



UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
PRÓ-REITORIA DE GRADUAÇÃO
CENTRO MULTIDISCIPLINAR DE ANGICOS
CURSO ENGENHARIA DE PRODUÇÃO

THALIA DE MOURA DE QUEIROZ

**INTELIGÊNCIA ARTIFICIAL PARA GESTÃO DA SEGURANÇA PÚBLICA NO
ESTADO DO RIO GRANDE DO NORTE – RN**

ANGICOS / RN
2023

THALIA DE MOURA DE QUEIROZ

**INTELIGÊNCIA ARTIFICIAL PARA GESTÃO DA SEGURANÇA PÚBLICA NO
ESTADO DO RIO GRANDE DO NORTE – RN**

Trabalho Final de Graduação apresentado à
Universidade Federal Rural do Semi-Árido
como requisito para obtenção do título de
Bacharel em Engenharia de Produção.

Orientador (a): Prof. Dr. Ciro José Jardim de
Figueiredo

© Todos os direitos estão reservados a Universidade Federal Rural do Semi-Árido. O conteúdo desta obra é de inteira responsabilidade do (a) autor (a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei nº 9.279/1996 e Direitos Autorais: Lei nº 9.610/1998. O conteúdo desta obra tomar-se-á de domínio público após a data de defesa e homologação da sua respectiva ata. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu (a) respectivo (a) autor (a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

Q3i Queiroz, Thalia.
 Inteligência artificial para gestão da
segurança pública no estado do Rio Grande do Norte
- RN / Thalia Queiroz. - 2023.
64 f. : il.

 Orientador: Ciro José Jardim de Figueiredo.
Monografia (graduação) - Universidade Federal
Rural do Semi-árido, Curso de Engenharia de
Produção, 2023.

1. Inteligência artificial. 2. Segurança
pública. 3. Análise de dados. 4. Padrões
criminais. I. José Jardim de Figueiredo, Ciro,
orient. II. Título.

Ficha catalográfica elaborada por sistema gerador automático em conformidade
com AACR2 e os dados fornecidos pelo(a) autor(a).

Biblioteca Campus Angicos
Bibliotecário: Helder Romero Maia Duarte
CRB: 15/673

O serviço de Geração Automática de Ficha Catalográfica para Trabalhos de Conclusão de Curso (TCC's) foi desenvolvido pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (USP) e gentilmente cedido para o Sistema de Bibliotecas da Universidade Federal Rural do Semi-Árido (SISBI-UFERSA), sendo customizado pela Superintendência de Tecnologia da Informação e Comunicação (SUTIC) sob orientação dos bibliotecários da instituição para ser adaptado às necessidades dos alunos dos Cursos de Graduação e Programas de Pós-Graduação da Universidade.

THALIA DE MOURA DE QUEIROZ

**INTELIGÊNCIA ARTIFICIAL PARA GESTÃO DA SEGURANÇA PÚBLICA NO
ESTADO DO RIO GRANDE DO NORTE – RN**

Trabalho Final de Graduação apresentado à
Universidade Federal Rural do Semi- Árido
como requisito para obtenção do título de
Bacharel em Engenharia de Produção.

Defendida em: 17/05/2023.

BANCA EXAMINADORA



Documento assinado digitalmente
CIRIO JOSE JARDIM DE FIGUEIREDO
Data: 19/05/2023 15:24:26-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Ciro José Jardim de Figueiredo
(UFERSA) Presidente



Documento assinado digitalmente
SAMIRA YUSEF ARAUJO DE FALANI BEZERRA
Data: 20/05/2023 19:12:27-0300
Verifique em <https://validar.iti.gov.br>

Prof.^a. Dra. Samira Yusef Araújo de Falani Bezerra (UFERSA)
Membro Examinador



Documento assinado digitalmente
LETICIA SOARES TEIXEIRA DE SOUZA
Data: 19/05/2023 15:45:34-0300
Verifique em <https://validar.iti.gov.br>

Prof.^a. Letícia Soares Teixeira de Souza (UFERSA)
Membro Examinador

AGRADECIMENTOS

Ao escrever estes agradecimentos, me pego refletindo sobre a trajetória que me trouxe até aqui e percebo que não seria possível sem o apoio e amor incondicional de minha família. Meus pais, Adriana e Miguel, minha irmã, Monique, e familiares que sempre estiveram presentes em todos os momentos, me apoiando e incentivando a seguir em frente. Não há palavras suficientes para expressar minha gratidão.

E como falar de agradecimentos sem mencionar os amigos? Vocês foram meus pilares de sustentação em momentos de dúvidas e incertezas, sempre me encorajando e acreditando em mim. São as lembranças dos momentos divertidos e desafiadores que levarei para sempre em meu coração.

Ao meu orientador, que dedicou seu tempo e esforço para me orientar durante todo o processo de elaboração da monografia. Sua orientação, conselhos e críticas construtivas foram fundamentais para o sucesso deste trabalho. Obrigada por me ensinar a pensar de maneira crítica, por me desafiar e me ajudar a encontrar minha própria voz acadêmica.

Não poderia deixar de agradecer à minha instituição de ensino, a UFERSA, por ter me proporcionado todas as condições para que eu pudesse desenvolver este trabalho, e a todo corpo docente, por todo apoio e conhecimento que me proporcionaram durante minha jornada acadêmica, obrigada por me encorajarem a explorar novas áreas de conhecimento, e por terem dedicado tanto tempo e energia para me ajudarem a crescer e alcançar meus objetivos. Vocês foram uma parte fundamental na minha jornada educacional, levarei para sempre seus ensinamentos e valores.

Em resumo, eu sou grata por todas as pessoas que fizeram parte da minha jornada até este momento, e que continuam me apoiando em todos os aspectos da minha vida. Obrigada por terem sido minha base sólida em tempos incertos. Este trabalho é dedicado a vocês.

RESUMO

O presente estudo tem como objetivo principal avaliar o uso da inteligência artificial na gestão da segurança pública no estado do Rio Grande do Norte (RN). Para alcançar esse objetivo, foram definidos objetivos específicos, que incluem avaliar o crescimento quantitativo da violência no estado e identificar e analisar padrões criminais por meio de ferramentas de análise de dados. Em seguida, pretende-se identificar e analisar padrões criminais por meio de ferramentas avançadas de análise de dados, a fim de oferecer insights acionáveis para a tomada de decisões no âmbito das políticas de segurança pública. Ressaltando-se a importância da análise de dados para o combate à violência e propondo ações para contornar a situação no estado do RN. A pesquisa foi conduzida por meio da utilização de uma base de dados, bem como estudos revisados, revelando uma alta correlação entre variáveis socioeconômicas e criminalidade.

Acredita-se que os resultados deste estudo fornecerão *insights* valiosos para gestores públicos e tomadores de decisão no campo da segurança, contribuindo para o aprimoramento das políticas e práticas de segurança pública no estado do RN.

Palavras-Chave: Inteligência artificial; segurança pública; análise de dados; padrões criminais.

LISTA DE FIGURAS

Figura 1 - Ocorrências criminais anuais no estado do Rio Grande do Norte (2015 – 2021)	28
Figura 2 - Variação mensal de CVLI's no Rio Grande Do Norte (2015-2021)	29
Figura 3 - Ranking dos 10 municípios mais violentos no Rio Grande do Norte (2015 - 2021).....	30
Figura 4 - Taxa de ocorrências criminais a cada 100 mil habitantes.....	31
Figura 5 - Mapa de calor referente ao coeficiente de correlação	35
Figura 6 - Mapa de calor filtrado	35
Figura 7 - Método do Cotovelo.....	37
Figura 8 - Método da Silhueta.....	37
Figura 9 - População estimada versus Ano	40
Figura 10 - Pessoa ocupada versus Ano.....	42
Figura 11 - Número absoluto de escolas que oferecem educação básica no município versus Ano.....	44
Figura 12 - Número absoluto de escolas que oferecem ensino médio no município versus Ano	46
Figura 13 - Pessoa com acesso ao Sistema Único de Saúde versus Ano	49

LISTA DE TABELAS

Tabela 1 - Estatísticas da taxa de CVLI's nos anos de 2015 - 2021 no estado do RN.....	28
Tabela 2 - Tipo de morte registradas e número de ocorrências.....	30
Tabela 3 - Definição dos crimes	30
Tabela 3 - Variáveis e descrição presentes no banco de dados	32

SUMÁRIO

1. INTRODUÇÃO	10
1.1 Contextualização do Tema	10
1.2 Objetivos.....	11
1.2.1 Objetivo geral	12
1.2.2 Objetivos específicos.....	12
1.3 Problema.....	12
1.4 Justificativa.....	14
1.5 Estrutura do trabalho	15
2. FUNDAMENTAÇÃO TEÓRICA.....	17
2.1 Teorias sobre crime	17
2.2 Segurança pública e modelos quantitativos.....	19
2.3 Clusterização e K-means	21
2.4 Análise de dados.....	22
3. METODOLOGIA.....	24
3.1 Classificação da pesquisa	24
3.2 Etapas da pesquisa.....	24
4. RESULTADOS	27
4.1 Análise Descritiva	27
4.2 Análise Exploratória.....	32
5. DISCUSSÃO	51
6. CONSIDERAÇÕES FINAIS.....	53
REFERÊNCIAS	55
APÊNDICE A – Mapa de calor com coeficientes de correlação	62
APÊNDICE B – Código fonte utilizado.....	63

1. INTRODUÇÃO

1.1 Contextualização do Tema

A criminalidade no cotidiano vem sendo um problema persistente para a sociedade, afligindo vários países do mundo, inclusive o Brasil (MIRABETE, 2010).

Segundo a Lei Nº 3.914 (Brasil,1941), constitui infração penal, condutas em que a lei julga como sendo ilícitas, são ações proibidas que, se executadas, a lei prevê pena privativa de liberdade ou prisão, isolada ou subsidiariamente ou em conjunto com multa; contravenção, crime para o qual a lei impõe, designadamente, pena de prisão simples ou multa, ou ambas, alternativamente ou cumulativamente.

É notório que a definição dada pelo referido Decreto - Lei não é suficiente, afinal, apenas realiza a diferenciação dos crimes das contravenções penais. Assim, cabe aos estudiosos e cultores do direito definir o que é crime. Dentre os conceitos que são apresentados pela doutrina, cabe ressaltar os três mais importantes, a saber: conceito material, formal e analítico.

Segundo Capez (2010), o conceito material define crime como, todo e qualquer fato humano que, intencionalmente ou por negligência, danifica ou põe em perigo a propriedade dos princípios jurídicos, considerados essenciais para a existência do coletivo e paz social. Já no aspecto formal, o conceito de crime surge da mera inclusão de ações do tipo jurídico e tudo o que o legislador julgar como tal, independentemente do seu conteúdo. Ainda segundo o autor, avaliar a existência de uma infração penal, sem levar em consideração o seu conteúdo ou dano patrimonial é uma violação do princípio constitucional da dignidade humana.

De acordo com Capez (2003), o aspecto analítico, que é o conceito utilizado no Brasil, busca sob uma perspectiva jurídica estabelecer os elementos estruturais do crime, isto é, proporcionar a exata e mais justa decisão sobre a infração e seu autor, desta forma fazendo com que o julgador ou intérprete progrida com seu raciocínio em etapas.

BRASIL. Lei Nº 7.209 (Brasil,1984), o Código Penal Brasileiro lista, em seu componente Parte Especial, os seguintes crimes: crimes contra a pessoa, crimes contra o patrimônio, crimes contra a propriedade imaterial, crimes contra a organização do trabalho, crimes contra o sentimento religioso e contra o respeito aos mortos, crimes contra a família, crimes contra a incolumidade pública, crimes contra a dignidade sexual, crimes contra a fé

pública, crimes contra a administração pública e crimes contra a paz pública, os quais são subdivididos em títulos.

Segundo Neto (1991), como efeitos de uma sociedade industrial, o crescimento desgovernado da população, bem como sua desordenada concentração urbana, tornou possível a marginalização da comunidade no Brasil.

Com isso, torna-se evidente a importância da segurança pública, para que se possa garantir a proteção aos direitos individuais de cada ser, para que assim seja possível efetivar seu direito à cidadania e segurança.

Segundo o Artigo 144 da Constituição Federal, a segurança pública é um dever do estado, sendo um direito e responsabilidade de todo ser, é efetuada para a manutenção da ordem pública e segurança das pessoas e do patrimônio, sendo executada através dos seguintes órgãos: polícia federal, polícia rodoviária federal, polícia ferroviária federal, policiais militares e corpos de bombeiros militares, polícias penais federal, estadual e distrital. (BRASIL, 1988)

Nesse sentido, é perceptível que a segurança pública, além de ser um dever do estado, conforme é descrito no Art. 144 da Constituição, é também uma responsabilidade de todos enquanto cidadãos.

Nesse contexto, a inteligência artificial (IA), tem se mostrado uma forte aliada na área da pesquisa e tomadas de decisão. Segundo Russel e Norvig (2016), a IA abrange uma variedade de técnicas que possibilitam que os sistemas automatizados tomem decisões, resolvam problemas complexos e realizem tarefas que antes eram apenas dos seres humanos.

Autores como Bursteinz et al. (2019), chamam atenção para o fato de que a IA pode contribuir para a análise preditiva de padrões criminais, auxiliando na identificação de áreas de risco e na otimização do policiamento preventivo.

Com isso, a aplicação da IA no âmbito da segurança pública tem despertado grande interesse, pois oferece a possibilidade de melhorar a eficiência e a eficácia na gestão e prevenção de crimes, além de ser uma forte aliada na análise de grandes volumes de dados relacionados à segurança pública, como registros de ocorrências, informações demográficas e socioeconômicas.

1.2 Objetivos

O presente trabalho será desenvolvido no intuito de alcançar os seguintes objetivos:

1.2.1 Objetivo geral

Construir um painel de *analytics data* para os dados sobre violência no estado do Rio Grande do Norte – RN, para auxiliar na construção de ações de segurança pública.

1.2.2 Objetivos específicos

- Avaliar o crescimento quantitativo da violência no estado;
- Avaliar as formas de violência praticadas, e relacionar com os fatores sociais;
- Identificar e analisar padrões criminais através de ferramentas de análise de dados;
- Ressaltar a importância da análise de dados acerca do tema, para o combate à violência;
- Propor ações para que se possa contornar a violência no estado do RN.

1.3 Problema

De acordo com Singh e Josi (2018), a medida em que a população aumenta, novos tipos de crimes são efetivados, dessa forma as cidades se mostram cada vez mais expostas e desprotegidas, o que implica de maneira direta na segurança da população. Consequentemente, medidas de prevenção são extremamente importantes para auxiliar na redução de novos crimes.

De acordo com dados do Instituto de Pesquisa Econômica Aplicada (Ipea), a taxa de homicídios por Unidade Federal (UF), com base em óbitos por local de residência da vítima em 2019, indica que os números variam entre 10 e 42,7 homicídios por 100 mil habitantes.

A criminalidade afeta todas as sociedades, em especial as da América Latina, que se mostra como uma região violenta. Segundo Asmann e O’reilly (2020), o ano de 2019 foi apontado como o ano em que houve um aumento sistemático de mortes violentas na América Latina, com destaque para homicídio, que reativou novas discussões sobre os problemas motivados pela violência.

No que se trata de crimes contra o patrimônio, no ano de 2018, foram registrados 490.956 roubos e furtos de veículos no Brasil, o que corresponde a uma taxa de 33,26% de todos os crimes patrimoniais registrados no período. Isso corresponde a mais de mil e trezentos furtos diários, ou seja, mais de cinquenta e sete veículos roubados e furtados por hora. Deste modo, são mudanças de números que se firmam em graus muito elevados há algum tempo (LIMA; BUENO, 2019).

De acordo com Farias et al. (2018), no Brasil, o governo surge investindo no quesito segurança pública, bem como nos departamentos de inteligência policial, para que assim se possa lidar com a crescente ocorrência criminal. Esses investimentos são uma tendência mundial, tendo em vista o impacto que o fenômeno ‘criminalidade’ acarreta a economia.

Segundo Cerqueira et al. (2019), em 2018 os gastos com segurança pública no Brasil totalizaram R\$91,2 bilhões, o que correspondeu a 1,34% do Produto Interno Bruto (PIB). Se comparado ao ano de 2017, houve um aumento real de 3,9% nas despesas empenhadas, uma vez que o crescimento aconteceu de maneira diferente entre os entes federativos.

Previsões assertivas de crimes em tempo real, auxiliam na redução a taxa de criminalidade, porém, é um problema bastante desafiador para o corpo social científico, pois os episódios criminais dependem de vários fatores complexos (TOPPIREDDY; SAINI; MAHAJAN,2018).

A aplicação da inteligência artificial nas previsões de crimes tem se mostrado promissora na identificação de padrões e tendências que possam auxiliar na prevenção e combate à criminalidade. Vários estudos destacam os resultados positivos obtidos por meio do uso da IA nesse contexto.

Segundo Bursztein et al. (2019), pesquisas demonstram que algoritmos de aprendizado de máquina podem ser eficazes na previsão de áreas de maior risco para crimes. Essas previsões baseadas em IA permitem uma alocação mais estratégica dos recursos de segurança pública, direcionando o policiamento preventivo para os locais mais propensos a ocorrências criminais.

