

UNIVERSIDADE NOVE DE JULHO – UNINOVE
Programa de Pós-Graduação em Cidades Inteligentes e Sustentáveis

Giovanna Petrica Neves

**ANÁLISE DAS RELAÇÕES SOCIOESPACIAIS DOS CRIMES DE FURTO E
ROUBO NA SUBPREFEITURA DE ITAQUERA, SÃO PAULO/SP - BRASIL: UMA
ABORDAGEM UTILIZANDO INTELIGÊNCIA ARTIFICIAL**

São Paulo
2023

GIOVANNA PETRICA NEVES

**ANÁLISE DAS RELAÇÕES SOCIOESPACIAIS DOS CRIMES DE FURTO E
ROUBO NA SUBPREFEITURA DE ITAQUERA, SÃO PAULO/SP - BRASIL: UMA
ABORDAGEM UTILIZANDO INTELIGÊNCIA ARTIFICIAL
ANALYSIS OF THE SOCIO-SPATIAL RELATIONS OF THEFT AND ROBBERY
CRIMES IN THE ITAQUERA/SP SUB-PREFECTURE, SÃO PAULO/SP - BRAZIL:
AN APPROACH USING ARTIFICIAL INTELLIGENCE.**

Dissertação de mestrado apresentado ao Programa de Pós-Graduação em Cidades Inteligentes e Sustentáveis da Universidade Nove de Julho – UNINOVE, como requisito obrigatório para obtenção do grau de Mestre em Cidades Inteligentes e Sustentáveis

Orientador: Prof. Dr. Cristiano Capellani Quaresma

Coorientador: Prof. Dr. Sidnei Alves de Araújo

São Paulo

2023

Neves, Giovanna Petrica.

Análise das relações socioespaciais dos crimes de furto e roubo na Subprefeitura de Itaquera, São Paulo/SP - Brasil: uma abordagem utilizando inteligência artificial. Giovanna Petrica Neves. 2023.

106 f.

Dissertação (Mestrado) - Universidade Nove de Julho - UNINOVE, São Paulo, 2023.

Orientador (a): Prof. Dr. Cristiano Capellani Quaresma.

1. Criminologia espacial. 2. Roubo. 3. Furto. 4. Crimes contra o patrimônio. 5. Inteligência artificial. 6. Mineração de dados. 7. Análise espacial.

**ANÁLISE DAS RELAÇÕES SOCIOESPACIAIS DOS CRIMES DE FURTO E
ROUBO NA SUBPREFEITURA DE ITAQUERA, SÃO PAULO/SP - BRASIL: UMA
ABORDAGEM UTILIZANDO INTELIGÊNCIA ARTIFICIAL**

Por

Giovanna Petrica Neves

Dissertação de mestrado apresentado ao Programa de Pós-Graduação em Cidades Inteligentes e Sustentáveis da Universidade Nove de Julho – UNINOVE, como requisito obrigatório para obtenção do grau de Mestre em Cidades Inteligentes e Sustentáveis.

Prof. Dr. Cristiano Capellani Quaresma – Universidade Nove de Julho –
UNINOVE - orientador

Orientador Prof. Dr. Sidnei Alves de Araújo – Universidade Nove de Julho –
UNINOVE – coorientador

Profa. Dra. Alessandra Cristina Guedes Pellini – Universidade Nove de Julho - UNINOVE

Prof. Dr. Peterson Adriano Belan – Universidade Nove de Julho –
UNINOVE

Prof. Dr. Vitor Vieira Vasconcelos – Universidade Federal do ABC – UFABC

São Paulo, 11 de setembro de 2023

Dedicatória

Primeiramente, dedico este trabalho a Deus, que me guiou em mais esta fase significativa da minha trajetória. À minha querida família - Célia, Pietra, Sidnei, Nicole, Henry, Tio Antônio, Vô João e Vó Tereza - meu eterno agradecimento pelo apoio incondicional.

Não poderia deixar de expressar minha profunda gratidão ao Dr. Cristiano Capellani Quaresma, meu orientador, cuja dedicação, sabedoria e generosidade foram fundamentais na realização deste trabalho.

Não poderia deixar de expressar também minha profunda gratidão ao meu coorientador, Dr. Sidnei Alves de Araújo, cuja dedicação, sabedoria e generosidade foram fundamentais na realização deste trabalho.

Agradecimentos

Meus sinceros agradecimentos, em especial, aos meus orientadores, Dr. Cristiano Capellani Quaresma e Prof. Dr. Sidnei Alves de Araújo, pelo incentivo, paciência e aconselhamento que contribuíram decisivamente para o meu desenvolvimento acadêmico.

À Universidade Nove de Julho – UNINOVE, sou grata pela infraestrutura, pela oportunidade oferecida e pela bolsa de estudos concedida.

À CAPES, pelo apoio financeiro na forma de bolsa de estudos.

Expresso minha profunda gratidão à minha família: minha mãe, Célia Cristina Petrica, e ao meu padrasto, Sidnei, pelo constante acolhimento e por serem minha referência de estudo e trabalho. A minha irmã gêmea, Pietra Petrica, que esteve ao meu lado, sempre disposta a ajudar, agradeço por compartilhar sua generosidade e amor incondicionalmente.

Aos meus irmãos, Nicole Petrica Araújo e Henry Petrica Araújo, meus avós, Tereza de Jesus Petrica e João Batista Petrica, e meus tios, agradeço o suporte emocional e auxílio na estruturação psicológica deste estudo.

E, por fim, expresso minha gratidão aos membros da banca pelas suas inestimáveis contribuições.

RESUMO

Introdução: Os crimes de furto e roubo, categorizados como contra o patrimônio, ameaçam a segurança pública e a qualidade de vida nas cidades. Estudos buscam entender suas causas e padrões, visando melhorar políticas de segurança urbana. Teorias apontam variáveis individuais, sociodemográficas e espaciais como explicativas. Devido às particularidades de cada local, o tema exige mais pesquisas, especialmente na criminologia espacial, que estuda a relação entre crimes e elementos socioespaciais. Identificar tais elementos e sua ligação com os padrões criminais é crucial para criar políticas de prevenção e combate à violência urbana.

Objetivo: Este estudo analisou os padrões de distribuição espacial de crimes de furto e roubo e suas relações com atributos socioespaciais urbanos na subprefeitura de Itaquera, Zona Leste de São Paulo/SP – Brasil. A escolha de Itaquera deveu-se ao alto registro de crimes patrimoniais na subprefeitura e à presença de elementos associados a esses crimes na literatura.

Metodologia: Trata-se de um estudo exploratório quantitativo, baseado na análise documental e de dados secundários dos crimes de furto e roubo ocorridos entre 2017 e 2021, e de indicadores sociodemográficos. As análises apoiaram-se em técnicas de análise espacial, Sistema de Informação Geográfica (SIG), Inteligência Artificial (IA) e Mineração de Dados (MD).

Resultados: Com apoio da IA e MD, a abordagem encontrou padrões contextuais com boas precisões (acima de 70%) que explicam relações entre variáveis sociodemográficas e as ocorrências de furtos e roubos. Comparados a outros estudos, os achados revelaram convergências e divergências. Em adição, foram revelados padrões associados aos roubos com base no perfil das vítimas.

Conclusão: Ao aprofundar a compreensão de como os padrões espaciais dos crimes de furto e roubo são influenciados por características sociodemográficas, essa abordagem traz uma valiosa contribuição científica, podendo ser aplicada em outros contextos socioespaciais e orientar futuras políticas de segurança urbana. Além disso, este estudo contribui para o alcance de Objetivos do Desenvolvimento Sustentável (ODS) estabelecidos pela ONU, incluindo a Redução da Desigualdade (ODS 10), Cidades e Comunidades Sustentáveis (ODS 11), e Paz, Justiça e Instituições Eficazes (ODS 16).

Palavras-chave: criminologia espacial; roubo; furto; crimes contra o patrimônio; inteligência artificial; mineração de dados; análise espacial.

ABSTRACT

Introduction: Theft and robbery crimes, categorized as property-related offenses, pose threats to public safety and the quality of life in cities. Studies aim to comprehend their causes and patterns, with the intention of enhancing urban security policies. Theories point to individual, sociodemographic, and spatial variables as explanatory factors. Due to the unique characteristics of each locality, the topic demands further research, especially within spatial criminology, which investigates the relationship between crimes and socio-spatial elements. Identifying such elements and their connection to criminal patterns is crucial for devising policies for urban violence prevention and intervention. **Objective:** This study analyzed the spatial distribution patterns of theft and robbery crimes and their associations with urban socio-spatial attributes in the subprefecture of Itaquera, Eastern Zone of São Paulo/SP – Brazil. The selection of Itaquera was driven by the high incidence of property-related crimes in the subprefecture and the presence of elements associated with these crimes in the literature. **Methodology:** This is a quantitative exploratory study, based on documentary analysis and secondary data of theft and robbery crimes that occurred between 2017 and 2021, along with sociodemographic indicators. The analyses were supported by spatial analysis techniques, Geographic Information System (GIS), Artificial Intelligence (AI), and Data Mining (DM). **Results:** With the assistance of AI and DM, the approach identified contextual patterns with good accuracies (above 70%) that explain relationships between sociodemographic variables and occurrences of theft and robbery. In comparison to other studies, the findings revealed both convergences and divergences. Additionally, patterns associated with robberies were unveiled based on victim profiles. **Conclusion:** By deepening the understanding of how spatial patterns of theft and robbery crimes are influenced by sociodemographic characteristics, this approach provides a valuable scientific contribution, which can be applied in other socio-spatial contexts and guide future urban security policies. Furthermore, this study contributes to achieving the Sustainable Development Goals (SDGs) established by the UN, including Reduced Inequality (SDG 10), Sustainable Cities and Communities (SDG 11), and Peace, Justice, and Strong Institutions (SDG 16).

Keywords: spatial criminology; robbery; theft; property crimes; artificial intelligence; data mining; spatial analysis.

Lista de Figuras

FIGURA 1 – INTERAÇÕES ENTRE IA, CIÊNCIA DE DADOS E MINERAÇÃO DE DADOS	33
FIGURA 2 – PASSOS DO KDD	35
FIGURA 3 – MAPA DE LOCALIZAÇÃO DA SUBPREFEITURA DE ITAQUERA, SÃO PAULO/SP-BRASIL, 2023.....	41
FIGURA 4 – DIAGRAMA DE FUNCIONAMENTO DA ABORDAGEM DE IA PROPOSTA.....	46
FIGURA 5 – DISTRIBUIÇÃO ESPACIAL DA RENDA MÉDIA (REM).....	52
FIGURA 6 – DISTRIBUIÇÃO ESPACIAL DO PERCENTUAL DE POPULAÇÃO JOVEM (PPJ).....	53
FIGURA 7 – DISTRIBUIÇÃO ESPACIAL DE DENSIDADE POPULACIONAL (DEP).....	54
FIGURA 8 – DISTRIBUIÇÃO ESPACIAL DO PERCENTUAL DE POPULAÇÃO MASCULINA (PPM).....	55
FIGURA 9 – DISTRIBUIÇÃO ESPACIAL DO PERCENTUAL DE POPULAÇÃO ALFABETIZADA (PPA)	56
FIGURA 10 – DISTRIBUIÇÃO ESPACIAL DAS OCORRÊNCIAS DE FURTOS E ROUBOS NA SUBPREFEITURA DE ITAQUERA, SÃO PAULO/SP-BRASIL NO ANO DE 2017	57
FIGURA 11 – DISTRIBUIÇÃO ESPACIAL DAS OCORRÊNCIAS DE FURTOS E ROUBOS NA SUBPREFEITURA DE ITAQUERA, SÃO PAULO/SP-BRASIL NO ANO DE 2018	58
FIGURA 12 – DISTRIBUIÇÃO ESPACIAL DAS OCORRÊNCIAS DE FURTOS E ROUBOS NA SUBPREFEITURA DE ITAQUERA, SÃO PAULO/SP-BRASIL NO ANO DE 2019	58
FIGURA 13 – DISTRIBUIÇÃO ESPACIAL DAS OCORRÊNCIAS DE FURTOS E ROUBOS NA SUBPREFEITURA DE ITAQUERA, SÃO PAULO/SP-BRASIL NO ANO DE 2020	59
FIGURA 14 – DISTRIBUIÇÃO ESPACIAL DAS OCORRÊNCIAS DE FURTOS E ROUBOS NA SUBPREFEITURA DE ITAQUERA, SÃO PAULO/SP-BRASIL NO ANO DE 2021	59
FIGURA 15 – DISTRIBUIÇÃO ESPACIAL DAS OCORRÊNCIAS DE FURTO POR SETOR CENSITÁRIO DA SUBPREFEITURA DE ITAQUERA/SP – BRASIL, NO PERÍODO DE 2017 A 2021	60
FIGURA 16 – DISTRIBUIÇÃO ESPACIAL DAS OCORRÊNCIAS DE ROUBO POR SETOR CENSITÁRIO DA SUBPREFEITURA DE ITAQUERA/SP – BRASIL, NO PERÍODO DE 2017 A 2021	61
FIGURA 17 – DISTRIBUIÇÃO ESPACIAL DAS OCORRÊNCIAS DE CRIMES (FURTO + ROUBO) POR SETOR CENSITÁRIO DA SUBPREFEITURA DE ITAQUERA/SP – BRASIL, NO PERÍODO DE 2017 A 2021	62
FIGURA 18 – DISTRIBUIÇÃO DOS FURTOS NOS SETORES CENSITÁRIOS	65
FIGURA 19 – DISTRIBUIÇÃO DOS ROUBOS NOS SETORES CENSITÁRIOS.....	66

FIGURA 20 – FURTOS OCORRIDOS NOS SETORES CENSITÁRIOS COM RELAÇÃO À VARIÁVEL REM	66
FIGURA 21 – ROUBOS OCORRIDOS NOS SETORES CENSITÁRIOS COM RELAÇÃO À VARIÁVEL REM	67
FIGURA 22 – FURTOS OCORRIDOS NOS SETORES CENSITÁRIOS COM RELAÇÃO À VARIÁVEL PPJ.....	68
FIGURA 23 – ROUBOS OCORRIDOS NOS SETORES CENSITÁRIOS COM RELAÇÃO À VARIÁVEL PPJ.....	69
FIGURA 24 – FURTOS OCORRIDOS NOS SETORES CENSITÁRIOS COM RELAÇÃO À VARIÁVEL DEP	69
FIGURA 25 – ROUBOS OCORRIDOS NOS SETORES CENSITÁRIOS COM RELAÇÃO À VARIÁVEL DEP	70
FIGURA 26 – FURTOS OCORRIDOS NOS SETORES CENSITÁRIOS COM RELAÇÃO À VARIÁVEL PPM.....	71
FIGURA 27 – ROUBOS OCORRIDOS NOS SETORES CENSITÁRIOS COM RELAÇÃO À VARIÁVEL PPM.....	71
FIGURA 28 – FURTOS OCORRIDOS NOS SETORES CENSITÁRIOS COM RELAÇÃO À VARIÁVEL PPA	72
FIGURA 29 – ROUBOS OCORRIDOS NOS SETORES CENSITÁRIOS COM RELAÇÃO À VARIÁVEL PPA	73
FIGURA 30 – AD FURTO	76
FIGURA 31 – AD ROUBO.....	79
FIGURA 32 – AD CRIMES.....	83
FIGURA 33 – AD GERADA PARA MAPEAR PADRÕES DOS CRIMES A PARTIR DO PERFIL DAS VÍTIMAS	87

Lista de Tabelas

TABELA 1 – TÉCNICAS DE IA UTILIZADAS NOS TRABALHOS DA LITERATURA.....	32
TABELA 2 – VARIÁVEIS DE RESPOSTA.....	43
TABELA 3 – VARIÁVEIS EXPLICATIVAS COM BASE NA TEORIA DA DESORGANIZAÇÃO SOCIAL	44
TABELA 4 – ORIGEM DOS DADOS RELATIVOS ÀS VARIÁVEIS EXPLICATIVAS BASEADAS NA TEORIA DA DESORGANIZAÇÃO SOCIAL.....	45
TABELA 5 – EXEMPLO DE MATRIZ DE CONFUSÃO	49
TABELA 6 – NÍVEIS DE CONCORDÂNCIA DO COEFICIENTE KAPPA	49
TABELA 7 – SÍNTESE METODOLÓGICA.....	50
TABELA 8 – CATEGORIZAÇÃO DAS VARIÁVEIS EXPLICATIVAS.....	64
TABELA 9 – CATEGORIZAÇÃO VARIÁVEIS DE RESPOSTA.....	64
TABELA 10 – MATRIZ DE CONFUSÃO REFLETINDO OS RESULTADOS DA AD PARA A VARIÁVEL “FURTO”	77
TABELA 11 – MATRIZ DE CONFUSÃO REFLETINDO OS RESULTADOS DA AD PARA A VARIÁVEL “ROUBO”	80
TABELA 12 – MATRIZ DE CONFUSÃO REFLETINDO OS RESULTADOS DA AD PARA A VARIÁVEL “CRIMES”	84
TABELA 13 – DADOS EXTRAÍDOS DOS BOLETIM DE OCORRÊNCIA.....	86
TABELA 14 – PADRÕES IMPORTANTES ORIUNDOS DAS ANÁLISES USANDO ADS	90

Lista de Abreviaturas e Siglas

AD	Árvores de Decisão
AR	Árvores de Regressão
AM	Aprendizado de Máquina
BO	Boletim de Ocorrência
COHAB	Companhia Metropolitana de Habitação de São Paulo
CP	Código Penal
CPTM	Companhia Paulista de Trens Metropolitanos
CSV	<i>Comma Separated Values (Valores Separados por Vírgulas)</i>
FBSP	Fórum Brasileiro de Segurança Pública
HNN	<i>Hierarchical Nearest Neighbour</i>
KDD	<i>Knowledge Discovery in Databases</i>
KNN	<i>K-Nearest Neighbour</i>
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IPVS	Índice Paulista de Vulnerabilidade Social
MD	Mineração de Dados
NGA	<i>National Geospatial-Intelligence Agency</i>
NetKDE	<i>Network Kernel Density Estimation</i>
ODS	Objetivos de Desenvolvimento Sustentável
ONU	Organização das Nações Unidas
PBDA	Portal Brasileiro de Dados Abertos
PGVs	Polos Geradores de Viagens
PIB	Produto Interno Bruto
QGIS	<i>Quantum Geographic Information System</i>
RNA	Redes Neurais Artificiais
SEADE	Sistema Estadual de Análise de Dados
SIG	Sistema de Informação Geográfica
SIRGAS	Sistema de Referência de Geocêntrico
SSP	Secretaria de Segurança Pública do Estado de São Paulo
SVM	Support Vector Machines
WEKA	Waikato Environment for Knowledge Analysis

SUMÁRIO

1. INTRODUÇÃO	13
1.1 CONTEXTUALIZAÇÃO DO TEMA	13
1.2 PROBLEMA DE PESQUISA	15
1.3 OBJETIVO GERAL	16
1.4 OBJETIVOS ESPECÍFICOS.....	16
1.5 JUSTIFICATIVA E MOTIVAÇÃO	16
1.6 ESTRUTURA DO TRABALHO	17
2. REFERENCIAL TEÓRICO	19
2.1 PADRÕES ESPACIAIS DOS CRIMES DE FURTO E ROUBO E VARIÁVEIS SOCIODEMOGRÁFICAS	19
2.2 PADRÕES ESPACIAIS DOS CRIMES E IA	23
2.3 SUMARIZAÇÃO DA REVISÃO DE LITERATURA	30
2.4 INTELIGÊNCIA ARTIFICIAL (IA).....	32
2.5 MINERAÇÃO DE DADOS	34
2.5.1 <i>Árvores de Decisão – AD’S</i>	36
2.5.2 <i>WEKA – Waikato Environment for Knowledge Analysis</i>	37
3. MÉTODOS E MATERIAIS	38
3.1 CARACTERIZAÇÃO METODOLÓGICA DA PESQUISA	38
3.2 CONTEXTO HISTÓRICO E ÁREA DE ESTUDO: SUBPREFEITURA DE ITAQUERA, SÃO PAULO/SP- BRASIL	39
3.3 DADOS E VARIÁVEIS	42
3.3.1 <i>Dados cartográficos</i>	42
3.3.2 <i>Variáveis de reposta</i>	42
3.3.3 <i>Variáveis explicativas</i>	43
3.4 ABORDAGEM PROPOSTA PARA ANÁLISE DE DADOS E VARIÁVEIS USANDO IA	45
3.4.1 <i>Seleção de dados</i>	46
3.4.2 <i>Pré-processamento</i>	47
3.4.3 <i>Espacialização de dados em SIG</i>	47
3.4.4 <i>Transformação</i>	47
3.4.5 <i>Mineração de dados</i>	48
3.4.6 <i>Interpretação/avaliação</i>	48
3.5 SÍNTESE METODOLÓGICA	50
4. RESULTADOS.....	51
4.1 ESPACIALIZAÇÃO DE DADOS	51
4.1.1 <i>Espacialização dos dados sociodemográficos</i>	51
4.1.2 <i>Espacialização das ocorrências de furto e roubo</i>	57
4.2 ANÁLISE DA DISTRIBUIÇÃO DOS DADOS SOCIODEMOGRÁFICOS E DOS CRIMES DE FURTO E ROUBO	63
4.3 DESCOBERTA DE PADRÕES USANDO MD (ÁRVORES DE DECISÃO).....	74
4.3.1 <i>Análise das relações entre as variáveis explicativas e de resposta</i>	74
4.3.2 <i>Análise de padrões dos crimes com base no perfil das vítimas</i>	85
4.4 DISCUSSÃO DOS RESULTADOS	88
5. CONSIDERAÇÕES FINAIS.....	96
REFERÊNCIAS BIBLIOGRÁFICAS.....	100

1. INTRODUÇÃO

1.1 Contextualização do tema

Crimes são ações ou omissões que violam as leis estabelecidas por uma sociedade. Eles podem variar em gravidade e abranger desde pequenas infrações de trânsito até crimes graves como assassinato e terrorismo. Os crimes podem ser cometidos por indivíduos ou grupos, e seus impactos podem afetar não apenas a vítima direta, mas também a sociedade como um todo. Alguns crimes são mais comuns em determinadas regiões ou grupos sociais, enquanto outros são mais universais (DORNELLES, 2017). O estudo dos crimes e sua prevenção é uma área importante no campo da criminologia (SUMARIVA, 2023). Neste contexto, referências como as obras de Cesare Lombroso, Émile Durkheim e Michel Foucault, além de estudos mais recentes como o "Relatório Mundial sobre a Prevenção da Violência" da Organização Mundial da Saúde, são importantes para compreendermos as complexidades dos crimes e como preveni-los (LOMBROSO, 2020; DURKHEIM, 2017; FOUCAULT, 2012; OMS, 2015).

Há diversos tipos de crimes que podem ser praticados contra a pessoa, o patrimônio, a família, a dignidade sexual e a administração pública, entre outros. Os crimes de furto e roubo, considerados nesta pesquisa, são comuns em todo o mundo, e geram um impacto significativo na vida das vítimas e nas comunidades afetadas (GRECO, 2017). No Brasil, esses crimes são considerados como "crimes contra o patrimônio" tipificados no Código Penal (CP), e constituem uma preocupação constante das autoridades de segurança pública (AZEVEDO et al., 2016).

O crime de furto, de acordo com o Art. 155 do Código Penal (CP), é caracterizado pela subtração, para si ou para outrem, de coisa alheia móvel. Já o crime de roubo, conforme definido pelo Art. 157 do CP, é caracterizado pela subtração de coisa móvel alheia para si ou outrem, empregando grave ameaça ou violência à pessoa.

Em outras palavras, ambos os crimes tratam de subtração de coisa alheia, mas o crime de roubo tem como característica o emprego de violência ou grave ameaça à vítima. Adicionalmente, ambos causam prejuízos materiais e emocionais às vítimas e geram insegurança na comunidade.

De acordo com dados do Fórum Brasileiro de Segurança Pública, no ano de 2021 que foi marcado pela retomada de atividades pós-pandemia, houve também crescimento nos roubos a estabelecimentos, residências, roubo a instituições financeiras e roubo de carga. Já no ano de 2020, foram registrados mais de 1,2 milhão de casos de roubo no país, com uma média de 3,4 mil casos por dia. Com relação ao furto, registraram-se mais de 1,1 milhão de casos no mesmo período (FBSP, 2022).

No estado de São Paulo, o roubo e o furto são problemas crônicos que afetam a segurança pública e o bem-estar da população. Localizada nesse estado, a cidade de São Paulo é a maior e mais próspera do Brasil, com 12,4 milhões de habitantes, domiciliados em 96 distritos. Ela ocupa o primeiro lugar entre os municípios mais populosos e possui o maior Produto Interno Bruto (PIB) do Brasil, com quase R\$ 764 bilhões (IBGE,2021). Conforme Soares et al. (2020), a incidência de crimes patrimoniais na região metropolitana de São Paulo é significativa, tornando necessárias ações integradas e estratégicas de prevenção e combate ao crime.

Com a falta de planejamento e a urbanização muito rápida ocorrida nas últimas décadas, é possível verificar as desigualdades socioespaciais na cidade. Notam-se centros urbanos modernos e ricos, refletindo a cidade globalizada, mas também se percebe áreas periféricas com infraestrutura precária (SCARLATO E ALVES, 2018). De acordo com Da Silva Furtado et al. (2020), esse cenário de desigualdade e exclusão social tende a gerar mais violência urbana. No estudo, os autores citam a cidade de São Paulo como exemplo.

No período de 20 anos compreendido entre 2001 e 2021, a cidade de São Paulo registrou mais de 2,6 milhões de roubos e mais de 3,6 milhões de furtos, dos quais cerca de 1,7 milhão dos registros estavam relacionados a furto de veículo. Com base nos dados da Secretaria de Segurança Pública do Estado de São Paulo (SSP) é possível verificar que, apesar da redução dos casos de furto e roubo no período pandêmico de COVID-19, houve um aumento do número de ocorrências desses crimes nos últimos 10 anos. Esses números indicam a gravidade do problema e a necessidade de políticas públicas efetivas para combatê-los.

Nesse contexto, destaca-se a importância de estudos que visem à compreensão das dinâmicas desses crimes, bem como a ampliação da compreensão dos elementos

e atributos do espaço urbano que possam influenciar em suas ocorrências, visando subsidiar políticas públicas para prevenção e o combate da violência urbana.

1.2 Problema de pesquisa

Na cidade de São Paulo, o rápido e desordenado crescimento urbano também gerou um incremento das desigualdades socioespaciais. A criminologia crítica e a teoria humanista do direito entendem que as práticas de infração de crimes estão associadas às condições de reprodução da desigualdade, tais como as condições de renda e moradia, a segregação e a inacessibilidade à cidade, acesso às políticas públicas, à educação, à escolaridade e ao emprego, que, juntas, frustram as expectativas e potencializam as condições de criminalidade (DE PAULA FALEIROS, 2022). Assim, o entendimento das relações entre as características socioespaciais e a ocorrência de crimes, principalmente nas grandes metrópoles, constitui um problema de pesquisa que vem sendo cada vez mais explorado na literatura (DE ANDRADE CARNEIRO, 2022).

A cidade de São Paulo é subdividida em cinco regiões (Norte, Sul, Leste, Oeste e Centro) e 32 subprefeituras regionais. Essa divisão é realizada para garantir que toda a cidade receba o devido cuidado em todos os aspectos socioeconômicos. Atualmente, a região Norte conta com sete subprefeituras, a Sul com nove, a Leste com doze, a Oeste com três e a Centro com apenas uma subprefeitura. No que tange ao número de ocorrências de furto e roubo, a região Leste é a que mais se destaca.

Na zona Leste de São Paulo, a subprefeitura de Itaquera se destaca por registrar o maior número de ocorrências de crimes contra o patrimônio. Portanto, o que ocorre em relação aos crimes está diretamente ligado aos ambientes locais e aos contextos nos quais essas ocorrências acontecem, incluindo o bairro e a cidade.

A presente pesquisa tem como foco o recorte espacial referente à subprefeitura de Itaquera, composta por quatro distritos: Cidade Líder, Itaquera, José Bonifácio e Parque do Carmo, que abrigam, juntos, mais de 523 mil habitantes e ocupam uma área de 55,10 km². Considerando que as desigualdades socioespaciais na cidade tendem a gerar mais violência urbana, conforme destacado por Da Silva Furtado et al. (2020) e outros trabalhos na literatura; que a subprefeitura de Itaquera se destaca negativamente em relação ao número de crimes na cidade de São Paulo; que a

compreensão da influência dos elementos e atributos do espaço urbano na ocorrência de crimes pode subsidiar políticas públicas de prevenção e combate à violência; e que a Inteligência Artificial pode ser uma alternativa para identificar padrões e relações espaciais ao cruzar dados socioeconômicos e ocorrências de crimes, surgem as seguintes questões de pesquisa:

1. Que indicadores sociodemográficos estão relacionados à distribuição espacial dos crimes de furto e roubo na subprefeitura de Itaquera?
2. Como a Inteligência Artificial (IA) pode ser empregada para analisar tais relações e revelar padrões espaciais?

