



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO ACADÊMICO DO AGRESTE
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

JEFFERSON CARLOS DE OLIVEIRA RIBEIRO COSTA

**IDENTIFICAÇÃO DE GRUPOS DE MUNICÍPIOS PERNAMBUCANOS PARA
RECOMENDAÇÃO DE POLÍTICAS DE SEGURANÇA PÚBLICA UTILIZANDO
UMA TÉCNICA DE CLUSTERIZAÇÃO**

Caruaru

2020

JEFFERSON CARLOS DE OLIVEIRA RIBEIRO COSTA

**IDENTIFICAÇÃO DE GRUPOS DE MUNICÍPIOS PERNAMBUCANOS PARA
RECOMENDAÇÃO DE POLÍTICAS DE SEGURANÇA PÚBLICA UTILIZANDO
UMA TÉCNICA DE CLUSTERIZAÇÃO**

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

Área de concentração: Otimização e Gestão da Produção.

Orientadora: Prof.^a Dr.^a Maísa Mendonça Silva.

Caruaru
2020

Catálogo na fonte:
Bibliotecário – Raul César de Melo - CRB/4 - 1735

C837i Costa, Jefferson Carlos de Oliveira Ribeiro.
Identificação de municípios pernambucanos para recomendação de políticas de segurança pública utilizando uma técnica de clusterização / Jefferson Carlos de Oliveira Ribeiro Costa. – 2020.
69 f. : il. ; 30 cm.

Orientadora: Maísa Mendonça Silva.
Dissertação (Mestrado) – Universidade Federal de Pernambuco, CAA, Programa de Pós-Graduação em Engenharia de Produção, 2020.
Inclui Referências.

1. Segurança pública - Pernambuco. 2. Política pública. 3. Cluster (Sistema de computador). 4. Mineração de dados (Computação). 5. Algoritmos computacionais. I. Silva, Maísa Mendonça (Orientadora). II. Título.

CDD 658.5 (23. ed.) UFPE (CAA 2020-054)

JEFFERSON CARLOS DE OLIVEIRA RIBEIRO COSTA

**IDENTIFICAÇÃO DE GRUPOS DE MUNICÍPIOS PERNAMBUCANOS PARA
RECOMENDAÇÃO DE POLÍTICAS DE SEGURANÇA PÚBLICA UTILIZANDO
UMA TÉCNICA DE CLUSTERIZAÇÃO**

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

Aprovada em 14/05/2020.

BANCA EXAMINADORA

Prof.^a Dr.^a Maísa Mendonça Silva (Orientadora)
Universidade Federal de Pernambuco

Prof. Dr. Thyago Celso Cavalcante Nepomuceno (Examinador Interno)
Universidade Federal de Pernambuco

Prof.^a Dr.^a Eduarda Asfora Frej (Examinadora Externa)
Universidade Federal de Pernambuco

À memória da minha avó, Izabel Paiva.

AGRADECIMENTOS

Diante do árduo caminho que percorri para chegar até aqui, algumas pessoas foram essenciais para que eu conseguisse alcançar cada etapa.

Gostaria de agradecer primeiramente a minha família, que sempre me apoiou e me incentivou em meus estudos. Em especial a minha mãe, que a tenho como exemplo de pessoa altruísta, resistente e destemida.

Agradeço a minha orientadora e amiga Prof.^a Dra. Maísa Mendonça, que teve papel crucial em toda minha formação, desde a graduação. Agradeço por toda partilha de conhecimento, paciência e incentivo.

À coordenação e secretaria do PPGEP que sempre se dispuseram a me auxiliar nas resoluções de questões administrativas e burocráticas. Não menos importante, à CAPES, que exerce uma responsabilidade primordial para o desenvolvimento do nosso país na expansão e consolidação da pós-graduação.

A todos os meus amigos, principalmente aqueles que estiveram sempre nessa jornada comigo. Gostaria de agradecer a minha eterna parceira de estudos, Ana Paula, que além de ter me apoiado nos piores e melhores momentos, também fez parte desse caminho comigo em busca de um futuro melhor. Um agradecimento especial a cada um dos meus amigos pertencentes ao grupo “NTNB” e aos meus amigos da minha cidade natal, Garanhuns. Cada conversa, abraço e *happy hour* fez parte do combustível necessário à minha corrida nessa trajetória.

Por fim, gostaria de finalizar meus agradecimentos expressando minha felicidade e gratidão por estar obtendo o título de mestre numa área em que sou apaixonado e num centro universitário interiorizado, que desde 2006 vem ofertando oportunidades e mudando a realidade deste país. Agradeço a Deus e as boas energias do universo.

RESUMO

O direcionamento de políticas públicas é de fundamental importância para a sociedade como um todo, em especial na segurança, que além de ser considerada como uma necessidade de todo cidadão, é garantida constitucionalmente. Esta dissertação tem por objetivo utilizar uma abordagem de aprendizagem não supervisionada para o estabelecimento de clusters entre os municípios do estado de Pernambuco, considerando alguns tipos de crimes representativos no estado, visando direcionar ações para prevenção e combate ao crime de forma a apoiar os formuladores de políticas de segurança pública. Os dados foram obtidos a partir da Lei nº 12.527/2011 que regulamenta o direito constitucional de acesso às informações públicas e foi utilizado o algoritmo *k-means*, que é uma ferramenta de mineração de dados multivariados consolidada na literatura, como principal ferramenta considerada no estudo. Foram realizadas aplicações do algoritmo para seis diferentes valores do parâmetro *k* (2, 3, 4, 5, 12 e 26); parâmetro esse que é exigido anteriormente a execução do algoritmo, e que diz respeito ao número de clusters ou grupos. Para demonstrar o direcionamento das ações, foi utilizado como base o agrupamento realizado para $k=26$, fazendo referência a áreas de integração de segurança já existentes. Toda execução do algoritmo foi realizada através do *software* R 3.6.1 e foram direcionadas recomendações de ações a cada um dos clusters considerados. A formulação das ações se deu a partir de uma investigação a documentos específicos, como o Plano Nacional de Segurança Pública, Plano Nacional sobre Drogas e outros fornecidos pela Secretaria de Política de prevenção à Violência e às Drogas e a Secretaria de Defesa Social de Pernambuco. A partir dos resultados foi percebido que o uso de uma abordagem de clusterização de municípios fornece uma efetividade maior no direcionamento de ações de combate e prevenção de crimes, diante do fato de que são agrupados num mesmo cluster aqueles municípios que apresentam maiores semelhanças no que se refere a criminalidade.

Palavras-chave: Políticas públicas. Segurança pública. Cluster. Algoritmo k-means.

ABSTRACT

The direction of public policies is of fundamental importance for a society as a whole, in particular security, which in addition to being considered a necessity for every citizen, is constitutionally guaranteed. This dissertation aims to use an unsupervised learning approach for the establishment of clusters among municipalities in the state of Pernambuco, considering some types of representative crimes in the state, aiming to direct actions for preventing and combating crime in order to support policy makers. public security policies. The data were obtained from Law 12.527/2011 that regulates the constitutional right of access to public information and the k-means algorithm was used, which is a tool for mining multivariate data consolidated in the literature, as the main tool considered in the study. The algorithm was applied to six different values of parameter k (2, 3, 4, 5, 12 and 26); a parameter that is previously required to execute the algorithm, and which concerns the number of clusters or groups. All execution of the algorithm was performed using software R 3.6.1 and recommendations for actions were directed to each of the considered clusters. The formulation of the actions was based on an investigation of specific documents, such as the National Plan for Public Security, National Plan on Drugs and others provided by the Secretariat for Policy on Prevention of Violence and Drugs and the Secretariat for Social Defense of Pernambuco. From the results, it was realized that the use of a clustering approach to municipalities provides greater effectiveness in directing actions to combat and prevent crimes, given the fact that those municipalities that present greater similarities in terms of refers to crime.

Keywords: Public policies. Public security. Cluster. K-means algorithm.

LISTA DE ILUSTRAÇÕES

Figura 1 –	Estrutura da Dissertação.....	15
Figura 2 –	Pirâmide do conhecimento.....	16
Figura 3 –	Etapas do KDD.....	17
Figura 4 –	Interdisciplinaridade da mineração de dados	19
Mapa 1 –	Fragile Index State.....	23
Mapa 2 –	Infográfico das taxas de MVI referentes a 2018.....	24
Fluxograma 1 –	Etapas do processo de armazenamento em cluster.....	27
Figura 5 –	Categorização dos principais métodos de agrupamento	28
Mapa 3 –	Mapa do Brasil de acordo com a qualidade dos dados da criminalidade	36
Quadro 1 –	Pseudocódigo para o k-means	37
Gráfico 1 –	Gráfico do relacionamento entre variáveis e componentes principais.....	40
Gráfico 2 –	Localização de municípios de acordo com a incidência de crimes.....	40
Gráfico 3 –	Clusters 2018 (k=2, 3, 4 e 5).....	42
Mapa 4 –	Regiões de desenvolvimento do estado de Pernambuco.....	45
Gráfico 4 –	Clusters 2018 (k=12).....	45
Gráfico 5 –	Clusters 2018 (k=26).....	45
Mapa 5 –	Mapa do estado segregado por clusters.....	50

LISTA DE TABELAS

Tabela 1 –	Trabalhos sobre análise de crimes através da mineração de dados.....	20
Tabela 2 –	Áreas Integradas de Segurança do estado de Pernambuco.....	25
Tabela 3 –	Caracterização e descrição das variáveis.....	38
Tabela 4 –	Proporção de variação explicada por componentes.....	39
Tabela 5 –	Municípios que compõem os clusters 2018 (k=2 e k=3).....	43
Tabela 6 –	Municípios que compõem os clusters 2018 (k=4).....	43
Tabela 7 –	Municípios que compõem os clusters 2018 (k=5).....	44
Tabela 8 –	Ações por tipo de crime.....	52
Tabela 9 –	Ações por tipo de crime.....	53
Tabela 10 –	Crimes representantes dos clusters.....	54

SUMÁRIO

1	INTRODUÇÃO.....	11
1.1	Objetivos.....	13
1.2	Justificativa.....	13
1.3	Estrutura da Dissertação.....	14
2	BASE CONCEITUAL E REVISÃO DA LITERATURA.....	16
2.1	O processo de descoberta de conhecimento em base de dados.....	16
2.2	Mineração de Dados.....	17
2.2.1	Mineração de dados no contexto da segurança pública.....	19
2.3	Segurança Pública.....	21
2.4	A segurança pública no estado de Pernambuco.....	24
2.5	Análise de Cluster.....	26
2.5.1	Algoritmos por particionamento.....	29
2.6	O algoritmo K-means.....	30
2.6.1	Medidas de similaridade ou distância (dissimilaridade).....	31
2.7	Análise de Componentes Principais – PCA.....	32
3	DELINEAMENTO METODOLÓGICO.....	35
3.1	Obtenção dos dados.....	35
3.2	Obtenção dos resultados.....	37
4	APLICAÇÃO DA TÉCNICA.....	38
4.1	Obtenção dos clusters.....	38
4.1.1	Análise de Componentes Principais – Criminalidade.....	39
4.1.2	Clusters.....	42
4.2	Direcionamento das Políticas de Segurança Pública.....	51
5	CONCLUSÕES.....	55
5.1	Sugestões para trabalhos futuros.....	56
	REFERÊNCIAS.....	57
	APÊNDICE A – DADOS DE CRIMINALIDADE MUNICIPAIS (2018).....	64
	APÊNDICE B – COEFICIENTES DE PONDERAÇÃO DAS COMPONENTES PRINCIPAIS.....	69

1 INTRODUÇÃO

A complexidade da tomada de decisão na segurança pública tem se demonstrado crescente, o que é natural em vista dos diversos aspectos que influenciam decisões deste âmbito, sendo desde aspectos relacionados a questões políticas, assim como, econômicas, socio culturais, entre outros. De acordo com o 13º Anuário Brasileiro de Segurança Pública, publicado em 2019 pelo Fórum Brasileiro de Segurança Pública, houve 57.341 mortes violentas intencionais no Brasil em 2018, o que corresponde a aproximadamente 157 mortes por dia.

Indicadores como o *Global Peace Index* refletem a crise que está associada aos problemas de segurança pública enfrentados pelo país, tendo este ocupado a 106º posição em um ranking com 163 países, na última pesquisa realizada até a data presente deste trabalho. Num período de 30 anos, de 1980 a 2010, mais de 1 milhão de homicídios foram registrados no Brasil, demonstrando uma magnitude espantosa para um país que atualmente não vive nenhum conflito étnico, religioso e/ou territorial (WAISELFISZ, 2012; PEREIRA et al., 2016).

Koegl e Day (2018) argumentam que o crime é considerado caro em quase todos os pontos de vista. Eles citam vários custos públicos e privados que estão associados ao crime, mas que não se limitam a estes, como os custos de perdas tangíveis e intangíveis para as vítimas, dor e sofrimento, perda de produtividade e perda de custos de oportunidade, prêmios de seguro, segurança privada, atividades de vigilância e prevenção ao crime, além de custos relacionados a morte, abuso, policiamento, tribunais, administração de justiça e etc.

Esse problema envolve diversas causas; algumas delas comentadas no 11º Anuário Brasileiro de Segurança Pública (2017) são as falhas no sistema prisional, ineficiência na alocação de recursos financeiros, redução do número de policiais e corrupção na gestão de políticas públicas. No cenário do estado de Pernambuco, de Lima et al. (2002) sugerem que a violência está associada às altas taxas de analfabetismo, desemprego, concentração de alta renda e desigualdade social. Pereira et al. (2016) complementam informando que embora o estado de Pernambuco tenha experimentado um aumento significativo nos últimos anos em seus indicadores socioeconômicos, a taxa de homicídios ainda é significativamente superior à média nacional.

A violência é uma questão alarmante no estado de Pernambuco, mesmo diante do fato em que quedas abruptas nas taxas criminais vem se intensificando desde 2017/2018, não somente nesta unidade federativa, mas na maioria dos estados brasileiros. Não é o caso aqui entrar em detalhes acerca das possíveis causas destas reduções, visto que especialistas comentam que não há de fato uma explicação precisa e que apesar dos avanços na segurança, como a formalização

do Sistema Único de Segurança Pública (SUSP), é difícil identificar um conjunto de medidas federais significativas que possam ter impactado a dinâmica criminal.

Dados oficiais do Fórum Brasileiro de Segurança Pública mostraram que em 2017 houve um surto de homicídios em Pernambuco com um crescimento de quase 40%. Esse surto não se limitou a crimes contra a vida, também foi observado em relação àqueles crimes considerados como crimes contra a propriedade; um exemplo é o caso de roubos a veículos, em que foi constatado um aumento expressivo de 69%. De acordo com Weisburd e Amram (2014), é preciso investigar não apenas criminosos crônicos, mas também lugares crônicos. Podendo ser tratado como crônica aquela região em que, dado um específico tipo de crime, apresente números significativamente superior a outras regiões. Dessa forma, uma região pode ser crônica em um específico tipo de crime, mas pode ser pacífica em outro.

Farias et al. (2008) comentam que questões ligadas ao agrupamento de municípios são de suma importância na análise da criminalidade, ainda que não se considere o padrão espacial, pois, esses agrupamentos podem revelar algumas características, como por exemplo, quais grupos apresentam maior intensidade de crimes violentos, orientando assim, a formação de políticas públicas.

A partir de uma investigação visando identificar lugares crônicos é possível direcionar ações de combate e prevenção ao crime de uma forma mais adequada e é nesse contexto em que o presente trabalho surge visando abordar uma maneira de agrupar os municípios do estado de Pernambuco. É utilizada uma abordagem de aprendizagem não supervisionada, mais precisamente o algoritmo *k-means*, que se trata de uma técnica difundida na literatura, a fim de direcionar de maneira assertiva ações que visam combater e prevenir os crimes de maior incidência para cada um dos *clusters* obtidos.

1.1 Objetivos

O objetivo geral do trabalho é utilizar uma abordagem de aprendizagem não supervisionada para o estabelecimento de *clusters* entre os municípios pernambucanos, considerando diversas variáveis representativas da criminalidade no estado, com o intuito de auxiliar o direcionamento de ações para prevenção e combate ao crime a fim de apoiar os formuladores de políticas de segurança pública. De modo sumário, para atingir o objetivo geral, os objetivos específicos a serem alcançados são:

- ✓ Demonstrar e discutir o problema da criminalidade, em especial, no estado de Pernambuco;
- ✓ Obter um agrupamento dos municípios do estado a partir do algoritmo de agrupamento *K-Means*;
- ✓ Avaliar, interpretar os resultados obtidos e demonstrar o direcionamento de ações de combate e prevenção de crimes a partir das características semelhantes de violência dos diferentes *clusters*.

1.2 Justificativa

A exigência de uma maior atenção de pesquisadores e formuladores de políticas na área da segurança pública é algo urgente, visto que a violência tem consequências onerosas e prejudiciais para os indivíduos e a sociedade como um todo. No cenário brasileiro, atualmente, as regiões Norte e Nordeste apresentam-se no topo do ranking da violência. O Atlas da Violência – retrato dos municípios brasileiros 2019, elaborado em parceria com o Fórum Brasileiro de Segurança Pública – mostra que dos 20 municípios mais violentos do Brasil, 18 estão no Norte-Nordeste, além de identificar que há uma heterogeneidade na prevalência da violência letal nos municípios. Diante desse pensamento, é possível utilizar do argumento de Waiselfisz (2014) que comenta sobre a necessidade de se reexaminar urgentemente as práticas empregadas nessa conjuntura, na tentativa de implementar novos modelos dentro de uma realidade que comprovadamente não encontra resolutividade em si mesmo.

Em todo o país, há uma escassez de condutas gerenciais associada a uma cultura organizacional que se demonstra falha na defesa da sociedade, atreladas a projetos de prevenção e combate à criminalidade. Como comentam de Lima et al. (2016), a história da segurança pública no Brasil tem sido marcada por demandas acumuladas e mudanças incompletas, confirmando que há a ausência de um projeto de governança de alinhamento das políticas de segurança pública aos requisitos da democracia e à garantia dos direitos humanos.

O estado de Pernambuco tem um certo destaque nacional a respeito do combate ao crime contra a vida devido a implementação da política pública Pacto Pela Vida (PPV-PE), que surgiu no ano de 2007 e vivenciou alguns momentos de significativa redução de homicídios, até momentos de retomada do crescimento das taxas de homicídios, como em 2015 e 2016.