Além disso, a IA tem sido aplicada com sucesso na análise preditiva de padrões criminais. Mohler et al. (2018), realizaram estudos controlados em campo que demonstraram a capacidade dos modelos de IA em identificar tendências temporais e espaciais nos crimes, permitindo uma melhor compreensão dos fatores que contribuem para a criminalidade. Esses estudos revelaram a utilidade da IA na identificação de áreas e períodos de maiores atividades criminosas, o que pode subsidiar a implementação de estratégias de prevenção mais eficazes.

Outro aspecto relevante é a capacidade da IA em identificar correlações não óbvias entre variáveis socioeconômicas e criminalidade. Estudos como o de Mohler et al. (2018), destacam-se que os algoritmos de IA podem revelar relações complexas entre diferentes fatores, como desemprego, densidade populacional e incidência de crimes, que podem ser utilizadas para uma análise mais precisa e estratégica.

Diante do exposto, o presente trabalho tem em seu escopo a aplicação da análise de dados, com o intuito de avaliar as formas de violência praticadas a fim de relacionar com fatores sociais, econômicos e geográficos.

1.4 Justificativa

Segundo dados do Instituto de Geografia e Estatística (2010), o Brasil pode ser denominado como um país urbanizado, em média 84% da população, do total de 214,3 milhões de pessoas, vivem em cidades, com uma taxa de crescimento populacional de 1,17% ao ano. Em um período de cinquenta anos, a população urbana no Brasil aumentou cinco vezes. Com um crescimento de maneira rápida e desordenada, a população urbana cresceu velozmente, em contrapartida, a população rural diminuiu substancialmente. Acredita-se que as informações sejam possivelmente corroboradas com a saída do novo censo de 2022.

Segundo Menegat e Almeida (2004), o desenvolvimento urbano brasileiro deu-se de maneira desequilibrada em inúmeros aspectos. Conforme a distribuição territorial se dá de maneira bastante heterogênea, com maior concentração da população residindo em grandes cidades, a desigualdade social nas cidades também se elevou, evidencia-se que os setores populacionais mais pobres, foram os que mais se multiplicaram em relação aos ricos.

De acordo com Rocha e Silva (2005), o espaço urbano se reflete na maneira como a sociedade se organiza. No Brasil, as áreas urbanas são tidas por grandes desigualdades sociais, o que acarreta desequilíbrios socioeconômicos, que por sua vez fragmentam o espaço urbano em incontáveis territórios com características próprias e excludentes, que influenciam na criminalidade e no declínio da cidadania.

Observa-se que com esse modelo de urbanização crescente no território brasileiro, os grandes centros são também áreas de maior desigualdade social, podendo realizar um comparativo entre o processo de crescimento e os índices de criminalidade observados.

De acordo com Cerqueira et al. (2021), o Sistema de Informação sobre Mortalidade do Ministério da Saúde (SIM/MS), em 2019 houve 45.503 homicídios no Brasil, o que condiz a uma taxa de 21,7 mortes por 100 mil habitantes. Os dados oficiais do SIM/MS mostram que, em 2017, ocorreram cerca de 65.602 homicídios no Brasil, o que gira em torno de 31,6 mortes para cada cem mil habitantes, tratando-se do mais alto nível histórico de letalidade violenta intencional no país.

Segundo Waiselfiz (2018), o Brasil possui por algumas das cidades mais violentas do mundo, especialmente alguns dos centros urbanos no Nordeste, que constantemente surgem nos rankings internacionais de violência.

Com o avanço crescente e constante de tecnologias, o Estado não pode se abster de implementá-la no intuito de reduzir custos a longo e a curto prazo, tendo como resultado uma maior eficiência dos órgãos responsáveis pela realização das atividades de segurança pública.

Segundo Máximo (2004), no mapeamento de atos criminosos a associação dos dados no âmbito geográfico possibilita a oportunidade de exploração e análise dos dados que não se encontram quando faltam dados geográficos.

Dessa forma, fatores que apresentam relação com aspectos urbanísticos e ambientais contribuem para a concentração de crimes em uma determinada área, corroborando para a necessidade da implementação de maneiras que auxiliem na superação desse fenômeno, possibilitando a realização de diagnósticos mais precisos, por meio da percepção de padrões a respeito dos dados de incidentes criminais, para que possam intervir no planejamento estratégico governamental, bem como o método de tomada de decisão, a fim de combater e atenuar a criminalidade.

Existem na literatura diversos trabalhos que utilizam a inteligência artificial na área da segurança pública. Essas pesquisas têm mostrado resultados promissores e comprovados em diferentes locais. Um dos principais benefícios do uso da IA na segurança pública é a capacidade de previsões de crimes, identificando padrões criminais, tendências temporais e espaciais, além de correlações não óbvias entre as variáveis socioeconômicas e criminalidade.

Corroboram com as informações apresentadas, os autores Endrass e Hayes (2017), onde destacam a importância da aplicação da IA na segurança pública, por meio do uso de técnicas de previsão, as quais tem se mostrado promissora na identificação de áreas de risco e na alocação eficiente de recursos policiais.

1.5 Estrutura do trabalho

Este trabalho se estrutura em seis seções principais. A primeira seção é a introdução, onde são apresentados o contexto do estudo, o problema de pesquisa e os objetivos a serem alcançados. A segunda seção é a fundamentação teórica, na qual são explorados os conceitos e teorias relevantes relacionados ao tema da segurança pública e à aplicação da inteligência artificial nesse contexto. A terceira seção é dedicada à metodologia, descrevendo os

procedimentos e técnicas utilizadas para a coleta e análise dos dados. Na quarta seção, são apresentados os resultados obtidos a partir da aplicação da inteligência artificial na gestão da segurança pública, destacando as principais descobertas e achados do estudo. A quinta seção é a discussão, na qual os resultados são interpretados e analisados em relação aos objetivos propostos, relacionando-os com a literatura existente e oferecendo *insights* adicionais. Por fim, a sexta seção é reservada às considerações finais, onde são apresentadas as conclusões do estudo, suas limitações e sugestões para pesquisas futuras.

2. FUNDAMENTAÇÃO TEÓRICA

Nesta seção apresenta-se a Fundamentação Teórica que foi utilizada como suporte para o desenvolvimento do trabalho, serão abordados temas essenciais para o estudo da gestão da segurança pública utilizando inteligência artificial. Serão apresentadas teorias sobre crime, explorando os principais conceitos e abordagens teóricas relacionadas à compreensão e análise da criminalidade. Além disso, serão discutidos aspectos da segurança pública e modelos quantitativos aplicados nesse contexto, destacando como essas abordagens podem contribuir para a tomada de decisões efetivas. Serão apresentados também conceitos e aplicações da clusterização e do algoritmo K-means, bem como a importância da análise de dados na obtenção de *insights* relevantes para a gestão da segurança pública.

2.1 Teorias sobre crime

De acordo com Brantingham e Brantingham (1991), a criminologia ambiental defende que os eventos criminosos são resultado da interação entre infratores, vítimas ou alvos de crimes e princípios legais, em cenários específicos, sucedendo em um tempo e local específico. O foco da criminologia ambiental se deu ao estudo do crime e suas dimensões, e não ao “desenho” de um criminoso, ou as suas motivações inerentes para cometer crimes.

Segundo Molina e Gomes (1997), anteriormente os profissionais compreendiam delito como um fenômeno causal, repentino, já os novos pesquisadores da criminologia ambiental respaldam o crime como um fenômeno seletivo, onde o infrator buscava o lugar e momento apropriado e uma vítima conveniente.

Como consequência dessa discussão, surgiram algumas teorias acerca da criminologia ambiental, dentre elas, Teoria das Atividades Rotineiras, a Teoria da Escolha Racional, Teoria do Padrão Criminal, e a Teoria da Oportunidade. (FORTES; TASCA, 2014)

A Teoria da Atividade de Rotina teve seus primeiros estudos desenvolvidos pelos pesquisadores norte-americanos Lawrence Cohen e Marcus Felson, a partir de pesquisas a respeito do crescimento da criminalidade nos Estados Unidos, especialmente nas décadas de sessenta e setenta. (VITO; MAAHS; HOLMES, 2007)

De acordo com os autores Cohen e Felson (1979), grande parte dos atos criminosos convergem no espaço-temporal de um provável agressor, um alvo adequado, bem como a ausência de um guardião que seja capaz de impedir o crime. Dessa forma, o crime está

vinculado à ocorrência destes três elementos, os quais devem estar presentes ao mesmo tempo, e em um mesmo local. Do contrário, não se pode afirmar que existe a possibilidade de o crime acontecer.

Percebe-se em poucas linhas que, mesmo sendo uma teoria aparentemente simples, ela se mostra de grande utilidade. Com poucos conceitos a teoria pode explicar fenômenos demasiadamente complexos, envolvendo tanto a ocorrência criminal, quanto a ausência de crimes.

Segundo Brantingham e Brantingham (1993), o possível agressor é uma pessoa propensa a cometer o crime, ou seja, aquele que está inclinado a praticar determinada conduta criminosa. Já de acordo com Clarke e Eck (2021), no que se refere ao alvo adequado, este pode ser tanto uma pessoa, quanto um objeto, afinal, ambos podem servir a propósito do infrator. Além de definir como guardião capacitado, atores humanos ou dispositivos de segurança

De acordo com Cohen e Felson (1979), as razões para o crescimento de incidentes criminais não teriam ligações com questões demográficas, ou nos índices de pobreza e desemprego, mas sim no aumento das oportunidades geradas com a mudança na rotina das pessoas.

Segundo Cronish e Clarck (2008), a Teoria da Escolha Racional, em conjunto com a Teoria das Atividades Rotineiras, apresentam influência na prevenção do crime e na criminologia, mostrando que essas teorias se complementam.

De acordo com Rolim (2006), os criminosos tomam suas decisões através de um processo racional, ou seja, a teoria pressupõe que os indivíduos irão levar em consideração os custos ou as consequências do crime como benefício a ser obtido, antes de realizar a prática do ato criminoso.

Segundo Viapiana (2006), durante a escolha para executar determinado delito, são levados em consideração benefícios líquidos esperados, bem como os custos implícitos para a ação, como por exemplo, a probabilidade de prisão e condenação. Corroborando com a ideia defendida pelos autores, Felson e Clarck (1998), determinam que, a mera ganância do criminoso não é o suficiente, dessa forma, ele age com a racionalidade, baseada em fatores como, percepção de risco e de lucro.

De acordo com Brantingham e Brantingham (2008), a teoria do padrão criminal sugere que todos os indivíduos seguem rotinas, ou executam atividades diárias repetitivas. Inclui-se nesse meio os criminosos, que em durante sua rotina diária estão envolvidos em atividades não

criminais. Dessa maneira, os locais de crime estão relacionados a locais que o criminoso convive.

Dessa forma, percebe-se que os indivíduos se movem durante a realização de determinadas atividades, criando assim, um modelo de deslocamento em um espaço geográfico, que será observado por indivíduos que praticam atos criminosos, e procuram um alvo ou vítima de acordo com o seu objetivo.

Segundo Boba (2005), eventos criminosos ocorrem comumente em áreas onde o espaço de atividades dos criminosos coincide com o espaço de atividades dos alvos ou vítimas. Dessa forma, ao analisar a quantidade de alvos ou vítimas, bem como a ocorrência de crimes, é possível construir um padrão criminal, que exhibe um agrupamento geográfico com a distribuição temporal e espacial das vítimas, alvos e ofensores.

De acordo com Lopes, Santos e Hoffmann (2012), a teoria da oportunidade é tida como uma das fundamentais causas do crime, ou seja, a conduta criminosa é resultado da interação entre o indivíduo com o ambiente, e pelas oportunidades que são criadas.

Diante do exposto, entende-se que a oportunidade ideal se torna fundamental para a realização de um delito criminoso, já que ela desempenha um papel causal em todos os crimes. Para isso, focar na redução das oportunidades, pode reduzir a prática de crimes.

2.2 Segurança pública e modelos quantitativos

Existem na literatura estudos que consideram a utilização de modelos quantitativos para a análise da criminalidade.

Durante o estudo realizado por Wang et al. (2013), os autores utilizaram uma estrutura de mineração de dados espaciais, a fim de estudar pontos críticos por meio de variáveis relacionadas, para então analisar a relação entre variáveis socioeconômicas e criminais. Foram utilizados Padrões Discriminativos Geoespaciais, para o desenvolvimento de um novo modelo para a melhor identificação de *hotspots* de crimes, o *Hotspot Optimization Tool* (HOT). O intuito principal por trás dos métodos utilizados foi conseguir identificar os pontos críticos, e apresentar padrões geográficos, os resultados obtidos comprovam que o modelo HOT superou o modelo *Hotspot Analysis*, ferramenta implementada pela Esri ArcGIS, identificando com precisão pontos críticos de crime.

Figueiredo e Mota (2016), utilizaram as ferramentas do Sistema de Informações Geográficas (SIG), em conjunto com a Tomada de Decisão Multicritério (MCDM) para mapear

os locais mais propensos para a ocorrência de roubos em uma cidade brasileira, avaliando seu nível de segurança. A partir do estudo, os autores conseguiram determinar a viabilidade de adaptar o sistema SIG-MCDM para o auxílio à tomada de decisão e o planejamento para solucionar os problemas de segurança pública, tendo como base a implementação adequada dos resultados obtidos.

Nepomuceno e Costa (2019), chama atenção para o fato de que, a maioria dos estudos recentes utilizam os dados de maneira agrupada, ou seja, sem uma diferenciação cuidadosa de qual tipo de crime os resultados correspondem, com isso, é proposta a utilização de medidas de autocorrelação para visualizar pontos frios e quentes do crime, para possibilitar ações rápidas onde o crime está concentrado, além de delinear padrões que podem ser correlacionados com fatores sociais, ambientais, econômicos, auxiliando na tomada de decisões eficientes, bem como na alocação de recursos de segurança pública.

Para Figueiredo e Mota (2019), é relevante compreender a relação entre violência e as diversas variáveis ambientais, dentre elas, densidade demográfica, alfabetização, investimento econômico e número de domicílios com iluminação pública, para que assim, seja possível entender a relevância que essas variáveis exercem para a segurança pública, como alocação de recursos, e avisos sobre a criminalidade. Diante disto, os autores utilizam de preferências de Sistemas de Informação Geográficas e o conhecimento em tomadores de decisão para apresentar um modelo de identificação de áreas públicas que apresentam fragilidade quanto a violência.

De acordo com Mota, Figueiredo e Pereira (2020), as questões de segurança pública, bem como violência, são temas frequentemente discutidos, tendo em vista que afetam diretamente todos enquanto sociedade. No que se refere aos casos de homicídio, são considerados graves, pois envolvem o óbito de um indivíduo da comunidade. Se tratando de um problema presente em todo o mundo, com mais evidência nos países em desenvolvimento.

De acordo com Nepomuceno, Daraio e Costa (2021), para que a sociedade se desenvolva de maneira sustentável é imprescindível que os recursos sejam utilizados de maneira otimizada, para que assim se forneça bens e serviços, bem como os objetivos sociais desejados. Com isso, os autores destacam a relevância de um julgamento justo e sem compensação a respeito das unidades policiais, com a utilização de critérios de decisão fundamentados em eficiência e eficácia.

Borba et al. (2022), a criação de novas instalações policiais é um projeto cauteloso, tendo em vista os custos envolvidos durante o processo, com isso, é preciso realizar uma análise

minuciosa se as instalações são de fato viáveis, de acordo com seus objetivos operacionais e estratégicos. Dessa forma, os autores propõem a utilização de Sistemas de Informações Gerenciais para auxiliar nas decisões a respeito de melhorias que podem ser efetuadas na segurança pública, no intuito de propor locais para a criação de instalações policiais que proporcionem uma maior cobertura, reduzindo a distância média percorrida, e como consequência será possível uma melhor prestação de serviço à sociedade.

2.3 Clusterização e K-means

Um cluster consiste em um grupo de objetos que se assemelhem e estejam mais próximos uns dos outros, se diferenciando assim dos demais (JAIN; DUBES, 1988).

Dessa forma, um cluster no contexto da análise de dados, é um grupo ou subgrupo de elementos que compartilham características semelhantes entre si e se distinguem de outros elementos fora desse grupo. Os clusters são formados com base na similaridade ou proximidade dos dados, de modo que os elementos dentro de um mesmo cluster sejam mais semelhantes entre si do que com os elementos de outro clusters.

Segundo Hastie, Tibshirani e Friedman (2009), o agrupamento é um método de aprendizado não supervisionado, utilizado para agrupar pontos de dados ou objetos em subconjuntos, ou clusters, com base em sua similaridade. O clustering pode ser utilizado para várias finalidades, dentre elas, a exploração de dados.

Dentre as técnicas de clusterização amplamente utilizadas na análise de dados, destacam-se: a clusterização hierárquica (Jain & Dubes, 1988), o agrupamento baseado em densidade (Ester et al., 1996), a clusterização baseada em modelos (Fraley & Raftery, 2002), o agrupamento fuzzy, como o algoritmo Fuzzy C-means (Bezdek, 1981), e a clusterização espectral (Ng et al., 2002). Essas técnicas oferecem abordagens distintas para identificar padrões e estruturas em conjuntos de dados, permitindo uma compreensão mais aprofundada e *insights* valiosos.

Existem diferentes abordagens e algoritmos para a clusterização, conforme citado anteriormente, mas neste trabalho, será abordado especificamente o algoritmo K-means.

O algoritmo K-means é amplamente empregado em mineração de dados e aprendizado de máquina para realizar a tarefa de clusterização. Sua aplicação é eficiente e sua implementação é relativamente simples (MACQUEEN, 1967).

