

Análise descritiva de perfis de violência doméstica contra mulheres em Minas Gerais (2023)

Arthur Y. R. Codama¹, Caique B. Fortunato¹, Fabio C. M. Filho¹,
Victor G. M. Oliveira¹, Wesley M. D. Chaves¹

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)
Minas Gerais – MG – Brasil

Abstract. *This study analyzes domestic violence reports against women in Minas Gerais (2023) using FP-Growth and Subgroup Discovery (SD) techniques. The goal was to identify frequent and exceptional patterns related to recurrence and offense consummation. Results highlight high recurrence in threat cases, especially in rural areas, and a higher incidence of consummated violence on weekends. The findings reveal critical territorial and temporal contexts, offering insights to support targeted public policies and preventive strategies.*

Resumo. *Este estudo analisa registros de violência doméstica contra mulheres em Minas Gerais (2023), aplicando os métodos FP-Growth e Descoberta de Subgrupos (SD). O objetivo foi identificar padrões frequentes e excepcionais relacionados à reincidência e à consumação dos delitos. Os resultados revelam alta reincidência em casos de ameaça, especialmente no interior, e maior ocorrência de violência consumada nos fins de semana. Os achados fornecem subsídios para políticas públicas e ações preventivas mais eficazes.*

1. Introdução

A violência doméstica contra mulheres é uma realidade dura e persistente no Brasil. Apesar de avanços importantes como a criação da Lei nº 11.340/2006, conhecida como Lei Maria da Penha, muitas mulheres ainda enfrentam situações de agressão dentro de seus próprios lares, frequentemente de forma silenciosa e repetida. A reincidência dos casos e a dificuldade de romper o ciclo de violência revelam um cenário complexo, que desafia tanto o sistema de justiça quanto as redes de apoio social.

Entender esse cenário vai muito além de contabilizar estatísticas. É preciso olhar com atenção para os dados buscar respostas para padrões que se repetem, contextos específicos como certos dias da semana, regiões ou tipos de violência em que o risco é maior. Tais análises estendem também na identificação de regularidades que podem fazer a diferença entre uma política pública genérica e uma ação realmente eficaz na proteção de mulheres em situação de vulnerabilidade.

Utilizando dados públicos sobre violência doméstica contra mulheres em Minas Gerais no ano de 2023, este trabalho aplica técnicas de mineração de dados como a Mineração de Regras de Associação e a Descoberta de Subgrupos para explorar padrões relevantes relacionados à reincidência, à gravidade dos casos e às circunstâncias em que ocorrem. O objetivo é transformar esses dados de modo que seja possível apoiar decisões mais conscientes e estratégias de enfrentamento mais sensíveis à realidade das vítimas.

2. Referencial Teórico

A presente seção apresenta os fundamentos teóricos que sustentam a análise realizada neste trabalho.

2.1. Violência Doméstica contra Mulheres

De acordo com o artigo 5º da Lei nº 11.340/2006 (Lei Maria da Penha) [Brasil 2006], configura-se como violência doméstica e familiar contra a mulher qualquer ação ou omissão baseada no gênero que lhe cause danos físicos, sexuais, psicológicos, morais ou patrimoniais.

Segundo a OMS [Waiselfisz 2015], em 2013, o Brasil ocupava a 5ª posição entre 83 países com maiores taxas de homicídio de mulheres (4,8 por 100 mil), sendo cerca de 30% dos casos no ambiente doméstico. A pesquisa do DataSenado [DataSenado 2013] apontou que uma em cada cinco brasileiras já foi vítima de violência doméstica praticada por um homem. Já a Fundação Perseu Abramo [Abramo and do Comércio (SESC) 2010] estimou que, a cada dois minutos, cinco mulheres sofrem agressões violentas no país.

A violência de gênero tende a se repetir de maneira cíclica em três fases: tensão, agressão e reconciliação, envolvendo diferentes formas de abuso, que podem ocorrer isoladamente ou de forma combinada.

Trata-se, portanto, de um problema estrutural e recorrente, cuja superação requer estratégias baseadas em evidências. A análise dos padrões temporais, territoriais e comportamentais mostra-se importante para subsidiar políticas públicas mais eficazes.

2.2. Mineração de Regras de Associação

A Mineração de Regras de Associação é uma técnica de aprendizado não supervisionado voltada à identificação de relações frequentes entre variáveis em grandes bases de dados, extraíndo padrões do tipo se... então... [Han et al. 2012].

