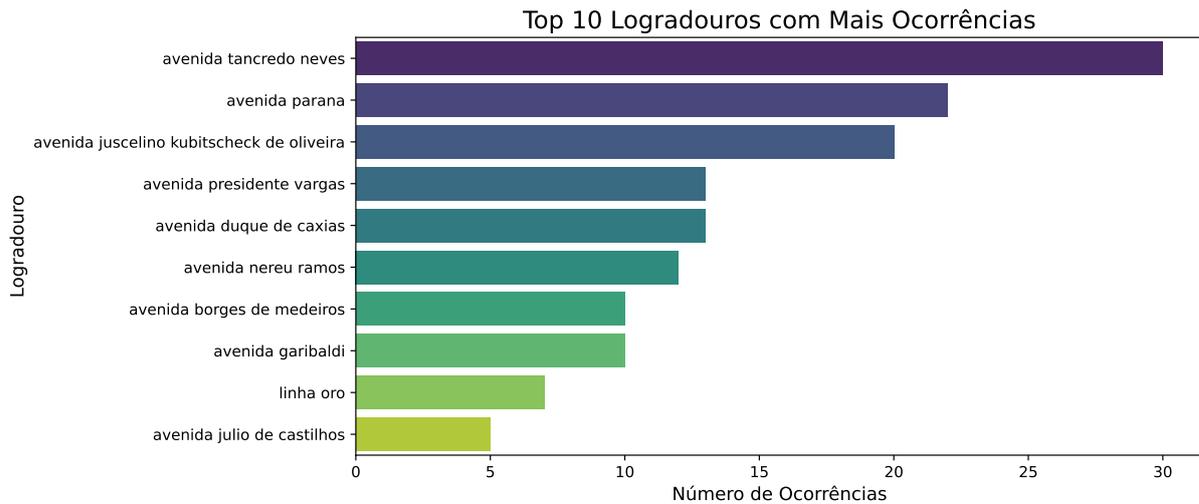


Figura 15 – Top 10 ocorrências distribuídos por ruas

Na figura 15 é possível verificar que a rua com mais incidências criminais é a Avenida Tancredo Neves, com 30 incidência e que esta rua está localizada no bairro Agro Cafeeira, seguida pela Avenida Paraná, com 22 ocorrências que é localizada no Centro de Matelândia e em terceiro lugar como rua com mais registros é a Avenida Juscelino Kubitscheck de Oliveira, localizada na Vila Pasa, com 20 fatos.

Ao distribuir geograficamente os crimes ocorridos na cidade de Matelândia, usando latitude e longitude para isto, fica visível na figura 16 que as maiores ocorrências são de fato dentro da cidade e em seu distrito de Agro Cafeeira, sendo que a área rural e o distrito de Vila Esmeralda ficam em segundo plano em relação ao número de crimes registrados.

Este estudo utiliza o método k-means e o banco de dados proposto, foram manipulados e rodados na linguagem Python através do google colab. Será trabalhado os dados de forma igualitária, sem graus de importância para diferenciar os crimes, ou seja, nesta primeira etapa o crime de furto terá a mesma importância que um crime de estupro, sem distinguir qual é mais danoso pra sociedade ou qual é mais relevante de ser atendido por uma equipe policial militar de plantão. Nesta primeira rodada, também será respondido um dos objetivos propostos para este trabalho, que será concluir através do método escolhido, se é necessário uma segunda base policial militar na cidade de Matelândia, e se esta segunda base teria a localização no distrito de Agro Cafeeira.

Para isto primeiramente será rodado os dados criminais, devidamente geo-localizados, no mapa do município de Matelândia, e o resultado obtido é descrito na figura 17a.

Como a intenção é descobrir onde seria o local ideal da base policial militar, caso todos os crimes fossem tratados de forma igualitária, apenas um cluster foi requerido. O

