

**Data Science aplicada à análise criminal baseada nos dados abertos governamentais de Minas Gerais**

**Data Science applied to criminal analysis based on Minas Gerais open government data**

**Ciencia de Datos aplicada al análisis criminal basado en datos abiertos del gobierno de Minas Gerais**

Recebido: 08/11/2020 | Revisado: 11/11/2020 | Aceito: 13/11/2020 | Publicado: 17/11/2020

**Kleber Henrique de Jesus Prado**

ORCID: <https://orcid.org/0000-0002-3555-9475>

Universidade Federal de Sergipe, Brasil

E-mail: [kleprado@hotmail.com](mailto:kleprado@hotmail.com)

**Methanias Colaço Junior**

ORCID: <https://orcid.org/0000-0002-4811-1477>

Universidade Federal de Sergipe, Brasil

E-mail: [mjrse@hotmail.com](mailto:mjrse@hotmail.com)

**Resumo**

Contexto: Crime é um problema social comum e complexo, que afeta a qualidade de vida, o crescimento econômico e a reputação de uma nação. Governantes e a sociedade em geral têm tido enormes problemas causados por esse fenômeno. A cada ano, os governos gastam milhões de dólares combatendo a violência e, conseqüentemente, a prevenção e o controle do crime são questões de grande preocupação para as agências de segurança pública. Objetivo: Aplicar fundamentos de Data Science e fornecer um modelo automatizado, constantemente atualizado, para analisar dados abertos governamentais relacionados aos crimes ocorridos em Minas Gerais. Método: Um experimento foi executado para descoberta de associações entre os municípios, Regiões Integradas de Segurança Pública (RISPs), crimes e alvos de roubo e furto. Adicionalmente, foram desenvolvidos rankings com os municípios mais perigosos. Resultados: Do ponto de vista geral, com ponderações para os crimes, os municípios de Belo Horizonte, Confins e Contagem estiveram, constantemente, entre os cinco mais perigosos. Além disso, ficou evidenciado que existem dependências entre: crimes e municípios, crimes e RISPs, alvos de roubo e municípios, e alvos de roubo e RISPs. Conclusão: A Ciência de Dados possibilita a execução de diagnósticos mais precisos e mais céleres, auxiliando o

planejamento estratégico e a tomada de decisão em Segurança Pública. Com algumas particularidades e indo além dos homicídios, Minas Gerais segue parcialmente a tendência nacional de ter índices de criminalidade mais baixos em áreas ao redor de regiões com maior desenvolvimento econômico.

**Palavras-chave:** Análise criminal; Ciência de dados; Dados abertos governamentais.

### **Abstract**

**Context:** Crime is a common and complex social problem that affects a nation's quality of life, economic growth and reputation. Governments and society in general have had enormous problems caused by this phenomenon. Each year, governments spend millions of dollars fighting violence and, consequently, crime prevention and control are issues of great concern to public security agencies. **Objective:** To apply fundamentals of Data Science and provide an automated model, constantly updated, to analyze open government data related to crimes occurred in Minas Gerais. **Method:** We have performed an experiment to discover associations between municipalities, Integrated Public Security Regions (IPSRs), crimes, robbery targets, and theft targets. Additionally, we have developed rankings with the most dangerous municipalities. **Results:** From a general point of view, with scores for crimes, Belo Horizonte, Confins and Contagem were always among the five most dangerous. In addition, it became evident that there are dependencies between: crimes and municipalities, crimes and IPSRs, robbery targets and municipalities, and robbery targets and IPSRs. **Conclusion:** Data Science enables the execution of more accurate and faster diagnoses, helping strategic planning and decision making in Public Security. With some peculiarities and going beyond homicides, Minas Gerais partially follows the national trend of having lower crime rates in areas around regions with greater economic development.

**Keywords:** Criminal analysis; Data science; Open government data.

### **Resumen**

**Contexto:** El crimen es un problema social común y complejo que afecta la calidad de vida, el crecimiento económico y la reputación de una nación. Los gobiernos y la sociedad en general han tenido enormes problemas provocados por este fenómeno. Cada año, los gobiernos gastan millones de dólares en la lucha contra la violencia y, en consecuencia, la prevención y el control del delito son temas de gran preocupación para las agencias de seguridad pública. **Objetivo:** Aplicar los fundamentos de la ciencia de datos y proporcionar un modelo automatizado, constantemente actualizado, para analizar datos gubernamentales abiertos

relacionados con delitos ocurridos en Minas Gerais. Método: Se llevó a cabo un experimento para descubrir asociaciones entre municipios, Regiones de Seguridad Pública Integrada (RISPs), delitos y robos y objetivos de hurto. Además, se desarrollaron rankings con los municipios más peligrosos. Resultados: Desde el punto de vista general, considerando los delitos, los municipios de Belo Horizonte, Confins y Contagem estuvieron constantemente entre los cinco más peligrosos. Además, se hizo evidente que existen dependencias entre: delitos y municipios, delitos y RISPs, objetivos de robo y municipios, y objetivos de robo y RISPs. Conclusión: La ciencia de datos permite la ejecución de diagnósticos más precisos y rápidos, lo que ayuda a la planificación estratégica y la toma de decisiones en seguridad pública. Con algunas peculiaridades y más allá de los homicidios, Minas Gerais sigue parcialmente la tendencia nacional de tener menores índices de criminalidad en áreas alrededor de las regiones con mayor desarrollo económico.

**Palabras clave:** Análisis criminal; Ciencia de datos; Datos abiertos gubernamentales.

## 1. Introdução

Com o aumento da urbanização, diversas transformações sociais, econômicas e ambientais têm ocorrido em todas as partes do mundo, com desafios cada vez maiores enfrentados pelos governantes. Áreas como mobilidade urbana, saúde e segurança pública têm recebido uma atenção especial (Catlett et al., 2018).

Nos últimos anos, a criminalidade vem se destacando como um problema social predominante e sua mitigação é extremamente importante. Segundo (ToppiReddy et al., 2018), a prevenção e controle do crime são questões de grande preocupação para os governos e agências de segurança pública, uma vez que, se não forem bem controladas e gerenciadas, podem afetar drasticamente a economia de um país ao longo do tempo, uma vez que mais emigração ocorrerá naturalmente. A cada ano, os governos gastam milhões de dólares combatendo a violência, fornecendo equipamentos, treinamento e adquirindo ferramentas para auxiliar o trabalho policial. Consequentemente, governantes e a sociedade, em geral, têm tido enormes problemas causados por esse fenômeno. É responsabilidade das agências de aplicação da lei monitorar e reduzir a taxa de atividades criminosas que estão acontecendo continuamente nos dias de hoje (Damasceno et al., 2012) (Marzan et al., 2017).

Um outro grande desafio enfrentado por essas organizações é lidar com um grande volume de informações referentes aos crimes e criminosos. De acordo com (Catlett et al., 2018), uma quantidade significativa de dados com informações espaciais e temporais é obtida

diariamente. Consequentemente, novas abordagens e sistemas avançados são necessários para melhorar a análise de crimes e para proteger suas comunidades, permitindo uma maior compreensão da dinâmica das atividades criminosas e fornecendo respostas como “onde”, “quando” e “por que” certos crimes são prováveis de acontecer (Phillips & Lee, 2011)(ToppiReddy et al., 2018). Neste contexto, a *Data Science*, aliada aos sistemas computacionais inteligentes, vem desempenhando um papel vital na melhoria dos resultados das investigações e detecções criminais, facilitando o registro, a análise de recuperação e o compartilhamento das informações (Gupta et al., 2014).

Essa evolução da Ciência de Dados e da Tecnologia da Informação e Comunicação (TIC) em geral produziu novas maneiras de disponibilizar informações públicas à população. Novos sistemas foram desenvolvidos, novos serviços oferecidos e integrações sistêmicas aconteceram, gerando mudanças nos processos internos e nas relações do governo com o público externo. Por meio do Governo Eletrônico, ou simplesmente *e-Gov*, os órgãos públicos passaram a ser os maiores criadores e coletores de dados em muitos domínios. Estes domínios de dados variam entre tráfego, clima, informações geográficas, turísticas, segurança pública, estatísticas, negócios, orçamentação do setor público e vários outros (Janssen et al., 2012).

Entretanto, apesar das ações realizadas pelos governos federais e estaduais brasileiros para publicitar informações sobre estatísticas oficiais de segurança pública, nenhuma dessas iniciativas que promovem a transparência parece ser suficiente para produzir informações claras, consistentes e transparentes ao grande público. São milhares de fluxos de dados publicados frequentemente, sem a certeza de que sejam utilizáveis e sem uma aplicação inteligente sobre os dados avaliados. Muitas vezes, sem métricas que possibilitem ao cidadão inferir conclusões acerca da publicação.

Em razão disso, a proposta deste artigo é aplicar *Data Science* apoiada em um processo experimental, para realizar uma avaliação sobre dados abertos governamentais relacionados a incidentes criminais, dos municípios de Minas Gerais (MG), disponibilizados pela Secretaria de Estado de Justiça e Segurança Pública (Sejusp), que, apesar da limitação ainda presente, é uma das secretarias estaduais que melhor detalha os eventos criminais. Logo, a escolha do estado se deu pelo seu porte e pela disponibilidade das informações, objetivando detectar associações entre cidades e crimes, *rankings* de perigosidade, padrões e modelos replicáveis em outros estados, bem como promover maior transparência, visando auxiliar o processo de apoio às tomadas de decisões estratégicas e operacionais dos governantes e agentes da lei, no combate efetivo da criminalidade. Ainda neste contexto, este trabalho contribui para a produção e disponibilidade de bases de dados abertas e consistentes

pelas Secretarias Estaduais, inclusive para o melhoramento dos dados da Sejust-MG, bem como para que outras organizações, cidadãos e pesquisadores produzam e possuam modelos para automatização da extração, produção de conhecimento e publicação atualizada constantemente sobre dados abertos criminais, intensificando as suas atuações como controladores sociais.

Para alcançar esse fim, o restante deste artigo está organizado da seguinte forma. Na seção 2, os trabalhos relacionados sobre o tema são apresentados. Na seção 3, a metodologia adotada é abordada. A seção 4 descreve alguns conceitos básicos necessários para o entendimento deste trabalho. Na seção 5, a definição e o planejamento do experimento são apresentados. A seção 6 detalha a operação do experimento. Na seção 7, os resultados são analisados. E, finalmente, na seção 8, a conclusão e os trabalhos futuros são apresentados.

## 2. Trabalhos Relacionados

De acordo com (Singh et al., 2018), as técnicas de *Data Mining* têm se mostrado eficazes na análise de conjuntos de dados e na coleta de informações úteis em muitos domínios. No campo criminal, a mineração de dados está recebendo maior atenção para descobrir padrões subjacentes nos dados sobre crimes. A seguir, são descritos alguns trabalhos que utilizaram bases de dados que, minimamente, contêm informações semelhantes às disponibilizadas pelo governo de Minas Gerais. Na maioria dos casos, em uma situação diferente do Brasil, o maior detalhamento dos dados fornecido pelos governos permite o uso de mais opções de algoritmos, não explorados neste artigo pela incompletude dos dados disponíveis.

Neste contexto, Agrawal & Sejwar (2017) utilizaram o algoritmo chamado Multi Objective Particle Swarm Optimization (MOPSO). Essa nova proposta gera padrões otimizados, utilizando o algoritmo para descoberta de associações *FP-Growth* em bancos de dados criminais. Primeiramente, os padrões foram extraídos por meio da execução do algoritmo *FP-Growth*, o qual tentou constatar os padrões frequentes dos tipos de crimes em relação às cidades (localidade). Em seguida, os padrões otimizados foram gerados, utilizando o MOPSO. Segundo os autores, os resultados evidenciaram que a abordagem proposta é promissora, indicando com que frequência um padrão de crime aparece no banco de dados e auxiliando os analistas de crimes a obter conhecimento sobre a ocorrência de crimes em locais específicos, em menos tempo de execução.

Ainda sobre combate à criminalidade e Regras de Associação (RAs), os pesquisadores

de (Marzan et al., 2017) tiveram como objetivo identificar áreas com maior ocorrência de crimes (*hotspots*) da cidade de Manila, Filipinas. Adicionalmente, usaram o algoritmo *Apriori* para descoberta de padrões frequentes, ajudando os policiais a formar uma ação preventiva. Este trabalho também avaliou vários métodos de previsão de séries temporais, tais como regressão linear, processos gaussianos, multicamada *Perceptron* e *SMOreg*, para prever tendências futuras do crime. Como resultado, o algoritmo multicamada *Perceptron* foi capaz de prever o número de crimes na maioria dos locais em Manila, com mais precisão do que as outras técnicas. De forma semelhante, Yadav et al. (2017) utilizaram os algoritmos *Apriori*, *K-Means* e *Naive Bayes*, bem como análise de correlação e regressão, para tentar ajudar os especialistas criminais a descobrir padrões, tendências, realizar previsões, encontrar relacionamentos e possíveis explicações, mapear redes criminosas e identificar suspeitos. De acordo com os autores, o modelo desenvolvido reduzirá os crimes e auxiliará o processo de detecção de crimes de várias maneiras. O trabalho aqui apresentado, dentro das possibilidades que os dados abertos proporcionam, também utilizou o algoritmo *Apriori* para descoberta de associações e produziu um *ranking* de perigosidade para as cidades de Minas Gerais.