1.3 Objetivo geral

Analisar as relações espaciais dos crimes de furto e roubo com indicadores sociodemográficos na subprefeitura de Itaquera, Zona Leste de São Paulo/SP – Brasil.

1.4 Objetivos específicos

1.4.1 Mapear a distribuição espacial dos crimes de furto e roubo e de dados sociodemográficos na subprefeitura de Itaquera;

1.4.2 Identificar padrões de distribuição espacial das ocorrências de furto e roubo na subprefeitura de Itaquera;

1.4.3 Analisar as relações entre as variáveis sociodemográficas e os padrões espaciais de distribuição dos casos de furto e roubo na subprefeitura de Itaquera;

1.4.4 Identificar padrões das ocorrências de crimes na subprefeitura de Itaquera com base nas características pessoais das vítimas;

1.5 Justificativa e motivação

Trabalhos científicos dedicados ao estudo das dinâmicas dos padrões de crimes são fundamentais para um país, pois a redução da criminalidade tem um impacto positivo no bem-estar da população e no desempenho econômico. Alinhado a esse contexto, estudos que investigam crimes contra o patrimônio têm relevância para o planejamento urbano, pois esses crimes podem estar associados a variáveis

socioeconômicas, sendo a renda per capita e o nível de urbanização fatores frequentemente relacionados (MARTINS, TEIXEIRA E SILVA, 2021).

Teorias e estudos prévios demonstram que a redução das ocorrências de crimes melhora a sustentabilidade das áreas e bairros, gerando resultados sociais e de saúde mais favoráveis para a população (MATIJOSAITIENE e MCDOWALD, 2019).

Embora haja na literatura estudos que buscam correlações entre características socioespaciais e o número de ocorrências de crimes de furto e roubo, são escassos os trabalhos que exploram o uso de ferramentas de IA, especialmente aquelas voltadas para mineração de dados (MD), em suas análises. Cita-se como exemplo o trabalho de Rodrigues (2023) que analisou as relações socioespaciais dos crimes de furto e roubo na subprefeitura de Itaquera, com base em análise visual e estatística descritiva. Destaca-se que, apesar da relevância de tais técnicas, elas apresentam limitações quando aplicadas à análise de padrões de eventos que envolvem muitas variáveis. Nesse sentido, o emprego de ferramentas de IA propicia a descoberta de padrões nos dados, transcendendo as análises multivariadas produzidas pela estatística descritiva, comumente utilizada nos trabalhos da literatura.

Assim, visando preencher a lacuna existente, esta pesquisa se justifica pela necessidade de compreender os padrões de distribuição espacial dos crimes de furto e roubo na cidade de São Paulo, especialmente na subprefeitura de Itaquera, bem como verificar a correlação desses crimes com características socioespaciais (variáveis explicativas) definidas em Rodrigues (2023). Tal compreensão é fundamental à orientação de políticas públicas de segurança urbana consonantes com os Objetivos do Desenvolvimento Sustentável da ONU e com a melhoria da qualidade de vida humana.

1.6 Estrutura do trabalho

A estrutura deste trabalho é organizada da seguinte maneira: o Capítulo 1 abrange a introdução, onde são expostos o problema de pesquisa, as questões de pesquisa, os objetivos e a justificativa. O capítulo 2 introduz o embasamento teórico, construído com base nas pesquisas relevantes ao tópico "análise dos padrões espaciais dos crimes de furto e roubo", e explora as ferramentas de Inteligência Artificial utilizadas. No capítulo 3, é apresentado o método de pesquisa, que inclui a caracterização e o

delineamento da pesquisa, os procedimentos metodológicos adotados, a consolidação metodológica e a abordagem proposta para a análise de dados e variáveis mediante o uso de IA. O capítulo 4 revela os resultados obtidos e, por fim, o Capítulo 5 apresenta as considerações finais da pesquisa, seguido pelo Capítulo 6, no qual são apresentadas as referências bibliográficas que fundamentam a presente dissertação.

2. REFERENCIAL TEÓRICO

Nas seções 2.1 e 2.2 são apresentados trabalhos da literatura que visam investigar a existência de relações entre crimes e atributos socioespaciais, empregando estatística descritiva e técnicas oriundas da inteligência artificial (IA). Na seção 2.3 é feita uma sumarização da revisão da literatura apresentada nas seções 2.1 e 2.2. Por fim, nas seções 2.5 e 2.6, são apresentados conceitos de IA e de MD, além das técnicas e software empregados nesta pesquisa.

2.1 Padrões espaciais dos crimes de furto e roubo e variáveis sociodemográficas

Adewuyi et al. (2017) exploraram os padrões espaciais dos crimes de roubo à mão armada e de furto e suas relações com a distribuição geográfica na Nigéria. Os autores destacaram principalmente a incidência de crimes e o nível de violência nas cidades nigerianas, nas quais houve um aumento significativo nas últimas décadas em decorrência da privação econômica, que tem provocado um ambiente hostil com insegurança e que impossibilita investimentos estrangeiros e a expansão dos negócios.

No estudo de Adewuyi et al. (2017), foram consideradas quatro variáveis de resposta (roubo à mão armada, furto, estupro e sequestro) e seis variáveis explicativas (taxas de desemprego, densidade populacional, índices de educação, renda nacional bruta, porcentagem populacional masculina com idade de 18 a 35 anos e estrutura de policiamento). Para análise de dados, os autores empregaram técnicas estatísticas como análise de regressão combinada com um modelo misto de Poisson, além da abordagem Bayesiana e simulação de Monte Carlo, utilizando Cadeias de Markov para estimar os parâmetros do modelo.

De acordo com Adewuyi et al. (2017), foi possível observar que as taxas de desemprego apresentaram correlação positiva com estupro, sequestro e roubo à mão armada e negativa com furto. Os sequestros, furtos e roubos à mão armada tiveram correlações positivas com a população masculina jovem e com a renda nacional bruta. Além disso, os resultados mostraram que o roubo à mão armada, furto e sequestro não apresentaram correlações com taxas de desemprego e níveis de educação. No entanto, todos os crimes tiveram correlações positivas significativas com a renda. Os

autores destacaram ainda que a presença de postos policiais tem a capacidade de reduzir as atividades criminosas de todos os tipos.

Khalid et al. (2018) investigaram as atividades criminosas em determinados locais e suas distribuições no tempo e no espaço. Eles consideraram que o comportamento criminoso depende de fatores sociodemográficos, como pobreza, nível de educação e outros atributos espaço-temporais. Os autores utilizaram a técnica *Network Kernel Density Estimation - NetKDE* para identificar as áreas com maior frequência de crimes. O mapa resultante foi utilizado para visualização, alocação de recursos e previsão de futuros padrões para a ocorrência de crimes. Eles analisaram os *hotspots* dos crimes de furto de bicicleta, furto de carros e roubo na cidade de Faisalabad, na região central da província de Punjab, no Paquistão. A combinação dos métodos de estimativa de densidade de rede por Kernel e os resultados como valores de entrada para calcular estatísticas *Getis-Ord GI* possibilitou encontrar os pontos de acesso restritos à rede e que são estatisticamente significativos.

Os resultados obtidos por Khalid et al. (2018) demonstraram ainda *hotspots* significativos nas redes rodoviárias, concentração dos crimes na área central da cidade e que as ocorrências de crimes podem ser influenciadas pelas características de uso e ocupação do solo, como áreas comerciais, residenciais, industriais, espaços públicos e parques públicos.

Quick, Li e Brunton-Smith (2018) pontuam que vários tipos de crimes possuem padrões espaciais semelhantes e que, juntos, são associados a fatores iguais de risco e com as mesmas teorias ecológicas, como a teoria da desorganização social, empregada para explicar os padrões de crimes na escala de bairros. Ainda segundo os autores, crimes violentos e de propriedade podem estar associados a características estruturais, como privações socioeconômicas, instabilidade residencial e heterogeneidade étnica.

Quick, Li e Brunton-Smith (2018) tiveram como objetivo examinar se e como os padrões espaciais de crime podem ser explicados a partir de um ou mais padrões subjacentes do crime geral. Para tanto, eles consideraram uma pequena área em Londres e utilizaram um conjunto de modelos espaciais multivariados bayesianos para análise de diferentes tipos de crimes, como furto e roubo. A variabilidade residual de cada tipo de crime foi dividida em componentes compartilhados e específicos do tipo

de crime, após controlar os efeitos da densidade populacional, privação, instabilidade residencial e heterogeneidade étnica. Os componentes que foram compartilhados respondem pelas correlações entre os tipos de crimes e identificam os padrões gerais do crime compartilhados entre outros crimes.

Em seus resultados, roubos e crimes violentos tiveram associações positivas mais fortes com privação, instabilidade e heterogeneidade étnica, enquanto componentes compartilhados explicaram as maiores proporções de variabilidade para todos os tipos de crimes. Já os crimes de arrombamento, furto e crimes envolvendo veículos representaram padrões específicos do tipo que divergem dos padrões gerais do crime.

Quick, Li e Brunton-Smith (2018) concluíram que os padrões gerais dos crimes são importantes para entender o padrão espacial de vários tipos de crime em pequenas áreas. Os modelos espaciais multivariados forneceram uma estrutura para quantificar diretamente as estruturas de correlação entre os crimes e revelaram os padrões gerais desses crimes subjacentes compartilhados entre vários tipos de crimes.

Buil-Gil, Moretti e Langton (2021), assim como Quick, Li e Brunton-Smith (2018), utilizaram pequenas escalas espaciais em seus estudos para analisar se as estatísticas criminais agregadas nessas escalas são afetadas por um viés maior do que as produzidas para áreas maiores. No estudo, uma população sintética foi simulada, consistente com as características de Manchester, empregando-se o Censo do Reino Unido e parâmetros derivados do *Crime Survey* para a Inglaterra e País de Gales. Foram simulados crimes sofridos por indivíduos e a probabilidade desses crimes serem registrados pela polícia. Isso permitiu a comparação das diferenças entre todos os crimes e incidentes registrados pela polícia em diferentes escalas. Os resultados mostraram que a porcentagem de crimes desconhecidos pela polícia varia em pequenas áreas, subestimando os crimes em certos lugares e superestimando em outros. De acordo com os autores, a análise criminal em nível micro é afetada por um risco maior de viés do que os crimes agregados em escalas maiores.

Andresen, Curman e Linning (2017) investigaram as concentrações espaciais e a estabilidade das trajetórias dos tipos de crimes ocorridos entre 1991 e 2006 em segmentos de rua e intersecções em Vancouver, Canadá. No estudo, os segmentos de rua e as intersecções, denominados "micro lugares" pelos autores, foram

analisados com o auxílio do *software R*, utilizando o algoritmo *k-means* para estimar os *clusters* longitudinais. Os autores consideraram oito tipos de crimes (crimes totais, furto, roubo, assalto, arrombamento, furto de veículo, furto dentro de veículo e outros). Como resultados, eles apontaram que os crimes estavam concentrados em "micro lugares" específicos, mostrando um aumento na concentração de crimes ao longo do tempo analisado. Em 1991, 50% dos crimes haviam se concentrado em 4,35% dos segmentos de ruas e intersecções, enquanto em 2006 o valor de concentração foi para 3,25%. Esta disposição também foi notada para crimes específicos. Os autores apontaram ainda que a lei de concentração dos crimes em locais foi válida para Vancouver, mas houve diferenças notáveis entre os tipos de crimes, fato que deve ser considerado.

Ceccato e Oberwittler (2008) realizaram uma análise embasada nas teorias da atividade de rotina e da desorganização social dos padrões espaciais e temporais dos crimes de furto e roubo nas cidades de Tallinn, Estônia, e Colônia, Alemanha. Os autores empregaram dados georreferenciados de ocorrências dos crimes, Sistema de Informação Geográfica (SIG), técnicas de estatística espacial, além de um conjunto de dados transversais para examinar os padrões de roubo em Tallinn. Um processo semelhante foi empregado para os dados da cidade de Colônia. A análise estatística foi baseada na técnica de cluster do vizinho mais próximo hierárquico (*Hierarchical Nearest Neighbour - HNN*), a qual permite identificar grupos de incidentes que são espacialmente próximos.

Ceccato e Oberwittler (2008) notaram que, embora as ocorrências dos crimes de roubo fossem mais altas em Tallinn do que em Colônia, suas razões seguiram os mesmos componentes de sobreposição de contextos sociais, como a desorganização social e a atividade de rotina. Os grupos de roubo ficaram concentrados nas áreas centrais das cidades e, especificamente em Tallinn, a concentração estava nas estradas principais, estações e centros locais. Ambas as cidades mostraram uma distribuição do roubo mais associada às variáveis de atividade de rotina do que às variáveis de desorganização social, que mesmo assim mostraram uma contribuição significativa. Verificou-se também nos resultados que não houve grandes diferenças entre as variáveis significativas na distribuição diurna e noturna dos crimes.

Baumer e Arnio (2012) avaliaram as influências da heterogeneidade espacial dos indicadores demográficos tradicionais nas taxas de roubo, arrombamento e homicídio em Chicago, EUA. Ao abordar a variabilidade espacial dos crimes e considerar dados censitários de 821 setores na cidade de Chicago, os autores mostraram os principais atributos socioespaciais que respondem pelas ocorrências de crimes, como a prevalência de estabelecimentos comerciais adultos, a presença de mercado de drogas ilícitas, estabelecimentos que vendem bebidas alcoólicas, desordem física e social, policiamento, laços sociais, eficácia coletiva, força institucional, gentrificação e densidade de comércio. No estudo, foram empregadas técnicas de estatística descritiva, como correlação e modelo de regressão geográfica, que permitiram estimar os efeitos dos indicadores demográficos sobre a criminalidade local. Como indicadores demográficos, foram adotados a estabilidade residencial, a concentração de imigrantes, a desvantagem socioeconômica, a composição racial, a densidade da população, a porcentagem de divórcios, a faixa etária (15 a 29 anos) e as habitações desocupadas.

O estudo de Baumer e Arnio (2012) constatou que modelos globais explicativos dos padrões espaciais dos crimes abordados pela literatura não são adequados para analisar totalmente a complexidade existente, visto que a variabilidade local sugeriu estimativas positivas e negativas para as variáveis.

2.2 Padrões espaciais dos crimes e IA

Hassani et al. (2016) realizaram uma revisão das aplicações de mineração de dados em crimes. Eles analisaram mais de 100 aplicações de mineração de dados em crimes e observaram fortes evidências, com base no número de aplicações, de que as técnicas de classificação são a forma mais popular de mineração de dados nessa área. Assim, destacaram a mineração de dados como uma importante ferramenta para descobrir informações ocultas do Big Data que podem ser utilizadas por instituições privadas e governamentais ao redor do mundo para coibir e prevenir crimes. Segundo os autores, *Support Vector Machines* (SVM), Redes Neurais Artificiais (RNA) e Regras de Associação são raramente utilizadas para minerar dados oficiais de maneira geral, mas na mineração de dados criminais, esses métodos são bem populares e explorados. No trabalho, foram apresentadas aplicações importantes da mineração de dados no domínio de crimes, tais como agrupamento, geração de

regras de associação, Árvores de Decisão (AD), SVM, *Naive Bayes*, RNA e análise de redes sociais, entre outras.

Thongsatapornwatana (2016) constatou que, nos últimos anos, a mineração de dados tem sido uma técnica de IA utilizada para analisar dados criminais previamente armazenados em várias fontes, a fim de encontrar tendências e padrões nos crimes. Além disso, essa técnica pode ser aplicada para aumentar a eficiência na resolução dos crimes de forma mais rápida e notificar automaticamente sobre eles. Para tanto, o autor destaca que, para aumentar a eficácia da detecção de crimes, é preciso selecionar adequadamente as técnicas de mineração de dados, e que os desafios incluem a modelagem de crimes para encontrar algoritmos adequados, bem como o tempo de processamento para preparação, transformação e análise de dados. O trabalho forneceu informações sobre formas que podem ser utilizados na mineração de dados criminais para encontrar padrões e tendências.

David e Suruliandi (2017) realizaram um levantamento sobre as técnicas de aprendizagem de máquina que têm sido aplicadas na identificação criminal. Eles apresentaram uma revisão sobre a análise e previsão dos crimes utilizando técnicas de mineração de dados. Os autores apresentaram um gráfico que representa o grupo ou organização criminosa como uma rede, a partir de dados de comunicação ou com base nos registros criminais. Eles compreendem que a mineração de dados é um procedimento que inclui a avaliação e o exame de grandes bases de dados pré-existentes para a geração de novas informações essenciais. Ainda segundo os autores, as delegacias de polícia e outras agências de justiça criminal semelhantes possuem muitos bancos de dados e outras informações que podem ser usadas para prever ou analisar os crimes. Por fim, eles concluíram que as técnicas de mineração de dados podem ser aplicadas para identificar os criminosos e fornecer previsões sobre a criminalidade.

Dutta, Gupta e Narayan (2017) propuseram uma nova camada de defesa contra crimes de falsidade ideológica em pedidos de cartão de crédito, utilizando mineração de dados. Os autores aplicaram dois algoritmos - Detecção Comunal e Detecção de *Spike* - para detectar fraudes. O estudo se baseou em três conceitos: resiliência, fornecendo múltiplas camadas de segurança; dados de qualidade e adaptabilidade, para distinguir fraudes de comportamentos permitidos por lei. No entanto, na prática,

a busca de muitos atributos e tipos de links em grandes janelas de tempo pelos algoritmos pode consumir um tempo significativo para o sistema gerar resultados. Apesar da atualização regular dos algoritmos, a avaliação não é completamente precisa, pois os invasores não têm tempo para mudar sua tática em resposta aos algoritmos, como seria possível se fossem implantados em tempo real. Além disso, a limitação do estudo é que o conceito de adaptabilidade não é adequadamente demonstrado.

O estudo de Oatley (2022) teve como objetivo analisar o impacto das novas tecnologias relacionadas à Inteligência Artificial (IA) na mineração de dados e no *big data* sobre as questões importantes da análise criminal. O autor observou que a análise de crimes está fragmentada em absorção da governança de dados e governança de *big data*, em que se utilizam dados forenses, *modus operandi*, geoespaciais e dados infracionais. O estudo conclui que, à medida que os dados se tornam mais fáceis de coletar e os algoritmos de computador se tornam mais sofisticados, métodos preditivos mais avançados e novas tecnologias serão desenvolvidos. Os pontos fortes do estudo são a análise cuidadosa do impacto das novas tecnologias na análise criminal e o reconhecimento da necessidade de métodos preditivos mais avançados e novas tecnologias. No entanto, um ponto fraco é a falta de detalhes sobre como a inovação e as medidas policiais serão impulsionadas.

Ates, Bostanci e Serdar (2020) propuseram em seu estudo uma revisão do uso da mineração de dados e aprendizado de máquinas em dados criminais, com o objetivo de fornecer uma nova perspectiva sobre os processos de tomada de decisão e prevenção do crime. O estudo apresentou exemplos práticos de mineração de dados e aprendizado de máquinas nas áreas de segurança e crime, e propôs uma estrutura conceitual para o uso dessas tecnologias, com seus processos e métodos que podem ser aplicados no sentido preventivo. Os autores destacam que a mineração de dados e aprendizado de máquinas podem melhorar a eficiência da aplicação da lei, prevenindo a criminalidade, auxiliando na investigação e resolução de crimes, bem como na identificação de pontos fracos na segurança pública. No entanto, uma das limitações apontadas pelos autores é a necessidade de um conjunto de dados de qualidade e a preocupação com a privacidade dos indivíduos envolvidos nos dados criminais.

Os autores Xu e Feng (2022) utilizaram um método qualitativo para estudar a evolução do comportamento dos crimes de ameaça à segurança pública ao longo do tempo, utilizando algoritmos de mineração de dados. Um algoritmo aprimorado foi desenvolvido com base nas necessidades específicas da mineração de dados jurídicos e do sistema legal. A aplicação desse algoritmo permitiu a construção de um modelo de análise das questões legais dos crimes contra a segurança pública, com atribuição de IP dinâmico a cada processo. Os resultados obtidos indicaram que o modelo proposto foi efetivo na mineração dos dados jurídicos relacionados aos crimes contra a segurança pública, permitindo uma análise mais aprofundada das questões legais envolvidas nesses casos.

O objetivo do artigo de Feng et al. (2019) é aplicar a análise de *big data* a dados criminais para identificar padrões e prever tendências. Para isso, os autores utilizaram técnicas de mineração de dados e aprendizado de máquinas, adotando o *model Prophet* e o *Keras stateful LSTM*, em dados de criminalidade de três cidades dos EUA, com uma janela de dados de três anos. Os resultados do estudo foram interessantes, permitindo a identificação de padrões e tendências em crimes, o que pode ser útil para a prevenção e rastreamento de atividades criminosas, bem como para otimizar o processo de tomada de decisões. Os pontos fortes do artigo incluem a aplicação de técnicas de análise de big data a um problema real e a utilização de modelos de aprendizado de máquinas avançados para analisar os dados. Além disso, os resultados do estudo apresentaram previsões precisas e úteis, o que pode ajudar na tomada de decisões por parte das autoridades policiais e governamentais. No entanto, alguns pontos fracos podem ser apontados, como a falta de informações sobre a coleta e a qualidade dos dados utilizados no estudo. Além disso, o estudo foi realizado apenas em três cidades dos EUA, o que pode limitar a aplicação dos resultados em outras áreas geográficas.

O estudo de Ratul e Rab (2020) teve como objetivo identificar as tendências de ocorrências criminais em Denver, nos EUA, no período de janeiro a maio de 2019, utilizando técnicas de análise estatística e visualização de dados, bem como algoritmos de classificação. Os autores aplicaram diversos modelos de classificação, como *Random Forest*, *Decision Tree*, *AdaBoost Classifier*, *Extra Tree Classifier*, *Linear Discriminant Analysis*, *K-Nearest Neighbors Classifiers*, para classificar quinze

diferentes tipos de crimes. Os resultados foram avaliados utilizando *train-test split* e *k-fold cross-validation* e apresentaram os melhores resultados em todas as bases de avaliação. Os pontos fortes do estudo incluem a aplicação de várias técnicas de análise e classificação em dados criminais de Denver, bem como a utilização de métodos de avaliação rigorosos para verificar a eficácia dos modelos. Os resultados apresentados pelos autores são promissores e podem ser úteis para melhorar a prevenção e o combate ao crime, fornecendo informações valiosas para as autoridades responsáveis pela segurança pública.

O estudo de Dakalbab et al. (2022) é uma revisão da literatura que avaliou as estratégias de inteligência artificial (IA) na previsão de crimes. Os autores analisaram 120 trabalhos publicados entre 2008 e 2021, abordando vários aspectos dos modelos utilizados na análise de crimes, como tipo de análise, crimes estudados, técnicas de previsão, métricas e avaliações de desempenho, pontos fortes e fracos do método proposto, bem como limitações e direções futuras. O estudo identificou 34 categorias de crimes pesquisados e 23 metodologias distintas de análise de crimes. Foram encontradas 64 técnicas diferentes de aprendizado de máquina para a previsão de crimes, sendo que a abordagem de aprendizado supervisionado foi a mais aplicada para a prevenção de crimes. As métricas de avaliação e desempenho, bem como as ferramentas utilizadas na construção dos modelos, foram discutidas, destacando seus pontos fortes e fracos. Os resultados mostraram que há várias abordagens eficazes e promissoras que podem ser aplicadas no campo da segurança pública.

Saraiva et. al (2022) investigaram padrões espaciais e hotspots nos dados criminais da polícia de Porto, em Portugal, entre 2016 e 2018, utilizando técnicas geoespaciais combinadas com técnicas de mineração de dados e aprendizado de máquinas. Primeiramente, os autores utilizaram métodos espaciais que permitiram a identificação de padrões espaciais e *hotspots* relevantes. Logo em seguida, aplicaram processos de mineração de dados e aprendizagem de máquinas, como *Random Forest* e *Árvore de Decisão*. Por fim, coletaram *twets* relacionados à insegurança. Como resultado, verificou-se que o uso de tecnologias geoespaciais, combinadas com técnicas de mineração de dados e aprendizado de máquina, possibilitou avanços significativos na criminologia do lugar, onde foi possível a interpretação de padrões e previsões.

Matijosaitiene, McDonald e Juneja (2019) realizaram um estudo para investigar a identificação de padrões espaciais e temporais de roubos em Manhattan, Nova York, utilizando um modelo de previsão baseado em aprendizado de máquina. O objetivo do estudo foi revelar os fatores urbanos que contribuem para o roubo de veículos na cidade. Os autores utilizaram os segmentos de ruas como recorte espacial da análise e empregaram cinco conjuntos de dados, incluindo informações sobre furtos, restaurantes, entradas de metrô, grafite e classificação do pavimento. O conjunto de dados dos segmentos de ruas foi definido com base na teoria da Sintaxe Espacial, que leva em consideração os trechos de rua entre os cruzamentos. Foram incluídas diversas técnicas de análise, como análise de séries temporais e análise de *hotspots*, além de métodos de aprendizado de máquina, como regressão linear, *elastic-net*, SVM, Árvores de Decisão (AD), árvores de regressão, árvores de classificação, *Random Forest* e aumento de gradiente estocástico. Os resultados indicaram que um maior número de acessos ao metrô, locais com pichações e restaurantes nas ruas contribuem para um maior número de roubos de veículos. Além disso, os modelos lineares apresentaram menor desempenho na análise dos dados. Os autores concluíram que, embora o modelo de previsão para furtos tenha atendido a quase todas as hipóteses, sua precisão foi de 77%, o que sugere a possibilidade da existência de outros fatores não descobertos que contribuem para as ocorrências de roubos.

O estudo de Wheeler e Steenbeek (2021) mostra como os modelos de aprendizagem de máquina, em particular o modelo *Random Forest*, podem ser aplicados no mapeamento de riscos de crimes em microlocais de uma cidade. A partir da previsão espacial gerada pelo modelo, é possível obter previsões mais precisas em comparação a outras técnicas, como a densidade de kernel e a modelagem de área de risco. Os autores demonstraram que a utilização de resumos de modelos interpretáveis facilitou a compreensão do funcionamento do modelo, permitindo a identificação das variáveis geradoras de crimes e dos fatores demográficos que contribuem para a ocorrência de crimes em uma área específica. Além disso, o estudo mostrou que o modelo *Random Forest* superou outros modelos considerados no sentido de prever os crimes futuros usando diferentes medidas de precisão preditiva. No entanto, o estudo também ressalta que o uso de modelos de aprendizado de máquinas de caixa preta pode ser de difícil interpretação, mesmo que eles possam

identificar onde provavelmente estarão os focos de crimes no futuro. É necessário, portanto, interpretar os resultados de maneira que seja possível a compreensão espacial das áreas consideradas de alto risco.

Clancy et al. (2022) combinaram métodos estatísticos tradicionais com aprendizado de máquina para entenderem melhor os modelos contextuais localmente relevantes. Eles analisaram crimes em duas cidades americanas urbanas, Milwaukee (Wisconsin) e Chicago (Illinois). Na cidade de Milwaukee, o crime violento foi associado à desvantagem concentrada, estabilidade residencial, heterogeneidade étnica, população total e atraso espacial do crime violento. Já em Chicago, os crimes violentos estão associados à imigração, moradia ocupada pelo proprietário, taxa de desemprego, população com diploma universitário, heterogeneidade étnica, população total e defasagem espacial para crimes violentos. Os modelos de aprendizado de máquina sugeriram que o crime violento em Chicago teve como variáveis mais importantes população total, imigração, educação universitária, heterogeneidade étnica e proporção em ocupações profissionais. A variável mais importante para crimes contra a propriedade em Milwaukee foi população total e heterogeneidade étnica. Em Milwaukee, a heterogeneidade étnica é igualmente importante para crimes violentos e contra a propriedade, enquanto em Chicago, a heterogeneidade étnica não é uma variável muito importante para crimes violentos e não é um preditor significativo de crimes contra a propriedade.

Como pode ser observado nesta seção, os estudos da literatura que buscam investigar a existência de relações entre crimes e atributos socioespaciais utilizam tanto técnicas estatísticas quanto outras técnicas oriundas da inteligência artificial (IA). A mineração de dados (MD) é aplicada em boa parte dos trabalhos apresentados, sendo que as técnicas mais empregadas são Árvores de Decisão (AD), Regras de Associação e *Random Forest*.

Dos estudos apresentados nesta seção, os três últimos são os que mais se aproximam da pesquisa que se pretende conduzir no presente trabalho. Wheeler e Steenbeek (2021) utilizaram aprendizado de máquina no mapeamento de risco para crimes, embora seu foco não tenha sido especificamente em crimes de furto e roubo. Já o estudo de Matijosaitiene, McDowald e Juneja (2019) teve como objetivo identificar padrões espaciais e temporais de roubo em Manhattan - Nova York, por meio de um

modelo de aprendizado de máquina, para revelar fatores urbanos que contribuem para os roubos de veículos. Um detalhe interessante nesse estudo, complementado pelos autores, foi que, embora o modelo de previsão para furtos tenha atendido quase todas as hipóteses, sua precisão foi de 77%, o que pode sugerir que existam outros fatores não descobertos que contribuíram para as ocorrências de roubos. Clancy et al. (2022) combinaram métodos tradicionais com métodos de aprendizado de máquina para entender melhor os modelos contextuais localmente relevantes na análise de crimes em duas cidades americanas urbanas: Milwaukee e Chicago. Ao construir modelos locais para cada uma das duas cidades, pôde-se identificar que a regressão binomial negativa e modelos de floresta aleatória produzem resultados consistentes que não são influenciados pelo próprio método. Resta saber se esse padrão é específico do espaço geográfico considerado ou se pode ser verificado em outras cidades, como São Paulo.