Mapa Geográfico das Ocorrências

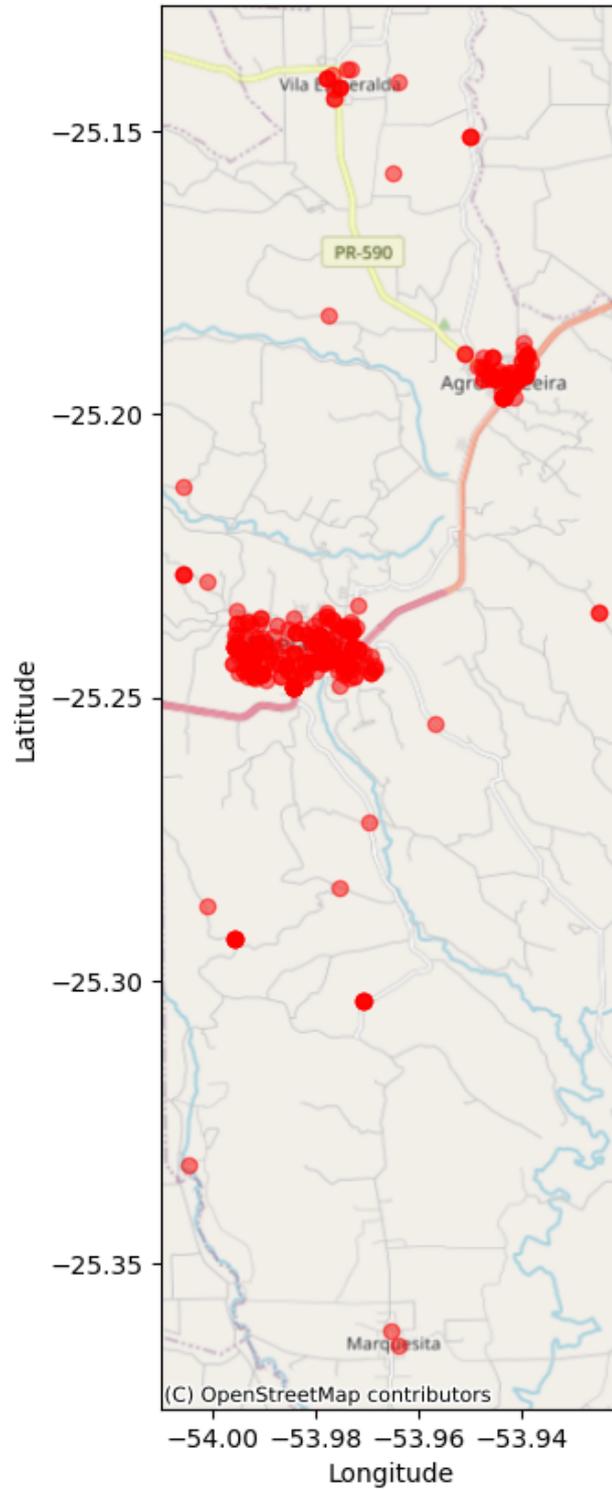
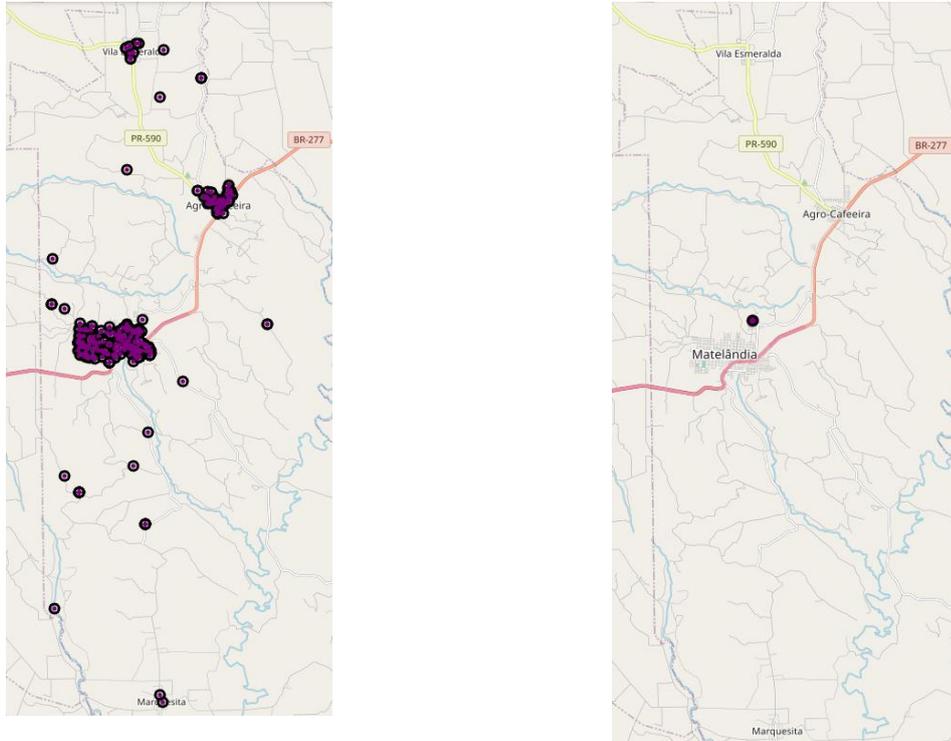


Figura 16 – Mapa geral dos crimes em Matelândia

próximo passo será calcular o centroide destas ocorrências, e assim analisar qual ponto representa a melhor localização dentre todas, a figura 17b mostra qual localização seria esta.



(a) Crimes ocorridos no município com um cluster

(b) Centroide com um cluster

Figura 17 – Estudo feito com 1 cluster

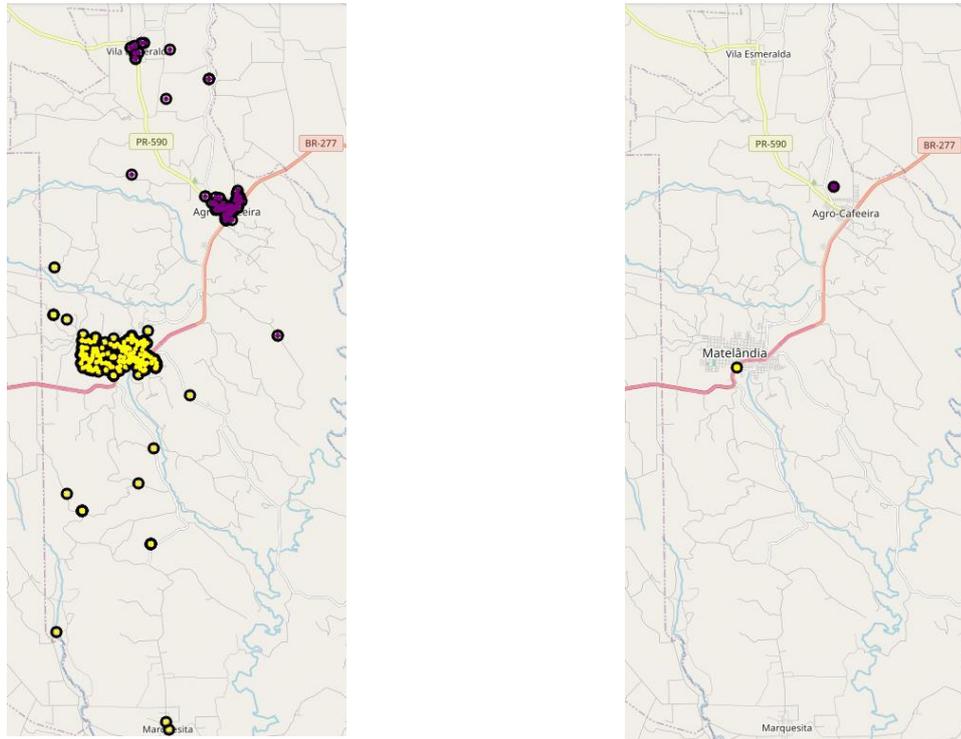
A tabela 3 explicita os dados encontrados como localização ideal segundo o método, o centroide encontrado, em comparação com a localização atual, base policial militar já existente no município.

Tabela 3 – Distância de todos os crimes com um cluster

	Base PM Matelândia	Centroide	Distância Base PM Mat.
Todos os crimes um cluster	-25.2407563,-53.9867984	-25.22958973, -53.97340081	2,5 Km

Na tabela 3 é possível observar que a diferença da posição atual para a posição ideal, segundo o método, está a 2,5 quilômetros de distancia.

Nesta segunda rodada será feita a mesma análise anterior, porém com dois clusters, pois assim poderá ser analisado onde seria o segundo melhor local para ter-se uma base policial militar na cidade. O resultado obtido nesta fase está explicitado na figura 18a.