Diversos estudos foram realizados a fim de avaliar o programa PPV-PE; um destes, destacado aqui, foi o de Cruz (2014) em que ele chama atenção ao fato de que, mesmo com uma redução do número de homicídios em Pernambuco no período de 2007 a 2010, houve um aumento do número de outros tipos de crime, em algumas regiões, informando que o tráfico de entorpecentes é um exemplo destes (com um aumento de 496%, nesta época). Esse fato alimenta a força do pensamento crítico a respeito da prática do direcionamento das políticas públicas para cada microrregião, surgindo alguns questionamentos associados a efetividade destas. Alguns autores já estudaram o fato de que muitas soluções de políticas públicas implicam ações locais, com foco no território. Contudo, não foram encontrados estudos que utilizem uma maneira de agrupar os municípios, em específico do estado de Pernambuco, a fim de direcionar políticas públicas visando as características das regiões no que diz respeito a incidência das categorias dos crimes.

O estado de Pernambuco é referência na coleta de dados atrelados a criminalidade, devido as melhorias significativas considerando novos procedimentos e incorporando mais tecnologia ao processo de coleta. Sendo assim, os dados são considerados confiáveis e mais precisos que os dados relatados pelo sistema de saúde (PEREIRA et al., 2016; SAURET, 2012).

Como afirmam Agarwal et al. (2013), não há um campo de estudo tão apropriado como a mineração de dados para realizar análise de crimes. Dispondo a abordagem de análise de cluster, dentro da mineração, a obtenção de uma dimensão de dados reduzida e permitindo um agrupamento dos municípios do estado de forma a auxiliar a extração de conhecimentos para apoiar os formuladores de políticas de segurança num direcionamento de ações, mais assertivo. Uma das técnicas de análise de cluster é o algoritmo *k-means*, que além de ser considerado prático, simples e ser difundido na literatura, possui diversos pacotes em ferramentas estatísticas e de mineração rápida, como é o caso do *software R*; o que facilita o desenvolvimento do estudo.

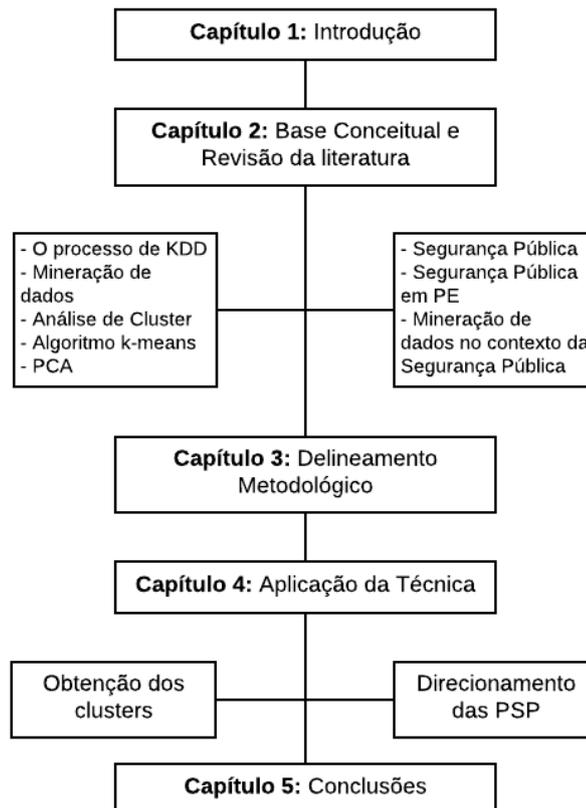
1.3 Estrutura da Dissertação

Além da introdução, o trabalho é dividido em mais quatro capítulos. O Capítulo 2 fornece uma estrutura envolvendo conceitos e revisão da literatura dos principais temas aqui tratados, como: o processo de conhecimento em base de dados, a mineração de dados no contexto da

segurança pública, a própria segurança pública, em especial a do estado de Pernambuco e a análise de cluster nesse cenário.

Em sequência, no Capítulo 3 é realizado um delineamento metodológico que constitui em um detalhamento acerca dos aspectos metodológicos utilizados para desenvolver o estudo, assim como é identificada a fonte de dados utilizada, fornecendo qual foi o esquema percorrido para seleção dos dados e obtenção dos resultados. No Capítulo 4 é fornecido todo o resultado advindo da aplicação do modelo proposto, além do acesso à discussão gerencial de direcionamento das políticas de segurança. Por fim, no Capítulo 5 são apresentadas as conclusões críticas acerca da aplicação da técnica e resultados obtidos, bem como as limitações e possíveis evoluções que podem ser tidas como sugestões para trabalhos futuros. O fluxograma, disposto na figura 1, fornece a estrutura da dissertação de forma a facilitar o entendimento.

Figura 1 - Estrutura da Dissertação



Fonte: O autor (2019)

2 BASE CONCEITUAL E REVISÃO DA LITERATURA

Este capítulo tem como intuito fornecer uma estrutura conceitual e referencial acerca das tomadas de decisões com o auxílio de uma abordagem não supervisionada, no contexto da segurança pública e conseqüentemente prover um suporte para o melhor entendimento do tema geral do trabalho. O capítulo inicia-se com uma relevante conceituação e revisão da literatura a respeito das técnicas gerais, aqui utilizadas, que visam fornecer os resultados. Em seqüência, conceitos sobre o assunto de segurança pública são analisados, tratando particularmente do cenário pernambucano. E por conseguinte, são abordadas uma base conceitual e uma revisão da literatura sobre a análise de cluster, oferecendo uma maior ênfase ao algoritmo *k-means* e suas peculiaridades.

2.1 O processo de descoberta de conhecimento em base de dados

O volume de dados armazenados em computadores tem tido um aumento significativo nos últimos anos e junto a isto uma capacidade imensa de geração de informações e consequentes conhecimentos. Segundo Safhi et al. (2019), o crescimento exponencial dos dados abriu uma gama de novas oportunidades e fez com que os dados se tornassem uma matéria-prima de extrema importância, considerada para muitas organizações como a mais valiosa. De forma sucinta, pode-se afirmar que os dados podem ser convertidos em informações, onde são identificadas mensagens consideradas significativas e posteriormente em conhecimento, oferecendo um apoio mais que relevante para a tomada de decisão. Diante da pirâmide do conhecimento, disposta na figura 2, é possível se ter uma ideia de como o fenômeno da considerada gestão do conhecimento ocorre.

Figura 2 - Pirâmide do conhecimento



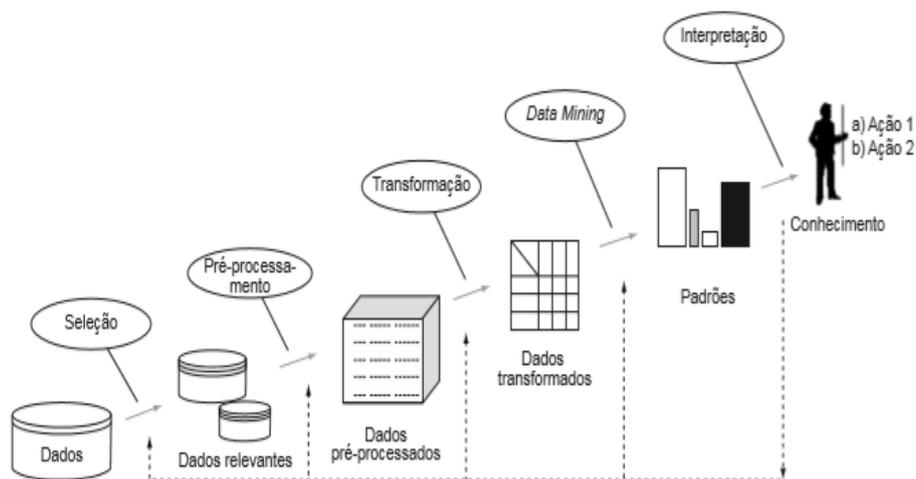
Fonte: Adaptado de Bernstein (2009)

Sob essa enorme quantidade de dados surge, então, a necessidade do uso de uma metodologia capaz de auxiliar a descoberta por conhecimentos. E é neste contexto em que é desenvolvida a KDD – *Knowledge Discovery in Databases*, conhecida também como processo

de descoberta de conhecimento em base de dados. De acordo com Fayyad et al. (1996) esse processo pode ser representado em seis etapas, vistas na figura 3, e descritas abaixo:

- **Seleção:** refere-se à seleção do conjunto de dados para a realização do processo de mineração. Dados esses que possuem inúmeras características e podem ser obtidos de diferentes bases de dados;
- **Pré-processamento:** basicamente nessa etapa é realizada uma separação dos dados, descartando aqueles considerados incompletos ou redundantes;
- **Transformação:** nessa fase os dados são transformados através de específicos processos a fim de organizá-los da forma adequada para que possam ser interpretados;
- **Mineração de dados (*Data Mining*):** é nessa etapa em que é realizada a geração de informação, onde são usadas técnicas computacionais com o intuito de se identificar possíveis padrões, anomalias e correlações;
- **Interpretação:** como última etapa, a partir das informações descobertas na etapa de mineração, é feita a interpretação das informações obtidas, possibilitando a descoberta de conhecimento.

Figura 3 - Etapas do KDD



Fonte: Fayyad et al. (1996)

2.2 Mineração de Dados

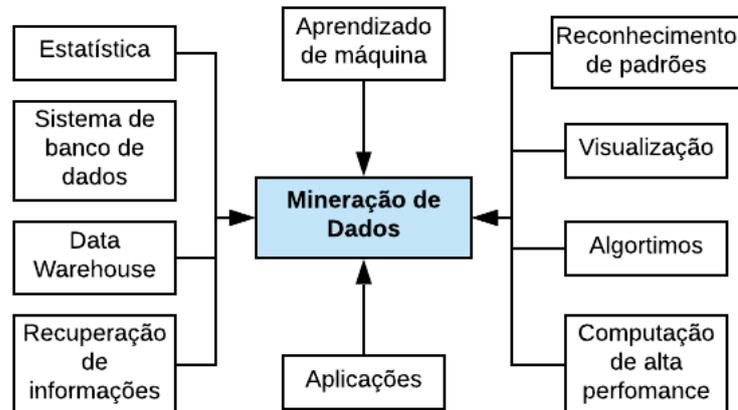
A mineração de dados, ou *data mining*, é considerada como uma forma de extrair conhecimento de um conjunto de dados geralmente grande. Trata-se de um subdomínio da inteligência artificial e pode apresentar descobertas, a partir dos dados, de forma abrangente para o usuário (KEYVANPOUR et. al, 2011; SCHUH et. al, 2019).

De acordo com Turban et al. (2005), a mineração de dados é um processo que utiliza técnicas estatísticas, matemáticas, de inteligência artificial e de aprendizado de máquina com o objetivo de extrair e identificar informações úteis e conhecimentos subsequentes de grandes bancos de dados. Eles argumentam que com a queda drástica nos custos, a relação custo/desempenho dos sistemas de computadores permitiu que muitas organizações começassem a aplicar algoritmos complexos de técnicas de mineração de dados, sendo cada classe de aplicativo de mineração suportada por um conjunto de abordagens algorítmicas para extrair relacionamentos relevantes nos dados. Dessa forma, algumas das classes ou tarefas são as seguintes:

- Classificação: trata-se de métodos que envolvem a propagação de um conjunto de dados com um conjunto conhecido de classes e é conhecido como uma abordagem supervisionada de dados;
- Clusterização (agrupamento): é uma abordagem não supervisionada de dados, ou seja, é uma abordagem que nenhuma característica predefinida é obtida, onde são identificados grupos de itens que compartilham uma característica específica;
- Associação: basicamente, trata-se de ferramentas que buscam encontrar uma relação entre dois ou mais itens, geralmente do mesmo tipo;
- Sequenciamento: são consideradas técnicas semelhantes as técnicas de associação, mas o relacionamento entre os itens ocorre ao longo de um período de tempo, ou seja, são métodos úteis para identificar ocorrências regulares de eventos semelhantes e/ou tendências;
- Regressão: técnicas utilizadas para mapear dados para um valor de predição, explorando e inferindo a relação entre uma variável com variáveis independentes específicas.

Ainda sobre a mineração de dados, de uma forma geral, Han et. al (2011) dizem que é um campo considerado interdisciplinar e que faz uso de conceitos de estatística, inteligência artificial, aprendizado de máquina, reconhecimento de padrões, inteligência artificial, sistema de recuperação de informação, entre tantas outras áreas de conhecimento, conforme ilustrado na figura 4.

Figura 4 - Interdisciplinaridade da mineração de dados



Fonte: Adaptado de Han et. al (2011)

2.2.1 Mineração de dados no contexto da Segurança Pública

Segundo Alves et. al (2018), compreender as causas dos crimes é uma questão de longa data na agenda de um pesquisador. Diversas técnicas são apresentadas na literatura com o objetivo de auxiliar não só na compreensão das causas dos crimes, como também na compreensão do padrão espacial destes. Wortley e Mazarolle (2008) comentam que perceber o papel do ambiente e compreender como se dá o padrão espacial do crime é primordial para o seu controle e investigação.

Oatley e Ewart (2003) realizaram um projeto no qual foi desenvolvido um sistema de software, que utiliza ferramentas de mapeamento e visualização, onde os métodos estatísticos empregados e as tecnologias de mineração de dados são usados para investigar o impacto dos tipos de evidência disponíveis e determinar a causalidade desse domínio. Já Qazi e Wong (2019) propuseram um esquema de descoberta de conhecimento centrado no homem e de mineração de dados para mineração de texto de crime, onde, de acordo com os autores, é um esquema capaz de extrair associações plausíveis entre crimes, identificando padrão de crime, agrupando crimes semelhantes, provocando rede de co-ofensores e lista de suspeitos com base na similaridade espacial-temporal e comportamental. No caso de Agarwal et. al (2013) foi realizado um projeto que se concentrou na análise dos crimes, implementando um algoritmo de agrupamento, nesse caso o *k-means*, no conjunto de dados de crimes usando ferramenta de mineração rápida e foram obtidas conclusões a respeito da tendência de alguns tipos de crimes ao longo dos anos.

Diversos outros trabalhos são encontrados na literatura utilizando a mineração de dados e abordando o tema em diferentes perspectivas. Dessa forma, com o intuito de demonstrar esse

uso, num âmbito geral, foi realizada uma pequena revisão da literatura. Além dos trabalhos já citados, foram selecionados dez trabalhos que demonstraram uma linha de pesquisa mais alinhada com o conteúdo aqui abordado, no portal de períodos da CAPES.

Na tabela 1 estão dispostos os autores, o ano de publicação e o objetivo de cada um dos trabalhos.

Tabela 1 - Trabalhos sobre análise de crimes através da mineração de dados

Autores, Ano	Objetivo
Qazi et al. (2019)	Propor um esquema de descoberta de conhecimento centrado no homem e de mineração de dados para mineração de texto de crime capaz de extrair associações plausíveis entre crimes, identificando padrão de crime, agrupando crimes semelhantes, provocando rede de co-ofensores e lista de suspeitos com base na similaridade espacial-temporal e comportamental.
Das e Das (2019)	Propor uma técnica de agrupamento baseada em gráficos para descobrir rótulos de relatórios de crimes com base em paráfrases extraídas de grandes corpos de crimes sem identificação.
Kirgel (2016)	Agrupar uma amostra de 190 países utilizando sete dimensões de crimes cibernéticos, além de medidas de PIB e uso da internet através de uma análise de cluster.
Khan et al. (2019)	Propor um modelo para o sistema de policiamento preditivo para crimes de rua na região de Carachi-Paquistão através da ferramenta <i>k-means</i> e de métodos bayesianos.
Hong et al. (2017)	Apresentar um sistema de apoio à decisão para a vinculação de crimes com base em várias características, inclusive comportamentais, de casos criminais. Sua técnica subjacente é a classificação aos pares com base na similaridade de recursos para calcular as semelhanças aos pares e construir um classificador para determinar se um par de casos deve pertencer a uma série.
Sjödin et al. (2018)	Identificar grupos com diferentes padrões de violência por parceiro íntimo (VPI) entre jovens infratores suecos condenados
Wang et al. (2018)	Propor um método de pistas atadas para obter resultados refinados a respeito do fenômeno conhecido na análise de crimes chamado de efeito de quase repetição, adotando uma perspectiva de ciência de dados, combinando coeficiente de correlação, cluster hierárquico e mineração de padrões de frequência em uma ordem específica.
Curman et al. (2015)	Realizar uma análise longitudinal de um conjunto de dados de 16 anos usando o segmento de rua como a unidade de análise, evidenciando as tendências dos crimes. E utilizando o modelo de trajetória baseado em grupos, através da técnica de análise de cluster <i>k-means</i> .
Das et al. (2019)	Propor um agrupamento adaptativo incremental em grupo de relatórios de crimes, onde este é adaptado para fornecer conjuntos atualizados de clusters. Além de realizar comparações com algoritmos de <i>clustering</i> para expressar a sua eficácia e significância estatística.
Tayal et al. (2014)	Propor uma abordagem para o design e implementação de detecção de crimes e identificação criminal para cidades indianas usando técnicas de mineração de dados

Fonte: O autor (2019)

2.3 Segurança Pública

Conforme prevê a Constituição Federal de 1988 (CF/1988), a definição constitucional de Segurança Pública, disposta no art. 144, se apresenta nos seguintes termos:

Art. 144. A segurança pública, dever do Estado, direito e responsabilidade de todos, é exercida para a preservação da ordem pública e da incolumidade das pessoas e do patrimônio, através dos seguintes órgãos:

I - polícia federal;

II - polícia rodoviária federal;

III - polícia ferroviária federal;

IV - polícias civis;

V - polícias militares e corpos de bombeiros militares.

[...]

O direito à Segurança Pública está interligado a vários outros direitos fundamentais. Segundo a Declaração Universal dos Direitos Humanos, em seu artigo nº III, promulgada na Assembleia Geral das Nações Unidas em 10 de dezembro de 1948, é estabelecido que “todo ser humano tem direito à vida, à liberdade e à segurança pessoal”. Tendo em vista a existência de uma efetiva prestação de segurança, de fato, é proporcionado ao ser humano uma maior garantia de vida, de liberdade, de segurança pessoal, assim como de propriedade. Diante disso, Buonamici (2011), afirma que o direito à segurança pública sempre esteve ligado a história da própria humanidade, presente em qualquer espécie de agrupamento humano. Ele argumenta que os integrantes destes agrupamentos sempre tiveram a necessidade de proteção social a fim de garantir a paz e a harmonia na convivência social. Entende-se, então, a segurança como uma componente chave para a convivência social, já que, de certa forma, estabelece uma salvaguarda da espécie humana.