As lacunas identificadas nos estudos de Wheeler e Steenbeek (2021), Clancy et al. (2022) e Matijosaitiene, McDowald e Juneja (2019) serão consideradas na pesquisa desenvolvida no presente trabalho.

2.3 Sumarização da revisão de literatura

Com base na revisão da literatura apresentada neste capítulo, constata-se que grande parte dos estudos dedicados ao entendimento dos padrões espaciais de crimes baseia-se na teoria da desorganização social.

Tal Teoria é definida a partir de uma perspectiva criminológica que investiga de que maneira a carência de coesão social, a fragilidade das instituições e a desordem presentes em uma comunidade podem contribuir para o aumento da atividade criminal. Ela ressalta que a falta de vínculos interpessoais sólidos e o funcionamento precário das estruturas sociais estabelecem um cenário propício para comportamentos delituosos. A abordagem da Teoria da Desorganização Social destaca a importância de intervenções que fortaleçam as interações comunitárias e aprimorem as condições locais como estratégias para mitigar a criminalidade e fomentar a coesão social (DE ANDRADE CARNEIRO, 2022).

Nos estudos que se apoiam na teoria da desorganização social, as variáveis mais frequentemente utilizadas para explicar os padrões espaciais dos crimes são:

1. Renda Média (Baumer e Arnio, 2012; Ceccato e Moreira, 2021; Adewuyi et al., 2017), 2. População Jovem (Baumer e Arnio, 2012; Ceccato e Moreira, 2021; Xu et al., 2021) 3. Densidade Populacional (Adewuyi et al., 2017; Baumer e Arnio, 2012;), 4. Taxa de Desemprego (Adewuyi et al., 2017; Khalid et al., 2018), 5. Privação Social (Quick Li e Brunton-Smith, 2018; Baumer e Arnio, 2012; Ceccato e Oberwitter, 2008), 6. População Masculina (Adewuyi et al., 2017; Ceccato e Moreira, 2021) e 7. escolaridade (Adewuyi et al., 2017). Ressalta-se que a ordem das variáveis no texto está relacionada com o número de trabalhos em que elas foram citadas.

Com relação à aplicação de IA e MD, as técnicas mais empregadas nos trabalhos encontrados na revisão da literatura foram: Árvores de Decisão (AD), Regras de Associação e Random Forest, conforme pode ser observado na tabela 1.

Tabela 1 – Técnicas de IA utilizadas nos trabalhos da literatura

Técnicas utilizadas	Autores
Árvores de Decisão (AD)	Hassani et al. (2016); David e Suruliandi (2017); Matijosaitiene, McDonald e Juneja (2019) Ratul e Rab (2020);
Geração de Regras de Associação	Hassani et al. (2016); Matijosaitiene, McDonald e Juneja (2019) ; Ratul e Rab (2020);
Random Forest	Matijosaitiene, McDonald e Juneja (2019); Ratul e Rab (2020); Wheeler e Steenbeek (2021)
Support Vector Machines (SVM)	Hassani et al. (2016); Matijosaitiene, McDonald e Juneja (2019)
Outras técnicas	Hassani et al. (2016); Dutta, Gupta e Narayan (2017); Feng et al. (2019); Matijosaitiene, McDonald e Juneja (2019); Ratul e Rab (2020).

Fonte: Aatoria do autor

Nas seções a seguir, são apresentados os conceitos e técnicas de IA e MD utilizados na presente pesquisa.

2.4 Inteligência Artificial (IA)

A Inteligência Artificial (IA) pode ser definida como um ramo da Ciência da Computação que busca construir mecanismos, métodos, abordagens, modelos e dispositivos que simulem a capacidade humana de resolver problemas, por meio de algoritmos computacionais. Quanto ao aprendizado, as técnicas de IA ou Aprendizado de Máquina (AM) podem ser classificadas em três categorias: supervisionadas, não-supervisionadas e por reforço. As técnicas supervisionadas exigem que, durante o treinamento, haja uma resposta esperada para cada instância de treinamento (dados

de treinamento), enquanto as técnicas não-supervisionadas buscam identificar grupos ou padrões a partir dos dados de treinamento, sem a necessidade de ter a resposta esperada para cada instância. Já as técnicas por reforço consideram incertezas causadas por mudanças no ambiente em que as ações são tomadas e escolhem a melhor ação dependendo das circunstâncias (DA SILVA e MAIRINK, 2019).

De acordo com Sayad (2019), Inteligência Artificial (IA) refere-se ao estudo de algoritmos de computador que imitam comportamentos inteligentes por meio da interpretação de dados externos, utilizando tal aprendizado para alcançar objetivos e realizar tarefas específicas. Um subcampo da IA é o Aprendizado de Máquina (*Machine Learning*), em que os algoritmos são capazes de ajustar automaticamente seus modelos enquanto lidam com os dados.

A área de Ciência de Dados, cujas interações com IA são ilustradas na figura 1, lida com diversos tipos de dados, incluindo aqueles estruturados em tabelas e aqueles não estruturados, como textos, imagens e sons. O processo envolve a limpeza, preparação e análise dos dados, e combina princípios da ciência da computação, matemática e estatística. Em resumo, é um conjunto de técnicas utilizadas para extrair informações e insights valiosos dos dados (SOARES, 2020).

Figura 1 – Interações entre IA, Ciência de Dados e Mineração de Dados



Fonte: SOARES (2020)

Nos últimos anos, a IA e a Ciência de Dados têm sido cada vez mais incorporadas às metodologias de combate ao crime. Departamentos de polícia, juízes e outros profissionais do campo jurídico e de planejamento de cidades podem usar conhecimento preditivo para apoiá-los em decisões em suas respectivas áreas.

Exemplos de como a criminologia espacial explora as ferramentas IA e de Ciência de Dados podem ser vistos nos trabalhos de David e Suruliandi (2017); Dutta, Gupta e Narayan (2017); Feng et al. (2019); Matijosaitiene, McDonald e Juneja (2019); Ratul e Rab (2020); Wheeler e Steenbeek (2021); Xu e Feng (2022) e Clancy et al. (2022).

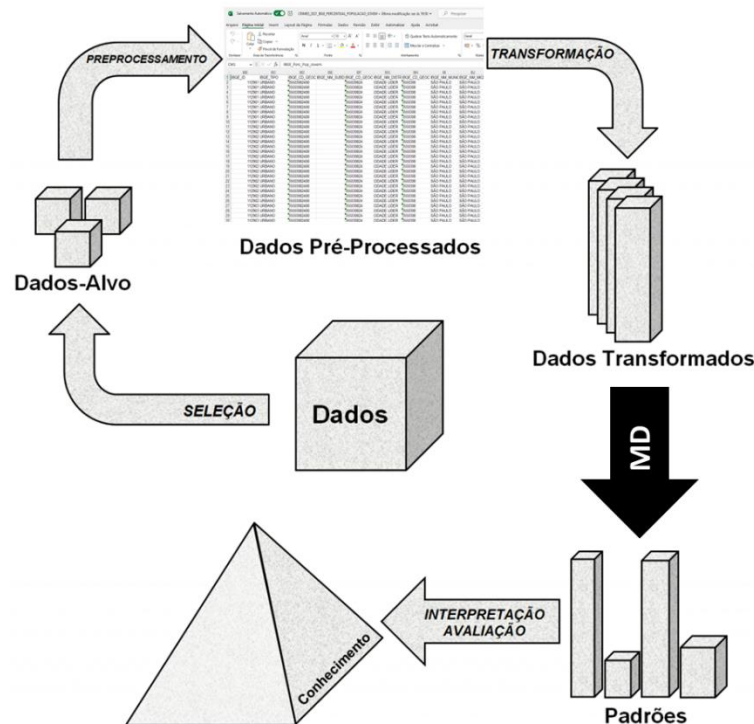
2.5 Mineração de Dados

A Mineração de Dados (MD) pode ser definida como um processo automatizado que tem como objetivo a descoberta de conhecimento valioso em grandes bases de dados. Ela se baseia na utilização de algoritmos de Inteligência Artificial e/ou estatísticos capazes de vasculhar grandes bases de dados de modo eficiente e revelar padrões interessantes escondidos dentro da "montanha de dados" (HAN et al., 2011) (WITTEN et al., 2016) (FAYYAD., 1996).

Um exemplo de aplicação da MD é o caso de uma grande loja de departamentos dos EUA, que descobriu que muitos consumidores que compravam fraldas nas noites de quinta-feira também compravam cerveja. Esse tipo de informação pode ser útil para a loja em sua estratégia de marketing e promoções. Outro exemplo é a área de criminologia espacial visando a descoberta de padrões e correlações espaciais, como demonstrado nos trabalhos de Hassani et al. (2016); Thongsatapornwatana (2016); Feng et al. (2019); Matijosaitiene, McDonald e Juneja (2019); Ratul e Rab (2020); Wheeler e Steenbeek (2021); Xu e Feng (2022); Clancy et al. (2022).

Quando se fala de MD, é importante mencionar que embora *Knowledge Discovery in Databases* – KDD e MD sejam frequentemente entendidos como sinônimos, MD é apenas uma etapa do KDD, que compreende todos os passos necessários para a descoberta do conhecimento a partir do conjunto de dados analisados, conforme ilustrado na figura 2.

Figura 2 – Passos do KDD



Fonte: Adaptado de Fayyad et.al (1996)

A **seleção** está relacionada com a obtenção e segmentação dos dados para a análise a partir de fontes adequadas; no **pré-processamento** é feita a limpeza dos dados; a **transformação** consiste na padronização e categorização dos dados para suprir possíveis limitações das técnicas de MD. Após esses passos é que são conduzidos os passos relativos à **Mineração de Dados** e **interpretação e avaliação** dos resultados obtidos pela aplicação das técnicas de IA.

As principais tarefas da MD são a classificação, a geração de regras de associação, e o agrupamento de dados. Na classificação o objetivo é produzir descrições das características dos dados para cada classe. Na geração de regras de associação busca-se identificar relações entre os dados, representadas por regras do tipo SE...ENTÃO. Já no agrupamento, o objetivo é dividir os dados em grupos com base nas suas similaridades.

Algumas técnicas utilizadas na MD incluem redes neurais artificiais, Árvores de Decisão, Algoritmos Genéticos, Lógica *fuzzy*, *K*-nearest neighbors (KNN), *Support Vector Machine* (SVM) e Redes Bayesianas. Essas técnicas são utilizadas para

processar e analisar os dados, identificando padrões e relações que possam ser úteis para a tomada de decisões.

2.5.1 Árvores de Decisão – AD'S

A árvore de decisão é um algoritmo de aprendizado de máquina supervisionado que pode ser utilizado tanto para classificação quanto para regressão, ou seja, pode ser usado para prever categorias discretas e para prever valores numéricos. As ADs são uma representação simples do conhecimento e têm sido amplamente utilizadas na construção sistemas de inteligentes. Elas permitem construir classificadores eficientes que predizem classes com base nos valores dos atributos, podendo ser aplicadas em diversas áreas, como a criminologia espacial (ZHOU, 2021).

O algoritmo AD é um dos mais simples em aprendizado de máquina e pode ser interpretado como um conjunto de regras do tipo SE...ENTÃO. As ADs são um dos modelos mais práticos e utilizados em inferência indutiva. A chave para o sucesso de um algoritmo AD é como gerar a árvore, ou seja, como escolher os atributos mais importantes para gerar as regras e quais regras podem ser descartadas da árvore. O ideal é gerar a AD com base na importância dos atributos, desta forma, o atributo mais importante deve ficar na raiz da árvore para que o problema possa ser resolvido com o menor número de regras.

Uma árvore de decisão é composta por um conjunto de nós, onde cada nó representa uma decisão a ser tomada. A cada nó, o valor da instância é avaliado e o processo segue por um dos nós internos. Uma árvore de decisão possui um nó raiz, nós internos e nós folha, onde a classificação é definida.

Para construir uma AD, é necessário definir quais atributos compõem um nó e qual a condição de particionamento, que se dá por meio de um teste executado a partir de uma expressão lógica. É importante definir também uma condição de parada, onde não haverá mais particionamentos e o modelo levará à uma classe. Um algoritmo de AD busca sempre construir uma árvore menor possível (KINGSFORD, 2008).

Entre os algoritmos de AD estão: ID3 (QUINLAN, 1986), C4.5 (QUINLAN, 1993) e CART (BREIMAN et al., 1984), sendo o primeiro o mais utilizado. O algoritmo C4.5, uma evolução do ID3, constrói ADs a partir de um conjunto de dados de treinamento da mesma forma que o algoritmo ID3, utilizando o conceito de entropia para definir a

importância dos atributos. Em cada nó da árvore, o algoritmo C4.5 escolhe o atributo que melhor particiona o conjunto de amostras em subconjuntos tendendo a uma categoria ou a outra. O atributo com maior ganho de informação normalizado é escolhido para tomar a decisão (QUINLAN, 1993).

Durante o processo de geração de uma AD é importante conhecer alguns parâmetros que podem ser modificados para proporcionar resultados mais precisos. Dentre os parâmetros relevantes, destacam-se, o uso de podas na árvore, o número mínimo de instâncias por folha e a construção de árvore binária. Esses parâmetros podem ser ajustados de acordo com as características dos dados e do problema em questão.

As ADs têm sido amplamente utilizadas em diversas áreas do conhecimento, como em estudos de criminologia. Hassani et al. (2016); David e Suruliandi (2017); Matijosaitiene, McDonald e Juneja (2019) Ratul e Rab (2020) são exemplos de trabalhos nos quais as ADs foram empregadas para auxiliar no entendimento das relações espaciais dos crimes estudados.

2.5.2 WEKA – *Waikato Environment for Knowledge Analysis*

O WEKA é uma ferramenta que reúne algoritmos de IA e aprendizado de máquina (AM), sendo uma das mais populares ferramentas de mineração de dados no ambiente acadêmico. O desenvolvimento do WEKA teve início em 1993 na Universidade de Waikato – Nova Zelândia, e foi posteriormente adquirido por uma empresa no final de 2006. O WEKA é licenciado sob a General Public License, permitindo o estudo e a modificação do código fonte.

Entre os algoritmos de AD disponibilizados pelo WEKA está o J48, que é uma implementação do algoritmo C4.5, que é amplamente utilizado na literatura. Outras técnicas disponíveis na ferramenta são redes neurais artificiais e algoritmos genéticos, que têm sido aplicados em diversas áreas de pesquisa. Dakalbab et al. (2022) mencionam, em sua revisão da literatura sobre criminologia espacial, o uso do WEKA na análise de dados e predição de crimes.

3. MÉTODOS E MATERIAIS

3.1 Caracterização metodológica da pesquisa

Do ponto de vista de sua natureza, trata-se de uma pesquisa aplicada com o objetivo de gerar conhecimentos para aplicação prática dirigida à solução de problemas específicos, envolvendo fatos e interesses locais (CRESWELL, 2021).

A metodologia adotada pode ser classificada como exploratória, visto que busca proporcionar maior familiaridade com o problema de pesquisa para tornar mais explícitas as hipóteses e aprimorar ideias (CRESWELL, 2021).

Também pode ser considerada uma pesquisa experimental, visto que determina um objeto de estudo, seleciona as variáveis que podem influenciá-lo, além de definir as formas de controle e observação dos efeitos que a variável produz no objeto (CRESWELL, 2021).

Com relação ao objetivo, a pesquisa pode se enquadrar como descritiva, que, de acordo com Triviños (2008), é aquela que tem como objetivo descrever os fatos e fenômenos de determinada realidade.

Já do ponto de vista da abordagem, a pesquisa pode ser classificada como quantitativa, pois emprega métodos de investigação científica para medir variáveis preestabelecidas, a fim de verificar e explicar a influência de variáveis, por meio de análises utilizando a IA (CRESWELL, 2021).

Quanto aos procedimentos técnicos, adotou-se a revisão de literatura, estudo de caso e análise documental, sendo esta última amparada por dados de ocorrência de crimes de furto e roubo, oriundos da Secretaria de Segurança Pública do Estado de São Paulo (SSP/SP¹); dados sociodemográficos, provenientes do Censo Demográfico de 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE²); e Camadas de dados vetoriais obtidos junto ao IBGE e ao Portal GeoSampa.

¹ <http://www.ssp.sp.gov.br/transparenciassp/Consulta.aspx>

² <https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama>

De acordo com o recorte adotado neste estudo, o conjunto de dados inclui os casos de furto e roubo notificados nos distritos policiais da subprefeitura de Itaquera entre os anos de 2017 e 2021.

3.2 Contexto histórico e área de Estudo: subprefeitura de Itaquera, São Paulo/SP-Brasil

Itaquera, cujo nome significa "pedra dura" em Tupi-guarani, é uma região geologicamente composta por rochas ígneas e metamórficas duras, como granitos e gnaisses, localizada na Subprefeitura de Itaquera, em São Paulo/SP, Brasil (Prefeitura de São Paulo, 2022). Embora a primeira menção da região tenha sido registrada em uma Carta de Sesmaria em 1686, o principal marco do povoamento da área foi em 1820, com o estabelecimento de um rancho chamado "Casa Pintada" para viajantes (Prefeitura de São Paulo, 2022). O desenvolvimento de Itaquera ocorreu por meio da subdivisão de fazendas em lotes e surgiu a Vila Santana, dedicada à adoração de Santa Ana (FREITAS, 2020).

O primeiro loteamento na região ocorreu no sítio Caaguassu em 1919, para diversos fins, incluindo veraneio e chácaras, e outros grandes loteamentos foram realizados para abrigar os imigrantes japoneses que trabalhavam em atividades rurais. A chegada da ferrovia em 1875, na estação Ferroviária de Itaquera, impulsionou o crescimento comercial, especialmente em torno da estação, graças ao cultivo de café. A construção de casas e apartamentos pela Companhia Metropolitana de Habitação de São Paulo (COHAB) também teve um impacto significativo na região (FREITAS, 2020).

A construção da estação Corinthians-Itaquera do metrô, inaugurada em 1º de agosto de 1988, foi um marco importante no desenvolvimento da região, que se tornou ainda mais estratégica com a inauguração das estações Dom Bosco e José Bonifácio da linha Coral da Companhia de Trens Metropolitanos (CPTM) em 27 de maio de 2000. Isso aumentou significativamente o número diário de usuários em ambos os sentidos, consolidando a importância da região como um ponto estratégico de mobilidade urbana (CPTM, 2022).

Além disso, o Shopping Metrô Itaquera foi inaugurado em 2007, tornando-se um importante centro comercial e de lazer, atraindo um grande fluxo de pessoas para a

área. Tal atração também é impulsionada pela construção do estádio de futebol Neoquímica Arena, que comporta 49.205 pessoas, sediou grande eventos esportivos como a Copa do Mundo de 2014 e os Jogos Olímpicos de 2016, colocando a área no cenário nacional e internacional (FREITAS, 2020).

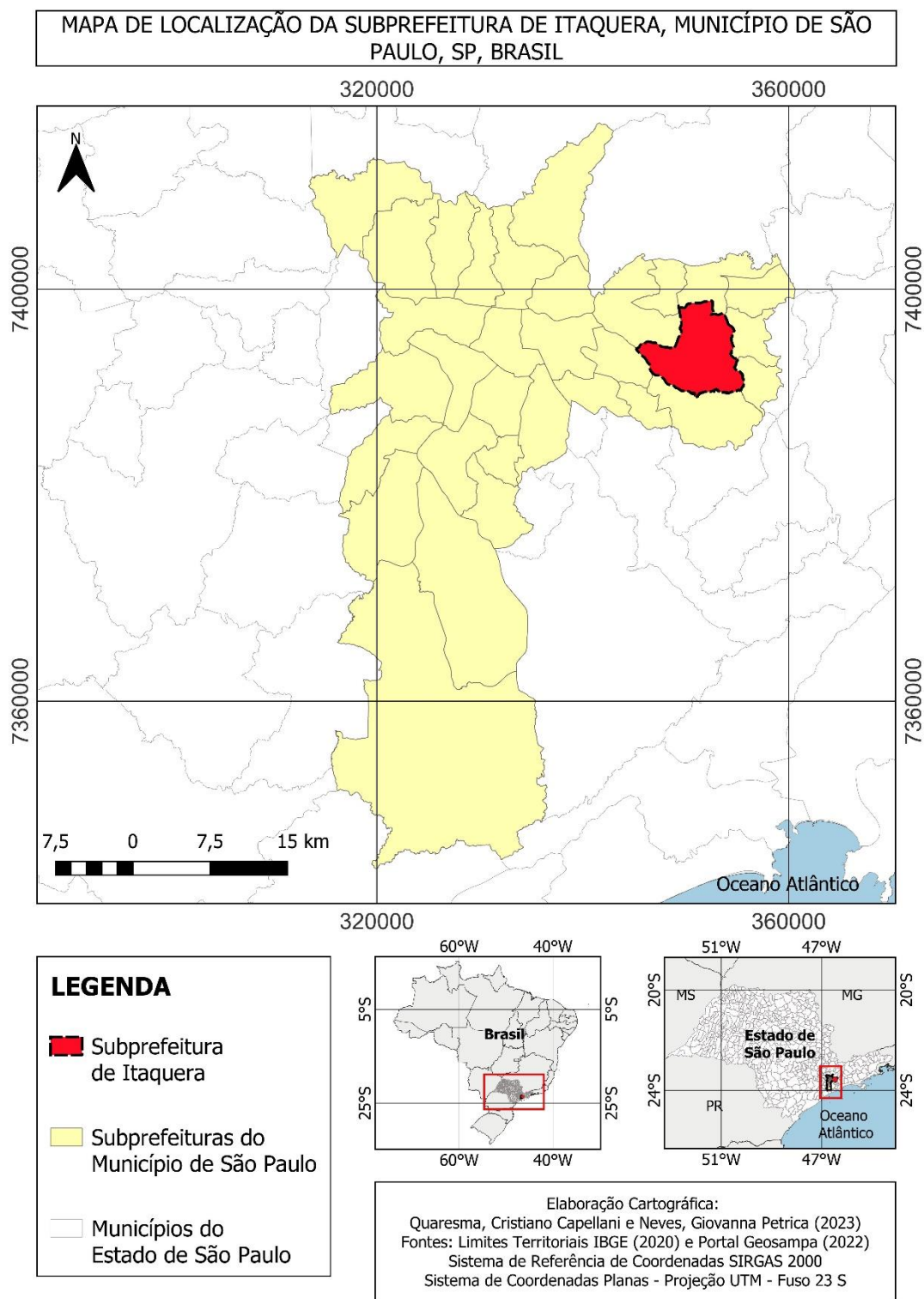
A Subprefeitura de Itaquera possui uma área total de 54,30 km² e uma população total de 560.055 habitantes, com uma densidade demográfica de 10.123,9 habitantes por km², sendo 268.398 (47,9%) do sexo masculino e 291.345 (52,1%) do sexo feminino (SEADE, 2023). A subprefeitura é adjacente a outras subprefeituras em termos administrativos e territoriais, incluindo São Mateus, Cidade Tiradentes, Guaianases, Aricanduva/Formosa, São Miguel Paulista e Penha. Cada distrito apresenta características espaciais distintas que atraem pessoas para suas atividades diárias. Por exemplo, o Parque Urbano do Carmo e o Sesc Itaquera são atrações proeminentes no distrito de Parque do Carmo, enquanto o Shopping Metrô Itaquera, a estação Corinthians-Itaquera (metrô/CPTM), o estádio de futebol NeoQuímica Arena, o Hospital Santa Marcelina, o Parque Raul Seixas e o Aquário de Itaquera são os destaques do distrito de Itaquera (RODRIGUES, 2023).

No que concerne à população, é relevante destacar que o distrito de Itaquera é o mais populoso da subprefeitura, com 212.875 habitantes. Em segundo lugar, está o distrito de José Bonifácio, com 138.247 habitantes, seguido pelo Cidade Líder, com 136.600, e pelo Parque do Carmo, com 72.333 habitantes (SEADE, 2023).

De acordo com informações da Prefeitura de São Paulo (2022), a densidade demográfica é mais alta no distrito de Itaquera (14.032 hab/km²), seguido pelo Cidade Líder (12.411 hab/km²), José Bonifácio (8.803 hab/km²) e Parque do Carmo (4.432 hab/km²).

A figura 3 destaca a localização da subprefeitura de Itaquera no município de São Paulo/SP-Brasil.

Figura 3 – Mapa de localização da Subprefeitura de Itaquera, São Paulo/SP-Brasil, 2023



Fonte: elaboração própria. Mapa Cartográfico Digital: IBGE (2020); GeoSampa, PMSP, 2023.

Para cerca de 73% da população total dos distritos administrativos da subprefeitura de Itaquera a renda média é de 1/3 do salário-mínimo, devido ao trabalho informal e ao baixo nível educacional (SEADE, 2023). No entanto, os distritos diferem significativamente em termos de renda e educação. Por exemplo, enquanto apenas 8,7% da população de Itaquera possui emprego formal, a porcentagem sobe para 10,1% na Cidade Líder, 10,2% no Parque do Carmo e 14,9% em José Bonifácio. Em relação à educação, o distrito de Parque do Carmo tem a maior porcentagem de pessoas com nível superior, com 17,0% da população total, seguido por Itaquera com 12,3%, José Bonifácio com 6,8% e Cidade Líder com apenas 5,4% (SEADE, 2023).

3.3 Dados e variáveis

3.3.1 Dados cartográficos

Para a elaboração dos mapas, utilizamos camadas vetoriais obtidas do Instituto Brasileiro de Geografia e Estatística (IBGE) e do Sistema de Consulta do Mapa Digital da Cidade de São Paulo (GeoSampa). No site do IBGE, obtivemos as camadas vetoriais de Unidades de Federação do Brasil e Limites dos Setores Censitários. Já no GeoSampa, adquirimos as camadas vetoriais de Limites dos Municípios do Estado de São Paulo, Limites das Subprefeituras do Município de São Paulo e Distritos do Município de São Paulo.

3.3.2 Variáveis de resposta

Seguindo a metodologia proposta em Rodrigues (2023), foram adotados, como variáveis de resposta (tabela 2), os casos de furto e roubo registrados na subprefeitura de Itaquera durante os anos de 2017 a 2021. Os dados são oriundos da Segurança Pública de São Paulo (SSP). As tabelas possuem informações como a identificação da delegacia, nome da seccional, nome da delegacia, ano e número do boletim de ocorrência, descrição do tipo do boletim de ocorrência, data e hora da ocorrência, rubrica do crime, descrição da conduta, local do crime, sexo, cor da pele, profissão, grau de instrução, nacionalidade e naturalidade. No entanto, em muitos casos campos como sexo, cor da pele, profissão e grau de instrução não foram registrados.

Tabela 2 – Variáveis de resposta

Variável	Descrição	Fonte / Referência
Roubo	“Subtrair coisa móvel alheia, para si ou para outrem, mediante grave ameaça ou violência à pessoa, ou depois de havê-lo, por qualquer meio, reduzido à impossibilidade de resistência”	Art. 157 do Decreto-Lei n.º 2.848, de 07 de dezembro de 1940 (CP)
Furto	“Subtrair, para si ou para outrem, coisa alheia móvel”	Art. 155 do Decreto-Lei n.º 2.848, de 07 de dezembro de 1940 (CP)

Fonte: Adaptado de Rodrigues (2023)

3.3.3 Variáveis explicativas

As variáveis explicativas consideradas nesta pesquisa são aquelas denominadas "variáveis explicativas baseadas na teoria da desorganização social", as quais foram definidas com base na revisão da literatura, conforme indicado na seção 2.3. Os dados referentes a essas variáveis foram extraídos das bases do IBGE. Destaca-se que as variáveis "Privação Social" e "Taxa de Desemprego" foram excluídas devido à dificuldade de localização de dados públicos. Tanto as variáveis quanto a origem dos dados relativos a elas estão detalhados nas tabelas 3 e 4.

Tabela 3 – Variáveis explicativas com base na teoria da desorganização social

Variável explicativas – Teoria da desorganização social	Descrição IBGE	Fonte / Referência
Renda Média (REM)	Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento)	(Adewuyi et al., 2017; Baumer e Arnio, 2012; Ceccato e Moreira, 2021).
Percentual de População Jovem (PPJ)	Proporção de indivíduos da faixa etária de 15 a 35 anos	(Baumer e Arnio, 2012; Ceccato e Moreira, 2021; Adewuyi et al., 2017).
Densidade Populacional (DEP)	Média de moradores em domicílios particulares permanentes.	(Adewuyi et al., 2017; Baumer e Arnio, 2012;).
Percentual de População Masculina (PPM)	Proporção de indivíduos do sexo masculino, homens residentes em domicílios particulares permanentes	(Adewuyi et al., 2017; Ceccato e Moreira, 2021)
Percentual de População Alfabetizada (PPA)	Proporção de pessoas alfabetizadas com 5 ou mais anos de idade	(Adewuyi et al., 2017).

