



**UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA,
CONTABILIDADE**

PROGRAMA DE ECONOMIA PROFISSIONAL – PEP

AGUIMAR MENDONÇA NETO

**“A CRIMINALIDADE TOMOU CONTA DA CIDADE”:
O CINTURÃO DE INSEGURANÇA EM FORTALEZA, CEARÁ**

FORTALEZA

2024

AGUIMAR MENDONÇA NETO

“A CRIMINALIDADE TOMOU CONTA DA CIDADE”:
O CINTURÃO DE INSEGURANÇA EM FORTALEZA, CEARÁ

Dissertação submetida à Coordenação do Programa de Economia Profissional – PEP, da Universidade Federal do Ceará - UFC, como requisito parcial para a obtenção do grau de Mestre em Economia. Área de Concentração: Economia do Setor Público.

Orientador: Prof. Dr. Christiano Modesto
Penna

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- M494" Mendonça Neto, Aguiar.
"A criminalidade tomou conta da cidade": o cinturão de insegurança em Fortaleza, Ceará / Aguiar
Mendonça Neto. – 2024.
51 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de Economia, Administração,
Atuária e Contabilidade, Mestrado Profissional em Economia do Setor Público, Fortaleza, 2024.
Orientação: Prof. Dr. Christiano Modesto Penna.
1. Fortaleza. 2. Economia do Crime. 3. Clusterização. 4. Análise dos Componentes Principais. I. Título.
CDD 330
-

AGUIMAR MENDONÇA NETO

“A CRIMINALIDADE TOMOU CONTA DA CIDADE”:
O CINTURÃO DE INSEGURANÇA EM FORTALEZA, CEARÁ

Dissertação submetida à Coordenação do Programa de Economia Profissional – PEP, da Universidade Federal do Ceará - UFC, como requisito parcial para a obtenção do grau de Mestre em Economia. Área de Concentração: Economia do Setor Público.

Orientador: Prof. Dr. Christiano Modesto Penna

Aprovada em: 18/01/2024

BANCA EXAMINADORA

Prof. Dr. Christiano Modesto Penna
Universidade Federal do Ceará (UFC)

Prof. Dr. Marcos Renan Vasconcelos Magalhães
Secretaria de Saúde do Estado do Ceará (SESA/CE)

Prof(a). Dr(a). Isadora Gonçalves Costa Osterno
Secretaria de Planejamento e Gestão do Estado do Ceará (SEPLAG/CE)

AGRADECIMENTOS

Ao me debruçar sobre as páginas desta tese, reflito sobre a jornada que percorri para chegar até aqui, uma jornada que não teria sido possível sem o apoio inestimável de algumas pessoas especiais. Primeiramente, gostaria de expressar minha profunda gratidão à minha esposa Aglay e meus filhos Davi e Vinícius. Vocês foram a força silenciosa por trás de cada palavra escrita, cada análise realizada. As horas que passei longe, assistindo aulas e estudando, foram suportadas pela paciência e compreensão de vocês. Vocês sempre estiveram ao meu lado, não apenas fisicamente, mas em cada gesto de encorajamento e em cada sorriso compreensivo nos momentos de cansaço.

Aos meus pais, palavras não são suficientes para agradecer o suporte contínuo e a crença inabalável nas minhas capacidades. Vocês me ensinaram o valor da educação e do trabalho árduo, e foram a minha inspiração e o meu porto seguro em cada etapa deste caminho. Seu amor incondicional e apoio foram fundamentais para a minha formação, não apenas acadêmica, mas como ser humano.

Um agradecimento especial ao meu orientador e aos colegas Marcos Renan e Felipe Pinto, pelo apoio nos modelos e obtenção dos dados. A orientação de vocês, conhecimento e insights foram essenciais para a realização deste trabalho. Sua habilidade em me guiar, sem nunca tirar de mim a responsabilidade da aprendizagem e da descoberta, foi um presente inestimável. Sua paciência e sabedoria me encorajaram a perseguir a excelência e a profundidade em minha pesquisa.

Por fim, a todos que, direta ou indiretamente, contribuíram para este estudo, meu sincero agradecimento. Cada um de vocês foi parte fundamental desta jornada.

“Àquele que puder ser sábio, não lhe perdoamos que não o seja”.

“Se tens de servir a Deus com a tua inteligência, estudar para ti é uma obrigação grave”.

São Josemaria Escrivá

RESUMO

Esse trabalho investiga a criminalidade na cidade de Fortaleza, capital do Ceará, utilizando dados de boletins de ocorrências policiais para realizar uma análise de *clusterização*, que identificou a existência de uma espécie de cinturão de insegurança na cidade. O cinturão de insegurança é formado por bairros que apresentam altos índices de criminalidade, como homicídios, roubos e furtos. Os bairros que compõem o cinturão estão localizados em regiões periféricas da cidade, com baixa renda e infraestrutura precária. Adicionalmente foram realizadas Análises de Componentes Principais (ACP), de modo a reduzir a dimensionalidade dos dados. Esta identificou que os principais fatores que estão relacionadas à criminalidade em Fortaleza são a pobreza, desigualdade social e baixa educação. A criminalidade em Fortaleza é um problema complexo, que requer políticas públicas integradas para resolvê-lo. Essas políticas devem abordar os fatores socioeconômicos que contribuem para o crime, como a pobreza, a desigualdade social e a falta de oportunidades. Por fim, são sugeridas políticas públicas para reduzir a criminalidade, devendo estas abordarem os fatores socioeconômicos relacionados a este.

Palavras-chave: Fortaleza; Economia do Crime; Clusterização; Análise dos Componentes Principais.

JEL: C21; C38; O11.

ABSTRACT

This paper investigates crime in the city of Fortaleza, capital of Ceará, using data from police reports to perform a cluster analysis, which identified the existence of a kind of belt of insecurity in the city. The insecurity belt is formed by neighborhoods with high crime rates, such as homicides, robberies and thefts. The neighborhoods that make up the belt are located in peripheral regions of the city, with low income and precarious infrastructure. Additionally, Principal Component Analyzes (PCA) were performed in order to reduce the dimensionality of the data. This identified that the main factors that are related to crime in Fortaleza are poverty, social inequality and low education. Crime in Fortaleza is a complex problem that requires integrated public policies to solve it. These policies must address the socio-economic factors that contribute to crime, such as poverty, social inequality and lack of opportunities. Finally, public policies are suggested to reduce crime, which should address the socioeconomic factors related to it.

Keywords: Fortaleza; Economy of Crime; Clustering; Principal Component Analysis.

JEL: C21; C38; O11.

LISTA DE FIGURAS

Figura 1 - Distribuição do IDH de Renda segundo os bairros de Fortaleza.	18
Figura 2 - Distribuição de Roubos, Furtos, Homicídios e Apreensão de Armas de Fogo por bairros de Fortaleza, em 2021.	19
Figura 3 - Fluxo de Implementação da Estratégia Empírica.....	25
Figura 4 - Cinturão de insegurança em Fortaleza – formação dos clusters, a partir do k-means.	36
Figura 5 - CP1 × CP2 sobre índices de criminalidade por bairros de Fortaleza, em 2021.	38

LISTA DE TABELAS

Tabela 1 - Quartis do IDH de Renda, Longevidade e Educação e média de variáveis de delitos por bairros de Fortaleza, em 2021.	17
Tabela 2 - Ranking dos 10 bairros com maiores incidências em cada tipo de crime em ordem alfabética, em 2021.	21
Tabela 3 - Quantidades de Bairros por cluster e suas respectivas médias de IDH e de crimes.	34

SUMÁRIO

1 INTRODUÇÃO.....	10
2 CRIMINALIDADE EM FORTALEZA E BASE DE DADOS.....	15
3 METODOLOGIA.....	23
3.1 ESTRATÉGIA EMPÍRICA.....	23
3.2 ALGORITMO DE CLUSTERIZAÇÃO <i>k-MEANS</i>.....	25
3.3 ANÁLISE DOS COMPONENTES PRINCIPAIS (PCA).....	29
4 ANÁLISE E DISCUSSÃO DOS RESULTADOS.....	33
4.1 ALGORITMO DE CLUSTERIZAÇÃO <i>k-MEANS</i>.....	33
4.2 ANÁLISE DOS COMPONENTES PRINCIPAIS (PCA).....	36
5 CONSIDERAÇÕES FINAIS E PROPOSTAS DE POLÍTICAS PÚBLICAS.....	40
APÊNDICE A – QUADRO DOS BAIRROS PERTENCENTES A CADA CLUSTER.....	42
APÊNDICE B – GRÁFICO COM O RESULTADO DO MÉTODO DA SILHUETA E NÚMERO ÓTIMO DE CLUSTERS.....	43
APÊNDICE C – GRÁFICO COM O RESULTADO DA DISSIMILARIDADE E NÚMERO ÓTIMO DE CLUSTERS (MÉTODO DE ELBOW).....	44
APÊNDICE D – GRÁFICO COM O RESULTADO DO MÉTODO DO GAP ESTATÍSTICO E PARA O CÁLCULO DO NÚMERO ÓTIMO DE CLUSTERS.....	45
REFERÊNCIAS.....	46

1 INTRODUÇÃO

O 15º Anuário Brasileiro de Segurança Pública revelou que, em 2020, o Ceará teve a maior taxa proporcional de Mortes Violentas Intencionais (MVIs) do Brasil, com uma taxa de 45,2 por 100 mil habitantes. Além disso, o Ceará registrou a maior taxa de assassinatos de crianças no país e teve um aumento significativo no número de assassinatos de 2019 para 2020 (FBSP, 2022). O aumento nos homicídios foi atribuído a uma série de fatores, incluindo um motim da Polícia Militar, que alterou a dinâmica local da criminalidade e desarranjou as políticas públicas que estavam em curso (O Povo, 2021).

Fortaleza, capital do estado do Ceará, é uma cidade que atrai tanto visitantes nacionais quanto internacionais, com um movimento intenso de passageiros pelo sistema aeroportuário, totalizando 2.268.613 passageiros, e pelo sistema rodoviário, com 2.267.266 passageiros, em 2022. Além disso, entre os municípios, Fortaleza é o maior Produto Interno Bruto (PIB) do Nordeste e o 11º maior do Brasil, com R\$ 73,43 bilhões em 2021, representando 37,68% do PIB do Ceará no mesmo ano (Ibge, 2023). Contudo, além de seu potencial turístico e econômico, a cidade enfrenta desafios significativos relacionados à criminalidade, sendo considerada a quinta cidade mais violenta do Brasil em termos de homicídios dolosos por 100 mil habitantes, de acordo com o Anuário Brasileiro de Segurança Pública de 2022, além de registrar altas taxas de roubo, furto e tráfico de drogas (FBSP, 2022).

É importante salientar que, nos anos mais recentes, Fortaleza e o Ceará como um todo têm registrado uma tendência de redução nos homicídios (O Povo, 2023). O esforço conjunto das forças policiais, aliado ao aumento das apreensões de armas de fogo, tem contribuído para a redução da criminalidade e para a melhoria dos indicadores de segurança na região. Tal melhoria é atribuída às operações de combate à criminalidade realizadas pelas Forças de Segurança do Estado, que incluem estratégias e investimentos em segurança pública, trabalho integrado das forças de segurança e uma operação focada em coibir Crimes Violentos Letais e Intencionais (CVLIs) e Crimes Violentos Contra o Patrimônio (CVPs) (O Povo, 2023).

Uma das principais preocupações na atual situação de criminalidade em Fortaleza é a presença de facções criminosas em disputa por territórios, resultando em confrontos violentos, incluindo tiroteios e assassinatos, que ameaçam a segurança local. Essas facções surgiram no Ceará devido a transformações sociais nas dinâmicas do crime em cidades como Fortaleza e em sua zona metropolitana (Paiva, 2019; Melo; Paiva, 2021; Moraes; Siqueira, 2019). Até 2015, mencionar facções no Ceará não implicava na observação cotidiana acerca de

prisões e periferias, como ocorreu após o surgimento dos Guardiões do Estado (GDE), em 2016. Grandes facções como o Primeiro Comando da Capital (PCC) e o Comando Vermelho (CV) também passaram a ter papéis significativos e consolidaram suas posições, com o PCC aliado à GDE e o CV como principal rival do grupo cearense. A Família do Norte (FDN) desempenhou um papel importante, fortalecendo o CV e tornando-se sua principal aliada nos confrontos com a GDE e o PCC. Desde meados de 2016, o Ceará é palco de uma disputa complexa entre facções pelo controle de prisões e territórios periféricos, com GDE e CV desempenhando papéis de destaque nesse cenário repleto de conflitos, alianças e dissidências. Essa dinâmica complexa das facções aumenta os desafios da segurança pública em Fortaleza e na região (Paiva, 2019; Melo; Paiva, 2021; Moraes; Siqueira, 2019).

A incidência da criminalidade causa um impacto significativo na vida dos residentes da cidade, resultando em alterações nas suas interações sociais devido ao sentimento de insegurança que surge da possibilidade ou experiência de um assalto, por exemplo (Costa, 2018). Esse aspecto é crucial, pois a segurança desempenha um papel fundamental na estabilidade social. Um ambiente seguro garante que as pessoas possam satisfazer suas necessidades mais básicas, como morar em segurança e circular livremente na área urbana (Ceccato; Nalla, 2020).

Para combater a criminalidade, diversas intervenções e políticas de segurança têm sido implementadas pelas autoridades e órgãos competentes. Essas ações visam dismantelar as facções criminosas, intensificar o policiamento em áreas de risco e promover programas de prevenção e reabilitação para jovens em situação de vulnerabilidade (FBSP, 2022).

A atual situação da criminalidade em Fortaleza constitui um desafio complexo e preocupante que requer a cooperação e coordenação de diversas esferas da sociedade. Este contexto enfatiza a importância dos arranjos institucionais na implementação de políticas públicas efetivas, destacando a necessidade de interação entre diferentes agentes, resolução de conflitos, utilização de instrumentos e tecnologias que moldam as relações na execução dessas políticas (Longuinho, 2022). Além disso, a violência urbana, os conflitos entre facções e os desafios sociais e econômicos são questões intrinsecamente ligadas que exigem abordagens holísticas. Portanto, a eficácia das medidas de segurança pública deve estar alinhada com políticas sociais inclusivas e investimentos em educação e infraestrutura.

A criminalidade não se limita apenas a afetar a propriedade e a vida das pessoas, mas também tem repercussões nas atividades de outros atores econômicos, como as empresas, e nas variáveis macroeconômicas. As firmas que constituem o cenário de negócios são impactadas de diversas formas pela criminalidade: algumas são excluídas da atividade

econômica, outras enfrentam perdas de produtividade, e todas lidam com um ambiente de incerteza causado pela violência. Além disso, a criminalidade resulta em interrupções na cadeia produtiva e reduz a qualidade do capital humano disponível (Goldberg; Kim; Ariano, 2014). Adicionalmente, a violência tem um forte efeito negativo nos níveis de investimento e no crescimento econômico, conforme indicam estudos empíricos consistentes apresentados por Loureiro e Silva (2010).