Como este trabalho, apresentou dados sobre diversos furtos que ocorrem em Minas Gerais, em (Carazza et al., 2020), os pesquisadores investigaram os efeitos da adoção de uma política de toque de recolher juvenil, implantada por diversas cidades do estado de São Paulo. Foi utilizada a estimativa “difference-in-differences”, para comparar seu impacto em relação aos municípios que não adotaram tal medida. Os resultados indicaram uma redução de 17,9% da taxa de furto, nas cidades que implementaram a política.

No que diz respeito a trabalhos nacionais sobre criminalidade e regionais sobre dados criminais de Minas Gerais, independente dos procedimentos usados terem sido totalmente automatizados, é preciso ficar atento às bases científicas locais e a alguns trabalhos mais antigos. Por exemplo, na nossa Revisão Sistemática da Literatura publicada, a faixa utilizada foi dos últimos dez anos e a prioridade foram as maiores bases de dados de pesquisa do mundo, tal como a *Scopus*, a qual não indexa muitos dos trabalhos nacionais sobre crimes. Isto pode ser um indicativo da necessidade de maior renovação dos trabalhos ou de mais publicações em outras línguas, nesta área específica de automação da produção de conhecimento. Além disso, o foco desta pesquisa foi a busca por trabalhos que utilizaram dados abertos criminais governamentais e que, de alguma forma, automatizaram a coleta, extração, transformação, geração e publicação de conhecimento sobre estes dados, dentro da limitação do que é disponibilizado.

Não obstante não terem sido encontrados trabalhos relacionados dentro do escopo da

Revisão Sistemática efetuada, fonte de entrada para esta pesquisa, em outros indexadores, tais como Scielo e DOAJ, é possível encontrar trabalhos que possuem uma relação indireta com a proposta deste trabalho, mas são de grande relevância para o relacionamento dos resultados aqui expostos com fatores econômicos, sociais e demográficos. Neste contexto, mais recentemente, destacam-se dois trabalhos. O trabalho de de Barros et al. (2019) analisou a relação entre a taxa de homicídios e o nível de desenvolvimento econômico dos municípios brasileiros, observando-se que regiões, com alto nível de desenvolvimento, tendem a ser cercadas por municípios com baixo índice de criminalidade. No entanto, para alguns municípios, o nível de desenvolvimento econômico não é capaz de barrar o avanço do crime.

Com uma relação mais direta com este trabalho, do ponto de vista da localidade, Ervilha & Lima (2019) verificaram em que magnitude, e relevância, variáveis socioeconômicas e demográficas específicas afetaram as taxas de criminalidade e suas variações delituosas, nos municípios do estado de Minas Gerais, no período de 2000 a 2014. As análises foram realizadas por meio da modelagem econométrica de dados em painel e evidenciaram que políticas de combate à criminalidade devem ser conjugadas com outras políticas públicas relacionadas à educação e assistência social, considerando a faixa etária e a vulnerabilidade socioeconômica da população. Estes trabalhos podem ser usados como base para inferências sobre os níveis de perigosidade e associações encontradas.

### 3. Metodologia

A metodologia adotada para o trabalho envolveu, inicialmente, uma Revisão Sistemática (RS) (Pereira et al., 2018) quantitativa da literatura, publicada em (Prado et al., 2020), tendo por finalidade encontrar o estado da arte das pesquisas sobre análise inteligente de dados abertos governamentais relacionados a incidentes criminais. Para operacionalizar a revisão, acessamos a base *Scopus*, a qual inclui buscas em diferentes bancos de dados científicos (IEEE, ACM, Elsevier e outros), considerando publicações dos últimos dez anos.

Ato contínuo, para realização do objetivo principal desta pesquisa, foram utilizados os dados abertos criminais disponibilizados pela Sejusp/MG. Todavia, algumas dessas informações são fornecidas em arquivos muito grandes, com dados dispersos, limitando o entendimento do cidadão com relação ao significado ou à importância dos dados abertos. Desta forma, este trabalho permitiu e permitirá fazer a transição dos dados brutos do governo para informações estruturadas, perfazendo o *download* do(s) arquivo(s), nos formatos CSV, bem como a leitura, interpretação do conteúdo e armazenamento numa base de dados

estruturada. A seção 4.5 ilustra uma visão geral da arquitetura utilizada para realizar as etapas, que vão desde o *download* do *dataset* até a detecção dos padrões.

Do ponto de vista da classificação, em Computação, especificamente na Ciência de Dados, os objetos de estudo são dados e algoritmos, os quais foram analisados em laboratório, ambiente controlado, com a averiguação e o teste de hipóteses, antes da publicação da aplicação. Desta forma, este estudo pode ser classificado como experimental (Pereira et al., 2018), pela ocorrência de testes de hipóteses (Juristo & Moreno, 2013) (Blackburn, 2016), e tem características de um estudo “*in vitro*”, pela experiência em laboratório, do ponto de vista da Computação. Por outro lado, não pode ser classificado como “*in virtuo*” ou “*in silico*”, pois não há simulação computacional de pessoas e/ou de outros aspectos do mundo real (Travassos & Barros, 2003) (Wohlin et al., 2012).

Neste contexto, este trabalho conduziu um experimento, o qual tem o seu método descrito de forma autocontida, no seu planejamento, detalhado na seção 5, com 3 etapas macro: (1) identificação e *download* dos arquivos nos formatos CSV; (2) construção de programas ETL (*Extract, Transform, Load*) para as cargas dos dados e da base de dados para tratamento, estruturação e armazenamento das informações contidas nos arquivos; (3) seleção, exploração, análise, testes de hipóteses e validação das informações oriundas dos dados estruturados.

## **4. Base Conceitual**

### **4.1 Transparência Pública e Dados Abertos**

A transparência na gestão pública é um mecanismo de controle social, associado ao princípio constitucional da publicidade, que prevê obrigatoriedade na divulgação das contas públicas em governos democráticos e estabelece um conjunto de aspectos que sugerem a existência de políticas, procedimentos e tecnologias que proporcionem acesso, uso, qualidade, compreensão e auditabilidade de processos e informações (Albuquerque et al., 2016) (Nascimento, 2019).

Segundo Fraga et al. (2019), o acesso dos cidadãos e órgãos fiscalizadores às atividades desenvolvidas pela administração pública é de grande relevância, pois em algumas situações, os gestores tomam decisões em benefício próprio, em detrimento de toda a sociedade, o que acaba por influenciar nos resultados econômicos e sociais. Dessa forma, torna-se importante que ocorra transparência de informações entre as partes, para que os



cidadãos possam visualizar as ações realizadas pelos governantes e tenham informações suficientes sobre a aplicação de suas contribuições.

Neste mesmo contexto, o conceito de dados abertos baseia-se no fato de que existem certas informações cujo acesso deve estar acima de *copyright*, patentes, censura ou qualquer outro acesso privado (Bertot et al., 2014). De acordo com Hardy & Maurushat (2017), dados abertos são dados acessíveis gratuitamente ou a um custo mínimo, por qualquer pessoa, sendo reutilizados para qualquer finalidade. Os governos comprometidos com o movimento de dados abertos estão postando milhares de conjuntos de dados em portais *on-line*. O objetivo de liberar esses dados não é meramente fornecer informações ao público, mas impulsionar a inovação por meio da análise destes “grandes repositórios de dados”. Portanto, os governos, ao “abrirem” seus *datasets*, permitem que diversas empresas, pesquisadores e o público em geral possam extrair novas informações (*insight*) desse “mar de dados” e contribuir com soluções inovadoras para políticas complexas.

No contexto brasileiro, o portal da Transparência (Portal Transparência, 2020) e o portal de Dados Abertos (Dados Abertos, 2020) do Governo Federal são ferramentas que funcionam como um grande catálogo que facilita a busca e o uso de todo e qualquer tipo de dado publicado pelos órgãos do governo. Informações sobre saúde suplementar, sistema de transporte, segurança pública, indicadores de educação, gastos governamentais, processo eleitoral, programas sociais e outros podem ser facilmente encontradas. Com isso, abre-se um canal direto entre o cidadão e governo, com o intuito de melhorar a utilização dos dados, proporcionando o fortalecimento do processo democrático e melhorando a qualidade de vida da população.

Na esfera criminal, o Governo Federal instituiu, pela Lei nº 12.681, de 4 de Julho de 2012, o Sistema Nacional de Informações de Segurança Pública, Prisionais e de Rastreabilidade de Armas e Munições, de Material Genético, de Digitais e de Drogas (SINESP). Este sistema é um portal de informações integradas, possibilitando consultas operacionais, investigativas e estratégicas, sobre drogas, segurança pública, justiça, sistema prisional, entre outras, implementado em parceria com os entes federados (Dados Abertos, 2020). Considerando que os dados do SINESP são muito gerais, nesta mesma linha, alguns estados passaram a disponibilizar, normalmente, por meio das suas secretarias de segurança pública, informações mais detalhadas, relacionadas aos incidentes criminais ocorridos em seus territórios. Dessa forma, utilizamos para este experimento, os dados criminais de Minas Gerais. A escolha pelo conjuntos de dados desse estado é resultado de uma busca minuciosa de *datasets* estaduais disponibilizados de forma *online* na web, a qual constatou que esta

Unidade Federativa, até o momento da pesquisa, apresentou, em relação aos outros estados, uma melhor qualidade na disponibilização e composição dos seus *datasets*, os quais são publicados em formatos abertos (arquivos .csv), padronizados, contextualizados e podem ser lidos facilmente por máquinas. Adicionalmente, ficou constatado que o número de *datasets* (anos) disponíveis, relacionados a dados criminais, era maior em relação aos demais estados, o que nos possibilitou realizar uma análise mais robusta.

#### **4.2 Região Integrada de Segurança Pública (Risp)**

A Região Integrada de Segurança Pública (Risp) é a instância responsável pelo planejamento estratégico das áreas de coordenação integrada da segurança pública, que estabelece as diretrizes de enfrentamento da criminalidade, a partir da promoção da articulação entre as polícias civil e militar (Sejusp, 2020). Para Maciel et al. (2019), Risp é uma divisão geográfica que permite a articulação e integração regional, no nível tático e operacional, dos órgãos competentes da pasta de segurança pública, para realizar o planejamento, controle, supervisão, avaliação e monitoramento corretivo das atividades de segurança pública. Normalmente, uma Rip é constituída por um ou mais municípios.

Atualmente, as Risps encontram-se constituídas em diversos estados brasileiros. No contexto de Minas Gerais, foco do nosso trabalho, o governo estadual assinou uma resolução conjunta, no dia 15 de fevereiro de 2008, que instituiu a integração geográfica entre as polícias civil e militar. Na época, tal resolução definiu que a correspondência circunscricional das instituições ficava estruturada em 16 Risps, em todo território estadual (Sejusp, 2020). Com o passar dos anos, mais 3 regiões foram criadas, totalizando, atualmente, 19 regiões integradas. A Figura 1 mostra um mapa territorial de Minas Gerais que detalha a distribuição atual das Regiões Integrada de Segurança Pública.

**Figura 1.** Divisão territorial por Risp no estado de Minas Gerais.



Fonte: Diagnóstico da Violência Doméstica e Familiar contra a Mulher em Minas Gerais (2013-2015).

### 4.3 Regras de Associação

Conhecida como Análise de Cesta de Mercado (*Market Basket Analysis*), as Regras de Associação (RAs) são técnicas de mineração de dados que representam combinações de itens que ocorrem com determinada frequência em uma base de dados. Segundo (Marzan et al., 2017), uma das abordagens de mineração de dados mais conhecidas para encontrar conjuntos de itens frequentes e gerar regras de associação é o algoritmo *Apriori*. Isso abriu o caminho para os pesquisadores resolverem problemas sobre como encontrar padrões ocultos em análises criminais. Vale ressaltar que este experimento utilizou o algoritmo *Apriori* para a detecção das regras de associação.

De maneira formal, Agrawal et al. (1993) definiram as regras de associação da seguinte forma: Sejam  $I = i_1, i_2, \dots, i_m$  um conjunto de  $m$  itens distintos e  $D$  uma base de dados formada por um conjunto de transações, onde cada transação  $T$  é composta por um conjunto de itens (*itemset*), tal que  $T \subseteq I$ . Uma regra de associação é uma expressão na forma  $A \Rightarrow B$ , onde  $A \subset I$ ,  $B \subset I$ ,  $A \neq \emptyset$ ,  $B \neq \emptyset$  e  $A \cap B = \emptyset$ .  $A$  é denominado antecedente e  $B$  denominado conseqüente da regra. Tanto o antecedente quanto o conseqüente de uma regra de associação podem ser formados por conjuntos contendo um ou mais itens. Como exemplo,  $A$  e  $B$  podem ser produtos ou eventos, então, poderíamos ter a regra, “quem compra o produto  $A$  também compra o produto  $B$ ”, ou a regra, “quando ocorre um roubo ( $A$ ) também ocorre um

assassinato (B)”.