Muitos autores tem buscado entender como a criminalidade reage frente a intervenções políticas visando estabelecer uma alocação eficiente de recursos. Um dos primeiros modelos econômicos, que foi proposto originalmente por Gary Becker em 1968, buscou analisar quais seriam as quantidades ótimas de recursos dispendidos e punições a serem adotadas para as leis serem cumpridas. Para Becker (1968), um crime é cometido se o benefício esperado pelo indivíduo com esse ato for maior que o de outras atividades, elencando fatores como a probabilidade de detenção, a severidade de punição, a renda obtida no caso da prática de atividades lícitas, entre outros fatores. Seguindo na mesma linha de raciocínio que Becker, Jawadi et al. (2019) comentam que países com maior taxa de desemprego podem ter maior taxa de criminalidade.

Giddens (2004) comenta que umas das primeiras tentativas de explicação do crime tinham caráter essencialmente biológico, concentrando-se nas qualidades inatas dos indivíduos como fonte de crime. Além disso, ele também comenta sobre as teorias psicológicas do crime, sendo

estas, concentradas nos tipos de personalidades. Logo, entende-se que as justificativas a respeito da violência e do crime estão traçadas sobre diferentes perspectivas. De fato, explicar as causas associadas a violência e o crime não é tarefa fácil. O crime não é singular e o seu universo é heterogêneo (SOARES, 2006).

No Brasil, o acelerado crescimento da violência pode ser observado cotidianamente. A soma desse vasto crescimento com a incapacidade do Estado em promover políticas de controle da criminalidade fez com que questões sobre segurança pública se tornassem uma das principais preocupações da sociedade brasileira, colocando o tema como central no debate eleitoral do ano de 2018, como é afirmado no documento “O Novo Sistema Único de Segurança Pública”, publicado pelo Fórum Brasileiro de Segurança Pública, em 2019.

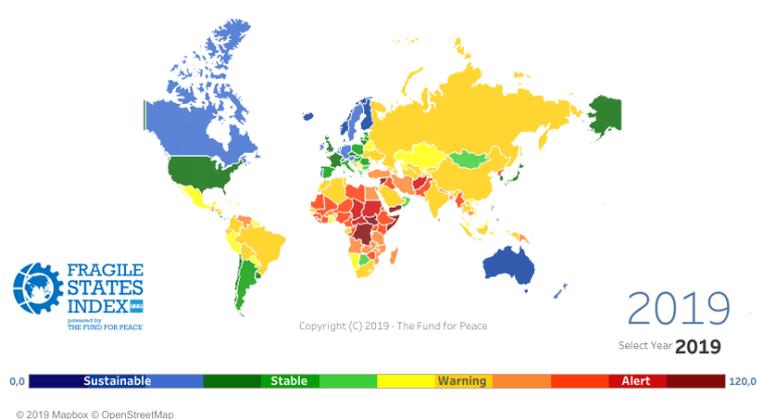
A criminalidade no país pode ser destacada através de indicadores, um destes, o *Global Peace Index*, que foi desenvolvido pela revista *The Economist* em parceria com a Universidade de Sidney, Universidade de Londres, Universidade de Uppsala e o Instituto Internacional de Pesquisas pela Paz de Estocolmo, apresenta o Brasil na 116ª posição em um ranking contendo 163 países. É notória a necessidade de atenção no que diz respeito à segurança pública no Brasil e não é de hoje que a criminalidade está no centro de debate. Além disso, as altas taxas de criminalidade têm influência direta sobre a imagem do país e sua reputação internacional.

No mapa 1, onde os países são apresentados de acordo com sua fragilidade, através do índice conhecido como *fragile states index*, que é publicado pelo Fundo de Paz dos Estados Unidos e pela revista americana *Foreign Policy* desde 2005, o Brasil é visto como um país classificado como *elevated warning*, ou seja, basicamente é um estado avaliado com um *score* alto em relação a sua vulnerabilidade ao conflito ou colapso. São levados em consideração doze indicadores de risco de conflito para medir a condição de um estado a qualquer momento. Dentro desses indicadores são levados em conta segurança, serviço público, direitos humanos, desenvolvimento econômico, queixa do grupo, entre outros aspectos. São todos aspectos que, de certa forma, estão intimamente relacionados uns com os outros. E sendo assim, não se pode ser descartado o fato de que a segurança pública tem forte impacto sobre o dito índice.

Diversos autores realizaram estudos a fim de entender qual/quais as maiores causas das altas taxas de criminalidade no país. Dos Santos e Kassouf (2013) mostraram que o crime letal tem uma conexão negativa com o aprimoramento da lei, da mesma forma que está positivamente correlacionado com o desemprego. Freitas e Freitas (2017) defendem que muitas são as causas apontadas para a origem e o crescimento da violência em sociedade, citando a condição de pobreza e miserabilidade, a desigualdade social, bem como o acesso e o uso de armas de fogo, como possíveis causas. Já Ingram e Costa (2019) realizaram um estudo que

demonstra um relato político geográfico da violência letal no Brasil, oferecendo uma análise estatística do efeito desigual e geograficamente variável de preditores políticos de violência nos municípios brasileiros, enquanto controla contas socio estruturais dominantes. Segundo os dois últimos autores citados, os três principais partidos políticos brasileiros afetam a violência nos municípios de maneira diferentes, mostrando que partidos políticos, políticos e o processo político também contribui para a incidência da violência.

Mapa 1 - Fragile Index State



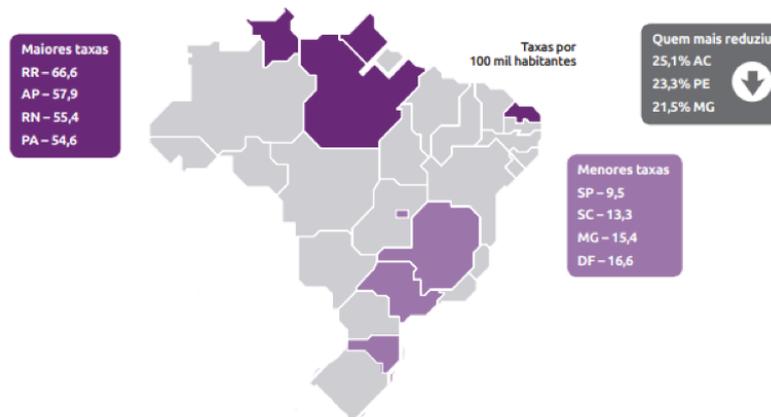
Fonte: Adaptado de Fund For Peace (2019)

Assim como em qualquer governo que tem como uma das pautas a segurança, o governo brasileiro tenta investir na área, de forma efetiva, a fim de lidar com o crescente número de crimes que, além de tudo, tem impacto direto sobre a economia. De acordo com o infográfico apresentado no mapa 2, é possível perceber que houve uma redução significativa dos crimes violentos intencionais nos estados do Acre, Pernambuco e Minas Gerais, crimes esses que são compostos por homicídio doloso, latrocínio, lesão corporal seguida de morte e mortes decorrentes de intervenções policiais em serviço e fora. Como sugerem especialistas da área de segurança, não há um fator único que seja capaz de justificar as reduções. Algumas ações pontuais podem ser consideradas, como por exemplo, o aumento do efetivo de policiais no estado de Pernambuco no início de 2018. Porém, é necessário a realização de um trabalho muito mais minucioso no que diz respeito a identificar quais os fatores que levaram a determinada redução, para que assim sejam replicadas tais ações.

Uma das ações mais atuais realizadas pelo governo brasileiro foi a criação do Sistema Único de Segurança Pública (SUSP), que foi implantado pela Lei nº 13.675/2018, sancionada em 11 de junho. Esse sistema tem a capacidade de desenvolver uma governança no campo da segurança pública através da padronização dos dados, integração tecnológica, de inteligência e

operacional, abrindo uma real possibilidade de finalmente ser alcançado uma Política Nacional de Segurança Pública (MINISTÉRIO DA SEGURANÇA PÚBLICA, 2018).

Mapa 2 - Infográfico das taxas de MVI referentes a 2018



Fonte: Adaptado de FBSP (2019)

2.4 A segurança pública no estado de Pernambuco

O Brasil é composto por 27 unidades federativas, dentre estas, está o estado de Pernambuco. Localizado no centro-leste da região nordeste, que é a terceira maior região do país em extensão territorial, estimada em 1.554.000 km² e segunda mais populosa, com cerca de 50 milhões de habitantes (IBGE, 2010).

Mesmo diante das melhorias socioeconômicas de toda a região Nordeste, Pernambuco é destacado como uma das Unidades Federativas com mais altos índices de criminalidade. De acordo com dados da Secretaria de Defesa Social do estado de Pernambuco (SDS-PE), apenas durante o horizonte de tempo entre janeiro a outubro de 2019, houve 2881 vítimas de crimes violentos letais intencionais, além de 67382 ocorrências de crimes violentos contra o patrimônio registradas. Como destacado por Nóbrega Júnior (2014), nos últimos dez anos anteriores a sua pesquisa, a segurança pública em Pernambuco, como serviço prestado pelo estado, passou a ser destaque diante de um quadro de violência exacerbada, onde antes disso, era vista como exclusividade das polícias que, muitas vezes, não dialogavam com o governo. Foi então, a partir de pressão social por segurança que houve alterações no serviço provocadas por mudanças no conceito de gestão pública.

A partir da portaria GAB/SDS nº 240/2004, foram definidos os limites territoriais e o quantitativo de áreas e circunscrições de segurança, com a proposta de fortalecer as políticas de defesa social do estado (PERNAMBUCO, 2004). Atualmente, os territórios dividem-se em oito, sendo eles: capital, região metropolitana, zona da mata norte, zona da mata sul, agreste 1,

agreste 2, sertão 1 e sertão 2. Cada um desses territórios é composto por Áreas Integradas de Segurança (AIS), que podem ser vistas a partir da tabela 2. O território agreste 1, por exemplo, é composto pela AIS 14 – Caruaru e a AIS 17 – Santa Cruz do Capibaribe. Além disso, cada uma dessas AIS é composta por específicas circunscrições, sejam elas, bairros e/ou municípios e tem suas divisões baseadas nas proximidades das regiões. Como disposto no decreto nº 1197, de 11 de junho de 2010, o limite entre os territórios, áreas e circunscrições consideram os critérios técnicos-cartográficos fornecidos por instituições como o Instituto Brasileiro de Geografia e Estatística, a Agência Estadual de Planejamento e Pesquisas de Pernambuco, entre outras.

Tabela 2 - Áreas Integradas de Segurança do estado de Pernambuco

Território	AIS	Nomenclatura da AIS
Capital	1	Santo Amaro
	2	Espinho
	3	Boa Viagem
	4	Várzea
	5	Apipucos
Região Metropolitana	6	Jaboatão dos Guararapes
	7	Olinda
	8	Paulista
	9	São Lourenço da Mata
Zona da Mata Norte	10	Cabo de Santo Agostinho
	11	Nazaré da Mata
Zona da Mata Sul	12	Limoeiro
	13	Vitória de Santo Antão
Agreste 01	14	Palmares
	15	Caruaru
Agreste 02	16	Santa Cruz do Capibaribe
	17	Belo Jardim
Sertão 01	18	Garanhuns
	19	Arcoverde
	20	Afogados da Ingazeira
	21	Serra Talhada
Sertão 02	22	Floresta
	23	Salgueiro
	24	Ouricuri
	25	Cabrobó
	26	Petrolina

Fonte: O autor (2019)

Para Lopes (2016), essas divisões territoriais foram estabelecidas com o objetivo de distribuir de maneira isométrica a atuação das polícias militar e civil, estimulando a ação integrada de ambas as instituições. Além de permitir que sejam usadas como uma forma de direcionar as políticas de segurança pública.

De acordo com Macedo (2012), no ano de 2007, o estado de Pernambuco reestruturou sua política de segurança pública. Foi então criado um programa com ações que unia os esforços dos principais atores sociais pertencentes à área de segurança no estado, o Pacto Pela Vida (PPV), sendo considerado um “divisor de águas” no estilo de gestão da segurança pública no estado. Segundo Lopes (2016), uma das inovações proposta pelo PPV foi o fortalecimento da gerência de análise criminal e estatística, dentro da Secretaria de Defesa Social (SDS) a partir da produção de informações mais precisas e de boa qualidade capazes de nortear as ações policiais de maneira precisa para a consolidação de diagnósticos seguros.

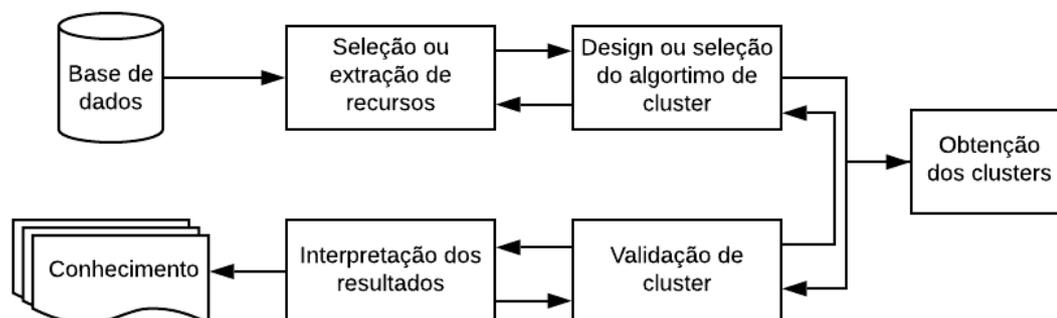
2.5 Análise de Cluster

A análise de cluster, ou de agrupamentos, trata-se de uma técnica clássica de análise de dados que desempenha um papel importante na identificação de estruturas naturais dos conjuntos de dados de destino (ZHU et. al, 2019). É uma técnica considerada distinta da classificação por não ser realizada nenhuma suposição a respeito dos grupos advindos do uso da técnica. Não contando com classes predefinidas, a forma de aprendizado da técnica é considerada não supervisionada. De acordo com Hair et al. (2009), a análise de cluster se assemelha a análise fatorial em seu objetivo de avaliar estrutura, mas diferem no sentido de que a primeira agrega objetos e a segunda está prioritariamente interessada em agregar variáveis. Além disso, a análise fatorial faz os agrupamentos com base em padrões de variação nos dados, enquanto a análise de cluster faz agregados baseados em distância (proximidade).

Segundo Xu e Tian (2015), o *cluster* é considerado a questão mais importante do aprendizado não supervisionado, no entanto, a definição completa para *clustering* ainda não chegou a um acordo. Porém, muitos autores tratam a lógica da similaridade e dissimilaridade como parte de uma clássica definição para *clustering*, onde, é dito que elementos que compõem um mesmo *cluster* devem ser o mais semelhantes possível, da mesma forma que devem ser o mais dissimilares possível de elementos pertencentes a outros *clusters*. Ahmad e Starkey (2017) asseguram que a análise envolve o processo de organização de objetos em grupos naturais, encontrando a classe dos objetos de forma que eles, em uma classe, sejam semelhantes entre si e diferentes dos objetos de outra classe.

Para Xu e Wunschl (2005) o processo padrão de armazenamento em *cluster* pode ser dividido em algumas etapas, vistas no fluxograma 1 e descritas logo abaixo.

Fluxograma 1 - Etapas do processo de armazenamento em cluster



Fonte: Adaptado de Xu e Wunschll (2005)

Etapa 1 – Seleção ou extração de recursos: extração e seleção dos recursos mais representativos do conjunto de dados original;

Etapa 2 – Design ou seleção do algoritmo de cluster: essa etapa é combinada com uma seleção de medida de proximidade e a construção de uma função critério. Quase todos os algoritmos de *clustering* estão conectados a alguma definição de medida de proximidade, descritas no tópico 2.6.1. Logo, uma vez que a medida de proximidade é escolhida, a construção de uma função critério de agrupamento torna a partição de clusters um problema de otimização, que além de ser bem definido matematicamente, apresenta soluções ricas na literatura.

Etapa 3 – Validação de cluster: essa etapa tem uma grande importância devido a alguns pontos como o fato de diferentes abordagens levarem a diferentes aglomerados, os parâmetros e apresentações dos dados de entrada de um determinado algoritmo pode afetar os resultados finais, entre outros pontos, mais detalhados na descrição do algoritmo utilizado. Portanto, padrões efetivos de avaliação e critérios são importantes para fornecer aos usuários um grau de confiança para os resultados de agrupamentos derivados dos algoritmos usados.

Etapa 4 – Interpretação dos resultados: nessa etapa são fornecidas explicações práticas para o resultado do agrupamento. Análises de especialistas, experimentos, entre outras análises, podem ser necessárias para garantir a confiabilidade do conhecimento extraído.

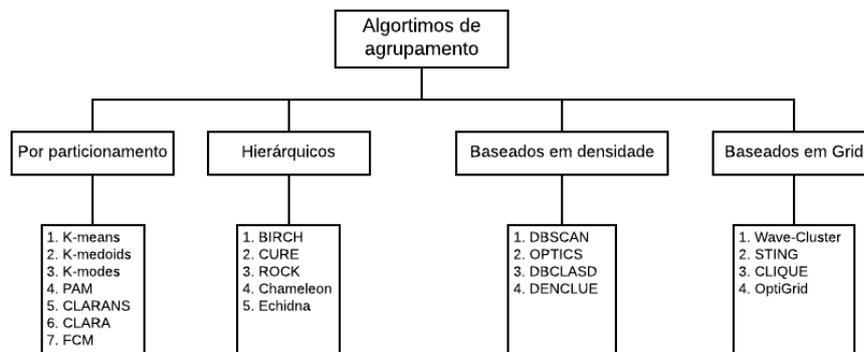
De acordo com Manning, Raghavan e Schütze (2009), são diversas as situações em que a análise de *cluster* pode ser utilizada, como por exemplo, no agrupamento de resultados de uma busca textual, na análise de criminalidade, na organização de documentos, etc. A análise de *cluster* vem atraindo a atenção de muitos pesquisadores em diferentes disciplinas (VELMURUGAN, 2014).

Os algoritmos de *clustering* podem ser classificados de acordo com diversas características, como por exemplo, pelo tipo de resultado a ser produzido e pelas diferentes

formas de definir a proximidade entre um indivíduo e um grupo. Fasulo (1999) fornece uma classificação para os algoritmos em métodos hierárquicos e não hierárquicos.

É difícil fornecer uma nítida categorização desses métodos, devido à sobreposição que pode existir a respeito das categorias, podendo um método ter recursos de várias delas. Porém, é importante apresentar uma imagem relativamente organizada dos métodos de agrupamento. De um modo geral, os principais métodos podem ser categorizados de acordo com a forma dos agrupamentos gerados: em hierárquicos, por particionamento, baseado em densidade e baseado em grid (HAN, KAMBER e PEI, 2012; FAHAD et al., 2014). Na figura 5 podem ser vistos os principais métodos de agrupamento de forma categorizada.