Fonte: Adaptado de Rodrigues (2023)

Os dados relativos às variáveis explicativas baseadas na teoria da desorganização social, apresentadas na Tabela 3, foram extraídos da base de informações do Censo Demográfico 2010 do IBGE. Eles estão agrupados por setor censitário e descrevem as características socioeconômicas da população e dos domicílios pertencentes a cada setor. A aquisição dos dados foi feita com auxílio do plugin “Censo IBGE” instalado no QGIS (versão 3.28.5). É importante destacar que identificação da planilha Excel de origem dos dados relativos a cada variável foi feita por meio do documento do IBGE intitulado “Base de informações do Censo Demográfico 2010: Resultados do Universo por setor censitário”³. As planilhas, das quais foram extraídos os dados relativos a cada variável, estão indicadas na tabela 4.

³ https://www.cidadessustentaveis.org.br/arquivos/SIG/Senso2010_Dados/Documenta%C3%A7%C3%A3o-IBGE_Censo2010.pdf

Tabela 4 – Origem dos dados relativos às variáveis explicativas baseadas na teoria da desorganização social

Variável explicativas – Teoria da desorganização social	Tabela	Campo(s)
Renda Média (REM)	Arquivo Básico	V005 – Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento)
Percentual de População Jovem (PPJ)	Arquivo Idade, total	V049 – Pessoas com 15 anos de idade a V069 – Pessoas com 35 anos de idade
Densidade Populacional (DEP)	Arquivo Básico	V003 – Média do número de moradores em domicílios particulares permanentes
Percentual de População Masculina (PPM)	Arquivo Idade, homens	V002 – Homens residentes em domicílios particulares permanentes
Percentual de População Alfabetizada (PPA)	Arquivo Alfabetização, total	V001 – Pessoas alfabetizadas com 5 ou mais anos de idade

Fonte: elaboração própria

3.4 Abordagem proposta para análise de dados e variáveis usando IA

A abordagem proposta nesta pesquisa, cujo diagrama de funcionamento é ilustrado na figura 4, tem como entrada dados de furto e roubo oriundos da SSP, dados socioeconômicos do IBGE que serão analisados por meio de técnicas de MD e IA visando a descoberta de padrões para auxiliar no entendimento de como a ocorrência dos crimes se relacionam com atributos socioespaciais urbanos. As etapas componentes da abordagem, que incluem os passos do KDD, são apresentadas nas subseções 3.4.1 a 3.4.6 e foram projetadas para contemplar os objetivos específicos apresentados na seção 1.4.

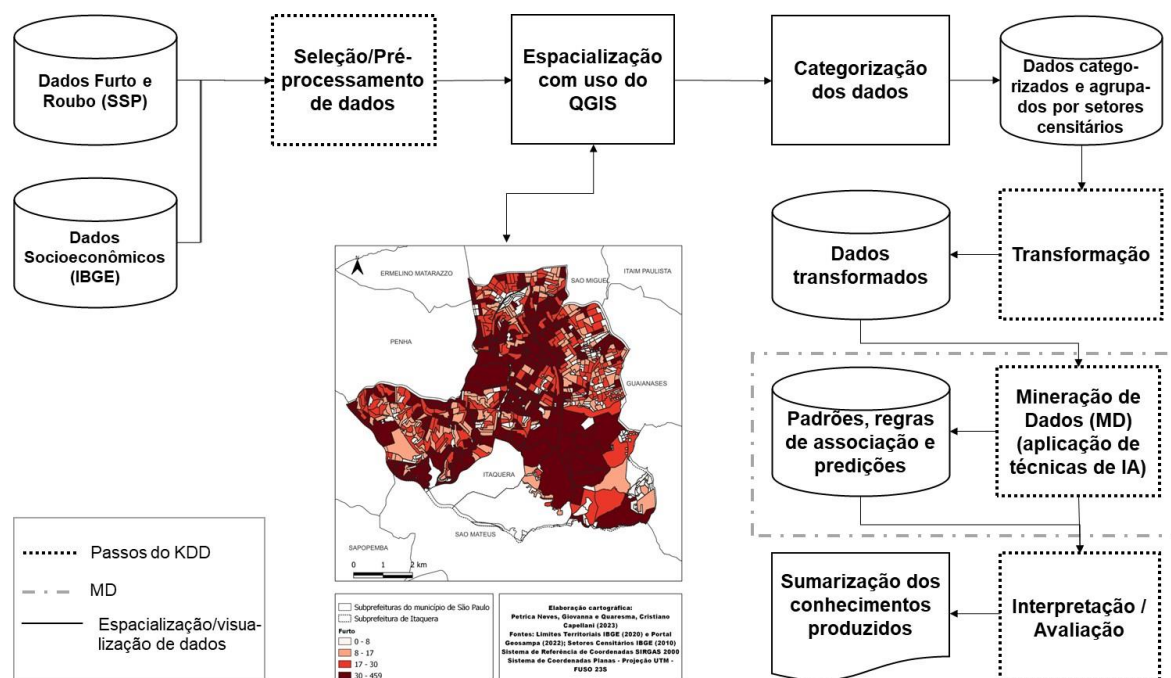


Figura 4 – Diagrama de funcionamento da abordagem de IA proposta

3.4.1 Seleção de dados

Os dados relativos às ocorrências de furto e roubo (variáveis de resposta) são oriundos do Registro Digital de Ocorrências utilizado pelas delegacias de polícia e pela delegacia eletrônica da Secretaria de Estado dos Negócios da Segurança Pública de São Paulo (SSP). Cada tabela (em formato Excel) contendo registros de um determinado ano possui 52 campos nos quais são armazenados dados como a identificação da delegacia, ano e número do boletim de ocorrência (BO), data e hora da ocorrência, rubrica do crime, descrição da conduta, coordenadas geográficas indicando o local do crime (latitude e longitude), alguns dados da vítima (sexo, cor da pele, profissão, grau de instrução, nacionalidade e naturalidade), entre outros.

Os dados socioeconômicos relativos às variáveis explicativas baseadas na teoria da desorganização social (apresentadas na tabela 3) foram extraídos da base de informações do Censo Demográfico 2010 do IBGE. Eles são agrupados por setor censitário e descrevem as características socioeconômicas da população e dos domicílios pertencentes a cada setor. A aquisição dos dados foi feita com auxílio do plugin “Censo IBGE” instalado no QGIS. É importante destacar que identificação da planilha Excel de origem dos dados relativos a cada variável foi feita por meio do

documento do IBGE intitulado “Base de informações do Censo Demográfico 2010: Resultados do Universo por setor censitário”, como descrito na tabela 4.

3.4.2 Pré-processamento

Inicialmente, as planilhas referentes as ocorrências de furto e roubo no período 2017 a 2021 na subprefeitura de Itaquera foram analisadas visando a exclusão de registros repetidos (associados às pessoas envolvidas em um mesmo BO) e contendo dados faltantes, como por exemplo as coordenadas geográficas, imprescindíveis para associação espacial de cada registro a um setor censitário.

3.4.3 Espacialização de dados em SIG

Após o procedimento de limpeza, as planilhas foram salvas em formato CSV e importadas no QGIS, utilizando a função "Gerenciador de Fonte de Dados – Texto Delimitado". Após a espacialização dos pontos dos crimes no SIG, a partir da inserção das camadas de furto e roubo, foi verificado que algumas ocorrências extrapolavam os limites territoriais da subprefeitura de Itaquera, sendo necessário suas exclusões com auxílio da ferramenta de exclusão de geometrias duplicadas e de correção de geometrias, ambas pertencentes à caixa de ferramentas de processamento do QGIS. Apesar dessas operações pertencerem à etapa de pré-processamento, suas execuções só foram possíveis após a espacialização dos dados.

Com os arquivos salvos, a ferramenta do QGIS para associação de atributos por localização foi utilizada para unir as tabelas de dados dos crimes (variáveis de resposta) com aquelas contendo os dados relativos às variáveis explicativas. Assim, os dados foram agrupados por setores censitários.

3.4.4 Transformação

Para viabilizar a aplicação do algoritmo de AD C4.5, todos os dados representados por números contínuos precisaram ser categorizados a partir das classes definidas na etapa de espacialização de dados. Assim, para cada variável os dados foram agrupados em intervalos de valores gerando as categorias. É importante mencionar que existem algoritmos de IA, como as árvores de regressão (AR), capazes de lidar com dados contínuos. No entanto, os conhecimentos produzidos por AR não são tão

intuitivos como aqueles produzidos pelas ADs, que podem ser lidas facilmente como um conjunto de regras do tipo SE...ENTÃO, como explicado na subseção 2.6.1.

3.4.5 Mineração de dados

Realizada a partir das tarefas de classificação usando AD, algoritmo C4.5, a aplicação das técnicas de IA nos dados transformados e agrupados por setor censitário foi feita com auxílio do Software WEKA. Esta etapa visa a identificação de padrões que explicam as relações entre as ocorrências de furto e roubo e as variáveis que caracterizam o espaço urbano.

3.4.6 Interpretação/avaliação

Os resultados produzidos na etapa de MD foram analisados e interpretados a partir das regras produzidas pelas ADs, visando o atingimento dos objetivos específicos 1.4.2 a 1.4.4. Não obstante, as avaliações dos resultados das ADs na produção de padrões foram feitas com base na taxa de acerto (sensibilidade), na precisão e no índice Kappa, que podem ser calculadas a partir das equações 1 a 3.

$$\text{Taxa de acerto} = \frac{VP}{VP + FN} \quad (1)$$

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (2)$$

$$K = 1 - \frac{1 - (VP + VN) / (VP + FP + VN + FN)}{1 - (VP + FP) \times (VP + FN) + (FN + VN) \times (FP + VN) / (VP + FP + FN + VN)^2} \quad (3)$$

Essas medidas podem ser obtidas a partir de uma matriz de confusão, como a apresentada na Tabela 5, a qual reflete os acertos (verdadeiros positivos: VP e verdadeiros negativos: VN) na sua diagonal principal. Os demais elementos da matriz contemplam os erros do classificador (falsos positivos: FP e falsos negativos: FN).

Tabela 5 – Exemplo de matriz de confusão

Classe real		Classe predita
+	-	
VP	FP	+
FN	VN	-

Fonte: elaboração própria

O coeficiente Kappa é muito utilizado em MD para avaliar a concordância dos resultados ou ainda expressar a confiabilidade do classificador, conforme indicado na Tabela 6.

Tabela 6 – Níveis de concordância do coeficiente Kappa

K	Nível de concordância
Menor que zero	Insignificante (<i>poor</i>)
0 – 0,2	Fraca (<i>slight</i>)
0,21 – 0,4	Razoável (<i>fair</i>)
0,41 – 0,6	Moderada (<i>moderate</i>)
0,61 – 0,8	Forte (<i>substantial</i>)
0,81 – 1	Quase perfeita (<i>almost perfect</i>)

Fonte: Adaptado de Landis e Koch (1977)

3.5 Síntese metodológica

Na tabela 7 é apresentada a síntese metodológica deste trabalho, contendo os objetivos da pesquisa, a fundamentação teórica e os procedimentos metodológicos aplicados.

Tabela 7 – Síntese metodológica

	Objetivos específicos	Fundamentação teórica	Metodologia
1.3 Objetivo geral Analisar as relações espaciais dos crimes de furto e roubo com indicadores sociodemográficos na subprefeitura de Itaquera, Zona Leste de São Paulo/SP – Brasil.	1.4.1 Mapear a distribuição espacial dos crimes de furto e roubo e de dados sociodemográficos na subprefeitura de Itaquera;	2.1 Padrões espaciais dos crimes de furto e roubo, e variáveis sociodemográficas	3.4.1 Seleção de dados 3.4.2 Pré-processamento 3.4.3 Espacialização de dados em SIG
	1.4.2 Identificar padrões de distribuição espacial das ocorrências de furto e roubo na subprefeitura de Itaquera;	2.1 Padrões espaciais dos crimes de furto e roubo e variáveis sociodemográficas 2.2 Padrões espaciais dos crimes e IA 2.4 Inteligência Artificial 2.5 Mineração de Dados	3.4.1 Seleção de dados 3.4.2 Pré-processamento 3.4.3 Espacialização de dados em SIG 3.4.4 Transformação 3.4.5 Mineração de dados 3.4.6 Interpretação/avaliação
	1.4.3 Analisar as relações entre as variáveis sociodemográficas e os padrões espaciais de distribuição dos casos de furto e roubo na subprefeitura de Itaquera;		
	1.4.4 Identificar padrões das ocorrências de crimes na subprefeitura de Itaquera com base nas características pessoais das vítimas;		

Fonte: elaboração própria

4. RESULTADOS

Neste capítulo são apresentados e discutidos os resultados obtidos nas diferentes etapas da abordagem proposta, incluindo a espacialização de dados (seção 4.1), análise da distribuição dos dados sociodemográficos e dos crimes de furto e roubo (seção 4.2), descoberta de padrões usando MD (seção 4.3) e a discussão dos resultados (seção 4.4).

4.1 Espacialização de dados

4.1.1 Espacialização dos dados sociodemográficos

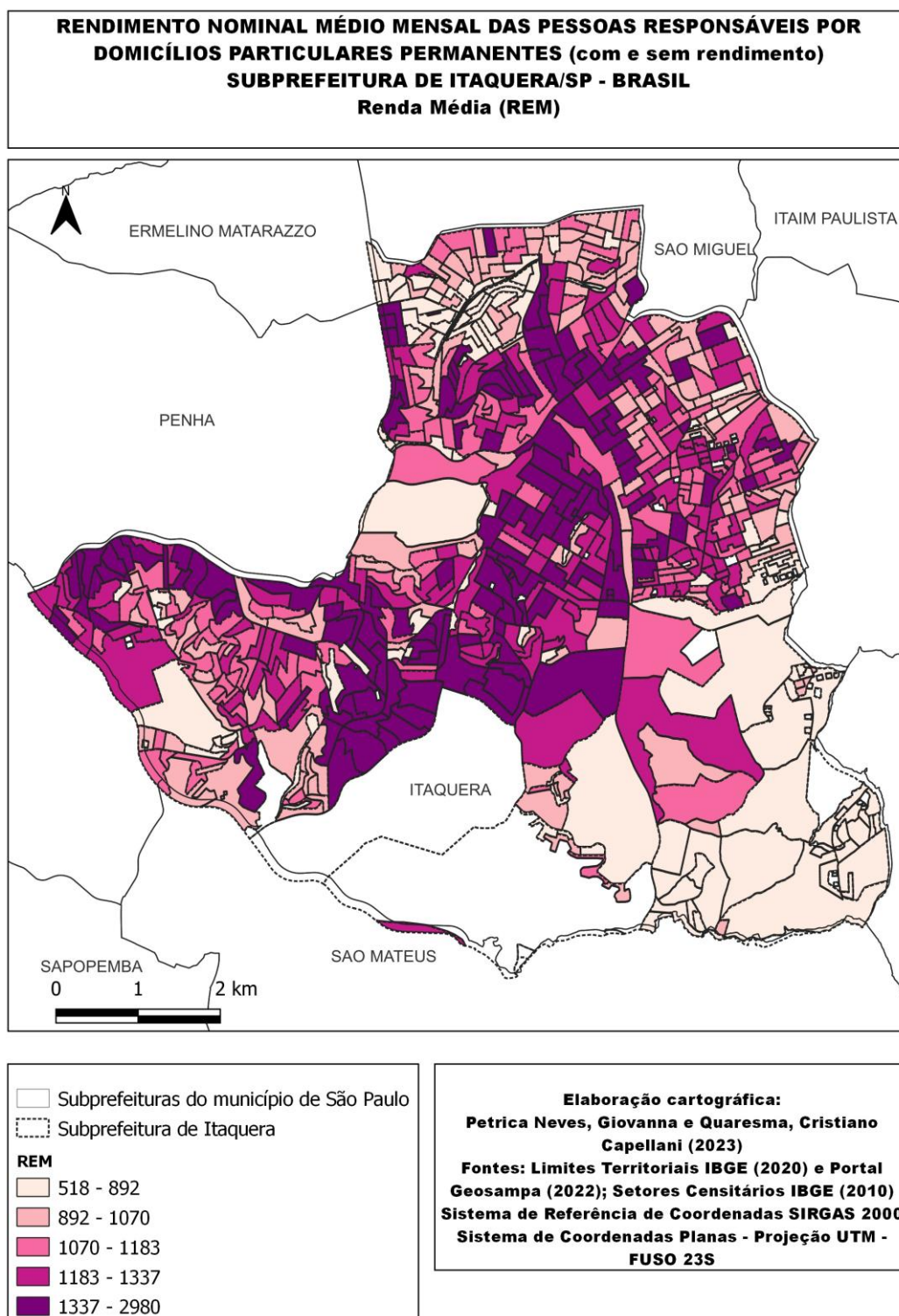
Essa seção apresenta a distribuição espacial de dados sociodemográficos visando satisfazer o objetivo específico 1.4.1. Os registros de crimes de furto e roubo foram organizados por setor censitário e categorizados utilizando o software QGIS, por meio da função "Iguar Contagem (quartil)". Essa função foi utilizada para garantir uma distribuição equitativa entre todas as categorias.

Os mapas ilustrados nas Figuras de 5 a 9 representam mapas temáticos que sintetizam os valores das variáveis explicativas renda média (REM), percentual de população jovem (PPJ), densidade populacional (DEP), percentual de população masculina (PPM) e percentual de população alfabetizada (PPA) para os setores censitários que compõem a subprefeitura de Itaquera. Os valores para essas variáveis são oriundos do censo IBGE de 2010, conforme explicado na seção 3.3.3.

Os mapas gerados mostram a distribuição dos valores das variáveis distribuídos em cinco intervalos representando classes, cujas cores se tornam progressivamente mais escuras à medida que os valores aumentam.

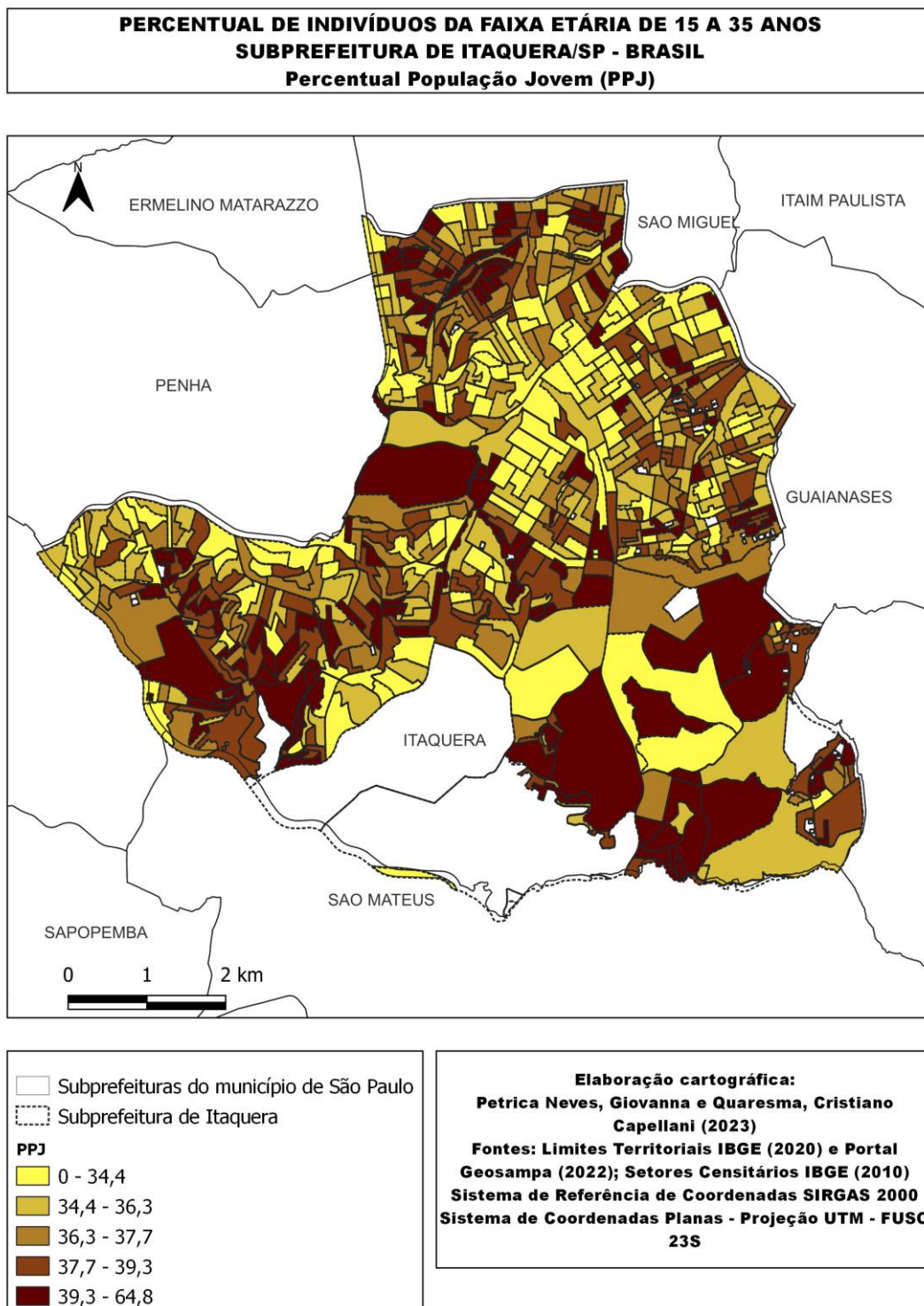
É importante destacar que os intervalos de valores definidos para as variáveis, representando as classes, são usados na mineração de dados como uma forma de categorizar os dados.

Figura 5 – Distribuição espacial da renda média (REM)



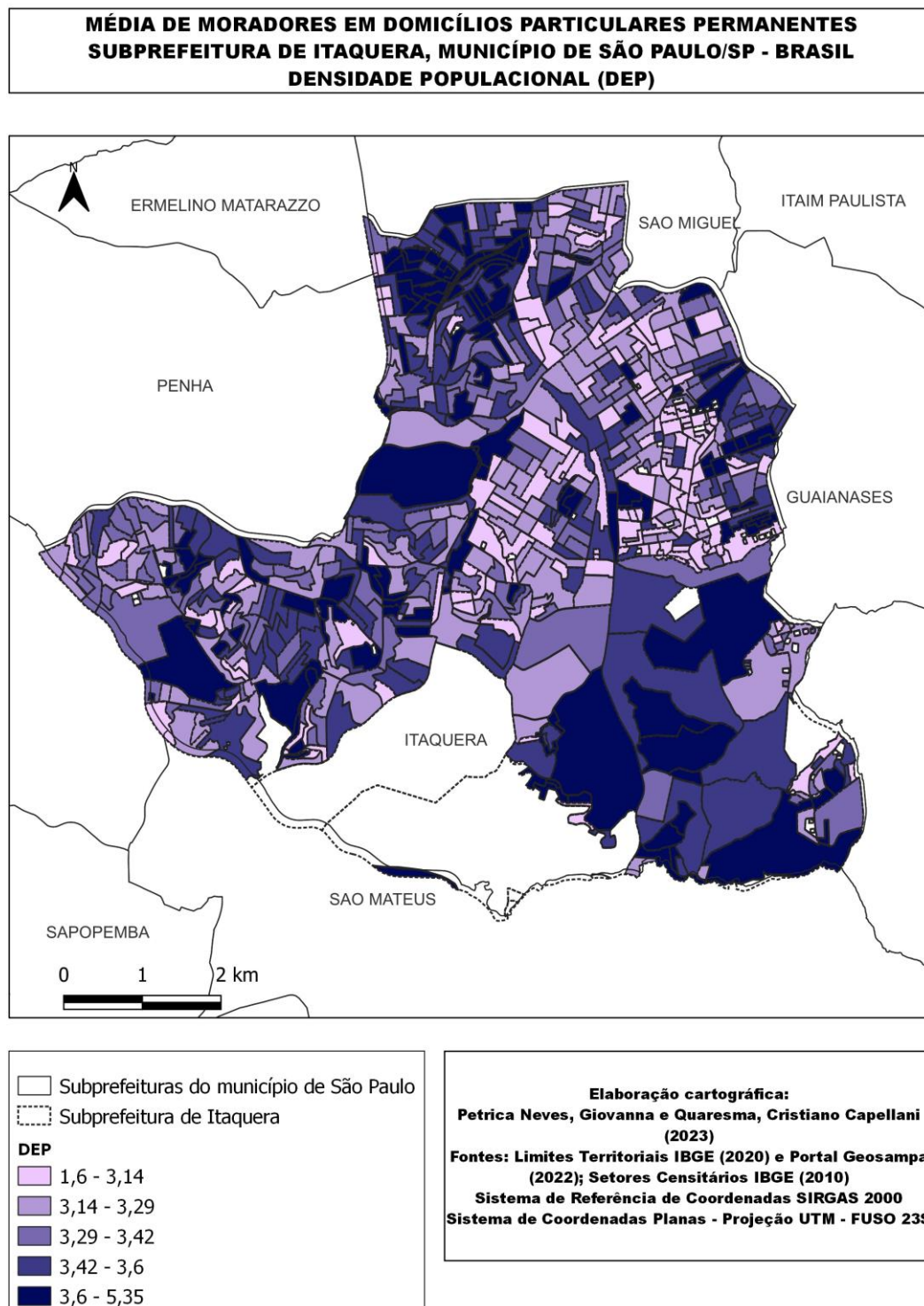
Fonte: elaboração própria. Mapa Cartográfico Digital: IBGE (2020).

Figura 6 – Distribuição espacial do percentual de população jovem (PPJ)



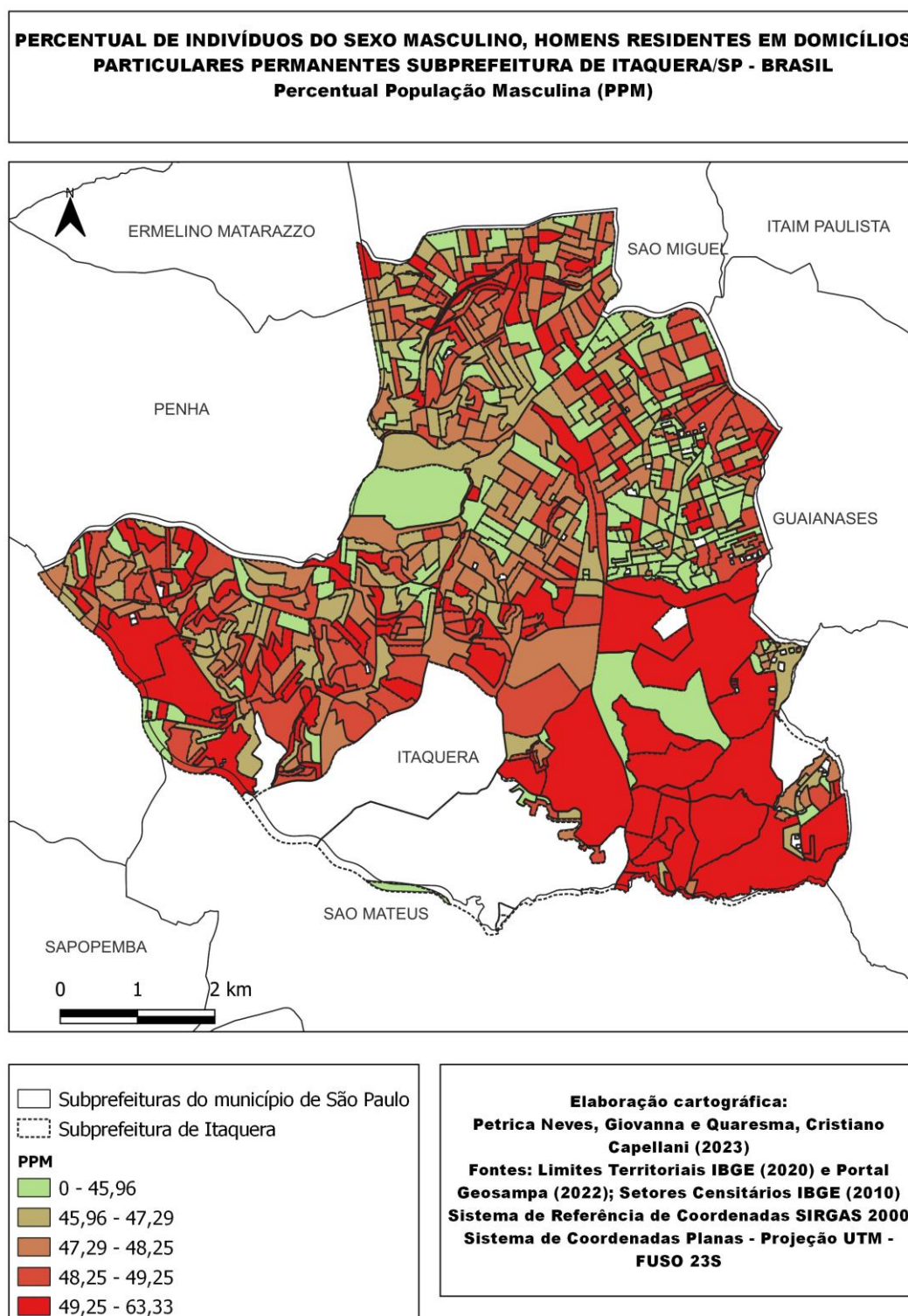
Fonte: elaboração própria. Mapa Cartográfico Digital: IBGE (2020).