Existem várias medidas de interesse que avaliam regras de acordo com as restrições iniciais do usuário. Diversos pesquisadores propõem medidas que objetivam extrair um padrão específico dos dados (Campos, 2018). Em nossa pesquisa, utilizamos as seguintes medidas de interesse:

- **Suporte:** O suporte de um conjunto de itens Z,  $\text{Sup}(Z)$ , representa o percentual de transações da base de dados que contêm os itens do conjunto Z. O suporte de uma regra de associação  $A \Rightarrow B$ ,  $\text{Sup}(A \Rightarrow B)$ , é dado por  $\text{Sup}(A \cup B)$ , e pode ser visto na Equação 1.

$$\text{Sup}(A \Rightarrow B) = \frac{\text{Transações que contêm A e B}}{\text{Total de transações}} \quad (1)$$

- **Confiança:** A confiança da regra  $A \Rightarrow B$ ,  $\text{Conf}(A \Rightarrow B)$  representa, dentre as transações que contêm A, a porcentagem de transações que também contêm B, é dada por  $\text{Conf}(A \Rightarrow B) = \text{Sup}(A \cup B) \div \text{Sup}(A)$ , e pode ser vista na Equação 2.

$$\text{Conf}(A \Rightarrow B) = \frac{\text{Número total de transações que contêm A e B}}{\text{Total de transações de contêm A}} \quad (2)$$

- **Lift:** A medida de interesse *Lift*, também conhecida como *Interest*, é utilizada para indicar, dado uma associação  $A \Rightarrow B$ , o quanto mais frequente torna-se B quando ocorre em conjunto com A. O *Lift* de uma regra de associação  $A \Rightarrow B$  é dado pela Equação 3. Quando  $\text{Lift}(A \Rightarrow B) = 1$ , significa que A e B são independentes, ou seja, não existe associação entre eles. Se  $\text{Lift}(A \Rightarrow B) > 1$ , então A e B são positivamente dependentes. Se  $\text{Lift}(A \Rightarrow B) < 1$ , então, a regra é enganosa e A e B são negativamente dependentes.

$$\text{Lift}(A \Rightarrow B) = \frac{\text{Conf}(A \Rightarrow B)}{\text{Sup}(B)} \quad (3)$$

- **Count:** Representa a frequência da ocorrência de um determinado conjunto de itens.
- **Coefficiente de correlação de Pearson (r):** É uma medida de associação bivariada (força) do grau de relacionamento entre duas variáveis. O coeficiente de correlação

*Pearson (r)* varia de -1 a 1. O sinal indica a direção da correlação (negativa ou positiva), enquanto que o valor indica a magnitude. Uma correlação positiva ( $r=1$ ), por exemplo, indica que quando X aumenta, Y também aumenta, ou seja, valores altos de X estão associados a valores altos de Y (Paranhos et al., 2014).

- **Qui-Quadrado ( $X^2$ ):** É um teste não paramétrico utilizado para avaliar a correlação entre vários itens. Nas regras de associação, é utilizado para testar a independência entre os itens das regras (Campos, 2018). O valor crítico da distribuição Qui-Quadrado, com 1 grau de liberdade (tabela de contingência 2x2), em  $\alpha= 0,05$ , é 3,84; quanto maior o valor do Qui-Quadrado, mais provável é a correlação das variáveis. O nível de significância é indicado pelo valor do *p-value*, um *p-value* abaixo de 0,05 indica que há dependência entre os itens (Wu et al., 2016).

#### 4.4 Definição do Ranking de Perigosidade

Um grande problema encontrado na análise criminal é quantificar quanto um determinado local é mais perigoso que outro. Isso torna-se ainda mais problemático quando envolvemos diversos tipos de crimes dentro da mesma análise. Diariamente, diferentes regiões são alvos de diversos tipos de crimes e em várias proporções. Então, como é possível determinar que uma determinada região “A” é mais perigosa que uma região “B”?

Neste contexto, (de Farias et al., 2018) criaram um *ranking* de perigosidade para os bairros de uma cidade brasileira chamada Mossoró, localizada no Estado do Rio Grande do Norte. Com o uso dos dados criminais combinado com uma fórmula pré-definida, eles definiram um grau de perigosidade para cada bairro da cidade. Para a definição dessa fórmula, os autores ponderaram cada tipo de crime, a fim de permitir uma comparação entre os diferentes bairros. Esta ponderação foi definida usando o bom senso e a opinião de especialistas. Por exemplo, é senso comum que um homicídio é um crime mais grave que um roubo. A Tabela 1 mostra os pesos que foram definidos para cada crime.

**Tabela 1.** Pesos dos crimes.

Tipo de crime	Peso
Furto	1,0
Roubo	2,0
Homicídio	3,0
Tráfico de drogas	4,0

Fonte: de Farias et al., (2018).

Baseada nos pesos acima, o Índice de Criminalidade (IC) para cada bairro foi calculado por meio de uma média ponderada, ou seja, o somatório do produto entre o peso e o número de incidentes de cada tipo de crime, dividido pela soma dos pesos. A Equação 4 mostra a fórmula utilizada, onde  $n$  representa o número de tipos de crime,  $P_i$  o valor do peso e  $N_i$  o total de ocorrências daquele tipo de crime.

$$IC = \frac{\sum_{i=1}^n P_i \times N_i}{\sum_{i=1}^n P_i} \quad (4)$$

Após o computo do IC, o Índice de Criminalidade Normalizado (ICN) foi obtido, utilizando o processo de normalização no intervalo [0,1] (Lima et al., 2020). Então, os bairros foram classificados do menos para o mais perigoso, onde ICN=0 significa o menos perigoso e ICN=1 representa o mais perigoso. Porém, a fim de permitir uma classificação mais interpretável e clara, os autores transformaram o ICN, que são representados por valores contínuos, em níveis de perigosidade representados por valores categóricos. A definição desses níveis foi realizada da seguinte forma:

- Não Perigoso:  $0,00 \leq ICN < 0,15$ ;
- Pouco Perigoso:  $0,15 \leq ICN < 0,30$ ;
- Perigoso:  $0,30 \leq ICN < 0,50$ ;
- Muito Perigoso:  $0,50 \leq ICN \leq 1,00$

Em nosso trabalho, foi utilizada a mesma estratégia adotada por (de Farias et al., 2018) para encontrar os níveis de perigosidade. Entretanto, a fim de obter resultados mais precisos, alguns ajustes pontuais no processo foram realizados. Primeiramente, percebe-se que a Equação 4 utiliza o número de incidentes criminais. De certa forma, isso favorece regiões (bairros) menos populosas, pois locais com maior concentração populacional podem ter mais crimes. Logo, locais mais populosos poderiam alcançar os maiores IC por este fator e não necessariamente pelos níveis de segurança e criminalidade. Para mitigar essa ameaça, foi considerada a estimativa populacional anual no cálculo do IC. Ao invés de utilizar diretamente o número de ocorrências criminais, foi utilizada a taxa de crimes por 100.000 habitantes. Este procedimento atenuará as discrepâncias existentes, trazendo todas as regiões analisadas para o mesmo patamar, ao diminuir o potencial de confundimento com a população (fator de confusão). A Equação 5 representa o cálculo para encontrar a taxa de crimes por 100.000 habitantes ( $TC_{HAB}$ ). A estimativa populacional anual de cada estado foi obtida no portal do Instituto Brasileiro de Geografia e Estatística (IBGE), localizado no site

<https://ibge.gov.br>.

$$TC\_HAB = \frac{\text{Número de crimes}}{\text{Estimativa Populacional}} \times 100.000 \quad (5)$$

A Equação 6 mostra a Equação 4 adaptada, utilizando a taxa de crimes por 100.000 habitantes em seu computo.

$$IC = \frac{\sum_{i=1}^n P_i \times TC\_HAB_i}{\sum_{i=1}^n P_i} \quad (6)$$

Um outro ponto ajustado foram as faixas dos níveis de criminalidade. Diferente do trabalho de (de Farias et al., 2018), foram adotadas cinco faixas de perigosidade (Baixíssimo, Baixo, Intermediário, Alto ou Altíssimo), com o intuito de também mensurar um nível intermediário. Neste contexto, é importante ressaltar que o ICN sempre indicará o nível de perigosidade relativo ao estado de Minas Gerais e ao ano analisado, ou seja, é um *ranking* de perigosidade de Minas Gerais, pois, ainda que os municípios obtivessem níveis de criminalidade desejáveis e compatíveis com os lugares mais seguros do mundo, sempre haverá um município ocupando a posição Altíssimo, representando a pior posição em relação ao resto do estado. Desta forma, os níveis adotados por este artigo foram:

- Baixíssimo:  $0,00 \leq ICN \leq 0,15$ ;
- Baixo:  $0,15 < ICN \leq 0,40$ ;
- Intermediário:  $0,40 < ICN \leq 0,60$ ;
- Alto:  $0,60 < ICN \leq 0,85$ ;
- Altíssimo:  $0,85 < ICN \leq 1,00$ .

Com intuito de otimizar o *ranking* aqui desenvolvido, vale ressaltar que quatro especialistas em segurança pública foram consultados, para ponderar os novos tipos de crime e calibrar os pesos existentes descritos na Tabela 1. Os especialistas trouxeram as visões da Polícia, do Ministério Público, do Judiciário e dos advogados, com as perspectivas dos crimes de um Coordenador de um Grupo de Combate ao Crime Organizado, de um Promotor, de um Juiz e de um Advogado Criminalista. Os desempates do debate consideraram sempre os intervalos das penas, sem majorantes. Desta forma, para que não houvesse arbitrariedade nos pesos, estes intervalos foram considerados.

O debate dos pesos dos crimes é amplo e profundo, o qual não tem e talvez nunca terá uma conclusão simples, uma vez que as penas nem sempre refletirão a costumaz gravidade de um crime e as opiniões divergem veementemente quando os crimes lidam com vítimas fatais.

Neste sentido, apenas a realização de um Survey com juristas e especialistas em segurança pública de todo o Brasil poderá estimar o que a sociedade espera e como considera os níveis de gravidade dos crimes, coletando as medianas de pesos atribuídos pelos entrevistados e ampliando a discussão sobre o que pode ser considerada uma localidade mais perigosa e sobre as penas dos crimes. Isto também poderá embasar o poder público e os legisladores, em busca de um código penal que reflita melhor o clamor da sociedade brasileira. A árdua e complexa pesquisa de campo será um trabalho futuro do grupo de pesquisa ao qual os autores deste artigo pertencem.

Em linhas gerais, foram consideradas as médias dos intervalos das penas dos tipos de crimes, sem majorantes, bem como a uniformidade de distâncias entre os pesos (considerando uma aproximação proporcional às distâncias entre as penas). Pensando de maneira simplista, poderia ser considerado, para todos os crimes, apenas a primeira pena, sem majorantes, todavia, infelizmente, os dados abertos fornecidos pelo governo não tipificam detalhadamente os crimes. Em outras palavras, não são contemplados os casos em que o legislador criou um subtipo do crime, por entender que há uma maior reprovabilidade da conduta, ou seja, essa maior reprovabilidade, diferentemente da majorante, faz com que a pena base deixe de ser a do caput para iniciar-se já em um valor maior, análogo a uma nova tipificação. Para estes casos, foi usada a média do intervalo da pena do caput e a média dos intervalos das penas dos subtipos principais, com maior reprovabilidade. No caso de Homicídio, por exemplo, não é possível distinguir entre os Homicídios Culposos e Dolosos, ou seja, não é razoável considerar a menor pena, referente ao Homicídio Culposo.

Exemplificando, considerando as penas de Furto e Roubo, temos o intervalo de pena, para Roubo, sem majorantes, de 4 a 10 anos, bem como temos o intervalo de pena, para Furto, também sem majorantes, de 1 a 4 anos, e, para Furto Qualificado (maior reprovabilidade), de 2 a 8 anos ([http://www.planalto.gov.br/ccivil\\_03/decreto-lei/del2848compilado.htm](http://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm)). Considerando a média 7 do intervalo para Roubo e a média da média dos intervalos de Furto, de 3,75, temos uma razão de 1,87, aproximadamente 2, mesma razão/proporção utilizada entre os pesos de Furto Consumado e Roubo Consumado. Após essas validações, os novos tipos de crime e os pesos adotados estão descritos na Tabela 2.