(a) Crimes ocorridos no município com dois clusters

(b) Centroides com dois clusters

Figura 18 – Estudo feito com dois clusters

Algumas informações devem ser consideradas a cerca desta análise, como por exemplo as distâncias aqui aferidas. Pois por uma limitação do método utilizado para calcular os centroides de nossos dados, o calculo dos centroides não levaram em consideração a distância sendo percorrida através de vias, sendo estas médias calculadas de forma a ignorar esta informação, portanto podendo gerar algumas diferenças caso seja calculado o centroide dos clusters se valendo desta informação, de que o centroide deve respeitar a menor distância levando-se em conta as ruas percorridas. Já as distâncias aqui explicitadas em tabelas, das bases até os centroides encontrados, foram medidas levando em conta esta informação, respeitando portanto a distância sendo percorrida por vias terrestres.

Nesta segunda abordagem, conforme a figura 18b, é possível observar que a posição ideal para a base de Matelândia, agora está apenas à 1,4 quilômetros da posição em que se encontra atualmente, enquanto em relação ao segundo cluster, que teve o centroide localizado no distrito de Agro Cafeeira, a distância sobe para 10,1 quilômetros. Foi possível identificar também que segundo o método, caso fossem dividido as ocorrências em duas bases, que atenderiam estas ocorrências de acordo com a distância em que ocorrem, esta

segunda base que otimizaria o atendimento, estaria localizada em Agro Cafeeira, e o centroide que indica o local ideal para alocar esta base de Agro Cafeeira ficou apenas a 1,5 quilômetros do local onde já está construída esta estrutura, que apenas aguarda o efetivo. Os dados podem ser conferidos na tabela 4 e tabela 5.

Tabela 4 – Distância dos dois centroides em comparação a base PM Matelândia

	Base Policia Militar	Centroide	Distância Base PM
Todos os crimes cluster um	-25.2407563,-53.9867984	-25.24533494, -53.98264292	1,4 Km
Todos os crimes cluster dois	-25.2407563,-53.9867984	-25.18562783, -53.94759608	10,1 Km

Tabela 5 – Distância do centroide em comparação a base PM Agro Cafeeira

	Base PM Agro Cafeeira	Centroide	Distância Base PM
Todos os crimes cluster dois	-25.1922386,-53.9465905	-25.18562783, -53.94759608	1,4 Km

Com os resultados encontrados nesta primeira parte do estudo, foi alcançado um dos objetivos inicialmente proposto neste presente trabalho, onde foi possível concluir que com o volumes de registros criminais atual e sua dispersão pela cidade, é justificada que acaso fosse aberta uma nova base policial militar no município de Matelândia, esta base deveria ser localizada no distrito de Agro Cafeeira, pois isto otimizaria o atendimento das ocorrências nesta cidade.

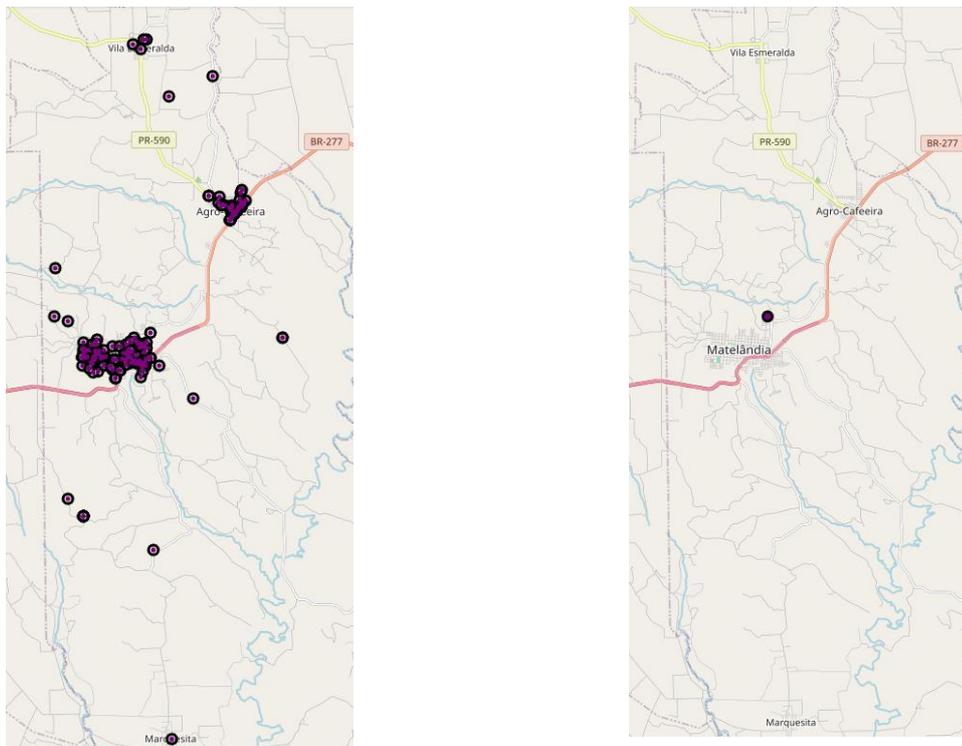
Nesta etapa do estudo, será feita uma análise um pouco diferente em relação a anterior. O objetivo é investigar se a localização atual da única base policial militar presente no município atende de forma satisfatória as ocorrências policiais quando a categorizamos de maneira a dar prioridade as ocorrências que forem consideradas mais urgentes. Como os dados de estudo foram separadas quatro tipificações criminais, sendo elas furto, roubo, “Maria da Penha” e estupro, deverá ser criado uma maneira de classificar por ordem de importância estes crimes.

Conforme debatido anteriormente neste trabalho no capítulo 3, crimes que agridem a pessoa humana terão sempre prioridade em comparação a crimes que são cometidos contra objetos. Por esta razão, para efeitos de estudo será classificado as ocorrências de furto como tendo grau de importância “um”, situações de “Maria da Penha” como grau de importância ”dois”, roubo como grau de importância ”três” e estupro como sendo o mais grave com grau de importância “quatro”.

Estes graus de importância aqui definidos servem para termos uma base com a qual trabalhar, podendo ser modificados de acordo com o objetivo do gestor de segurança pública do local a ser estudado, definindo assim qual a prioridade de acordo com a realidade em que se encontra, ou o objetivo a qual pretende alcançar.