Neste trabalho, foi utilizado o algoritmo FP-Growth [Han et al. 2000], que se destaca por extrair padrões frequentes de forma eficiente, sem gerar candidatos, utilizando uma estrutura compacta chamada *FP-tree*. As regras são avaliadas por métricas como *suporte*, que indica a frequência de ocorrência do conjunto de itens; *confiança*, que mede a probabilidade de o consequente ocorrer dado o antecedente; e *lift*, que avalia a relevância da associação, indicando se a coocorrência é maior do que o esperado pelo acaso.

Essa abordagem é especialmente útil em contextos sociais por revelar padrões latentes entre fatores como tipo de delito, localização e período da semana, oferecendo suporte à tomada de decisão.

2.3. Descoberta de Subgrupos

A Descoberta de Subgrupos (Subgroup Discovery) é uma técnica de mineração de dados focada na identificação de padrões estatisticamente distintos em subconjuntos (subgrupos) de dados, com relação a uma variável-alvo [Herrera et al. 2011].

Diferente da mineração de regras de associação, que busca padrões frequentes, essa abordagem foca em padrões *excepcionais*, ou seja, que se desviam da média do conjunto de dados. Neste contexto, uma métrica comum é a *WRAcc* (Weighted Relative

Accuracy), que equilibra a cobertura do subgrupo com o ganho de precisão em relação à distribuição geral da variável [Lavrač et al. 1999].

Entre os algoritmos utilizados, destaca-se o *SimpleDFS*, que realiza uma busca em profundidade limitada para compor regras com até três atributos, explorando de forma eficiente o espaço de busca [Langenberg 2023]. Essa técnica tem sido aplicada com sucesso em áreas como saúde, segurança pública e ciências sociais, contribuindo para a identificação de perfis de risco e o direcionamento de políticas públicas [Herrera et al. 2011].

3. Metodologia

Este estudo segue uma abordagem quantitativa e descritiva, com foco na análise exploratória de dados públicos relacionados à violência doméstica contra mulheres em Minas Gerais no ano de 2023. O processo metodológico compreendeu as etapas de coleta, pré-processamento, análise e interpretação dos dados, conforme descrito a seguir.

3.1. Coleta dos dados

Este trabalho baseia-se na análise de dados públicos sobre violência doméstica contra mulheres em Minas Gerais, disponibilizados pelo Governo do Estado através do Portal de Dados Abertos¹. O conjunto de dados contém 61.536 registros referentes ao ano de 2023.

Cada linha da base de dados representa um registro individual de ocorrência policial, contendo cinco variáveis principais. A variável `municipio_cod` (int) representa o código numérico do município onde o fato foi registrado, enquanto `municipio_fato` (string) indica o nome do município. A variável `data_fato` (string) corresponde à data da ocorrência, posteriormente processada para extração de informações temporais. Já `mes` (int) identifica o mês do registro e `ano` (int) indica o ano da ocorrência, observe que todos os registros pertencem a 2023. A estrutura simples da base foi complementada por variáveis derivadas ao longo do processo analítico, como `dia_da_semana`, `fim_de_semana`, `natureza_delito`, `rmbh`, e `reincidencia`, possibilitando análises mais ricas sob múltiplas perspectivas temporais, geográficas e criminais.

3.2. Pré-processamento dos Dados

Os dados foram carregados e processados em Python 3, no ambiente Google Colab. A coluna de datas (`data_fato`) foi corrigida, uma vez que apresentava valores mal formatados (números seriais originados de planilhas). A partir dessa variável, foram extraídas novas informações, como o `dia_da_semana` e uma variável booleana `fim_de_semana`. Além disso, foi criada a variável categórica `dia_categoria`, que agrupa os dias da semana em faixas temporais, facilitando análises agregadas.

Foram selecionados atributos categóricos como município, mês, tipo de delito, status do crime (tentado ou consumado), RISP, RMBH e as variáveis temporais derivadas. Variáveis contínuas, como datas e quantidades, foram transformadas em categorias (ex.: dias da semana, fim de semana vs. dias úteis), viabilizando a extração de padrões com técnicas de mineração de dados. Por seguinte, os dados foram transformados no formato transacional (*atributo:valor*) para viabilizar a mineração de padrões.

¹<https://dados.mg.gov.br/dataset/violencia-contramulher>

3.3. Mineração de Regras de Associação

Para identificar padrões frequentes de coocorrência entre atributos das ocorrências de violência doméstica, foi utilizado o algoritmo FP-Growth (*Frequent Pattern Growth*), evitando a geração exaustiva de candidatos como ocorre no algoritmo Apriori. Adicionalmente, embora não seja o caso da aplicação empregada, o algoritmo é eficiente para extração de regras de associação em grandes volumes de dados transacionais

A base de dados foi transformada em uma estrutura transacional, onde cada linha representa uma ocorrência e os atributos categóricos selecionados formam os itens. Foram considerados, por exemplo, os atributos `natureza_delito`, `rmbh`, `fim_de_semana`, `tentado_consumado`, entre outros.