Um outro fenômeno que dialoga com a criminalidade são as atividades informais. A economia informal e as altas taxas de criminalidade em cidades grandes estão interligadas de forma complexa, refletindo aspectos socioeconômicos mais amplos. A falta de empregos formais muitas vezes empurra indivíduos para a economia informal, que, embora não seja intrinsecamente criminosa, pode coexistir com atividades ilegais devido à ausência de regulação e fiscalização. Essa informalidade, frequentemente resultado da desigualdade social e da exclusão de oportunidades econômicas formais, pode levar a um sentimento de alienação e a uma percepção de atividades criminosas como alternativa viável, especialmente em áreas carentes de investimento público e de uma relação construtiva entre a comunidade e as forças de segurança. Assim, a relação entre economia informal e criminalidade não apenas reflete a falta de oportunidades formais de emprego, mas também as falhas no investimento em infraestrutura e serviços públicos, na regulação efetiva de atividades econômicas, e na inclusão social e econômica das populações mais vulneráveis. Importante salientar que o setor informal é um setor que demanda menos capital físico e mão de obra, além de operar com uma função de produção com menor produtividade, o que resulta em menor produção agregada em a economia (Arbex; Corrêa; Magalhães, 2022).

Ademais Santos, Vasconcelos e Moreira Junior (2020) buscaram estimar os custos socioeconômicos provenientes da violência e criminalidade para o estado do Ceará no período compreendido de 2010 a 2018, considerando tanto os custos reais do setor público, que envolvem as despesas empenhadas em segurança pública, sistema penitenciário e no sistema de saúde (referente à criminalidade), quanto os custos reais do setor privado, que envolvem as despesas com prêmios pagos às seguradoras, segurança privada, roubo de cargas e veículos, além da perda de produtividade com as mortes prematuras de vítimas de crimes violentos letais e intencionais (CVLI). Para os autores, os custos reais, sociais e econômicos advindos da violência e criminalidade (V&C) no Ceará cresceram de forma substancial entre 2010 e 2018, passando de R\$5,08 bilhões para R\$8,08 bilhões, um crescimento de aproximadamente 59,14% entre o período analisado. Isso equivale a um incremento real médio de cerca 6,61% ao ano.

A luta contra o crime é uma tarefa complexa e requer esforços de diversas áreas, especialmente em grandes cidades com altos índices de violência. De fato, problemas como criminalidade e segurança pública não podem ser adequadamente compreendidos apenas por meio da perspectiva de uma única disciplina, como a criminologia (Ceccato, 2016). Em vez disso, uma abordagem mais abrangente que integre contribuições de várias ciências, como sociologia, geografia e economia, é necessária para abordar essas questões. Na ciência econômica, os estudos sobre criminalidade ganharam impulso com o trabalho seminal de Becker (1968). Nessa análise microeconômica, o autor postulou que as pessoas se envolvem em atividades criminosas somente se perceberem que a utilidade esperada dessas ações supera aquela que poderiam obter em outras atividades com o mesmo tempo e recursos. O objetivo do modelo era determinar políticas públicas e privadas ótimas para combater a criminalidade.

Com o avanço dos estudos sobre crime na economia, as pesquisas começaram a incorporar com frequência a dimensão espacial em suas análises (Anselin *et al.* 2000; Andresen, 2006). Esse enfoque foi adotado em várias áreas da economia, não apenas naquelas em que o espaço é um fator fundamental, como economia regional, mas também em campos tradicionais como organização industrial e economia do setor público (Scoot, 1996; Balaguer-Coll *et al.*, 2019). Isso significa que os economistas passaram a utilizar modelos para lidar com problemas relacionados a dados espaciais.

Apesar das informações estatísticas disponibilizadas em documentos oficiais das autoridades municipais e estaduais serem acessíveis e úteis para analisar a ocorrência de roubos e furtos a pedestres, é essencial avançar em técnicas de análise mais robustas, que permitam identificar padrões semelhantes entre as observações das variáveis, visando uma maior uniformidade nos dados (Favero; Belfiore, 2021).

Além disso, é importante destacar que existem lacunas nos estudos que consideram uma abordagem mais localizada sobre as ocorrências de violência urbana. É necessário fornecer subsídios e diagnósticos para a formulação de políticas públicas mais específicas e direcionadas a microrregiões dentro da cidade, como a de Fortaleza, por exemplo. Santos (2019) enfatiza que é preciso ir além das políticas de combate ao crime e à violência em âmbito federal e estadual, voltando a atenção também para o nível dos bairros.

A questão central abordada por esta pesquisa visa investigar, com base em uma perspectiva que considera fatores sociais e espaciais (Costa, 2005), a distribuição da violência urbana, exemplificada pelos homicídios, roubos e furtos de pertences, além da apreensão de drogas e de armas de fogo ao longo do território de Fortaleza em 2021, que é o corte temporal mais recente da base de dados.

Este estudo adota uma abordagem exploratória, focando na identificação de agrupamentos específicos de bairros, organizados em conglomerados (*clusters*), que compartilham padrões semelhantes em relação aos roubos de pertences. As contribuições desse trabalho abrangem: 1) o fornecimento de informações fundamentais para futuras análises probabilísticas, abrindo caminho para o direcionamento de políticas públicas customizadas e bem ajustadas a cada conjunto de bairros, de acordo com suas características criminais distintas; 2) a possibilidade de se estabelecer uma abordagem alternativa para a análise de dados criminais, ao considerar as semelhanças e diferenças entre os grupos de observações nas diferentes microrregiões municipais, utilizando a técnica de agrupamento não supervisionado, também conhecida como *clustering*; e 3) identificação dos bairros mais violentos da cidade a partir da Análise dos Componentes Principais (do inglês *Principal Component Analysis* - PCA) que reduz a dimensionalidade do conjunto de dados e preserva a maior parte das informações relevantes contidas nas variáveis.

Como hipótese teórica, compreende-se que os bairros de maior renda per capita na cidade de Fortaleza, apesar de sua proximidade com atrações turísticas, como praias, e, conseqüentemente, maior tráfego de pessoas, não apresentaram uma predominância numérica na uniformidade interna de relatos de delitos, independentemente do cenário real. Abordando essa questão por uma perspectiva distinta, Feitoza e Costa (2019) indicam que áreas com alta visibilidade turística não necessariamente se sobrepõem àquelas onde a atividade criminosa é mais intensa. De acordo com esses pesquisadores, tal situação ocorre porque, em muitos casos, as políticas públicas municipais favorecem espaços específicos que são considerados de interesse para o turismo, tornando-os frequentemente alvos de investimentos e iniciativas de segurança pública, o que reduz o risco de incidentes criminosos (Feitoza; Costa, 2019). Nesse sentido, espera-se identificar em Fortaleza um espécime de “cinturão de insegurança”: os bairros mais nobres da cidade são circundados pelos mais violentos.

O estudo está estruturado da seguinte maneira: além desta introdução, a seção 2 discute brevemente a literatura empírica e fatos estilizados da criminalidade local em Fortaleza, analisando a relação de causalidade entre a criminalidade (representada aqui pelos crimes de homicídios, roubos, furtos, apreensões de drogas e entorpecentes, e apreensões de armas de fogo) e os fatores sociais e econômicos (IDH). Na seção 3, são abordados o algoritmo de clusterização *k-means*, os procedimentos para determinar o número ótimo de clusters, a análise dos componentes principais (PCA) e o detalhamento da base de dados empregada. Na seção 4, os resultados são apresentados, e em seguida, as conclusões e propostas políticas são discutidas na seção 5.

2 CRIMINALIDADE EM FORTALEZA E BASE DE DADOS

A violência nas áreas urbanas da cidade de Fortaleza tornou-se algo comum na vida diária dos seus cidadãos. Esta seção se destina a apresentar e explorar as bases de dados que serão utilizadas pelas metodologias de agrupamento não supervisionado e da análise da PCA. A **Tabela 1** a seguir versa sobre variáveis relacionados ao crime (homicídios, roubos, furtos e apreensão de drogas e de armas de fogo) no ano de 2021 por bairros, além de trazer informações do Índice de Desenvolvimento Humano (IDH) segunda as dimensões/categorias¹, elaborado pela Secretaria Municipal de Desenvolvimento Econômico de Fortaleza (SDE/PMF, 2015) com base nos dados do Censo de 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE).

Com relação aos dados de criminalidade estes foi obtida junto à Secretaria de Segurança Pública e Defesa Social do Estado do Ceará (SSPDS/CE) que divulga mensalmente informações oficiais consolidadas sobre diversos tipos de crime, classificados por Áreas Integradas de Segurança (AIS). Neste estudo foi utilizado um painel dos 113 bairros híbridos², contendo informações referentes aos crimes de roubos³, furtos, apreensões de drogas/entorpecentes⁴ e apreensões de armas de fogo registrados oficialmente nos limites territoriais da capital de Fortaleza para o período de 2021. Por outro lado, as informações de óbitos por homicídios foram extraídas do Sistema de Informação sobre Mortalidade – SIM/DATASUS da Prefeitura Municipal de Fortaleza (PMF).

Os dados foram organizados para os quartis do IDH-Renda dos 113 bairros da capital. Bairros com IDH-Renda mais baixo (Q1) têm uma incidência significativamente maior de homicídios, roubos e apreensões de drogas e armas de fogo. Isso sugere uma correlação entre baixa renda e maior atividade criminal, corroborando a teoria de que áreas mais carentes podem estar mais propensas à criminalidade devido a fatores socioeconômicos desfavoráveis. A exceção é em relação aos furtos: percebe-se que a incidência de destes aumenta em bairros com maior renda *per capita*, tanto para os valores absolutos como para a taxa por 10 mil habitantes.

¹ Para mais detalhes da metodologia de cálculo ver em: Canal Desenvolvimento Econômico - Prefeitura Municipal de Fortaleza: <https://desenvolvimentoeconomico.fortaleza.ce.gov.br/sde/estudos-e-pesquisas.html>.

² Vale apenas ressaltar que as análises que serão realizadas, têm como foco uma abordagem que leva em consideração os transbordamentos espaciais entre os bairros, não o georreferenciamento dos delitos supracitados.

³ Crimes violentos contra o patrimônio (CVP) é a soma de todas as ocorrências de roubo, exceto o crime de roubo seguido de morte (latrocínio) dado que já é contabilizado na categoria de CVLI. Até meados de 2019 os crimes de CVP eram divididos em dois grupos: (i) CVP-1: roubo à pessoa, roubo de documentos e outros roubos que não estão incluídos no CVP-2. (ii) CVP 2: roubo a residência, roubo com restrição de liberdade da vítima, roubo de cargas e roubo de veículos.

⁴ A partir de 2019 houve mudanças na descrição da tabela para "Apreensão de derivados da Cannabis (Maconha e Haxixe)". (3) Inclui Maconha e Haxixe.

Tais evidências estão de acordo com os trabalhos de Beato Filho (1998), Beato Filho e Reis (2000) e Batella e Diniz (2010). Tais autores discutiram a relação entre desenvolvimento humano e criminalidade. Para eles, esperar-se-ia uma correlação mais significativa entre os crimes violentos contra o patrimônio e o nível de desenvolvimento econômico, medido pelo IDH municipal. Tal expectativa é explicada pelas relações que esta categoria de delito guarda com contextos espaciais marcados por elevados indicadores de desenvolvimento, como por exemplo, os bairros mais nobres ou com alta concentração comercial na cidade de Fortaleza. Estes ambientes, segundo os autores, são caracterizados pela coexistência de diversos fatores que contribuem para o fortalecimento dessa relação, tais como melhores condições econômicas, grandes concentrações populacionais e comerciais e enfraquecimento dos mecanismos de controle social, garantindo assim mais oportunidades ao ato criminoso.

Na mesma linha de raciocínio, a relação entre riqueza e crimes contra o patrimônio foi discutida por Beato (1998) e Batella e Diniz (2010). Os autores afirmam que, contraditoriamente ao proposto em inúmeros trabalhos, a explicação mais significativa para o crime não é o nível de pobreza, mas a riqueza. Ambientes mais prósperos – regiões mais nobres ou com uma elevada dinâmica econômica – são sinônimos de oportunidades para ação criminosa, uma vez que fornecem mais alvos viáveis e compensadores, além de enfraquecerem mecanismos tradicionais de controle social e vigilância.

Apesar de os dados estarem dispostos para os quartis do IDH-Renda por bairros, também se observam fenômenos similares para os IDH-Educação e IDH-Longevidade. Na medida em que o IDH-Educação aumenta nos quartis, os índices de homicídios, roubos e furtos diminuem. Isso está alinhado com a hipótese de que uma população mais educada pode estar mais consciente das consequências legais e sociais do crime, o que pode contribuir para a redução das taxas de criminalidade.

Por fim, observa-se uma possível relação inversa entre os índices de apreensão de drogas e armas de fogo e a renda *per capita*. Bairros com menor IDH-Renda apresentam maiores índices de apreensões de drogas e armas. Isso pode indicar que áreas com maior vulnerabilidade socioeconômica podem estar mais propensas ao tráfico de drogas e à posse ilegal de armas. Em resumo, a análise dos quartis do IDH-Renda e das variáveis relacionadas ao crime nos bairros de Fortaleza sugere que fatores socioeconômicos, como renda e educação, desempenham um papel importante na determinação das taxas de criminalidade.

Tabela 1 - Quartis do IDH de Renda, Longevidade e Educação e média de variáveis de delitos por bairros de Fortaleza, em 2021.

Quartis	IDH			Nº absolutos ⁽¹⁾					Taxa por 10 mil habitantes				
	Renda	Long.	Edu.	H	R	F	AD	AA	H	R	F	AD	AA
Q1	0,04	0,26	0,91	16	313	136	21	22	7	140	63,1	18	10
Q2	0,09	0,36	0,95	9	265	177	47	17	5	122	90	38	8,7
Q3	0,15	0,46	0,96	8	250	219	11	13	4	126	109	4,7	5,7
Q4	0,46	0,62	0,98	2	257	323	5,7	6,5	2	185	213	8,7	4,6

Fonte: IBGE/Censo (2010); SDE/PMF (2015); SSPDS/CE E (2021).

Nota: (1) H – Óbitos por Homicídios; R – Nº de ocorrências de Roubos; F – Nº de ocorrências de Furtos; AD – Massa apreendida de Drogas/Entorpecentes (kg); e AA – Nº de armas de fogo apreendidas.

Confirmando os postulados teóricos, os crimes contra a pessoa tendem a ser mais presentes em áreas economicamente mais deprimidas do estado, em detrimento dos crimes contra o patrimônio, que são mais recorrentes em regiões mais ricas, onde há um contexto de oportunidades para os autores dos delitos (Beato Filho, 1998; Beato Filho e Reis, 2000; Batella e Diniz, 2010).