Cada grau de importância nesta etapa, representará um cluster no conjunto de dados avaliado, por exemplo, o crime de estupro conta com 15 registros no banco de dados, e como foi definido com grau de importância "quatro", usará quatro clusters para esta porção de dados, com o objetivo de ao serem definidos os quatro centroides, referente aos quatro clusters, calcular o centroide dos centroides dados pelos clusters, definição esta que foi explicitada no capítulo 2. A posição final, será a localização onde a distancia entre todas as ocorrências está mais otimizada. Ou seja, caso tivesse uma base policial militar para atender apenas aquele tipo de crime, a base ideal seria localizada na posição onde foi definido o centroide dos quatro centroides dado pelos quatro clusters.

Esta análise terá início com os casos de furto, que neste banco de dados representa 169 registros. Como furto representa grau de importância um, um cluster será usado e assim temos a seguinte distribuição dos dados.



(a) Crimes de furto

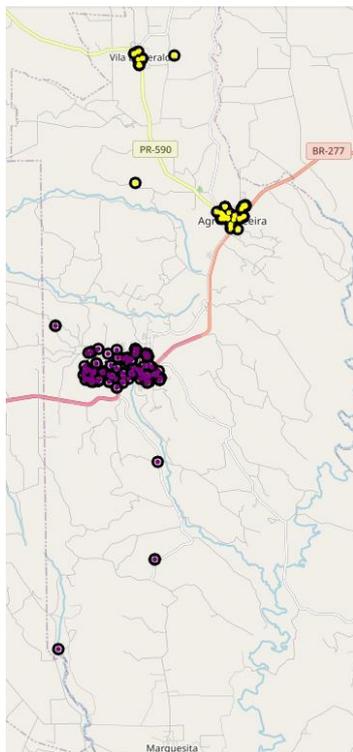
(b) Centroide dos crimes de furto

Figura 19 – Estudo feito com crimes de furto

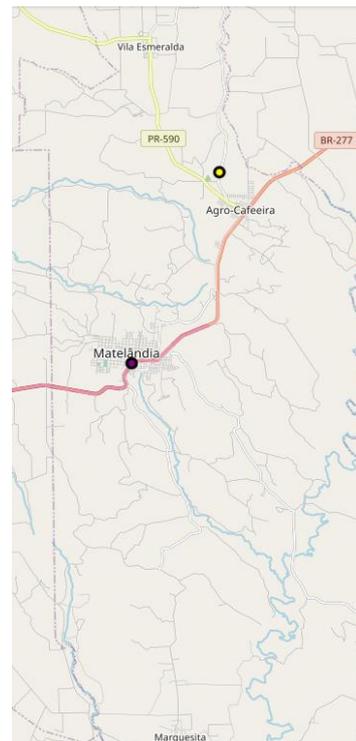
É possível analisar que a distribuição representada na figura 19a assemelha-se à distribuição obtida quando foi feita a distribuição com todos os crimes, conforme figura 17b, porém com menos registros a serem representados no mapa.

Quando tratado o crime de furto com grau de importância "um" de relevância, o centroide aparece dentro de Matelândia, como pode ser observado na figura 19b. A distância em relação a base policial militar do município de Matelândia estará exposta ao final de todos os estudos de caso propostos, em uma única tabela.

O segundo caso, será investigado os registros do crime de “Maria da Penha”, que possui grau de importância dois em nosso estudo. Após o método ser aplicado, é possível verificar na figura 20a as duas áreas referente aos dois clusters, sendo uma roxa que ficou na região de Matelândia e Zona rural ao sul da cidade e outro cluster que se concentrou na área do distrito de Agro Cafeeira e distrito de Vila Esmeralda.



(a) Crimes de “Maria da Penha”



(b) Centroides de crimes de “Maria da Penha”

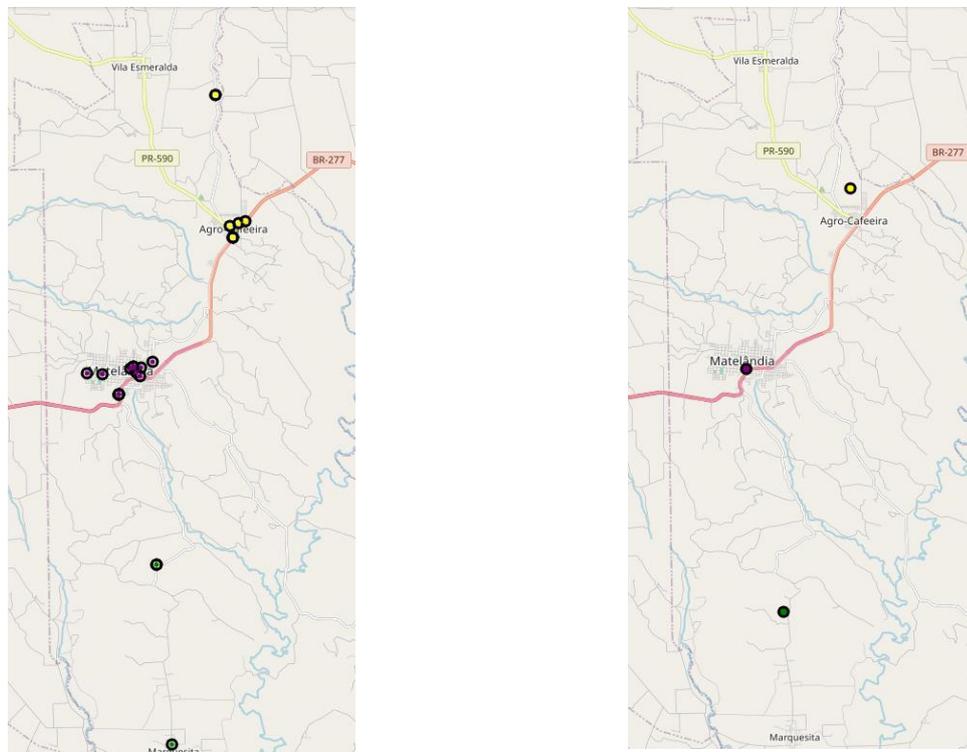
Figura 20 – Estudo feito com crimes de “Maria da Penha”

Calculando os dois centroides referentes a estas áreas temos, conforme figura 20b, foi verificado que um centroide, em roxo, está localizado no centro de Matelândia, enquanto o segundo centroide, em amarelo, ficou localizado no distrito de Agro Cafeeira, calculando

o centroide dos centroides, este será o local ideal para uma possível base policial que atenderia apenas este tipo de ocorrência. Então observa-se conforme a figura 23b.