**Tabela 2.** Pesos dos crimes utilizados neste trabalho.

Tipo de crime	Peso
Extorsão Tentado	1,0
Furto Consumado	1,0
Roubo Tentado	1,0
Sequestro e Cárcere Privado Tentado	1,0
Estupro Tentado	1,5
Lesão Corporal Consumado	1,5
Sequestro e Cárcere Privado Consumado	1,5
Estupro de Vulnerável Tentado	2,0
Extorsão Consumado	2,0
Homicídio Tentado	2,0
Roubo Consumado	2,0
Estupro Consumado	3,0
Estupro de Vulnerável Consumado	3,5
Extorsão Mediante Sequestro Consumado	3,5
Homicídio Consumado	4,0

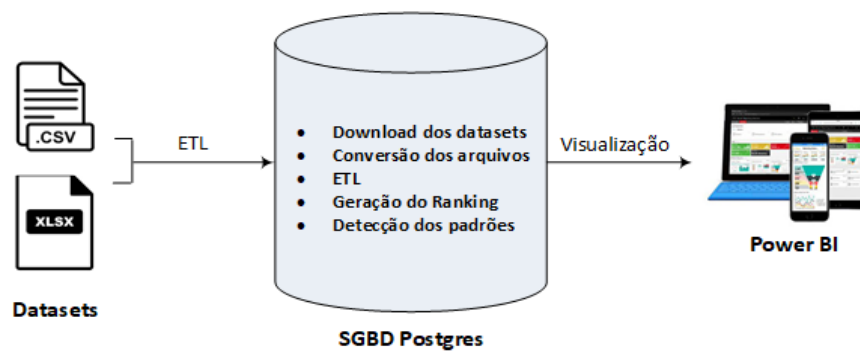
Fonte: Elaboração própria.

#### 4.5 Visão Geral da Arquitetura

O *PostgreSQL* é um Sistema de Gerenciamento de Banco de Dados (SGBD) que permite que as funções definidas pelo usuário sejam escritas em outras linguagens além de SQL e C. Estas linguagens são chamadas genericamente de Linguagens Procedurais (LPs). No caso de uma função escrita em uma Linguagem Procedural (LP), o servidor de banco de dados não possui nenhum conhecimento interno sobre como interpretar o texto do código fonte da função. Em vez disso, a tarefa é passada para um tratador especial que conhece os detalhes da linguagem. O próprio tratador pode fazer todo o trabalho (análise gramatical e sintática, execução, etc) ou pode servir como um “elo de ligação” entre o *PostgreSQL* e a implementação existente de uma linguagem de programação (PostgreSQL, 2019).

Para facilitar a obtenção, o tratamento, a manipulação e a análise dos dados realizados por esta pesquisa, uma arquitetura centralizada (unificada) foi desenvolvida, tendo como ponto principal o *PostgreSQL*. Dessa forma, procurou-se automatizar ao máximo os processos existentes, que vão desde a obtenção dos *datasets* até a detecção dos padrões criminais. Importante ressaltar que a arquitetura desenvolvida utilizou a ferramenta *Power BI* da *Microsoft* para a apresentação gráfica dos resultados. A Figura 2 exibe uma visão geral da arquitetura desenvolvida.

**Figura 2.** Visão Geral da Arquitetura.



Fonte: Elaboração própria.

Em nosso trabalho, três LPs foram utilizadas, são elas: *PL/PgSQL*, *PL/Python* e *PL/R*. A *PL/PgSQL* é uma LP nativa do *PostgreSQL*, a qual foi utilizada, principalmente, nos procedimentos de Extração, Transformação e Carga (ETL) dos dados e geração do *ranking* de perigosidade. Já a *PL/Python* permite a implementação de funções utilizando a linguagem *Python*, as quais poderão ser executadas dentro do *PostgreSQL*, sendo útil para realizar os *downloads* dos *datasets*, conversão, remoção e descompactação dos arquivos. E, por fim, a *PL/R*, a qual foi utilizada para detectar padrões de associação com o algoritmo *Apriori*, fazendo uso de códigos escritos na linguagem R.

## 5. Definição e Planejamento do Experimento

Nesta e nas duas próximas seções, este trabalho é apresentado como um processo experimental. O mesmo segue as diretrizes de Wohlin (Wohlin et al., 2012) e processos experimentais com publicações recentes (Santos et al., 2018) (Santos et al., 2020). Esta seção focará na definição do objetivo e planejamento do experimento.

### 5.1 Definição dos Objetivos

O objetivo principal deste experimento é analisar as possíveis associações entre municípios, Risps, tipos de crime, alvos de furto e alvos de roubo, utilizando os dados disponibilizados pelo governo do estado de Minas Gerais, por meio da Secretaria de estado de Justiça e Segurança Pública (Sejusp). Para atingi-lo, conduziu-se um experimento em ambiente controlado, no qual foram verificadas regras de associação (vide seção 4.3), utilizando o algoritmo *Apriori*, para determinação das combinações de itens que ocorrem com

determinada frequência, bem como para a geração das medidas de interesse que, estatisticamente, puderam servir de base para medir a força de tais regras. Adicionalmente, *rankings* de perigosidade dos municípios e das Risps também foram desenvolvidos.

Baseado no modelo GQM (*Goal Question Metric*) apresentado em (Basili & Weiss, 1984) (Basili et al., 2014), segue a formalização do objetivo desse trabalho: Analisar as ocorrências criminais no estado de Minas Gerais, com o propósito de avaliá-las, com respeito à detecção de padrões criminais, do ponto de vista de cientistas de dados, analistas criminais e cidadãos, no contexto de dados abertos da Sejusp/MG.

## 5.2 Planejamento

**Formulação das Hipóteses** - Não foram encontrados estudos experimentais que analisaram as possíveis associações entre municípios, Risps, tipos de crime, alvos de furto e alvos de roubo. Além disso, também não foram encontradas pesquisas científicas que descrevessem os níveis de perigosidade das Risps ou dos municípios mineiros. Baseadas nestas premissas, oito questões de pesquisa foram formalizadas para esse trabalho. Para as duas primeiras questões, **Q1** e **Q2**, serão realizadas análises estatísticas descritivas, enquanto que, para as questões **Q3**, **Q4**, **Q5**, **Q6**, **Q7** e **Q8** serão aplicados testes de significância das hipóteses. As questões de pesquisa são:

- **Q1:** Quais os municípios vêm se destacando como os mais perigosos?; **Q2:** Quais as Risps mais perigosas em 2019?; **Q3:** Há associações entre tipos de crime e municípios?; **Q4:** Há associações entre tipos de crime e Risps?; **Q5:** Há associações entre alvos de roubo e municípios?; **Q6:** Há associações entre alvos de roubo e Risps?; **Q7:** Há associações entre alvos de furto e municípios?; **Q8:** Há associações entre alvos de furto e Risps?.

Para responder às questões **Q3** a **Q8**, as seguintes hipóteses resumidas serão testadas individualmente, para combinação de antecedente e consequente:

- $H_0$ : Os tipos de crime, alvos de roubo e alvos de furto são independentes dos municípios e das Risps.
- $H_1$ : Os tipos de crime, alvos de roubo e alvos de furto são dependentes dos municípios e das Risps.

**Seleção do Contexto** – Para a realização do experimento, foram utilizados os conjuntos dos dados criminais de 853 municípios do estado de Minas Gerais.

**Seleção dos Participantes e Objetos** - Foram selecionados todos dados disponibilizados até o início dessa pesquisa. No total, 6 *datasets* foram obtidos, os quais foram distribuídos em três categorias de dados: “Ocorrências criminais”, “Alvos de roubo” e “Alvos de furto”. Cada uma dessas categorias é constituída por 2 conjuntos de dados. A categoria “Ocorrências criminais” é composta pelos dados de crimes violentos e crimes de outras naturezas. Já “Alvos de roubo” é formada pelas informações de roubos de veículos e roubos a outros alvos (estabelecimentos comerciais, residências, transporte coletivo, cargas e transeuntes). Por fim, o grupo “Alvos de furto” é formado pelos dados de furtos de veículos e furtos a outros alvos (estabelecimentos comerciais, residências, transporte coletivo, cargas e transeuntes). Tais dados englobam as informações do período de Janeiro de 2012 a Dezembro de 2019, as quais foram obtidas no portal de dados abertos da Secretaria de Estado de Justiça e Segurança Pública de Minas Gerais (Sejusp/MG), localizado em (Sejusp, 2020).

**Variáveis Dependentes** – As variáveis dependentes abordadas no experimento, para validação das hipóteses, foram as frequências dos conjuntos de itens analisados e as regras geradas, com seus Suportes e Confianças, das quais podem ser derivadas outras medidas de interesse objetivas para auxiliar na identificação das forças destas regras de associação: *Lift*, *r* (coeficiente de correlação de *Pearson*) e Qui-Quadrado ( $X^2$ ), com seu nível de significância (*p-value*).

**Variáveis Independentes** - Para este trabalho, foram consideradas, como variáveis independentes: o conjunto de registros compilados e disponibilizados em arquivos, contendo os dados dos incidentes criminais ocorridos nos municípios mineiros, o algoritmo *Apriori* utilizado, bem como Suporte e Confiança mínimos, intervalos aceitáveis de *Lift* e *p-value* máximo.

**Instrumentação** - O processo de instrumentação teve início com a configuração do ambiente para o experimento, planejamento de coleta de dados, construção de ETL (programa de Extração, Transformação e Carga) e o desenvolvimento dos algoritmos necessários. Os materiais/recursos utilizados foram:

- Arquivo com os incidentes criminais, disponibilizado pela Sejusp/MG; Arquivo com as estimativas populacionais dos municípios, disponibilizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE); Arquivos com o *script* de criação do projeto de banco de dados; Banco de Dados *PostgreSQL*, versão 11.6-3, para armazenar os dados e realizar o processo de ETL, com o uso da linguagem nativa *PL/PgSQL*; *Python*, versão 3.7.6; Software livre R, versão 3.6.0 e a Ferramenta *Power BI*.

## 6. Operação do Experimento

### 6.1 Execução

Os *datasets* disponibilizados pela Sejusp/MG estão sumarizados, ou seja, cada registro informa a quantidade de incidentes criminais ocorridos em um determinado mês, município e tipo de crime. Dessa forma, foi necessário realizar o desmembramento dessas informações, para que cada registro representasse apenas um crime, ou seja, passando a configurar uma transação, a qual é exigida neste formato, no R, e pôde ser avaliada pelo algoritmo *Apriori*. O nome transação representa a ligação existente em cada ocorrência de itens, no nosso caso, a ligação ocorre de forma geográfica e temporal, quando crimes ocorrem em uma determinada cidade, em determinado período. Por exemplo, para a avaliação das associações significativas entre Municípios e Alvo de Roubo, cada linha da nova tabela anual criada possui um par atômico de itens, tal como em (Belo Horizonte, Transeunte), representando cada crime ocorrido em um determinado ano.

De posse dessas transações, o *Apriori* verifica os pares que ocorrem de forma frequente no banco de dados. Essas associações frequentes possuirão um número de ocorrências, em relação ao número total de transações do banco de dados, maior ou igual ao suporte mínimo. Em seguida, a associação será considerada válida, se o seu número de ocorrências (Belo Horizonte, Transeunte), em relação ao número de vezes que o antecedente (Belo Horizonte) ocorrer, for superior à confiança mínima. Por fim, se o consequente é algo comum a todos e gerar regras enganosas, estas serão excluídas pela medida de interesse (*Lift*), pois possuirão um *Lift* menor que zero. Todas essas etapas, bem como a apresentação dos resultados em painéis gráficos na *web* foram automatizadas.

#### 6.1.1 Coleta de Dados

Após a execução, com o intuito de facilitar a coleta e análise dos dados, as informações foram publicadas em uma ferramenta gráfica (*Power BI*). Utilizando os diversos recursos deste *framework*, gráficos e tabelas foram construídos, a fim de elucidar as questões de pesquisa propostas.

Com relação às RAs, para cada associação, foi gerado um conjunto de medidas de interesse. Por se tratar de mineração em uma base de dados real, o número de regras gerados foi relativamente alto, como corroborado pelos ensaios apresentados em (Zheng et al., 2001).

Além disso, grande parte destes resultados minerados costuma ser composta por regras óbvias, redundantes ou, até mesmo, contraditórias. Para filtrar as regras interessantes para o estudo, uma confiança mínima de 70% foi adotada, para todas as categorias de dados analisadas (“Ocorrências criminais”, “Alvos de roubo” e “Alvos de furto”). Este valor foi adotado a partir de uma analogia com os intervalos considerados forte (0,70 - 0,89) e muito forte (0,90 - 1,00), para uma correlação (Guimarães et al., 2016) (Chaves & Shimizu, 2018). Assim, analogamente, a confiança mínima de 70% permitirá a geração de regras com confianças fortes e muito fortes.

Em se tratando de suporte mínimo, devido à diferença do número de transações, foram adotados valores individuais para cada categoria. Como visto na Equação 1, o valor do suporte de uma RA é dado pela relação entre a quantidade de transações, nas quais aparecem os itens A e B, e o total de transações. Para o cálculo, adotamos o valor aproximado (para cima) de uma amostra com população infinita, considerando uma margem de erro de 3,5% e 95% de confiabilidade, o que totaliza 784 ocorrências. Desta forma, a quantidade mínima de transações, nas quais aparecem os itens A e B (crime e município, por exemplo), deve ser de 800 (arredondamento para cima), para o ano em que houve menos transações, ou seja, para a categoria “Ocorrências criminais”, cuja a menor quantidade de transações ocorreu em 2019, 405.692 transações, o suporte mínimo foi obtido por meio da divisão,  $800/405.692$ , resultando, aproximadamente, no valor de 0.0019 (0,19%). Nos outros anos, com mais transações, a margem de erro será ainda menor, uma vez que o percentual aplicado foi o mesmo.