A Figura 1, por sua vez, apresenta o IDH-Renda dos 113 bairros de Fortaleza. Os bairros com os maiores IDH-Renda estão principalmente localizados na região leste da cidade, indicando uma concentração de desenvolvimento econômico nessa área. Bairros como Meireles, Guararapes, Cocó e Aldeota apresentam os maiores índices, sugerindo uma maior renda *per capita* e padrões de vida mais elevados.

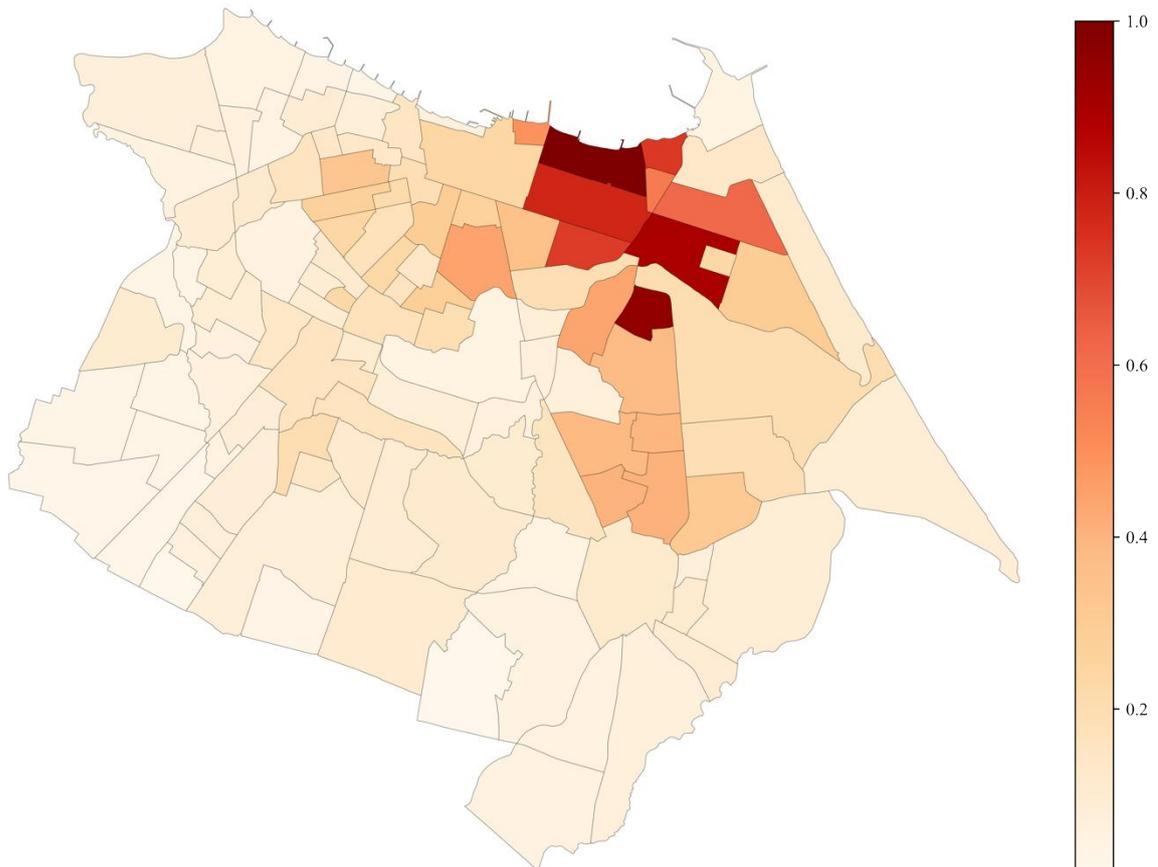
Notavelmente, os bairros com os maiores IDH-Renda parecem estar próximos entre si, formando uma espécie de aglomerado de áreas de maior desenvolvimento econômico na região leste da cidade. Isso pode ser um reflexo das oportunidades econômicas, acesso a serviços e qualidade de vida mais elevada nesses locais.

Nota-se também que as disparidades significativas entre as diferentes áreas da cidade para a variável em questão: as diferenças nos IDH-Renda podem ser atribuídas a uma série de fatores, incluindo acesso a empregos formais, educação de qualidade, infraestrutura e investimentos públicos e privado.

De acordo com o Jornal O Povo (2022), estudos conduzidos pelo Comitê de Prevenção e Combate à Violência destacam uma forte correlação negativa entre IDH e taxas de homicídios, evidenciando a relação entre condições socioeconômicas precárias e violência. Segundo a mesma matéria, apesar dos esforços contínuos para enfrentar essa realidade, a cidade enfrenta desafios persistentes, incluindo baixa responsabilização em casos de homicídio e um

foco predominante em policiamento ostensivo, o que levou a um encarceramento massivo direcionado principalmente a pequenos traficantes e ladrões, em vez de homicidas.

Figura 1 - Distribuição do IDH de Renda segundo os bairros de Fortaleza.



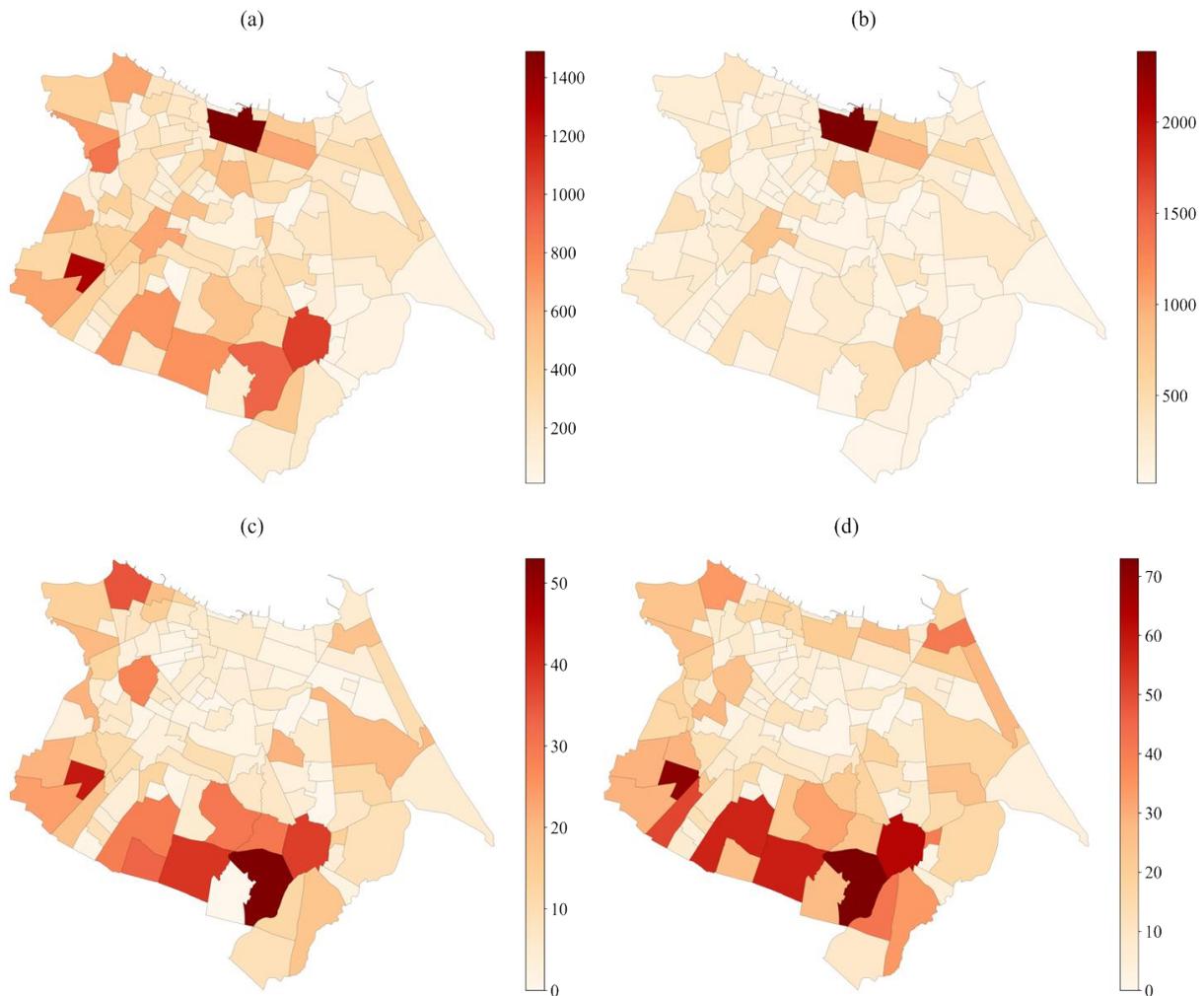
Fonte: SDE/PMF (2015) e IBGE/Censo (2010).

De outro lado, a Figura 2 apresenta os números de ocorrências de cada tipo de crime e suas respectivas quantidades por bairros na cidade de Fortaleza, em 2021. Os crimes considerados incluem roubos, furtos, homicídios e apreensões de armas de fogo, separados em cada um dos painéis.

Os números de roubos, furtos, homicídios e apreensões de armas de fogo variam significativamente entre os bairros, indicando uma distribuição desigual das atividades criminosas. Alguns bairros têm baixos números de crimes em todas as categorias consideradas, possivelmente indicando um ambiente mais seguro, enquanto outros parecem concentrar um número notável de diferentes tipos de crimes. Por exemplo, os bairros "Jangurussu" e "Bom

Jardim" apresentam altas incidências de roubos, furtos, homicídios e apreensões de armas de fogo.

Figura 2 - Distribuição de Roubos, Furtos, Homicídios e Apreensão de Armas de Fogo por bairros de Fortaleza, em 2021.



Fonte: SSPDS/CE (2021).

Nota: (a) Roubos; (b) Furtos; (c) Homicídios; (d) Apreensão de Armas de Fogo.

Bairros de todas as regiões têm sua parcela de atividades criminosas, de modo que não existe um padrão único de concentração de crimes em uma região específica da cidade. Notam-se padrões. Percebe-se uma maior concentração de homicídios e apreensão de armas de fogo na região sul/sudoeste da cidade, o que sugere potencial correlação entre esses diferentes tipos de crimes. Já em relação às ocorrências de roubos, parece que tais práticas “circundam” a região central da cidade, iniciando da região sul para a noroeste. Fugindo desse padrão, figura o bairro do Centro como o maior detentor de incidências de roubos, em 2021. Por fim, acerca

das ocorrências de furto, novamente o Centro aparece na primeira posição com mais de duas mil ocorrências, no mesmo ano. Os furtos não têm uma distribuição clara, mas vê-se bairros nobres entre os maiores detentores desses delitos, tais como Aldeota (2ª posição) e Meireles (7ª posição).

Segundo Santos (2020), duas coisas chamam a atenção com relação a estas composições de bairros quanto ao desempenho criminal. A primeira delas com relação à qualidade de vida, medido aqui pelo Índice de Desenvolvimento Humano (IDH), dos bairros que compõem o grupo de maior densidade. Conforme estudo da Secretaria Municipal de Desenvolvimento Econômico (SDE/PMF, 2015), os bairros da capital cearense considerados com IDH-Geral muito alto/alto e médio, são respectivamente: Meireles, Aldeota, Praia de Iracema, Fátima, Parquelândia, De Lourdes/Papicu e Centro. A maioria destes bairros, entre os anos 2010 e 2014, apresentaram uma média anual de 2.117 ocorrências de crimes contra o patrimônio, ou seja, os bairros mais nobres e medianos também podem ser considerados perigosos e suscetíveis às ações criminais. Por outro lado, os bairros com menor desempenho criminal composto pelos bairros Guajeru, Sabiaguaba, Coaçu, Floresta, Curió, Dendê e Parque Presidente Vargas, estão localizados numa região de periferia (baixo IDH), os mesmos concentraram apenas 0,48% das ocorrências de crimes contra a patrimônio na capital cearense entre 2010 e 2014.

De acordo com Santos (2020), nestas áreas periféricas, a polícia é constantemente desafiada pelas facções criminosas que atuam na cidade, principalmente no que se refere à ocupação dos condomínios residenciais de programas de imóveis populares como o “Minha Casa, Minha Vida”, onde as facções expulsam moradores de suas casas ou apartamentos bem como impõe o tribunal do crime “*Lei do vai pro saco*” caso algum morador cometa certos delitos, em especial, roubar ou furtar na região.

Ressalta-se que o mapa para as apreensões de drogas não foi apresentado por conta da preponderância do bairro Demócrito Rocha na dimensão, o que fez com que tal mapa não fornecesse subsídios visuais para identificação da heterogeneidade entre os bairros.

A Tabela 2, a seguir, a fim de complementar as informações fornecidas na Figura 2, apresenta em uma única lista os 10 bairros com maiores incidências em cada tipo de crime.⁵ Há uma mistura de bairros localizados em diferentes regiões da cidade (leste, oeste), sugerindo que as incidências de crimes não se concentram em uma única área. Pode-se observar que

⁵ São apresentados um total de 31 bairros na Tabela 2, por conta de mesmos bairros aparecem concomitantemente em mais de uma das cinco listas dos 10 bairros com maiores incidências criminosas por cada tipo.

alguns bairros têm altas incidências em múltiplos tipos de crimes, o que pode indicar a presença de atividades criminosas mais complexas e coordenadas.

Bairros como Jangurussu e Bom Jardim apresentam altas incidências em múltiplos tipos de crimes, incluindo roubos, homicídios e apreensão de armas de fogo. O primeiro, por exemplo, foi o que deteve o maior número de homicídios e apreensão de armas de fogo, em 2021.

Tabela 2 - Ranking dos 10 bairros com maiores incidências em cada tipo de crime em ordem alfabética, em 2021.

Bairros	Roubos	Furtos	Homicídios	Apreensão	
				Armas de Fogo	Drogas (kg)
Aldeota ²	641	958 ²	2	9	0,89
Ancuri	465	117	12	42	42,00
Antônio Bezerra	855	538	13	20	10,49
Barra do Ceará	654	370	35	34	9,54
Barroso	335	90	30	24	24,00
Bom Jardim ^{2, 2, 2}	1.320 ²	269	43 ²	70 ²	89,54
Canindezinho	397	157	18	51	51,00
Centro ^{1, 1}	1.489 ¹	2.388 ¹	6	21	5,89
Conjunto Ceará	611	427	3	16	12,96
Curió	33	26	14	40	40,00
Demócrito Rocha ¹	127	98	2	6	799,18 ¹
Fátima	551	768	4	3	1,75
Guajerú ³	1.320 ³	28	4	2	100,38
Henrique Jorge ³	403	266	7	28	157,52 ³
Jangurussu ^{1, 1}	932	398	53 ¹	73 ¹	123,26
Jardim Iracema	145	122	7	7	79,29
Lagoa Sapiranga (Coité) ²	139	128	12	25	176,35 ²
Meireles	449	663	2	26	2,21
Messejana ³	1.064	860 ³	38	63 ³	9,20
Mondubim	711	413	29	57	57,00
Papicu	290	513	4	21	10,36
Parangaba	644	789	3	7	4,69
Parque Iracema	33	41	4	2	54,14
Passaré	507	264	30	32	32,00
Paupina	175	100	17	34	125,04
Pici	250	86	28	25	25,00
Planalto Ayrton Senna	219	116	33	26	26,00
Prefeito José Walter	725	308	39 ³	58	7,58
Quintino Cunha	694	204	20	25	6,06
Siqueira	654	271	24	29	3,88
Vicente Pinzón	187	235	18	41	41,00

Fonte: Elaboração própria com base nos dados da SSPDS/CE (2021).