Figura 5 – Categorização dos principais métodos de agrupamento



Fonte: Adaptado de Fahad et al. (2014)

Uma breve definição de cada uma das categorias, de acordo com Han, Kamber e Pei (2012), é a seguinte:

Por particionamento: dado um conjunto de n objetos, um método de particionamento constrói k partições dos dados, onde cada partição representa um *cluster*. A maioria dos métodos de particionamento é baseado em distância e o critério geral utilizado para uma boa partição é que objetos no mesmo *cluster* estejam “próximos” ou relacionados entre si, enquanto objetos em diferentes grupos estejam “distantes” ou muito diferentes. Na seção 2.5.1 são estudados os métodos por particionamento, dando a devida ênfase ao método *k-means*.

Hierárquicos: esses métodos criam uma decomposição hierárquica de determinado conjunto de objetos de dados. Eles podem ser baseados em distância ou baseados em densidade e continuidade. São úteis para o resumo de dados e visualização, por exemplo, na organização dos funcionários de uma grande empresa, os particionando como executivos, gerentes e etc. A divisão dos dados advinda da aplicação do algoritmo hierárquico pode ser feita de duas formas

conhecidas na literatura, a *top-down* e a *botton-up*, também conhecidas como divisivos e aglomerativos.

Baseados em densidade: a ideia geral desses métodos é a de continuar crescendo um determinado *cluster* enquanto a densidade (número de objetos) na “vizinhança” excede algum limite. Por exemplo, para cada ponto de dados em um determinado *cluster*, a vizinhança de um determinado raio deve conter pelo menos um número mínimo de pontos.

Baseados em Grid: tais métodos adotam uma abordagem orientada a espaço, particionando o espaço de incorporação nas células, independentemente da distribuição dos objetos de entrada. Eles quantificam o espaço de dados em um número finito de células que forma uma estrutura em grid e todas as operações de agrupamento são aplicadas nessa estrutura.

2.5.1 Algoritmos por particionamento

De acordo com Han, Kamber e Pei (2012), a versão mais simples e mais fundamental da análise de *cluster* é o particionamento, que organiza os objetos de um conjunto em vários grupos ou grupos exclusivos. Para Goldschmidt, Bezerra e Passos (2015), os algoritmos por particionamento dividem o conjunto de dados em k grupos.

Inicialmente, tais algoritmos escolhem arbitrariamente k objetos como sendo o centro de k grupos e então os objetos são divididos entre os k grupos de acordo com uma medida de similaridade adotada previamente, de forma que cada um deles seja incluso no grupo que forneça o menor valor de distância entre o objeto e o centro do referido grupo.

Formalmente, dado um conjunto de dados, D , de n objetos, e k o número de clusters a serem formados, um algoritmo de particionamento organiza os objetos em k partições ($k \leq n$), em que cada partição representa um cluster, sendo esses clusters formados para otimizar um critério de particionamento objetivo, como uma função de dissimilaridade com base na distância, para que os objetos dentro de um cluster sejam “semelhantes” um ao outro e “diferentes” aos objetos de outros clusters em termos de atributos do conjunto de dados (HAN, KAMBER e PEI, 2012).

Segundo Pacífico e Ludermir (2014), cada cluster é representado pelo seu vetor centroide e então um algoritmo tenta minimizar uma função de critério por meio de um processo iterativo, no qual todos os centroides são atualizados na tentativa de melhorar a qualidade dos clusters finais. De acordo com Han, Kamber e Pei (2012), o método de particionamento mais conhecido e mais usado é o *k-means*. Onde, segundo Ayed, Halima e Alimi (2014), é o mais utilizado devido a sua simplicidade de implementação e sua eficácia.

2.6 O algoritmo *K-Means*

De acordo com Wu et. al (2009), o algoritmo *k-means* é um método iterativo simples para particionar um determinado conjunto de dados em um número especificado de clusters k . Foi proposto inicialmente em um trabalho de Lloyd em 1957, porém, apenas em 1982 o trabalho foi publicado. Trata-se de um método baseado em pontos que começa com os centros de cluster inicialmente colocados em posições arbitrárias e prossegue movendo os centros de cluster a cada etapa, a fim de minimizar o erro de cluster (LIKAS et al., 2003).

Segundo Jain (2010), mesmo tendo sido publicado há mais de 50 anos, o *k-means*, ainda é um dos algoritmos mais utilizados para clusterização devido a sua facilidade de implementação, eficiência, simplicidade e sucesso empírico.

Witten et al. (2011), apresenta que o método tem como entrada um conjunto, $CD = \{D_1, D_2, \dots, D_n\}$, contendo n instância de dados (ou pontos), e um valor inteiro atribuído ao parâmetro k . Inicia-se escolhendo, randomicamente, k instâncias de CD , que representam k centroides de grupos. A partir disso, cada instância de CD é atribuída ao grupo cujo centroide lhe seja mais próximo, por meio do cálculo da distância de cada instância a cada um dos k centroides considerados. Segundo Fávero et al. (2009), existem diversas maneiras de medir a similaridade ou a distância entre os objetos. O estudo a respeito das medidas consideradas mais importantes podem ser vistos na seção 2.6.1. Existem algumas vertentes do algoritmo *k-means* disponíveis na literatura, uma dessas, considerada como padrão, é o algoritmo Hartigan-Wong (1979), que define a variação total dentro do cluster a partir da soma das distâncias quadradas, utilizando as distâncias euclidianas entre os itens e o centroide correspondente.

Com a finalidade de minimizar a determinada função objetivo, mostrada na equação 1, este algoritmo segue, basicamente, os seguintes passos:

$$Z = \sum_{i=1}^k \sum_{j=1}^n d(X_i, C_j)^2 \quad (1)$$

Onde,

n é o número de instâncias;

k é o número de clusters previamente definidos;

X_i é o i -ésimo dado;

C_j é o centroide do j -ésimo cluster;

$d(X_i, C_j)$ é a distância entre X_i e C_j .

Passo 1: Deve-se particionar os objetos, de forma aleatória, em k grupos não vazios, onde cada partição será o centroide desses grupos;

Passo 2: De acordo com a função distância, usualmente a distância Euclidiana, cada objeto restante deve ser caracterizado como pertencente ao cluster ao qual apresenta-se mais próximo do seu centroide;

Passo 3: Atualizar os centroides de cada cluster, sendo calculada a média dos elementos que pertencem ao cluster, através da equação 2, como as novas posições dos centroides.

$$c_j = \frac{\sum_{i=1}^{n^{(j)}} x_i^{(j)}}{n^{(j)}} \quad (2)$$

Onde,

$x_i^{(j)}$ é um elemento i associado ao cluster j ;

$n^{(j)}$ é o número de elementos associados ao cluster j .

Passo 4: Caso algum elemento modifique de cluster no passo 2, repetem-se os passos 3 e 4. Caso contrário, o algoritmo é encerrado.

2.6.1 Medidas de similaridade ou distância (dissimilaridade)

Segundo Fávero et al. (2009), as medidas de distância são consideradas medidas de dissimilaridade. Quanto maiores os valores de distância, menor é a semelhança entre os objetos e vice-versa. Ainda, segundo os autores, as principais medidas de distância tratadas na literatura quando do estudo da análise de cluster são:

- a) Distância Euclidiana: trata-se da distância entre duas observações (i e j) que corresponde à raiz quadrada da soma dos quadrados das diferenças entre os pares de observações (i e j) para todas as p variáveis.

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (3)$$

Onde, x_{ik} é o valor da variável k referente à observação i e x_{jk} representa a variável k para a observação j .

- b) Distância Quadrática Euclidiana: trata-se da distância entre duas observações (i e j) que corresponde à soma dos quadrados das diferenças entre i e j para todas as p variáveis.

$$d_{ij}^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2 \quad (4)$$

- c) Distância de Minkowski: a distância euclidiana é um caso particular da distância de Minkowski, que é uma distância mais geral. E é dada pela seguinte expressão:

$$d_{ij} = \left(\sum_{k=1}^p (|x_{ik} - x_{jk}|)^n \right)^{1/n} \quad (5)$$

Onde, d_{ij} é a distância de Minkowski entre as observações i e j , p é o número de variáveis e $n=1, 2, \dots, \infty$. Quando aplicado $n=2$ na formulação de Minkowski, chegaremos à distância euclidiana. Entretanto, para $n=1$, tem-se uma nova distância, denominada City-Block, também chamada de Manhattan Distance.

- d) Distância absoluta, City-Block ou Manhattan: trata-se da soma das diferenças absolutas entre os valores das p variáveis para os dois casos.

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (6)$$

- e) Mahalanobis: trata-se da distância estatística entre dois indivíduos i e j , considerando a matriz de covariância para o cálculo das distâncias.

$$d_{ij} = \sqrt{(x_i - x_j)' S^{-1} (x_i - x_j)} \quad (7)$$

Onde, S é a estimativa amostral da matriz de variância-covariância Σ dentro dos agrupamentos.

- f) Chebychev: trata-se da diferença absoluta máxima entre todas as p variáveis entre duas observações.

$$d_{ij} = \max_k |x_{ik} - x_{jk}| \quad (8)$$

2.7 Análise de Componentes Principais - PCA

De acordo com Jolliffe (2002), a ideia central da análise de componentes principais é reduzir a dimensionalidade de um determinado conjunto de dados, mantendo o máximo possível a variação presente nesse conjunto. Basicamente, isso é alcançado transformando o conjunto em um novo conjunto de variáveis, os componentes principais, que não são correlacionados e que são ordenados para que os primeiros retenham a maior parte da variação presente em todas as variáveis originais. Essa técnica foi inicialmente descrita por Pearson (1901) e aprimorada, através de métodos computacionais, por Hotelling (1933).

Na PCA, a maior parte das informações significativas estão presentes nas primeiras componentes devido ao fato de que elas são construídas de forma decrescente em variância explicada (BERRUETA et al., 2007).

Para Myatt e Johnson (2009), a PCA é uma combinação linear de pesos das variáveis originalmente observadas no problema em análise. O desenvolvimento matemático de obtenção das componentes principais tem como síntese de aplicação, a respeito do assunto aqui tratado, os seguintes termos e passos:

A partir de uma matriz de dados iniciais X formada por p tipos de crimes obtidos de n municípios pernambucanos e representada pela equação 9, considera-se que o vetor aleatório \vec{x} , apresentado na equação 10, representa o conjunto de variáveis selecionadas para a análise.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \quad (9)$$

$$\vec{x} = [x_1, x_2, \dots, x_n] \quad (10)$$

Então, é calculada a matriz de covariância para todas as variáveis, cujos os autovalores λ e os autovetores a de k componentes principais são:

$$\lambda = [\lambda_1, \lambda_2, \dots, \lambda_k] \quad (11)$$

$$\vec{a} = \begin{bmatrix} \vec{a}_1 \\ \vdots \\ \vec{a}_k \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kp} \end{bmatrix} \quad (12)$$

Os autovetores são vetores com os pesos das combinações lineares utilizadas para calcular o valor das componentes principais y e os autovalores são o valor da variância de cada componente. As equações apresentadas no grupo de equações 13 representam o valor das componentes principais:

$$\begin{aligned} y_1 &= a_1x = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\ y_2 &= a_2x = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\ &\vdots \\ y_k &= a_kx = a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kp}x_p \end{aligned} \quad (13)$$

Após a realização da PCA em um determinado conjunto de variáveis, é obtido um novo conjunto de variáveis composto de k elementos, onde estes são os k componentes principais, cujo autovalor λ_k é o valor da variância apresentada pelo componente y_k e o somatório dos autovalores, representado por $\sum_{i=0}^k \lambda_i$ é utilizado para indicar a representatividade do conjunto

obtido. De acordo com Johnson e Wichern (1992), a PCA é mais um meio para se atingir um resultado final do que um resultado por si só. Para ilustrar a aplicação desta técnica no contexto de crimes, Neisse e Hongyu (2016) utilizaram a PCA como instrumento de redução de massa de dados de diferentes categorias de crimes com frequências observadas em 26 estados nos Estados Unidos da América (EUA) e obtiveram efetividade em sua aplicação, reduzindo a quantidade de variáveis de sete para dois componentes principais que explicaram uma parte satisfatória dos dados originais.

3 DELINEAMENTO METODOLÓGICO

Este capítulo tem como intuito descrever, de forma clara e objetiva, os aspectos metodológicos utilizados para desenvolver o presente estudo. O trabalho trata-se de uma pesquisa aplicada, visto que ele se dedica a gerar conhecimento para apoiar a solução de um problema em específico, que é o problema do direcionamento de políticas de segurança pública. A abordagem utilizada pode ser considerada quantitativa, mas sem desconsiderar o uso da abordagem qualitativa em alguns momentos, percebidos ao longo do estudo e, além disso, é considerada como uma pesquisa descritiva e exploratória.

O acesso aos dados se deu a partir da Lei nº 12.527/2011, conhecida como Lei de Acesso à Informação - LAI, que regulamenta o direito previsto na Constituição, de qualquer pessoa solicitar e receber dos órgãos e entidades públicos, de todos os entes e Poderes, informações públicas por eles produzidas e custodiadas.

3.1 Obtenção dos dados

O Fórum Brasileiro de Segurança Pública (FBSP) classifica, de forma bienal, as unidades federativas de acordo com a qualidade dos dados criminais. Esse estudo de qualidade dos dados, de uma forma resumida, leva em consideração cinco eixos de avaliação da qualidade da informação, como visto no 12º Anuário Brasileiro de Segurança Pública, que são eles: (1) o conceito de homicídios, (2) as informações registradas sobre vítima, fato e suspeito/autor, (3) a perda de informações sobre vítima, fato e suspeito/autor, (4) o grau de convergência dos dados das secretarias estaduais com a fonte oficial da área da saúde e (5) a transparência. E a partir da classificação já definida, os estados são agrupados em 4 grupos. No grupo 1 estão aqueles estados com a maior qualidade de informações, no grupo 2 aqueles com qualidade intermediária das informações, no grupo 3 aqueles com menor qualidade das informações e no grupo 4 aqueles em que não há como atestar a qualidade dos dados.

No mapa 3 tem-se a apresentação de todos os estados brasileiros de acordo com a qualidade dos dados, que é obtida por parte do FBSP. É possível perceber que os estados que compõem o grupo 1, estados com maior qualidade de dados, são aqueles com uma tonalidade mais forte. Sendo assim, este grupo 1 é composto por: Pernambuco, Alagoas, Ceará, Espírito Santo, Goiás, Maranhão, Mato Grosso, Pará, Paraíba, Piauí, Rio de Janeiro, Rio Grande do Norte e Santa Catarina.

Em contexto apenas a respeito do estado de Pernambuco, como disposto no Boletim Trimestral da Conjuntura Criminal em Pernambuco – 1º trimestre de 2018, a sistemática de

3.2 Obtenção dos resultados

A fim de obter os resultados de forma rápida, gratuita e com forte aparato à visualização de resultados, foi usado o *software R 3.6.1*, que tem sido utilizado de forma ampla por pesquisadores em diversas áreas, em especial na estatística e mineração de dados. Sendo um *software* de contribuição aberta, foram utilizados alguns pacotes disponíveis que possuem específicas funções e comandos necessários a aplicação do algoritmo *k-means* e visualização dos dados.

Dentre os pacotes utilizados, o de maior apoio para a aplicação foi o pacote *cluster*, que implementa a técnica de clusterização e o *factoextra*, que facilita a extração e visualização da saída de análises exploratórias de dados multivariados, além de permitir que seja implementada a análise de componentes principais com o objetivo de reduzir a dimensionalidade dos dados, condensando a informação contida em suas várias variáveis originais em um conjunto menor de variáveis (componentes) e com uma perda mínima de informação. Um pseudocódigo foi elaborado, visando demonstrar de forma simples a codificação do algoritmo utilizado. O pseudocódigo está disposto no quadro 1.

Quadro 1 - Pseudocódigo para o *k-means*

O algoritmo <i>k-means</i>	
Inputs:	$X = \{x_{11}, x_{12}, x_{21}, \dots, x_{ij}\}$: conjunto de dados de criminalidade por município k : número de grupos
Outputs:	$P = \{G_1, G_2, G_3, \dots, G_k\}$: partição com k grupos
1	Selecionar aleatoriamente k dados como centroides iniciais
2	Repita
3	para cada valor x pertencente a X faça
4	computar a dissimilaridade de x para cada centroide C ;
5	atribuir x ao centroide mais próximo;
6	fim
7	recalcular o centroide de cada grupo;
8	até atingir uma convergência;

Fonte: O autor (2019)

4 APLICAÇÃO DA TÉCNICA

No capítulo anterior foram descritos os aspectos metodológicos utilizados nesse estudo para atender aos objetivos. Para demonstrar a aplicação, neste capítulo são estabelecidos passos que levam em consideração uma certa organização dos dados e respeitam as etapas do armazenamento em cluster, descritos no capítulo 2. Ademais, neste capítulo, a partir do conjunto de *clusters* obtidos são geradas discussões e explicações práticas. Também são direcionadas ações de combate e prevenção aos crimes considerados no estudo, levando em conta as características dos agrupamentos obtidos.

4.1 Obtenção dos clusters

Primeiramente, como comentado no delineamento metodológico, foram selecionadas algumas variáveis, dentre elas estão os 185 municípios pertencentes ao estado de Pernambuco. Todas as variáveis selecionadas, que podem ser vistas na tabela 3, são consideradas numéricas por se tratar do número de ocorrências, com exceção dos municípios. Ainda na tabela 3, podem ser visualizadas as descrições de cada uma das variáveis, levadas em consideração no estudo, para um melhor entendimento.

Com o intuito de reduzir a dimensionalidade do problema e permitir uma interpretação dos dados de forma simplificada, reduzindo o número de variáveis sem perda significativa de informação a respeito do conjunto de variáveis originais, foi aplicada a ACP. Por conseguinte, foram obtidos os *clusters* a partir do algoritmo *k-means*.

Tabela 3 - Caracterização e descrição das variáveis

Variável	Tipo de dados	Descrição
Município	Caracteres	Municípios do estado de Pernambuco
Tráfico de entorpecentes	Numéricos	Comércio de substâncias consideradas ilícitas
Homicídio doloso	Numéricos	Ato de matar outra pessoa intencionalmente
Lesão corporal dolosa	Numéricos	Ofensa a integridade corporal ou à saúde de uma pessoa (dolo)
Lesão corporal seguida de morte	Numéricos	Crime com dolo no ato antecedente e culpa no fato subsequente
Latrocínio	Numéricos	Crime em que há dolo em matar a vítima para dela subtrair algo
Furto de veículo	Numéricos	Apoderação de um veículo em que a vítima não está presente
Roubo de veículo	Numéricos	Apoderação de um veículo por meio de violência ou ameaça
Roubo a Instituição Financeira	Numéricos	Modalidade de crime voltada especialmente a agências bancárias
Roubo de carga	Numéricos	Crime que ocorre no processo de transferência de mercadorias
Roubo a estabelecimento comercial	Numéricos	Assalto a estabelecimento comerciais, no geral
Roubo a residência	Numéricos	Invasão e roubo a residência

Fonte: O autor (2019)

4.1.1 Análise de Componentes Principais – Criminalidade

A seguir é descrita a análise de componentes principais em relação às variáveis representativas da criminalidade do estado, levadas em consideração no estudo, para os dados do ano de 2018. Na ACP, a variabilidade explicada é vista de forma decrescente por parte das componentes principais, ou seja, decresce da 1ª para a 2ª, da 2ª para a 3ª e assim sucessivamente. Sendo cada componente principal definida como uma combinação linear de todas as variáveis originais e sendo representada por um novo conjunto de variáveis artificiais que são função lineares daquelas originais e que tem variância máxima.