O centroide dos centroides dos dois clusters referente as ocorrências do tipo “Maria da Penha” está representado na cor azul, conforme figura 23b e ficou localizado em uma área rural do município.

O método para a situação que envolvem roubos, e como definido anteriormente, este tipo de registro terá grau de importância três, portanto serão usados três clusters, o que resulta na seguinte distribuição mostrada na figura 21a:



(a) Crimes de roubo

(b) Centroides dos crimes de roubo

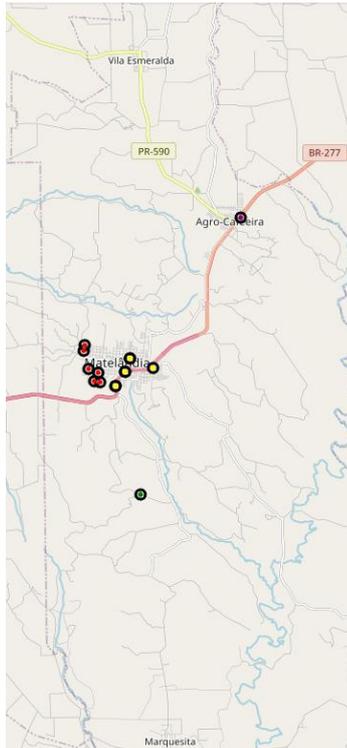
Figura 21 – Estudo feito com crimes de roubo

Nesta situação, é possível averiguar que o maior cluster se manteve na cidade de Matelândia, em roxo, o segundo maior cluster, em amarelo, teve predominância no distrito de Agro Cafeeira e o terceiro cluster ficou localizado na Zona rural ao sul da cidade, em verde, como é possível observar na figura 21b. Para finalizar, será feito centroide dos centroides encontrados, vide figura 23c.

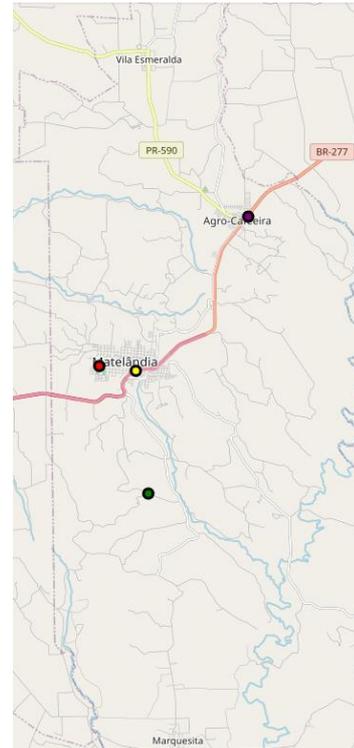
Oberseva-se que o centroide dos centroides referente aos três clusters do crime de roubo, apesar de estar um pouco afastado de Matelândia, está visivelmente mais próximo

do centro do que das demais áreas da cidade, isto pode ser analisado na figura 23c.

Para o ultimo cenário desta análise, será feito dos registros de estupro, que foi classificado como tendo grau de importância quatro, portanto será estudado com quatro clusters como mostra a figura 22a.



(a) Crimes de estupro



(b) Centroides dos crimes de estupro

Figura 22 – Estudo feito com crimes de estupro

Neste caso, pode-se observar um fato que ainda não havia ocorrido em outros crimes, dois dos quatro clusters estão localizados no centro de Matelândia, sendo eles representados pelas cores vermelho e amarelo. O terceiro cluster, em roxo, está no distrito de Agro Cafeeira e o ultimo cluster localizado na Zona rural da cidade, representado em verde, vide figura 22b.

Seguindo o jogo de cores usado nos casos anteriores, observa-se o centroide dos centroides dos quatro clusters do crime de estupro, representado na cor azul. Este cluster ficou localizado próximo à rodovia federal 277 (BR-277) que transecciona a cidade de Matelândia, conforme mostra figura 23d.

Por fim, será feito o calculo do centroide dos quatro centroides obtidos em todos os crimes estudados, representado na figura 24a.

Nesta imagem, figura 24a, os quatro centroides referente aos quatro crimes estudados neste trabalho, sendo o centroide representado na cor roxa referente ao crime de furto, centroide na cor amarela representando os crimes referentes a “Maria da Penha”, na cor verde o centroide que representa os centroides dos crimes de roubo e na cor vermelha temos o centroide que representa os centroides do crime de estupro.

É possível analisar onde essas coordenadas obtidas através do método, estão localizadas fisicamente dentro da cidade de estudo. Os centroides dos crimes de furto e crimes do tipo “Maria da Penha” por exemplo, ficaram localizados em zonas rurais entre o centro do município e o distrito de Agro Cafeeira. Isto se deve ao fato do alto volume de incidência deste tipo de delito, em ambas as localidades, o que naturalmente posiciona o centroide dos centroides entre ambos os locais.

Porém, quando analisasse os crimes de roubo e estupro, visualiza-se que o centroide dos centroides encontrado, é muito próximo a endereços onde há uma maior incidência destes crimes. O crime de roubo por exemplo, teve o centroide dos centroides em um bairro chamado “Vila Pasa”, que é bastante coerente para se construir uma base policial militar exclusiva para este tipo de crime, que por ser tratar de uma zona periférica de baixa renda, e com bastante casas em situação irregular, vulgarmente conhecidas como “invasões”, é bastante usada por criminosos para esconder-se após efetuarem roubos às margens da rodovia federal 277 (BR-277).

Em relação ao crime de estupro, o centroide dos centroides ficou posicionado em uma rua próxima ao centro de Matelândia, a Rua Professora Lerides Pagnuceli Lima, a qual dá acesso a zona rural da cidade. Por se tratar de um local com pouco movimento de automóveis, é muito usada pelos moradores locais para se fazer caminhadas em horários variados, o que causa um cenário propício para bandidos com intenção de cometer crimes de estupro, visto que é um local afastado e região de matas, com muitas rotas de fuga por estradas rurais. Sendo assim, um base policial militar com intuito de ser exclusiva para este tipo de crime, estaria localizada na coordenada coincidente ao centroide dos centroides.