Nota: ¹ indica que o bairro tem a primeira posição em um tipo de crime, enquanto ² e ³ denotam as 2^a e 3^a posições, respectivamente.

De outro lado, bairros como Demócrito Rocha, Sapiranga/Coité e, novamente, Jangurussu têm incidências significativamente altas de apreensões de drogas, indicando possíveis atividades relacionadas ao tráfico. O bairro Centro, por sua vez, foi o que apresentou o maior número de roubos e furtos haja vista a dinâmica econômica e comercial da região. Não menos importante, é interessante citar que também bairros nobres, como Aldeota e Meireles, aparecem na tabela uma vez que possuem altos indícios de roubos e furtos, possivelmente por conta do maior poder aquisitivo e posse de bens de maior valor agregado da população que habita nessas localidades conforme em Beato Filho (1998), Beato Filho e Reis (2000), Batella e Diniz (2010).

Essa ideia de concentração geográfica do comportamento transgressor, conforme nas palavras de Guerra (2016), remonta a exemplos clássicos da literatura sociológica, como a teoria da estrutura social, que aborda a importância do ambiente urbano na incidência de crimes – maiores níveis de delinquência estão associados a comunidades compostas por uma população heterogênea, com baixo nível de renda e educação, pela presença maciça de atividades industriais e comerciais, entre outras características, como, por exemplo, é o caso dos bairros inseridos na Região Grande Aldeota (Aldeota, Dionísio Torres, Meireles, Papicu, Varjota e Praia de Iracema) e o bairro Centro.

3 METODOLOGIA

3.1 Estratégia empírica

A Análise dos Componentes Principais (PCA - *Principal Component Analysis*) e *k-means* são duas técnicas estatísticas amplamente usadas na análise de dados, mas servem a propósitos distintos e possuem metodologias diferentes. Tanto PCA quanto *k-means* podem ser utilizados para reduzir a complexidade dos dados. O PCA faz isso ao identificar as direções (componentes principais) que maximizam a variação nos dados, enquanto o *k-means* agrupa os dados em *k clusters*, reduzindo a quantidade de dados pela representação de cada grupo.

O PCA uma técnica de transformação linear que visa identificar as direções (componentes principais) ao longo das quais a variação nos dados é máxima. O objetivo principal é reduzir a dimensionalidade dos dados, preservando o máximo de informação (variância) possível (Jolliffe, 2002). Tal metodologia fornece os componentes principais, que são combinações lineares das variáveis originais. Os primeiros componentes principais retêm a maior parte da variabilidade dos dados.

De outro lado, o *k-means* um método de agrupamento que visa particionar os dados em *k* grupos distintos, de modo que os pontos de dados dentro de cada grupo estejam mais próximos uns dos outros do que dos pontos de dados em outros grupos (Bishop, 2006). O foco está em identificar agrupamentos ou padrões nos dados. Assim, o foco dessa metodologia está em fornecer centróides para cada *cluster*, que são pontos centrais, e rótulos para cada ponto de dados indicando a qual *cluster* o ponto pertence.

Uma vez que, ao destacar as direções de maior variância, o PCA pode ajudar a identificar padrões nos dados dos bairros, como características socioeconômicas, demográficas, de infraestrutura ou mesmo de aspectos ligados ao crime que diferenciam um bairro de outro; o *k-means* é eficaz na formação de *clusters* de bairros com características similares, facilitando a análise segmentada da cidade e permitindo intervenções direcionadas e políticas públicas mais eficazes.

Quando utilizados em conjunto, o PCA e o K-means demonstram uma sinergia poderosa (Mouravian Anaraki, et al., 2021). O método, denominado por Mouravian Anaraki, et al. (2021) como P-K, envolve a aplicação inicial do PCA para redução da dimensionalidade dos dados e remoção de ruídos. Os componentes principais resultantes, que capturam a maior parte da variância nos dados, são então utilizados como entrada para o K-means. Essa abordagem não apenas otimiza o processo de clusterização, tornando-o mais eficiente, mas

também aprimora a qualidade dos clusters formados, uma vez que o K-means é aplicado a um conjunto de dados simplificado e mais representativo.

Nesse contexto, diversos estudos abordam objetivos variados, como o trabalho de Behiry e Aly (2024), que propõem um modelo híbrido visando aprimorar a segurança de Redes de Sensores Sem Fio. Esse modelo utiliza a técnica de Análise de Componentes Principais (PCA) para a redução de dimensionalidade dos dados, juntamente com o algoritmo de agrupamento K-means para a extração de características das redes analisadas. Por outro lado, Marsha et al. (2024) investigaram o uso do sistema de imagem hiperespectral (HSI) para identificar, de maneira rápida e não destrutiva, a autenticidade e origem de três variedades de arroz iraniano. Eles empregaram técnicas como a Análise de Componentes Principais (PCA) para redução da dimensionalidade da base de dados, seguida da aplicação de algoritmos de agrupamento, especialmente o K-means. Os resultados destacaram a eficácia do PCA e do algoritmo K-means na discriminação visual das variedades de arroz, evidenciando a capacidade do HSI em autenticar e diferenciar as origens geográficas das amostras.

Por outro lado, é possível aplicar o método K-P, em que a aplicação do PCA ocorre após a clusterização. Isso proporciona insights mais profundos sobre a estrutura interna de cada cluster, permitindo uma interpretação mais rica e detalhada dos resultados da clusterização, conforme Mouravian Anaraki, et al (2021). O PCA auxilia na identificação das variáveis que mais contribuem para a variação dentro de cada cluster, sendo útil para compreender as características distintas de diferentes grupos de bairros e embasar decisões com base em evidências. Além disso, o PCA pode ser utilizado para visualizar dados de alta dimensão em um espaço de dimensões inferiores (geralmente 2D ou 3D), facilitando a interpretação visual dos clusters e de suas características intrínsecas.

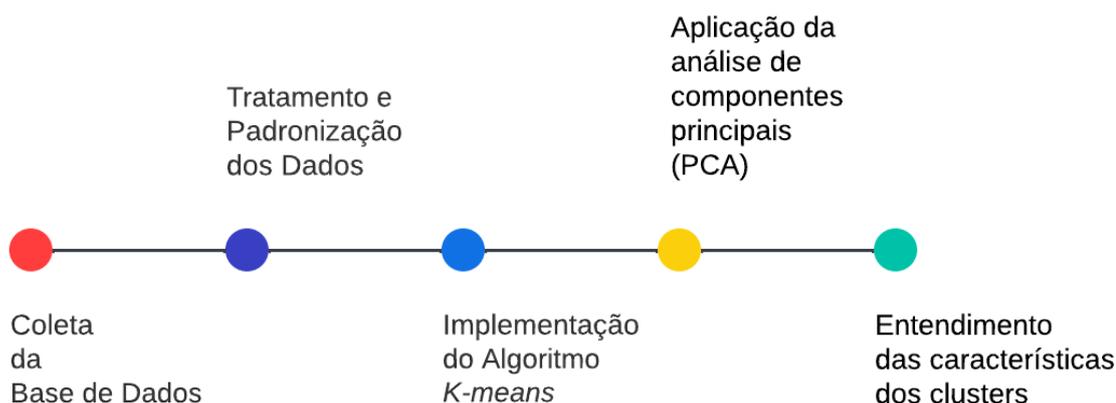
É crucial ressaltar que, assim como na abordagem PCA seguida por K-means, é necessário considerar a escala das variáveis tanto para o K-means quanto para o PCA. A normalização dos dados pode ser necessária para garantir que a análise seja justa e não seja distorcida por variáveis com maior amplitude. Embora a aplicação de PCA após K-means forneça insights detalhados, há também o risco de perder informações importantes sobre as relações entre os clusters, uma vez que o PCA é aplicado de forma independente em cada cluster.

Um exemplo dessa abordagem pode ser observado em Wu (2023), que explora a integração de dispositivos IoT e métodos de big data na gestão educacional universitária para aprimorar a eficácia e segurança dos dados. Inicialmente, o algoritmo K-means é utilizado para

agrupar os estudantes, seguido pela aplicação do PCA para obter uma compreensão abrangente dos dados coletados, investigando a relação entre visitas à biblioteca e notas finais.

Dessa forma, dada a quantidade relativamente baixa de variáveis, como observado no capítulo anterior, a implementação do algoritmo não apresentou desafios significativos devido à dimensionalidade reduzida da base de dados. Assim, optou-se pela segunda estratégia empírica para o presente estudo. Conforme desenvolvido por Mouravian Anaraki, et al (2021), o fluxo de implementação empírica do estudo (Figura 3) compreendeu: 1) a coleta de dados; 2) tratamento e padronização dos dados; 3) implementação do algoritmo de agrupamento K-means; 4) aplicação da análise de componentes principais (PCA); e 5) entendimento das características dos clusters.

Figura 3 - Fluxo de Implementação da Estratégia Empírica.



Fonte: Elaboração própria.

3.2 Algoritmo de clusterização *k-means*

O *clustering* é uma técnica sofisticada de mineração de dados multivariados que tem como propósito agrupar um conjunto de N observações provenientes de um banco de dados em k grupos distintos, conhecidos como *clusters*. Dentro de cada *cluster*, as observações possuem a máxima semelhança entre si, enquanto a similaridade entre as observações de *clusters* diferentes é minimizada. Os algoritmos de agrupamento são empregados para revelar padrões em dados espaciais, mas sua aplicação em larga escala requer uma análise minuciosa das diversas técnicas de *clusterização* existentes (Jain; Dubes, 1988). O objetivo primordial é

segmentar a amostra de dados originais em grupos consideravelmente homogêneos (Hair *et al.*, 2009).

No âmbito deste estudo, o algoritmo de agrupamento selecionado foi o *k*-means, empregando a métrica de distância euclidiana, amplamente utilizada na análise de agrupamentos (Favero; Belfiore, 2021). A essência do método *k*-means reside em estabelecer *clusters* de modo a minimizar a variação total intragrupo (*W*), também conhecida como variação interna do *cluster*. Diversos algoritmos *k*-means estão disponíveis, sendo o algoritmo padrão o Hartigan-Wong (1979). Neste algoritmo, a variação total interna do *cluster* é definida como a soma dos quadrados (SS) das distâncias euclidianas entre as observações e o centróide correspondente, conforme demonstrado na equação (1):

$$\sum_{k=1}^K W(C_k) = \sum_{x \in C_k} (x_i - \mu_k)^2 \quad (1)$$

onde, x_i representa uma dada variável da unidade de análise i pertencente ao *cluster* C_k ; μ_k é a média de uma dada variável x pertencente ao *cluster* C_k ; e K é o total de *clusters*.

Cada observação (x_i) é atribuída a um *cluster* específico de forma que a soma das distâncias ao quadrado da observação até os centros de *cluster* atribuídos μ_k seja mínima. Define-se a variação total dentro do *cluster* da seguinte forma:

$$\sum_{k=1}^K W(C_k) = \sum_{k=1}^K \sum_{x \in C_k} (x_i - \mu_k)^2 \quad (2)$$

Para determinar o número ótimo de *clusters* no algoritmo não supervisionado *k*-means, é necessário definir a quantidade de *clusters* desejada. A literatura oferece diversos métodos estatísticos para a determinação destes *clusters*, sendo os principais: o método do Cotovelo (Elbow), o método da Silhueta e o *Gap* Estatístico. Entre eles, o método do Cotovelo destaca-se como um dos mais populares e amplamente utilizado. Os três métodos foram efetuados para fins de comparação: o método do Cotovelo indicou a existência de três *clusters*, assim como o método da Silhueta. Apenas o método do *Gap* Estatístico indicou a existência de dois *clusters*. Os resultados dos três métodos são apresentados no Apêndice. Desta feita, conforme será apresentado, optou-se por três agrupamentos de bairros.

O método do Cotovelo envolve a realização repetida do algoritmo de *clusterização* para diferentes valores de k (número de clusters). Em seguida, é traçada uma representação gráfica da variância explicada ou da soma dos quadrados *intracluster* (ou do inglês *within-clusters sum-of-squares*, comumente abreviado para WCSS) em relação ao número de *clusters*. O objetivo é identificar o ponto na curva onde ocorre uma mudança abrupta, semelhante a um

“cotovelo”. Esse ponto sugere o número ótimo de *clusters*, pois a adição de mais *clusters* após esse ponto não resulta em uma diminuição substancial do WCSS. A expressão do WCSS é dada a seguir:

$$WCSS = \sum d(P_1, C_1) + \sum d(P_2, C_2) + \sum d(P_3, C_3) + \dots + \sum d(P_i, C_k) \quad (3)$$

onde C_k representa o centro de cada *cluster* k , enquanto P_i representa os elementos que estão agrupados juntos dentro de um *cluster* específico k .

O método da silhueta, introduzido por Rousseeuw (1987), avalia a adequação de cada unidade de análise ao seu respectivo *cluster*. Uma alta largura de silhueta média indica um agrupamento sólido. O cálculo da silhueta média é realizado para várias observações e diferentes valores de k (número de *clusters*). O número ótimo de *clusters* k é aquele que maximiza a silhueta média dentro de uma faixa de valores possíveis para i .

Conforme destacado por Batool e Henning (2021), a utilização da silhueta média é recomendada devido à sua capacidade de fornecer uma medida objetiva, intuitiva e resistente a valores atípicos. O cálculo dessa métrica é realizado inicialmente ao calcular o coeficiente de silhueta, $s(i)$, para cada elemento i , conforme a equação a seguir:

$$s(i) = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (4)$$

Nesta equação, a_i representa a média das distâncias entre o elemento i e todos os outros elementos no mesmo *cluster*, enquanto b_i é a menor média das distâncias entre o elemento i e os elementos de qualquer outro *cluster*, distintos do próprio *cluster* do elemento i . O coeficiente de silhueta $s(i)$ varia de -1 a 1 , onde valores próximos de 1 indicam que a observação está bem alocada no seu *cluster* e distante de outros *clusters*, valores próximos de 0 indicam sobreposição ou ambiguidade, e valores próximos de -1 indicam que a observação foi alocada ao *cluster* errado. Assim, o método da silhueta oferece uma abordagem analítica para determinar o número ideal de *clusters*, considerando a adequação das unidades de análise dentro de seus *clusters* e o grau de separação entre os *clusters*.