De forma semelhante, para a categoria de “Alvos de roubo”, cuja a menor quantidade de transações foi de 49.502, em 2019, o suporte mínimo foi obtido por meio da divisão,  $800/49.502$ , resultando, aproximadamente, no valor de 0.0161 (1,61%). E, finalmente, para o conjunto de dados “Alvos de furto”, no qual a menor quantidade de transações ocorreu, também, em 2019, o suporte mínimo foi obtido por meio da divisão,  $800/187.188$ , resultando, aproximadamente, no valor de 0.0042 (0,42%).

Para o *Lift*, o valor deverá estar nas seguintes faixas:  $Lift > 1$  (significando que existe uma dependência positiva) ou  $0 < Lift < 0.1$  (indicando que existe uma dependência negativa considerável). Regras foras destes limites serão consideradas enganosas e descartadas, pois indicam que, em verdade, o antecedente diminui a probabilidade do conseqüente ocorrer, ainda que numa pequena proporção.

## 6.1.2 Validação dos Dados

Para assegurar a análise, interpretação e validação dos resultados das regras de associação, foram aplicados os testes de significância das hipóteses, utilizando o teste do Qui-Quadrado ( $X^2$ ). Além disso, também foi calculado o coeficiente de correlação de Pearson ( $r$ ).

## 7. Análise dos Resultados

### 7.1 Resultados Brutos

Como exposto anteriormente, foi feita uma avaliação que contempla os incidentes criminais de todos os municípios do estado de Minas Gerais. Devido ao espaço reduzido e ao alto número de municípios que Minas Gerais contém (853 municípios), serão apresentadas apenas as evoluções das taxas de criminalidade (de 2012 a 2019) dos municípios que obtiveram as 10 maiores taxas, no ano de 2019. Como foi descrito, o valor desta taxa foi calculado, utilizando a estimativa populacional de cada localidade, a fim de ser proporcional a 100.000 habitantes, utilizando a Equação 5.

**Tabela 3.** Dez municípios com as maiores taxas de criminalidade por 100.000 habitantes em 2019 (ordenados da maior para a menor).

Município	2012	2013	2014	2015	2016	2017	2018	2019	Varição (2018-2019)
Belo Horizonte	4092,19	4267,93	4353,28	4664,63	4940,54	4676,13	4103,93	3677,09	-10,40%
Água Comprida	1538,46	1884,06	2806,00	2228,68	3105,29	3449,95	3192,02	3301,65	3,43%
Confins	5479,68	6755,05	6537,68	6143,87	6157,37	4963,68	4581,64	3239,23	-29,30%
Buritit	3715,73	4074,40	4315,45	3108,70	3090,85	3272,71	2846,37	3176,20	11,59%
Uberaba	3468,67	3794,08	4235,08	4360,72	4429,43	4097,82	3293,67	3055,88	-7,22%
Campos Altos	2004,72	2345,63	2380,95	2258,66	3453,46	3015,53	2897,89	2923,48	0,88%
Juatuba	2959,27	3121,01	3596,63	4049,91	4876,42	3806,91	3190,61	2905,81	-8,93%
Delfinópolis	2198,28	2156,14	1952,25	2393,95	2568,04	3145,00	3043,11	2881,64	-5,31%
Corinto	4538,39	4039,37	3393,71	3667,32	4109,48	3580,22	3475,23	2878,09	-17,18%
Rio Novo	1865,63	2285,59	2934,01	2099,45	2900,63	2653,02	2762,55	2815,96	1,93%

Fonte: elaboração própria.

A Tabela 3 apresenta a evolução dos 10 municípios com as maiores taxas de criminalidade. A tabela está ordenada, de forma **decrecente**, pelos valores das taxas de 2019. Como pode ser visto, apesar de Belo Horizonte atingir a maior taxa em 2019, nos demais anos (2012 a 2018), o município de Confins liderou o *ranking*. Esta mudança, ocorrida no último ano, pode ser justificada pela queda constante do número de crimes ocorridos em Confins, ao longo dos anos, e, principalmente, pela redução considerável da taxa de criminalidade (29,30%) alcançada em 2019. Além disso, excetuando Água Comprida, Buritit, Campos

Altos e Rio Novo, é notória a redução da taxa de criminalidade em 2019, relativa ao ano anterior, seguindo uma tendência geral de queda em 2018, em menores proporções que em 2019.

Neste contexto, os destaques vão para Confins, Corinto e Belo Horizonte. Nestes municípios, houve uma manutenção da tendência de queda, nos últimos três anos, sendo mais acentuada em 2019. Apesar da queda geral, uma análise detalhada destes resultados é interessante, considerando as ações que foram efetivadas, bem como se estas ações realmente trouxeram efeitos positivos mais significativos para estes municípios. O raciocínio inverso vale para os municípios de Água Comprida, Buritis, Campos Altos e Rio Novo. No Rio Novo, os valores sobem desde 2018, nas outras cidades, o aumento ocorreu em 2019.

A Tabela 4 apresenta, mais detalhadamente, a taxa de criminalidade, por tipo de crime, dos 10 municípios com as maiores taxas criminais em 2019. Vale a pena destacar que analisando apenas os homicídios e considerando somente as cidades listadas na Tabela 3, os municípios de Corinto e Juatuba assumiriam, respectivamente, a primeira e segunda colocações, merecendo, dessa forma, uma atenção especial por parte das políticas públicas do governo estadual. Adicionalmente, também vale ressaltar o alto número de estupros consumados, i. e., “Estupros Consumados” somados aos “Estupros de Vulneráveis Consumados”, ocorridos nas cidades de Juatuba, Campos Altos e Buritis, em 2019. Destacar os estupros é importante, uma vez que a discussão e o questionamento sobre as análises, que consideram a perspectiva única dos assassinatos, indagam a ausência, por exemplo, do registro da barbárie e da natureza hedionda deste tipo de crime. Os homicídios e estupros explicam a posição de Juatuba no *ranking* de perigosidade (vide Figura 3), levando-se em consideração ainda que um estupro pode resultar em morte e que, infelizmente, os dados disponíveis não detalham todos os subtipos de estupro. De qualquer forma, ainda que não tenha sido letal, o estupro pode mudar uma vida ou, metaforicamente, ceifá-la para sempre.



**Tabela 4.** Taxa dos tipos de crime, dos 10 municípios com as maiores taxas de criminalidade, no ano de 2019, proporcional a 100.000 habitantes.

Tipos de Crime	Municípios									
	Belo Horizonte	Água Comprida	Confins	Buritís	Uberaba	Campos Altos	Juatuba	Delfinópolis	Corinto	Rio Novo
Homicídio Consumado	1,43	0,00	0,00	0,00	0,90	0,00	3,71	0,00	4,21	0,00
Extorsão Mediante Sequestro Consumado	8,68	0,00	0,00	4,03	4,19	0,00	11,13	0,00	4,21	0,00
Estupro de Vulnerável Consumado	14,89	0,00	14,86	28,18	19,47	25,87	37,11	14,06	12,64	11,17
Estupro Consumado	37,98	0,00	14,86	32,20	28,76	51,74	44,53	14,06	16,86	22,35
Roubo Consumado	207,76	100,05	118,87	213,36	379,29	245,78	322,87	70,28	181,20	100,57
Homicídio Tentado	0,88	0,00	0,00	0,00	0,90	0,00	0,00	0,00	0,00	0,00
Extorsão Consumado	10,63	0,00	0,00	4,03	5,69	6,47	14,84	0,00	8,43	0,00
Estupro de Vulnerável Tentado	14,65	0,00	14,86	24,15	11,68	12,94	25,98	0,00	12,64	0,00
Sequestro e Cárcere Privado Consumado	0,80	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Lesão Corporal Consumado	676,37	1200,60	297,18	382,43	420,93	763,21	653,16	702,84	741,65	502,85
Estupro Tentado	14,09	0,00	14,86	4,03	6,89	6,47	22,27	0,00	12,64	0,00
Sequestro e Cárcere Privado Tentado	0,08	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Roubo Tentado	0,88	0,00	0,00	0,00	0,60	0,00	0,00	0,00	0,00	0,00
Furto Consumado	2684,76	2001,00	2763,74	2483,80	2173,57	1811,01	1766,50	2080,40	1879,40	2179,01
Extorsão Tentado	3,22	0,00	0,00	0,00	3,00	0,00	3,71	0,00	4,21	0,00

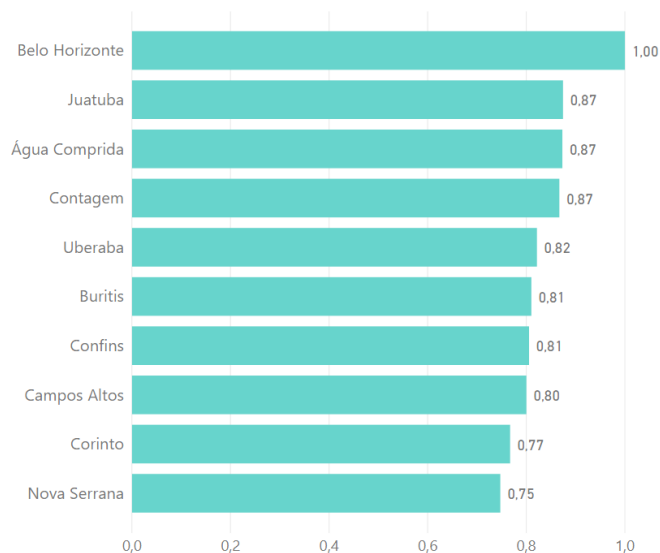
Fonte: elaboração própria.

## 7.2 Análise e Interpretação do Ranking de Perigosidade

Com relatado na seção 4.4, um grande problema encontrado na análise criminal é quantificar quanto um determinado local é mais perigoso que outro. Isso torna-se ainda mais problemático quando envolvemos diversos tipos de crimes dentro da mesma análise. Para responder quais os municípios e Risps vêm se destacando como os mais perigosos (questões **Q1** e **Q2**), um Índice de Criminalidade (IC) foi criado, baseado em uma média ponderada, calculada a partir da taxa da criminalidade (proporcional à população) e dos pesos adotados para os diversos tipos de crimes. Posteriormente, para permitir uma classificação mais interpretável e clara, o IC foi normalizado e reenquadrado em classes categóricas.

### 7.2.1. Ranking de Perigosidade dos Municípios

**Figura 3.** 10 municípios com os maiores Índices de Criminalidade Normalizados (ICN), em 2019.



Fonte: Elaboração própria.

A Figura 3 apresenta os 10 maiores Índices Normalizados, no ano mais recente (2019). No nível altíssimo, enquadraram-se Belo Horizonte, Juatuba, Água Comprida e Contagem. Ao confrontarmos esses dados com a Tabela 3, a qual apresenta os 10 municípios com as maiores taxas de criminalidade por 100.000 habitantes em 2019, notam-se algumas alterações, evidenciando que nem sempre os locais que alcançaram altas taxas de criminalidade, apresentaram, também, segundo os pesos adotados, altos níveis de perigosidade. Por esta nova análise, o município de Confins, terceiro local com a maior taxa criminal em 2019, foi classificado como sétimo mais perigoso e seu nível obtido foi “Alto”. Por outro lado, o município de Contagem, que não aparece entre os 10 municípios com as maiores taxas em 2019, foi classificado como quarto mais perigoso, atingindo o nível “Altíssimo”.

O ICN indica um *ranking*, uma ordem de perigosidade, relativa aos municípios e ao ano analisado (vide seção 4.4), podendo indicar, por exemplo, que as políticas públicas do governo estadual devem considerar uma atenção maior para Belo Horizonte, Juatuba, Água Comprida e Contagem, i.e., cidades com grau “Altíssimo”, cumulativas às desenvolvidas para 2019. A classificação de um município, por exemplo, em um ano como alto e em outro ano como altíssimo (vide Tabela 5) não implica um aumento absoluto no nível de perigosidade. Esta troca de posição pode indicar, apenas, uma menor distância para o primeiro colocado daquele ano, aumentando seu nível de perigosidade relativo ao resto do estado. No entanto,

essa concorrência salutar só beneficia o cidadão e deve estimular prefeitos na busca pela liderança das cidades mais seguras.