A etapa de pré-processamento incluiu a seleção e transformação de variáveis categóricas; conversão do conjunto de dados para o formato binário transacional e aplicação do algoritmo *FP-Growth*, com parametrização dos valores mínimos de suporte e confiança.

A implementação foi realizada com a biblioteca `mlxtend`, sendo os parâmetros de suporte mínimo ajustados de forma iterativa para equilibrar a quantidade de regras geradas e a relevância dos padrões. As regras extraídas foram avaliadas com base em três métricas: suporte, que representa a proporção de transações em que o conjunto de itens ocorre; confiança, que expressa a probabilidade condicional de o consequente ocorrer dado o antecedente; e lift, que corresponde à razão entre a confiança da regra e a probabilidade esperada do consequente, sendo que valores superiores a 1 indicam correlação positiva entre os itens.

As regras com maior *lift* e suporte foram analisadas qualitativamente, com foco em identificar combinações de fatores contextuais, territoriais e temporais associados a determinados tipos de violência.

3.4. Descoberta de Subgrupos

A técnica de *descoberta de subgrupos* (*Subgroup Discovery* – SD) foi aplicada com o objetivo de identificar subconjuntos de dados cujos comportamentos se destacam estatisticamente em relação a variáveis-alvo específicas. A abordagem exploratória busca padrões locais, ou seja, regras condicionais que definem subconjuntos com distribuição significativamente diferente da população total para um atributo de interesse.

Neste contexto, descoberta de subgrupos foi utilizada para investigar os aspectos:

- **Reincidência de violência doméstica**, definida a partir da repetição de registros no mesmo município, com base na mesma natureza de delito;
- **Perfis Contextuais com Alta Probabilidade de Consumo**.

A tarefa foi conduzida por meio da construção de regras condicionais do tipo *condição.1 AND condição.2 AND ... → comportamento*, nas quais o antecedente define o subgrupo (por exemplo, `fim_de_semana == 'Sim' AND rmbh == 'Interior de MG'`) e o consequente está associado ao comportamento a ser explicado ou analisado (por exemplo, alta taxa de reincidência).

Para implementação, foi utilizada a biblioteca `subgroup_discovery`, com os dados previamente tratados e codificados. As variáveis explanatórias utilizadas incluíram:

- `fim_de_semana` (booleana derivada do `dia_da_semana`);
- `rmbh` (região da ocorrência: capital, RMBH ou interior);
- `natureza_delito` (tipo de crime);
- entre outras variáveis categóricas relacionadas ao contexto das ocorrências.

O algoritmo utilizado baseia-se em busca heurística com avaliação por função de qualidade, considerando atributos como cobertura, precisão e *lift*. Foram calculadas quatro métricas principais: *coverage*, que representa a proporção do conjunto de dados coberto pela regra; *target share*, que indica a proporção do alvo positivo dentro do subgrupo; *lift*, que expressa a razão entre o *target share* do subgrupo e o do conjunto total, refletindo o ganho informativo; e *quality*, uma medida composta usada para ranqueamento dos subgrupos, que considera o desvio em relação à média global, ponderado pela cobertura.

A interpretação dos subgrupos foi complementada por aspectos socioterritoriais, com o intuito de contextualizar os padrões identificados em termos de localização e dinâmica social.

Os resultados foram ordenados pelo valor da função de qualidade, permitindo a seleção dos 10 subgrupos mais relevantes para cada variável-alvo. Esses padrões foram posteriormente interpretados à luz de aspectos sociais e territoriais, considerando implicações práticas para ações de prevenção e intervenção.

3.5. Ferramentas Utilizadas

Toda a análise foi realizada em notebooks Jupyter no Google Colab, utilizando Python e as bibliotecas `pandas`, `mlxtend`, `pysubgroup`, `matplotlib` e `seaborn`.

4. Resultados

Nesta seção, são apresentados e discutidos os principais resultados obtidos a partir da aplicação das etapas descritas na seção de Metodologia. Os achados incluem padrões extraídos por meio do algoritmo *FP-Growth*, bem como subgrupos estatisticamente relevantes identificados com o uso da técnica de descoberta de subgrupos.