Segundo Arthur e Vassilvitskii (2017), consiste em uma técnica popular de agrupamento de dados, em que os dados são particionados em K clusters, minimizando a soma das distâncias quadradas entre cada ponto de dados e o centro de seu cluster atribuído, sendo assim um processo de otimização iterativo, que começa com um conjunto inicial de centros de cluster e os atualiza iterativamente até que a convergência seja alcançada.

2.4 Análise de dados

Segundo Xia (2014), a análise de dados se trata de um método de inspeção, limpeza, transformação e modelagem de dados, que tem como objetivo encontrar informações importantes, para o apoio à tomada de decisões, desempenhando um papel importante para a tomada de decisões. A análise de dados é um procedimento em que se adquire dados brutos, e em seguida realiza-se a conversão dos mesmos em informações pertinentes para a tomada de decisão dos usuários (JUDD; MCCLELLAND, 1989).

Segundo Pierson (2019), a análise de dados pode ser dividida em quatro tipos de análises, que são elas, análise descritiva, análise preditiva, análise prescritiva e análise diagnóstico.

De acordo com Kuhn e Silge (2022), o objetivo de um Modelo Descritivo é ilustrar ou descrever as características de um grupo de dados, com isso, torna-se possível enxergar padrões, tendências ou qualquer qualidade e atributo de uma variável. Segundo Reis (2002), a Análise Descritiva é a fase em que se estuda os dados que foram coletados, utilizando-se de métodos de Estatística Descritiva para estruturar, sintetizar, e caracterizar os pontos mais relevantes de um conjunto de características que foram examinadas, ou até mesmo realizar o comparativo entre dois ou mais conjuntos.

Os Modelos Preditivos buscam investigar padrões que são vistos em dados históricos, para realizar a identificação de riscos e oportunidades, os modelos fazem a relação entre muitos fatores para possibilitar a avaliação do risco ou potencial agregado a um determinado conjunto de condições, aconselhando a tomada de decisão (ECKERSON, 2007). De acordo com Finlay (2014), a Análise Preditiva é a área de estudo estatístico que tem como finalidade extrair informações dos dados, e utilizá-los para identificar padrões de comportamento, bem como antecipar tendências. Auxiliando a previsão de eventos desconhecidos no futuro, presente ou passado.

Segundo Gandomi e Haider (2015), a Análise Prescritiva emprega a otimização de maneira que identifique as melhores alternativas objetivando maximizar ou minimizar algum objetivo. Tem como característica testar vários cenários, avaliando cada um deles para indicar qual ação fornece o resultado mais desejável (ROWE, 2017). De acordo com Victorino (2017), nesse tipo de análise o projeto experimental busca responder perguntas sobre o porquê algo aconteceu, por meio da utilização de experimentos, com isso, a otimização tenta encontrar o nível ideal de uma variável definida, bem como suas relações com outra variável.

Por fim, o Modelo de Análise Diagnóstica tem como objetivo encontrar relações de causa e efeito para solucionar um determinado acontecimento, o processo é baseado em probabilidades, a fim de estabelecer uma relação baseada em um acontecimento passado (FINLAY, 2014).

Durante o desenvolvimento deste estudo, foram aplicadas diferentes abordagens analíticas para obter *insights* valiosos. Foram utilizadas técnicas de análise descritiva para explorar os dados e identificar padrões e tendências significativas relacionadas à criminalidade. Além da análise preditiva, que foi empregada para realizar previsões sobre o comportamento das variáveis, com base em dados históricos. A análise prescritiva desempenhou um papel importante na identificação de estratégias e recomendações para aprimorar as políticas de segurança pública, a fim de reduzir a criminalidade. Por fim, a análise diagnóstica permitiu compreender as causas subjacentes dos padrões observados e investigar os fatores que contribuíram para o comportamento dos índices criminais.

Ao integrar essas quatro abordagens analíticas, foi possível obter uma compreensão abrangente e aprofundada dos dados criminais, fornecendo suporte para a formulação de estratégias eficazes de segurança pública.

3. METODOLOGIA

Nesta seção serão apresentadas a classificação da metodologia utilizada e as etapas que serão realizadas para a execução do trabalho. Apresenta-se primeiramente a classificação desta pesquisa, e em seguida descreve-se os procedimentos utilizados para a coleta e análise dos dados.

3.1 Classificação da pesquisa

A seguinte pesquisa apresenta uma abordagem quantitativa, visto que recorre a linguagem matemática para caracterizar as causas de um fenômeno, bem como a relação entre variáveis. A pesquisa quantitativa, tem sua base no pensamento lógico, buscando enfatizar o raciocínio dedutivo, as regras da lógica, bem como os atributos mensuráveis da experiência humana (SILVEIRA; CORDOVA, 2009).

Classificando-se como pesquisa de levantamento, tendo em vista que envolve a coleta e análise de dados, com o objetivo de obter *insights* e embasar a tomada de decisões no campo da segurança pública.

O trabalho é definido como sendo de natureza aplicada, uma vez que sua construção tem a finalidade de conceber conhecimento para a aplicação prática e imediata, direcionadas à solução de problemas específicos.

Quanto ao objetivo, este estudo se caracteriza como sendo exploratório e descritivo. Para Gil (2018), o objetivo da pesquisa, definida como exploratória, é proporcionar uma melhor compreensão de um problema com o objetivo de expô-lo e criar possibilidades de soluções, ainda segundo o autor, a pesquisa de caráter descritivo visa detalhar um fenômeno, as características de um grupo, e detectar associações existentes entre variáveis sugeridas na população estudada.

3.2 Etapas da pesquisa

Para a elaboração da pesquisa foi realizada a coleta de dados referentes aos homicídios no estado do Rio Grande do Norte – RN, disponibilizados pela secretaria pública e da defesa

social, conforme regulamentado pela Lei de Acesso à Informação, Lei nº 12.527. É importante reforçar que todas as informações são de caráter sigiloso, sem identificação dos envolvidos.

Dentre os indicadores utilizados para análise, incluíram variáveis como incidência de CVLI's por localização geográfica, tipos cometidos, períodos anuais, além de informações socioeconômicas.

Posteriormente, foi realizada a análise dos dados coletados, levando em consideração variáveis sociais, demográficas e econômicas, a fim de estruturar um fluxo de análise sobre a violência, auxiliando em políticas públicas de prevenção e reação. Durante a análise foram selecionadas variáveis com base em sua relevância, para compreender os diferentes aspectos da criminalidade e identificação de possíveis padrões, além do tratamento prévio dos dados, a fim de remover variáveis não numéricas, bem como objetos nulos, que não apresentavam relevância para o estudo.

Com os dados coletados e tratados, realizou-se a análise utilizando a linguagem de programação Python. De acordo com Foundation (2001), Python é uma linguagem de programação interpretada, de alto nível, com semântica dinâmica. Possui estruturas de dados que a torna atraente para o desenvolvimento rápido de aplicativos, ou para ser usada como uma linguagem de *script*, servindo para conectar componentes existentes juntos.

Além da utilização da linguagem de programação, também foram utilizadas as ferramentas de análise e manipulação de dados. Para as análises descritivas foram utilizados gráficos e tabelas, já para a realização das análises prescritivas e preditivas, se fez o uso das bibliotecas Python. Segundo Pierson (2019), no Python, uma biblioteca funciona como uma coleção especializada de *scripts* criados por outra pessoa para desempenhar conjuntos específicos de tarefas, ou seja, são uma coleção de módulos e funções imprescindíveis que buscam reduzir a utilização de códigos no programa.

Segundo Pydata (2008), a biblioteca Pandas proporciona estruturas de dados de simples utilização, porém de alto desempenho, além de ferramentas de análise de dados para a linguagem de programação Python. Essas estruturas podem ser observadas como tabelas, além de possuir maneiras distintas de serem acessadas e futuramente modificadas. A biblioteca foi utilizada no trabalho para a manipulação dos dados em geral.

A biblioteca *Scikit Learn* funciona como um conjunto de ferramentas simples, porém eficientes para a análise preditiva de dados, que oferece uma gama de variedades de algoritmos integrados, dentre eles, Algoritmos de Classificação que indicam a qual categoria o objeto em estudo pertence, Algoritmos de Regressão que busca entender a relação entre os dados de

entrada e saída prevendo um atributo de valor contínuo relacionado a um objeto, já os Algoritmos de Agrupamento são utilizados em sua maioria para o agrupamento de maneira automática de dados que possuem as mesmas características em conjuntos (SCIKIT-LEARN, 2011).

De acordo com Vanderplas (2016), Matplotlib é uma biblioteca bastante eficaz e flexível utilizada para a visualização de dados em Python. Ela oferece inúmeros recursos, desde gráficos simples até gráficos mais robustos e personalizados. O autor destaca também, a importância da biblioteca para a análise de dados em Python.

Para a produção de gráficos informativos, foi utilizada a biblioteca Seaborn, que, segundo Waskom (2017), a biblioteca visa tornar a visualização uma parte central da exploração e compreensão dos dados. Sua principal função são as plotagens orientadas a conjunto de dados, operando em quadrados de dados e matrizes contendo conjunto de dados inteiros e executando internamente o mapeamento semântico necessário e a agregação estatística para gerar gráficos informativos.

A pesquisa foi realizada no estado do Rio Grande do Norte – RN, levando em consideração os cento e sessenta e sete municípios do estado, não foram analisados nem bairros, nem zonas.

4. RESULTADOS

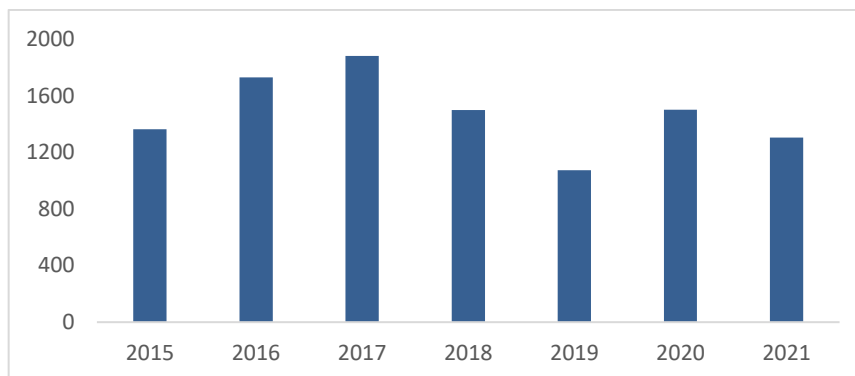
Na presente seção serão apresentados os resultados obtidos através da análise dos dados, que foram divididos em Análise Descritiva, a qual oferece uma visão geral dos dados coletados, e Análise Exploratória, que busca investigar as relações entre as variáveis, identificar tendências e padrões. Esses subtópicos fornecerão uma compreensão detalhada dos dados e possibilitarão a interpretação dos resultados obtidos.

4.1 Análise Descritiva

Durante a análise dos dados, constata-se que o ano de 2017 apresentou o maior índice de incidentes criminais no estado do Rio Grande do Norte (RN), conforme apresentado na Figura 1. Segundo o Atlas da Violência (2018), a taxa de homicídios que anteriormente correspondia a 9,6 pessoas por 100 mil habitantes no ano de 2000, alavancou para 62,8 no ano de 2017, ano esse em que ficou marcado pelas mais altas taxas nas Condutas Violentas Letais Intencionais (CVLI's). Segundo Rodrigues (2019), a política de repressão às drogas e às armas de fogo estabelecida em Brasília não foi implementada de maneira eficaz pelo sistema de justiça do Rio Grande do Norte, atrelado a isso, a superlotação e os massacres nos estabelecimentos penais contribuíram para esse processo de expansão da violência no estado. Como exemplo, o ano de 2017 em que ficou marcado pela tragédia na Penitenciária de Alcaçuz.

Conforme observa-se na Figura 1, nos anos seguintes nota-se uma redução da violência. Segundo a Revista Obvium (2020), dentre os fatores relacionados à redução da violência, destaca-se o resgate do controle prisional, a integração das ações de segurança pública, além de avanços na gestão das investigações da Divisão Especializada de Homicídios e Proteção à Pessoa, atrelado a isso, houve também a otimização de recursos humanos do policiamento ostensivo e a maior utilização dos dados criminais, para que se fosse possível realizar uma análise bem como o direcionamento de ações de policiamento e investigação.

Figura 1 - Ocorrências criminais anuais no estado do Rio Grande do Norte (2015 – 2021)



Fonte: Autoria Própria

Na Tabela 1, é exposto as estatísticas mensais para cada ano analisado, em números absolutos com relação ao número de ocorrências de CVLI's, dentre as variáveis apresentadas estão médias, moda, mediana, desvio padrão, bem como seus mínimos e máximos. Evidencia-se que o ano de 2019 tem a menor média de ocorrências entre os anos estudados sendo o ano menos violento, em contrapartida, o ano de 2017 tem a maior média, caracterizando-se como o ano mais violento dentre o período de 2015 a 2021.

Já com relação a variável Moda os anos de 2015, 2017, 2020, 2021 não apresentou frequência (repetição) de valores durante seus doze meses analisados, com isso, os seus campos estão em branco. Ao observarmos a Mediana, mais uma vez temos que o ano de 2017 se destaca dos demais, isso porque, sua medida central é a mais elevada, corroborando com a afirmação de que é o ano mais violento.

O ano de 2015 apresentou o menor desvio padrão dentre as amostras analisadas, isso indica que, é o ano em que os dados apresentam maior uniformidade, diferentemente do ano de 2018, onde teve o maior desvio padrão dentre as amostras analisadas, ou seja, houve meses em que o número de ocorrências apresentou um valor bastante elevado, já em outros meses, esse número caiu de maneira significativa.

Com relação aos mínimos e máximos, eles reforçam a afirmação dos anos mais violentos, bem como os menos violentos. Basta analisar, por exemplo, o ano de 2017, onde apresentou o maior mínimo dentre os anos analisados, bem como o maior máximo. Ao comparar os anos 2017 e 2019, percebe-se que, o ano de 2017 apresenta o mínimo de 121, já o ano de 2019, tem como seu máximo 121, evidenciando os seus extremos.

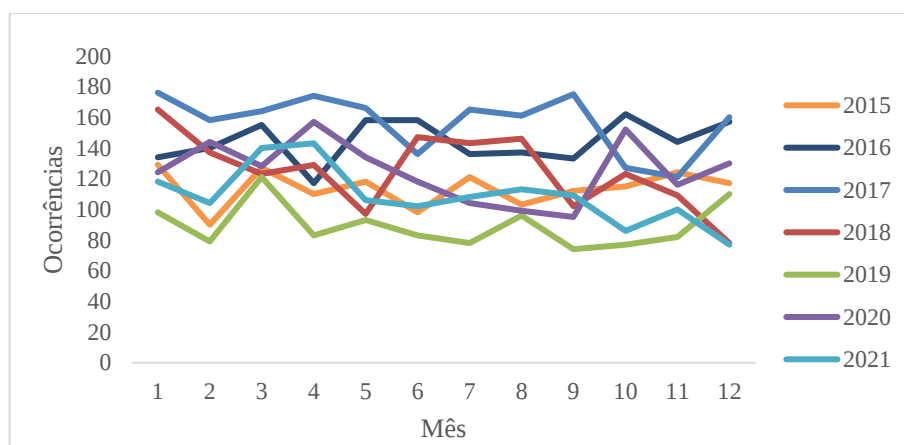
Tabela 1 - Estatísticas da taxa de CVLI's nos anos de 2015 - 2021 no estado do RN

ANO	Média	Moda	Mediana	Desvio padrão	Mínimo	Máximo
2015	113,7	-	116	11,8	90	129
2016	144,3	158	142	13,8	117	162
2017	156,9	-	162	18,7	121	176
2018	124,9	123	126	24,9	78	165
2019	89,5	83	83	14,5	74	121
2020	125,1	-	126	19,9	95	157
2021	108,8	-	107	18,9	77	143

Fonte: Autoria Própria

A Figura 2 permite observar a evolução mensal no número de CVLI's em cada ano, no período de 2015 a 2021. Verifica-se claramente que, exceto nos anos de 2016 e 2020, o mês de janeiro se inicia com um índice elevado, e apresenta uma queda no mês seguinte. Outro fator importante a ser analisado é que, ao ano de 2019 apresentou a série histórica com o menor número de ocorrências. O conjunto de dados mostra um padrão de sazonalidade, com um pico no número de crimes ocorrendo no mês de janeiro, bem como no meio do ano, por volta de junho e agosto, além de uma queda no número de crimes no início e no final do ano.