A fim de fornecer uma visualização bidimensional ao longo de todo o estudo, de forma a simplificar a análise, foi optado por utilizar apenas os dois primeiros componentes principais. Como exposto na tabela 4, que contém as proporções de variação explicada por cada componente, além das proporções acumuladas, somente os componentes 1 e 2 já explicam aproximadamente 80,57% de toda a variação dos dados, sendo 70,46% correspondente ao primeiro componente principal e 10,11% ao segundo componente.

A partir da tabela de coeficientes de ponderação das componentes principais, disposta no anexo B, é possível obter as equações que representam cada uma das componentes, usando como padrão a equação 13.

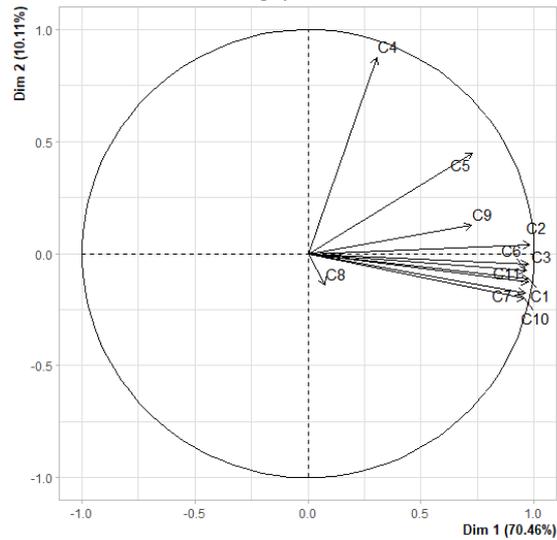
Tabela 4 - Proporção de variação explicada por componentes

Componente Principal	Proporção (%)	Proporção Acumulada (%)
CP1	70,46	70,46
CP2	10,11	80,57
CP3	9,36	89,93
CP4	5,70	95,63
CP5	2,58	98,21
CP6	0,67	98,88
CP7	0,39	99,27
CP8	0,33	99,60
CP9	0,20	99,79
CP10	0,12	99,91
CP11	0,08	100,00

Fonte: O autor (2019)

O relacionamento estrutural entre as variáveis e os componentes principais pode ser analisado a partir do gráfico de variáveis, disposto no gráfico 1, que apresenta uma visão das projeções das variáveis observadas projetadas no plano medido pelos dois primeiros componentes. Para interpretar o gráfico, é indicado que sua dedução seja focada nas extremidades (superior, inferior, esquerda e direita).

Gráfico 1 - Gráfico do relacionamento entre variáveis e componentes principais



Fonte: O autor (2019)

Dessa forma, ao observar o gráfico é percebido que a primeira componente principal reflete, principalmente, atos violentos de qualquer tipo, com menor contribuição dos crimes 4 (lesão corporal seguida de morte), 5 (latrocínio) e 8 (roubo a instituição financeira). Já no caso da segunda componente principal, justamente os crimes 4, 5 e 8 são os de maior contribuição.

Fazendo um paralelo com o mapa de municípios de acordo com a incidência de crimes (gráfico 2), fica claro que os municípios localizados à direita são aqueles onde a maioria dos atos violentos são mais frequentes. Listados a direita, em destaque, apresentam-se os municípios de Recife (mostrado com disparidade na localização em relação aos demais), Jaboatão dos Guararapes, Araripina, Petrolina, São Vicente Ferrer, Salgueiro, Belo Jardim, Serra Telhada, Paulista, Caruaru, Vitória de Santo Antão, Cabo de Santo Agostinho, Lajedo, Abreu e Lima, Afogados da Ingazeira, Taquaritinga do Norte, Moreno, São Lourenço da Mata, São Bento do Uma, Olinda e Santa Cruz do Capibaribe.

Gráfico 2 - Localização de municípios de acordo com a incidência de crimes



Fonte: O autor (2019)

Logo, no gráfico 2, fica demonstrado que a partir da aplicação da ACP, por si só, já é visível um certo agrupamento dos municípios. Um exemplo disto é o caso dos municípios de Bom Jardim, Flores, São Vicente Ferrer, Belo Jardim e Serra Talhada que se apresentam próximos. Essa situação é confirmada no próximo tópico, 4.1.2, onde ao ser aplicado algoritmo de *clusterização*, fica confirmada a presença desses municípios no mesmo agrupamento, até mesmo para diferentes números de grupos.

A capital do estado, Recife, disposta na extremidade inferior à direita, apresenta-se com significativa distância dos demais municípios. Tal feito está relacionado ao alto índice de criminalidade em comparação aos demais municípios. O crime de lesão corporal dolosa, por exemplo, foi o crime com o maior número de ocorrências no município de Jaboatão dos Guararapes, no ano de 2018. Com exceção da capital do estado, em relação a esse tipo de crime, Jaboatão dos Guararapes se apresentou no topo do ranking no estado. Porém, o número de ocorrências para o mesmo tipo de crime na capital do estado foi de aproximadamente 3,2 vezes maior.

Dessa forma, fica evidente que existe uma necessidade de um detalhamento maior no que diz respeito a região da capital, para que seja possível realizar uma identificação de “pontos quentes” em relação a específicos crimes na região.

É conhecida a divisão do território da capital, estabelecida pela Secretária de Defesa Social de Pernambuco, que a tem segregada em 5 áreas, que são elas: Santo Amaro, Espinheiro, Boa Viagem, Várzea e Apipucos. Onde, cada uma dessas áreas agrega algumas circunscrições que são compostas por específicos bairros. Logo, seria plausível a consideração do município dividido por tais regiões para realizar o estudo na forma em que o trabalho é pautado, porém, há uma certa limitação no que diz respeito a desagregação dos dados, já que não são fornecidos dessa maneira.

Há também, no gráfico 2, uma concentração grande de municípios na extremidade esquerda. Essa concentração, evidentemente, está relacionada a municípios com um número de ocorrências baixas comparados aos demais municípios, em relação aos crimes em que os demais municípios se destacam, como é o caso de Alagoinha, Brejo da Madre de Deus, Correntes, entre outros. Outro ponto que pode ser observado é o fato de que os municípios de Araripina, Petrolina e Jaboatão dos Guararapes apresentam-se na extremidade superior um pouco à direita, isso faz com que nitidamente seja percebido que esses municípios tem um certo destaque em relação aos crimes que mais contribuem com a componente principal 2, em relação aos outros municípios, sendo eles os crimes de lesão corporal seguida de morte e latrocínio.

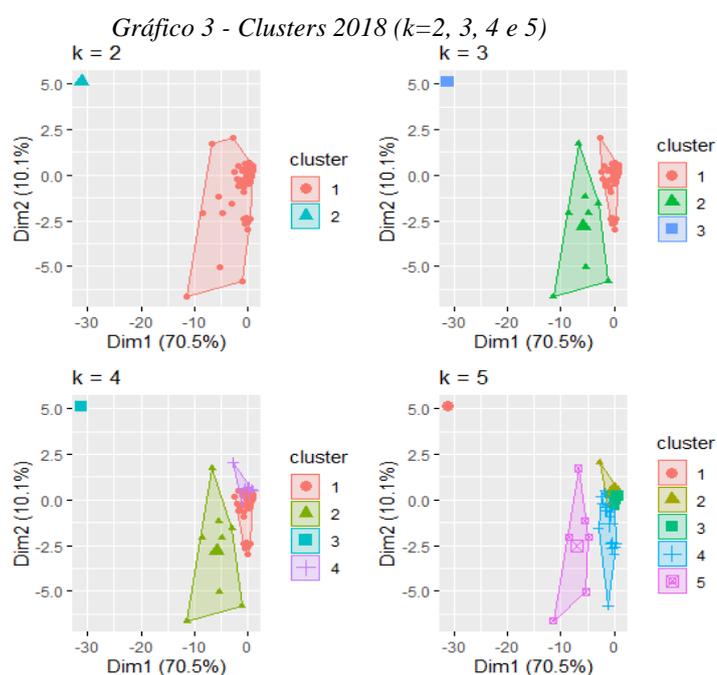
4.1.2 Clusters

Com a dimensionalidade dos dados já devidamente reduzida, a partir da aplicação da análise de componentes principais, foram obtidos os clusters, utilizando a técnica de clusterização *k-means*; onde, primeiramente, foi necessário estabelecer a quantidade de grupos/clusters. Como afirmam Witten et al. (2011), deve ser atribuído, previamente, um número inteiro ao parâmetro k que será justamente a quantidade de cluster.

Esse parâmetro, normalmente, é definido de forma *ad hoc* pelo usuário. De acordo com Silva e Ribeiro (2018), tipicamente, o valor de k é escolhido com base no conhecimento a priori do problema; necessitando da visão crítica de um especialista. Devido as circunstâncias já discutidas, não é pauta deste trabalho fornecer um valor ótimo para esse parâmetro.

Diante disso, foi optado pela implementação do algoritmo para diferentes valores do parâmetro k sendo analisados e comentados os agrupamentos obtidos para cada um dos casos.

De início, foi aplicado o algoritmo para os seguintes valores de k : 2, 3, 4 e 5. Os municípios que compõem cada um clusters obtidos (gráfico 3) podem ser vistos a partir das tabelas 5, 6 e 7.



Como se pode observar, alguns agrupamentos que são esperados acontecem, a partir do que é demonstrado com o uso da PCA, como é o caso dos municípios de Petrolina, Jaboatão dos Guararapes e Araripina, que fazem parte do mesmo cluster para o caso de $k=2$ e $k=3$. A partir da tabela 7 já se pode observar que o município de Araripina passa a fazer parte de um outro *cluster*, mostrando que o algoritmo é bastante sensível à escolha dos pontos iniciais. Essa

mudança ocorre devido as novas distâncias euclidianas que são calculadas a partir da quantidade de cluster pré-estabelecidas.

Tabela 5 - Municípios que compõem os clusters 2018 (k=2 e k=3)

Para k=2		Para k=3		
Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 3
Demais municípios	Recife	Demais municípios	Araripina	Recife
-	-	-	Cabo de Santo Agostinho	-
-	-	-	Jaboatão dos Guararapes	-
-	-	-	Vitória de Santo Antão	-
-	-	-	Paulista	-
-	-	-	Caruaru	-
-	-	-	Olinda	-
-	-	-	Petrolina	-

Fonte: O autor (2019)

Tabela 6 - Municípios que compõem os clusters 2018 (k=4)

Para k=4			
Cluster 1	Cluster 2	Cluster 3	Cluster 4
Demais municípios	Araripina	Recife	Alagoinha
-	Cabo de Santo Agostinho	-	Brejo da Madre de Deus
-	Jaboatão dos Guararapes	-	Correntes
-	Vitória de Santo Antão	-	Buenos Aires
-	Paulista	-	Capoeiras
-	Caruaru	-	Pombos
-	Olinda	-	Ribeirão
-	Petrolina	-	São Bento do Una
-	-	-	Sirinhaém
-	-	-	Surubim
-	-	-	Tacaratu
-	-	-	Bezerros
-	-	-	Carnaíba
-	-	-	Itacuruba
-	-	-	Santa Filomena
-	-	-	Verdejante
-	-	-	Cumarú
-	-	-	Santa Cruz do Capibaribe

Fonte: O autor (2019)

Tabela 7 - Municípios que compõem os clusters 2018 (k=5)

Para k=5				
Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Recife	Alagoinha	Demais municípios	Araripina	Cabo de Santo Agostinho
-	Brejo da Madre de Deus	-	Belo Jardim	Jaboatão dos Guararapes
-	Correntes	-	Camargibe	Paulista
-	Buenos Aires	-	Escada	Caruaru
-	Capoeiras	-	Ipojuca	Olinda
-	Pombos	-	Moreno	Petrolina
-	Ribeirão	-	Palmares	-
-	Santa Cruz do Capibaribe	-	Paudalho	-
-	São Bento do Una	-	São Vicente Ferrer	-
-	Sirinhaém	-	Sertânia	-
-	Surubim	-	Vitória de Santo Antão	-
-	Tacaratu	-	Gravatá	-
-	Bezerros	-	Serra Talhada	-
-	Carnaíba	-	Bom Jardim	-
-	Cumarú	-	Flores	-
-	Itacuruba	-	Goiana	-
-	Santa Filomena	-	Salgueiro	-
-	Verdejante	-	São Lourenço da Mata	-
-	-	-	Igarassu	-
-	-	-	Lajedo	-

Fonte: O autor (2019)

A fim de observar o agrupamento dos municípios para outros valores do parâmetro k , foram consideradas algumas divisões que já existem e são praticadas no direcionamento de políticas de segurança pública e na avaliação de desenvolvimento do estado. Sendo as AIS's, comentadas na sessão 2.4, as divisões relacionadas ao direcionamento de políticas de segurança pública e as Regiões de Desenvolvimento do Estado de Pernambuco (mapa 4), relacionadas ao estudo de desenvolvimento do estado, num contexto amplo. E sendo consideradas, respectivamente, 26 e 12 áreas integradas no que diz respeito às AIS's e Regiões de desenvolvimento. Vale salientar que a região referente ao sertão Pajeú/Moxotó, presente no mapa 4, é dividido em sertão Pajeú, sertão Moxotó e sertão Central.

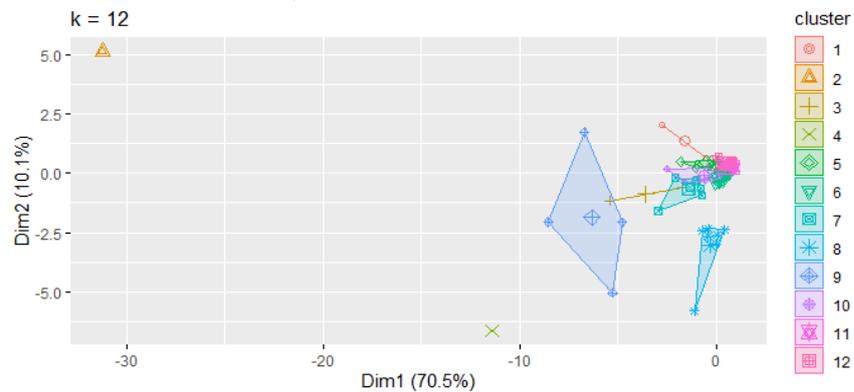
Mapa 4 - Regiões de desenvolvimento do estado de Pernambuco



Fonte: Adaptado de Lima et al. (2005)

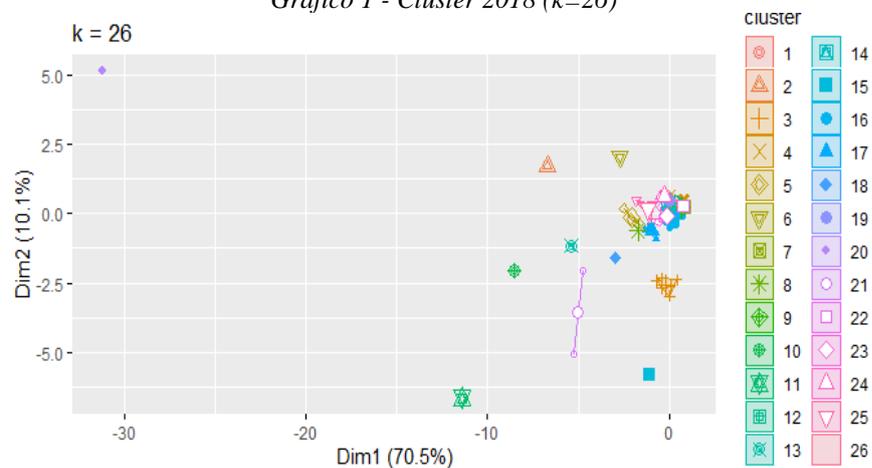
Dessa forma, novamente foi utilizado o algoritmo com o intuito de avaliar a obtenção de clusters para o parâmetro $k=12$ e $k=26$. Os clusters obtidos podem ser vistos a partir dos gráficos 4 e 5.

Gráfico 4 - Clusters 2018 ($k=12$)



Fonte: O autor (2019)

Gráfico 1 - Cluster 2018 ($k=26$)



Fonte: O autor (2019)

De forma listada, tem-se abaixo os clusters obtidos a partir do parâmetro $k=12$ e $k=26$.