Figura 7 – Distribuição espacial de densidade populacional (DEP)



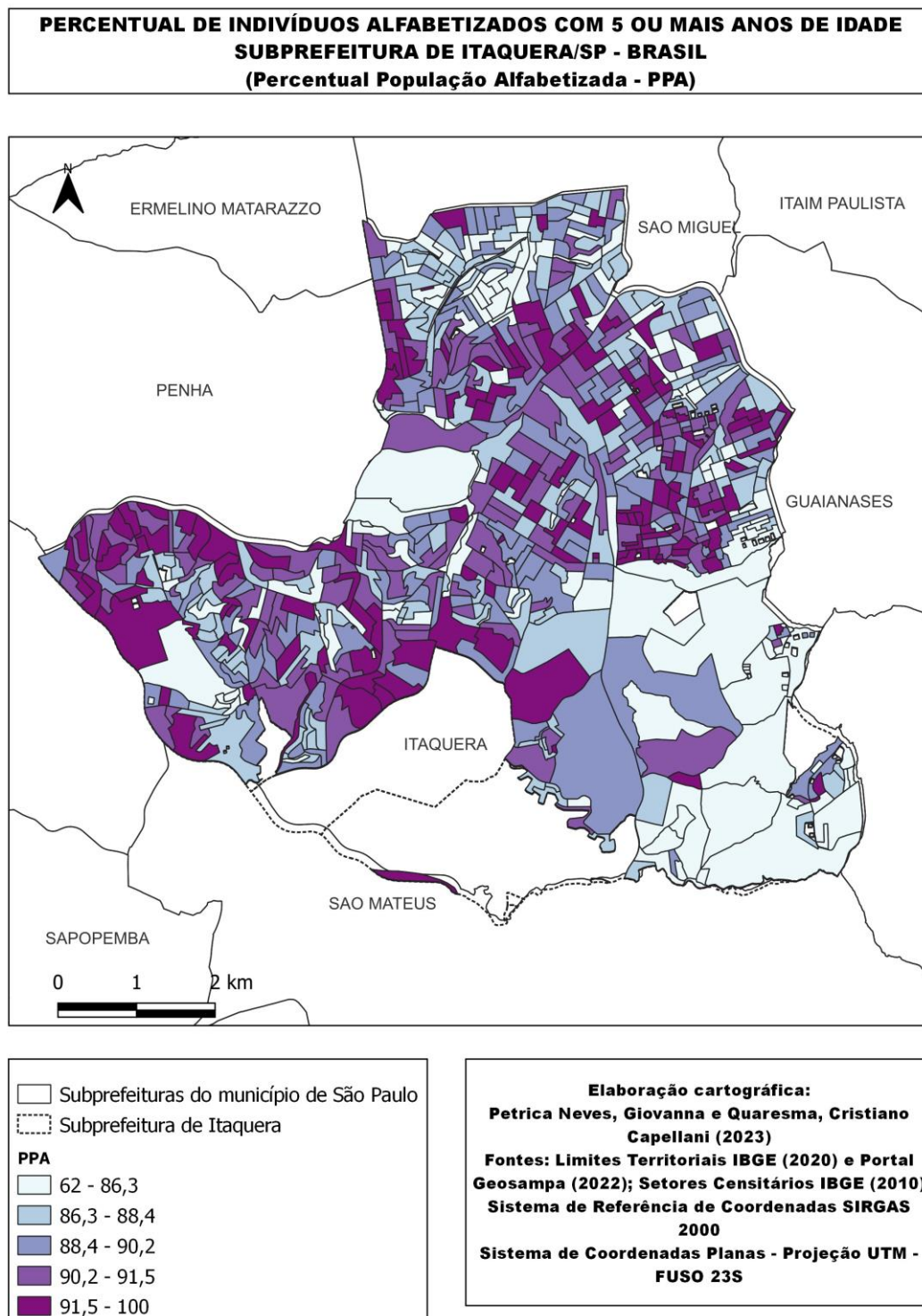
Fonte: elaboração própria. Mapa Cartográfico Digital: IBGE (2020).

Figura 8 – Distribuição espacial do percentual de população masculina (PPM)



Fonte: elaboração própria. Mapa Cartográfico Digital: IBGE (2020).

Figura 9 – Distribuição espacial do percentual de população alfabetizada (PPA)



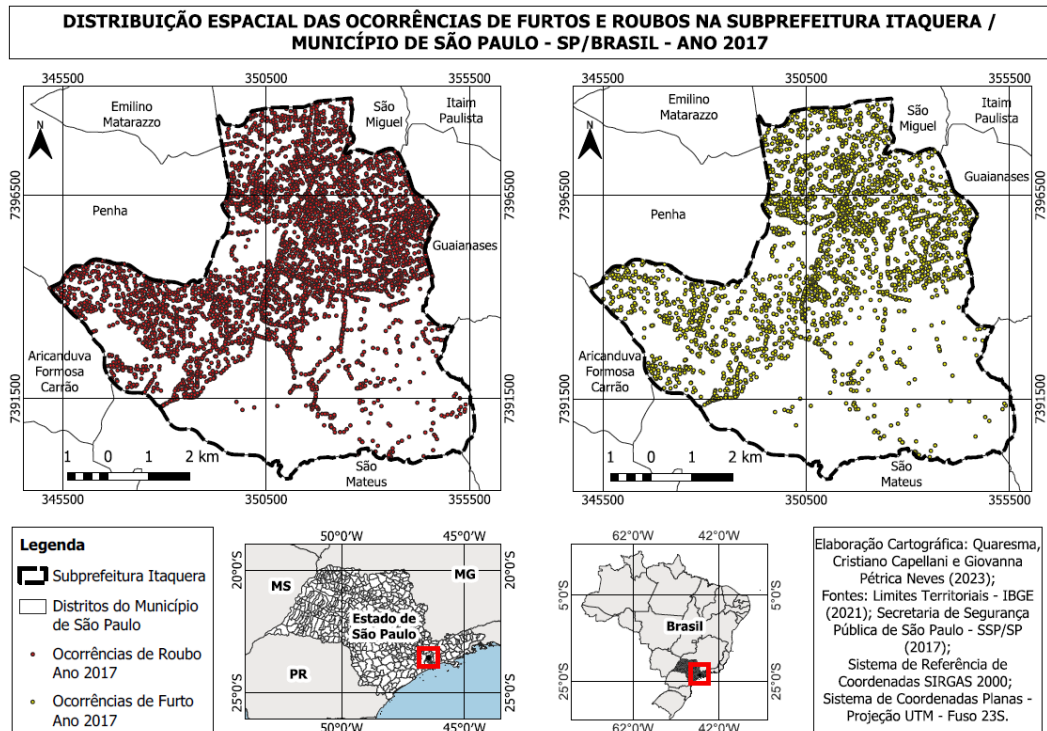
Fonte: elaboração própria. Mapa Cartográfico Digital: IBGE (2020).

4.1.2 Espacialização das ocorrências de furto e roubo

Para atender ao segundo objetivo específico (item 1.4.2) foi realizada a espacialização das ocorrências de furto e roubo registradas entre 2017 e 2021 (figuras 10 a 17), também usando o QGIS.

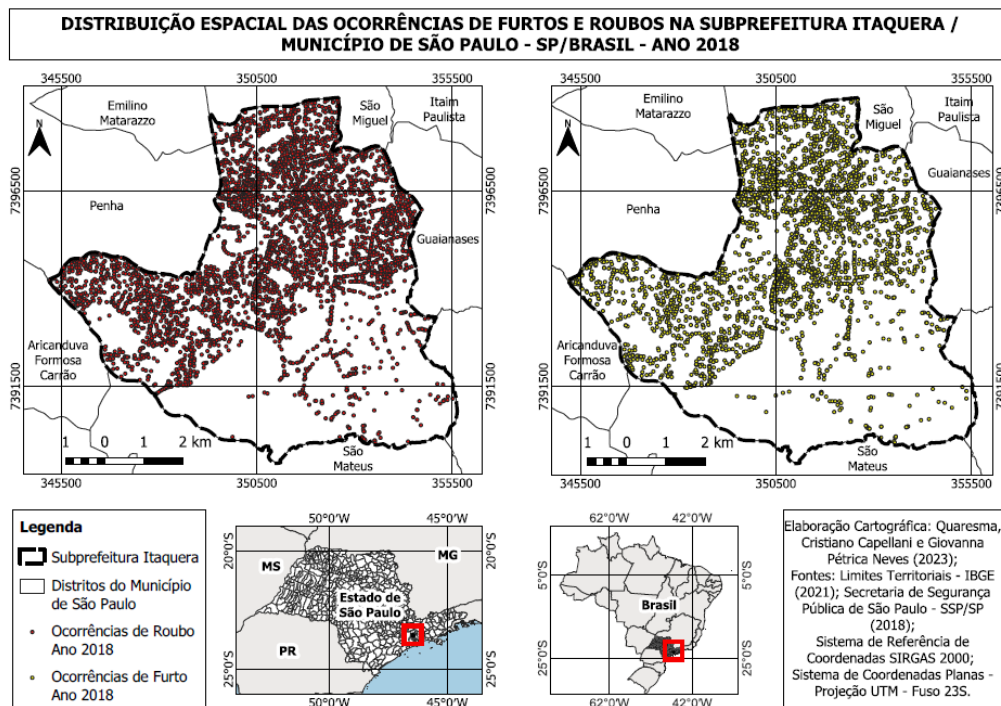
Nas figuras 10 a 14 são apresentadas as distribuições anuais de furtos e roubos, enquanto as figuras 15 a 17 apresentam as distribuições acumuladas no período considerado.

Figura 10 – Distribuição espacial das ocorrências de furtos e roubos na Subprefeitura de Itaquera, São Paulo/SP-Brasil no ano de 2017



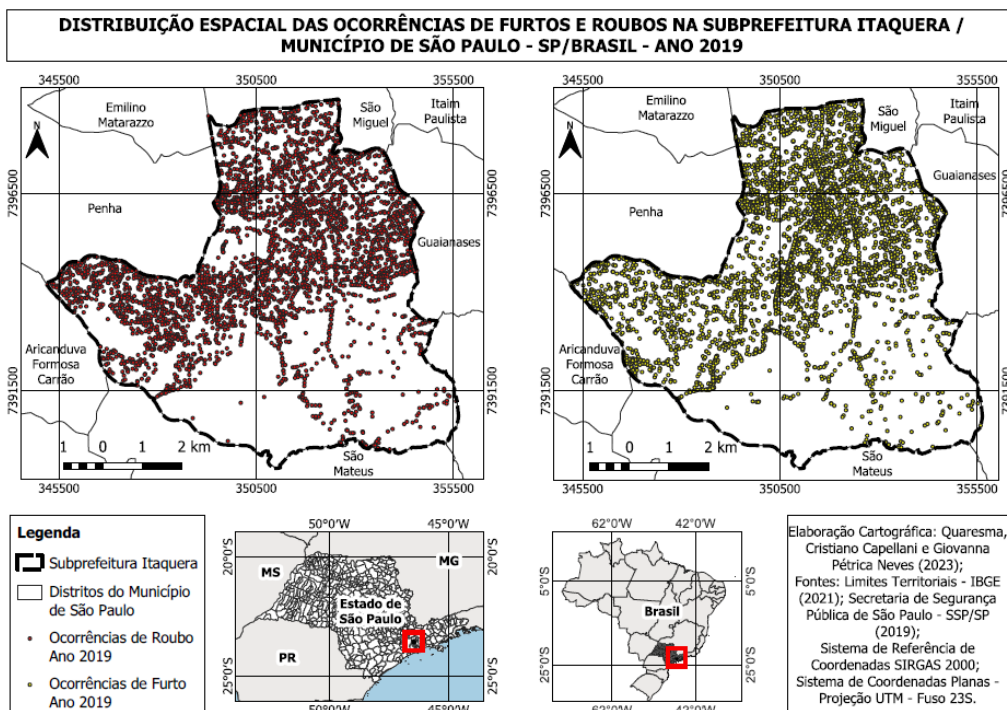
Fonte: elaboração própria.

Figura 11 – Distribuição espacial das ocorrências de furtos e roubos na Subprefeitura de Itaquera, São Paulo/SP-Brasil no ano de 2018



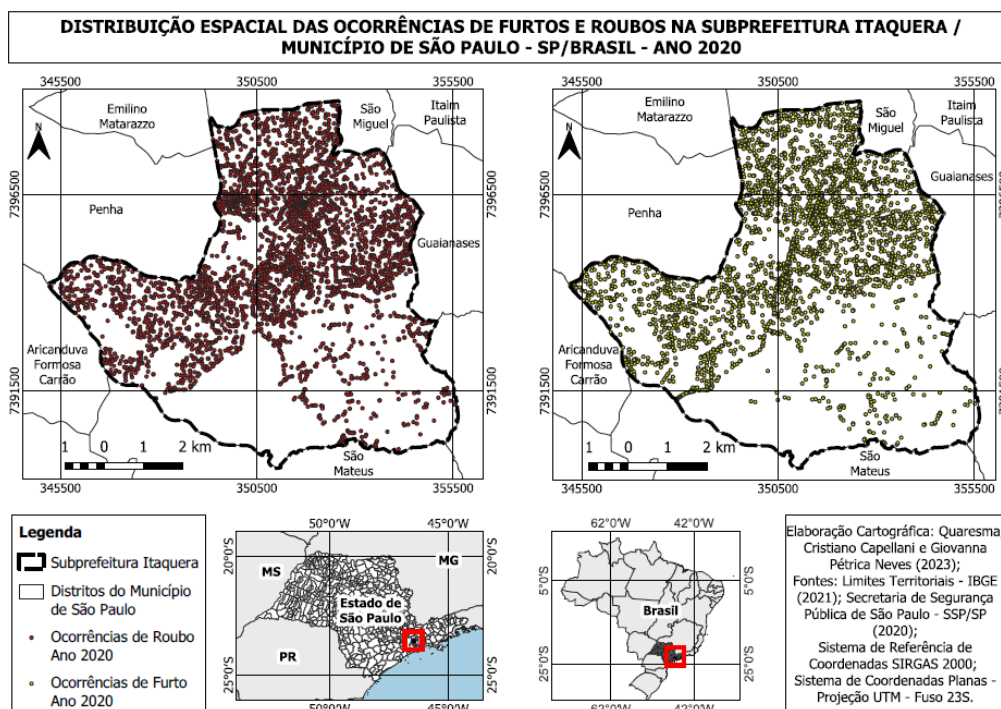
Fonte: elaboração própria

Figura 12 – Distribuição espacial das ocorrências de furtos e roubos na Subprefeitura de Itaquera, São Paulo/SP-Brasil no ano de 2019



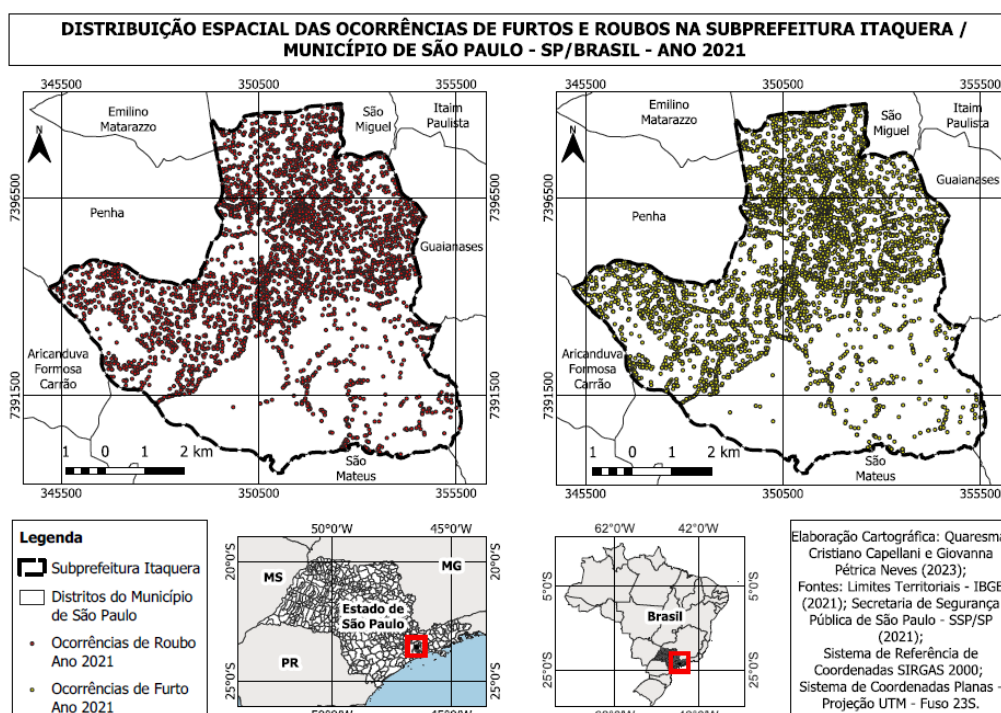
Fonte: elaboração própria

Figura 13 – Distribuição espacial das ocorrências de furtos e roubos na Subprefeitura de Itaquera, São Paulo/SP-Brasil no ano de 2020



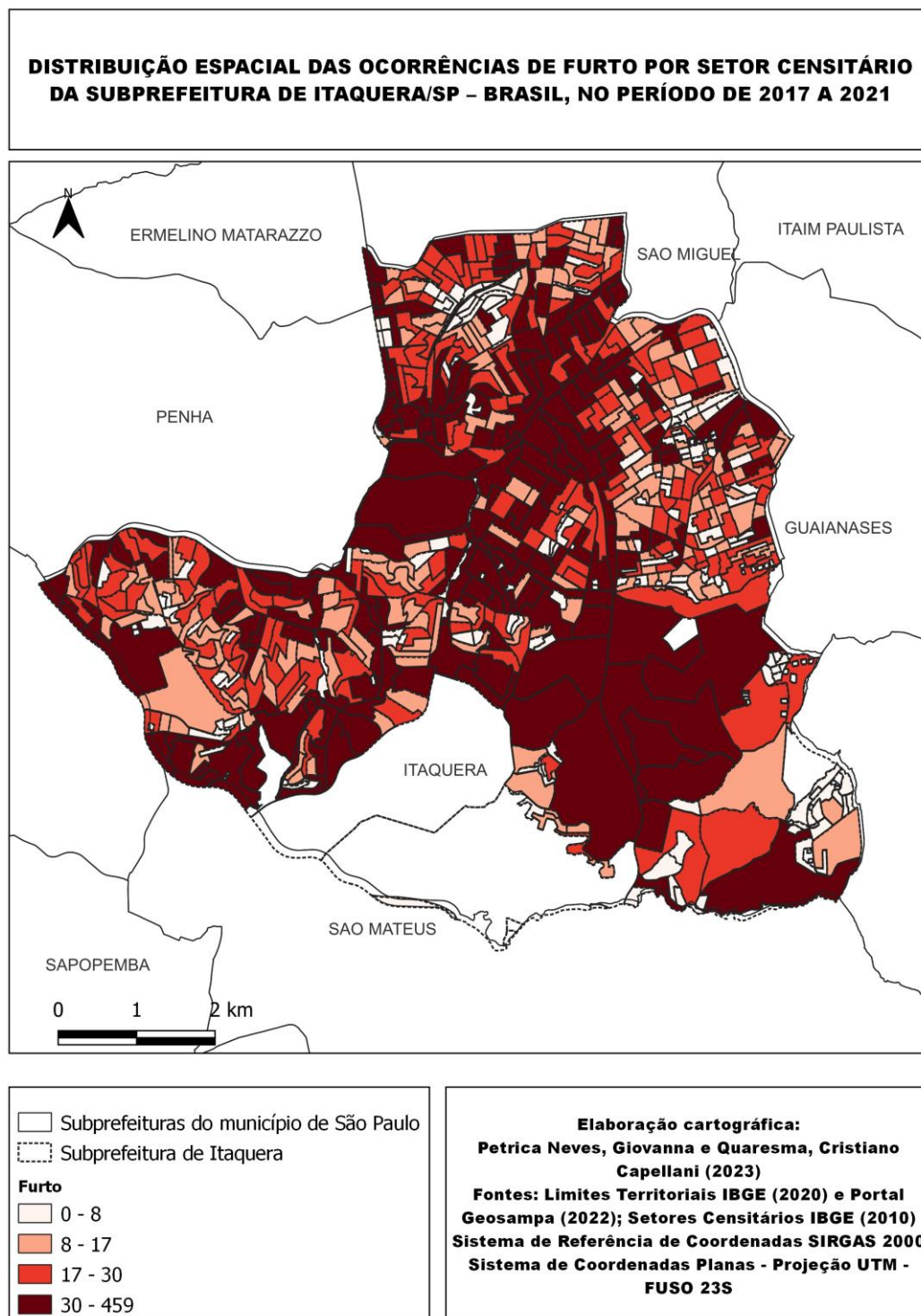
Fonte: elaboração própria

Figura 14 – Distribuição espacial das ocorrências de furtos e roubos na Subprefeitura de Itaquera, São Paulo/SP-Brasil no ano de 2021



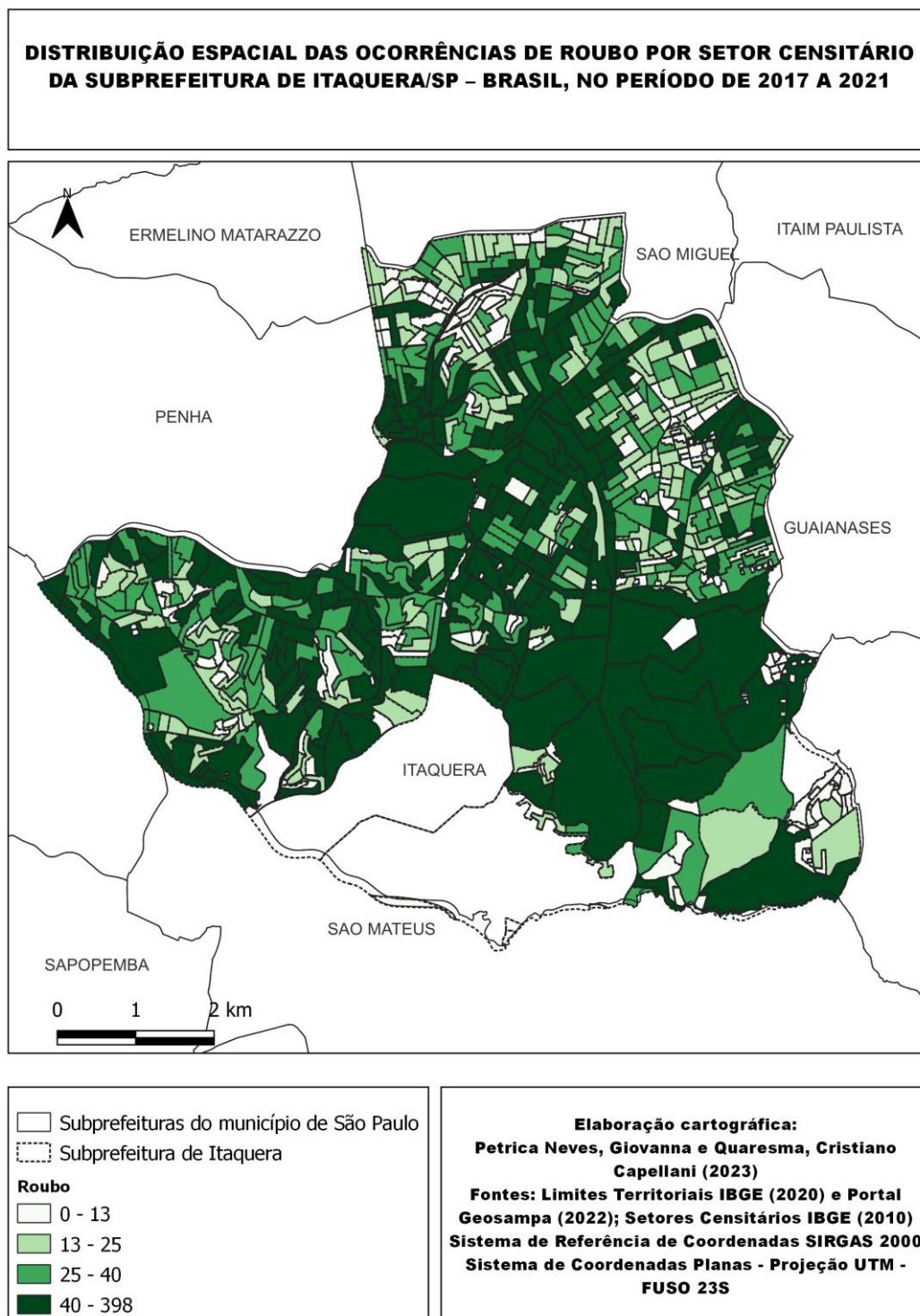
Fonte: elaboração própria

Figura 15 – Distribuição espacial das ocorrências de furto por setor censitário da subprefeitura de Itaquera/SP – Brasil, no período de 2017 a 2021



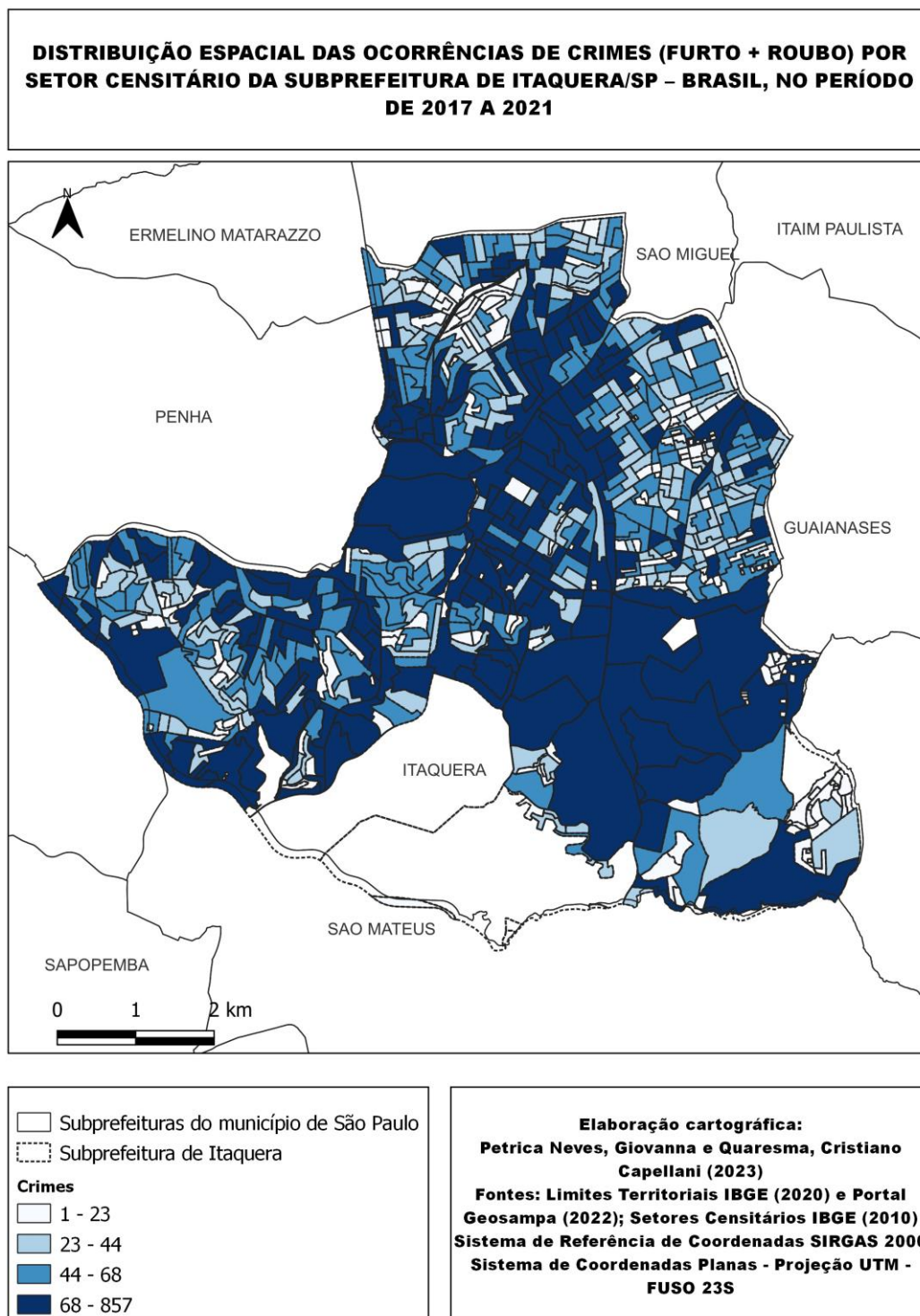
Fonte: elaboração própria.

Figura 16 – Distribuição espacial das ocorrências de roubo por setor censitário da subprefeitura de Itaquera/SP – Brasil, no período de 2017 a 2021



Fonte: elaboração própria.

Figura 17 – Distribuição espacial das ocorrências de crimes (furto + roubo) por setor censitário da subprefeitura de Itaquera/SP – Brasil, no período de 2017 a 2021



Fonte: elaboração própria.