Será feito uma análise de distâncias nas tabelas 6 e 7 para visualizar os resultados obtidos. Estas distâncias foram obtidas através da ferramenta Google Maps, e trata-se de distâncias rodoviárias, não em linha reta.

Verifica-se através dos resultados obtidos, que quando comparasse os centroides dos crimes estudados, que são representações da localização ideal de cada base policial militar

Tabela 6 – Distancias da Base PM dos centroides dos crimes

	Base PM Matelândia	Centroides finais dos crimes	Dist. Base PM Mat. à Centroide dos crimes
Furto	-25.2407563,-53.9867984	-25.22956188, -53.97319218	2,6 Km
“Maria da Penha”	-25.2407563,-53.9867984	-25.21304454, -53.9661402	5,7 Km
Roubo	-25.2407563,-53.9867984	-25.24999259, -53.96508139	3,0 Km
Estupro	-25.2407563,-53.9867984	-25.24042538, -53.9717263	1,6 Km

Tabela 7 – Distancias da Base PM Agro Cafeeira dos centroides dos crimes

	Base PM Agro Cafeeira	Centroides finais dos crimes	Dist. Base PM Agro à Centroide dos crimes
Furto	-25.1922386,-53.9465905	-25.22956188, -53.97319218	9 Km
“Maria da Penha”	-25.1922386,-53.9465905	-25.21304454, -53.9661402	5,8 Km
Roubo	-25.1922386,-53.9465905	-25.24999259, -53.96508139	9,5 Km
Estupro	-25.1922386,-53.9465905	-25.24042538, -53.9717263	7,5 Km

para atender cada tipo de crime, e avaliasse as distâncias que estas bases ideias estão em relação a atual base policial militar ativa no município, conforme tabela 6, observa-se que com exceção do crime relacionado a “Maria da Penha”, o qual teve pior desempenho, os outros crimes estão há uma distância máxima de 3 quilômetros.

Quando feita a mesma comparação com a base policial militar do distrito de Agro Cafeeira, tabela 7, que está desativada no momento, o desempenho torna-se inferior, sendo que com exceção dos crimes relacionados a “Maria da Penha” o qual observa-se uma diminuição na distancia, os demais crimes seriam piores assistidos caso esta base de Agro Cafeeira fosse ativa e única, e a base de Matelândia fosse desativada.

Para última análise será feito o estudo do centroide que será gerado pelos quatro centroides dos crimes aqui propostos. As cores anteriormente definidas foi respeitada, e adicionado o ponto em cor azul, que representa o centroide gerado a partir dos outros quatro centroides, conforme exposto na figura 24b.

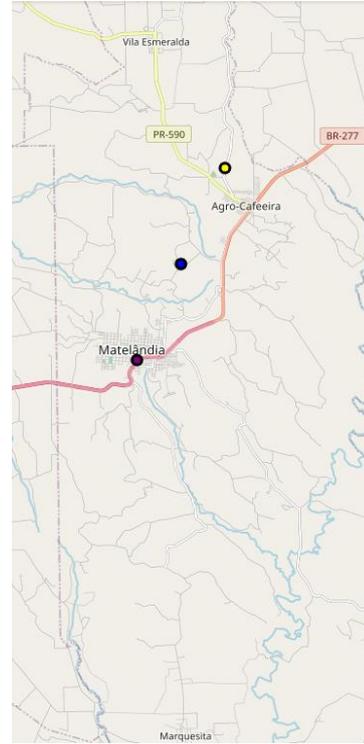
Segundo o método analisado, a base ideal estaria localizada na coordenada do ponto em cor azul, conforme tabela 8, e quando comparado esta localização com a localização da base atual de policia militar de Matelândia, a diferença de distância é de apenas 2,7 quilômetros.

Tabela 8 – Distancias da Base PM em relação ao centroide dos centroides

	Base Policia Militar Mat.	Centroide dos centroides	Distância Base PM Mat.
Todos os crimes estudados	-25.2407563,-53.9867984	-25.2332561 , -53.96903502	2,7 Km



(a) Centroide dos centroides furto



(b) Centroide dos centroides "Maria da Penha"

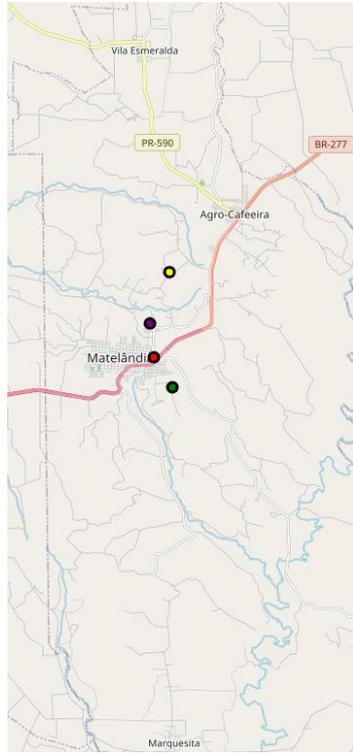


(c) Centroide dos centroides roubo

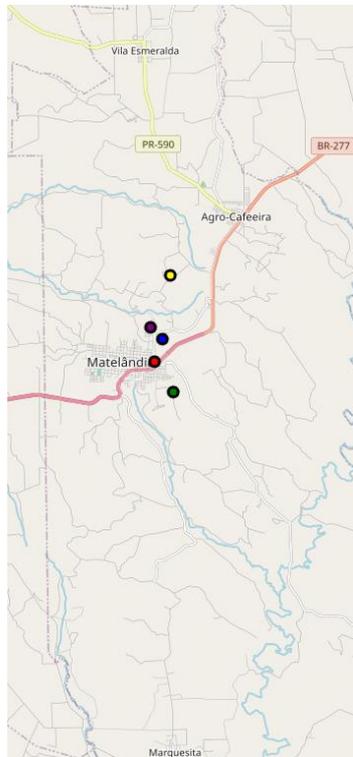


(d) Centroide dos centroides estupro

Figura 23 – Todos os centroides dos centroides



(a) Centroide dos quatro crimes estudados



(b) Centroide dos centroides de todos os crimes

Figura 24 – Centroide dos centroides

5 CONCLUSÃO

Foi possível através do método k-means, realizar um estudo sobre a localização das bases policiais militares do município de Matelândia. Conforme pudemos observar e avaliar no tópico anterior, os objetivos aqui propostos neste trabalho foram todos atingidos, sendo seguro afirmar que a atual localização da base policial militar da cidade, atende satisfatoriamente as ocorrências criminais que venham acontecer, sendo isto válido para as duas formas de se estudar os dados.