Para $k=12$:

Cluster 1: Santa Cruz do Capibaribe e Surubim

Cluster 2: Recife

Cluster 3: Cabo de Santo Agostinho e escada

Cluster 4: Jaboatão dos Guararapes

Cluster 5: Abreu e Lima, Camaragibe, Garanhuns, Pesqueira, Timbaúba, Carpina, Itambé e São Caitano

Cluster 6: Agrestina, Cupira, Itapetim, Riacho das Almas, Sertânia, Taquaritinga do Norte, Trindade, Aliança, Arcoverde, Calçado, Catende, Custódia, Ferreiros, Glória do Goitá, Ipubi, Lagoa de Itaenga, Nazaré da Mata, Barra de Guabiraba, Gameleira, Jurema, Ouricuri, Panelas e São José da Coroa Grande

Cluster 7: Ipojuca, Paudalho, Vitória de Santo Antão, Gravatá, Bezerros, Igarassu e Lajedo

Cluster 8: Araripina, Belo Jardim, São Vicente Ferrer, Serra Talhada, Bom Jardim, Flores e Salgueiro

Cluster 9: Paulista, Caruaru, Olinda e Petrolina

Cluster 10: Moreno, Palmares, Ribeirão, Xexéu, Águas Belas, Belém de São Francisco, Goiana e São Lourenço da Mata

Cluster 11: Demais municípios

Cluster 12: Alagoinha, Brejo da Madre de Deus, Correntes, Buenos Aires, Capoeiras, Pombos, São Bento do Una, Sirinhaém, Tacaratu, Carnaíba, Cumarú, Itacuruba, Santa Filomena e Verdejante

Para $k=26$, tem-se:

Cluster 1: Barreiros, Bonito, Canhotinho, Limoeiro, Altinho, Buíque, São Joaquim do Monte, Tamandaré, Vertentes e Vicência

Cluster 2: Olinda

Cluster 3: Belo Jardim, São Vicente Ferrer, Bom Jardim, Salgueiro, Serra Talhada e Flores

Cluster 4: Brejo da Madre de Deus, Correntes, Alagoinha, Buenos Aires, Capoeiras, Carnaíba, Pombos, Santa Filomena, Sirinhaém, Tacaratu, Verdejante, Cumaru e Itacuruba

Cluster 5: São Lourenço da Mata, Goiana, Igarassu

Cluster 6: Santa Cruz do Capibaribe

Cluster 7: Caetés, Casinhas, Chã de Alegria, Ibimirim, Itaíba, Lagoa do Carro, Água Preta, Angelim, Araçoiaba, Bom Conselho, Camocim de São Félix, Jaqueira, Lagoa Grande, Primavera, Rio Formoso, Tupanatinga, Venturosa, Cortês, Exú, Iati e Itaquitinga

Cluster 8: Escada

Cluster 9: Amaraji, Floresta, Paranatama, Quipapá, Saloá, Santa Maria da Boa Vista, São José do Belmonte, Cabrobó, Condado, Joaquim Nabuco, Orocó, Parnamirim, Petrolândia, Ilha de Itamaracá e Itapissuma

Cluster 10: Caruaru

Cluster 11: Jaboatão dos Guararapes

Cluster 12: Timbaúba e Itambé

Cluster 13: Cabo de Santo Agostinho

Cluster 14: Pesqueira e São Caitano

Cluster 15: Araripina

Cluster 16: Agrestina, Cupira, Itapetim, Riacho das Almas, Sertânia, Taquaritinga do Norte, Trindade, Aliança, Arcoverde, Barra de Guabiraba, Calçado, Catende, Jurema, Lagoa de Itaenga, Nazaré da Mata, Ouricuri, Panelas, São José da Coroa Grande, Custódia, Ferreiros, Gameleira, Glória do Goitá e Ipubi

Cluster 17: Paudalho, Bezerros, Lajedo e Gravatá

Cluster 18: Vitória de Santo Antão

Cluster 19: Carpina e Toritama

Cluster 20: Recife

Cluster 21: Paulista e Petrolina

Cluster 22: Bodocó, Orobó, Sairé, São João, Tacaimbó, Afogados da Ingazeira, Cachoeirinha, Camutanga, Chã Grande, João Alfredo, Jupi, Macaparana, Passira, Sanharó, Santa Maria do Cambucá, São José do Egito, Tabira, Tracunhaém, Feira Nova e Frei Miguelinho

Cluster 23: Moreno, Palmares, Águas Belas, Belém de São Francisco, Ribeirão e Xexéu

Cluster 24: São Bento do Una e Surubim

Cluster 25: Abreu e Lima, Camaragibe, Garanhuns e Ipojuca

Cluster 26: Carnaubeira da Penha, Fernando de Noronha, Granito, Iguaraci, Jatobá, Lagoa dos Gatos, Manari, Poção, Santa Cruz da Baixa Verde, São Benedito do Sul, Solidão, Tuparema, Vertente do Lério, Afrânio, Belém de Maria, Betânia, Brejão, Brejinho, Calumbi, Cedro, Jataúba, Lagoa do Ouro, Machados, Maraial, Mirandiba, Palmerina, Pedra, Quixaba, Salgadinho, Santa Cruz, Santa Terezinha, Serrita, Moreilândia, Terezinha, Terra Nova, Triunfo, Dormentes, Ibirajuba, Inajá, Ingazeira e Jucati

Para a primeira situação, em que $k=12$, o cluster 11 foi o grupo com a maior quantidade de municípios, tendo um total de 108 municípios fazendo parte desse grupo, já para o caso em que $k=26$, um total de 41 municípios foi a maior quantidade presente em um *cluster*, nesse caso, o *cluster* 26. A quantidade de municípios presentes em um cluster pode ser atrelada a indicadores como população e extensão territorial para direcionamento de específicas políticas de segurança, como é o caso da ação de realocar o quadro efetivo de policiais na região.

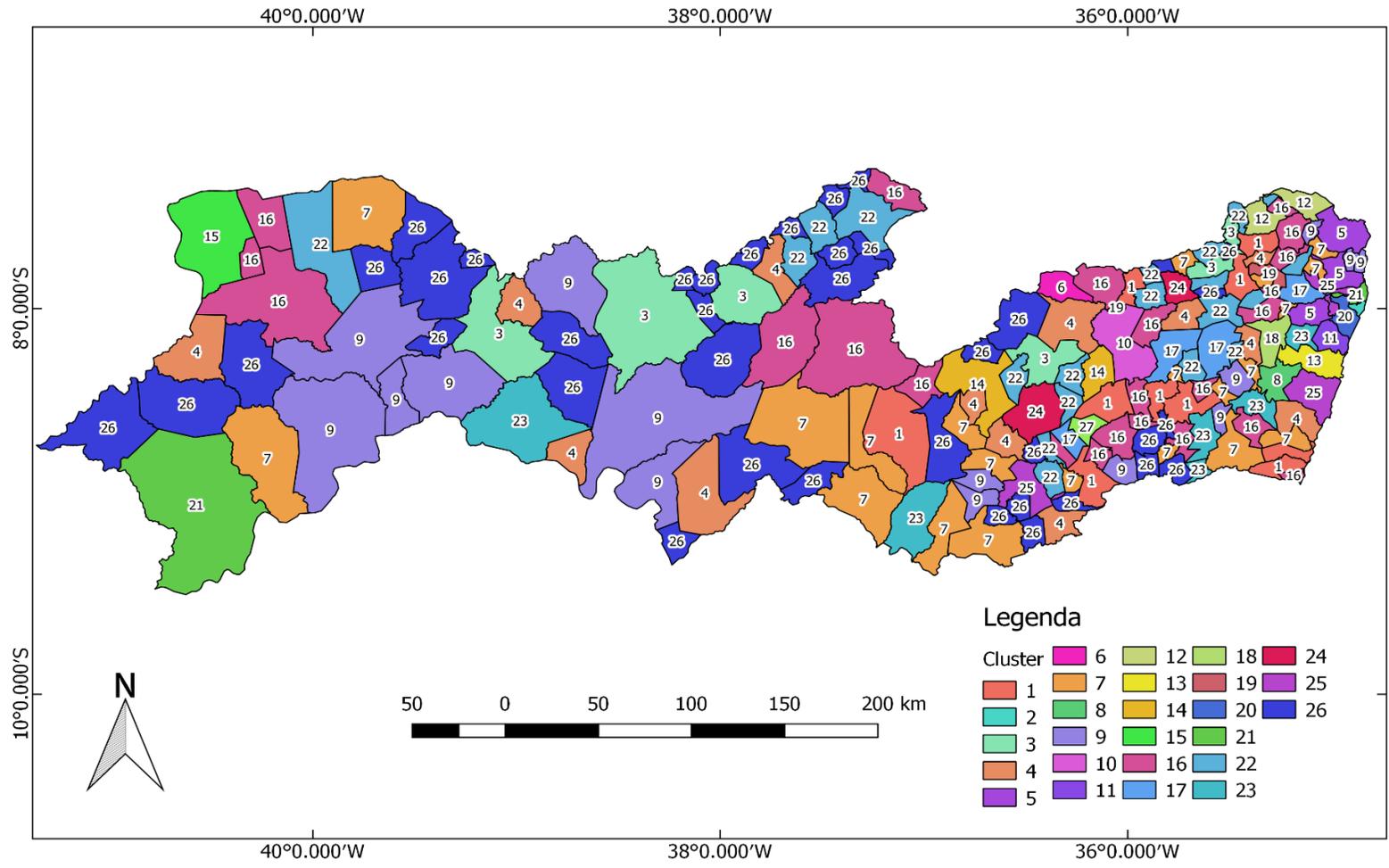
Outras discussões podem ser realizadas a respeito dos clusters, como por exemplo, sobre o *cluster* 9, para $k=12$, que é composto pelos municípios de Paulista, Caruaru, Olinda e Petrolina, onde ao se analisar o número de ocorrência dos 11 tipos de crimes, considerados no estudo, para os 4 municípios que compõem o cluster, é observada uma certa semelhança no número de ocorrências entre os crimes, com um destaque em relação aos crimes de lesão corporal dolosa e roubo de veículo, tendo uma média de ocorrências no ano de aproximadamente 873 e 586, respectivamente. Diferente dos números para os crimes de lesão corporal seguida de morte e roubo a instituição financeira, em que se apresentam com uma média de 1 ocorrência anualmente. Dessa forma, é admissível considerar os crimes com maior incidência em cada um dos clusters para realizar um direcionamento de políticas de segurança pública adequadas.

O mapa 5 apresenta o mapa do estado dividido a partir dos clusters obtidos utilizando o parâmetro $k=26$. Pode ser percebido que não necessariamente existe uma relação entre proximidade, no que se refere a localidade dos municípios, para o agrupamento das regiões. Um exemplo claro disto é o *cluster* apresentado na cor verde de numeração 21, sendo este composto pelos municípios de Petrolina e Paulista. Como pode ser observado, o município de Petrolina está localizado a 712 km a oeste da capital do estado, já o município de Paulista está a uma distância de 18 km do Recife, o tendo como um dos seus municípios limítrofes.

O fato desses dois municípios estarem presentes no mesmo grupo tem relação direta com a distância euclidiana, calculada a partir dos números de ocorrências por tipo de crimes, de

forma que foram agrupados baseado na semelhança identificada. Essa constatação permite que seja percebida a necessidade do uso de abordagens de agrupamento, nesse sentido, para direcionar ações de combate ao crime e não somente o uso de regiões estabelecidas de acordo com a proximidade entre municípios.

Mapa 5 - Mapa do estado segregado por clusters



Fonte: O autor (2019)

4.2 Direcionamento das Políticas de Segurança Pública

É de senso comum a existência de práticas difusas de crimes entre cidades; diante disso, a partir da obtenção dos *clusters*, foi possível alcançar grupos com ocorrências basicamente semelhantes.

Com a finalidade de demonstrar o direcionamento de políticas adequadas para cada um desses grupos foi realizado um levantamento de algumas políticas e programas nacionais e estaduais para cada um dos tipos de crimes aqui tratados, como a Política Nacional sobre Drogas, os documentos e programas definidos pela Secretaria de Política de Prevenção à Violência e às Drogas do estado de Pernambuco, Secretaria de Defesa Social de Pernambuco e o Plano Nacional de Segurança Pública. Logo, as ações aqui propostas foram definidas com base em informações de diversos âmbitos.

A partir desse levantamento foi possível listar específicas ações de combate e prevenção a cada um dos crimes e com isso direcionar categorias de ações que mais se adequam a cada um dos clusters. As categorias estão listadas de acordo com cada um dos tipos de crimes e podem ser vistas, junto com as suas respectivas ações nas tabelas 8 e 9.

Para exemplificar o direcionamento das categorias de ações, foi considerado o agrupamento realizado com 26 clusters, já que essa quantidade de grupos faz referência a quantidade utilizada atualmente, sendo as Áreas Integradas de Segurança do estado de Pernambuco. Dessa forma, foi verificado quais os três tipos de crimes mais corriqueiros no(s) município(s) de cada um dos clusters e definidos como representantes do grupo a fim de priorizar as ações frente a urgência. Na tabela 10 são mostrados quais os crimes considerados como representantes para cada um dos grupos, estando eles em ordem de incidência. A definição dos crimes representantes foi realizada através da aplicação de uma média móvel simples com os números de ocorrências de cada um dos crimes para todos os municípios pertencentes a cada um dos clusters.

Tabela 8 - Ações por tipo de crime

Categoria de Ações	Ações
Tráfico de entorpecentes	<p>Investir em programas educacionais</p> <p>Promover ação contínua de desmonte das organizações criminosas</p> <p>Investir em tecnologias para monitoramento de plantações de drogas</p> <p>Estabelecer uma divisão de inteligência estadual no combate ao tráfico</p> <p>Implantar (ou investir, caso exista) delegacias de repressão ao narcotráfico</p>
Homicídio doloso	<p>Implementar um plano de combate a crimes contra à vida</p> <p>Implantar câmeras de monitoramento de homicídios</p> <p>Requalificar e/ou estruturas ou ampliar os programas de proteção a grupos sociais em situação de alta vulnerabilidade à violência</p> <p>Fortalecer a capacidade investigativa dos policiais</p>
Lesão corporal dolosa	<p>Implementar um plano de combate a crimes contra à vida</p> <p>Criar projetos sociais em que sejam tratados os temas</p> <p>Aumentar a ronda policial</p> <p>Requalificar e/ou estruturas ou ampliar os programas de proteção a grupos sociais em situação de alta vulnerabilidade à violência</p>
Lesão corporal seguida de morte	<p>Implementar um plano de combate a crimes contra à vida</p> <p>Ampliar gradativamente o efetivo de policiais</p> <p>Capacitação do efetivo através de cursos de mediação de conflitos e atendimento humanizado</p> <p>Cursos de apoio e qualificação para grupos em vulnerabilidade social</p>
Latrocínio	<p>Realizar operações em locais de maior incidência no interior dos municípios</p> <p>Desburocratizar e fortalecer a capacidade investigativa dos policiais</p> <p>Capacitar profissionais em técnicas de investigação, entrevista e interrogatório</p> <p>Implantar câmeras de monitoramento em vias estratégicas</p>
Furto de veículo	<p>Expandir a troca de informações e o compartilhamento de sistemas entre os estados</p> <p>Implantar/expandir o sistema de videomonitoramento e acompanhamento de placas de veículos</p> <p>Ampliar o uso de tecnologia da informação no monitoramento de veículos</p> <p>Definir operações de verificação de veículos em pontos estratégicos</p>

Fonte: O autor (2019)

Tabela 9 - Ações por tipo de crime

Categoria de Ações	Ações
Roubo de veículo	Estabelecer força-tarefa na investigação de roubo de veículos Implantar (ou investir) em delegacias de combate ao roubo de veículos Expandir a troca de informações e o compartilhamento de sistemas entre os estados Implantar/expandir o sistema de videomonitoramento e acompanhamento de placas de veículos
Roubo a Inst. Financeira	Investir no gerenciamento de risco em conjunto com a Federação Brasileira de Bancos Determinar ações de inteligência com uma equipe multidisciplinar Investir em sistemas de compartilhamento de imagens das câmeras de segurança dos bancos com as autoridades policiais
Roubo de carga	Promover ação contínua da desarticulação de organizações criminosas responsáveis por roubos de cargas Ampliar a fiscalização em rodovias Investir em pessoal especializado, como por exemplo em auditores fiscais
Roubo a estabelecimento comercial	Investigar e desarticular quadrilhas praticantes de roubos desta natureza Aumentar a coordenação de patrulhas em áreas estratégicas periodicamente Promover ação de aconselhamento e parceria com comerciantes das regiões
Roubo a residência	Expandir o sistema de monitoramento por câmeras Identificar as regiões com maior incidência e realizar patrulhamento Realizar ações em parceria com as prefeituras na identificação da necessidade de iluminação em pontos estratégicos

Fonte: O autor (2019)

Os crimes de lesão corporal dolosa e roubo de veículo destacaram-se como os mais corriqueiros em todo o estado de Pernambuco, sendo considerados como representantes em todos os clusters e demonstrando urgência no combate e prevenção. Esse fato pode ser confirmado através dos números de ocorrência dos crimes dispostos no Anexo A.

Já os crimes de lesão corporal seguida de morte, latrocínio e roubo a instituição financeira não foram considerados como representantes para nenhum dos clusters. Porém não devem ser desconsiderados num possível plano estadual de segurança pública, já que, de forma pontual, alguns municípios apresentaram número de ocorrência significativo quando comparados aos demais.

Tabela 10 - Crimes representantes dos clusters

Cluster	Crimes representantes
1	Lesão corporal dolosa Roubo de veículo Homicídio doloso
2	Lesão corporal dolosa Roubo de veículo Tráfico de entorpecentes
3	Lesão corporal dolosa Roubo de veículo Tráfico de entorpecentes
4	Lesão corporal dolosa Roubo de veículos Homicídio doloso
5	Furto de veículo Roubo de veículo Lesão corporal dolosa
6	Roubo de veículo Furto de veículo Lesão corporal dolosa
7	Lesão corporal dolosa Roubo de veículo Homicídio doloso
8	Roubo de veículo Lesão corporal dolosa Roubo de carga
9	Lesão corporal dolosa Roubo de veículo Homicídio doloso
10	Roubo de veículo Lesão corporal dolosa Furto de veículo
11	Lesão corporal dolosa Roubo de veículo Furto de veículo
12	Lesão corporal dolosa Tráfico de entorpecentes Roubo de veículo
13	Roubo de veículo Lesão corporal dolosa Homicídio doloso
14	Lesão corporal dolosa Roubo de veículo Roubo a residência
15	Lesão corporal dolosa Tráfico de entorpecentes Roubo e Furto de veículo
16	Lesão corporal dolosa Roubo de veículo Furto de veículo
17	Roubo de veículo Lesão corporal dolosa Furto de veículo
18	Roubo de veículo Lesão corporal dolosa Tráfico de entorpecentes
19	Roubo de veículo Lesão corporal dolosa Furto de veículo
20	Lesão corporal dolosa Roubo de veículo Roubo a estab. comercial
21	Lesão corporal dolosa Roubo de veículo Tráfico de entorpecentes
22	Lesão corporal dolosa Roubo de veículo Furto de veículo
23	Lesão corporal dolosa Roubo de veículo Homicídio doloso
24	Roubo de veículo Lesão corporal dolosa Furto de veículo
25	Lesão corporal dolosa Roubo de veículo Tráfico de entorpecentes
26	Lesão corporal dolosa Roubo de veículo Homicídio doloso

Fonte: O autor (2019)

Dessa forma, é possível associar a tabela 10 com as tabelas 8 e 9 com o intuito de obter as recomendações de ações de segurança pública que mais se adequam aos municípios que compõem o determinado cluster. Ou seja, os municípios que compõem o cluster 1 devem seguir as recomendações de ações para os crimes de lesão corporal dolosa, roubo de veículo e homicídio doloso, já que foi representado por estes. Os municípios que compõem o cluster 2 devem seguir as recomendações de ações de lesão corporal dolosa, roubo de veículo e tráfico de entorpecentes, e assim sucessivamente.

5 CONCLUSÕES

A amplitude exacerbada da violência, em contexto geral, torna evidente a necessidade de estudos para investigar o tema sob diferentes perspectivas a fim de atingir contribuições significativas na redução e controle de taxas. É de extrema importância entender que para alcançar resultados significativos neste meio é crucial que as decisões sejam estruturadas de forma eficiente e que os recursos públicos tenham um direcionamento efetivo.