4.2 Análise da distribuição dos dados sociodemográficos e dos crimes de furto e roubo

As análises apresentadas nesta seção foram realizadas com o auxílio do WEKA, a partir da ferramenta de visualização de dados, e contribuem para o atingimento do objetivo específico 1.4.3, a saber, “Analisar as relações entre as variáveis sociodemográficas e os padrões espaciais de distribuição dos casos de furto e roubo na subprefeitura de Itaquera”.

Os números absolutos de furtos e roubos foram adotados, uma vez que essa é a abordagem utilizada na literatura, conforme observado nos estudos de Khalid et al. (2018), Quick, Li e Brunton-Smith (2018), Andresen, Curman e Linning (2017), LaRue e Andresen (2015), Ceccato e Moreira (2021), e Rodrigues (2023). Isso possibilitou uma comparação direta entre os trabalhos.

Para melhor entendimento das análises realizadas, primeiro são apresentadas as categorizações das variáveis explicativas e de resposta nas tabelas 8 e 9. Destaca-se que a categorização consiste em atribuir rótulos (categorias) aos intervalos de valores definidos na espacialização de dados, e apresentados nos mapas das figuras 5 a 9 (variáveis explicativas) e figuras 15 a 17 (variáveis de resposta).

Tabela 8 – Categorização das variáveis explicativas

Variáveis explicativas	Categorias/faixas de valores				
	Muito baixa (MB)	Baixa (B)	Média (M)	Alta (A)	Muito alta (MA)
Renda Média (REM) em R\$	518,00 a 892,00	892,00 a 1.070,00	1.070,00 a 1.183,00	1.183,00 a 1.337,00	1.337,00 a 2.980,00
Percentual População Jovem (PPJ)	0 a 34,4%	34,4 a 36,3%	36,3 a 37,7%	37,7 a 39,3%	39,3 a 64,8%
Densidade Populacional (DEP)	1,6 a 3,14	3,14 a 3,29	3,29 a 3,42	3,42 a 3,6	3,6 a 5,35
Percentual População Masculina (PPM)	0 a 45,96%	45,96 a 47,29%	47,29 a 48,25%	48,25 a 49,25%	49,25 a 63,33%
Percentual População Alfabetizada (PPA)	62 a 86,3%	86,3 a 88,4%	88,4 a 90,2%	90,2 a 91,5%	91,5 a 100%

Fonte: elaboração própria

Tabela 9 – Categorização variáveis de resposta

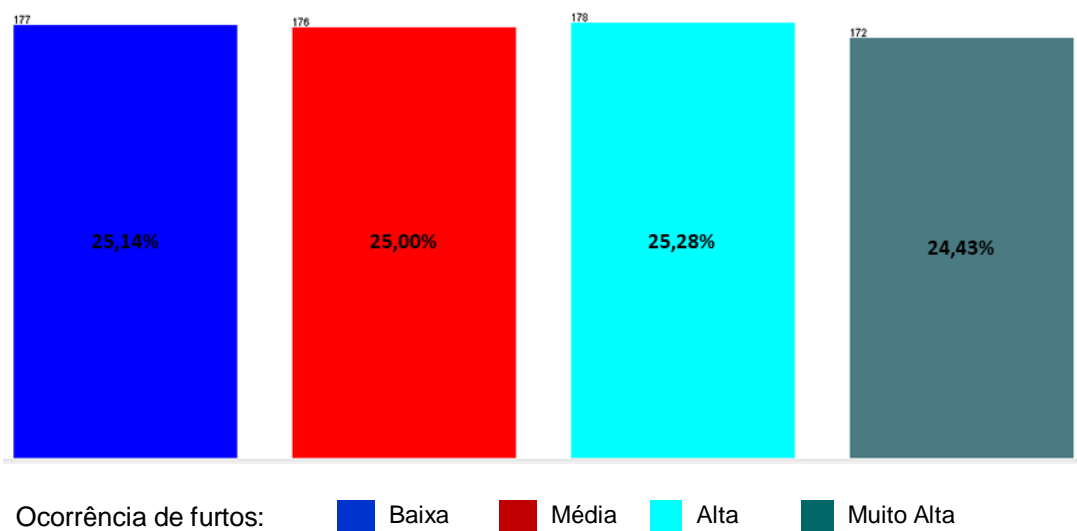
Variáveis de resposta	Categorias/faixas de valores			
	Baixa (B)	Média (M)	Alta (A)	Muito Alta (MA)
Furto	0 a 8	8 a 17	17 a 30	30 a 459
Roubo	0 a 13	13 a 25	25 a 40	40 a 398
Crimes (furto + roubo)	1 a 23	23 a 44	44 a 68	68 a 857

Fonte: elaboração própria

Os dados, após serem categorizados com auxílio do Excel, foram transferidos para o WEKA, no qual foi conduzida a análise das variáveis explicativas em função das variáveis de resposta (furtos, roubos, e crimes = furto + roubo), sendo os

resultados apresentados nas figuras 18 a 29. É importante destacar que essas análises são superficiais e foram baseadas nos dados criminais acumulados no período considerado nesta pesquisa (2017 a 2021). As análises mais aprofundadas, realizadas por meio da mineração de dados, estão apresentadas na próxima seção.

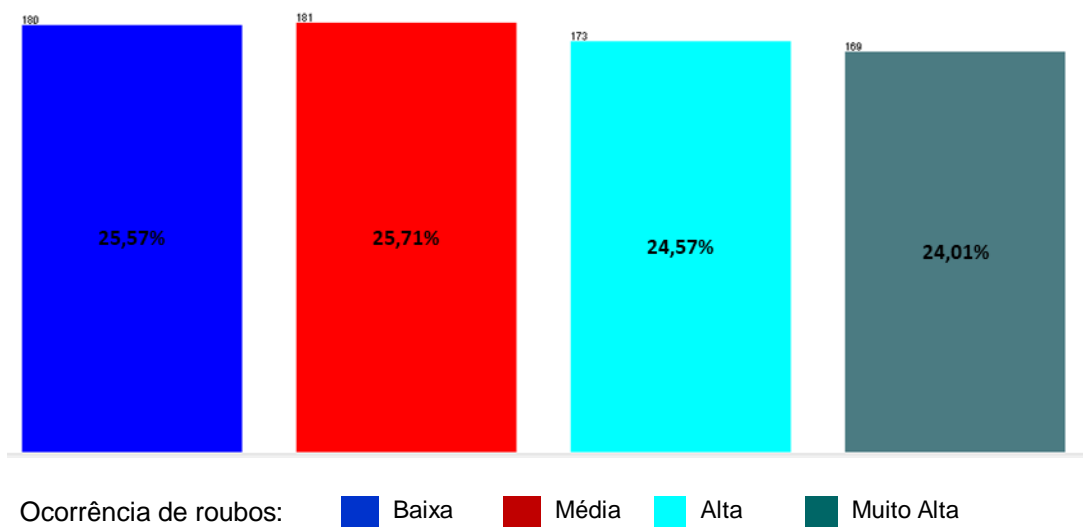
Figura 18 – Distribuição dos furtos nos setores censitários



Por meio da figura 18, pode-se observar que, para a variável de resposta furto, 25,14% dos setores censitários apresentam baixa ocorrência; aqueles com número médio de ocorrências totalizam 25,00%; com alta ocorrência são 25,28% dos setores; e 24,43% dos setores apresentam números muito alto de ocorrências.

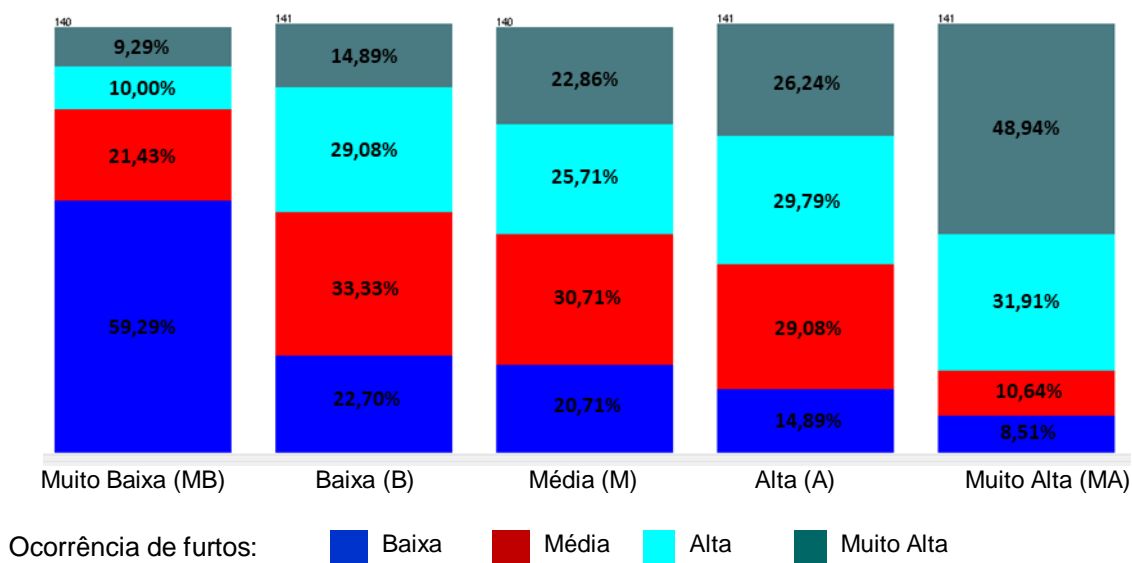
Já para a variável de resposta roubo, os setores censitários com baixa ocorrência representam 25,57%; com média ocorrência, 25,71%; com alta ocorrência, 24,57%; e com ocorrência muito alta, 24,01%, conforme figura 19.

Figura 19 – Distribuição dos roubos nos setores censitários



Os gráficos das figuras 18 a 29 apresentam os resultados das análises das variáveis explicativas em função das ocorrências de furto e roubo. A figura 20, por exemplo, mostra que do total de setores censitários com renda média (REM) muito baixa, 59,29% apresentam número baixo de ocorrências de furto; 21,43% apresentam número médio de ocorrências de furto; 10,00% apresentam alto número de ocorrências de furto; 9,29% apresentam um número muito alto de ocorrências de furto.

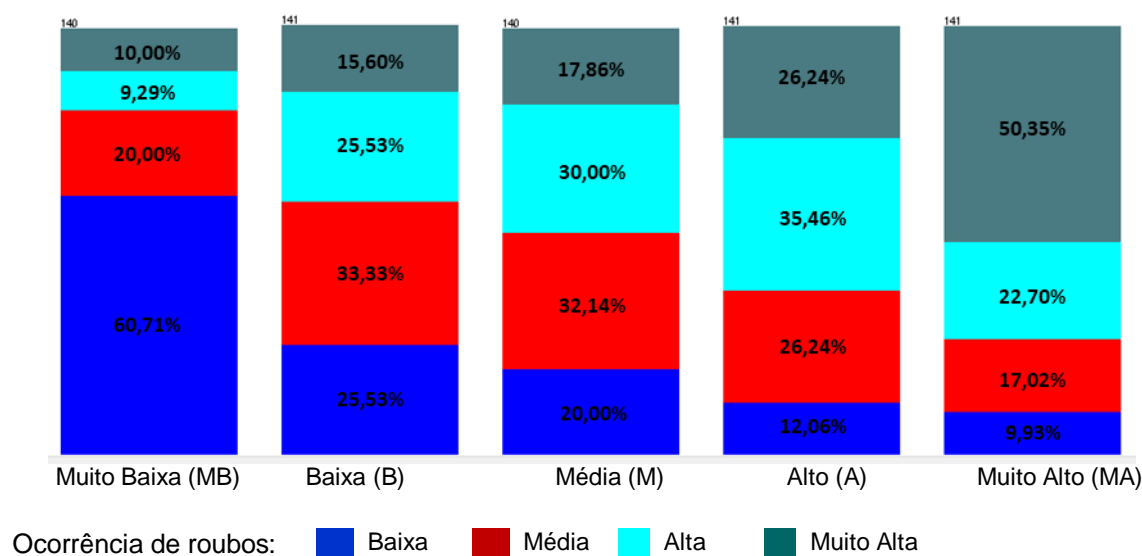
Figura 20 – Furtos ocorridos nos setores censitários com relação à variável REM



Ainda em relação à figura 20, pode-se observar que conforme aumenta a renda, há aumento da representatividade da classe Muito Alta das ocorrências de furtos (de 9,29 para 48,94%), ao passo que diminui gradualmente a representatividade da classe Baixa das mesmas ocorrências (de 59,29 para 8,51%).

Assim na análise da variável de renda média (REM), nota-se uma relação direta entre furto e renda. Renda média muito alta está relacionada com números muito altos de ocorrências de furto, enquanto renda média muito baixa está relacionada com números baixos de ocorrências de furto. Isso é indicado no gráfico da Figura 20 que mostra que 59,29% dos setores censitários com REM muito baixa apresentam baixa ocorrência de furtos, e que 48,94% dos setores censitários com REM muito alta apresentam número alto de ocorrência de furtos. A figura 21 apresenta o gráfico da variação das classes de roubo em função das classes de renda média.

Figura 21 – Roubos ocorridos nos setores censitários com relação à variável REM

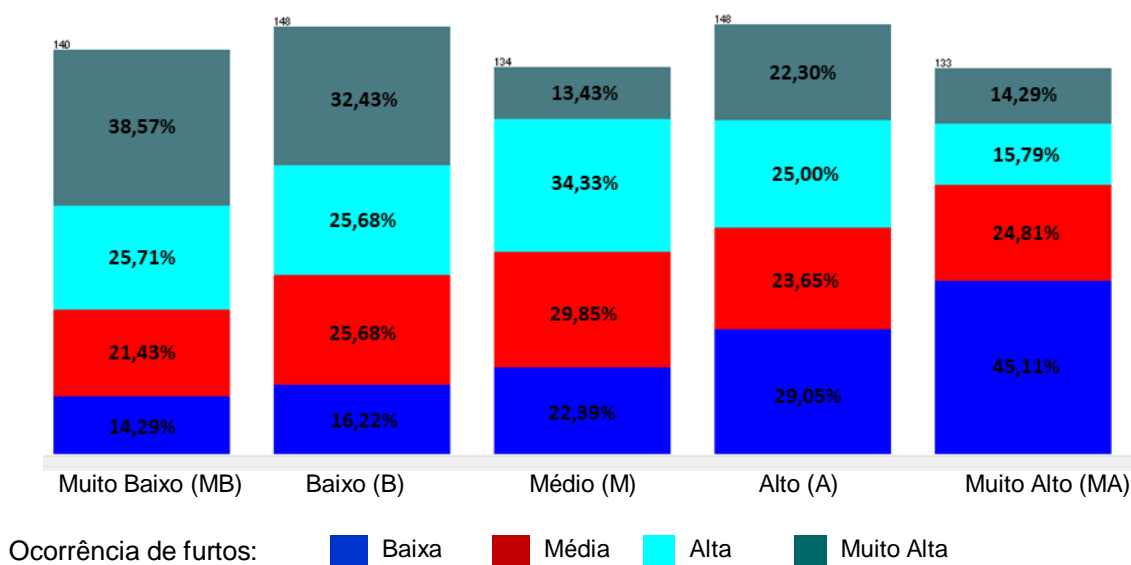


Com base na figura 21, observa-se padrão semelhante ao apontado na figura 20. Nos setores censitários com REM muito baixa, 60,71% apresentam números baixos de roubos. Por outro lado, em 50,35% dos setores com REM muito alta, as ocorrências de roubos também são muito altas.

Observa-se ainda que o aumento da renda é acompanhado pelo incremento da representatividade da classe Muito Alta das ocorrências de roubos (de 10,00 para 50,35%), ao passo que diminui gradualmente a representatividade da classe Baixa das mesmas ocorrências (de 60,71 para 9,93%).

A figura 22 permite a observação da variável percentual de população jovem (PPJ), identificada pela literatura como uma variável explicativa relevante para a incidência de crimes contra o patrimônio, conforme destacado nos estudos de Baumer e Arnio (2012), Ceccato e Moreira (2021) e Adewuyi et al. (2017).

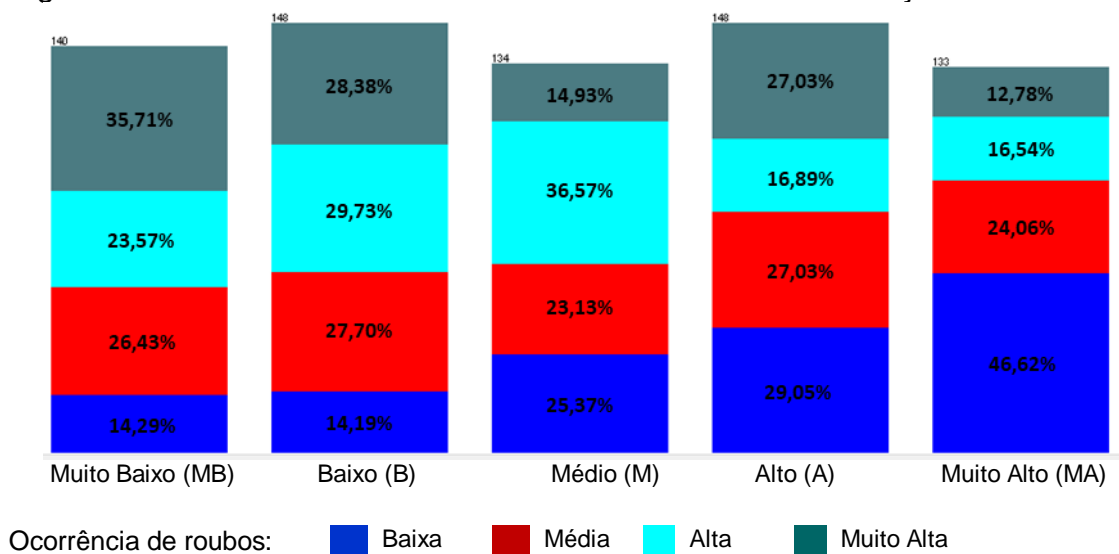
Figura 22 – Furtos ocorridos nos setores censitários com relação à variável PPJ



Por meio da figura 22, observa-se que 38,57% dos setores censitários com PPJ muito baixo apresentam altas ocorrência de furtos, e que 45,11% dos setores censitários com PPJ muito alta apresentam números baixos de ocorrência de furtos.

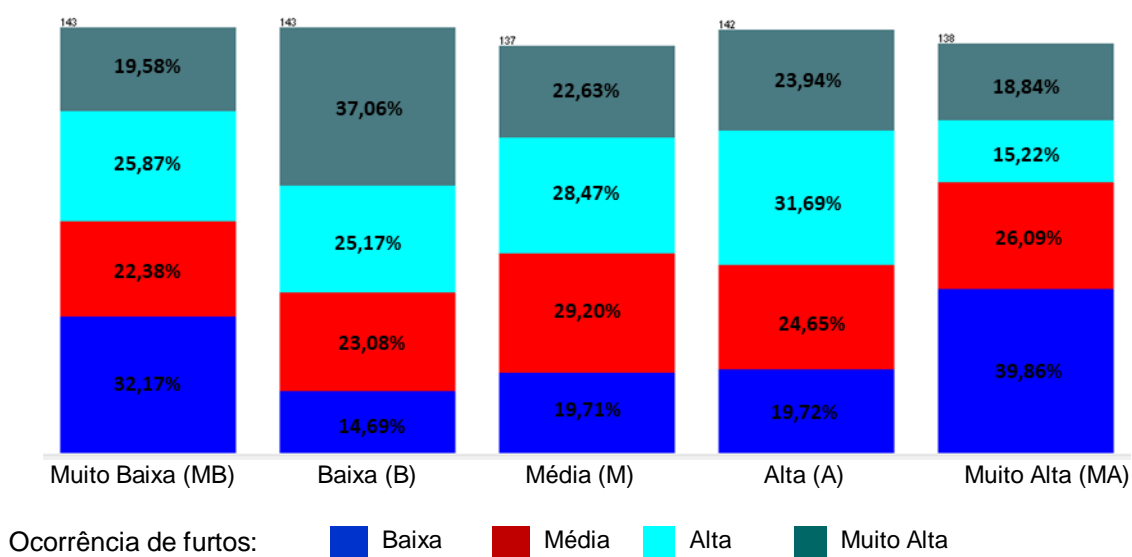
No contexto das ocorrências de roubo, observa-se padrão semelhante, segundo o qual, em 35,71% dos setores censitários com PPJ muito baixo, os números de roubos são muito altos. Por outro lado, em 46,62% dos setores com PPJ muito alto, as ocorrências de roubos são baixas, conforme figura 23.

Figura 23 – Roubos ocorridos nos setores censitários com relação à variável PPJ



Na análise da variável de densidade populacional (DEP) nota-se uma relação direta entre furto e densidade populacional. Densidade populacional muito alta está relacionada com números baixos de ocorrências de furto, enquanto densidade populacional baixa está relacionada com números muito altos de ocorrências de furto, conforme figura 24.

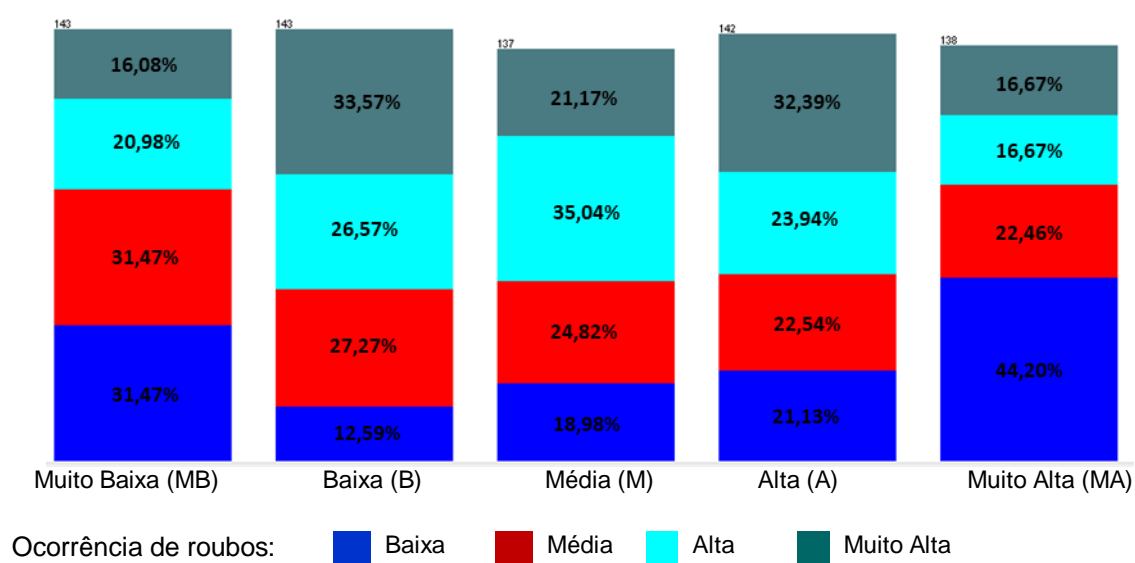
Figura 24 – Furtos ocorridos nos setores censitários com relação à variável DEP



A Figura 24 ilustra que 39,86% dos setores censitários com DEP muito alta apresentam baixas ocorrência de furtos, e que 37,06% dos setores censitários com DEP baixa apresentam números muito altos de ocorrência de furtos.

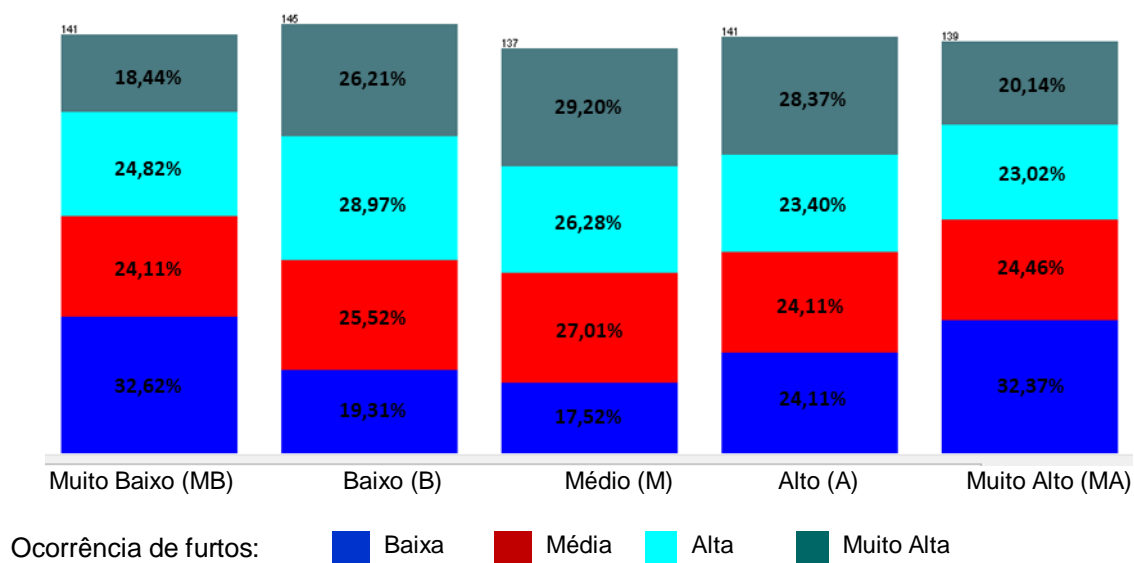
No contexto de roubo, observou-se comportamento semelhante, onde 33,57% dos setores censitários com DEP baixa corresponderam à classe de roubos Muito Alta. Por outro lado, em 44,20% dos setores com DEP muito alta as ocorrências de roubos são baixas.

Figura 25 – Roubos ocorridos nos setores censitários com relação à variável DEP



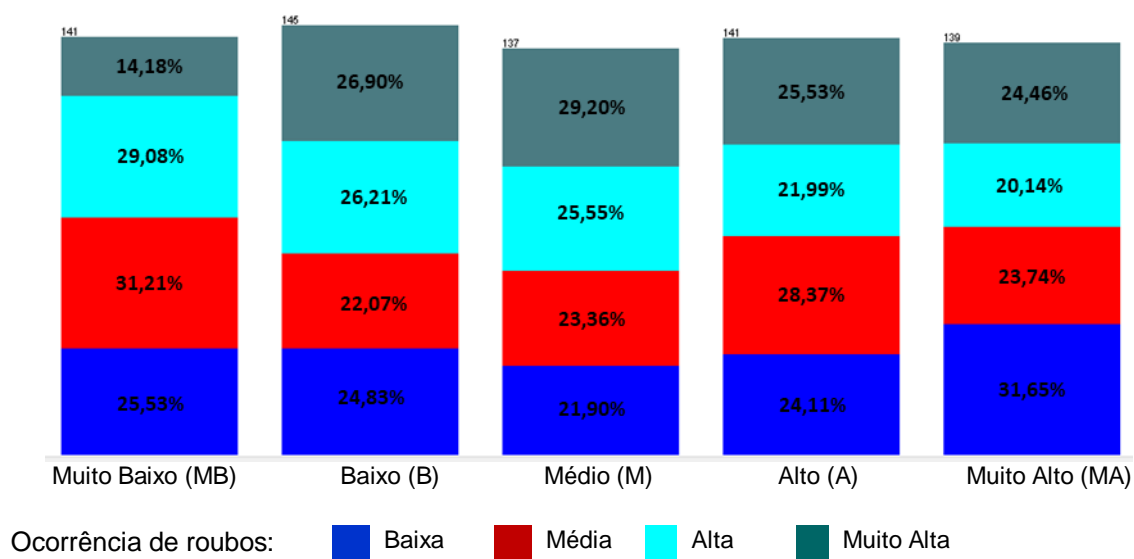
Na análise da variável de percentual de população masculina (PPM) no gráfico da Figura 26, observa-se que 32,62% dos setores censitários com um percentual muito baixo de população masculina apresentam uma baixa ocorrência de furtos, enquanto 32,37% dos setores censitários com um percentual muito alto de população masculina mostram um baixo número de ocorrências de furtos.