Em um primeiro momento foi proposto tratar todos os crimes de forma igualitária, sem distinção de importância a ser atendido, e conclui-se que com o centroide gerado a partir de um cluster, a distância deste centroide, que podemos interpretar como sendo o local que mais se aproxima de todos os dados ao mesmo tempo, ficou apenas a 2,5 quilômetros da posição atual da base policial militar. Quando abrangemos a mesma interpretação dos dados, porém gerando dois clusters, onde um cluster ficou localizado no centro do município de Matelândia, e o segundo ficou localizado no distrito de Agro Cafeeira, é possível afirmar que caso seja ativada a segunda base policial militar no distrito de Agro Cafeeira, esta estará localizada em um endereço estratégico, pois a distância em relação ao centroide do segundo cluster, foi de apenas 1,4 quilômetros, sendo observado que seria justificável e eficiente ter uma equipe de plantão neste distrito.

Agora, quando trabalhamos os dados de uma maneira diferente, a fim de realizar um estudo onde as ocorrências teriam graus de importância, de acordo com o potencial lesivo da natureza criminal praticada, e além disso, trabalhando apenas com o cenário atual, onde está em funcionamento apenas uma base policial militar na cidade de Matelândia, obtivemos resultados importantes.

Como descrito anteriormente, para cada natureza criminal estudada neste trabalho, foi definido com ajuda de literaturas que corroboraram estas escolhas, um grau de importância. Onde furto ficou com grau de importância 'um', fatos de "Maria da Penha" teve grau dois, roubo foi definido com grau três de importância e ao estupro atribui-se grau quatro. Desta forma foi possível analisar os centroides de cada tipo de ocorrência separadamente, e não agrupados como no primeiro caso. Esta forma de abordagem permitiu que fosse mais preciso em relação aos centroides obtidos, e ao calcularmos o centroide em relação aos quatro centroides referentes a cada natureza criminal, obtivemos a coordenada

que segundo o método aqui utilizado, seria a melhor posição para ter uma base policial militar, pois seria a representação com mais otimização de distância entre as ocorrências, levando em consideração que atenderia primeiro os crimes mais gravosos para a sociedade. E com este resultado chegou-se a conclusão que a atual base policial militar do município de Matelândia está muito bem localizada, pois a distância da base policial militar ideal segundo o método, para a atual base é de apenas 2,7 quilômetros, uma distância muito aceitável se for levar em consideração todas as limitações logísticas para se instalar uma base policial militar em uma cidade.

Além disso, foi possível observar que mesmo quando comparado cada natureza criminal de forma isolada, as distâncias aferidas continuam aceitáveis, com exceção dos crimes de "Maria da Penha", cuja a distância ficou em 5,7 quilômetros.

Este material acadêmico, pretende auxiliar de algum modo, gestores de segurança pública responsáveis pelo município de Matelândia, pois fornecem dados relevantes e fundamentados bibliograficamente, que elucidam e explicitam uma parcela da mancha criminal da cidade. Com as informações aqui expostas, é possível traçar estratégias de como enfrentar a criminalidade na região e de como aplicar o efetivo disponível, sem necessariamente depender de ajuda do governo do Estado do Paraná em disponibilizar efetivo para que possa ser alocada outras bases de polícia militar na cidade.

Uma limitação que podemos elencar neste trabalho, é a quantidade de crimes que foram escolhidos para ser feito o estudo. Existem outras naturezas criminais que são atendidas diariamente pela polícia militar em todo o estado do Paraná, incluindo a cidade de Matelândia. Porém para que o trabalho pudesse apresentar informações descritas de forma concisa, optou-se por filtrar estes dados e apresentar os que foram considerados mais importantes, sendo esta escolha embasada em literatura criminal especializada, um fator que contribuiu para esta escolha mais enxuta, foi a forma que os dados criminais foram disponibilizados, em um formato que demandou grande energia para que pudessem ser trabalhados pelos softwares escolhidos para o estudo.

O estudo aqui realizado pode ser visto como precursor de outros trabalhos a serem realizados na cidade de Matelândia, visto que até o momento, nada semelhante havia sido feito em relação a uma análise criminal do município, visando entender as características geográficas das ocorrências.

O tema é vasto e há muito a ser explorado, buscando sempre a otimização e o aperfeiçoamento do serviço prestado a população. Outras questões poderão ser exploradas, como

predição de crimes a partir dos crimes já ocorridos, um tema que poderá ser abordado em trabalhos futuros a partir deste estudo, usando os dados aqui apresentados como base para prever onde futuros crimes poderiam ocorrer. Outra proposta para trabalhos futuros seria aplicar o estudo aqui apresentado para outros tipos de ocorrências criminais. São questões relevantes e que agregariam informações pertinentes para os gestores de segurança pública daquela região.

Por fim recomenda-se uma instrução para os policiais militares que elaboram boletins de ocorrência dos crimes atendidos, salientando a importância de se atentar ao correto preenchimento de endereços, contendo nome do bairro, logradouro e numeração de onde ocorreram os fatos. Isto ajudará a aumentar a precisão de futuros estudos, como também aumentará a eficiência de como são realizados.