4.1. Regras de associação

Entre os padrões de associação extraídos utilizando o algoritmo *FP-Growth*, destaca-se um grupo de regras com alta coesão estatística, evidenciada por valores de *lift* superiores a 7. Essas regras indicam uma forte associação entre três dimensões: o território (representado pela jurisdição da RISP), a localização regional (regiões da RMBH fora de Belo Horizonte), e o tempo (ocorrências durante a semana). Em particular, foi observada uma concentração significativa de ocorrências de violência doméstica consumada sob responsabilidade do 2º Departamento da RISP, que abrange municípios como Contagem, Betim e Ribeirão das Neves. A regra com antecedente (*risp:2º Departamento - Contagem, tentado_consumado:CONSUMADO*) e consequente (*fim_de_semana:Não, rmbh:RMBH (sem BH)*) apresentou uma confiança de 67,1% e *lift* de 7,38, indicando que tais ocorrências são mais de sete vezes mais frequentes do que o esperado aleatoriamente.

Além disso, outras regras complementares revelaram que, quando uma ocorrência acontece fora do fim de semana e nos municípios da RMBH (sem incluir a capital), há

uma probabilidade superior a 58% de ela estar sob responsabilidade do 2º Departamento da RISP e resultar em crime consumado. Um padrão ainda mais forte foi encontrado com a seguinte composição de antecedentes: *fim_de_semana:Não*, *risp:2º Departamento - Contagem*, *tentado_consumado:CONSUMADO*, cuja confiança em relação ao consequente *rmbh:RMBH (sem BH)* foi de aproximadamente 99%. Esse padrão sugere que, quando esses três fatores estão presentes, é quase certo que a ocorrência se deu fora da capital mineira.

Esses achados evidenciam um perfil socioterritorial específico para a violência doméstica consumada: episódios que se concentram em municípios da RMBH, fora de Belo Horizonte, durante os dias úteis, e que envolvem ação policial sob jurisdição do 2º Departamento. A repetição desses padrões em diferentes regras indica que a violência nessas áreas não é apenas frequente, mas recorrente sob determinadas circunstâncias contextuais. Tais evidências podem ser particularmente úteis para informar políticas públicas direcionadas, como reforço de medidas preventivas em horários e dias úteis, além da alocação estratégica de recursos nas unidades policiais que operam nas áreas mais críticas. A natureza concentrada dos padrões também pode refletir desigualdades estruturais, como acesso limitado a redes de apoio e proteção às vítimas nos municípios periféricos da RMBH.

A Tabela 1 resume as principais regras identificadas, organizadas por suporte, confiança e *lift*.

Observa-se que os valores *rmbh:RMBH (sem BH)* e *risp:2º Departamento - Contagem* foram os mais recorrentes, indicando que essas características territoriais são frequentemente inferidas a partir de diferentes combinações de variáveis antecedentes. Esse padrão reforça a existência de um perfil socioterritorial associado à violência doméstica consumada nos municípios periféricos da Região Metropolitana de Belo Horizonte, especialmente em dias úteis e sob jurisdição policial específica. A alta frequência desses consequentes sugere que tais fatores são consistentes e estruturantes no conjunto de regras descobertas.

Adicionalmente e complementando os resultados, também foram analisadas regras com consequente *natureza_delito:AMEACA*, a fim de identificar perfis contextuais mais associados a esse tipo de ocorrência, conforme demonstrado na Tabela 1. Os resultados revelam que a presença simultânea de duas variáveis — *rmbh:Interior de MG* e *fim_de_semana:Não* — configura a regra com maior confiança (30,6%) e *lift* (1,09), indicando que essa combinação de fatores ocorre com frequência ligeiramente superior ao que seria esperado sob independência estatística.

Apesar de os valores de *lift* se manterem apenas moderadamente acima de 1 (variação entre 1,02 e 1,09), os respectivos valores de suporte são expressivos, oscilando entre aproximadamente 16% e 24% da amostra analisada. Isso sugere que essas configurações contextuais — especialmente eventos ocorridos em municípios da RMBH fora da capital e durante os dias úteis — são recorrentes nos registros de ameaça.

As variáveis *tentado_consumado:CONSUMADO* e *fim_de_semana:Não* surgem em diversas regras com bons indicadores de confiança, refletindo um padrão que associa a consumação do crime e a ocorrência em dias úteis ao delito de ameaça. Notadamente, a simples presença da variável *rmbh:Interior de MG* já é suficiente para compor uma regra

Table 1. Regras com alto *lift* envolvendo o 2º Departamento da RISP e RMBH (sem BH).