Por último, o método do *Gap* Estatístico, proposto por Tibshirani *et al.* (2001), baseia-se na ideia fundamental de que, se as unidades amostrais i não formassem agrupamentos, sua distribuição se assemelharia à distribuição uniforme dentro das amplitudes das r variáveis observadas. A intuição subjacente a esse método é que, se o número escolhido de *clusters* for apropriado, a estrutura dos dados reais será substancialmente mais proeminente do que a estrutura gerada aleatoriamente.

No entanto, é importante notar que a análise de agrupamentos não garante que as similaridades intra-*cluster* sejam sempre significativas. A técnica de agrupamento é mais adequada para diagnósticos pontuais, nos quais conhecimentos adicionais sobre o problema de negócios são aplicados durante a avaliação para orientar decisões (Fawcet; Provost, 2016).

Apesar de sua ampla aplicação e eficácia em diversos cenários, o método também apresenta algumas desvantagens, como destacado por Kaufman e Rousseeuw (1987) e Fawcet e Provost (2016):

1. Sensibilidade ao número de *clusters*: O algoritmo *k-means* requer a especificação antecipada do número de *clusters*. Contudo, determinar o número ideal de *clusters* pode ser desafiador em muitos casos, levando a agrupamentos subótimos caso *k* seja escolhido incorretamente.
2. Sensibilidade à inicialização dos centróides: O resultado do *k-means* pode variar de acordo com a escolha inicial dos centróides. A convergência para um mínimo local inadequado pode ocorrer se os centróides iniciais forem escolhidos de forma desfavorável. Técnicas de inicialização mais avançadas, como o *k-means++*, ajudam a mitigar esse problema.
3. Limitações em relação a formas complexas de *clusters*: o *k-means* assume *clusters* com formas esféricas e tamanhos iguais, o que pode ser limitante em dados com *clusters* de formas complexas, tamanhos variados ou sobreposições.
4. Sensibilidade a *outliers*: O *k-means* é sensível a *outliers*, ou seja, pontos de dados distantes dos centróides dos *clusters*. *Outliers* podem impactar negativamente a formação dos *clusters* e resultar em agrupamentos incorretos.
5. Incapacidade de lidar bem com diferentes densidades de *clusters*: O *k-means* assume densidades similares entre os *clusters*, o que pode ser problemático em dados com *clusters* de densidades bastante discrepantes. Além disso, o *k-means* não é capaz de identificar *clusters* aninhados dentro de outros *clusters*.
6. Dependência da métrica de distância: O *k-means* utiliza distâncias euclidianas para agrupar os dados. Essa abordagem pode não ser adequada para todos os tipos de dados, especialmente quando as características possuem escalas distintas ou não seguem uma distribuição gaussiana.
7. Falta de garantia de convergência para o ótimo global: Devido à natureza heurística do algoritmo, o *k-means* não garante a descoberta do ótimo global em todos os casos. Ele pode ficar preso em mínimos locais e oferecer uma solução subótima.

8. Dificuldade em lidar com dados de alta dimensionalidade: Com o aumento da dimensionalidade dos dados, as distâncias entre os pontos tendem a se igualar, o que pode resultar em agrupamentos menos significativos ou representativos.

3.3 Análise dos componentes principais (PCA)

A obtenção dos componentes principais envolve o processo de diagonalização de matrizes que são simétricas e semi-definidas positivas. A partir desse procedimento, é possível calcular os componentes principais de maneira direta, tornando-os aplicáveis em diversas áreas científicas variadas. Essa viabilidade deriva da disponibilidade de várias ferramentas computacionais capazes de realizar operações matriciais para efetuar a diagonalização de matrizes simétricas positivas semi-definidas.

O PCA oferece vantagens notáveis, como a capacidade de eliminar a multicolinearidade entre variáveis. Isso se traduz na habilidade de converter um conjunto original de variáveis que estão correlacionadas entre si em um novo conjunto de variáveis não correlacionadas, conhecidas como componentes principais. Ademais, ele permite reduzir o número de variáveis para um conjunto de eixos ortogonais que representam uma parcela das variáveis originais. Esses eixos explicam a variabilidade dos dados em ordem decrescente e independente (Hongyu, 2015; Regazzi, 2000).

Por outro lado, há algumas desvantagens associadas ao PCA. Ele é sensível a valores discrepantes (*outliers*) nos dados e não é recomendado quando há muitos valores ausentes na matriz ou quando se lida com dados ausentes. Também é inadequado quando o número de variáveis é maior do que o número de amostras. A redução de variáveis implica na perda de informações sobre a variabilidade original dos dados. Entretanto, essa parte retida pode ser considerada como o padrão de resposta, enquanto a parte descartada pode ser vista como ruído, ou seja, erro de medição e redundância. Vale notar que o PCA não é sempre eficaz – em algumas situações, mesmo após a redução de dimensões, a complexidade do problema pode não diminuir significativamente. Isso acontece especialmente quando as variáveis originais possuem pouca correlação, e no caso extremo quando a matriz de correlação é uma matriz identidade, nesse caso, os componentes principais se assemelham às próprias variáveis originais (Hongyu, 2015; Regazzi, 2000).

De acordo com Hongyu, Sandanielo e Oliveira Junior (2016), sejam as variáveis X_1, X_2, \dots, X_k em cada uma das n observações. Este conjunto de $n \times k$ medidas origina uma matriz de dados $\mathbf{X}(n \times k)$:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} \quad (5)$$

Para obter os componentes principais de uma forma geral, seja um conjunto de k variáveis X_1, X_2, \dots, X_k com médias $\mu_1, \mu_2, \dots, \mu_k$ e variância $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$, respectivamente. Tais variáveis são dependentes entre si e, portanto, possuem covariância entre a i -ésima e j -ésima variável definida por σ_{ij} , para $i \neq j = 1, 2, \dots, k$. Desse modo, as k variáveis podem ser expressas em forma vetorial por: $\mathbf{X} = [X_1, X_2, \dots, X_k]'$, com vetor de médias $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_k]'$ e matriz de covariância $\boldsymbol{\Sigma}$.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1k}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \cdots & \sigma_{nk}^2 \end{bmatrix} \quad (6)$$

Então, os pares de autovalores e autovetores $(\lambda_1, v_1), (\lambda_2, v_2), \dots, (\lambda_k, v_k)$ são encontrados, de modo que se pode definir o i -ésimo componente principal a partir dos autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$, associados à matriz $\boldsymbol{\Sigma}$ (Johnson; Wichern, 1998; Hongyu, 2015):

$$\mathbf{Z}_i = \mathbf{v}_i' \mathbf{X} = v_{i1}X_1 + v_{i2}X_2 + \cdots + v_{ik}X_k, \quad \forall i = 1, \dots, k. \quad (7)$$

O objetivo será determinar a variável \mathbf{Z}_i a partir da k variáveis contidas na matriz \mathbf{X} . A ideia envolve a transformação dos pontos de coordenadas originais em um plano, maximizando a distância entre eles, o que se traduz em aumentar ao máximo a variabilidade de \mathbf{Z}_i . A variância de \mathbf{Z}_i é dada por:

$$\text{Var}(\mathbf{Z}_i) = \text{Var}(\mathbf{v}_i' \mathbf{X}) = \mathbf{v}_i' \text{Var}(\mathbf{X}) \mathbf{v}_i = \mathbf{v}_i' \boldsymbol{\Sigma} \mathbf{v}_i \quad (8)$$

em que $i = 1, \dots, k$.

Utilizando a decomposição espectral⁶ da matriz Σ , tem-se que $\Sigma = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$, em que \mathbf{P} é a matriz composta de autovetores de Σ em suas colunas, enquanto $\mathbf{\Lambda}$ representa a matriz diagonal formada pelos autovalores de Σ , ou seja:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \quad (9)$$

e

$$\mathbf{P} = [v'_1 \quad v'_2 \quad \cdots \quad v'_k] \quad (10)$$

Deste modo:

$$tr(\Sigma) = tr(\mathbf{P}\mathbf{\Lambda}\mathbf{P}') = tr(\mathbf{\Lambda}\mathbf{P}\mathbf{P}') = tr(\mathbf{\Lambda}\mathbf{I}) = tr(\mathbf{\Lambda}) = \sum_{i=1}^k \lambda_i \quad (11)$$

Ou seja, $tr(\Sigma)$ é dado pela soma dos elementos da diagonal principal:

$$tr(\Sigma) = \sum_{i=1}^k \sigma_{ii} = \sum_{i=1}^k \lambda_i \quad (12)$$

Dessa forma, a quantidade total de variação presente nas variáveis originais equivale à quantidade total de variação presente nos componentes principais (Johnson; Wichern, 1998). A contribuição de cada componente principal (\mathbf{Z}_i) é expressa como uma porcentagem, e é possível calcular a explicação individual de cada componente, como por exemplo, para o k -ésimo componente principal, a proporção da explicação é determinada por:

$$C_k = \frac{Var(\mathbf{Z}_i)}{\sum_{i=1}^k Var(\mathbf{Z}_i)} = \frac{\lambda_i}{\sum_{i=1}^k \lambda_i} = \frac{\lambda_i}{tr(\Sigma)} \quad (13)$$

Através da avaliação da proporção pela qual o modelo de k componentes principais explica a variância total, é possível decidir quantos componentes devem ser mantidos. De acordo com Hongyu, Sandanielo e Oliveira Junior (2016), em geral escolhe-se o componente principal de maior importância (o primeiro componente principal) como sendo aquele de maior variância, que explique o máximo de variabilidade dos dados, o segundo

⁶ Seja A uma matriz simétrica $n \times n$, então A tem uma decomposição espectral $A = CDC'$, onde C é uma matriz $n \times n$ nas quais suas colunas são os autovetores C_1, C_2, \dots, C_n que correspondem aos autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$ de A e D é uma matriz diagonal $n \times n$ que possui como diagonal principal $\lambda_1, \lambda_2, \dots, \lambda_n$ (WANSBEEK; KAPTEYN, 1982).

componente de maior importância, o que apresenta a segunda maior variância e assim sucessivamente, até o componente principal de menor importância (Manly, 1986). Por outro lado, os últimos componentes principais serão responsáveis por direções que não estão associadas a muita variabilidade. Em outras palavras, esses últimos componentes principais identificarão relações lineares entre as variáveis originais próximo de constante (Johnson; Wichern, 1998; Anderson, 2003).

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

4.1 Algoritmo de clusterização *k-means*

Para iniciar a análise de agrupamentos, foi definida a métrica que mede a dissimilaridade (ou similaridade/semelhança) entre os elementos, assim como a abordagem estatística para a formação dos grupos, que se divide em métodos hierárquicos e não hierárquicos. Para tanto foram calculados três métodos para a escolha do número ótimo de *clusters*. A partir dessas métricas, o K ótimo ficou entre dois ou três (no Apêndice B, C e D): dois métodos indicaram três *clusters*, enquanto um indicou a existência de apenas dois. Dado que há 113 bairros, optou-se pela escolha de K=3, já que essa quantidade nos permite um agrupamento mais balanceado de bairros entre os *clusters*.

No que diz respeito à métrica de diferença, optou-se pela distância euclidiana⁷, uma escolha comum e tradicional na literatura. Quanto à estratégia hierárquica, foi selecionado o *Average Linkage*, que, em geral, gera agrupamentos mais eficazes do que os métodos de ligação simples ou completa. Para determinar o número ideal de grupos para a análise, realizou-se o teste de *Calinski e Harabasz*, cujo resultado, denominado pseudo-F, indica que quanto maior o valor, mais distintos são os grupos. Para a avaliação visual dos agrupamentos, a opção foi utilizar um dendrograma, que ilustra o processo de agrupamento e a distância entre os grupos.

Para efetuar o processo de agrupamento, foram empregadas como variáveis de referência as dimensões relacionadas ao contexto social, econômico e criminal dos 113 bairros situados em Fortaleza. A técnica de análise de *cluster* foi empregada com o propósito de categorizar os bairros com base nas suas características intrínsecas, de maneira que dentro de cada agrupamento estejam presentes aqueles bairros que compartilham traços mais similares. Deste modo, para agrupar os bairros foram utilizadas as cinco variáveis relacionadas à criminalidade e as três dimensões do IDH (vide Tabela 1). Em seguida, procedeu-se com o cálculo da matriz de distâncias (equação (3)), considerando a distância euclidiana. Desta feita, foram identificados três clusters.⁸

O primeiro *cluster* inclui 42 bairros (37% do total amostral) e se destaca por apresentar um patamar mais elevado de desenvolvimento em comparação com os demais. O

⁷ As principais métricas segundo a literatura correlata ao tema são: Distância Euclidiana; Distância de Manhattan; Distância de Minkowski e Distância de Mahalanobis.

⁸ Os gráficos da dissimilaridade e do método da silhueta identificaram a existência de três *clusters* e são expostos no Apêndice. O método do Gap Estatístico, por sua vez, indicou a existência de dois *clusters*. Desta feita, optou-se por trabalhar com três *clusters*.

IDH médio é mais substancial, com valores médios de 0,34 para IDH de Renda, 0,64 para IDH de Longevidade e 0,98 para IDH de Educação. Esses bairros aparentam ter um nível de desenvolvimento mais equilibrado nas diferentes dimensões do IDH. Quanto aos crimes, embora ainda existam ocorrências, as médias são relativamente menores. Esse *cluster* é o que detém os bairros mais nobres da cidade, com maiores níveis de desenvolvimento econômico local.

Tabela 3 - Quantidades de Bairros por cluster e suas respectivas médias de IDH e de crimes.

Cluster	Qtd Bairros	IDH Renda	IDH Long.	IDH Edu.	Valores absolutos				
					H	R	F	AD	AA
1	42	0,34	0,64	0,98	3	234	247	4,59	6
2	63	0,09	0,28	0,93	10	207	127	31,32	16
3	8	0,1	0,38	0,95	32	969	693	32,79	50

Fonte: Elaboração própria.

Nota: H – Homicídios; R – Roubos; F – Furtos; AD – Apreensão de Drogas/Entorpecentes (kg); e AA – Apreensão de Armas de Fogo.

O segundo *cluster* engloba a maioria dos bairros da cidade, um total de 63 (56% do total amostral). Nele, observa-se um padrão de IDH relativamente baixo, com valores médios de 0,09 para IDH de Renda, 0,28 para IDH de Longevidade e 0,93 para IDH de Educação. Esse *cluster* parece apresentar uma combinação de indicadores sociais menos desenvolvidos. Quanto à criminalidade, os números são baixos em relação a roubos e furtos, porém moderados em relação a homicídios e apreensão de armas e altos para apreensão de drogas, equiparando-se ao *cluster* 3 nessa dimensão.