**Tabela 5.** Detalhamento anual dos municípios mais perigosos (Top 5).

Ano	Total de municípios por nível	Município – ICN (Nível)
2012	Altíssimo : 6	Confins - 1,00 (Altíssimo)
	Alto : 35	Carmo do Paranaíba - 0,95 (Altíssimo)
	Intermediário : 127	Corinto - 0,90 (Altíssimo)
	Baixo : 472	Belo Horizonte - 0,89 (Altíssimo)
	Baixíssimo : 213	Pirapora - 0,88 (Altíssimo)
2013	Altíssimo : 1	Confins - 1,00 (Altíssimo)
	Alto : 12	Belo Horizonte - 0,74 (Alto)
	Intermediário : 64	Carmo do Paranaíba - 0,72 (Alto)
	Baixo : 489	Contagem - 0,67 (Alto)
	Baixíssimo : 287	Corinto - 0,64 (Alto)
2014	Altíssimo : 1	Confins - 1,00 (Altíssimo)
	Alto : 10	Belo Horizonte - 0,82 (Alto)
	Intermediário : 95	Contagem - 0,77 (Alto)
	Baixo : 477	Uberaba - 0,75 (Alto)
	Baixíssimo : 270	Juatuba - 0,71 (Alto)
2015	Altíssimo : 3	Confins - 1,00 (Altíssimo)
	Alto : 25	Belo Horizonte - 0,94 (Altíssimo)
	Intermediário : 101	Contagem - 0,89 (Altíssimo)
	Baixo : 487	Perdigão - 0,85 (Alto)
	Baixíssimo : 237	Juatuba - 0,85 (Alto)
2016	Altíssimo : 5	Confins - 1,00 (Altíssimo)
	Alto : 25	Juatuba - 0,97 (Altíssimo)
	Intermediário : 107	Belo Horizonte - 0,96 (Altíssimo)
	Baixo : 478	Contagem - 0,92 (Altíssimo)
	Baixíssimo : 238	Perdigão - 0,86 (Altíssimo)
2017	Altíssimo : 6	Belo Horizonte - 1,00 (Altíssimo)
	Alto : 29	Conceição do Pará - 0,96 (Altíssimo)
	Intermediário : 150	Nova Serrana - 0,94 (Altíssimo)
	Baixo : 494	Contagem - 0,94 (Altíssimo)
	Baixíssimo : 174	Confins - 0,94 (Altíssimo)
2018	Altíssimo : 3	Confins - 1,00 (Altíssimo)
	Alto : 34	Belo Horizonte - 0,99 (Altíssimo)
	Intermediário : 191	Contagem - 0,96 (Altíssimo)
	Baixo : 507	São Joaquim de Bicas - 0,84 (Alto)
	Baixíssimo : 118	Juatuba - 0,83 (Alto)
2019	Altíssimo : 4	Belo Horizonte - 1,00 (Altíssimo)
	Alto : 49	Juatuba - 0,87 (Altíssimo)
	Intermediário : 192	Água Comprida - 0,87 (Altíssimo)
	Baixo : 518	Contagem - 0,87 (Altíssimo)
	Baixíssimo : 90	Uberaba - 0,82 (Alto)

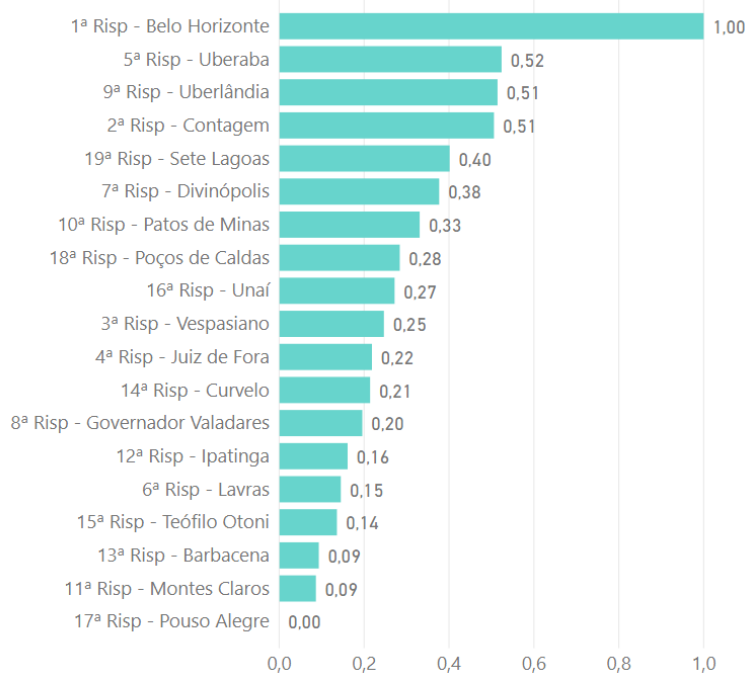
Fonte: Elaboração própria.

A Tabela 5 detalha, para cada ano, a quantidade de municípios por nível de perigosidade e os cinco mais perigosos, com seus respectivos índices de criminalidade normalizados. Em resposta à questão **Q1**, é importante destacar que Belo Horizonte esteve presente, entre os cinco mais perigosos, em todos os períodos avaliados, sendo classificado no grau “Altíssimo”, em seis dos oito anos. Além disso, os municípios de Confins e Contagem foram classificados entre os cinco mais perigosos, em sete dos oito anos. Uma atenção

especial também para Juatuba, presente entre os mais perigosos, em cinco dos últimos 6 anos.

### 7.2.2. Ranking de Perigosidade das Risps

**Figura 4.** Índice de criminalidade normalizado das Risps em 2019.



Fonte: elaboração própria.

Em resposta à **Q2**, o gráfico da Figura 4 detalha os índices normalizados, obtidos por cada Risp, no ano de 2019. Nota-se que a região “1ª Risp - Belo Horizonte”, única Risp a ser classificada no nível “Altíssimo” (ICN=1,00), superou consideravelmente as demais regiões, já que estas foram classificadas no grau “Intermediário” ou em níveis inferiores.

### 7.3 Análise e Interpretação das Regras de Associação

Esta seção tratará de responder as questões que envolvem a averiguação das regras de associação entre municípios, Risps, tipos de crime, alvos de furto e alvos de roubo: **Q3, Q4, Q5, Q6, Q7 e Q8**. Para realizar estas análises, as transações utilizadas em cada uma das questões foram agrupadas por ano e, posteriormente, analisadas separadamente pelo algoritmo de regra de associação (vide seção 4.3). Entretanto, é preciso que analisemos sob a luz da estatística, para encontrarmos evidências que corroborem, ou não, com os resultados encontrados. Para isto, foi aplicado, em cada questão de pesquisa, um teste de significância da hipótese, utilizando um nível de confiança igual a 95% ( $\alpha=0,05$ ).

### 7.3.1 Associações entre Tipos de Crime e Municípios/Risps

No total, 136 associações foram encontradas entre “Municípios  $\Rightarrow$  Tipos de Crimes”, as quais estão distribuídas, por ano, da seguinte forma: 2012 (19), 2013 (19), 2014 (13), 2015 (12), 2016 (15), 2017 (15), 2018 (20) e 2019 (23). Vale ressaltar que todas as associações encontradas atenderam aos requisitos mínimos exigidos pelo experimento (vide seção 6.1.1). Adicionalmente, ficou constatado que as associações “Unaí  $\Rightarrow$  Furto Consumado”, “Ouro Preto  $\Rightarrow$  Furto Consumado” e “Araguari  $\Rightarrow$  Furto Consumado” foram encontradas nos oito anos analisados.

**Tabela 6.** Regras encontradas para associações entre municípios e tipos de crime, em 2019.

Regra	supp	conf	lift	count	R	X <sup>2</sup>	p-value do X <sup>2</sup>
Unaí $\Rightarrow$ Furto Consumado	0,0030	0,78	1,16	1.217	0,014	80,03	0,00
Alfenas $\Rightarrow$ Furto Consumado	0,0039	0,77	1,15	1.590	0,015	93,39	0,00
Poços De Caldas $\Rightarrow$ Furto Consumado	0,0075	0,76	1,12	3.037	0,018	126,84	0,00
Ouro Preto $\Rightarrow$ Furto Consumado	0,0026	0,75	1,11	1.061	0,009	34,97	0,00
Uberlândia $\Rightarrow$ Furto Consumado	0,0347	0,74	1,10	14.081	0,031	394,86	0,00
Cataguases $\Rightarrow$ Furto Consumado	0,0027	0,74	1,10	1.081	0,008	28,41	0,00
Itaúna $\Rightarrow$ Furto Consumado	0,0038	0,73	1,09	1.551	0,010	36,91	0,00
Araguari $\Rightarrow$ Furto Consumado	0,0044	0,73	1,09	1.797	0,010	38,98	0,00
Governador Valadares $\Rightarrow$ Furto Consumado	0,0120	0,73	1,09	4.850	0,016	103,27	0,00
Belo Horizonte $\Rightarrow$ Furto Consumado	0,1662	0,73	1,09	67.443	0,067	1.824,80	0,00
Patos de Minas $\Rightarrow$ Furto Consumado	0,0066	0,73	1,09	2.691	0,012	56,24	0,00
Ipatinga $\Rightarrow$ Furto Consumado	0,0082	0,73	1,09	3.338	0,013	68,98	0,00
Nova Lima $\Rightarrow$ Furto Consumado	0,0031	0,72	1,08	1.260	0,007	21,11	0,00
São Sebastião do Paraíso $\Rightarrow$ Furto Consumado	0,0026	0,72	1,07	1.065	0,006	14,58	0,00
Ituiutaba $\Rightarrow$ Furto Consumado	0,0036	0,72	1,07	1.470	0,007	18,31	0,00
Araxá $\Rightarrow$ Furto Consumado	0,0042	0,72	1,06	1.718	0,007	20,36	0,00
Uberaba $\Rightarrow$ Furto Consumado	0,0179	0,71	1,06	7.255	0,013	72,69	0,00
Varginha $\Rightarrow$ Furto Consumado	0,0041	0,71	1,06	1.650	0,006	16,15	0,00
Pirapora $\Rightarrow$ Furto Consumado	0,0021	0,71	1,06	860	0,004	7,96	0,00
Itajubá $\Rightarrow$ Furto Consumado	0,0024	0,71	1,05	970	0,004	7,16	0,01
Pará de Minas $\Rightarrow$ Furto Consumado	0,0030	0,70	1,05	1.215	0,004	7,75	0,01
Passos $\Rightarrow$ Furto Consumado	0,0048	0,70	1,05	1.955	0,005	12,08	0,00
Pouso Alegre $\Rightarrow$ Furto Consumado	0,0045	0,70	1,04	1.809	0,005	10,29	0,00

Fonte: Elaboração própria.

Em razão do espaço reduzido, a Tabela 6 apresenta as associações detectadas para o ano de 2019. Percebe-se que todas estão associadas ao crime “Furto Consumado”, entre as quais se destacaram as regras dos municípios Unaí, Alfenas e Poços de Caldas, alcançando uma confiança superior a 75%. **Apesar de não necessariamente implicar causa**, essas regras coadunam com seus níveis de perigosidade mais baixos, uma vez que furto possui o menor peso para perigosidade. Além disso, como já publicado na literatura (Ervilha & Lima, 2019), em Minas Gerais, crimes menos violentos, tais como furtos, tendem a estar mais associados a

menores oportunidades de empregos para os jovens, menores gastos com segurança pública e um efetivo menor do policiamento militar em relação à população. **Estas associações foram avaliadas por meio da modelagem econométrica de dados em painel**, além disto, vale ressaltar mais uma vez que as regras aqui apresentadas não implicam causa e efeito, ou seja, todas os dados aqui apresentados são apenas contextuais às regras ou aos índices, sem implicar conclusões definitivas, mas apenas um embasamento maior para a tomada de decisão em segurança pública.

Nesta mesma linha, essas regras podem contribuir para a priorização das políticas públicas (Silvares, 2019) para estes municípios, bem como para o planejamento de experimentos e planos pilotos em locais com padrões mais explícitos. **Em algumas situações, a segurança pode imitar os negócios, nos quais as ações imediatas já partem do “O quê” para, depois, entender o “Porquê”.**

Finalmente, com relação às associações “Risps  $\Rightarrow$  Tipos de Crimes”, não foram encontradas RAs para os anos de 2014 e de 2017. Na Tabela 7, as informações encontradas para as associações são detalhadas, juntamente com os valores de suas medidas de interesse.