Diante disso, este trabalho propôs utilizar uma abordagem de aprendizagem não supervisionada com o intuito de estabelecer *clusters* entre os municípios do estado de Pernambuco, considerando algumas variáveis representativas da criminalidade no estado, a fim de direcionar ações para prevenção e combate aos crimes de forma a apoiar os formuladores de políticas de segurança pública.

Esse tipo de decisão é caracterizado por exigir capacidade de avaliação e perspicácia por parte do tomador de decisão e o uso da técnica em questão dar-se como um forte suporte nesse sentido, visto que o objetivo das formação de *clusters* é o de obter uma homogeneidade dentro dos grupos formados, de maneira que os componentes (nesse caso, os municípios) do grupo compartilhem características criminais comuns que os diferenciem dos municípios de outros *clusters*. Logo, a *clusterização* como ferramenta de auxílio a decisão de direcionamento de ações de segurança pública fornece uma melhor visão das regiões que apresentam uma necessidade maior de atenção no cenário do estado de Pernambuco.

Como visto, a partir da obtenção de *clusters* baseado no número de ocorrências criminais, os especialistas podem analisar agrupamentos de municípios simultaneamente, priorizar municípios no fornecimento de equipes, direcionar políticas de forma mais assertiva, além de que os *clusters* obtidos podem fornecer uma fonte de estudo interdisciplinar. Nesse sentido, foi demonstrado uma forma simples de direcionar as políticas baseado na incidência dos tipos de crimes. Ou seja, foram segregadas ações de combate e prevenção a crimes específicos e analisado quais os crimes com maior número de ocorrência no *cluster* em questão.

O trabalho alcançou os objetivos traçados para geração dos agrupamentos e proposição do direcionamento de ações, mesmo diante da sensibilidade por parte do algoritmo *k-means* na geração de diferentes agrupamentos em cada execução.

Algumas das limitações que podem ser levantadas dizem respeito aos dados fornecidos, através da Lei de Acesso à Informação, que apresentam lacunas quando observados em um maior horizonte de tempo. Além da não disposição de uma visão crítica por parte de um

especialista da área, pois apesar do uso de um algoritmo consolidado e uma forte revisão literária, a análise de um especialista se faz necessária para enriquecer ainda mais o trabalho.

5.1 Sugestões para trabalhos futuros

Como sugestão para trabalhos futuros incentiva-se o uso de outras técnicas de mineração de dados além da *clusterização*, pois como visto, técnicas visuais, intuitivas e de inteligência na investigação criminal podem aumentar a eficácia e eficiência das análises. Ainda, recomenda-se a análise de *cluster* considerando outros conjuntos de dados, além do número de ocorrência de crimes, como dados relacionados a grau de escolaridade, pobreza, entre outros fatores.

E por fim, é recomendado o uso de outros algoritmos visando realizar um estudo comparativo entre metodologias tradicionais e novas, identificando as principais características, vantagens e desvantagens associadas a cada uma delas.

REFERÊNCIAS

- AGARWAL, J.; NAGPAL, R.; SEHGAL, R.; Crime Analysis using K-Means Clustering. *International Journal of Computer Applications*, Vol.83, n.4, 0975-8887, 2013.
- AHMAD, A. U.; STARKEY, A.; Application of feature selection methods for automated clustering analysis: a review on synthetic datasets. *Neural Computing and Applications*, Vol.29, n.7, 317-328, 2017.
- ALVES, L. G. A.; RIBEIRO, H. V.; RODRIGUES, F. A.; Crime prediction through urban metrics and statistical learning. *Physica A: Statistical Mechanics and Its Applications*, Vol.505, 435-443, 2018.
- ANDERSHED, H.; KÖHLER, D.; LOUDEN, J. E.; HINRICHS, G.; Does the three-factor model of psychopathy identify a problematic subgroup of young offenders?. *International Journal of Law and Psychiatry*, Vol.31, n.3, 189-198, 2008.
- AYED, A. B.; HALIMA, M. B.; ALIMI, A. M.; Survey on clustering methods: Towards fuzzy clustering for big data. In: *Soft Computing and Pattern Recognition (SoCPaR), 6th International Conference of. IEEE*, 331-336, 2014.
- BECKER, G.; Crime and punishment: an economic approach. *Journal of Political Economy*, Chicago, v.76, n.2, 169-217, 1968.
- BERNSTEIN, J. H.; The data-information-knowledge-wisdom hierarchy and its antithesis. *Nasko*, Vol.2, n.1, 68-75, 2009.
- BERRUETA, L. A.; ALONSO-SALCES, R. M.; HÉBERGER, K.; Supervised pattern recognition in food analysis. *Journal of Chromatography A*, Vol.1158, n.1-2, 196-214, 2007.
- BRASIL. Constituição (1988). Constituição da República Federativa do Brasil. Brasília, DF: Senado Federal: Centro Gráfico, 1988.
- BUONAMICI, S. C.; Direito fundamental social à segurança pública. *Revista de estudos jurídicos UNESP*, Vol.15, n.21, 2011.
- COHEN-ADDAD, V.; KANADE, V.; MALLMANN-TRENN, F.; MATHIEU, C.; Hierarchical Clustering. *Journal of the ACM*, Vol.66, n.4, 1-42, 2019.
- CONDEPE/FIDEM - Agência Estadual de Planejamento e Pesquisas de Pernambuco. *Estatísticas da Criminalidade Violenta em Pernambuco*, 2015.
- CRUZ, J. P. da; Pacto Pela Vida: Avanços e retrocessos no combate à violência em Pernambuco. Dissertação (Mestrado Profissional em Gestão Pública para o Desenvolvimento do Nordeste). Programa de Pós-Graduação em Gestão Pública para o Desenvolvimento do Nordeste (MGP), Universidade Federal de Pernambuco, Recife, 2014.

- CURMAN, A. S. N.; ANDRESEN, M. A.; BRANTINGHAM, P. J.; Crime and Place: A Longitudinal Examination of Street Segment Patterns in Vancouver, BC. *Journal of Quantitative Criminology*, Vol. 31, 127-147, 2015.
- DAS, P.; DAS, A. K.; Graph-based clustering of extracted paraphrases for labelling crime reports. *Knowledge-Based Systems*, Vol. 179, 55-76, 2019.
- DAS, P.; DAS, A. K.; NAYAK, J.; PELUSI, D.; DING, W.; Group incremental adaptive clustering based on neural network and rough set theory for crime report categorization. *Neurocomputing*, 2019.
- DECLARAÇÃO UNIVERSAL DOS DIREITOS HUMANOS. Assembleia Geral das Nações Unidas em Paris, 217 A(III), 1948.
- de LIMA, M. L.; de SOUZA, E. R.; XIMENES, R.; de ALBURQUEQUE, M. de, F.; BITOUN, J.; BARROS, M. D.; Evolução de homicídios por área geográfica em Pernambuco entre 1980 e 1988. *Revista de Saúde Pública*. Vol.36, n.4, 462-469, 2002.
- de LIMA, R. S.; BUENO, S.; MINGARDI, G.; Estado, polícias e segurança pública no Brasil. *Revista Direito GV*. Vol.12, n.1, 49-85, 2016.
- DIÁRIO OFICIAL DO ESTADO DE PERNAMBUCO. Portaria GAB/SDS nº1197, de 11 de junho de 2010. Diário Oficial [do] Estado de Pernambuco, Recife, ano 87, n.111, 15 de junho de 2010.
- DIÁRIO OFICIAL DO ESTADO DE PERNAMBUCO. Portaria GAB/SDS nº240, de 19 de julho de 2004: Define a responsabilidade territorial, bem como os limites das áreas de atuação das polícias de Pernambuco, 2004.
- dos SANTOS, M. J.; KASSOUF, A. L.; A cointegration analysis of crime, economic activity, and police performance in São Paulo city. *Journal of Applied Statistics*, Vol.40, n.10, 2087–2109, 2013.
- FAHAD, A.; ALSHATRI, N.; TARI, Z.; ALAMRI, A.; KHALIL, I.; ZOMAYA, A. Y.; FOUFOU, S.; BOURAS, A. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE transactions on emerging topics in computing*, v. 2, n. 3, p. 267-279, 2014.
- FARIAS, C. A.; FIGUEIREDO, A. M.; de LIMA, J. E.; Dependência Espacial e Análise de Agrupamentos de Municípios para Diferentes Tipos de Crimes em Minas Gerais. *Reuna - Belo Horizonte*. Vol.13, n.3, 67-83, 2008.
- FASULO, D.; An Analysis of Recent Work on Clustering Algorithms. *Technical Report. Department of Computer Science & Engineering*. University of Washington, 1999.
- FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L.; Análise de dados: Modelagem multivariada para tomada de decisões. Campus: Elsevier, 2009.

- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; From data mining to knowledge discovery: An overview. In: *Advances in Knowledge Discovery and Data Mining*, AAAI Press/The MIT Press, Inglaterra, 1996.
- FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA – FBSP. Anuário brasileiro de segurança pública. Ano 13. Brasil, 2019.
- FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA – FBSP. Anuário brasileiro de segurança pública. Ano 11. Brasil, 2017.
- FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA - FBSP. O novo Sistema Único de Segurança Pública: questões sobre financiamento da segurança pública, Brasil, 2019.
- FREITAS, C. M. de; FREITAS, J. T. C. de.; Análises de direito comparado sobre segurança, criminalidade e políticas públicas - panorama Brasil e Itália. *JURIS*, Rio Grande, Vol.27, n.2, 65-86, 2017.
- FUND FOR PEACE. Fragile States Index Annual Report 2019.
- GARNICA, N. J.; MURILLO, A. M. O.; Exploration of sexual violence in the city of Bogota: application of a data mining technique. *Revista Criminalidad*, Vol.53, n.2, 145-173, 2011.
- GIDDENS, A.; Sociologia, 4.ª Edição. Lisboa, Fundação Calouste Gulbenkian, 2004.
- GOLDSCHMIDT, R.; BEZERRA, E.; PASSOS, E.; Data mining: conceitos, técnicas, algoritmos, orientações e aplicações. Brasil: Elsevier, 2015.
- HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L.; Análise Multivariada de Dados, 6. ed., São Paulo: Bookman, 2009.
- HAN, J.; KAMBER, M.; PEI, J.; Data Mining: Concepts and Techniques, 3. ed., EUA: Elsevier, 2012.
- HAN, J.; PEI, J.; KAMBER, M.; Data mining: concepts and techniques. [S. l.]: Elsevier, 2011.
- HARTIGAN, J. A.; WONG, M. A.; A K-means clustering algorithm. *Applied Statistics*, Vol. 28, 100-108, 1979.
- HONG, C.; ZHIHONG, L.; HUIDONG, J.; BAOGUANG, X.; MINGLIANG, Q.; A decision support system for detecting serial crimes, Vol. 123, 88-101, 2017.
- HOTELLING, H.; Analysis of a complex of statistical variable into principal component. *J. Educational Psychol*, Vol. 24, n.6, 417-441, 1933.
- IBGE-INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Censo Demográfico: Brasil/Pernambuco, 2010.
- INGRAM, M. C.; MARCHESINI, C. M.; Political geography of violence: Municipal politics and homicide in Brazil. *World Development*, Vol.124, 2019.

- INSTITUTE FOR ECONOMICS & PEACE. Global Peace Index: Measuring Peace in a Complex world. Sidney, 2019.
- INSTITUTO DE PESQUISA ECONÔMICA APLICADA - IPEA. Atlas da Violência: Retrato dos Municípios Brasileiros - 2019. Rio de Janeiro, 2019.
- JAIN, A. K.; Data Clustering: 50 years beyond K-means. *Pattern Recognition Letters*. Vol. 31, n.8, 651-666, 2010.
- JAIN, A. K.; DUBES, R. C.; Algorithms for clustering data. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988.
- JAWADI, F.; MALLICK, S. K.; IDI CHEFFOU, A.; AUGUSTINE, A.; Does higher unemployment lead to greater criminality? Revisiting the debate over the business cycle. *Journal of Economic Behavior and Organization*, 2019.
- JOHNSON, R. A.; WICHERN, D. W.; Applied Multivariate Statistical Analysis, 4. ed., NJ: Prentice hall Englewood, 1992.
- JOLLIFFE, I. T.; Principal Component Analysis, 2. ed., New York: Springer, 2002.
- KEYVANPOUR, M. R.; JAVIDEH, M.; EBRAHIMI, M. R.; Detecting and investigating crime by means of data mining: A general crime matching framework. *Procedia Computer Science*, Vol.3, 872-880, 2011.
- KHAN, J. R.; SAEED, M.; SIDDIQUI, F. A.; MAHMOOD, N.; UL ARIFEEN, Q.; Predictive Policing: A Machine Learning Approach to Predict and Control Crimes in Metropolitan Cities. *Journal of Information na Communication Technology*, Vol.3, n.1, 17-26, 2019.
- KIRGEL, A.; Cyber crime nation typologies: K-means clustering of countries based on cyber crime rates. *International Journal of Cyber Criminology*, Vol.10, n.2, 147-169, 2016.
- KOEGEL, C. J.; DAY, D. M.; The Monetary Costs of Crime for a Sample of Offenders in Ontario. *Canadian Journal of Criminology and Criminal Justice*, Vol.61, n.3, 1-24, 2018.
- LIKAS, A.; VLASSIS, N.; J. VERBEEK, J.; The Global k-means Clustering Algorithm. *Pattern Recognition*, Vol. 36, n.2, 451-461, 2003.
- LIMA, M. L. C.; XIMENES, R. A. A.; FEITOSA, C. L.; SOUZA, E. R.; ALBURQUEQUE, M. F. P. M.; BARROS, M. D. A.; SOUZA, W. V.; LAPA, T. M.; Conglomerados de Violência em Pernambuco, Brasil. *Rev. Panam Salud Publica/Pan Am J Public Health*, Vol. 18, n.2, 122-128, 2005.
- LLOYD, S.; Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, Vol. 28, n.2, 129-137, 1982.
- LOPES, J. M. A.; Políticas de Segurança Pública nos Estados de Minas Gerais e Pernambuco em Perspectiva Comparada.148f. Dissertação (Mestrado em Sociologia) - Universidade Federal de Pernambuco, Recife, 2016.

- MACEDO, A. O. "Polícia, quando quer, faz!": Análise da Estrutura de Governança do "Pacto Pela Vida" de Pernambuco. 164f. Dissertação (Mestrado em Sociologia) - Universidade de Brasília, Brasília, 2012.
- MANNING, C.; RAGHAVAN, P.; SCHÜTZE, M.; Probabilistic Information Retrieval. Press: Cambridge University, 2009.
- MINISTÉRIO DA SEGURANÇA PÚBLICA. Consulta Pública - Política e Plano Nacional de Segurança Pública, Brasil, 2018.
- MORANA, H. C. P.; CÂMARA, F. P.; ARBOLEDA-FLÓREZ, J.; Cluster analysis of a forensic population with antisocial personality disorder regarding PCL-R scores: Differentiation of two patterns of criminal profiles. *Forensic Science International*, Vol.164, 98-101, 2006.
- MOTA, J. C. da; VASCONCELOS, A. G. D.; ASSIS, S. G. de; Correspondence Analysis: a method for Classifying Similar Patterns of Violence Against Women. *Cadernos de Saúde Pública*, Vol.24, n.6, 1397-1406, 2008.
- MURRAY, A. T.; GRUBESIC, T. H.; Exploring Spatial Patterns of Crime Using Non-hierarchical Cluster Analysis. Crime Modeling and Mapping Using Geospatial Technologies. *Geotechnologies and the Environment*, Vol.8, 105-124, 2012.
- MYATT, G. J.; JOHNSON, W. P. Making Sense of Data II: A practical Guide to Data Visualization, Advanced Data Mining Methods, and Applications. New Jersey: John Wiley & Sons, 2009.
- NEISSE, A. C.; HONGYU, K.; Application of Principal Components and Factor Analysis to Crime Data From 26 Us States. *Engineering and Science*, Vol. 2, n.5, 2016.
- NÓBREGA JÚNIOR, J. M. P.; Políticas Públicas e Segurança Pública em Pernambuco: o case pernambucano e a redução da violência homicida. *Latitude*, Vol. 8, n.2, 315-335, 2014.
- OATLEY, G. C.; EWART, B. W.; Crimes analysis software: "pins in maps", clustering and Bayes net prediction. *Expert Systems with Applications*, Vol.25, n.4, 569-588, 2003.
- PACIFICO, L. D. S.; LUDERMIR, T. B.; A Group Search Optimization Method for Data Clustering. In: Intelligent Systems (BRACIS), Brazilian Conference on. IEEE, 342-347, 2014.
- PEARSON, K.; On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 559-572, 1901.
- PEREIRA, D. V. S.; MOTA, C. M. M.; ANDRESEN, M. A. Social Disorganization and Homicide in Recife, Brazil. *International Journal of Offender Therapy and Comparative Criminology*. Vol.61, n.14, 1570-1592, 2017.
- PEREIRA, D. V. S.; MOTA, C. M. M.; ANDRESEN, M. A.; The Homicide Drop in Recife, Brazil. *Homicide Studies*, Vol. 21, n. 1, 21-38, 2016.

- QAZI, N.; WONG, B. L. W.; An interactive human centered data science approach towards crime pattern analysis. *Information Processing & Management*, Vol.56, n.6, 2019.
- SAFHI, H. M.; FRIKH, B.; OUHBI, B. Assessing Reliability of Big Data Knowledge Discovery Process. *Procedia Computer Science*, Vol.148, 30-36, 2019.
- SAURET, G. Estatísticas pela Vida. Recife: Editora Bagaço, 2012.
- SCHUH, G.; REINHART, G.; PROTE, J. P.; SAUERMAN, F.; HORSTHOFER, J.; OPPOLZER, F.; KNOLL, D.; Data mining definitions and applications for the management of production complexity. *Procedia CIRP*, Vol.81, 874-879, 2019.
- SECRETARIA DE DEFESA SOCIAL DO ESTADO DE PERNAMBUCO - SDS/PE. Crimes Violentos Contra o Patrimônio: Número de ocorrências de CVP em Pernambuco por Município. Jan-Out de 2019, 2019.
- SECRETARIA DE DEFESA SOCIAL DO ESTADO DE PERNAMBUCO - SDS/PE. Crimes Violentos Letais Intencionais: Número de vítimas de CVLI em Pernambuco po Município. Jan-Out de 2019, 2019.
- SILVA, C.; RIBEIRO, B.; Aprendizagem Computacional em Engenharia, 1. ed., Imprensa da Universidade de Coimbra: Coimbra University Press, 2018.
- SJÖDIN, A. K.; WALLINIUS, M.; BILLSTEDT, B. H.; NILSSON, T.; Evidence for Two Levels of Intimate Partner Violence and Aggression Among Incarcerated Young Male Violent Offenders. *International Journal of Offender Therapy and Comparative Criminology*, Vol.62, n.10, 3097-3116, 2018.
- SOARES, L. E.; Segurança Pública: presente e futuro. *Estudos Avançados*, Vol.20, n.56, 91-106, 2006.
- TAYAL, D. K.; JAIN, A.; ARORA, S.; AGARWAL, S.; GUPTA, T.; TYAGI, N.; Crime Detection and Criminal Identification in India Using Data Mining Techniques. *AI & Society*, Vol. 30, n.1, 117-127, 2014.
- TURBAN, ARONSON, LIANG.; *Decision Support Systems and Intelligent Systems*, 7 ed., Prentice Hall, New Jersey, 2005.
- VELMURUGAN, T. Performance based analysis between k-Means and Fuzzy C-Means clustering algorithms for connection oriented telecommunication data. *Applied Soft Computing*, Vol. 19, 134-146, 2014.
- WASELFISZ, J. J.; Mapa da Violência (2012): Os novos padrões da violência no Brasil: Secretaria Geral da Presidência da República, 2012.
- WASELFISZ, J. J.; Mapa da Violência (2014): Os jovens do Brasil. Brasília, Brasil: Secretaria Geral da Presidência da República, 2014.