O terceiro *cluster* é composto por apenas 8 bairros (7% do total amostral) e se destaca por também ter baixos indicadores de desenvolvimento humano e altas taxas de criminalidade. Em relação à criminalidade, este *cluster* apresenta as médias mais altas: 32 homicídios, 969 roubos, 693 furtos, 32,79 kg de apreensão de drogas e 50 apreensões de armas de fogo. Isso sugere que esses bairros enfrentam desafios significativos tanto em termos de desenvolvimento humano quanto de segurança pública.

Em resumo, os *clusters* refletem diferentes níveis de desenvolvimento humano e incidência de crimes em diversos bairros da capital Fortaleza. O *cluster* 1 apresenta um desenvolvimento mais substancial e níveis de crime mais baixos, o *cluster* 2 mostra um padrão de menor desenvolvimento e criminalidade moderada, enquanto o *cluster* 3 enfrenta desafios significativos tanto no desenvolvimento humano quanto na segurança pública, com altas taxas de criminalidade.

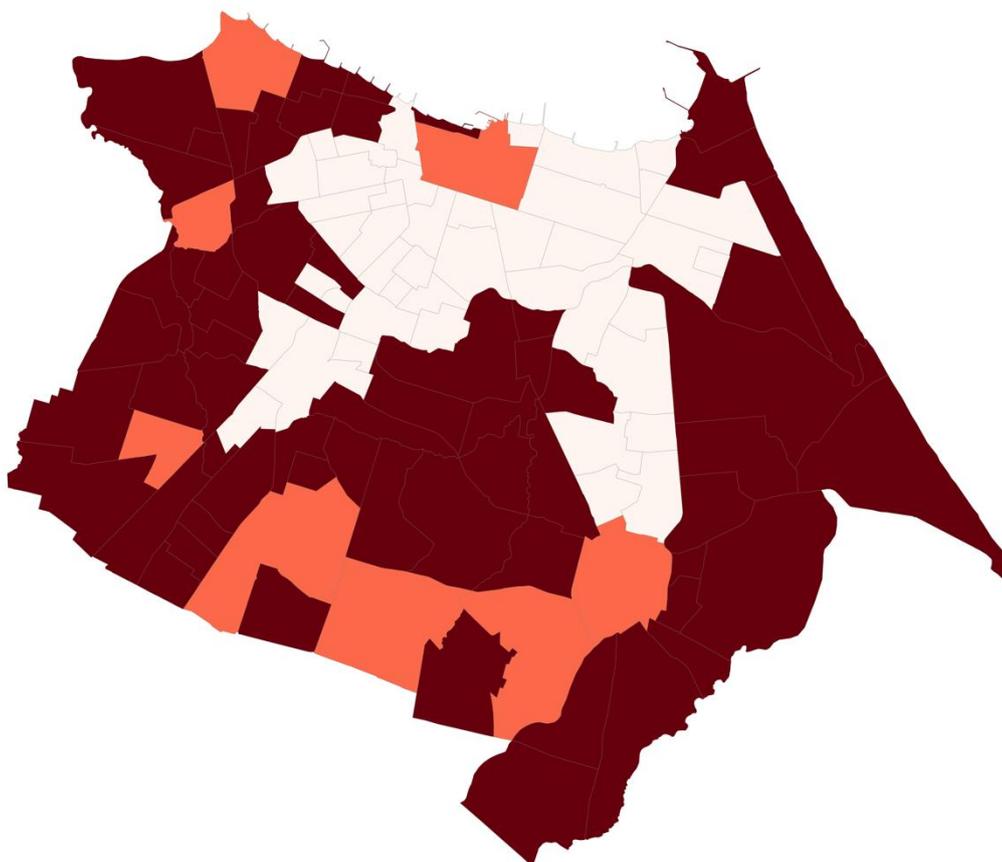
A Figura 4 apresenta o mapa de Fortaleza para os resultados da *clusterização*: os bairros brancos ao centro pertencem ao *cluster 1*, detém os bairros mais nobres da cidade e são caracterizados por menor atividade criminal, porém com índices moderados de roubos e furtos; bairros na cor vinho, por sua vez, remetem-se ao *cluster 2* e são os bairros com atividade criminal moderada; finalmente, os bairros na cor laranja são aqueles com maior atividade criminal e considerados os maiores perigosos da capital cearense. Os bairros pertencentes a cada *cluster* são apresentados no Apêndice A.

Observando o mapa anterior, fica patente que os bairros mais nobres da cidade são circundados por aqueles que detém maiores incidências de crimes, formando uma espécie de cinturão de insegurança. Metade dos bairros em laranja, caracterizados por deterem maiores níveis de criminalidade, estão localizados na zona sul da cidade: Acarapé (AIS 09), Mondubim (AIS 09), Prefeito José Walter (AIS 09), Messejana (AIS 03) e Jangurussu (AIS 03). Bom Jardim (AIS 02) aparece logo à oeste destes. No noroeste da cidade, aparecem Antônio Bezerra (AIS 06) e Barra do Ceará (AIS 08), mais acima. O bairro Centro (AIS 02), apesar de fazer fronteira com bairros nobres, tais como Aldeota (AIS 01), Meireles (AIS 01) e Praia de Iracema (AIS 01), também pertence ao grupo de bairros mais perigosos, em suma por conta dos altos índices de roubos e furtos. De outro lado, os bairros em vinho são mais contíguos e se estendem de leste a oeste da cidade.

As análises efetuadas podem ser usadas para identificar áreas onde o crime é mais comum, o que pode ajudar o poder público a direcionar esforços para essas áreas com intuito de reduzir o crime. Além da identificação dos bairros mais perigosos da cidade, os dados podem revelar fatores que contribuem para a propensão do crime, como pobreza e baixos níveis de educação, o que pode ajudar os formuladores de políticas no desenvolvimento de programas para abordar tais fatores e reduzir o risco de crime. Por fim, os resultados podem ser usados para avaliar a eficácia de políticas de segurança pública existentes. Isso pode ajudar o poder público na identificação de políticas que estão funcionando e que não estão funcionando, para que possam ser ajustadas ou descontinuadas, conforme necessário.

Ademais, conforme salientam Félix (1996; 2002) e Melara (2008), tais evidências justificam a tese de associação entre pobreza e a criminalidade pelo fato de determinados delitos concentrarem-se ou apresentarem maior visibilidade em regiões de baixo nível social e econômico, como por exemplo os 8 bairros que compõe o *cluster 3*.

Figura 4 - Cinturão de insegurança em Fortaleza – formação dos clusters, a partir do k-means.



Fonte: Elaboração própria.

Nota: Cluster 1 – criminalidade baixa (branco); Cluster 2 – criminalidade média (vinho); Cluster 3 – criminalidade alta (laranja).

4.2 Análise dos componentes principais (PCA)

Através da proporção da variância total que é explicada pelo modelo com k componentes principais, é possível decidir quantos componentes devem ser mantidos. Com base nos resultados obtidos pela técnica dos componentes principais, optou-se por reduzir a dimensionalidade das cinco variáveis relacionadas ao crime (Homicídios, Roubos, Furtos, Apreensão de Drogas/Entorpecentes (kg) e Apreensão de Armas de Fogo) e das três variáveis de IDH para duas dimensões, ou seja, dois componentes principais. Utilizou-se dois componentes principais que foram responsáveis por 67,56% da variação total sobre as incidências de crime em todos os bairros da cidade de Fortaleza, em que o CP1 foi responsável por 39,26% e o segundo, CP2, por 28,30% das variações dos dados. Como os dois primeiros componentes principais explicaram 67,56% da variação total dos dados (Figura 5), de acordo

com Rencher (2002), pelo menos 70% da variância total devem ser explicadas pelos primeiros e o segundo componentes principais.

Em trabalho similar para a taxa de criminalidade em algumas cidades dos Estados Unidos, Hongyu, Sandanielo e Oliveira Junior (2016) reduziram a dimensionalidade de sete variáveis para dois componentes principais, que foram responsáveis por 68,13% da variância total no conjunto de dados. Segundo tais autores, os dois primeiros componentes principais resumem efetivamente a variância amostral total e podem ser utilizados para o estudo do conjunto de dados, estratégia similar à seguida neste estudo.

Como intuito de se entender a importância de cada variável na construção dos dois componentes, na Tabela 4, são apresentados os coeficientes de ponderação de cada variável para a construção dos dois componentes principais. Com a seleção de dois componentes principais, a redução da dimensão de oito variáveis originais para dois componentes principais é bastante razoável. Portanto decidiu-se utilizar unicamente os dois primeiros componentes principais para a composição das equações (14) e (15).

Tabela 4 - Coeficientes de ponderação dos dois componentes principais.

X	Variáveis	CP_1	CP_2
X_1	Homicídios	0,48	0,18
X_2	Roubos	0,26	0,56
X_3	Furtos	0,02	0,58
X_4	Apreensão de Drogas	0,06	-0,04
X_5	Apreensão de Armas de Fogo	0,47	0,24
X_6	IDH Educação	-0,42	0,33
X_7	IDH Longevidade	-0,41	0,29
X_8	IDH Renda	-0,37	0,26

Fonte: Elaboração própria.

$$CP_1 = 0,48X_1 + 0,26X_2 + 0,02X_3 + 0,06X_4 + 0,47X_5 - (0,42X_6 + 0,41X_7 + 0,37X_8) \quad (14)$$

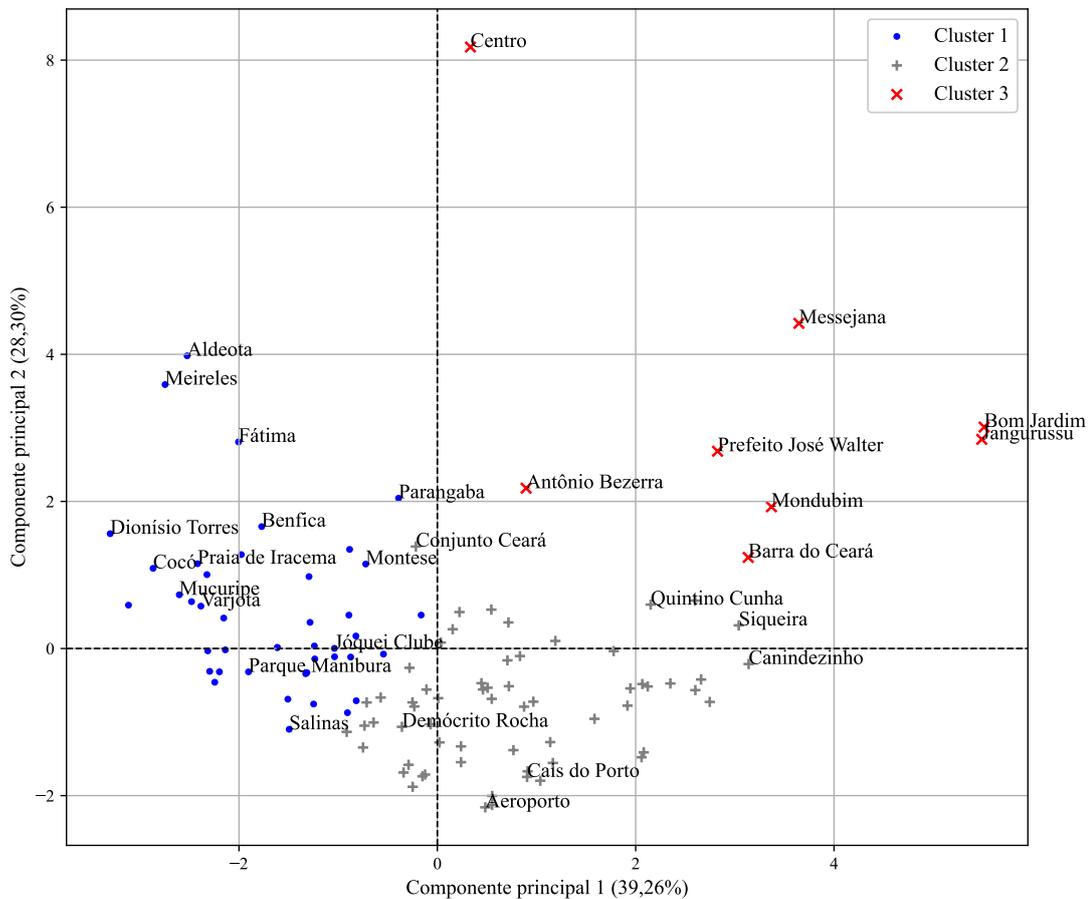
$$CP_2 = 0,18X_1 + 0,56X_2 + 0,58X_3 - 0,04X_4 + 0,24X_5 + 0,33X_6 + 0,29X_7 + 0,26X_8 \quad (15)$$

De acordo com a equação (14) e a Tabela 4, no primeiro componente principal, destacaram-se as variáveis X_1 (Homicídios) e X_5 (Apreensão de Armas de Fogo) e, neste caso, pode-se chamá-lo de componente de homicídios e apreensão de armas. De outro lado, no

segundo componente principal, equação (15), fica evidente a preponderância das variáveis X_2 (Roubos) e X_3 (Furtos), podendo ser chamado de componente de roubos e furtos. Ademais, fica evidente o contraste entre os dois componentes principais para as dimensões do IDH: bairros que apresentam maiores resultados para o CP_1 detêm níveis de IDH menores, enquanto os maiores pontuadores no CP_2 detêm maiores resultados de IDH.

A análise PCA foi aplicada para reduzir as dimensões das variáveis originais sem comprometer a integridade das informações. Por sua própria natureza, os principais componentes apresentam uma correlação nula, o que significa que a variação explicada no primeiro componente principal (CP1) é completamente independente da variação explicada no segundo componente principal (CP2) e assim por diante. Essa característica implica que a influência de um componente principal em particular não resultará em uma influência correlacionada nos outros componentes principais, evidenciando a ortogonalidade entre eles (Savegnago *et al.*, 2011; Fraga *et al.*, 2015).

Figura 5 - $CP_1 \times CP_2$ sobre índices de criminalidade por bairros de Fortaleza, em 2021.



Fonte: Elaboração própria.

Analisando a Figura 5, que é o gráfico de dispersão $CP_1 \times CP_2$ com os bairros sobre os índices de criminalidade e IDH, equações (14) e (15), fica patente que não há sobreposição entre os *clusters*, pode-se concluir que, de acordo com os dados de criminalidade por bairros da cidade de Fortaleza e com o PCA, Bom Jardim, Jangurussu, Messejana, Mondubim e Barra do Ceará, nessa ordem, são os que possuem maiores ocorrências de homicídios e apreensão de armas de fogo, a partir do CP_1 . De outro lado, a partir da análise da CP_2 , os bairros Centro, Messejana, Aldeota, Meireles e Bom Jardim foram os que apresentaram os maiores resultados para criminalidade relacionada a roubos e furtos, nessa ordem. Os bairros Bom Jardim e Messejana apareceram nas duas listagens, ou seja, apresentaram altos valores tanto para o CP_1 quanto para o CP_2 . Interessante observar que Aldeota e Meireles, bairros mais nobres da cidade, figuraram na lista dos cinco bairros com maiores resultados no CP_2 .