**Tabela 7.** Regras encontradas para associações entre Risps e tipos de crime.

Ano	Regra	supp	conf	lift	count	R	X <sup>2</sup>	p-value do X <sup>2</sup>
2012	16° Risp - Unaí $\Rightarrow$ Furto Consumado	0,0155	0,72	1,12	7.242	0,024	272,4	0,00
2013	16° Risp - Unaí $\Rightarrow$ Furto Consumado	0,0148	0,70	1,11	7.211	0,022	234,52	0,00
2015	9° Risp - Uberlândia $\Rightarrow$ Furto Consumado	0,0400	0,71	1,18	20.977	0,054	1.525,53	0,00
2016	17° Risp - Pouso Alegre $\Rightarrow$ Furto Consumado	0,0174	0,70	1,17	9.771	0,033	628,76	0,00
2016	9° Risp - Uberlândia $\Rightarrow$ Furto Consumado	0,0381	0,70	1,17	21.398	0,051	1.462,96	0,00
2018	16° Risp - Unaí $\Rightarrow$ Furto Consumado	0,0123	0,70	1,08	5.658	0,014	95,07	0,00
2018	9° Risp - Uberlândia $\Rightarrow$ Furto Consumado	0,0455	0,70	1,08	20.956	0,029	378,02	0,00
2019	16° Risp - Unaí $\Rightarrow$ Furto Consumado	0,0116	0,71	1,05	4.725	0,010	37,27	0,00
2019	18° Risp - Poços de Caldas $\Rightarrow$ Furto Consumado	0,0362	0,70	1,04	14.699	0,015	88,26	0,00
2019	1° Risp - Belo Horizonte $\Rightarrow$ Furto Consumado	0,1662	0,73	1,09	67.443	0,067	1824,80	0,00
2019	9° Risp - Uberlândia $\Rightarrow$ Furto Consumado	0,0474	0,73	1,08	19.210	0,031	384,67	0,00

Fonte: Elaboração própria.

Em resposta às questões **Q3** e **Q4** considerando as regras selecionadas pelas medidas de interesse adotadas, também ficou constatado que os *p-values* ficaram abaixo do nível de significância adotado ( $p\text{-value} < 0,05$ ). Dessa forma, para estas questões, a hipótese  $H_0$  pôde ser rejeitada, indicando que há uma dependência entre o antecedente e o consequente.

### 7.3.2 Associações entre Alvos de Roubo e Municípios/Risps

No contexto das associações entre “Municípios  $\Rightarrow$  Alvos de roubo”, 3 regras foram encontradas. A Tabela 8 detalha as informações encontradas para essas associações.

**Tabela 8.** Regras encontradas para associações entre municípios e alvos de roubo.

Ano	Regra	supp	conf	lift	count	r	X <sup>2</sup>	p-value do X <sup>2</sup>
2016	Montes Claros $\Rightarrow$ Transeunte	0,0204	0,70	1,27	2.535	0,052	332,06	0,00
2017	Montes Claros $\Rightarrow$ Transeunte	0,0189	0,70	1,27	2.022	0,052	281,24	0,00
2019	Belo Horizonte $\Rightarrow$ Transeunte	0,2180	0,71	1,25	10.792	0,189	1759,48	0,00

Fonte: Elaboração própria.

Além de serem municípios com alta atividade econômica, vale a pena destacar que a associação “Belo Horizonte  $\Rightarrow$  Transeunte”, ocorrida no último ano (2019), apresentou um suporte no valor aproximando de 21,80%, o que indica 10.792 ocorrências, de uma amostra de 49.502 transações. Além disso, o valor da confiança, ou seja, 0,71, confere-nos dizer que mais de 70% dos roubos do município de “Belo Horizonte” têm como alvo o “Transeunte”, pessoas em movimento pela capital mineira. Isto pode implicar necessidade de mudanças nas estratégias de policiamento e monitoramento por meios eletrônicos, como também pode implicar realocação de esforços de outros tipos de segurança, tal como a segurança patrimonial.

**Tabela 9.** Regras encontradas para associações entre Risps e alvos de roubo.

Ano	Regra	supp	conf	lift	count	r	X <sup>2</sup>	p-value do X <sup>2</sup>
2019	1ª Risp - Belo Horizonte $\Rightarrow$ Transeunte	0,2180	0,71	1,25	10.792	0,189	1.759,48	0,00

Fonte: Elaboração própria.

Por fim, com relação às associações “Risps  $\Rightarrow$  Alvos de roubo”, apenas uma associação foi encontrada, para o ano de 2019. A Tabela 9 detalha as informações encontradas para esta regra. A associação “1ª Risp - Belo Horizonte  $\Rightarrow$  Transeunte”, na Tabela 9, valida a associação “Belo Horizonte  $\Rightarrow$  Transeunte”, do ano de 2019, apresentada na Tabela 8, uma vez que a “1ª Risp - Belo Horizonte” é constituída apenas por este município.

Consequentemente, considerando as regras selecionadas pelas medidas de interesse

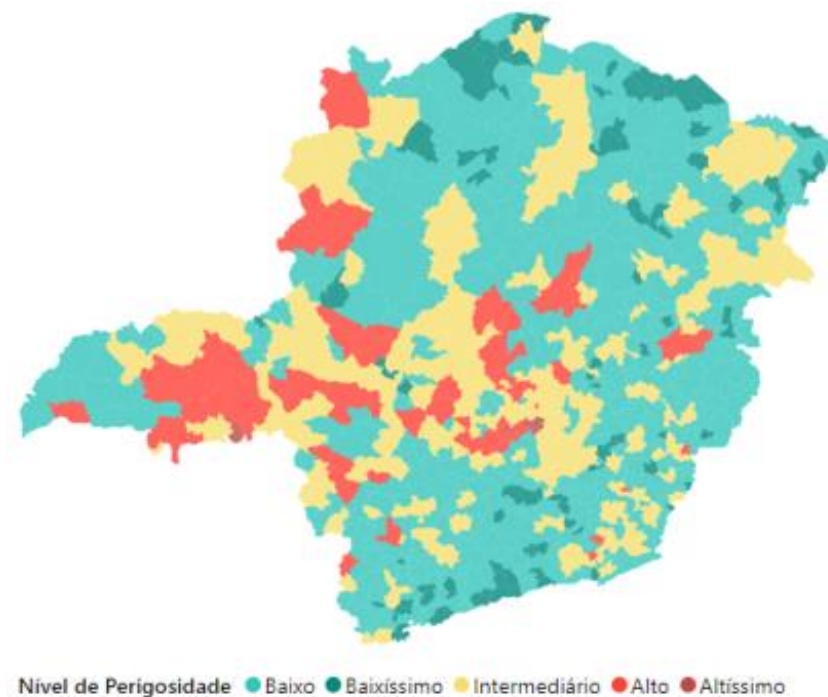
adotadas, também ficou constatado, para as questões **Q5** e **Q6**, que os *p-values* ficaram abaixo do nível de significância adotado ( $p\text{-value} < 0,05$ ). Dessa forma, para todas as regras aqui expostas, a hipótese  $H_0$  pôde ser rejeitada, indicando que há uma dependência entre o antecedente e o consequente.

### 7.3.3 Associações entre Alvos de Furto e Municípios/Risps

Não foram encontradas associações para “Municípios  $\Rightarrow$  Alvos de furto” e “Risps  $\Rightarrow$  Alvos de furto”. Logo, para as questões **Q7** e **Q8**, as hipóteses  $H_0$  não podem ser rejeitadas.

## 7.4. Análise Espacial e Socioeconômica

**Figura 5.** Índice de criminalidade por Localização em 2019.



Fonte: Elaboração própria.

Na Figura 5, os índices de perigosidade de 2019 podem ser vistos por município e sua localização geográfica. De imediato, é notória a confirmação de boa parte das evidências publicadas em (de Barros et al., 2019), as quais apontaram uma tendência de regiões com desenvolvimento econômico à baixa criminalidade, tais como as regiões de Juiz de Fora, Ipatinga e Montes Claros, todavia, que há casos em que o nível de desenvolvimento econômico não consegue conter o avanço do crime, tais como em Uberlândia e na grande



Belo Horizonte, cujos índices de perigosidade foram alto e altíssimo. Neste contexto, como os crimes contra pessoa têm uma influência maior no índice de perigosidade, em (Ervilha & Lima, 2019), usando modelagem de dados em painel, foram encontradas evidências de que, em Minas Gerais, as variáveis população de 15 a 24 anos (relação positiva com criminalidade), mortalidade padronizada (positiva), escolarização líquida (negativa), emprego formal (positiva) e razão de dependência (negativa) foram estatisticamente significativas em pelo menos 5%, bem como gasto *per capita* com assistência social (negativa), ao nível de 10%.

O índice de desenvolvimento econômico destes locais deveria refletir em investimentos na educação, políticas de proteção social (Silvares, 2019) e soluções para aumento da qualidade de vida e ocupação de jovens, os quais participam precocemente e são vítimas da criminalidade. Neste contexto, Carazza et al. (2020) evidenciaram que a adoção de uma política de toque de recolher juvenil reduziu a taxa de furto em 17,9%.

Fugindo um pouco às tendências já publicadas, uma vez que o nível de perigosidade não considera apenas os Homicídios, por exemplo, Chácara, vizinha a Juiz de Fora, alcançou nível de perigosidade alto, com destaque para as Lesões Corporais. Também chama a atenção as cidades do Noroeste do estado que alcançaram índices altos, Buritis e Paracatu, próximas ao Distrito Federal e ambas da Risp Unai, cuja a cidade de mesmo nome alcançou um índice intermediário. Essa tendência é clara no mapa, com cidades mais perigosas vizinhas a cidades com níveis intermediário e baixos, o que indica a necessidade de soluções preventivas contra o avanço da criminalidade e pode indicar um deslocamento da criminalidade para locais vizinhos, com maiores ofertas para prática criminal e/ou menor assistência social. Além disso, o município de Paracatu é o principal município da região, com presença de muitas empresas, mais empregos, escolas e mineração de ouro, o que atrai gente de toda a região e a presença de jovens, infelizmente, correlacionados com os crimes mais violentos. Por outro lado, Buritis, apesar de ser menor, chamou a atenção, conforme destacado anteriormente, pela alta taxa de Estupros, principalmente de vulneráveis. Para coibir os estupros, esta é uma cidade cujas particularidades municipais devem ser exploradas (Ervilha e Lima, 2019), além das variáveis que explicam o aumento da criminalidade já publicadas.

Na mesma linha, Água Comprida, localizada em uma região de grande desenvolvimento econômico, também chama atenção pelo seu índice Altíssimo. Em (Ervilha e Lima, 2019), também se verificou que os municípios com características intrínsecas não observadas e taxa de crimes contra pessoa positivamente relacionadas encontram-se, via de

regra, nas regiões com maiores desigualdades sociais. Neste sentido, por exemplo, na contramão, com melhor distribuição de renda, Água Comprida e Uberaba tiveram uma relação negativa entre os efeitos locais, fixos no tempo, não observáveis, e as taxas de crime. Isto implica que estas cidades podem concentrar-se em gerir melhor as variáveis explicativas já estudadas para a região de Minas Gerais. No caso de Água Comprida, é alarmante a taxa de Lesões Corporais Consumadas, sendo aproximadamente o dobro da taxa de Belo Horizonte (vide Tabela 4), para uma cidade que não teve nenhum Homicídio em 2019 e que já pertence a uma Risp com destaque para as lesões corporais. A administração pública pode analisar a viabilidade de mais políticas de proteção social e melhores investimentos em Educação.

Por fim, vale a pena situar Minas Geras em relação ao resto do país. De acordo com (Lima & Bueno, 2020), considerando o número de homicídios dolosos e um comparativo entre os primeiros semestres dos anos de 2019 e 2020, os estados do Amapá (0,0%) e Minas Gerais (0,6%) são destaques, com índices abaixo de 1%. No mesmo contexto, alguns estados se destacaram pela redução dos seus números. O estado de Roraima foi o que apresentou a maior redução, -31,1%. Em seguida, vieram Pará, com -24,5%, Goiás, com -16,6%, Rio de Janeiro, com -9,2%, e Distrito Federal, com -8,4%. Por outro lado, o Ceará mais que dobrou o número de homicídios, atingindo a maior majoração entre as Unidades Federativas, ou seja, 106,9%. Este índice pode ser justificado pelo fato de que este viveu uma crise de segurança pública no início de 2020, com a greve da Polícia Militar, a qual durou 13 dias, no mês de fevereiro. Essa crise teve impactos importantes nos indicadores da segurança pública estadual, no primeiro semestre.

### **7.5 Ameaças à validade**

De acordo com Chapetta (2006), ameaças à validade podem limitar a habilidade de interpretar e/ou descrever resultados dos dados obtidos em um experimento. Neste sentido, as seguintes ameaças, encontradas durante a experimentação, devem ser consideradas.

**Ameaças às validades de construção e interna** – Considerando que os dados foram obtidos, por meio de *download*, tratados e analisados pelos autores, existem ameaças a serem consideradas. Para mitigar possíveis erros, todos os artefatos de software construídos para o tratamento dos dados e os resultados por eles gerados foram homologados e revisados por mais de um pesquisador, considerando amostras de cálculos feitos pelos artefatos, contra amostras de cálculos replicadas em planilhas, manualmente e diretamente no banco de dados.

Tais testes foram feitos na fase de construção (validade de construção) dos artefatos e na fase de execução (validade interna).

Foi utilizada uma taxa bruta para o cálculo dos indicadores, ou seja, para municípios muito pequenos que se apresentem como muito perigosos, as autoridades públicas devem fazer uma avaliação melhor dos valores absolutos destes municípios e definir prioridades. Uma alternativa para isto é usar **taxas bayesianas**, no entanto, neste trabalho, essa situação foi relativamente mitigada pela consideração de diversos tipos de crimes para a criação do índice.