Figura 2 - Variação mensal de CVLI's no Rio Grande Do Norte (2015-2021)



Fonte: Autoria Própria

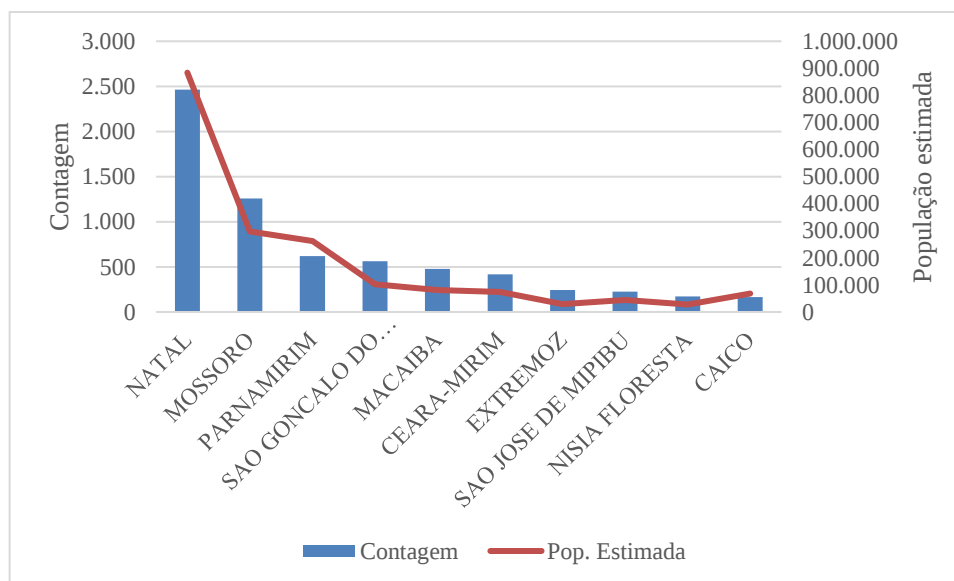
A Figura 3, apresenta um ranqueamento dos 10 municípios mais violentos do estado levando em consideração o período de 2015 a 2021, tendo como líder do *ranking* a capital Natal, o que se associa ao fato de ser a cidade mais populosa do estado.

Há uma relação positiva entre o tamanho da população e as taxas de criminalidade. De acordo com Martins (2010), uma maior concentração populacional em determinadas regiões contribui para que o índice de criminalidade se eleve, ou seja, regiões com densidade demográfica alta tendem a apresentar um índice mais alto de criminalidade. Segundo Hartung (2009), uma das razões para isso acontecer, tem relação com o maior anonimato das cidades grandes, dando ao criminoso mais facilidade para se esconder, quando procurados pela polícia.

Ao observar os seis primeiros municípios listados na Figura 3, nota-se que se trata dos seis municípios mais populosos do estado do RN de acordo com o censo demográfico divulgado pelo IBGE em 2010, onde Natal ocupa a primeira posição com a população estimada de 803.739 habitantes, seguida de Mossoró (259.815), Parnamirim (202.456), São Gonçalo do Amarante (87.668), Macaíba (69.467) e Ceará-Mirim (68.141).

Esses dados fortalecem a ideia de que o fator populacional estabelece uma relação direta com os incidentes criminais observados em uma determinada população.

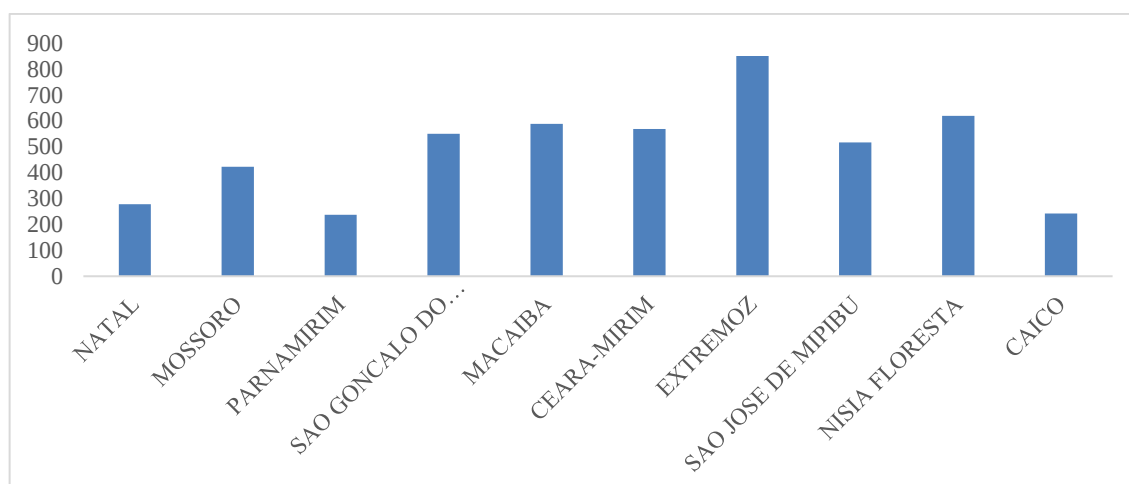
Figura 3 - Ranking dos 10 municípios mais violentos no Rio Grande do Norte (2015 - 2021)



Fonte: Autoria Própria

Ao observar a Figura 4, onde a taxa de ocorrências por cem mil habitantes é exposta, percebe-se que há uma mudança no *ranking*, isso porque é possível realizar uma comparação entre locais com diferentes tamanhos de população e neutralizar o crescimento populacional. Ou seja, o município de Natal é o mais populoso do estado do RN, portanto o número absoluto de CVLI's é naturalmente maior, por outro lado, utilizando a taxa por cem mil habitantes, Natal tem relativamente menos homicídios do que outros municípios analisados.

Figura 4 - Taxa de ocorrências criminais a cada 100 mil habitantes



Fonte: Autoria Própria

Agora, observando a Tabela 2 onde é exposto o tipo de morte levando em consideração o seu número de ocorrências, evidencia-se que o Homicídio Doloso, apresenta o maior expoente, foi o tipo mais observado correspondendo a quase 95% do total registrado. Seguido de Intervenção Policial, Latrocínio, Lesão Corporal Seguida de Morte, e por fim, Feminicídio.

Tabela 2 - Tipo de morte registradas e número de ocorrências

Tipo de morte	Contagem
Homicídio doloso	9836
Intervenção policial	295
Latrocínio	118
Lesão corporal seguida de morte	76
Feminicídio	33
Total	10358

Fonte: Autoria Própria

A seguir, exibe-se na Tabela 3 definições sucintas a respeito dos crimes citados na tabela anterior.

Tabela 3 - Definições dos crimes

Tipo de crime	Definição
Homicídio doloso	É o ato de matar alguém de forma intencional, com a vontade de causar a morte. É considerado um crime contra a vida e está previsto no Código Penal Brasileiro, no artigo 121.

Intervenção policial	Refere-se a ações realizadas por agentes de segurança pública, como policiais, no exercício de suas funções, que resultam na morte de uma pessoa.
Latrocínio	É o crime que consiste no roubo seguido de morte. Ou seja, ocorre quando um roubo é cometido e, durante sua execução, a vítima é assassinada. Está previsto no Código Penal Brasileiro, no artigo 157, §3º.
Lesão corporal seguida de morte	Refere-se a situações em que uma pessoa sofre lesões corporais graves que resultam em sua morte posteriormente. O crime está previsto no Código Penal Brasileiro, no artigo 129, §3º.
Feminicídio	Trata-se do homicídio de uma mulher em razão de sua condição de gênero. Envolve violência doméstica, discriminação de gênero e outras formas de opressão contra as mulheres. No Brasil, o feminicídio foi tipificado como crime pela Lei nº 13.104/2015.

Fonte: Autoria Própria

4.2 Análise Exploratória

Na Tabela 4, são apresentadas todas as variáveis presentes no banco de dados, bem como sua descrição. A mesma nomenclatura foi utilizada durante toda a análise exploratória dos dados.

Tabela 4 - Variáveis e descrição presentes no banco de dados

Variável	Descrição
Pop_estim	População absoluta em valores estimados por município (IBGE, 2020)
Pes_ocup	Porcentagem da população total que se encontra atualmente empregada ([Número de indivíduos ocupados no concelho / população total do concelho $\times$ 100] (2019)
Sal_médio	Rendimento médio mensal dos trabalhadores formais (2019) BRL USD
Pop_ocu	Número de pessoas ocupadas no município (%) (2019)
Pe_pop_mei	Percentual da população com renda nominal mensal per capita de até 1/2 salário-mínimo: [População residente em domicílios particulares permanentes com renda mensal de até 1/2 salário-mínimo / População total residente em domicílios particulares permanentes] $\times$ 100 (2010)
Num_Fund	Número absoluto de escolas que oferecem educação básica no município (2020)
Num_Emed	Número absoluto de escolas com ensino médio no município (2020)
IDHM	Índice de Desenvolvimento Humano Municipal, baseado nas seguintes dimensões: renda, educação e saúde (2010)
PIBpcapta	O PIB per capita é obtido dividindo o valor total do PIB pelo número de habitantes de uma região em específico. (2018)
Tx_Escol	[População residente no município de 6 a 14 anos matriculada no ensino regular / População total residente no município de 6 a 14 anos] $\times$ 100 (2010)
SUS	Estabelecimentos que oferecem serviços básicos de saúde e fazem parte do Sistema Único de Saúde (2009).
Esgotament	[População total residente em domicílios particulares permanentes com rede coletora geral e fossa séptica / População residente total em domicílios particulares permanentes] $\times$ 100 (2010)
Area	Área total do município com base em áreas rurais, áreas urbanas, núcleo urbano, núcleo rural e áreas urbanas com alta densidade de edificações
Area1	Área total do município com base em núcleo urbano, núcleo rural, áreas urbanas com alta densidade de edificações e áreas urbanas com baixa densidade de edificações
Area2	Área total do município com base no núcleo urbano e áreas urbanas com altas densidades de edifícios
den	[População absoluta em valores estimados por município/Área]
den1	[População absoluta em valores estimados por município/Área2]
den2	[População absoluta em valores estimados por município/Área3]

Fonte: Autoria Própria

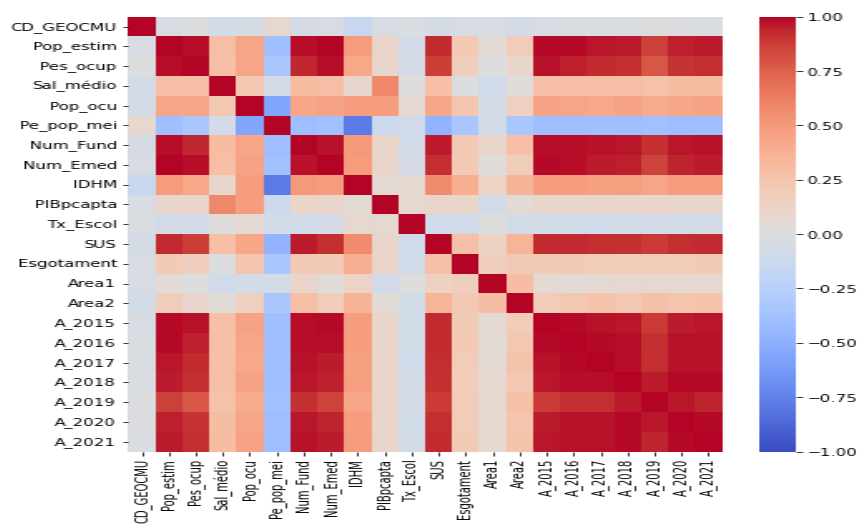
De acordo com Oliveira e Garcia (2017), uma matriz de correlação é uma ferramenta estatística que possibilita explorar as relações entre duas ou mais variáveis. Ou seja, a matriz de correlação é uma ferramenta estatística que permite explorar as relações entre duas ou mais variáveis. É uma tabela que mostra as correlações entre todos os pares possíveis de variáveis, tipicamente representadas pelos coeficientes de correlação de Pearson, que variam de -1 a 1.

Segundo Johnson e Bhattacharyya (2010), o coeficiente de correlação de Pearson é utilizado para medir a força de uma relação linear entre duas variáveis, onde o valor do coeficiente pode variar de -1 a 1. Assim, quanto mais próximo de 1 for o coeficiente de correlação, mais forte será a correlação positiva entre as variáveis, e quanto mais próximo de -1, mais forte é a correlação negativa entre as variáveis. Um coeficiente de correlação de 0 indica nenhuma correlação entre as variáveis.

No contexto dos dados criminais que foram analisados, a matriz de correlação que foi utilizada para explorar as relações entre diferentes variáveis relacionadas ao crime, como taxas de criminalidade, fatores demográficos, indicadores econômicos ou localização geográfica. Ao examinar as correlações entre essas variáveis, foi permitido identificar possíveis padrões ou tendências nos dados e gerar hipóteses para análise posterior.

Na Figura 5, está o Mapa de calor gerado a partir da Matriz de correlação, o Mapa de calor foi utilizado pois facilita a visualização dos dados, onde cada célula recebe uma cor com base em seu valor, dessa forma, permite a identificação rápida dos padrões e tendências, onde a cor azul escuro representa uma forte correlação negativa, já vermelha escura representa uma forte correlação positiva.

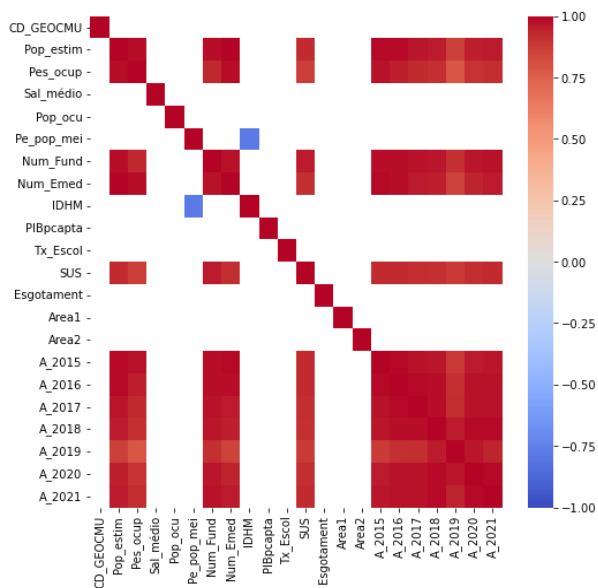
Figura 5 - Mapa de calor referente ao coeficiente de correlação



Fonte: Autoria Própria

Para auxiliar na visualização dos dados, foi criado um filtro para o mapa de calor, para que assim fosse possível focar em regiões específicas. Ao aplicar o filtro de mapa de calor, foi escolhido um coeficiente de correlação acima de 0.7 por ser considerado um valor de alta correlação, retirando assim todos os valores abaixo, auxiliando na visualização dos dados. Na Figura 6, pode-se visualizar a matriz filtrada, onde são exibidos os valores que apresentam alta correlação, seja ela positiva, ou, negativa.

Figura 6 - Mapa de calor filtrado



Fonte: Autoria Própria

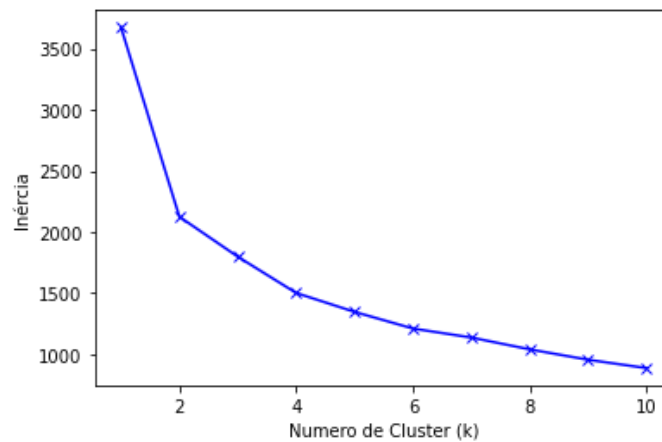
Com as variáveis correlacionadas, e filtradas de acordo com seu coeficiente de correlação, iniciou-se o agrupamento das variáveis que possuem padrões semelhantes de correlação entre si. Com isso, foi possível identificar grupos de variáveis com maior probabilidade de estarem relacionadas ao resultado de interesse, ou entre si. Para realizar esse agrupamento foi utilizado o algoritmo de clusterização K-means ou, como pode ser também chamado K-médias.

De acordo com Hastie, Tibshirani e Friedman (2009), o algoritmo K-means inicia pela seleção aleatória de K pontos de dados, esses dados serão os centros iniciais do cluster. Posteriormente, ele atribui em cada ponto de dados ao centro do cluster mais próximo, e recalcula o centroide de cada cluster levando em consideração as novas atribuições. Esse processo é repetido até que haja a convergência, ou até que um número máximo de iterações seja alcançado.

Uma abordagem comum para decidir o número de clusters é observar a mudança na soma dos quadrados dentro do cluster, à medida que o número aumenta. Para isso, utilizou-se o Método do cotovelo, que, segundo Hastie, Tibshirani e Friedman (2009), o método é uma heurística para determinar o número de clusters em um conjunto de dados, o método utiliza a soma dos quadrados dentro de um cluster em função do número de clusters. A partir disso, é possível reconhecer o ponto de inflexão, ou ‘cotovelo’, que consiste basicamente em uma curva da soma dos quadrados dentro do cluster em função do número de clusters.

Na Figura 7, podemos observar o gráfico onde foi plotado o resultado do Método do cotovelo com base nos dados utilizados para a pesquisa. Ao analisar o ponto de inflexão do gráfico, temos que o número ideal de Clusters é 2, ou seja, é onde a taxa de diminuição começa a se estabilizar.