Os oito bairros pertencentes ao *cluster 3* (vide Tabela A1) se localizam no extremo do primeiro quadrante da Figura 5, ou seja, foram os que apresentaram maiores valores para os dois componentes principais. Isso mostra que a natureza da base de dados aqui utilizada permite a associação do crime em Fortaleza com particularidades locais dos 113 bairros híbridos, e a captação de uma possível dependência espacial mesmo resultado encontrado por Guerra Jr (2016). Numa análise preliminar, ainda que parcial, nota-se que o crime não esteve distribuído uniformemente na capital cearense – apenas 7% dos bairros (*cluster 3*) agrupam aproximadamente 71,1% dos homicídios, 67,1% das ocorrências de roubos e furtos, 47,7% das apreensões de drogas/entorpecentes e 69,4% das apreensões de armas de fogo do total de ocorrências reportadas, no ano de 2021. Tais fatos evidenciam a causa de o *cluster 3* ter se relevado como o mais inseguro dentre os três, motivo que levou a distinção deste dos demais.

De outro lado, os bairros que formam o *cluster 1* se localizam basicamente no 2º e 3º quadrantes, enquanto os bairros que representam o *cluster 2* permanecem entre os outros dois *clusters*, indo do bairro Conjunto Ceará ao Canindezinho, a leste, e ao Aeroporto, ao sul, por exemplo.

5 CONSIDERAÇÕES FINAIS E PROPOSTAS DE POLÍTICAS PÚBLICAS

O Ceará é o estado que detém o maior número absoluto de homicídios e Fortaleza concentra boa parte deles. Este estudo usa técnicas estatísticas para agrupar bairros da capital que compartilham padrões similares de criminalidade. A metodologia empregada reduz a complexidade dos dados e mantém a maior parte das informações relevantes contidas nas variáveis, permitindo identificar os bairros com maiores índices de violência na cidade. Essa é uma abordagem inovadora para analisar dados criminais em Fortaleza, e considera as semelhanças e diferenças entre conjuntos de observações em várias áreas da cidade. Nossos resultados fornecem informações essenciais para análises vindouras e permitem o desenvolvimento de políticas públicas customizadas e ajustadas a cada grupo de bairros, de acordo com suas características criminais e socioeconômicas. É patente na pesquisa que os bairros mais perigosos da cidade estão majoritariamente localizados na região sul e oeste da capital, sendo uma exceção o bairro Centro, principalmente devido às altas taxas de roubos e furtos. Também se evidenciou que os bairros mais nobres da cidade são cercados por bairros com maior criminalidade, resultando em uma espécie de cinturão de insegurança.

Por meio da Análise dos Componentes Principais foram construídos dois indicadores que dão pesos diferentes para as diversas variáveis socioeconômicas e criminais utilizadas. Um critério dá maior peso às variáveis de homicídios e apreensões de armas de fogo e aponta os bairros Bom Jardim, Jangurussu, Messejana, Mondubim e Barra do Ceará como os cinco mais impactados. Já o segundo critério de ponderação dá maior peso a roubos e furtos e indica Centro, Messejana, Aldeota, Meireles e Bom Jardim como os bairros em situação mais preocupante. Note-se que Bom Jardim e Messejana figuram entre os maiores índices, independentemente do critério de ponderação, e que bairros privilegiados, como Aldeota e Meireles, também detém pontuações elevadas, de acordo com a ponderação acerbada nos roubos e furtos.

A análise da distribuição espacialmente heterogênea dos crimes apresentada aqui reforça as análises de Medeiros et al. (2013; 2015), Guerra Jr (2016), Muggah e Pellegrino (2020) e Santos (2020), além de dar um suporte focal às políticas e ações conduzidas pelo Estado e pela Prefeitura para a redução da Violência, Criminalidade, Marginalidade (VCM) no território da capital cearense.

Primeiramente, o estudo permite que os diferentes poderes tenham uma visão comum e unificada do problema, o que permite uma maior sinergia entre os governos para direcionar seus recursos e esforços para as regiões que mais necessitam de intervenção. A

análise também permite que as agências de segurança pública e outros órgãos de fomento possam alocar recursos de forma mais coordenada e eficiente. Isso inclui a distribuição de policiais, serviços sociais, programas de prevenção de crimes e outras medidas para áreas que enfrentam desafios significativos de segurança.

Segundo, ao entender as tendências de criminalidade em diferentes áreas, os formuladores de políticas podem desenvolver estratégias de prevenção mais direcionadas. Isso pode incluir a implementação de programas de educação, emprego, habitação e reabilitação em áreas de alto risco para abordar as causas subjacentes da criminalidade.

A análise também permite uma reavaliação ou reforço das políticas existentes. Se essa análise for sistematicamente realizada ao longo do tempo, e ao se constatar mudanças nos índices de criminalidade ao longo do tempo e em diferentes áreas, os governos podem determinar quais políticas estão funcionando e quais podem precisar de ajustes.

Por fim, o estudo promove a transparência e a responsabilidade das autoridades, permitindo ainda intervenções baseadas em evidências. Isso ajuda a evitar decisões arbitrárias ou baseadas em percepções subjetivas, permitindo que as políticas sejam mais fundamentadas em dados objetivos, além de permitir que os cidadãos monitorem as tendências de criminalidade em suas comunidades e responsabilizem os órgãos governamentais por suas ações no combate à criminalidade.

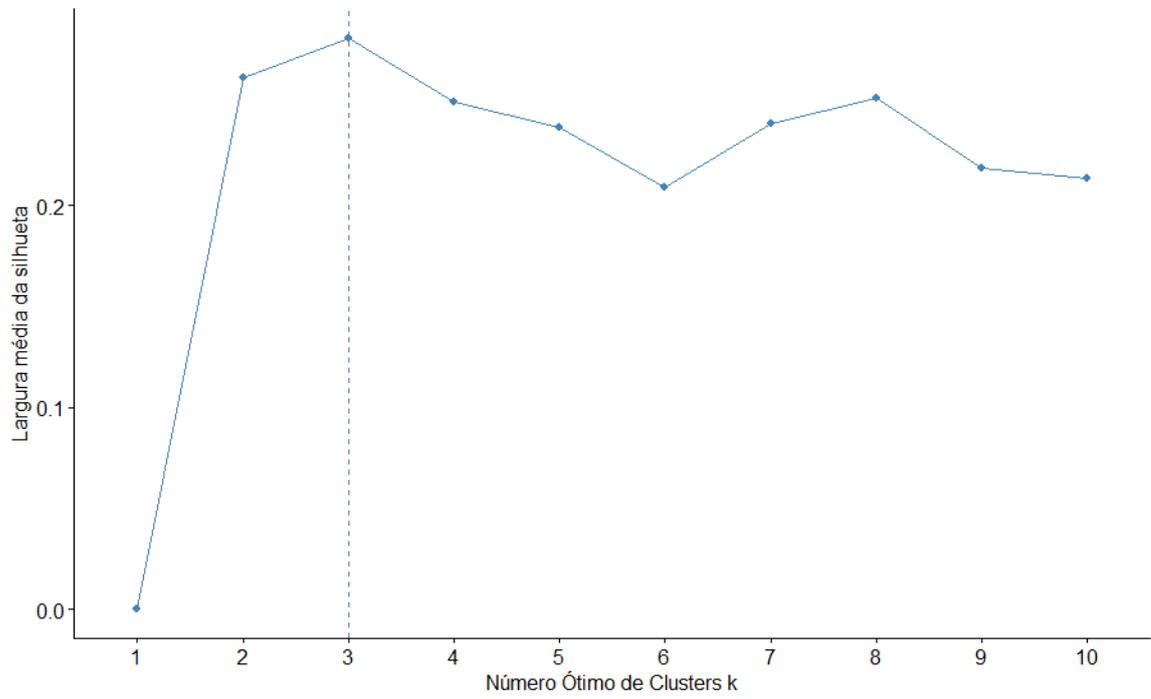
Dentro da agenda de pesquisa futura, está o foco em entender como esses dados dialogam com os resultados a serem apresentados pelo Censo Demográfico Brasileiro realizado em 2022. Uma vez que seja possível relacionar os dados estudados nessa pesquisa com outros indicadores presentes no Censo, como a informalidade e a pobreza por bairros de Fortaleza, torna-se factível identificar outros padrões existentes.

APÊNDICE A – Quadro dos Bairros pertencentes a cada cluster.

<i>Cluster 1</i>	Aldeota, Amadeu Furtado, Benfica, Bom Futuro, Cambeba, Cidade 2000, Cidade dos Funcionários, Cocó, Damas, Dionísio Torres, Engenheiro Luciano Cavalcante, Farias Brito, Fátima, Guararapes (Patriolino Ribeiro), Itaoca, Jacarecanga, Jardim América, Joaquim Távora, Jóquei Clube, José Bonifácio, Meireles, Monte Castelo, Montese, Mucuripe, Panamericano, Papicu, Parangaba, Parque Araxá, Parque Iracema, Parque Manibura, Parquelândia, Parreão, Praia de Iracema, Presidente Kennedy, Rodolfo Teófilo, Salinas, São Gerardo, São João do Tauape, Varjota, Vila Ellery, Vila Pery, Vila União.
<i>Cluster 2</i>	Aerolândia, Aeroporto, Alagadiço Novo (José de Alencar), Alto da Balança, Álvaro Weyne, Ancuri, Autran Nunes, Barroso, Bela Vista, Boa Vista (Castelão\Mata Galinha), Bonsucesso, Cais do Porto, Cajazeiras, Canindezinho, Carlito Pamplona, Coaçu, Conjunto Ceará, Conjunto Esperança, Conjunto Palmeiras, Couto Fernandes, Cristo Redentor, Curió, Demócrito Rocha, Dendê, Dias Macêdo, Dom Lustosa, Edson Queiroz, Floresta, Genibaú, Granja Lisboa, Granja Portugal, Guajerú, Henrique Jorge, Itaperi, Jardim Cearense, Jardim das Oliveiras, Jardim Guanabara, Jardim Iracema, João XXIII, Lagoa Redonda, Lagoa Sapiranga (Coité), Manuel Dias Branco, Maraponga, Moura Brasil, Padre Andrade, Parque Dois Irmãos, Parque Presidente Vargas, Parque Santa Rosa, Parque São José, Passaré, Paupina, Pedras, Pici, Pirambu, Planalto Ayrton Senna, Praia do Futuro, Quintino Cunha, Sabiaguaba, Serrinha, Siqueira, Vicente Pinzón, Vila Manoel Sátiro, Vila Velha.
<i>Cluster 3</i>	Antônio Bezerra, Barra do Ceará, Bom Jardim, Centro, Jangurussu, Messejana, Mondubim, Prefeito José Walter.

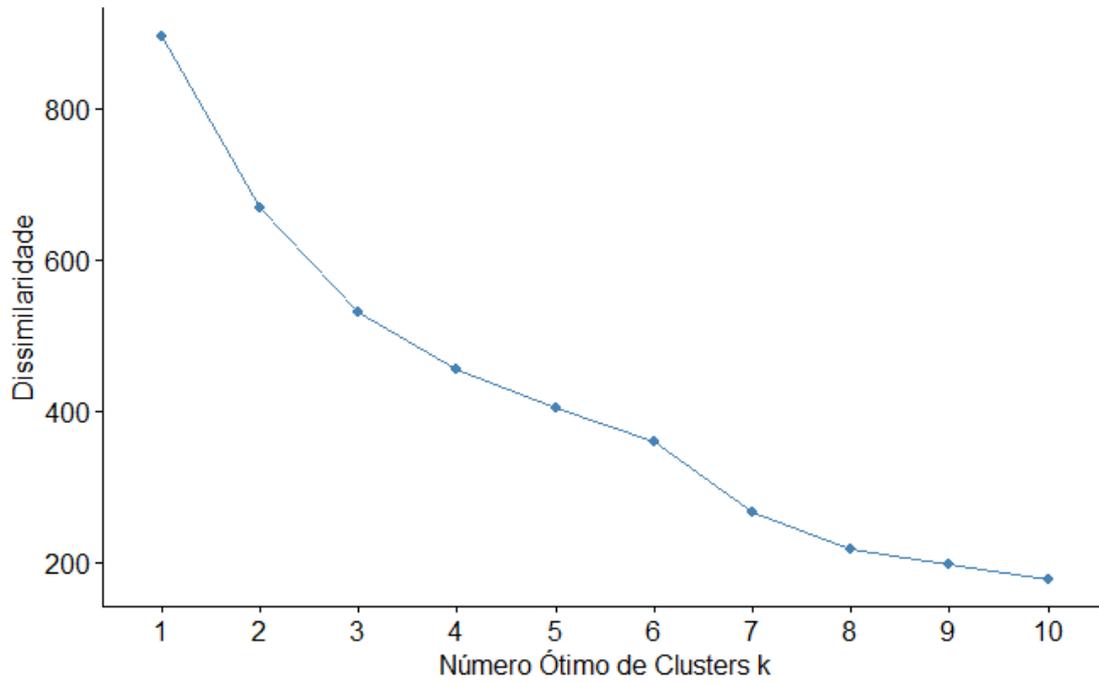
Fonte: Elaboração própria.

APÊNDICE B – GRÁFICO COM O RESULTADO DO MÉTODO DA SILHUETA E NÚMERO ÓTIMO DE CLUSTERS.



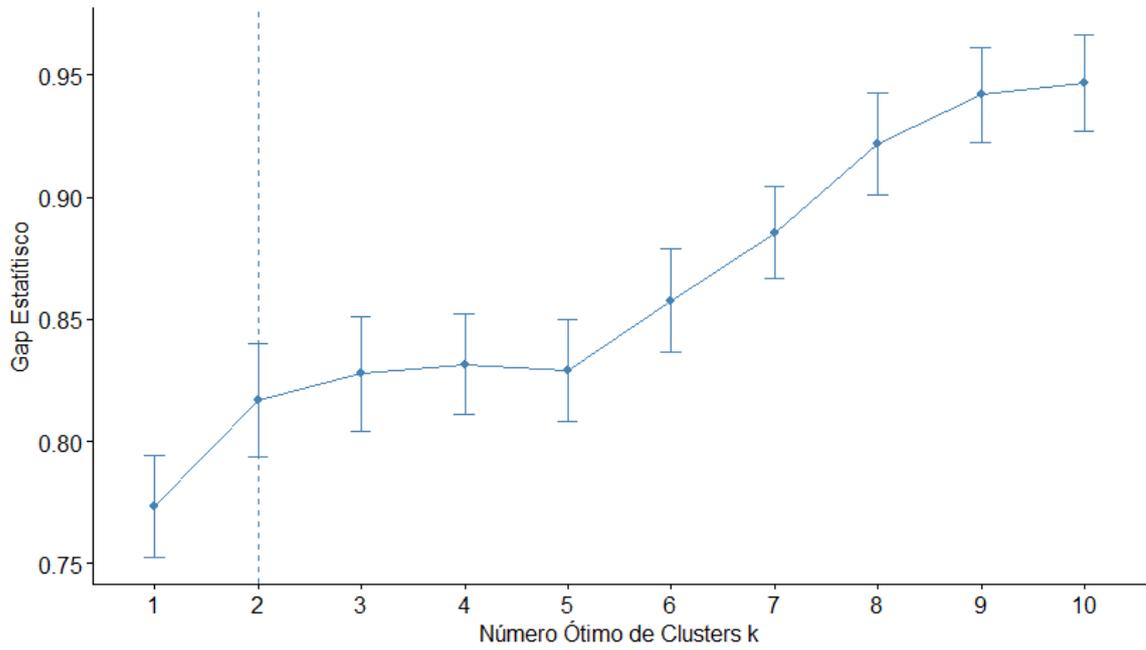
Fonte: Elaboração própria.