Figura 26 – Furtos ocorridos nos setores censitários com relação à variável PPM



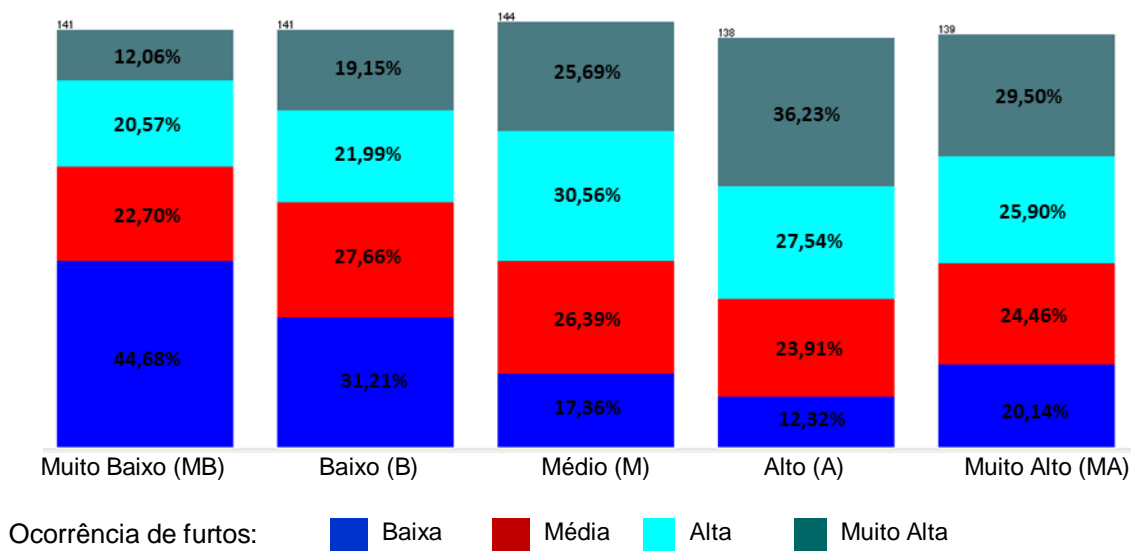
No contexto de roubos, considerando a variável explicativa Percentual da População Jovem, observa-se que 31,21% dos setores censitários com um percentual muito baixo de população masculina apresentam números médios de ocorrências de roubos. Por outro lado, em 31,65% dos setores com um percentual muito alto de população masculina, as ocorrências de roubos são baixas, conforme ilustrado na Figura 27.

Figura 27 – Roubos ocorridos nos setores censitários com relação à variável PPM



A análise permitiu identificar relações entre a variável Percentual de População Alfabetizada (PPA) com a ocorrência de furtos, conforme Figura 28.

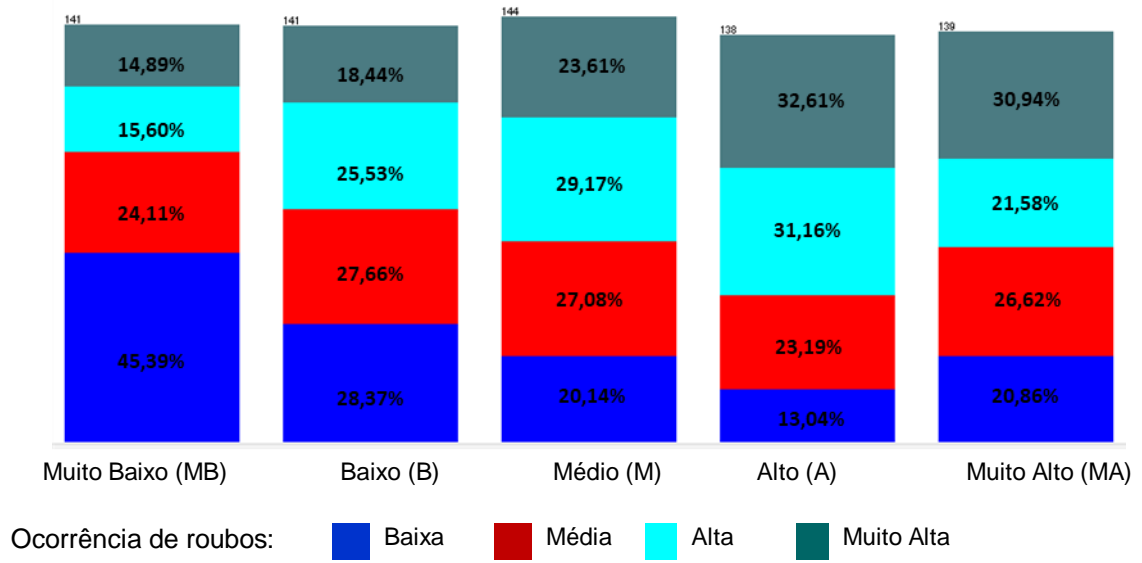
Figura 28 – Furtos ocorridos nos setores censitários com relação à variável PPA



Conforme pode ser observado por meio da Figura 28, a classe dos setores censitários correspondente ao PPA Muito Baixo apresentou os mais baixos percentuais de ocorrência de furtos (44,68%). Por outro lado, a classe de ocorrência de furtos Muito Alta teve maior expressão nos setores que apresentaram Alto e Muito Alto PPA, correspondendo aos valores 36,26% e 29,50%, respectivamente.

No contexto dos roubos, observou-se tendência semelhante, em que 45,39% dos setores censitários com PPA Muito Baixo apresentaram números de roubos também baixos. Por outro lado, em 30,94% dos setores com PPA Muito Alto, as ocorrências de roubos são Muito Altas. Além disso, conforme poderá ser observado, há um incremento gradual no número de ocorrências de roubos na medida em que se desloca da coluna referente à Classe Muito Baixo PPA em direção à da Classe Muito Alto PPA, conforme gráfico da Figura 29.

Figura 29 – Roubos ocorridos nos setores censitários com relação à variável PPA



4.3 Descoberta de padrões usando MD (Árvores de Decisão)

Essa seção contempla os resultados da mineração de dados (MD), com auxílio do WEKA, visando a descoberta de padrões que explicam as relações entre as variáveis explicativas e variáveis de resposta (seção 4.3.1), bem como padrões de crimes com base no perfil das vítimas (seção 4.3.2.). A MD foi realizada empregando-se a técnica de Árvore de Decisão (AD), usando o algoritmo C4.5 (J48 no WEKA), para classificação dos dados com intuito de investigar a existência de padrões.

Cada AD é dividida em diferentes ramificações com base nos valores dos atributos. Cada ramificação descreve um padrão por meio de uma regra do tipo **SE...ENTÃO**. Por exemplo, na árvore da Figura 36, a regra que descreve a primeira ramificação mais à esquerda é: **SE** renda média (REM) = “MB (muito baixa)” **ENTÃO** furto = “B_M (baixo ou médio)”.

O desempenho da AD é medido com base no número de instâncias classificadas corretamente (taxa de acerto/Recall), e no índice Kappa. Quanto maior forem essas medidas, mais consistentes são os padrões descobertos.

4.3.1 Análise das relações entre as variáveis explicativas e de resposta

Nesta seção são descritos os resultados das análises para atender o objetivo específico 1.4.3, a saber: “Analisar as relações entre as variáveis sociodemográficas e os padrões espaciais de distribuição dos casos de furto e roubo na subprefeitura de Itaquera”.

Após análise inicial no WEKA, pode-se observar que as categorias Baixa (B) e Média (M) e Alta (A) e Muito Alta (MA), associadas às variáveis de resposta, poderiam ser agrupadas para gerar melhores resultados. Assim, foi feito o agrupamento dando origem a duas categorias: baixo ou médio (B_M), alto ou muito alto (A_MA).

Foram criadas três ADs, sendo uma para furto, uma para roubo e a outra considerando crimes (furto + roubo). Para cada AD foram elencadas as regras (novos conhecimentos) com maiores graus de confiança (apresentados entre parênteses na sequência das regras).

a. Furto

A AD gerada para a análise das relações entre as variáveis explicativas e a variável de resposta furto está apresentada na Figura 30. Na classificação das 703 instâncias (dados dos 703 setores censitários), 509 foram corretamente classificadas, resultando em uma taxa de acerto (Recall/Sensibilidade) de 72,4%, precisão de 72,4% e concordância Kappa Moderada (44,81%). A matriz de confusão correspondente é apresentada na Tabela 10.

Figura 30 – AD furto

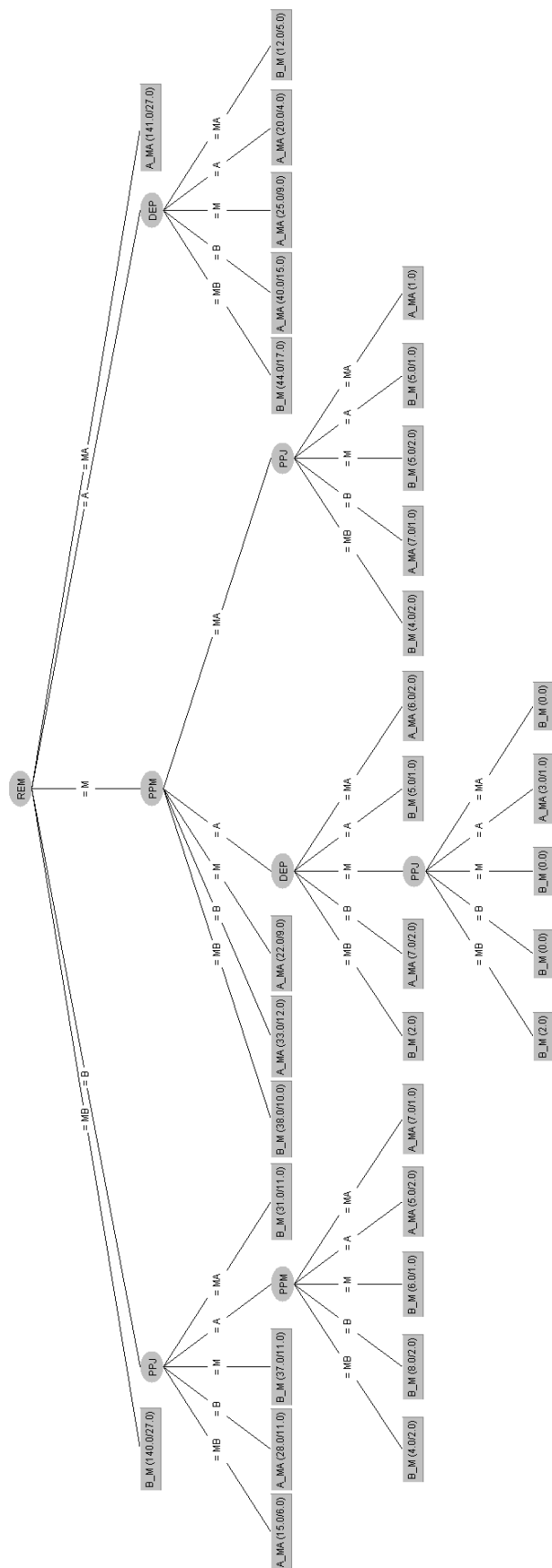


Tabela 10 – Matriz de confusão refletindo os resultados da AD para a variável “Furto”

Classe real		Classe predita
a	b	
251	102	a = B_M
92	258	b = A_MA

Fonte: elaboração própria

Ainda com base na mesma AD, podem-se derivar os seguintes conhecimentos sobre as relações das variáveis sociodemográficas com furto:

- 1) Setores censitários com valores médios de renda média, percentual de população masculina muito alto e percentual de população jovem baixo possuem ocorrências de furtos altas ou muito altas (A_MA) (6 de 7 casos possíveis = 90,00%).
- 2) Setores censitários com renda média muito alta possuem ocorrências de furtos altas ou muito altas (A_MA) (114 de 141 casos possíveis = 80,85%).
- 3) Setores censitários com renda média muito baixa têm ocorrências de furtos baixas ou médias (B_M) (113 de 140 casos possíveis = 80,71%).
- 4) Setores censitários com valores médios de renda média, percentual de população masculina muito alto e percentual de população jovem alto possuem ocorrências de furtos baixas ou médias (B_M) (4 de 5 casos possíveis = 80%).
- 5) Setores censitários com renda média alta e densidade populacional alta têm ocorrências de furtos altas ou muito altas (A_MA) (16 de 20 casos possíveis = 80,00%).
- 6) Setores censitários com renda média baixa, percentual de população jovem alto e percentual de população masculina muito alto possuem ocorrências de furtos altas ou muito altas (A_MA) (6 de 7 casos possíveis = 85,71%).
- 7) Setores censitários com valores médios de renda média e percentual da população masculina muito baixo possuem ocorrências de furtos baixas ou médias (B_M) (28 de 38 casos possíveis = 73,68%).
- 8) Setores censitários com renda média baixa e percentual de população jovem médio têm ocorrências de furtos baixas ou médias (B_M) (26 de 37 casos possíveis = 70,27%).
- 9) Setores censitários com valores médios de renda média, percentual de população masculina alto, densidade populacional média e percentual de população jovem

alto têm ocorrências de furtos altas ou muito altas (A_MA) (2 de 3 casos possíveis = 66,67%).

10) Setores censitários com valores médios de renda média, percentual de população masculina alto, densidade populacional média e percentual de população jovem muito baixo possuem ocorrências de furtos baixas ou médias (B_M) (2 de 2 casos possíveis = 100%).

É importante mencionar que se deve ter cuidado em afirmar que uma regra é boa apenas pelo percentual apresentado. Na última regra (regra 10), por exemplo, as premissas valor médio de renda média, percentual de população masculina alto, densidade populacional média e percentual de população jovem muito baixo aparecem juntas somente em dois casos possíveis. Assim, confiança de 100% está baseada apenas nesses dois casos.

b. Roubo

A AD gerada para a análise do roubo está apresentada na Figura 31. Na classificação das 703 instâncias de treinamento, 506 foram corretamente classificadas, resultando em uma taxa de acerto de 72,0%, precisão de 72,2% e concordância Kappa Moderada (43,74%). A matriz de confusão correspondente é apresentada na Tabela 11.

Figura 31 – AD Roubo

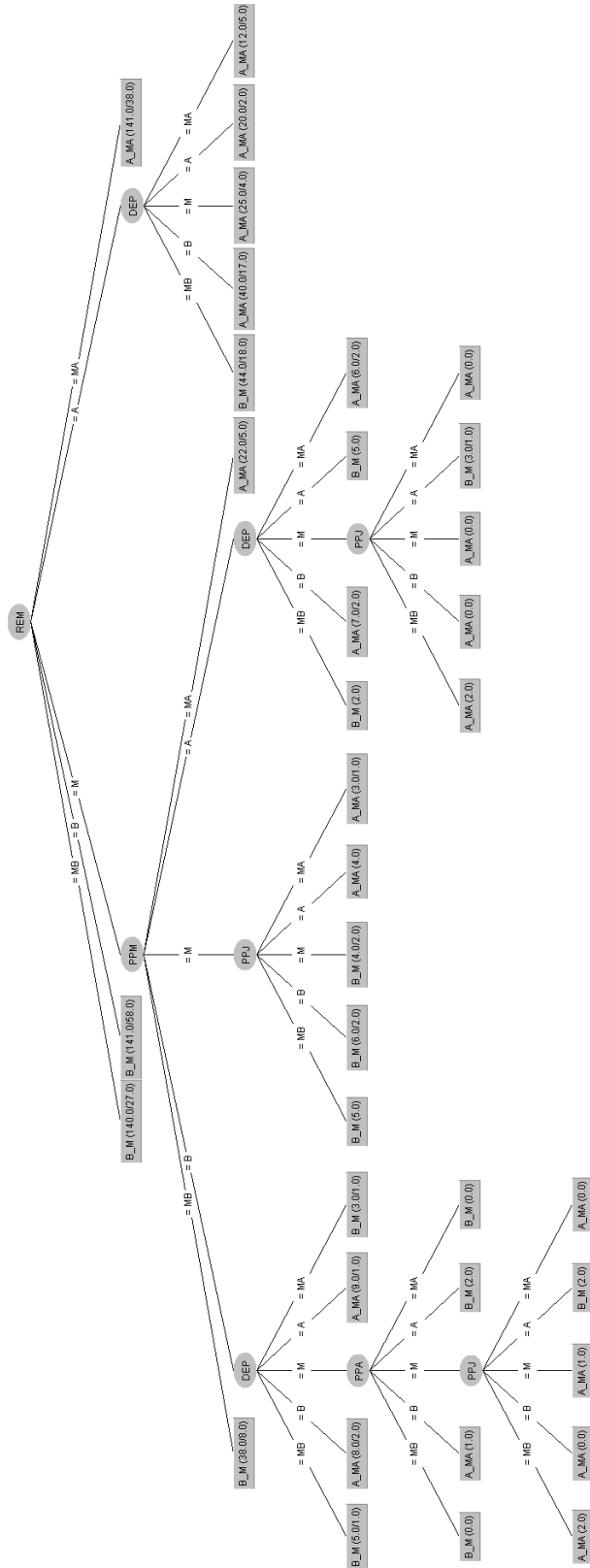


Tabela 11 – Matriz de confusão refletindo os resultados da AD para a variável “Roubo”

Classe real		Classe predita
a	b	
282	79	a = B_M
118	224	b = A_MA

Fonte: elaboração própria

A partir da AD apresentada na Figura 32, podem-se derivar os seguintes conhecimentos sobre as relações das variáveis sociodemográficas com roubo:

- 1) Setores censitários com valores médios de renda média, em combinação com percentual de população masculina baixo e densidade populacional alta, possuem ocorrências de roubos altas ou muito altas (A_MA) (8 de 9 casos possíveis = 90%).
- 2) Setores censitários com renda média muito baixa possuem ocorrências de roubos baixas ou médias (B_M) (113 de 140 casos possíveis = 80,71%).
- 3) Setores censitários com valores médios de renda média, aliados a um percentual de população masculina baixo e densidade populacional muito baixa, têm ocorrências de roubos baixas ou médias (B_M) (4 de 5 casos possíveis = 80,00%).
- 4) Setores censitários com valores médios de renda média e percentual de população masculina muito alto têm ocorrências de roubos altas ou muito altas (A_MA) (17 de 22 casos possíveis = 77,27%).
- 5) Setores censitários com valores médios de renda média e percentual de população masculina muito baixo têm ocorrências de roubos baixas ou médias (B_M) (30 de 38 casos possíveis = 78,94%).
- 6) Setores censitários com valores médios de renda média, aliados a um percentual de população masculina baixo e densidade populacional baixa, possuem ocorrências de roubos altas ou muito altas (A_MA) (6 de 8 casos possíveis = 75,00%).
- 7) Setores censitários com valores médios de renda média, aliados a um percentual de população masculina alto e densidade populacional baixa, possuem ocorrências de roubos altas ou muito altas (A_MA) (5 de 7 casos possíveis = 71,42%).

- 8) Setores censitários com valores médios de renda média, percentual de população masculina médio e percentual de população jovem baixo, têm ocorrências de roubos baixas ou médias (B_M) (4 de 6 casos possíveis = 66,66%).
- 9) Setores censitários com valores médios de renda média, aliados a um percentual de população masculina baixo e densidade populacional muito alta, têm ocorrências de roubos baixas ou médias (B_M) (2 de 3 casos possíveis = 66,66%).
- 10) Setores censitários com valores médios de renda média, em combinação com percentual de população masculina baixo, densidade populacional média e percentual de população alfabetizada entre muito baixo e médio, possuem ocorrências de roubos altas ou muito altas (A_MA) (2 de 2 casos possíveis = 100%).

c. Crimes

A AD gerada para a análise dos crimes está apresentada na Figura 32. Na classificação das 703 instâncias de treinamento, 512 foram corretamente classificadas, resultando em uma taxa de acerto de 72,8%, precisão de 72,8% e concordância Kappa Moderada (45,59%). A matriz de confusão de onde se originam tais medidas é apresentada na Tabela 12.

Figura 32 – AD Crimes

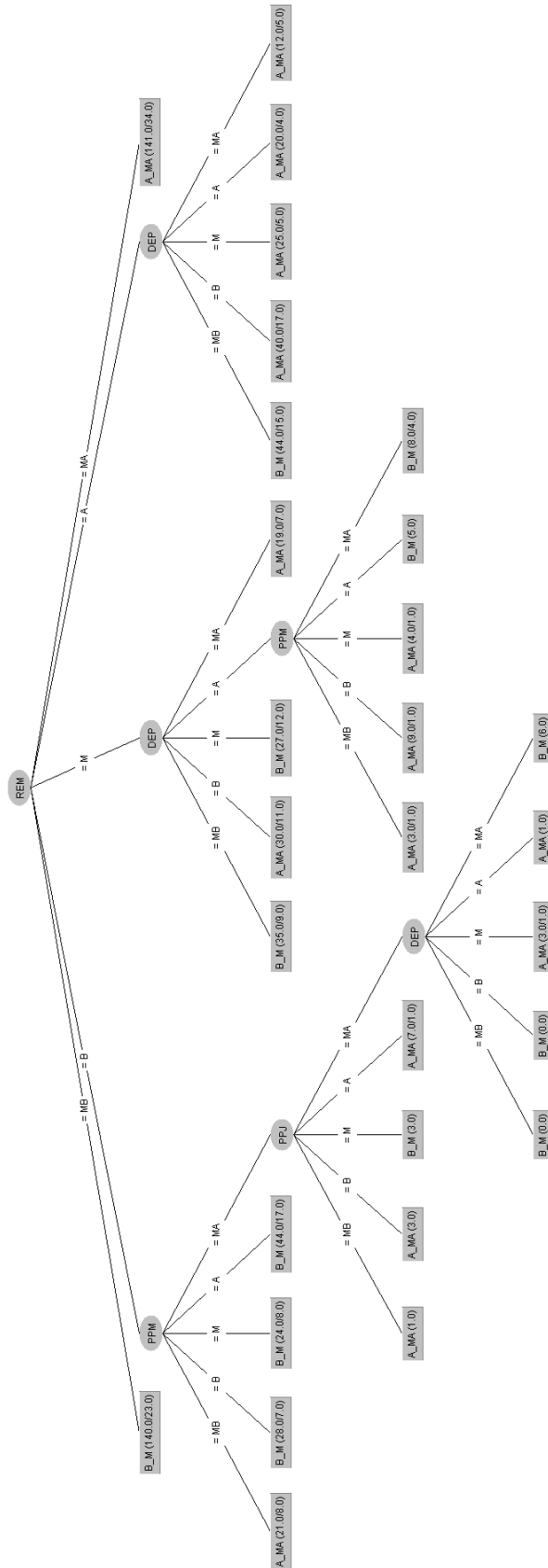


Tabela 12 – Matriz de confusão refletindo os resultados da AD para a variável “Crimes”

Classe real		Classe predita
a	b	
269	96	a = B_M
95	243	b = A_MA

Fonte: elaboração própria

A partir da AD mostrada, pode-se derivar os seguintes conhecimentos sobre as relações das variáveis sociodemográficas com crimes:

- 1) Setores censitários com valores médios de renda média, densidade populacional alta e percentual de população masculina baixo possuem ocorrências de crimes alta ou muito alta (A_MA) (8 de 9 casos possíveis = 88,88%).
- 2) Setores censitários com renda média baixa, percentual de população masculina muito alto e percentual de população jovem alto possuem ocorrências de crimes alta ou muito alta (A_MA) (6 de 7 casos possíveis = 85,71%).
- 3) Setores censitários com renda média muito baixa possuem ocorrências de crimes baixas ou médias (B_M) (117 de 140 casos possíveis = 83,57%).
- 4) Setores censitários com renda média alta e densidade populacional média possuem ocorrências de crimes alta ou muito alta (A_MA) (20 de 25 casos possíveis = 80,00%).
- 5) Setores censitários com valores médios de renda média, densidade populacional alta e percentual de população masculina médio possuem ocorrências de crimes baixas ou médias (B_M) (3 de 4 casos possíveis = 75,00%).
- 6) Setores censitários com valores médios de renda média, densidade populacional alta e percentual de população masculina alto possuem ocorrências de crimes baixas ou médias (B_M) (3 de 4 casos possíveis = 75,00%).
- 7) Setores censitários com renda média muito alta possuem ocorrências de crimes alta ou muito alta (A_MA) (107 de 141 casos possíveis = 75,88%).
- 8) Setores censitários com renda média baixa, percentual de população masculina baixo possuem ocorrências de crimes baixas ou médias (B_M) (21 de 28 casos possíveis = 75,00%).

- 9) Setores censitários com valores médios de renda média e densidade populacional muito baixa, têm ocorrências de crimes baixas ou médias (B_M) (26 de 35 casos possíveis = 74,28%).
- 10) Setores censitários com renda média baixa, percentual de população masculina muito alto, percentual de população jovem muito alto e densidade populacional média possuem ocorrências de crimes alta ou muito alta (A_MA) (2 de 3 casos possíveis = 66,66%).
- 11) Setores censitários com valores médios de renda média e densidade populacional muito alta possuem ocorrências de crimes altas ou muito altas (A_MA) (12 de 19 casos possíveis = 63,15%).
- 12) Setores censitários com valores médios de renda média e densidade populacional baixa, têm ocorrências de crimes alta ou muito alta (A_MA) (19 de 30 casos possíveis = 63,33%).

4.3.2 Análise de padrões dos crimes com base no perfil das vítimas

Nesta seção são apresentados os resultados dos experimentos envolvendo a aplicação de AD na tentativa de identificar padrões das ocorrências de crimes com base nas características pessoais das vítimas (sexo, faixa etária, cor, grau de instrução e naturalidade) e no horário de ocorrência dos crimes.

Para conduzir esses experimentos, que visam atender o segundo objetivo específico (subseção 1.4.2), foram empregados os dados de 40.552 boletins de ocorrência (BO) registrados no período de 01 de janeiro de 2017 a 31 de dezembro de 2021. Contudo, após a aplicação da etapa do KDD, pré-processamento (exclusão de registros com dados faltantes), restaram apenas 6.568 registros (16,2% dos BOs). Na etapa seguinte (transformação) foi feita a categorização dos dados conforme apresentado na tabela 13.

Tabela 13 – Dados Extraídos dos Boletim de Ocorrência

Dados	Categorias
Sexo	{(M)asculino), (F)eminino}
Faixa etária	{Até 17 anos, 18 a 30 anos, 31 a 60 anos, Acima 60 anos}
Cor e etnia	{Amarela, Branca, Parda, Preta}
Grau de instrução	{Analfabeto, 1º grau, 2º grau, Superior}
Naturalidade	{São Paulo – SP, Outra cidade}
Período	{Manhã, Tarde, Noite, Madrugada}

Fonte: elaboração própria

A partir da AD mostrada na figura 33, pode-se verificar existência de padrões relacionados ao Roubo, o que não aconteceu para os crimes de Furto. Isso pode estar relacionado ao fato de o criminoso escolher a vítima sobre a qual ele exercerá violência ou grave ameaça para subtrair-lhe algo, o que não acontece nos casos de furto. As taxas de acerto para roubo e furto foram, respectivamente, 91,7% e 24,7%. Quanto maior a taxa de acertos de uma classe, mais consistentes são os padrões descritos pelas AD para essa classe.

Um aspecto adicional a ser enfatizado é a quantidade de registros eliminados durante o pré-processamento devido à ausência de dados. Se a quantidade de dados disponíveis para mineração fosse maior, é possível que padrões relacionados a furtos também pudessem ser descobertos. Isso levanta a questão da necessidade de melhorar o preenchimento dos Boletins de Ocorrência (BOs), tanto em relação à integridade quanto à precisão e uniformidade dos dados. Esse cenário ressalta a importância de uma coleta mais abrangente e precisa de informações nos BOs, o que poderia potencialmente levar a descobertas mais amplas e representativas por meio da mineração de dados.

4.4 Discussão dos resultados

A espacialização dos dados relacionados a furto, roubo e variáveis sociodemográficas serviu como uma ferramenta essencial para a compreensão inicial da dinâmica dos referidos crimes na subprefeitura de Itaquera. O mapeamento das localizações e da distribuição espacial das ocorrências dos crimes permitiu uma prévia visualização de alguns padrões e tendências que serviram como apoio para as análises mais aprofundadas (análise das variáveis e mineração de dados – MD). Além disso, o mapeamento da distribuição espacial revelou áreas de maior concentração de crimes, o que pode servir como subsídio para direcionar esforços preventivos e alocação de recursos de segurança de maneira mais eficiente, nas regiões de maior risco.

Não obstante, a espacialização dos dados realizada neste trabalho propiciou a categorização dos dados, possibilitando as análises das suas distribuições em relação às variáveis explicativas e a aplicação das árvores de decisão (ADs) na etapa de MD. A categorização consiste basicamente em alocar etiquetas ou categorias específicas

(ex. “baixa”, “média”, “alta”) para os intervalos de valores das variáveis explicativas e de respostas, e foi conduzida por meio do software Excel. Posteriormente os dados categorizados foram exportados para o WEKA, no qual foi realizada a MD.

A análise das distribuições dos dados nas variáveis explicativas e de respostas consistiu em uma análise visual quantitativa que só permite identificar as relações uma-para-uma entre as variáveis explicativas e de resposta. Em outras palavras, ela permitiu observar a influência de cada variável sociodemográfica em cada uma das variáveis de resposta (furto e roubo). Por exemplo, a partir desta análise já foi possível evidenciar a influência da renda média nas ocorrências de furtos e roubos.

A etapa de MD foi além da visualização de dados e da análise das distribuições dos dados nas variáveis explicativas e de respostas, permitindo encontrar padrões contextualizados que explicam as relações entre as variáveis sociodemográficas e as ocorrências de furtos e roubos. Em resumo, a MD permitiu identificar como uma ou mais variáveis explicativas podem influenciar nos índices de criminalidade. Assim, esta análise não só proporcionou uma visão mais estruturada dos dados, como também estabeleceu um alicerce mais robusto para estabelecer comparações dos achados desta pesquisa com resultados da literatura.

Os principais achados acerca da dinâmica da criminalidade e suas relações com as variáveis sociodemográficas consideradas na presente pesquisa estão elencados na Tabela 14. Cabe destacar que há padrões que divergem dos relatados em estudos da literatura, como pode ser observado na terceira coluna da referida Tabela.