Figura 7 - Método do Cotovelo



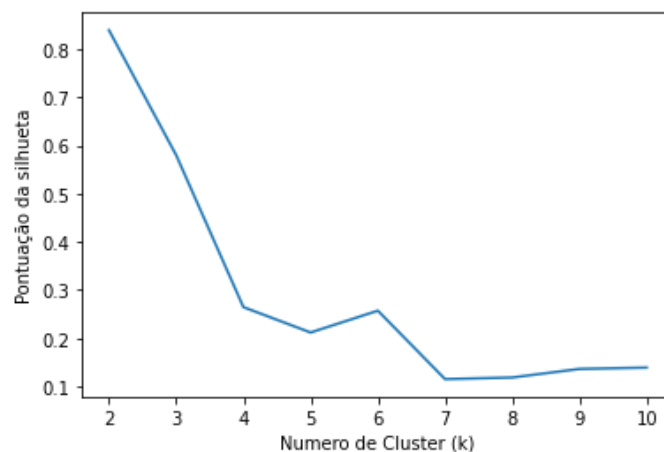
Fonte: Autoria Própria

No entanto, o método do cotovelo combinado com o método da silhueta fornece uma análise mais precisa dos dados, diferentemente de qualquer método utilizado sozinho. De acordo com Paudel e Jang (2020), o método da silhueta avalia a coerência dos agrupamentos baseados nas distâncias intra e intercluster.

Dessa forma, o método da silhueta é aplicado para determinar o número de clusters dentro do intervalo obtido pelo método do cotovelo (REVATHY; SUJATHA, 2018).

Na Figura 8, temos o gráfico onde foi plotado o resultado obtido através do método da Silhueta, aplicado aos dados utilizados durante esta pesquisa, observa-se que o resultado reafirma que o número ideal de Clusters é igual a 2, então a partir disso, podemos iniciar o processo de agrupamento dos dados.

Figura 8 - Método da Silhueta



Fonte: Autoria Própria

Com todas as variáveis filtradas, bem como o número ideal de clusters definido, iniciou-se a plotagem dos gráficos de dispersão. Onde as seguintes variáveis foram analisadas, População estimada, Pessoas ocupadas, Número de escolas com ensino básico no município, Número de escolas com educação secundária no município, bem como estabelecimentos que oferecem serviços básicos de saúde e fazem parte do Sistema Único de Saúde (SUS), durante os anos de 2015 a 2021.

Segundo Kelling e Coles (1996), o nível populacional exerce uma importante influência na compreensão de CVLI's, no geral, populações maiores apresentam taxas mais elevadas de crimes violentos.

Em áreas urbanas, onde a densidade populacional é mais elevada e a desorganização social é mais preponderante, a ocorrência criminal tende a ser maior (SHAW; MCKAY, 1942).

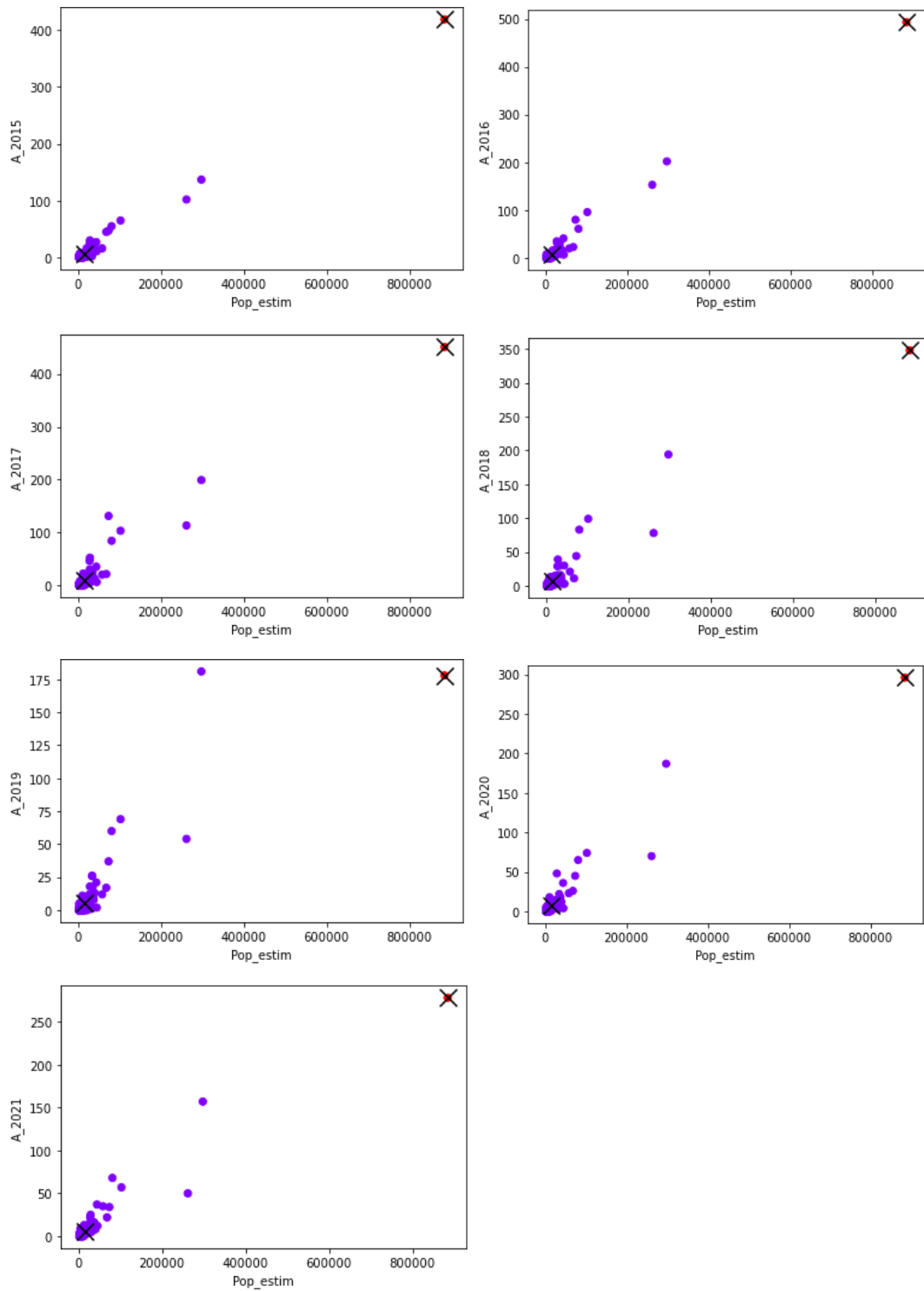
Contudo, a relação entre população e CVLI's é bastante complexa e inúmeros fatores podem influenciar. Uma conjectura comum é que, conforme a população aumenta, também se eleva a incidência de crimes violentos. A Organização Mundial de Saúde (OMS, 2014), determina que o número de homicídios em uma determinada população apresenta relação com seu tamanho, dessa forma, cidades maiores geralmente apresentam taxas mais altas do que as cidades menores.

Na Figura 9, estão plotados os gráficos de dispersão, pode-se observar que existe uma tendência geral de aumento nas estimativas populacionais ao longo dos anos. Existe uma correlação positiva entre as estimativas populacionais e o número de CVLI's, isso indica que, à medida que a população aumenta, há também um aumento correspondente no número de crimes violentos. Isso sugere que, áreas onde o crime é particularmente alto, cabe a alocação de recursos de maneira mais eficaz, considerando a implementação de intervenções direcionadas nessas áreas, como o aumento da presença policial, programas comunitários, bem como iniciativas para melhorar as oportunidades econômicas e o acesso a serviços sociais.

Com relação à análise de clustering, no primeiro cluster pode-se perceber que as variáveis mostram uma tendência semelhante de aumento de 2015 a 2017, seguida por uma queda em 2018 e 2019 e um aumento significativo em 2020. Já no segundo cluster essas variáveis mostram uma tendência mais consistente de se manterem estáveis, ou diminuem levemente ao longo do tempo. Esse cluster pode representar crimes com maior relação a fatores individuais, como comportamentos violentos.

De maneira geral, os agrupamentos auxiliam a identificar padrões e tendências nos dados, o que pode ser útil para desenvolver políticas e intervenções direcionadas para lidar com o tipos específicos de crime, os gráficos surgem como uma ferramenta útil para melhor entender a complexa relação entre crescimento populacional e crimes violentos, possibilitando o desenvolvimento bem como a alocação de recursos em evidências, a fim de reduzir as taxas de criminalidade, e melhorar a segurança pública.

Figura 9 - População estimada versus Ano



Fonte: Autoria Própria

De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE, 2010), é considerado como indivíduo empregado aquele que exerce algum tipo de atividade remunerada ou que obtém lucro, por pelo menos uma hora semanal, independente da situação empregatícia.

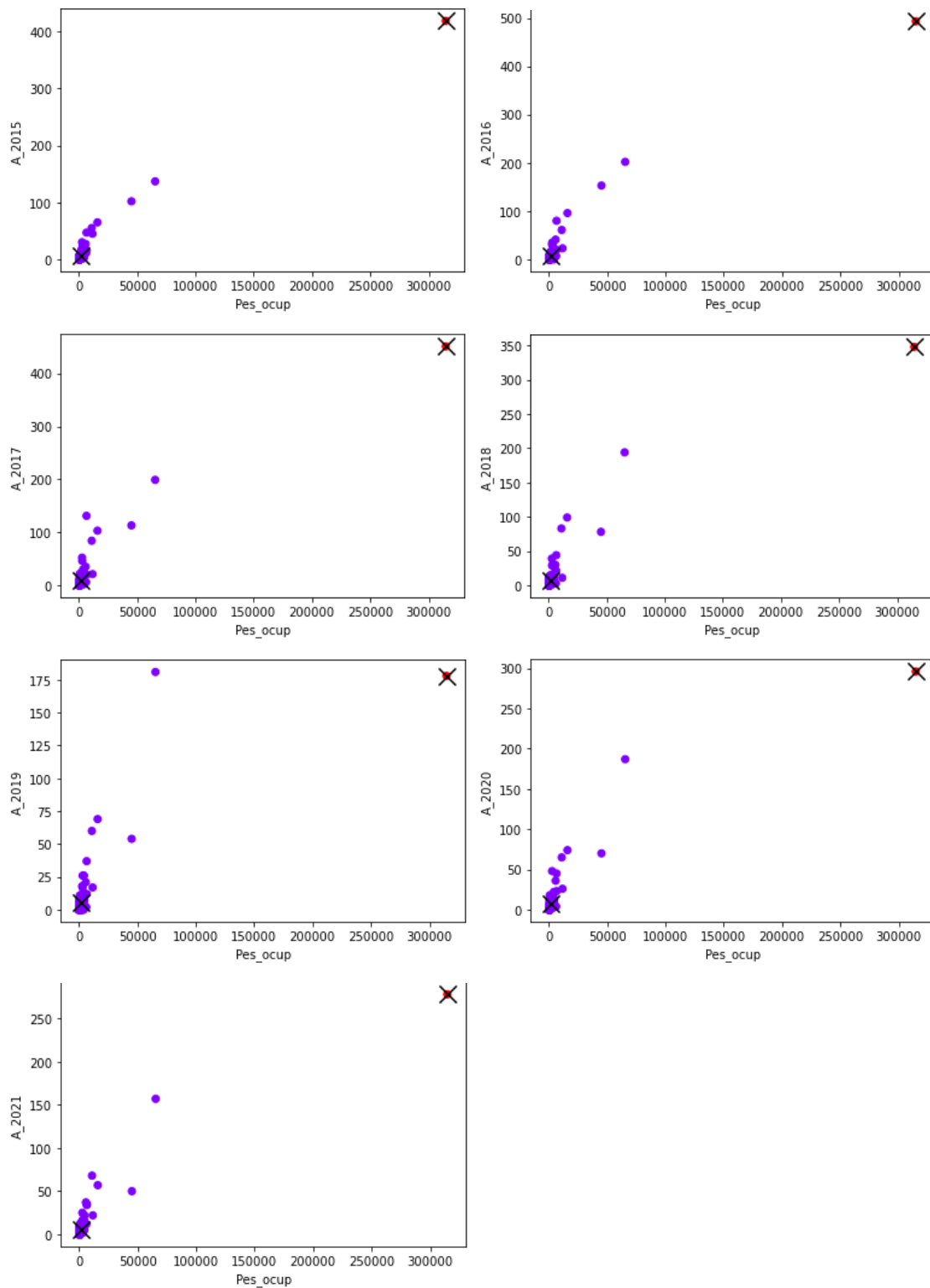
Dessa forma, pode-se entender como ‘Pessoa ocupada’ aquela que, está envolvida de maneira ativa em alguma forma de atividade produtiva, sendo ela um emprego remunerado, autônomo ou até mesmo voluntário.

De acordo com Alvazzi del Frate (2013), o desemprego é um dos fatores de maior relevância que contribuem para a ocorrência de crimes. Indivíduos desempregados correm maior risco de cometerem crimes, pois o desemprego desencadeia dificuldades financeiras, o que pode resultar em uma maior probabilidade de comportamentos criminosos, no intuito de obter dinheiro ou bens, recorrendo assim a atividades ilegais para sobreviver.

Na Figura 10, estão os gráficos de dispersão, onde o eixo X exibe o fator ‘Pessoa ocupada’ já o eixo Y representa a contagem de CVLÍ’s durante os anos analisados, que foram de 2015 a 2021, além disso, as variáveis estão divididas em dois clusters, conforme determinado anteriormente, com a utilização do método do cotovelo, e o método da silhueta, os pontos do centróide para cada cluster também são mostrados nos gráficos de dispersão. Os pontos lilás, representam os dados que pertencem ao Cluster 1, enquanto os pontos vermelhos representam os dados que pertencem ao Cluster 2.

O agrupamento indica que existe uma diferença significativa entre os valores, para os dois clusters. Isso sugere que, as variáveis apresentam uma forte relação entre si.

Figura 10 - Pessoa ocupada versus Ano



Fonte: Autoria Própria

Estudos têm demonstrado que existe uma correlação negativa entre os níveis de escolaridade e a incidência de CVLI's. De acordo com Kuhns e Maguire (1979), quanto mais baixo o nível de escolaridade, mais altas são as probabilidades de se envolver em comportamentos criminosos. Reafirmando isso, Tavares, Moraes e Gomes (2016), confirmaram que maiores níveis de escolaridade estão associados a menores níveis de violência e criminalidade.

Pertinente a isso, um estudo realizado por Werneck, Hasselmann e Oliveira (2017), salientou que indivíduos com menor escolaridade são mais propensos a serem vítimas de violência. Isso indica que a educação não apenas reduz a probabilidade de se envolver em comportamento violento, como também serve de fator de proteção contra a vitimização.

Na Figura 11, estão expostos os gráficos de correlação, que trata o fator número absoluto de escolas a nível fundamental versus incidentes criminais. Inicialmente pode-se observar que, em geral, o número de escolas que oferecem educação básica tende a diminuir ao longo do tempo, podemos ainda verificar que, existe uma diminuição no número de escolas no ano de 2017, ano em que ficou marcado como sendo o mais violento dentre os anos analisados, ao compararmos com o ano de 2016.

Pode-se constatar que o número de CVLI's foi mais elevado no ano de 2017, seguindo-se uma ligeira diminuição em 2018 e posteriormente um aumento considerável no ano de 2019, que se prolongou até 2020, apresentando uma breve diminuição no ano de 2021. De maneira geral, pode-se observar um padrão inconsistente, com flutuações em relação ao número de crimes, bem como ao número de indivíduos com ensino fundamental completo.

A correlação positiva levanta questões importantes sobre os fatores socioeconômicos e educacionais que podem influenciar a criminalidade. Embora à primeira vista possa parecer contraditório, essa relação pode ser compreendida considerando diversos aspectos.

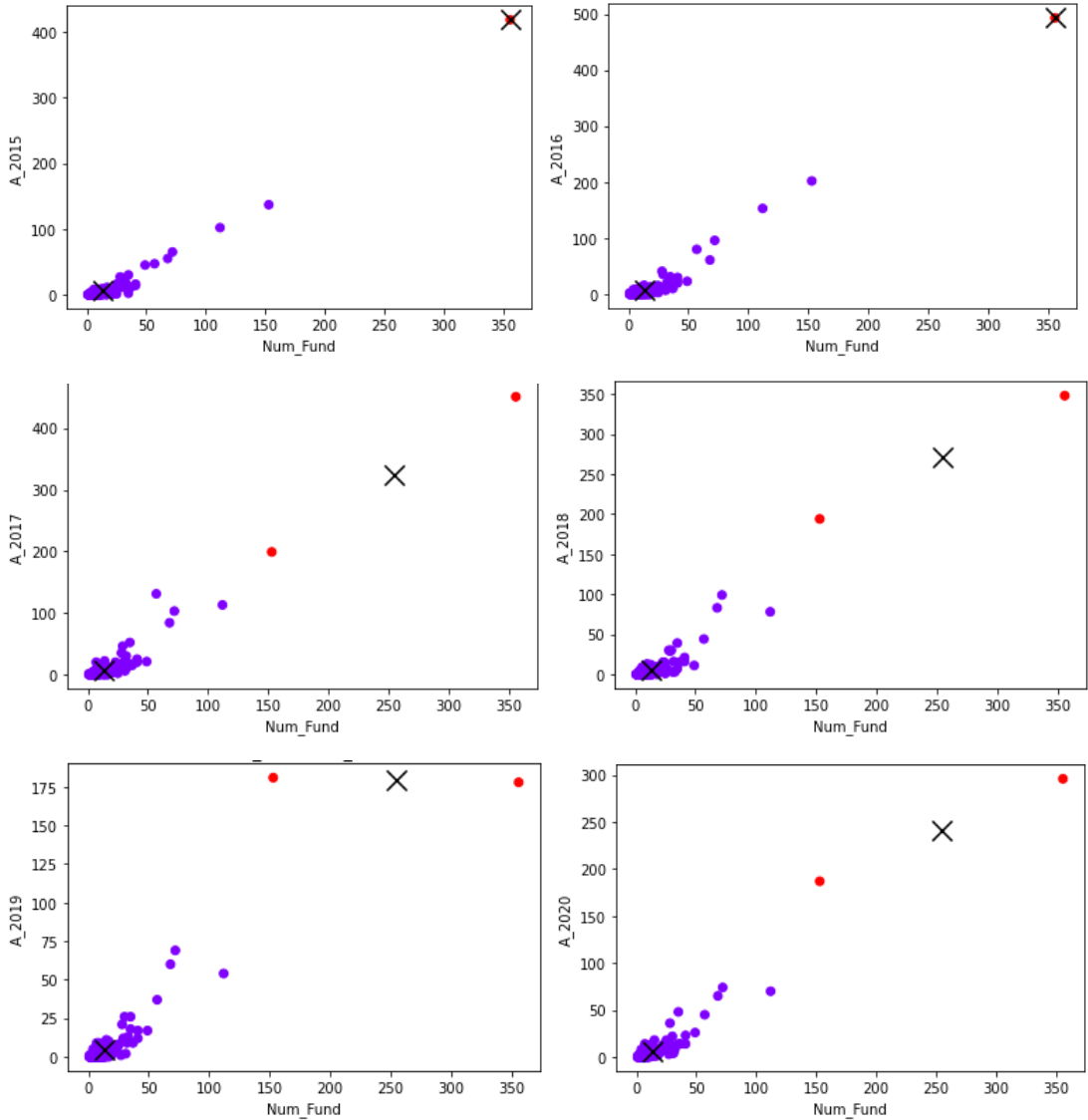
Primeiramente, o aumento no número de escolas com ensino fundamental pode refletir um maior acesso à educação em determinada região. No entanto, é importante considerar a qualidade da educação oferecida, pois a simples presença de escolas não garante necessariamente uma educação de qualidade. Desigualdades no sistema educacional, como a falta de recursos adequados, professores capacitados e infraestrutura adequada, podem limitar o impacto positivo da educação na redução da criminalidade.