REFERÊNCIAS

- ALI, H. H.; KADHUM, L. E. K-means clustering algorithm applications in data mining and pattern recognition. *International Journal of Science and Research (IJSR)*, v. 4, n. 8, p. 2804–2807, 2015. ISSN 2319-7064. Index Copernicus Value: 78.96 | Impact Factor: 6.391.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. 1st ed. 2006. corr. 2nd printing. ed. [S.l.]: Springer, 2006. (Information Science and Statistics). ISBN 9780387310732, 0387310738.
- BRASIL, C. da República Federativa do. *Constituição da República Federativa do Brasil de 1988*. [S.l.]: Senado Federal, 1988. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>.
- BRASIL, C. N. do. *Código Penal Brasileiro*. Atualizado até as últimas emendas e reformas. Brasília: Jornal do Brasil, 1940. Texto consolidado com as emendas e reformas mais recentes.
- BROOKS, A. B.; PEDERSEN, K.; MAEKAWA, C.; GEE, K. L.; TRANSTRUM, M. K.; JAMES, M. M.; SALTON, A. R. K-means clustering of inputs to a geospatial model for optimizing acoustic data collection. In: *Proceedings of the 176th Meeting of the Acoustical Society of America*. [S.l.]: Acoustical Society of America, 2018. Acoustics Week in Canada.
- CAPó, M.; PÉREZ, A.; LOZANO, J. A. An efficient k-means clustering algorithm for massive data. *Journal of LaTeX Class Files*, v. 14, n. 8, 2015.
- COSTA, D. H.; NJAINE, K.; SCHENKER, M. Repercussões do homicídio em famílias das vítimas: uma revisão da literatura. *Ciência & Saúde Coletiva*, v. 22, n. 9, p. 2951–2960, set 2017.
- DATAFOLHA. *Avaliação de um ano e três meses do presidente Lula*. 2024. Datafolha, São Paulo. Disponível em: <<https://datafolha.folha.uol.com.br/opiniao-e-sociedade/2024/04/em-alta-36-se-sentem-muito-inseguros-nas-ruas-do-proprio-bairro-ao-anoitecer.shtml>>. Acesso em: 05 ago. 2024.
- ESTATÍSTICA, I. B. de Geografia e. *Censo Demográfico 2022*. 2022. Acesso em: 15 set. 2024. Disponível em: <<https://www.ibge.gov.br/>>.
- FERNANDEZ, K. M. Emprego da análise criminal na atividade de inteligência de segurança pública. *Cadernos de Segurança Pública*, v. 12, n. 12, setembro 2020. Acesso em: 15 set. 2024. Disponível em: <<http://www.isprevista.rj.gov.br/download/Rev20201204.pdf>>.
- (FIRJAN), F. das Indústrias do Estado do Rio de J. *Panorama da Competitividade dos Países do G20: Brasil 2024*. [S.l.], 2024. Disponível em: <<https://www.firjan.com.br/data/files/18/85/13/ED/5065D810FABCEEC8C8284EA8/panorama-competitividade-G20-brasil-2024.pdf>>.
- FORGY, E. W. Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics*, v. 21, n. 3, p. 768–769, 1965.

- HARTIGAN, J. A. *Clustering Algorithms*. Wiley, 1975. (Out-of-print Books on demand). ISBN 9780471356455. Disponível em: <<https://books.google.com.br/books?id=NznvAAAAMAAJ>>.
- HASTIE, T.; TIBSHIRAN, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd. ed. [S.l.]: Springer, 2009. (Springer Series in Statistics). ISBN 9780387848570, 9780387848587, 0387848576.
- (IPEA), I. de P. E. A. *Atlas da Violência 2016: Dados sobre violência no Brasil*. Brasília, 2018. Disponível em: <<http://atlasvl.ipea.gov.br/atlasviolencia/publicacoes?page=2>>.
- (IPEA), I. de P. E. A. *Atlas da Violência 2022: Dados sobre violência no Brasil*. Brasília, 2024. Disponível em: <<https://www.ipea.gov.br/atlasviolencia/publicacoes>>.
- JUNIOR, J. B. C. Ocorrências de perturbação de sossego e as implicações no policiamento ostensivo das unidades operacionais de curitiba/pr. *RECIMA21 - Revista Científica Multidisciplinar*, v. 2, n. 9, p. 1–13, 2021. ISSN 2675-6218. Disponível em: <<https://recima21.com.br/index.php/recima21/article/view/785>>.
- LLOYD, S. P. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, v. 28, n. 2, p. 129–137, 1982.
- MACQUEEN, J. B. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. [S.l.]: University of California Press, 1967. v. 1, p. 281–297. Retrieved 2009-04-07.
- MAM, M.; G., L.; SAXENA, N. S. Improved k-means clustering based distribution planning on a geographical network. *I.J. Intelligent Systems and Applications*, MECS, v. 4, n. 4, p. 69–75, April 2017. Disponível em: <<http://www.mecs-press.org/>>.
- MARTANTO; ANWAR, S.; ROHMAT, C. L.; BASYSYAR, F. M.; WIJAYA, Y. A. Clustering of internet network usage using the kmedoid method. In: *Proceedings of the Annual Conference on Computer Science and Engineering Technology (AC2SET)*. [S.l.]: Conference Organizers, 2020.
- MURPHY, K. P. *Machine Learning: A Probabilistic Perspective*. [S.l.]: The MIT Press, 2012. (Adaptive Computation and Machine Learning). ISBN: 0262018020. ISBN 9780262018029.
- OLIVEIRA, H. K. S. d.; RESENDE, G. S. L. d. Violência sexual: uma análise social da cultura do estupro. *Perspectivas Em Diálogo: Revista De Educação E Sociedade*, v. 7, n. 14, p. 81–110, 2020. Recuperado de <https://periodicos.ufms.br/index.php/persdia/article/view/9329>. Disponível em: <<https://periodicos.ufms.br/index.php/persdia/article/view/9329>>.
- PARANÁ, P. M. do. *Ouvidoria*. 2024. Curitiba: Polícia Militar do Paraná. Disponível em: <<https://www.pmpr.pr.gov.br/>>. Acesso em: 01 ago. 2024.
- PRABHU, P.; ANBAZHAGAN, N. Improving the performance of k-means clustering for high dimensional data set. *International Journal on Computer Science and Engineering (IJCSE)*, v. 3, n. 6, p. 2286–2291, June 2011.

SILVA, J. A. da. *Análise Criminal: Teoria e Prática*. Salvador: Artpoesia, 2015. 288 p. Ilustrações; 24 cm. ISBN 978-85-66783-14-8.

STEINHAUS, H. Sur la division des corps matériels en parties. *Bulletin de l'Académie Polonaise des Sciences*, v. 4, n. 12, p. 801–804, 1957.

WILSON, J. Q.; KELLING, G. L. Broken windows: The police and neighborhood safety. *The Atlantic Monthly*, v. 249, n. 3, p. 29–38, 1982.