- WANG, S.; WANG, X.; YE, P.; YUAN, Y.; LIU, S.; WANG, F.; Parallel Crime Scene Analysis Based on ACP Approach. *IEEE Transactions on Computational Social Systems*, Vol. 5, n. 1, 2018.
- WEISBURD, D.; AMRAM, S.; The law of concentrations of crime at place: the case of Tel Aviv-Jaffa. *Police Practice and Research: International Journal*. Vol.15, 101-114, 2014.
- WITTEN, I. H.; FRANK, E.; HALL, M. A.; Data Mining: Practical Machine Learning Tools and Techniques, 2. ed., Amsterdam: Morgan Kaufmann Publishers, 2011.
- WORTLEY, R.; MAZEROLLE, L. Environmental Criminology and Crime Analysis: Situation the theory, analytic approach and application. In: R. Wortley & L. Mazerolle (eds). *Environmental criminology and crime analysis*, Ingleterra , 2008.
- WU, J.; CHEN, J.; XIONG, H.; XIE, M.; External validation measures for K-means clustering: A data distribution perspective. *Expert Systems with Applications*, Vol. 36, 6050-6061, 2009.
- XU, D.; TIAN, Y. A.; Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, Vol. 2, 165-193, 2015.
- XU, R.; WUNSCHLL, D.; Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, Vol.16, n.3, 645–678, 2005.
- ZHU, E.; ZHANG, Y.; WEN, P.; LIU, F.; Fast and stable clustering analysis based on Grid-mapping K-means algorithm and new clustering validity index. *Neurocomputing*, Vol.363, 149–170, 2019.

APÊNDICE A – DADOS DE CRIMINALIDADE MUNICIPAIS (2018)

Crime	Descrição
C1	Tráfico de entorpecentes
C2	Homicídio doloso
C3	Lesão corporal dolosa
C4	Lesão corporal seguida de morte
C5	Latrocínio
C6	Furto de veículo
C7	Roubo de veículo
C8	Roubo a instituição financeira
C9	Roubo de carga
C10	Roubo a estabelecimento comercial
C11	Roubo a residência

Município	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Abreu e Lima	97	41	171	0	1	28	170	0	1	60	33
Afogados da Ingazeira	12	2	138	0	0	10	2	0	0	4	7
Afrânio	3	2	11	0	0	3	14	0	0	1	1
Agrestina	1	13	56	0	1	17	142	0	0	11	11
Água Preta	8	10	31	0	0	8	13	0	1	1	0
Águas Belas	4	16	98	0	0	8	26	0	14	1	3
Alagoinha	2	6	38	0	0	2	9	1	0	2	3
Aliança	13	13	89	0	1	7	25	0	2	14	12
Altinho	5	11	52	0	0	10	68	0	0	4	9
Amaraji	0	17	23	0	0	4	25	0	2	2	1
Angelim	2	7	31	0	0	7	18	0	1	0	2
Araçoiaba	5	7	19	0	0	1	3	0	0	1	2
Araripina	32	20	140	2	3	31	31	0	1	23	12
Arcoverde	48	19	248	0	1	63	32	0	2	19	2
Barra de Guabiraba	1	10	22	0	1	4	27	0	0	4	8
Barreiros	22	14	46	0	0	9	19	0	0	5	4
Belém de Maria	2	4	11	0	0	2	23	0	0	0	4
Belém do São Francisco	13	6	67	0	1	0	4	0	11	2	1
Belo Jardim	24	30	178	1	1	41	117	0	0	29	24
Betânia	3	2	18	0	0	1	4	0	0	2	0
Bezerros	36	28	115	0	3	52	248	1	7	23	5
Bodocó	3	2	78	0	0	7	3	0	0	1	2
Bom Conselho	4	15	70	0	0	8	16	0	0	3	1
Bom Jardim	5	12	69	1	1	15	45	0	0	4	8
Bonito	10	19	72	0	0	7	74	0	1	6	13
Brejão	2	4	38	0	0	6	14	0	0	0	1
Brejinho	0	0	26	0	0	2	11	0	0	0	0

Município	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Brejo da Madre de Deus	16	28	97	0	0	25	138	1	0	16	12
Buenos Aires	1	2	50	0	0	1	11	1	0	0	2
Buíque	5	21	99	0	0	10	9	0	0	4	2
Cabo de Santo Agostinho	134	156	363	0	5	63	397	0	37	139	84
Cabrobó	34	6	112	0	0	4	11	0	3	10	0
Cachoeirinha	2	8	53	0	0	10	25	0	0	2	4
Caetés	2	7	24	0	0	11	17	0	1	2	4
Calçado	0	3	12	0	1	4	40	0	0	2	3
Calumbi	0	0	21	0	0	1	1	0	1	1	2
Camaragibe	125	52	284	0	1	80	245	0	4	81	48
Camocim de São Félix	8	9	52	0	0	13	31	0	0	10	0
Camutanga	3	5	25	0	0	1	21	0	1	1	7
Canhotinho	28	10	54	0	0	19	45	0	0	1	10
Capoeiras	3	7	17	0	0	17	17	1	1	1	2
Carnaíba	1	1	24	0	0	2	5	1	0	1	2
Carnaubeira da Penha	1	1	7	0	0	0	0	0	0	0	0
Carpina	57	27	177	0	0	62	149	0	0	28	17
Caruaru	272	140	749	1	5	466	1054	1	17	169	118
Casinhas	0	8	19	0	0	5	14	0	0	2	3
Catende	20	17	83	0	1	8	22	0	0	0	4
Cedro	1	1	20	0	0	0	3	0	0	0	0
Chã de Alegria	3	6	25	0	0	2	24	0	0	2	1
Chã Grande	4	6	46	0	0	5	22	0	0	5	4
Condado	22	14	61	0	0	6	28	0	2	5	4
Correntes	2	5	44	0	0	1	6	1	1	2	2
Cortês	0	4	17	0	0	3	9	0	1	1	0
Cumarú	2	3	21	0	0	4	21	1	0	5	5
Cupira	18	24	33	0	1	13	132	0	2	14	9
Custódia	12	11	60	0	2	11	18	0	4	1	6
Dormentes	1	5	13	0	0	2	3	0	0	2	1
Escada	30	32	81	0	1	21	189	0	44	12	13
Exu	5	4	59	0	0	3	3	0	0	1	0
Feira Nova	6	11	49	0	0	1	28	0	0	4	9
Fernando de Noronha	1	0	24	0	0	0	0	0	0	0	1
Ferreiros	8	4	27	0	1	2	19	0	0	3	3
Flores	5	3	25	1	0	11	7	0	3	0	0
Floresta	16	16	100	0	0	8	10	0	2	1	4
Frei Miguelinho	0	4	10	0	0	4	21	0	0	2	6
Gameleira	5	25	16	0	1	3	33	0	2	0	3
Garanhuns	53	34	384	0	1	118	151	0	3	23	18
Glória Do Goitá	6	10	39	0	2	14	70	0	1	7	12
Goiana	55	36	144	0	2	59	168	0	20	42	31
Granito	0	2	24	0	0	1	0	0	0	1	0
Gravatá	55	48	191	0	3	60	188	0	8	28	12
Iati	2	3	29	0	0	3	6	0	2	0	1

Município	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Ibimirim	4	11	34	0	0	0	5	0	1	2	1
Ibirajuba	0	0	18	0	0	1	25	0	0	2	5
Igarassu	92	89	164	0	2	50	210	0	13	54	44
Iguaraci	0	2	16	0	0	8	2	0	0	2	2
Inajá	0	3	18	0	0	4	4	0	0	0	0
Ingazeira	0	2	4	0	0	0	0	0	0	0	0
Ipojuca	75	75	175	0	2	15	69	0	4	34	14
Ipubi	9	5	54	0	1	9	12	0	0	2	2
Itacuruba	0	0	15	0	0	0	0	1	0	0	0
Itaíba	0	10	44	0	0	4	3	0	0	1	0
Ilha De Itamaracá	32	22	98	0	0	10	9	0	0	11	23
Itambé	10	16	64	0	0	25	254	0	1	15	20
Itapetim	1	1	43	0	1	4	5	0	0	3	0
Itapissuma	39	9	35	0	0	3	28	0	3	8	9
Itaquitinga	0	8	56	0	0	2	10	0	0	0	0
Jaboatão Dos Guararapes	335	261	1188	2	9	377	647	0	35	272	120
Jaqueira	2	7	12	0	0	2	8	0	0	2	0
Jataúba	7	2	40	0	0	9	13	0	0	0	1
Jatobá	3	1	39	0	0	1	6	0	1	3	0
João Alfredo	2	2	53	0	0	6	23	0	0	8	1
Joaquim Nabuco	12	16	33	0	0	3	17	0	4	1	2
Jucati	0	2	16	0	0	10	17	0	0	0	2
Jupi	3	3	15	0	0	8	59	0	1	3	3
Jurema	1	7	20	0	1	10	21	0	0	1	3
Lagoa Do Carro	10	10	47	0	0	1	8	0	0	5	0
Lagoa De Itaenga	2	15	24	0	1	3	74	0	1	10	17
Lagoa Do Ouro	0	4	19	0	0	0	3	0	0	0	0
Lagoa Dos Gatos	0	4	32	0	0	2	22	0	0	4	1
Lagoa Grande	6	7	39	0	0	7	12	0	0	1	0
Lajedo	5	30	93	0	4	33	183	0	1	14	11
Limoeiro	21	17	129	0	0	33	56	0	0	10	10
Macaparana	8	2	72	0	0	13	13	0	0	4	7
Machados	0	4	27	0	0	2	19	0	0	0	1
Manari	0	1	8	0	0	0	0	0	0	1	1
Maraial	0	5	17	0	0	1	1	0	0	1	2
Mirandiba	4	2	50	0	0	0	4	0	1	3	0
Moreno	20	27	60	0	1	3	89	0	13	13	13
Nazaré Da Mata	8	16	85	0	2	8	27	0	0	0	7
Olinda	392	136	955	0	1	230	623	0	6	221	160
Orobó	0	3	34	0	0	5	17	0	0	3	5
Orocó	12	3	22	0	0	4	6	0	2	1	0
Ouricuri	23	13	98	0	1	10	5	0	2	7	6
Palmares	48	32	103	0	1	25	36	0	18	5	5
Palmeirina	1	5	12	0	0	0	4	0	0	0	3
Panelas	0	10	41	0	1	8	73	0	0	5	11
Paranatama	0	3	20	0	0	5	9	0	2	1	1
Parnamirim	4	4	32	0	0	0	1	0	2	0	0
Passira	5	1	51	0	0	7	25	0	0	4	1
Paudalho	20	37	140	0	3	12	127	0	5	28	18
Paulista	290	111	687	1	3	128	432	1	5	108	67
Pedra	8	4	49	0	0	1	7	0	0	2	2
Pesqueira	32	20	111	0	0	32	67	0	6	16	16
Petrolândia	7	9	78	0	0	8	7	0	3	2	0
Petrolina	199	119	1099	2	3	249	234	0	4	107	54

Município	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Poção	3	5	27	0	0	2	9	0	0	1	0
Pombos	13	13	41	0	1	12	43	1	7	14	2
Primavera	0	12	13	0	0	7	12	0	0	0	1
Quipapá	0	7	30	0	0	8	38	0	4	2	1
Quixaba	1	0	8	0	0	2	0	0	0	0	0
Recife	1275	519	3775	0	7	1175	3054	0	51	1322	422
Riacho Das Almas	0	5	37	0	1	12	36	0	0	3	13
Ribeirão	11	21	52	0	0	13	70	1	24	11	9
Rio Formoso	2	13	26	0	0	1	22	0	0	1	2
Sairé	0	3	24	0	0	3	22	0	0	3	10
Salgadinho	0	0	12	0	0	2	4	0	0	0	2
Salgueiro	43	14	191	1	1	17	26	0	6	7	5
Saloá	6	6	44	0	0	8	5	0	5	0	0
Sanharó	4	3	46	0	0	13	38	0	0	4	6
Santa Cruz	3	0	14	0	0	0	1	0	0	0	3
Santa Cruz Da Baixa Verde	1	4	16	0	0	4	6	0	0	0	1
Santa Cruz Do Capibaribe	68	36	179	0	1	235	865	4	0	61	43
Santa Filomena	0	1	17	0	0	1	0	1	0	0	0
Santa Maria Da Boa Vista	7	18	65	0	0	2	7	0	7	4	2
Santa Maria Do Cambucá	3	3	27	0	0	5	27	0	0	0	4
Santa Terezinha	4	1	35	0	0	0	2	0	0	0	0
São Benedito Do Sul	0	2	14	0	0	1	7	0	0	2	2
São Bento Do Una	4	16	89	0	1	30	103	2	1	4	17
São Caitano	15	23	60	0	0	9	88	0	6	19	39
São João	0	3	40	0	0	15	47	0	0	3	4
São Joaquim Do Monte	7	15	47	0	0	6	67	0	0	3	7
São José Da Coroa Grande	12	15	29	0	1	5	25	0	0	0	2
São José Do Belmonte	1	5	52	0	0	8	3	0	2	3	1
São José Do Egito	14	3	90	0	0	19	11	0	0	4	4
São Lourenço Da Mata	77	72	200	0	1	258	229	0	17	49	32
São Vicente Ferrer	6	8	50	1	2	4	20	0	0	7	3
Serra Talhada	31	30	217	1	0	29	34	0	6	11	8
Serrita	2	0	35	0	0	2	6	0	1	0	0

Município	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Sertânia	10	9	76	0	2	10	23	0	7	4	8
Sirinhaém	9	19	36	0	1	3	22	1	0	6	5
Moreilândia	0	3	30	0	0	0	5	0	0	1	1
Solidão	0	1	7	0	0	2	0	0	0	0	1
Surubim	12	17	129	0	2	54	145	3	0	28	14
Tabira	4	5	68	0	0	10	1	0	0	3	3
Tacaimbó	3	2	17	0	0	6	44	0	1	4	14
Tacaratu	2	4	23	0	0	7	5	1	1	1	3
Tamandaré	11	17	70	0	0	7	24	0	0	2	7
Taquaritinga Do Norte	8	15	47	0	2	46	176	0	2	6	7
Terezinha	1	1	16	0	0	3	0	0	1	0	0
Terra Nova	0	1	33	0	0	0	0	0	0	0	0
Timbaúba	49	38	88	0	0	23	48	0	0	40	10
Toritama	19	26	93	0	0	44	133	0	0	13	1
Tracunhaém	3	5	35	0	0	2	10	0	0	6	4
Trindade	5	7	75	0	2	7	13	0	0	5	1
Triunfo	1	1	24	0	0	6	2	0	0	0	0
Tupanatinga	2	7	23	0	0	2	3	0	0	1	1
Tuparetama	2	1	19	0	0	4	2	0	0	0	0
Venturosa	4	8	42	0	0	4	10	0	0	2	0
Verdejante	3	2	17	0	1	0	3	1	0	1	1
Vertente Do Lério	0	5	14	0	0	2	8	0	0	1	0
Vertentes	3	13	34	0	0	15	62	0	0	6	4
Vicência	25	16	52	0	0	2	33	0	0	9	3
Vitória De Santo Antão	115	58	193	0	6	94	277	0	18	38	21
Xexéu	6	13	19	0	0	1	29	0	17	2	1

APÊNDICE B – COEFICIENTES DE PONDERAÇÃO DAS COMPONENTES PRINCIPAIS

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
C1	-0.35087536	0.10772783	-0.06130858	0.14446915	-0.011317707	-0.274341363	0.00756038	-0.052636743	-0.410174230	0.558307327	0.531490444
C2	-0.35292030	-0.03749828	-0.04508496	-0.04008314	0.033803190	-0.345677156	-0.11414289	0.561384006	0.632526261	0.146496875	-0.005922267
C3	-0.34975878	0.04417959	-0.05757038	0.20163358	-0.040677651	-0.235160664	0.48868343	0.072522577	-0.295720527	0.013428066	-0.665649560
C4	-0.10921453	-0.82942403	0.10291821	0.40027087	-0.324867789	0.035842191	-0.07783055	-0.115034904	0.046790143	0.004745067	0.031509227
C5	-0.26175416	-0.42689211	0.11176591	-0.38955765	0.744128042	0.005784535	0.07481786	-0.115682820	-0.096474229	-0.051220502	0.018167454
C6	-0.34760460	0.07091745	0.05470809	0.14041560	-0.018797692	0.637059738	0.26853711	0.449131589	-0.112355368	-0.271393424	0.291760684
C7	-0.34538113	0.16664577	0.09584419	0.05844585	0.039882386	0.515559843	-0.33849309	-0.300299007	0.204091675	0.455285251	-0.345327881
C8	-0.02772824	0.13227594	0.96938720	-0.05107017	-0.112770425	-0.151220199	0.05043292	0.002603998	-0.005240136	-0.026493543	0.023455829
C9	-0.26028265	-0.11829318	-0.09368450	-0.75594920	-0.567878105	0.040295223	0.04615509	-0.066128354	-0.085228505	-0.009799728	-0.007159239
C10	-0.34403068	0.19116256	-0.07924212	0.13598996	-0.025696291	-0.129602978	0.29748579	-0.594568225	0.418082693	-0.351774593	0.249644317
C11	-0.35009911	0.11841552	-0.02938978	0.09592126	-0.001411298	-0.188918535	-0.67571157	0.020310341	-0.310417091	-0.508451441	-0.075968756