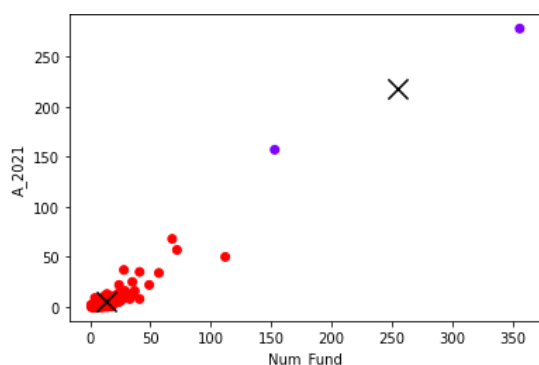
Além disso, a correlação pode estar relacionada a fatores socioeconômicos mais amplos. Áreas com maior número de escolas podem estar sujeitas a condições socioeconômicas desfavoráveis, como desigualdade de renda, falta de oportunidades de emprego e carência de

serviços básicos. Esses aspectos podem contribuir para o aumento da vulnerabilidade social e, consequentemente, para a ocorrência de crimes violentos.

Nesse contexto, é essencial realizar uma análise mais aprofundada para compreender as nuances dessa relação. Seria relevante considerar indicadores de qualidade da educação, como taxa de evasão escolar, desempenho dos alunos em testes padronizados e disponibilidade de recursos educacionais.

Figura 11 - Número absoluto de escolas que oferecem educação básica no município versus Ano





Fonte: Autoria Própria

Conforme desenvolvido anteriormente, estudos têm mostrado que indivíduos que possuem maior nível de escolaridade, como aqueles que concluíram o ensino médio de escolaridade, estão menos suscetíveis a cometer crimes violentos se comparado àqueles com menor escolaridade (ALVES, 2015).

Em um estudo realizado por Tavares et al. (2016), os autores realizaram uma revisão sistemática da literatura, onde encontraram evidências consistentes de que a educação está inversamente associada ao crime violento.

No Brasil, um estudo realizado entre 2000 e 2010, foi constatado que havia uma desigualdade na mortalidade por homicídios entre diferentes grupos raciais/étnicos e níveis de escolaridade, sendo indivíduos com menor nível de escolaridade mais vulneráveis à violência letal (WERNECK et al., 2017). Em estudo realizado nos Estados Unidos, constatou-se que o aumento nas taxas de conclusão do ensino médio acarretou uma redução nas taxas de homicídios (KUHNS; MAGUIRE, 1979).

Na Figura 12, estão plotados os gráficos de correlação com base nas variáveis número absoluto de escolas com ensino médio por município versus os incidentes criminais em seus respectivos anos, em geral, o número de crimes violentos está inversamente relacionado ao número absoluto de escolas com ensino secundário. Ou seja, conforme os números de escolas com ensino médio aumentam, o número de crimes tende a diminuir.

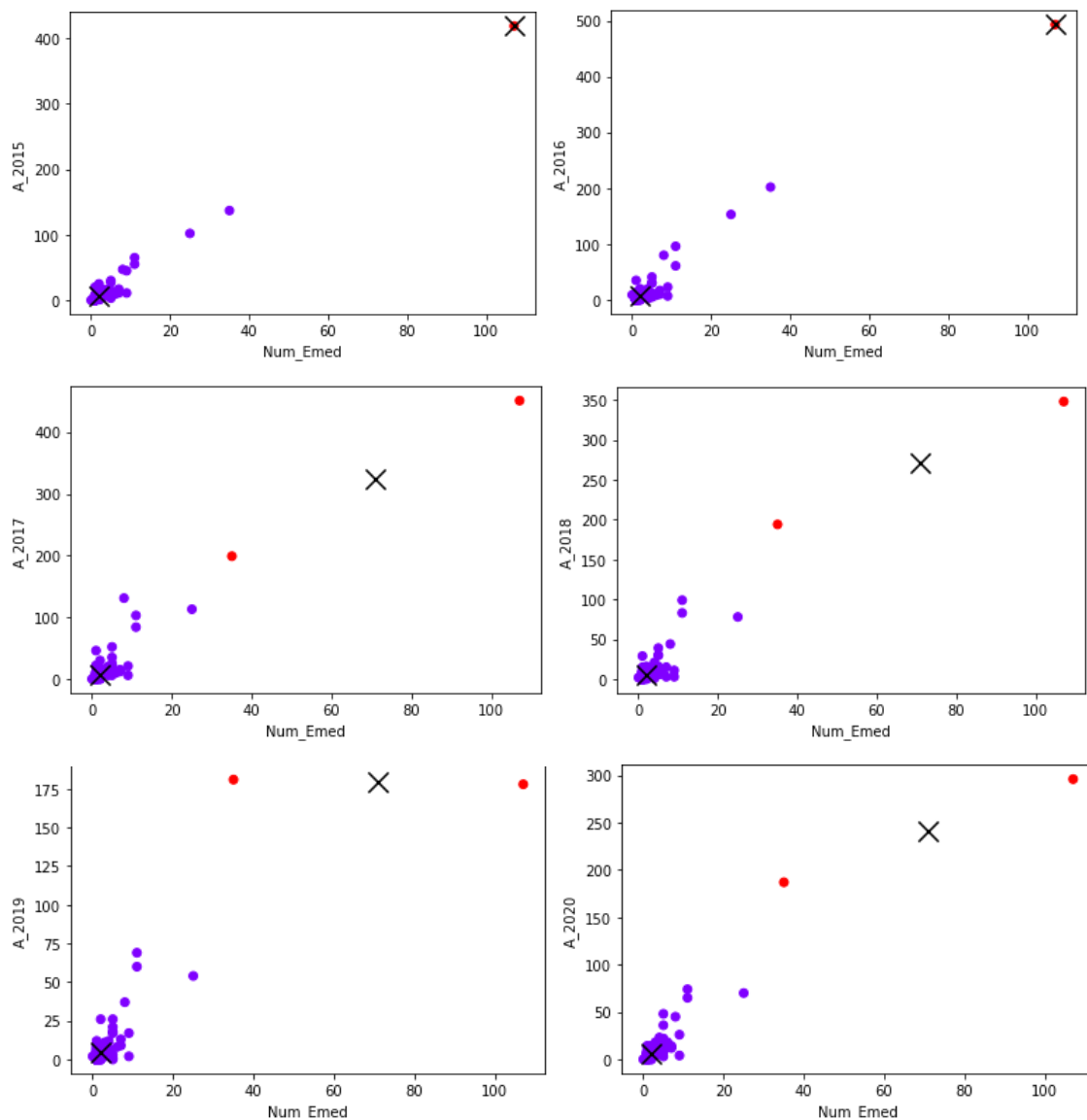
O ano de 2015, por exemplo, o gráfico mostra uma correlação positiva entre o número de crimes e o número de escolas com o ensino médio, já o ano de 2016, observa-se uma tendência diferente, onde o número de crimes aumentou de maneira significativa enquanto o número de escolas com ensino médio diminuiu, o que sugere uma correlação negativa entre as duas variáveis.

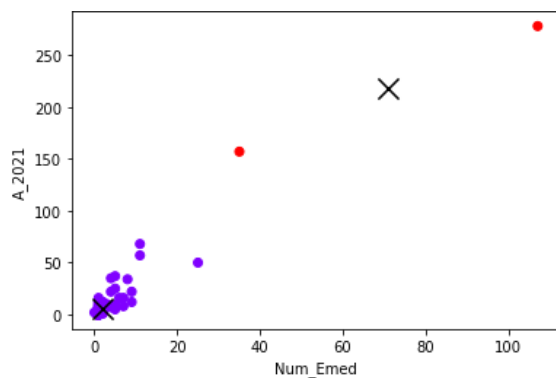
No ano de 2020, é possível visualizar uma queda no número de crimes, enquanto o número de escolas com ensino médio aumentou, enquanto, o ano de 2021, mostrou um pequeno

aumento do número de crimes, enquanto o número de escolas com ensino secundário continuou a aumentar.

De maneira geral, esses gráficos sugerem que existe uma ligação entre educação e taxas de criminalidade, especialmente quando se trata do ensino médio. No entanto, ressalta-se que outros fatores influenciam nas taxas de criminalidade, como fatores socioeconômicos, porém, compreender a relação entre educação e crime, possibilita o desenvolvimento de políticas e intervenções mais eficazes, promovendo maior igualdade social e econômica. Vale salientar que, não basta apenas aumentar o número de instituições de ensino, mas sim, garantir um ensino de qualidade.

Figura 12 - Número absoluto de escolas que oferecem ensino médio no município versus Ano





Fonte: Autoria Própria

Pesquisas relacionadas a relação entre acesso a serviços de saúde, como o Sistema Único de Saúde (SUS), e a incidência de CVLI's vem apresentando um crescimento considerável.

Em um estudo realizado por Santana e Santos (2018), onde foi investigado a relação entre a cobertura da Estratégia de Saúde da Família, serviço de atenção primária prestada pelo SUS, e as taxas de homicídio nos municípios brasileiros, foi verificado uma correlação negativa entre a cobertura do serviço e as taxas de homicídio, indicando que conforme o acesso aos serviços de atenção primária aumentam, auxiliam na redução de incidência de crimes violentos.

Corroboram com isso um estudo realizado por Coelho, Alvarenga e Lima (2019), onde os autores analisaram a relação entre o acesso aos serviços de saúde, incluindo o SUS, e a incidência de homicídios nos municípios brasileiros. Foi constatado que, os municípios com maior acesso aos serviços de saúde apresentam menores taxas de homicídios, sugerindo que um melhor acesso aos serviços de saúde pode contribuir para a redução dos crimes violentos.

Da mesma maneira, outros estudos também defendem a ideia de que, melhorar o acesso aos cuidados à saúde, especialmente para serviços de saúde mental, pode auxiliar a reduzir a incidência de crimes violentos. Por exemplo, um estudo realizado por Wilson et al. (2014), descobriu que indivíduos com doenças mentais não tratadas corriam maior risco de cometerem crimes violentos, destacar a importância de fornecer serviço de saúde mental adequado e acessível a todos pode prevenir tais incidentes.

Na figura 13, estão plotados os gráficos referentes a estabelecimentos que oferecem serviços básicos de saúde e fazem parte do Sistema Único de Saúde versus as ocorrências de CVLI's durante os anos de 2015 a 2021. Pode-se observar que o percentual de Estabelecimentos que oferecem serviços básicos de saúde e fazem parte do Sistema Único de Saúde apresenta um ligeiro aumento de 2015 a 2019, com uma leve queda em 2020 e uma queda significativa em 2021. Já em relação a incidência de CVLI's, pode-se observar que houve um aumento

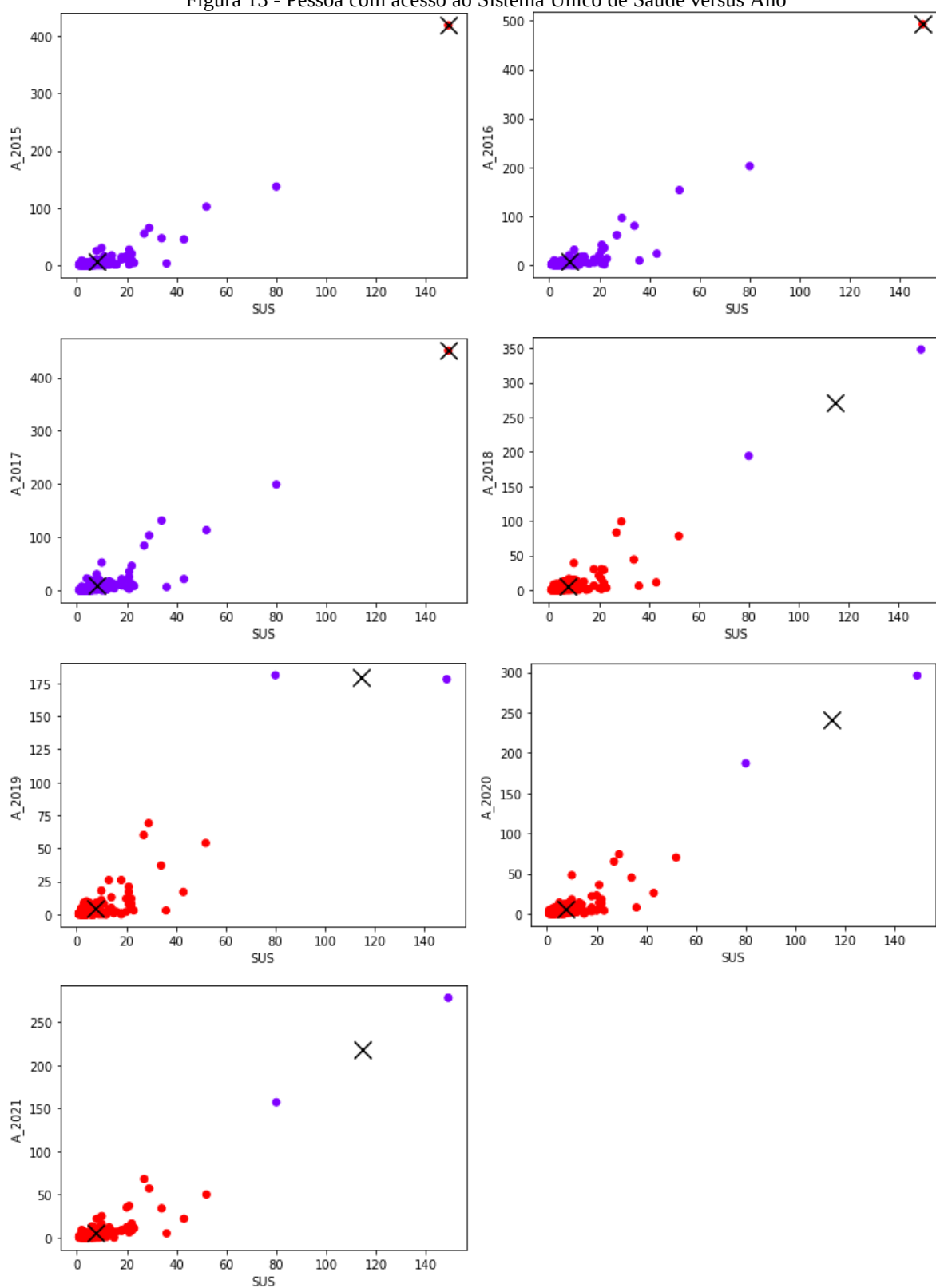
significativo de 2015 a 2017, com pico em 2017. No entanto, houve uma queda considerável no ano de 2019.

Paralelo a isso, o percentual de estabelecimentos que oferecem serviços básicos de saúde e fazem parte do Sistema Único de Saúde foi maior no ano de 2016, ano com menor número de CVLI's durante o período analisado. Isso sugere que o acesso à saúde pode estabelecer uma relação direta com a incidência de crimes.

Entender esses padrões possibilita a alocação de recursos com mais eficiência, a fim de reduzir a ocorrência de incidentes violentos, e proporcionar maior segurança para a sociedade. Por exemplo, durante a análise gráfica percebe-se que existe uma correlação positiva entre o acesso ao SUS e a diminuição do número de CVLI's, isso sugere que investir no SUS e ampliar o acesso à saúde pública pode ter um impacto positivo na redução de crimes.