APÊNDICE C – GRÁFICO COM O RESULTADO DA DISSIMILARIDADE E NÚMERO ÓTIMO DE CLUSTERS (MÉTODO DE ELBOW).



Fonte: Elaboração própria.

APÊNDICE D – GRÁFICO COM O RESULTADO DO MÉTODO DO GAP ESTATÍSTICO E PARA O CÁLCULO DO NÚMERO ÓTIMO DE CLUSTERS.



Fonte: Elaboração própria.

REFERÊNCIAS

- ANDERSON, T.W. **An introduction to multivariate statistical analysis**. New York: Wiley, 6 ed. 2003.
- ANDRESEN , M.A. Crime measures and the spatial analysis of criminal activity. **British Journal of criminology**, 46(2), p. 258-285, 2006.
- ANSELIN, L., COHEN, J., COOK, D., GOOR, W. and TITA, G. Spatial analyses of crime. **Criminal justice**, 4(2), p. 213-262, 2000.
- AIN, A. K. Agrupamento de dados: 50 anos além do K-Means. **Letras de reconhecimento de padrão**, v. 31, nº. 8, p. 651-666, 2010.
- ARBEX, M.; CORRÊA, M. V.; MAGALHÃES, M. R. V. Tolerance of informality and occupational choices in a large informal sector economy. **The BE Journal of Macroeconomics**, De Gruyter, v. 23, n. 1, p. 241–278, 2022.
- AZEVEDO, M. A. Concepções sobre Criminalidade e Modelos de Policiamento. **Revista Psicologia: Ciência e Profissão**. 23 (3), p. 18-25, 2003.
- BALAGUER-COLL, M.T., BRUN-MARTOS, M.I., MÁRQUEZ-RAMOS, L., PRIOR, D. Local government efficiency: determinants and spatial interdependence. **Applied Economics**, 51(14), p. 1478-1494, 2019.
- BATELLA, W. B.; DINIZ, A. M. A. Análise Espacial dos Condicionantes da Criminalidade Violenta no Estado de Minas Gerais. **Sociedade & Natureza**, Uberlândia, 22 (1): 151-163, 2010.
- BATOOL, F.; HENNING, C. Clustering with the average silhouette width. **Computational Statistics & Data Analysis**, 158, 107190, 2021.
- BEATO FILHO, C. C. Determinantes da Criminalidade em Minas Gerais. **Revista Brasileira de Ciências Sociais**. São Paulo, Vol.13, n.37, p. 74-89, 1998.
- BEATO FILHO, C. C.; REIS, I. A. Desigualdade, desenvolvimento socioeconômico e crime. In: HENRIQUES, R. **Desigualdade e Pobreza no Brasil**. Rio de Janeiro: IPEA, Cap. 13, p. 385-403, 2000.
- BEATO FILHO, C. C.; et al. Conglomerados de homicídios e o tráfico de drogas em Belo Horizonte, Minas Gerais, Brasil, de 1995 a 1999. **Cadernos de Saúde Pública**, v. 17, n. 5, p. 1163–1171, 2001.
- BEHIRY, M. H.; ALY, M. Detecção de ciberataques em redes de sensores sem fio usando uma técnica híbrida de redução de características com métodos de IA e aprendizado de máquina." **Journal of Big Data**, v. 11, n. 1, p. 16, 2024.
- BECKER, G. S. Crime and punishment: an economic approach. **Journal of Political Economy**, Chicago, v. 76, n. 2, p. 169-217, 1968.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. New York: Springer, 2006.

- CECCATO, V. Public space and the situational conditions of crime and fear. **International Criminal Justice Review**, Georgia, v. 26, n. 2, p. 69-79, 2016.
- CECCATO, V; NALLA, M. K. Crime and fear in public spaces: towards safe, inclusive and sustainable cities. New York: Routledge, 2020.
- COSTA, F. V. C. "**A violência impera nessa cidade**": reflexões sobre o medo de assaltos em Fortaleza, CE. 2018. 227 f. Tese (Doutorado) - Curso de Sociologia, Universidade Federal do Ceará, Fortaleza, 2018.
- EDRIS, M., GHASEMI-VARNAMKHASTI, M., KIANI, S., YAZDANPANA, H., IZADI, Z. Identificação da autenticidade e origem geográfica do arroz através da análise de imagens hiperespectrais utilizando algoritmos de agrupamento não supervisionado. **Journal of Food Composition and Analysis**, v. 125, p. 105737, 2024.
- FÁVERO, L. P.; BELFIORE, P. **Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®**. LTC. 2021.
- FAWCETT, T.; PROVOST, F. **Data Science para Negócios: O que você precisa saber sobre mineração de dados e pensamento analítico de dados**. Alta Books Editora, 2016.
- FEITOZA, B. M. B.; COSTA, J. H. Violência urbana, insegurança e turismo na 'Cidade do Sol' (Natal, RN, Brasil). **Anuario Turismo y Sociedad**, n. 25, p. 93-112, 2019.
- FELIX, A. S. Geografia do Crime. São Paulo. **Revista de Geografia**: Editora UNESP. V13, 1996.
- FELIX, A. S. **Geografia do crime: interdisciplinaridade e relevâncias**. Marília: Marília-UNESP-Publicações, 2002.
- FBSP, FORUM BRASILEIRO DE SEGURANÇA PÚBLICA. **Anuário Brasileiro de Segurança Pública 2022**. Ano 16, 2022. Disponível em: <https://forumseguranca.org.br/wp-content/uploads/2022/06/anuario-2022.pdf?v=4>. Acesso em: 04 ago. 2023.
- FRAGA, A. B.; SILVA, F. L.; HONGYU, K.; SANTOS, D. D. S.; MURPHY, T. W.; LOPES, F. B. Multivariate analysis to evaluate genetic groups and production traits of crossbred Holstein × Zebu cows. **Trop Anim Health Prod**. p. 1- 6, 2015.
- GOLDBERG, M.; KIM, K. W.; ARIANO, M. **How firms cope with crime and violence: experiences from around the world**. Washington, Dc: World Bank, 2014. Disponível em: <https://openknowledge.worldbank.org/handle/10986/16539>. Acesso em: 04 jul. 2023.
- GUERRA Jr., M. B. A Estrutura Social do Crime e o Spillover Espacial do Policiamento: O Caso de Fortaleza, Brasil. Dissertação de Mestrado – Universidade Federal do Ceará – UFC, Centro de Aperfeiçoamento dos Economistas do Nordeste – CAEN, Programa de Pós-Graduação em Economia, 2016.

HAIR JUNIOR, J. F., BLACK, W. C., BABIN, N. J., ANDERSON, R. E., & TATHAM, R. L. **Análise multivariada de dados**. 6ª ed., São Paulo: Bookman, 2009.

HARTIGAN, J.A.; WONG, M.A. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, 28, 100-108, 1979.

HONGYU, K. **Comparação do GGE-biplot ponderado e AMMI-ponderado com outros modelos de interação genótipo × ambiente**. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2015.

HONGYU, K.; SANDANIELO, V. L. M.; OLIVEIRA JUNIOR, G. J. Análise de componentes principais: resumo teórico, aplicação e interpretação. *E&S Engineering and science*, 5(1), pp.83-90, 2016.

IBGE (Instituto Brasileiro de Geografia e Estatística). **Produto Interno Bruto dos Municípios**. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9088-produto-interno-bruto-dos-municipios.html>. Acesso em: 04 ago. 2023.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. Madison: Prentice Hall International, 1998.

JOLLIFFE, I. T. **Principal Component Analysis**. Springer Series in Statistics. New York: Springer, 2002.

KAUFMAN, L.; ROUSSEEUW, P. **Finding groups in data: an introduction to cluster analysis**. John Wiley & Sons, 1987.

LONGUINHO, L. P. **Implementação de políticas públicas de prevenção à criminalidade em Minas Gerais: o Programa Fica Vivo! e o Programa de Inclusão Social de Egressos do Sistema Prisional**. Dissertação (Mestrado em Ciência Política) - Universidade Federal de Minas Gerais, Departamento de Ciência Política, 2022.

MANLY, B. F. J. **Multivariate statistical methods**. New York, Chapman and Hall, 1986.

MEDEIROS, C. N.; CARVALHO, J. R.; e OLIVEIRA, V. H. **Violência, desenvolvimento e demografia: uma análise espacial para a cidade de fortaleza em anos recentes**. In: BARRETO, Flávio Ataliba Flexa Daltro; MENEZES, Adriano Sarquis B. de; ASSIS, Dércio N. Chaves de; SOUSA, Fátima Juvenal de. (org.). *Economia do Ceará em Debate*, 2015. Fortaleza, CE: IPECE, v. 1, p. 8-41, 2015.

MEDEIROS, C. N.; PINHEIRO, F. S. A.; SOUZA, G. M.; e PINHEIRO, D. R. C. Correlações espaciais entre ocorrências de homicídios e concentração de aglomerados subnormais em Fortaleza/CE/Brasil. *Revista GeoUECE - Programa de Pós-Graduação em Geografia da UECE Fortaleza/CE*, v.2, nº1, p. 92-110, jan./jun.; 2013.

MELARA, E. **A dinâmica da violência criminal no espaço urbano de Santa Maria-RS**. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul.

Instituto de Geociências. Programa de Pós-Graduação em Geografia (UFRGS/PPGEA), Porto Alegre, RS - BR, 2008.

MELO, J; PAIVA, L. F. S. Violências em territórios faccionados do Nordeste do Brasil: notas sobre as situações do Rio Grande do Norte e do Ceará. **Revista USP**, v. 1, n. 129, 2021, p. 47-62.

MORAES, S. S. DE; SIQUEIRA, I. B. L. As margens como centro no bairro Benfica: falas da violência e do matar na cidade de Fortaleza. **RBSE Revista Brasileira de Sociologia da Emoção**, v. 18, n. 53, agosto de 2019, p. 143-159. ISSN 1676 8965.

MOUSAVIAN ANARAKI, S. A.; HAERI, A.; MOSLEHI, F. (2021). A hybrid reciprocal model of PCA and K-means with an innovative approach of considering sub-datasets for the improvement of K-means initialization and step-by-step labeling to create clusters with high interpretability. *Pattern Analysis and Applications*, 24(3), 1387–1402. doi:10.1007/s10044-021-00977-x

MUGGAH, R.; e PELLEGRINO, A. P. **Prevenção da violência juvenil no Brasil: uma análise do que funciona**. Instituto Igarapé, Rio de Janeiro, fevereiro de 2020.

REGAZZI, A. J. **Análise multivariada, notas de aula INF 766**, Departamento de Informática da Universidade Federal de Viçosa, v. 2, 2000.

O POVO. Nove fatos sobre a violência no Ceará apontados no Anuário de Segurança Pública. Fortaleza, Ceará, 2021. Disponível em: <https://www.opovo.com.br/noticias/fortaleza/2021/07/16/nove-fatos-sobre-a-violencia-no-ceara-apontados-no-anuario-de-seguranca-publica.html>. Acesso em: 20 jan. 2024.

O POVO. A correlação entre homicídios e indicadores sociais em Fortaleza. Fortaleza, Ceará, 2022. Disponível em: <https://mais.opovo.com.br/jornal/reportagem/2022/09/22/a-correlacao-entre-homicidios-e-indicadores-sociais-em-fortaleza.html>. Acesso em: 20 jan. 2024.

O POVO. Duas cidades da Grande Fortaleza estão entre as 50 mais violentas do Brasil. Fortaleza, Ceará, 2023. Disponível em: <https://www.opovo.com.br/noticias/ceara/2023/07/20/duas-cidades-da-grande-fortaleza-estao-entre-as-50-mais-violentas-do-brasil.html>. Acesso em: 20 jan. 2024.

PAIVA, L. F. S. “Aqui não tem gangue, tem facção”: as transformações sociais do crime em Fortaleza, Brasil. **Caderno CRH**, v. 32, 2019, p. 165- 184.

RENCHER, A. C. **Methods of Multivariate Analysis**. A John Wiley & Sons, Inc. Publication, p. 727, 2 ed., 2002.

SANTOS, A. L. C. Políticas criminais, direito à cidade e capital social. Reflexões sobre a sustentabilidade de políticas públicas de tratamento da criminalidade e da violência. **Revista de Direito da Cidade**, v. 11, n. 3, p. 727-765, 2019.

SANTOS, M. D.; VASCONCELOS, R. A.; MOREIRA JÚNIOR, F. N. **Gastos socioeconômicos da violência e criminalidade (V&C) para o estado do Ceará:**

evidências para o período 2010-2018. Boletim de Gestão Pública - Nº 17; Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE); Fortaleza/Ceará; 2020.

SANTOS, M. D. **Ensaio sobre violência & criminalidade (V&C) em uma análise teórica, empírica e espacial: evidências para o Estado do Ceará, capital de Fortaleza e bairros.** Tese (Doutorado) – Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade, Programa de Pós-Graduação em Economia, Fortaleza, 2020.

SAVEGNAGO, R. P., CAETANO, S. L., RAMOS, S. B., NASCIMENTO, G. B., SCHMIDT, G. S., LEDUR, M. C. MUNARI, D. P. Estimates of genetic parameters, and cluster and principal components analyses of breeding values related to egg production traits in a White Leghorn population, **Poultry Science**, 90, p. 2174-2188, 2011.

SCOOT, A. J. Industrial organization and location: Division of labor, the firm, and spatial process. **Economic geography**, 62(3), p. 215-231, 1986.

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. **Journal of The Royal Statistical Society B**, 63, p. 411–423, 2001.

WANSBEEK, T.; KAPTEYN, A. A simple way to obtain the spectral decomposition of variance components models for balanced data. **Communications in Statistics-Theory and Methods** 11, no. 18, p. 2105-2112, 1982.

W.U., S. Research on innovation and development of university instructional administration informatization in IoT and big data environment. **Soft Computing**, v. 27, n. 24, p. 19075-19094, 2023.