**Ameaças à validade de conclusão** – Os pesos considerados para os crimes podem não refletir a cultura, o nível de violência e o clamor social de algum lugar específico. Esta ameaça foi mitigada com a consideração das penas dos crimes. Além disso, as distâncias dos pesos são aproximadamente proporcionais às distâncias das penas. Vale ressaltar que os pesos podem variar a fórmula de cálculo, de acordo com o objetivo do estudo.

**Ameaças à validade externa - (1)** A falta de correção dos dados criminais fornecidos pelo governo de Minas Gerais e das estimativas populacionais fornecidas pelo IBGE, bem como possíveis subnotificações poderão influenciar diretamente o resultado do experimento. **(2)** A agregação, imposta pelo governo de Minas Gerais, de diversos tipos de crime, também pode influenciar os resultados, uma vez que não é possível pesar e ponderar com maior precisão pela reprovabilidade. Por fim, os *rankings* se limitam aos tipos de crimes disponibilizados.

## 8. Conclusão e Trabalhos Futuros

Este trabalho teve o intuito de detectar padrões, bem como promover maior transparência, visando auxiliar o processo de apoio às decisões estratégicas e operacionais dos governantes e agentes da lei, no combate efetivo da criminalidade. Foram utilizados dados dos 853 municípios, relacionados a incidentes criminais, disponibilizados, de forma aberta, pelo Governo de Minas Gerais, por meio da Secretaria de Estado de Justiça e Segurança Pública (Sejusp).

Como relação ao nível de perigosidade, ficou constatado que Belo Horizonte esteve presente entre os cinco mais perigosos, em todos os períodos avaliados. Além disso, os municípios de Confins e Contagem também foram classificados entre os cinco mais perigosos, em sete dos oito anos. Como destaque fora desse eixo, Água Comprida e Buritis apresentaram casos muito particulares a serem explorados, considerando localização,

desenvolvimento regional, educação e políticas sociais.

No âmbito das regras de associação, considerando as associações “Municípios ⇒ Tipos de Crimes”, destaque para as associações “Unaí ⇒ Furto Consumado”, “Ouro Preto ⇒ Furto Consumado” e “Araguari ⇒ Furto Consumado”, as quais foram encontradas nos oito anos analisados. Para as associações “Municípios ⇒ Alvos de Roubo”, o destaque ficou por conta da associação “Belo Horizonte ⇒ Transeunte”, ocorrida no último ano (2019), a qual obteve o suporte de 21,80% e uma confiança de 0,71, indicando a necessidade do poder público investigar as causas e ações necessárias para coibir o roubo a este tipo de alvo.

Além da pesquisa desenvolvida neste trabalho e como trabalhos futuros, outros possíveis desdobramentos poderão ser analisados. Dados criminais levando em consideração regiões menores, como, por exemplo, as regiões metropolitanas ou os bairros, poderão ser investigados, desde que estados ou municípios os disponibilizem. Nesta linha, podem ser encontradas as associações entre crimes, considerando a localidade e os espaços temporais disponíveis, bem como permitindo a prioridade na inibição de crimes que são antecedentes de outras ocorrências dentro de uma mesma cidade. Sobre pesos de crimes e suas penas, sugerimos a realização de um Survey com juristas, criminalistas, profissionais e especialistas em segurança pública em todo o Brasil, coletando as medianas de pesos atribuídos pelos entrevistados e ampliando a discussão sobre o que pode ser considerada uma localidade mais perigosa e sobre as penas dos crimes. Isto também poderá embasar o poder público e os legisladores, em busca de um código penal que reflita melhor o clamor da sociedade brasileira.

À guisa de conclusão, destacamos a relevância desta pesquisa na produção de evidências da necessidade da disponibilidade de melhores e mais completas arquiteturas de dados abertos para os governos estaduais e de um modelo para outros estados de extração, produção de conhecimento e publicação automatizadas, baseadas nestes dados. Em paralelo, também é importante a produção de indicadores de perigosidade atualizados, que não se baseiam apenas em homicídios, e regras que sirvam para apoiar a definição de prioridades nas avaliações e aplicações de políticas públicas, auxiliando o planejamento estratégico e o processo de apoio à decisão, a fim de otimizar o orçamento empregado na área de segurança pública. Os dados aqui apresentados poderão ser consultados, nessas e em outras perspectivas, no site do projeto Transparência Traduzida, em ([www.transparenciatraduzida.ufs.br](http://www.transparenciatraduzida.ufs.br)).

## Referências

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *In Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, 207-216.

Agrawal, S., & Sejwar, V. (2017). Crime identification using FP-growth and multi objective particle swarm optimization. *In 2017 International Conference on Trends in Electronics and Informatics (ICEI)*, 727-734. IEEE.

Albuquerque, D. J. S., Nunes, V. T., Cappelli, C., & Ralha, C. G. (2016). Implementing E-government Processes Distribution with Transparency using Multi-Agent Systems. *iSys-Revista Brasileira de Sistemas de Informação*, 9(1), 118-138.

Basili, V., Trendowicz, A., Kowalczyk, M., Heidrich, J., Seaman, C., Münch, J., & Rombach, D. (2014). *Aligning Organizations Through Measurement: The GQM+ Strategies Approach*. Springer.

Basili, V. R., & Weiss, D. M. (1984). A methodology for collecting valid software engineering data. *IEEE Transactions on software engineering*, (6), 728-738.

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information polity*, 19(1, 2), 5-16.

Blackburn, S. (2016). *The Oxford Dictionary of Philosophy*. (2a ed.), Oxford University Press. Online version. Recuperado de <http://www.oxfordreference.com/view/10.1093/acref/9780199541430.001.0001/acref-9780199541430-e-1645>.

Campos, O. S. F. (2018). Data analytics transparente para descoberta de padrões e anomalias na realização de convênios e contratos de repasse federais.

Carazza, L., da Mota Silveira Neto, R., & Emanuel, L. (2020). Juvenile curfew and crime reduction: Evidence from Brazil. *Papers in Regional Science*.

Catlett, C., Cesario, E., Talia, D., & Vinci, A. (2018, June). A data-driven approach for spatio-temporal crime predictions in smart cities. *In 2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, 17-24. IEEE.

Chapetta, W. A. (2006). *Uma Infra-estrutura para Planejamento, Execução e Empacotamento de Estudos Experimentais em Engenharia de Software* (Doctoral dissertation, Dissertação de Mestrado, Programa de Engenharia de Sistemas e Computação, COPPE/UFRJ, Universidade Federal do Rio de Janeiro. Rio de Janeiro, RJ, Brasil).

Chaves, M. S. R. S., & Shimizu, I. S. (2018). Síndrome de burnout e qualidade do sono de policiais militares do Piauí. *Revista Brasileira de Medicina do Trabalho*, 16(4), 436-441.

Dados Abertos. (2020). *Portal brasileiro de dados abertos*. Recuperado de <http://www.dados.gov.br>.

Damasceno, M., Teixeira, J., & Campos, G. (2012). A prediction model for criminal levels using socio-criminal data. *International Journal of Electronic Security and Digital Forensics* 7, 4(2-3), 201-214.

de Barros, P. H. B., Baggio, I. S., Stege, A. L., & Hilgemberg, C. M. D. A. T. (2019). Economic development and crime in Brazil: a multivariate and spatial analysis. *Revista Brasileira de Estudos Regionais e Urbanos*, 13(1), 1-22.

de Farias, A. M. G., Cintra, M. E., Felix, A. C., & Cavalcante, D. L. (2018). Definition of strategies for crime prevention and combat using fuzzy clustering and formal concept analysis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 26(03), 429-452.

Ervilha, G. T., & Lima, J. E. D. (2019). Um método econométrico na identificação dos determinantes da criminalidade municipal: a aplicação em Minas Gerais, Brasil (2000-2014). *Economía, sociedad y territorio*, 19(59), 1059-1086.

Fraga, L. D. S., Medeiros, A. P. D., & Vieira, K. M. (2019). Transparência da gestão pública:

Análise em pequenos municípios do Rio Grande o Sul. *Gestão & Planejamento-G&P*, 20.

Guimarães, F. F., Joaquim, S. F., Manzi, M. P., Silva, R. C. D., Bruder-Nascimento, A. C. M. D. O., Costa, E. O., & Langoni, H. (2016). Comparison phenotypic and genotypic identification of *Staphylococcus* species isolated from bovine mastitis. *Pesquisa Veterinária Brasileira*, 36(12), 1160-1164.

Gupta, M., Chandra, B., & Gupta, M. P. (2014). A framework of intelligent decision support system for Indian police. *Journal of Enterprise Information Management*.

Hardy, K., & Maurushat, A. (2017). Opening up government data for Big Data analysis and public benefit. *Computer law & security review*, 33(1), 30-37.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management*, 29(4), 258-268.

Juristo, N., & Moreno, A. M. (2013). *Basics of software engineering experimentation*. Springer Science & Business Media.

Lima, A.; Vignatti, A. & Silva, M. (2020). Reconhecimento de grafos Power-law por algoritmos de aprendizagem de máquina utilizando um conjunto reduzido de propriedades estruturais. *Encontro nacional de inteligência artificial e computacional*, 611-621.

Lima, R. S., & Bueno, S. (2020). Anuário brasileiro de segurança pública 2020. *Fórum Brasileiro de Segurança Pública*. São Paulo.

Maciel, G. S. (2019). Eficiência técnica da polícia militar: um estudo dos comandos de policiamento regionais do Distrito Federal por meio da análise envoltória de dado.

Marzan, C. S., Baculo, M. J. C., de Dios Bulos, R., & Ruiz Jr, C. (2017). Time series analysis and crime pattern forecasting of city crime data. *International conference on algorithms, computing and systems*, pp. 113-118.

Nascimento, D. O. do. (2019). Transparência pública: uma análise do município de Alagoa

Nova-PB.

Paranhos, R., Figueiredo Filho, D. B., da Rocha, E. C., da Silva Júnior, J. A., Neves, J. A. B., & Santos, M. L. W. D. (2014). Desvendando os mistérios do coeficiente de correlação de Pearson: o retorno. *Leviathan (São Paulo)*, (8), 66-95.

Pereira, A. S., Shitsuka, D. M., Parreira, F. J., & Shitsuka, R. (2018). Metodologia da pesquisa científica.

Phillips, P., & Lee, I. (2011). Crime analysis through spatial areal aggregated density patterns. *Geoinformatica*, 15(1), 49-74.

Portal Transparência. (2020). *Portal da Transparência do Governo Federal*. Recuperado de <http://www.portaltransparencia.gov>.

PostgreSQL. (2020). *Installing procedural languages*. Recuperado de <https://www.postgresql.org/docs/current/xplang-install.html>.

Prado, K. H. de J., Souza, L. S., de Jesus Junior, I. D., & Colaço Júnior, M. (2020). Applied Intelligent Data Analysis to Government Data Related to Criminal Incident: A Systematic Review. *Journal of Applied Security Research*, 1-35.

Santos, B. S., Júnior, M. C., & de Souza, J. G. (2018). A initial experimental evaluation of the neuromessenger: a collaborative tool to improve the empathy of text interactions. In *Information Technology-New Generations*, 411-419. Springer, Cham.

Santos, R. M., Santos, I. M., Júnior, M. C. R., & de Mendonça Neto, M. G. (2020). Long Term-short Memory Neural Networks and Word2vec for Self-admitted Technical Debt Detection. In *ICEIS (2)* (pp. 157-165).

Sejusp. (2020). *Portal da Secretaria de Estado de Justiça e Segurança Pública de Minas Gerais*. Recuperado de <http://www.seguranca.mg.gov.br>.

Silvares, A. C. (2019). Políticas Públicas em Segurança no Brasil: Avanços e Novos



Desafios. *Revista Científica Doctum Direito*, 1(3).

Singh, N., Kaverappa, C. B., & Joshi, J. D. (2018). Data mining for prevention of crimes. In *International Conference on Human Interface and the Management of Information* (pp. 705-717). Springer, Cham.

ToppiReddy, H. K. R., Saini, B., & Mahajan, G. (2018). Crime prediction & monitoring framework based on spatial analysis. *Procedia computer science*, 132, 696-705.

Travassos, G. H., & Barros, M. O. (2003). Contributions of in virtuo and in silico experiments for the future of empirical studies in software engineering. In *2nd Workshop on Empirical Software Engineering the Future of Empirical Studies in Software Engineering* (pp. 117-130).

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.

Wu, J., He, Z., Gu, F., Liu, X., Zhou, J., & Yang, C. (2016). Computing exact permutation p-values for association rules. *Information Sciences*, 346, 146-162.

Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017). Crime pattern detection, analysis & prediction. In *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, 1, 225-230.

Zheng, Z., Kohavi, R., & Mason, L. (2001). Real world performance of association rule algorithms. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 401-406.

#### **Porcentagem de contribuição de cada autor no manuscrito**

Kleber Henrique de Jesus Prado – 50%

Methanias Colaço Junior – 50%