De maneira geral, analisar e entender esses padrões e relacionamentos pode auxiliar os formuladores de políticas a tomarem decisões baseadas em evidências, e com isso, alocar recursos de maneira mais eficaz para reduzir o crime e melhorar a segurança pública.

Figura 13 - Pessoa com acesso ao Sistema Único de Saúde versus Ano



Fonte: Autoria Própria

Os resultados na análise mostram uma correlação positiva significativa entre a incidência de CVLI's e várias das variáveis examinadas, incluindo renda per capita, taxa de desemprego, escolaridade e densidade populacional, bem como o acesso ao Sistema Único de Saúde. Em particular, áreas com maior desigualdade de renda, maiores taxas de desemprego e maiores densidades populacionais apresentam maior incidência de CVLI's. Essas descobertas sugerem que abordar as disparidades socioeconômicas pode ser um fator-chave na redução da incidência de crimes nessas áreas.

Ao considerar a distribuição espacial dos CVLI's, a pesquisa revela que esses crimes geralmente se concentram em áreas específicas, com índices mais elevados de violência ocorrendo em determinados bairros ou regiões, isso sugere que, fatores como desorganização social, falta de acesso a recursos e oportunidades, atrelado a altos níveis de desigualdade contribuem para esses acontecimentos. Compreender a distribuição espacial do crime, auxilia os formuladores de políticas e agências de aplicação da lei a desenvolver estratégias mais direcionadas para e eficazes para a prevenção e redução de crimes.

5. DISCUSSÃO

Ao recapitular sucintamente as pesquisas abordadas ao longo deste trabalho, percebe-se que elas se concentram em modelos de decisão capazes de guiar a tomada de decisões em questões de segurança pública, ao analisar a interdependência entre variáveis socioeconômicas e criminais, com o intuito de identificar áreas com maior incidência de criminalidade.

Ao comparar com os resultados obtidos neste trabalho, é interessante notar que, também foram encontradas alta correlação entre variáveis como população estimada ($r = 0.95$), pessoas ocupadas ($r = 0.91$), número de escolas que oferecem ensino fundamental ($r = 0.96$) e médio ($r = 0.81$), bem como estabelecimentos que oferecem serviços básicos de saúde e fazem parte do Sistema Único de Saúde ($r = 0.91$) e a incidência de CVLI's. Isso fundamenta as conclusões dos estudos revisados, que mostram que existe uma relação entre variáveis socioeconômicas e criminalidade.

Com isso, o presente estudo foca especificamente nos CVLI's no estado do Rio Grande do Norte, ao analisar dados de 2015 a 2021, isso enriquece a literatura existente sobre o tema, e fornece uma compreensão mais abrangente dos fatores que contribuem para o crime violento na região, fornecendo novos *insights*.

Durante o desenvolvimento do estudo, confirma-se a importância de considerar variáveis socioeconômicas e demográficas no que se refere a análise da criminalidade, como mostram estudos anteriores. Com base nos coeficientes de correlação apresentados, pode-se inferir que as variáveis que mais influenciam a incidência de CVLI's são a população estimada, o número de escolas que oferecem ensino fundamental e médio e a taxa de acesso ao SUS. Essas variáveis estão diretamente relacionadas ao desenvolvimento socioeconômico da região e podem estar associados a fatores como desigualdade social, violência urbana e acesso limitado a serviços de saúde, o que pode contribuir para o aumento da criminalidade violenta. É importante ressaltar que os coeficientes de correlação indicam uma relação estatística entre as variáveis, mas não necessariamente uma relação de causa e efeito.

A utilização de técnicas de análise descritiva e cluster na presente pesquisa permitiu uma compreensão mais profunda dos padrões e tendências dos dados analisados. Por meio da análise descritiva, foi possível identificar as características gerais das variáveis investigadas, como média, moda, mediana, desvio padrão, mínimo e máximo. Já a técnica de cluster possibilitou a identificação de grupos de municípios com características semelhantes em relação

aos indicadores de criminalidade e socioeconômicos, o que pôde contribuir para a implementação de políticas mais direcionadas e eficazes.

No contexto da segurança pública, essas técnicas podem ser utilizadas para a identificação de áreas com maior incidência de crimes, auxiliando na tomada de decisões mais assertivas na alocação de recursos e estratégias de prevenção. Além disso, a análise de cluster pode ser útil para a identificação de grupos de indivíduos com características semelhantes, o que pode auxiliar na identificação de perfis criminais e na implementação de políticas públicas mais efetivas. Já a análise descritiva pode ser útil para entender as principais características de uma população ou fenômeno, auxiliando assim, na identificação de problemas e na elaboração de soluções.

Ao analisar os dados, foram identificadas variáveis-chave altamente correlacionadas com o crime violento. Esses resultados fornecem evidências adicionais da utilidade da análise de correlação na identificação de fatores que contribuem para a ocorrência de CVLI's no estado. A partir disso, é possível utilizar os dados de forma diferenciada, ou seja, com diferenciação criteriosa do tipo de crime, sua localização, para possibilitar ações rápidas onde o crime está concentrado. A alocação efetiva de recursos e tomada de decisão na segurança pública é fundamental e deve ser baseada em critérios de eficiência e eficácia, como destacado por estudos anteriores. Essas conclusões reforçam a importância de dos sistemas de informação gerencial no apoio à tomada de decisão em segurança pública e indicam caminhos para ações mais efetivas e focadas na prevenção e combate ao crime violento.

6. CONSIDERAÇÕES FINAIS

Diante dos resultados apresentados, pode-se concluir que a utilização de técnicas de análises de dados descritiva e de clusterização, são de grande relevância para a análise da ocorrência de crimes violentos letais intencionais no estado do Rio Grande do Norte. A análise abrangeu todos os municípios do estado, o que permitiu uma visão mais ampla e detalhada sobre a situação da violência em cada região.

Os resultados obtidos apontam para a necessidade de medidas mais efetivas de combate à violência no estado, especialmente em áreas onde a incidência de crimes é mais elevada. Essas informações podem ser utilizadas para orientar os responsáveis para a adoção de políticas públicas voltadas para a prevenção da violência, e para o fortalecimento da segurança pública, direcionando assim, os recursos para as áreas que mais necessitam.

Além disso, a utilização de técnicas de análise de dados pode ser uma importante aliada na identificação de padrões e na antecipação de situações de risco, contribuindo para a redução da violência no estado. Para isso, é necessário o investimento em tecnologia e em capacitação de profissionais para que sejam ampliadas as possibilidades de utilização da inteligência artificial no combate à violência.

O estudo levanta reflexões importantes sobre a necessidade de investimento em tecnologias avançadas para a análise de criminalidade, bem como a importância de considerar fatores socioeconômicos e demográficos em futuras pesquisas, além de incentivar discussões sobre as melhores formas de abordar questões a respeito da segurança pública, incluindo a implementação de políticas preventivas e a utilização de tecnologias avançadas para o combate à criminalidade. É importante ressaltar que os coeficientes de correlação indicam uma relação estatística entre as variáveis, mas não necessariamente uma relação de causa e efeito.

Outros fatores não analisados neste estudo podem influenciar a incidência de CVLI's na região, e devem ser considerados em pesquisas futuras. Uma possível linha de pesquisa seria investigar mais a fundo os fatores socioeconômicos e demográficos que podem influenciar a incidência de crimes violentos letais intencionais, além de investigar questões como faixa etária dos envolvidos, raça e gênero. Essa análise mais detalhada permitiria uma compreensão mais abrangente dos determinantes da criminalidade, possibilitando a criação de políticas públicas mais efetivas para prevenção e redução dos crimes.

Dentre as limitações, pode-se ressaltar a falta de acesso a dados mais detalhados sobre as ocorrências criminais. Além disso, os resultados obtidos foram baseados em dados históricos e não em tempo real, o que pode não refletir a situação atual da criminalidade no estado.

Por fim, acredita-se que este trabalho contribui para a compreensão mais profunda da situação da criminalidade no Rio Grande do Norte, espera-se que esses resultados possam embasar a tomada de decisões estratégicas no âmbito da segurança pública no estado. Com base nos resultados obtidos, é possível considerar uma série de medidas e ações que podem contribuir para a redução da criminalidade e a promoção de um ambiente mais seguro e pacífico.

Portanto, os resultados deste trabalho podem subsidiar decisões estratégicas, tais como a alocação de recursos, o desenvolvimento de políticas preventivas, intervenções específicas e o monitoramento contínuo da criminalidade. A aplicação prática dessas decisões tem o potencial de promover um ambiente mais seguro e pacífico e todo o estado do Rio Grande do Norte.

REFERÊNCIAS

ALVAZZI DEL FRATE, Anna. **A reflection on women, girls and armed violence.** Springer International Publishing, 2016.

ALVAZZI DEL FRATE, Anna. **Crime and development in the post-2015 development agenda.** United Nations Office on Drugs and Crime, 2013.

ALVES, Elvira Maria Farias. **Violência e escolarização: uma análise da percepção de jovens estudantes de uma escola pública da cidade de João Pessoa - PB.** Revista Interdisciplinar de Estudos em Educação, Araraquara, v. 1, n. 2, pág. 45-61, 2015.

ASMANN, P., & O'REILLY, E. (2020). **Insight Crime Balance of Homicides in 2019.** *InSight Crime*, 8. <https://insightcrime.org/news/analysis/insight-crime-2019-homicide-round-up/>. Acesso em 24 de janeiro de 2023

ARTHUR, David; VASSILVITSKII, Sergei. **K-means++ the advantages of careful seeding.** In: **Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms.** 2007. p. 1027-1035.

BEZDEK, J. C. **Pattern recognition with fuzzy objective function algorithms.** Plenum Press, 1981.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. **Latent Dirichlet Allocation.** Journal of Machine Learning Research, v. 3, p. 993-1022, 2003.

BOBA, Rachel. **Crime analysis and crime mapping.** United States: Sage Publications, 2005.

BORBA, Bruno Ferreira da Costa et al. **Otimização de Locais de Instalações Policiais com Base na Análise de Cluster e no Problema de Localização de Cobertura Máxima.** Inovação de Sistemas Aplicados, v. 5, n. 4, pág. 74, 2022.

BRANTINGHAM, Paul J; BRANTINGHAM, Patricia L. **Environmental Criminology.** 2. ed. United States: Waveland Pr Inc, 1991.

BRANTINGHAM, Paul; BRANTINGHAM, Patricia. 5. **Crime pattern theory.** *Environmental criminology and crime analysis*, p. 78, 2008.

BRANTINGHAM, P. L.; BRANTINGHAM, P. J. Nodes, Paths, and Edges: Considerations on the Complexity of Crime and the Physical Environment (1993). In: **Classics in Environmental Criminology.** Routledge, 2010. p. 289-326.

BRASIL. **Constituição** (1988). **Constituição** da República Federativa do Brasil. Brasília, DF: Senado Federal: Centro Gráfico, 1988.

BRASIL. Lei nº 7209, de 11 de julho de 1984. Altera dispositivos do Decreto-Lei nº 2848, de 7 de dezembro de 1940 – Código Penal, e dá outras providências.

BRASIL. Decreto-Lei 2.848, de 07 de dezembro de 1940. Código Penal. Diário Oficial da União, Rio de Janeiro, 31 dez. 1940.

BRASIL. Código Penal. Decreto-Lei nº 2.848, de 7 de dezembro de 1940. Brasília, DF: Senado Federal, 1940.

BRASIL. Decreto-Lei 3.914, de 09 de dezembro de 1941. Introdução ao Código Penal. Diário Oficial da União, Rio de Janeiro, 9 dez. 1941.

BRASIL. Lei nº 13.104, de 9 de março de 2015. Altera o art. 121 do Decreto-Lei nº 2.848, de 7 de dezembro de 1940 - Código Penal, para prever o feminicídio como circunstância qualificadora do crime de homicídio, e o art. 1º da Lei nº 8.072, de 25 de julho de 1990, para incluir o feminicídio no rol dos crimes hediondos. Diário Oficial da União, Brasília, DF, 10 mar. 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13104.htm. Acesso em: 19 de maio de 2023.

BURSZTEIN, E. et al. **Predictive Policing and Civil Liberties. In Oxford Handbook of Policing and Artificial Intelligence.** Oxford University Press, 2019.

CAPEZ, Fernando. **Curso de direito penal, volume 1, parte geral** – 23ª ed. – São Paulo: Saraiva Educação, 2010.

CAPEZ, Fernando. **Curso de direito penal** – parte geral, v. 1. 6ª ed. São Paulo: Saraiva, 2003.

CERQUEIRA, Daniel Ricardo de Castro et al. **Atlas da violência 2019.** 2019.

CERQUEIRA, Daniel Ricardo de Castro et al. **Atlas da violência 2021.** 2021.

CLARKE, Ronald V.; ECK, John E. Análise de crime para solucionadores de problemas em 60 pequenos passos. **Tradução em português Alessandro Souza Soares. Revisão Elenice De Souza**, v. 23, 2021.

COELHO, H.L.G.; ALVARENGA, A. T.; LIMA, L. C. **Acesso aos serviços de saúde e mortalidade por homicídios no Brasil.** Revista de Saúde Pública, São Paulo, v. 53, p. 70, 2019. doi: 10.11606/S1518-8787.2019053001067.

COHEN, L. E.; FELSON, M. **Social change and crime rate trends: a routine activity approach.** *American Sociological Review*, v. 44, 1979.

D. Cornish; R. Clarke. **The rational choice perspective.** Em R. Wortley; L. Mazerolle (eds) *Environmental Criminology and Crime Analysis.* Cullompton: Willan, United Kingdom, 21-47, 2008.

COPPIN, Ben. **Inteligência Artificial.**: Grupo GEN, 2010. *E-book.* ISBN 978-85-216-2936-8. Disponível em: <https://app.minhabiblioteca.com.br/#/books/978-85-216-2936-8/>. Acesso em: 08 fev. 2023.

ECKERSON, Wayne W. **PREDICTIVE ANALYTICS Extending the Value of Your Data Warehousing Investment.** United States: Tdwi Research, 2007.

ENDRASS, B.; HAYES, C. **Predictive Policing: Taking Stock and Moving Forward**. Policing: A Journal of Policy and Practice, v. 11, n. 1, p. 1-12, 2017.

ESTER, M.; KRIEGEL, H. P.; SANDER, J.; XU, X. **A density-based algorithm for discovering clusters in large spatial databases with noise**. In: KDD. Vol. 96, No. 34. 1996, p. 226-231.

FARIAS, A. M. G. de et al. **Definition of strategies for crime prevention and combat using fuzzy clustering and formal concept analysis**. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. World Scientific, v. 26, n. 03, p. 429–452, 2018.

FELSON, M.; CLARKE, R.V. **Opportunity Makes the Thief: Practical theory for crime prevention**. Police Research Series Paper 98. Londres: Home Office, 1998.

FIGUEIREDO, Ciro José Jardim de; MOTA, Caroline Maria de Miranda. **Um modelo de classificação para avaliar o nível de segurança em uma cidade baseado em GIS-MCDA**. Problemas Matemáticos em Engenharia, v. 2016, 2016.

FIGUEIREDO, C. J. J.; PEREIRA, D. V. S.; MOTA, C. M. M. **Avaliação de áreas em segurança pública: uma abordagem GIS-MCDA com sensoriamento remoto**. In: XLIX SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 2017, Blumenau. Anais do XLIX Simpósio Brasileiro de Pesquisa Operacional, 2017.

FIGUEIREDO, Ciro; MOTA, Carolina. **Preferências de aprendizagem em uma abordagem de decisão espacial com múltiplos critérios: uma aplicação no planejamento de segurança pública**. Jornal Internacional de Tecnologia da Informação e Tomada de Decisões , v. 18, n. 04, pág. 1403-1432, 2019.

FINLAY, Steven (2014). **Predictive Analytics, Data Mining and Big Data. Myths, Misconceptions and Methods** (1st ed.). Basingstoke: Palgrave Macmillan.

FORTES, Wanderson Alessandro; TASCA, Jorge Eduardo. **Uma análise sobre a aplicação de dados de business intelligence (BI) nas ações de prevenção situacional do crime**. Revista Ordem Pública, v. 7, n. 1, p. 159-177, 2014.

FOUNDATION, P. S. **What is Python? Executive Summary**. 2001. Disponível em: <https://www.python.org/doc/essays/blurp/> . Acesso em: fev. de 2023

FRALEY, C.; RAFTERY, A. E. **Model-based clustering, discriminant analysis, and density estimation**. Journal of the American Statistical Association, v. 97, n. 458, p. 611-631, 2002.

GANDOMI, A., & HAIDER, M. (2015). **Beyond the hype: Big data concepts, methods, and analytics**. International Journal of Information Management, 35 (2), 137-144.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007.

HARTUNG, Gabriel Chequer. **Ensaio em demografia e criminalidade**. 2009. Tese de Doutorado.

HASTIE, Trevor et al. **The elements of statistical learning: data mining, inference, and prediction**. New York: springer, 2009.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010: Aglomerados subnormais**. Rio de Janeiro, p.1-259, 2010.

IBGE. Instituto Brasileiro de Geografia e Estatística. (2006). **Censo**. Rio de Janeiro, RJ.

IBGE. Instituto Brasileiro de Geografia e Estatística. (2010). **Classificação e características ocupacionais: CBO 2002**. Rio de Janeiro: IBGE.

JAIN, Anil K.; DUBES, Richard C. **Algorithms for clustering data**. Prentice-Hall, Inc., 1988.