Tabela 14 – Padrões importantes oriundos das análises usando ADs

Variáveis explicativas	Padrões observados	Comparação com estudos da literatura (concordam/discordam)
REM	<ul style="list-style-type: none"> • A <u>renda média</u> dos setores censitários é um fator determinante nas ocorrências de crimes. • Setores com <u>renda média</u> muito baixa mostram predominantemente ocorrências de crimes baixas e médias. • Setores com <u>renda média</u> alta, particularmente quando associados a alta <u>densidade populacional</u>, apresentam predominantemente ocorrências de crimes altas ou muito altas. 	<p>De acordo com Adewuyi et al. (2017) furtos e roubos à mão armada estão positivamente correlacionados com a renda nacional bruta.</p> <p>O modelo de regressão de Rodrigues (2023) mostrou uma relação muito baixa entre a renda e a ocorrência de crimes de furto e roubo. Mesmo considerando isso, o coeficiente da regressão foi positivo, mostrando que a ocorrência de crimes tende a aumentar com o aumento da renda.</p>
PPJ	<ul style="list-style-type: none"> • Setores com alto <u>percentual de população jovem</u>, particularmente quando combinados com <u>renda média</u> baixa e alto <u>percentual de população masculina</u>, tendem a exibir altas ou muito altas ocorrências de crime. 	<p>Adewuyi et al. (2017) tiveram como resultados que furtos e roubos à mão armada estão positivamente correlacionados com a população masculina jovem e associaram a presença de uma população jovem a índices mais altos de criminalidade.</p>
DEP	<ul style="list-style-type: none"> • Setores com valores médios de <u>renda média</u> e diferentes níveis de <u>densidade populacional</u>, especialmente alta, apresentam uma variação nas taxas de crimes, mas inclinam-se majoritariamente para ocorrências baixas e médias. • A <u>densidade populacional</u> em associação com outras variáveis desempenha um papel importante nas correlações. 	<p>Clancy et al. (2022) em Chicago, os crimes violentos estão associados à população total.</p> <p>Adewuyi et al. (2017) e Rodrigues (2023) enquanto o primeiro sugere uma relação direta entre maior densidade populacional e criminalidade, o segundo observou que setores com mais moradores concentravam a maioria dos crimes.</p>
PPM	<ul style="list-style-type: none"> • <u>Percentual de população masculina</u> quando combinado com outras variáveis, como <u>renda média</u> e <u>densidade populacional</u>, permite evidenciar o impacto nas ocorrências de crimes. • O <u>percentual de população masculina</u> parece ser um indicador significativo nas correlações, pois aparece com frequência, tanto nos altos quanto nos baixos números de ocorrências de crimes. 	<p>Adewuyi et al. (2017) que tiveram como resultados que furtos e roubos à mão armada estão positivamente correlacionados com a população masculina jovem.</p>
PPA	<ul style="list-style-type: none"> • <u>Percentual de população alfabetizada</u> não foi tão frequentemente citada, mas onde aparece, setores com baixos níveis de alfabetização, em combinação com outras variáveis, mostram ocorrências de crimes altos e muito altos. 	<p>De acordo com Clancy et al. (2022) em Chicago a proporção de indivíduos com diploma universitário foi negativa associada com crimes violentos e a proporção daqueles com diploma universitário foi negativamente associado a crimes contra a propriedade.</p>

Fonte: elaboração própria

A seguir, para cada uma das variáveis sociodemográficas, são apresentadas algumas possíveis justificativas que explicam, do ponto de vista das características do espaço urbano, como ela(s) influencia(m) na ocorrência de furtos e roubos.

Variável REM

O estudo de Adewuyi et al. (2017) revelou que os padrões de criminalidade em uma área são predominantemente aleatórios e que para combater a criminalidade de forma duradoura na região, é necessário um esforço conjunto de todas as partes para reduzir o desemprego e a pobreza.

Além disso, LaRue e Andresen (2015) observaram que aumentos na renda média e no valor das habitações estão associados ao roubo. No entanto, a presença de jovens homens, geralmente relacionada ao aumento da criminalidade, teve efeitos negativos e relações estatisticamente significativas com todos os tipos de crime analisados.

Rodrigues (2023), por exemplo, baseou sua análise em estatísticas descritivas, identificando maior prevalência de crimes em setores censitários de menor renda, o que contrasta com as análises do presente estudo. É importante mencionar que métodos estatísticos, como os empregados por Rodrigues (2023), podem apresentar limitações, especialmente em cenários com diversas variáveis interagindo entre si.

Os padrões observados neste estudo indicam que a renda média dos setores censitários é um fator determinante nas ocorrências de crimes. Setores com renda média muito baixa apresentam predominantemente ocorrências de crimes baixas e médias, enquanto setores com renda média alta, especialmente quando associados a alta densidade populacional, apresentam predominantemente ocorrências de crimes altas ou muito altas.

No entanto, a relação entre a renda média e a ocorrência de furtos e roubos em um espaço urbano pode ser complexa e influenciada por vários fatores. Não é à toa que os resultados aqui descritos divergem dos resultados reportados em estudos como os de Rodrigues (2023), LaRue e Andresen (2015), Adewuyi et al. (2017) e Ceccato e Moreira (2021), os quais apontaram que locais com menor renda tendem a registrar níveis mais elevados de criminalidade.

Variável PPJ

Os resultados de Adewuyi et al. (2017) indicam que furtos e roubos à mão armada estão positivamente relacionados com a presença de uma população masculina jovem. Além disso, eles também estabeleceram uma associação entre a presença de uma população jovem e níveis mais elevados de criminalidade.

Contudo, nos resultados deste estudo os setores com um alto percentual de população jovem, especialmente quando combinados com baixa renda média e alto percentual de população masculina, tendem a apresentar altas ou muito altas ocorrências de crimes.

Variável DEP

Clancy et al. (2022) encontraram em Chicago uma associação entre crimes violentos e a população total, corroborando os achados de Adewuyi et al. (2017) que sugere uma ligação direta entre maior densidade populacional e criminalidade e Rodrigues (2023) que observou que setores com maior número de residentes concentraram a maioria dos crimes.

É crucial lembrar que a relação entre densidade populacional e criminalidade é complexa e multifatorial. Diversos outros fatores, como cultura, políticas públicas e dinâmicas sociais, também desempenham papéis cruciais quando combinados com essas variáveis.

Todavia, os achados neste trabalho observaram que setores com valores médios de renda e diferentes níveis de densidade populacional, especialmente alta, apresentam variações nas taxas de crimes, mas inclinam-se majoritariamente para ocorrências baixas e médias. A densidade populacional, em associação com outras variáveis, desempenha um papel importante nessas relações, diferindo dos estudos de Clancy et al. (2022), Adewuyi et al. (2017) e Rodrigues (2023).

Variável PPM

Os padrões identificados nesta pesquisa em relação ao percentual de população masculina revelam que, quando combinado com outras variáveis como renda média e densidade populacional, é possível destacar o impacto nas ocorrências de crimes.

O percentual de população masculina emerge como um indicador significativo nas correlações, pois é frequentemente observado tanto em números elevados quanto baixos de ocorrências de crimes.

No entanto, esses padrões não estão diretamente alinhados com as conclusões do estudo de Adewuyi et al. (2017), que indica uma correlação positiva entre a presença de uma população masculina jovem e altos índices de ocorrências de crimes.

Variável PPA

Os resultados neste estudo em relação ao Percentual de população alfabetizada, embora não tenham sido frequentemente citados, quando mencionados, indicam que setores com baixos níveis de alfabetização, combinados com outras variáveis, apresentam ocorrências de crimes altas e muito altas.

Esses achados contradizem as observações dos estudos de Adewuyi et al. (2017) e Rodrigues (2023) em relação ao percentual de população alfabetizada. De acordo com essas pesquisas, educação não mostrou correlação significativa com os crimes estudados. Já de acordo com Clancy et al. (2022) em Chicago os crimes violentos estão associados a população com diploma universitário.

É importante enfatizar que as variáveis sociodemográficas, por si só, não é uma causa direta de criminalidade. É a interação das variáveis com outras condições sociodemográficas e características urbanas que pode criar um ambiente mais ou menos propício ao surgimento de atividades criminosas.

No que tange a análise de padrões dos crimes com base no perfil das vítimas, neste estudo foram revelados padrões consistentes associados a casos de roubo, o que não se observou no contexto de crimes de furto. Essa discrepância pode ser atribuída ao fato de que, no roubo, o criminoso seleciona especificamente a vítima para exercer violência ou ameaça grave, algo que não se aplica aos casos de furto. Outro ponto que merece destaque nessa análise é a quantidade de registros excluídos na etapa de pré-processamento, em virtude de dados faltantes. Se o volume de dados minerados fosse maior, provavelmente haveria a possibilidade de descoberta de padrões também relacionados aos crimes de furtos. Isso suscita a questão de que os

BOs precisam ser melhor preenchidos tanto no que tange a completude quanto a corretude e padronização dos dados.

Da mesma forma que os resultados expostos nas subseções 4.3.1 e 4.3.2, os estudos conduzidos por Ratul e Rab (2020), Saraiva et al. (2022), Wheeler e Steenbeek (2021), bem como Clancy et al. (2022), utilizando ferramentas de Inteligência Artificial, contribuíram para a compreensão do aspecto espacial e para a identificação de padrões criminais.

É importante destacar que a complexidade do fenômeno criminal demanda um olhar meticuloso, principalmente ao correlacionar variáveis sociodemográficas com taxas de criminalidade. As divergências encontradas entre o estudo e pesquisas anteriores apontam para a necessidade de continuidade nas investigações, considerando especificidades regionais, diferenças metodológicas e outras variáveis não contempladas neste trabalho.

Por fim, cabe destacar que este estudo pode trazer contribuições para o atingimento de alguns Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU), principalmente as relacionadas à redução da desigualdade, promoção de cidades seguras e sustentáveis, e combate à violência, das seguintes maneiras:

Redução da Desigualdade (ODS 10): Ao analisar as relações entre atributos socioespaciais e ocorrências de furto e roubo, é possível identificar áreas mais vulneráveis a crimes. Com esse conhecimento, os governos e as organizações podem direcionar recursos para comunidades de baixa renda ou regiões mais afetadas, reduzindo a desigualdade na distribuição de segurança e proteção.

Cidades e Comunidades Sustentáveis (ODS 11): A análise das relações entre crimes e atributos socioespaciais pode ajudar a planejar o desenvolvimento urbano de forma mais sustentável. Esses insights podem orientar políticas para melhorar a infraestrutura, iluminação pública, acesso a serviços e a distribuição de recursos de segurança, criando ambientes urbanos mais seguros e habitáveis a longo prazo.

Paz, Justiça e Instituições Eficazes (ODS 16): Ao entender os fatores que contribuem para o aumento das ocorrências de furto e roubo, os governos podem implementar estratégias mais eficazes de aplicação da lei e prevenção do crime. Isso

pode envolver a alocação de recursos para patrulhamento policial em áreas críticas, adoção de tecnologias de vigilância inteligente e implementação de políticas de prevenção do crime baseadas em dados.

Parcerias para atingir os Objetivos (ODS 17): A abordagem de inteligência artificial requer colaboração entre governos, organizações da sociedade civil, instituições de pesquisa e empresas de tecnologia. A análise de dados e a implementação de soluções exigem a criação de parcerias para garantir o acesso a informações relevantes, recursos tecnológicos e conhecimento especializado.

Saúde e Bem-Estar (ODS 3): Reduzir as ocorrências de furto e roubo em áreas urbanas pode contribuir indiretamente para o bem-estar da população, já que a violência e a criminalidade afetam negativamente a saúde mental e física das pessoas. Comunidades mais seguras promovem um ambiente propício para um melhor estado de saúde e bem-estar.

Educação de Qualidade (ODS 4) e Trabalho Decente e Crescimento Econômico (ODS 8): A redução da criminalidade pode criar um ambiente mais propício para o acesso à educação de qualidade e para o crescimento econômico sustentável. A presença de crimes pode impactar negativamente o funcionamento das escolas e atração de investimentos econômicos.

5. CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo analisar as relações entre os atributos socioespaciais urbanos e as ocorrências de furto e roubo ocorridos na subprefeitura de Itaquera – São Paulo/Brasil, entre os anos de 2017 e 2021. Para tanto, foi proposta uma abordagem empregando Inteligência Artificial (IA) e Mineração de Dados (MD).

A escolha dessa área de estudo se baseou no fato dela registrar o maior número de ocorrências de crimes contra o patrimônio na cidade e por apresentar fatores, considerados importantes na literatura, associados às ocorrências de furto e roubo.

Cabe destacar que poucos estudos, à exemplo de Rodrigues (2023), consideraram a subprefeitura de Itaquera como recorte espacial. Tal autor analisou as relações socioespaciais dos crimes de furto e roubo por meio de estatística descritiva, que pode ser limitada para identificar padrões, principalmente quando se considera múltiplas variáveis. Nesse cenário, esta pesquisa visa preencher uma lacuna ao explorar a potencialidade da IA na análise de padrões de crimes e sua relação com atributos socioespaciais urbanos.

Este estudo se caracteriza como um estudo exploratório, embasado em pesquisa bibliográfica e análise de dados secundários sobre crimes de furto e roubo oriundos da SSP de São Paulo, e dados sociodemográficos do IBGE, os quais foram georreferenciados por meio de um SIG.

Para descoberta dos padrões os dados foram categorizados e agrupados por setores censitários, permitindo análises de possíveis correlações das variáveis sociodemográficas com as ocorrências de furto e roubo, utilizando os softwares QGIS e WEKA. É importante mencionar que, embora existam estudos correlatos na literatura, são escassos aqueles que exploram o uso de ferramentas de IA, especialmente aquelas voltadas para MD.

Os resultados foram organizados em 4 subseções que destacam os seguintes conteúdos: 1) Espacialização de dados; 2) Análise da distribuição dos dados sociodemográficos e criminais com relação às variáveis explicativas e de resposta e 3) Descoberta de padrões usando MD.

A espacialização dos dados relacionados a furto, roubo e variáveis sociodemográficas serviu como uma ferramenta essencial para a compreensão inicial da dinâmica dos referidos crimes na subprefeitura de Itaquera, além de propiciar a categorização dos dados, para posterior análise na etapa de MD.

Em relação à análise da distribuição dos dados sociodemográficos e criminais em relação às variáveis explicativas e de resposta, o presente estudo possibilitou algumas descobertas sobre a dinâmica da criminalidade e suas possíveis correlações com variáveis sociodemográficas.

Por meio da MD com o uso de Árvores de Decisão (AD), este estudo apresentou uma análise detalhada das relações entre as variáveis sociodemográficas e as ocorrências de crimes nos setores censitários. O uso da IA mostrou ser uma abordagem valiosa na identificação de padrões criminais. A técnica de AD foi muito útil para fornecer insights sobre as nuances dos crimes de furto e roubo.

Os resultados aqui apresentados mostrando que a realidade socioespacial e sua relação com os crimes contra o patrimônio na subprefeitura de Itaquera têm especificidades que devem ser consideradas por estudos, bem como por políticas públicas de prevenção ao crime. Descobertas sobre a dinâmica da criminalidade, especialmente a relação inversa entre renda média e taxas de criminalidade, bem como a redução de crimes em áreas com predominância de população jovem, reforçam tais especificidades.

Porém, alguns desafios surgiram durante a realização deste estudo. Foi observado que a atualidade, a qualidade e a integridade dos dados coletados podem influenciar diretamente a precisão e a profundidade das análises. Um exemplo foi o uso do censo do IBGE de 2010 para extrair os dados das variáveis explicativas, enquanto os dados dos crimes se referiam ao período 2017-2021. Em um cenário ideal, dados atuais e completos poderiam fornecer uma visão mais elucidativa dos padrões de criminalidade na área considerada. Isso pode destacar a importância de futuros estudos de abordagem metodológica semelhante à adotada nessa dissertação, mas que se utilizem de dados sociodemográficos mais atuais, como os dados do mais recente Censo Demográfico, realizado em 2022 e que ainda não se encontra totalmente disponível.

Também é fundamental salientar que a correlação não implica causalidade. A dinâmica do crime é complexa e é influenciada por uma multiplicidade de fatores. As correlações observadas neste estudo devem, portanto, ser interpretadas com cautela e usadas como um ponto de partida para investigações mais profundas e abrangentes.

Os lugares devem ser tratados de maneira distinta, tendo em vista, que a formação socioespacial é única, o que impede que fenômenos como crimes sejam explicados de forma homogênea para todos os lugares. Sendo assim, os resultados também destacam a importância das especificidades socioespaciais que devem ser levadas em consideração para formulação de políticas públicas para combate e prevenção.

No que tange as contribuições da pesquisa, os padrões obtidos a partir da investigação das relações entre os atributos socioespaciais urbanos e as ocorrências de furto e roubo, ocorridos na subprefeitura de Itaquera, podem ser citados como uma contribuição teórica tendo em vista seu potencial para guiar estratégias de combate à criminalidade. Em adição, a abordagem empregando IA também pode ser considerada como uma contribuição científica desta pesquisa já que, apesar de algumas limitações nos resultados produzidos, em virtude dos dados analisados, ela pode ser aplicada na análise de dados em outros contextos da criminologia espacial.

Além disso, a abordagem proposta utilizando IA para analisar a relação entre fatores socioespaciais urbanos e crimes como furto e roubo pode contribuir para vários Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU). Isso inclui reduzir desigualdades (ODS 10), promover cidades sustentáveis (ODS 11), alcançar paz e justiça (ODS 16), estabelecer parcerias (ODS 17), melhorar saúde e bem-estar (ODS 3), e proporcionar educação de qualidade e crescimento econômico (ODS 4 e 8).

Para estudos futuros sugere-se a ampliação do número de variáveis explicativas sociodemográficas, bem como da influência de elementos espaciais, como os polos geradores de viagens (PGVs) na distribuição e concentração de crimes. Além disso, sugere-se que sejam incluídos outros tipos de crime nas análises, tendo em vista que cada crime possui dinâmicas próprias e distintas. Por fim, sugere-se expandir a análise para outras subprefeituras do município de São Paulo, a fim de auxiliar na compreensão da variabilidade espacial dos dados e na formulação de futuras políticas

públicas direcionadas ao combate da criminalidade, tornando as cidades mais seguras, inclusivas e sustentáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

Adeyemi, T. O., Eneji, P. A., Baduku, A. S., & Olofin, E. A. (2017). Spatio-temporal analysis of urban crime pattern and its implication for Abuja municipal area council, Nigeria. *The Indonesian Journal of Geography*, 49(2), 145-154.

Aher, S. B., & Lobo, L. M. R. J. (2011, March). Data mining in educational system using weka. In *International conference on emerging technology trends (ICETT)* (Vol. 3, pp. 20-25).

Andresen, M. A., Curman, A. S., & Linning, S. J. (2017). The trajectories of crime at places: Understanding the patterns of disaggregated crime types. *Journal of quantitative criminology*, 33, 427-449.

Ates, E. C., Bostanci, G. E., & Serdar, M. S. G. (2020). Big Data, data mining, machine learning, and deep learning concepts in crime data. *Journal of Penal Law and Criminology*, 8(2), 293-319.

Azevedo, R., Garcia, A. M., & de Oliveira, A. (2016). *Criminalidade e segurança pública no Brasil*. Editora Contexto.

Baumer, E. P., Wolff, K. T., & Arnio, A. N. (2012). A multicity neighborhood analysis of foreclosure and crime. *Social Science Quarterly*, 93(3), 577-601.

Buil-Gil, D., Moretti, A., & Langton, S. H. (2021). The accuracy of crime statistics: Assessing the impact of police data bias on geographic crime analysis. *Journal of Experimental Criminology*, 1-27.

Ceccato, V., & Moreira, G. (2021). The dynamics of thefts and robberies in São Paulo's Metro, Brazil. *European Journal on Criminal Policy and Research*, 27, 353-373.

Ceccato, V., & Oberwittler, D. (2008). Comparing spatial patterns of robbery: Evidence from a Western and an Eastern European city. *Cities*, 25(4), 185-196.

Ceccato, V., & Paz, Y. (2017). Crime in São Paulo's metro system: Sexual crimes against women. *Crime Prevention and Community Safety*, 19(3-4), 211-226. <https://doi.org/10.1057/s41300-017-0027-2>

Ceccato, V., & Uittenbogaard, A. C. (2014). Space–Time Dynamics of Crime in Transport Nodes. *Annals of the Association of American Geographers*, 104(1),131–150. <https://doi.org/10.1080/00045608.2013.846150>

Clancy, K., Chudzik, J., Snowden, A. J., & Guha, S. (2022). Reconciling data-driven crime analysis with human-centered algorithms. *Cities*, 124, 103604.

Companhia Paulista de Trens Metropolitanos. (2022). Transparência. <https://www.cptm.sp.gov.br/Transparencia/>

Creswell, J. W., & Creswell, J. D. (2021). Projeto de pesquisa-: Métodos qualitativo, quantitativo e misto. Penso Editora.

Cundiff, K. (2021). Colleges and community crime: an analysis of campus proximity and neighborhood crime rates. *Crime & Delinquency*, 67(3), 431-448.

da Silva, J. A. S., & Mairink, C. H. P. (2019). Inteligência artificial. *LIBERTAS: Revista de Ciências Sociais Aplicadas*, 9(2), 64-85.

da Silva Furtado, D. B., Quaresma, C. C., Oliveira, E. A. D. A. Q., de Melo Conti, D., & da Cunha Calado, J. (2020). Iniciativas sociais na superação da crise de mobilidade urbana em áreas segregadas: o caso da Jaubra na Brasilândia, São Paulo, Brasil. *Humanidades & Inovação*, 7(5), 52-67.

Dakalbab, F., Talib, MA, Waraga, OA, Nassif, AB, Abbas, S., & Nasir, Q. (2022). Inteligência artificial e previsão do crime: uma revisão sistemática da literatura. *Ciências Sociais e Humanas Open*, 6 (1), 100342.

David, H., & Suruliandi, A. (2017). SURVEY ON CRIME ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES. *ICTACT journal on soft computing*, 7(3).

de Andrade Carneiro, L. (2022). CAUSAS E CONSEQUÊNCIAS DA CRIMINALIDADE NO BRASIL: UMA REVISÃO DA LITERATURA. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, 8(7), 20-44.

de Oliveira, T. L. (2022). A ANÁLISE CRÍTICA DO MOMENTO CONSUMATIVO NOS DELITOS DE FURTO E ROUBO. RECIMA21-Revista Científica Multidisciplinar-ISSN 2675-6218, 3(1), e31952-e31952.

de Paula Faleiros, V. (2022). Criminalidade, desigualdade social e penalização de adolescentes e jovens. Revista Direito, Estado e Sociedade, (61).

Dornelles, J. R. W. (2017). O que é crime. Brasiliense.

Durkheim, É. (2017). A educação moral. Editora Vozes Limitada.

Dutta, S., Gupta, A. K., & Narayan, N. (2017, October). Identity crime detection using data mining. In 2017 3rd International Conference on Computational Intelligence and Networks (CINE) (pp. 1-5). IEEE.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37-37.

FBSP – Fórum Brasileiro de Segurança Pública. (2022). Anuário Brasileiro de Segurança Pública 2022. Disponível em: <https://forumseguranca.org.br/anuario-brasileiro-seguranca-publica/>. Acesso em: 05/04/2023

Feng, M., Zheng, J., Ren, J., Hussain, A., Li, X., Xi, Y., & Liu, Q. (2019). Big data analytics and mining for effective visualization and trends forecasting of crime data. IEEE Access, 7, 106111-106123.

Foucault, M. (2012). Discipline and punish: The birth of the prison. Vintage.

Freitas, C. A. de O. (2020). Transformações na produção da periferia: O caso de Itaquera e o Conjunto Habitacional José Bonifácio. Revista Brasileira de Estudos Urbanos e Regionais. <https://doi.org/10.22296/2317-1529>.

SEADE – Fundação Seade. POPULAÇÃO. Disponível em: <https://populacao.seade.gov.br/>. Acesso em 07/04/2023.

Glasner, P., & Leitner, M. (2016). Evaluating the impact the weekday has on near-repeat victimization: A spatio-temporal analysis of street robberies in the city of Vienna, Austria. ISPRS International Journal of Geo-Information, 6(1), 3.

- Greco, R. (2017). Curso de Direito Penal: parte geral, volume I. Niterói, RJ: Impetus, 48-50.
- Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3^a ed.). Morgan Kaufmann.
- Hassani, H., Huang, X., Silva, E. S., & Ghodsi, M. (2016). A review of data mining applications in crime. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(3), 139-154.
- Khalid, S., Shoaib, F., Qian, T., Rui, Y., Bari, A. I., Sajjad, M., ... & Wang, J. (2018). Network constrained spatio-temporal hotspot mapping of crimes in Faisalabad. *Applied Spatial Analysis and Policy*, 11, 599-622.
- Kingsford, C., & Salzberg, S. L. (2008). What are decision trees?. *Nature biotechnology*, 26(9), 1011-1013.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.
- LaRue, E., & Andresen, M. A. (2015). Spatial patterns of crime in Ottawa: The role of universities. *Canadian Journal of Criminology and Criminal Justice*, 57(2), 189-214.
- Lombroso, C. (2020). O homem delinquente. Editora Edijur.
- Martinsa, A. C., Teixeirab, E. C., & Silvac, G. D. (2021) Determinantes da Probabilidade de Subnotificação de Crimes Contra o Patrimônio no Brasil.
- Matijosaitiene, I., McDowald, A., & Juneja, V. (2019). Predicting safe parking spaces: A machine learning approach to geospatial urban and crime data. *Sustainability*, 11(10), 2848.
- Oatley, G. C. (2022). Themes in data mining, big data, and crime analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(2), e1432.
- Oliveira, E. D., Couto, M. T., Separavich, M. A. A., & Luiz, O. D. C. (2020). The contribution of intersectionality on understanding young men's health-disease and care in contexts of urban poverty. *Interface-Comunicacao Saude Educacao*.

OMS – Organização Mundial da Saúde. Relatório Mundial sobre a Prevenção da Violência 2014. Núcleo de Estudos da Violência (Trad.) São Paulo: 2015. Disponível em: <https://nev.prp.usp.br/publicacao/relatorio-mundial-sobre-a-prevencao-da-violencia-2014/>. Acesso em 07/04/2023.

Quick, M., Li, G., & Brunton-Smith, I. (2018). Crime-general and crime-specific spatial patterns: A multivariate spatial analysis of four crime types at the small-area scale. *Journal of Criminal Justice*, 58, 22-32.

Quinlan, J. R., & Cameron-Jones, R. M. (1993). FOIL: A midterm report. In *Machine Learning: ECML-93: European Conference on Machine Learning Vienna, Austria, April 5–7, 1993 Proceedings 6* (pp. 1-20). Springer Berlin Heidelberg.

Ratul, M., & Rab, A. (2020). A comparative study on crime in Denver City based on machine learning and data mining. *arXiv preprint arXiv:2001.02802*.

Ristea, A., Kurland, J., Resch, B., Leitner, M., & Langford, C. (2018). Estimating the Spatial Distribution of Crime Events around a Football Stadium from Georeferenced Tweets. *ISPRS International Journal of Geo-Information*, 7(2), 43. <https://doi.org/10.3390/ijgi7020043>

Rodrigues, Luís Flávio. (2023). *Relações socioespaciais dos crimes de furto e roubo na subprefeitura de Itaquera, São Paulo/sp – brasil*. Universidade Nove de Julho.

Saraiva, M., Matijošaitienė, I., Mishra, S., & Amante, A. (2022). Crime Prediction and Monitoring in Porto, Portugal, Using Machine Learning, Spatial and Text Analytics. *ISPRS International Journal of Geo-Information*, 11(7), 400.

Sayad, Y. O., Mousannif, H., & Al Moatassime, H. (2019). Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire safety journal*, 104, 130-146.

Scarlato, F. C., & da Anunciação Alves, G. (2018). São Paulo: uma metrópole em constante mutação. *Boletim Paulista de Geografia*, (100), 156-172.

Silva, R. C. F., Caleffi, F., & Ruiz-Padillo, A. (2020). Diagnóstico das Relações Existentes em Estudos Recentes sobre Polos Geradores de Viagens.

Smith, T. A., & Sandoval, J. O. N. (2020). An Exploratory Spatial Analysis of the Urban Crime Environment around the Next National Geospatial-Intelligence Agency West Development. *The Geographical Bulletin*, 61(1), 37-53.

Soares, G. F. Ciência de dados aplicada à Auditoria Interna. *Revista da CGU*, v. 12, n. 22, p. 196-208, 2020.

Sumariva, P. (2023). *Criminologia-Teoria e Prática*. Editora Foco.

Thongsatapornwatana, U. (2016, January). A survey of data mining techniques for analyzing crime patterns. In *2016 Second Asian Conference on Defence Technology (ACDT)* (pp. 123-128). IEEE.

Wheeler, A. P., & Steenbeek, W. (2021). Mapping the risk terrain for crime using machine learning. *Journal of Quantitative Criminology*, 37, 445-480.

Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4^a ed.). Morgan Kaufmann.

Xu, Y., & Feng, T. (2022). Analysis of Legal Issues of the Crime of Endangering Public Safety Based on Data Mining Algorithm. *Security and Communication Networks*, 2022.

Zhou, Z. H. (2021). *Machine learning*. Springer Nature.