JOHNSON, R. A.; BHATTACHARYYA, G. **Statistics: principles and methods**. 6th ed. New York: John Wiley & Sons, 2010.

JUDD, Charles M.; MCCLELLAND, Gary H. **Data analysis: a model-comparison approach**. San Diego: Harcourt Brace Jovanovich, 1989.

KELLING, George L.; COLES, Catherine M. **Fixing broken windows: Restoring order and reducing crime in our communities**. Simon and Schuster, 1997.

KUHN, Max; SILGE, Julia. **Tidy Modeling with R**. " O'Reilly Media, Inc.", 2022.

KUHNS, Joseph B.; MAGUIRE, Kathleen. **Educational achievement and homicide rates**. Criminology, v. 17, n. 3, p. 305-324, 1979.

LIMA, R. S.; BUENO, S. **13 ° anuário brasileiro de segurança pública**. São Paulo: Fórum Brasileiro de Segurança Pública, 2019.

LOPES, Moisés José; SANTOS, William Cezar Sales dos; HOFFMANN, Maria Helena. **Conselhos Comunitários de segurança e programas de prevenção ao crime**, 2012.

HOFFMANN, Maria Helena; HAMMERSCHMIDT, Rodrigo (orgs.). **Segurança Pública: discurso permanente**. Florianópolis: DIOESC, 2012. p. 63-89.

MARTINS, Daisy Cristina Castilhos. **Os principais fatores que influenciam o crime no Brasil: uma análise estatística das variáveis**. 2010.

MALCZEWSKI, Jacek. **GIS-based multicriteria decision analysis: A survey of the literature**. *International Journal of Geographical Information Science*, v. 20, n. 7, p. 703–726, 2006.

MÁXIMO, Alexandre Alves et al. **A importância do mapeamento da criminalidade utilizando-se tecnologia de sistema de informação geográfica para auxiliar a segurança pública no combate à violência**. 2004.

MACQUEEN, J. **Classification and analysis of multivariate observations**. In: 5th Berkeley Symp. Math. Statist. Probability. Los Angeles LA USA: University of California, 1967. p. 281-297.

MENEGAT, Rualdo; ALMEIDA, Gerson. Sustentabilidade, democracia e gestão ambiental urbana. _____ (Org.). **Desenvolvimento Sustentável e Gestão Ambiental nas Cidades, Estratégias a partir de Porto Alegre**. Porto Alegre: UFRGS Editora, p. 171-194, 2004.

MIRABETE, Julio Fabbrini; FABBRINI, Renato N. **Manual de direito penal: parte geral, arts. 1º a 120 do CP**. São Paulo: Atlas, v. 1, 2010.

MOHLER, G. et al. **Randomized controlled field trials of predictive policing**. Journal of the American Statistical Association, v. 113, n. 522

MOLINA, Antonio García-Pablo de; GOMES, Luiz Flávio. **Criminologia**. 2.ed. São Paulo: RT, 1997

MOTA, Caroline Maria de Miranda; FIGUEIREDO, Ciro José Jardim de; PEREIRA, Débora Viana e Sousa. **Identifying areas vulnerable to homicide using multiple criteria analysis and spatial analysis**. ???: Omega, 2020.

NEPOMUCENO, Thyago Celso C.; COSTA, Ana Paula Cabral Seixas. **Visualização espacial em padrões de roubos desagregados**. Pesquisa Operacional, v. 19, p. 857-886, 2019.

NEPOMUCENO, Thyago CC; DARAIO, Cinzia; COSTA, Ana Paula CS. **Ranking multicritério para a avaliação eficiente e eficaz de departamentos de polícia**. Sustentabilidade, v. 13, n. 8, pág. 4251, 2021.

NETO, DIOGO DE FIGUEIREDO MOREIRA. **A segurança pública na Constituição**. O Alferes, v. 9, n. 28, 1991.

NG, Andrew; JORDAN, Michael; WEISS, Yair. **On spectral clustering: Analysis and an algorithm**. Advances in neural information processing systems, v. 14, 2001.

OBVIUM. Rio Grande do Norte: Obvio, v. 17, 21 jun. 2020. Disponível em: https://issuu.com/obvium/docs/obvium_especial_17_jun_2020. Acesso em: 28 fev. 2023.

Organização Mundial de Saúde. **Relatório Mundial de Violência e Saúde**. Genebra: OMS, 2014. ONU - Organização das Nações Unidas.

OLIVEIRA, A. B.; GARCIA, J. S. **Análise estatística de dados: teoria e prática**. São Paulo: Atlas, 2017.

PAUDEL, S., & JANG, J. (2020). **Optimal cluster identification using elbow and silhouette methods for k-means clustering**. Journal of Big Data, 7(1), 1-16. <https://doi.org/10.1186/s40537-020-00332-5>

PIERSON, Lillian. **Data science para leigos**. Alta Books Editora, 2019.

PYDATA. **Python data analysis toolkit**. 2008. Disponível em: <<https://pandas.pydata.org/pandas-docs/stable/>>. Acesso em: fev. 2023.

REIS, Edna Afonso; REIS, Ilka Afonso. **Análise Descritiva de Dados**. Minas Gerais, 2002.

REVATHY, K., & SUJATHA, S. S. (2018). **An efficient approach to clustering using k-means algorithm with Silhouette coefficient**. 2018 International Conference on Communication and Signal Processing (ICCSP), 1171-1175.

ROCHA, A. P. B. **Expansão urbana de Mossoró (período de 1980 a 2004): geografia dinâmica e reestruturação do território**. Natal, RN: EDUFRN Ed. Da UFRN, 2005.

RODRIGUES, Filipe Azevedo. **A “CRISE” DA VIOLÊNCIA E SUAS CAUSAS OU SOLUÇÕES: LEIS DE DROGAS E ARMAS COMO VETORES DA VIOLÊNCIA HOMICIDA**. 2019.

ROLIM, Marcos. **A Síndrome da Rainha Vermelha: policiamento e segurança pública no século XXI**. Rio de Janeiro: Jorge Zahar, 2006.

ROWE, S. D. **Leveraging the Three Stages of analytics. customer relationship management**, 20-23, 2017.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 3rd ed. Upper Saddle River, NJ: Pearson, 2016.

SANTANA, P. H.; SANTOS, S. M. **Atenção primária à saúde e taxas de mortalidade por homicídios em municípios brasileiros**. Revista de Saúde Pública, São Paulo, v. 52, p. 17, 2018. doi: 10.11606/S1518-8787.2018052000255.

SHAW, Clifford Robe; MCKAY, Henry Donald. **Juvenile delinquency and urban areas**. 1942.

SILVA, Fabrício M.; LENZ, Maikon L.; FREITAS, Pedro H C.; e outros **Inteligência artificial**. Grupo A, 2018. *E-book*. ISBN 9788595029392. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9788595029392/>. Acesso em: 08 fev. 2023.

SCIKIT-LEARN. Aprendizado de máquina em Python , Pedregosa *et al.* , JMLR 12, pp. 2825-2830, 2011.

SILVEIRA, Denise Tolfo; CÓRDOVA, Fernanda Peixoto. A pesquisa científica. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

SINGH, Neetu; BELLATHANDA KAVARAPPA, Chengappa; JOSHI, Jehan D. Mineração de dados para prevenção de crimes. In: **Conferência Internacional sobre Interface Humana e Gestão da Informação**. Springer, Cham, 2018.

TAVARES, Rita; MORAES, Claudia Leite; GOMES, Romeu. **Education and homicide: A systematic review of literature**. Ciência & Saúde Coletiva, v. 21, n. 4, p. 1061-1072, 2016.

TOPPIREDDY, Hitesh Kumar Reddy; SAINI, Bhavna; MAHAJAN, Ginika. Previsão do crime e estrutura de monitoramento com base na análise espacial. **Ciência da Computação Procedia**, v. 132, 2018.

VANDERPLAS, J. **Python Data Science Handbook**. O'Reilly Media, Inc., 2016.

VIAPIANA, Luiz Tadeu. **Economia do crime: uma explicação para a formação do crime**. Porto Alegre: AGE, 2006.

VICTORINO, Marcio et al. **Uma proposta de ecossistema de big data para a análise de dados abertos governamentais conectados**. Informação & sociedade, v. 27, n. 1, 2017.

VITO, Gennaro F.; MAAHS Jeffrey A.; HOLMES, Ronald M. **Criminology: theory, research and policy**. 2.ed. United States: Jones and Bartlett Publishers, 2007.

WASELFISZ, Julio Jacobo. **Mapa da violência 2013: mortes matadas por armas de fogo**. Centro Brasileiro de Estudos Latino-Americanos/Faculdade Latino-Americana de Ciências Sociais Brasil, 2013.

WANG, Dawei e cols. **Compreender a distribuição espacial do crime com base em suas variáveis relacionadas usando padrões discriminativos geoespaciais**. Computadores, Meio Ambiente e Sistemas Urbanos, v. 39, 2013.

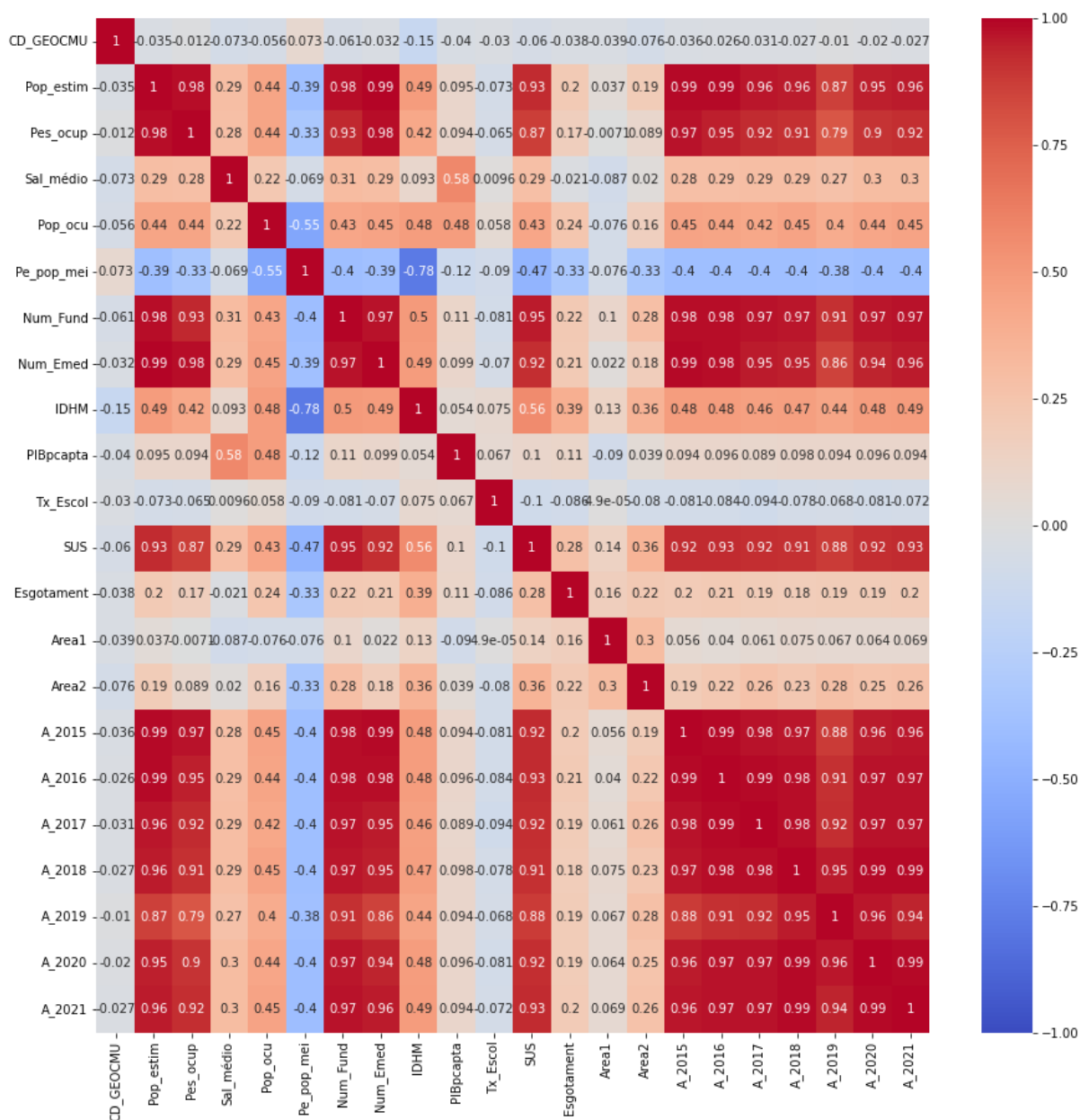
WERNECK, Guilherme L.; HASSELMANN, Maria H.; OLIVEIRA, Cecília D. **Desigualdade na mortalidade por homicídio entre diferentes grupos raciais/étnicos e níveis de escolaridade no Brasil, 2000-2010**. Journal of Public Health, Oxford, v. 39, n. 3, pág. 545-553, 2017.

WASKOM, M. et al. **Seaborn: statistical data visualization**. Journal of Open Source Software, v. 2, n. 16, p. 352, 2017.

WILSON, A. B. et al. **Examining the impact of mental illness and substance use on recidivism in a county jail**. International Journal of Law and Psychiatry, Amsterdam, v. 37, n. 4, p. 366-373, 2014. doi: 10.1016/j.ijlp.2014.02.031.

XIA, Belle Selene; GONG, Peng. Review of business intelligence through data analysis. **Benchmarking: An International Journal**, v. 21, n. 2, p. 300-311, 2014.

APÊNDICE A – Mapa de calor com coeficientes de correlação



Fonte: Autoria Própria

APÊNDICE B – Código fonte utilizado

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

# Ler dados
df = pd.read_csv(r'C:\Users\Thalia\Desktop\Material do TCC\DADOS\csv.thalia_dados.csv', sep=';')

# Remove as colunas não numéricas
df_numeric = df.select_dtypes(include=['float64', 'int64'])

# Padroniza os dados
scaler = StandardScaler()
X = scaler.fit_transform(df_numeric)

# Aplicando o método do cotovelo para encontrar o número ideal de clusters
inertia = []
k_range = range(1, 11)
for k in k_range:
    model = KMeans(n_clusters=k)
    model.fit(X)
    inertia.append(model.inertia_)

# Plotando os resultados
plt.plot(k_range, inertia, 'bx-')
plt.xlabel('Número de Cluster (k)')
plt.ylabel('Inércia')
plt.title('Elbow Method')
plt.show()

# Determinando o número ótimo de Cluster
import numpy as np
diff = np.diff(inertia)
diff_r = diff[1:] / diff[:-1]
k_opt = k_range[np.argmin(diff_r)+1]

print("O número ideal de clusters é:", k_opt)

from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_samples, silhouette_score

# Eliminando todas as colunas não numéricas
df = df.select_dtypes(include=np.number)

# Dimensionando os dados
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df)

# Executando o agrupamento de k-means para diferentes valores de k
k_values = range(2, 11)
silhouette_scores = []
for k in k_values:
    kmeans = KMeans(n_clusters=k, random_state=0)
    cluster_labels = kmeans.fit_predict(df_scaled)
    silhouette_avg = silhouette_score(df_scaled, cluster_labels)
    silhouette_scores.append(silhouette_avg)

# Plotando as pontuações da silhueta para cada valor de k
plt.plot(k_values, silhouette_scores)
plt.xlabel('Número de Cluster (k)')
plt.ylabel('Pontuação da silhueta')
plt.title('silhouette method')
plt.show()

# Determinando o número ideal de clusters baseado no método da silhueta
optimal_k = k_values[np.argmax(silhouette_scores)]
print("O número ideal de clusters é: ", optimal_k)

# Computando a matriz de correlação
corr_matrix = df.corr()

# Exibindo a atriz de correlação
print(corr_matrix)

import seaborn as sns

# Computando a matriz de correlação
corr_matrix = df.corr()

# Criando mapa de calor
plt.figure(figsize=(8, 8))
sns.heatmap(corr_matrix, cmap='coolwarm', vmin=-1, vmax=1)
```

```

# Computando a matriz de correlação
corr_matrix = df.corr()

# Criando mapa de calor
plt.figure(figsize=(15, 15))
sns.heatmap(corr_matrix, cmap='coolwarm', vmin=-1, vmax=1, annot=True)

# Computando a matriz de correlação
corr_matrix = df.corr()

# Filtrando a matriz de correlação para incluir apenas valores acima de 0,7
corr_matrix_07 = corr_matrix[corr_matrix.abs() > 0.7]

# Criando um mapa de calor com a matriz de correlação filtrada
plt.figure(figsize=(8, 8))
sns.heatmap(corr_matrix_07, cmap='coolwarm', vmin=-1, vmax=1, annot=True)

# realizando o agrupamento para cada par de variáveis e traçando gráfico de dispersão
for i in range(len(corr_vars)-1):
    for j in range(i+1, len(corr_vars)):
        col1 = corr_vars[i]
        col2 = corr_vars[j]
        X = df[[col1, col2]]
        kmeans = KMeans(n_clusters=2, random_state=42).fit(X)
        centroids = kmeans.cluster_centers_
        plt.scatter(X[col1], X[col2], c=kmeans.labels_, cmap='rainbow')
        plt.scatter(centroids[:,0], centroids[:,1], marker='x', color='black', s=200)
        plt.title(f'{col1} vs {col2}')
        plt.xlabel(col1)
        plt.ylabel(col2)
        plt.show()

```