ID	Antecedente	Consequente	sup.	conf.	lift
588	{RISP: Contagem, Consumado}	{Dias úteis, RMBH sem BH}	0,053	0,672	7,38
585	{Dias úteis, RMBH sem BH}	{RISP: Contagem, Consumado}	0,053	0,586	7,38
575	{Dias úteis, RMBH sem BH}	{RISP: Contagem}	0,054	0,592	7,38
578	{RISP: Contagem}	{Dias úteis, RMBH sem BH}	0,054	0,672	7,38
582	{Dias úteis, RMBH sem BH, Consumado}	{RISP: Contagem}	0,053	0,591	7,37
591	{RISP: Contagem}	{Dias úteis, RMBH sem BH, Consumado}	0,053	0,665	7,37
579	{RMBH sem BH}	{Dias úteis, RISP: Contagem}	0,054	0,397	7,30
574	{Dias úteis, RISP: Contagem}	{RMBH sem BH}	0,054	0,990	7,30
592	{RMBH sem BH}	{Dias úteis, RISP: Contagem, Consumado}	0,053	0,393	7,30
581	{Dias úteis, RISP: Contagem, Consumado}	{RMBH sem BH}	0,053	0,990	7,30

com suporte elevado (24,3%) e *confidence* próxima a 29,6%, o que reforça a importância dessa variável na composição do perfil territorial desse tipo de violência.

Embora tais associações não apresentem forte poder discriminativo, os padrões extraídos são consistentes e estatisticamente relevantes, e contribuem para delinear um perfil sociodemográfico e territorial da violência doméstica caracterizada como “ameaça”. Esses achados podem fundamentar ações de prevenção mais direcionadas, em especial voltadas a municípios periféricos da RMBH e com foco em dias úteis, período em que há maior incidência dessas ocorrências.

Table 2. Regras com consequente AMEAÇA e seus respectivos indicadores.

ID	Antecedente	sup.	conf.	lift
36	{fim_de_semana:Não, Interior}	0,165	0,306	1,09
42	{fim_de_semana:Não, CONSUMADO, Interior}	0,162	0,305	1,08
13	{rmbh:Interior de MG}	0,243	0,296	1,06
17	{CONSUMADO, Interior}	0,240	0,295	1,05
14	{fim_de_semana:Não}	0,191	0,288	1,03
30	{fim_de_semana:Não, CONSUMADO}	0,188	0,287	1,02

4.2. Descobertas de subgrupos

A técnica de descoberta de subgrupos permitiu identificar padrões contextuais relevantes nas ocorrências de violência doméstica. A análise destacou subconjuntos com comportamentos distintos da base geral, com foco na reincidência ao longo do ano e nos tipos de violência mais frequentes. Os resultados contribuem para a identificação de perfis críticos e para o direcionamento de ações preventivas.

4.2.1. Reincidência de violência doméstica

O objetivo dos resultados a seguir é identificar subgrupos de ocorrências de violência doméstica contra mulheres em Minas Gerais (2023) que apresentem alta taxa de reincidência, ou seja, situações onde um tipo de violência volta a ocorrer.

Table 3. Subgrupos com maior reincidência identificados por SD.

Subgrupo	Tam.	Reinc.	WRAcc	Interpretação
natureza_delito == <i>AMEACA</i>	17.285	17.238	0,0198	Quase todos os casos de ameaça se repetem no ano.
<i>AMEACA</i> no interior de MG	14.963	14.916	0,0170	Reincidência altíssima, indicando possível ausência de resposta efetiva.
<i>AMEACA</i> em dias úteis	11.726	11.696	0,0135	Forte padrão de reincidência em dias da semana.
<i>VIAS DE FATO / AGRESSAO</i>	11.495	11.385	0,0119	Alto índice de reincidência em agressões físicas.
<i>AMEACA</i> em dias úteis (refinado)	10.141	10.111	0,0116	Reforça padrão temporal de reincidência.
<i>LESAO CORPORAL</i>	9.780	9.679	0,0100	Lesões corporais também se repetem com frequência.

A análise revelou que os casos de *AMEAÇA* são os que mais se repetem, indicando que essa forma de violência tende a ocorrer diversas vezes, mesmo após uma primeira notificação. Entre os padrões mais relevantes, destacam-se subgrupos como *AMEAÇA no interior de MG*, que apresentou uma taxa de reincidência próxima de 100%, e *AMEAÇA durante os dias úteis*, reforçando a hipótese de que esse tipo de delito ocorre de maneira recorrente e previsível em certos contextos temporais e territoriais. Além disso, delitos como *LESAO CORPORAL* e *VIAS DE FATO / AGRESSAO* também apresentaram forte reincidência, sugerindo que formas de violência física de menor potencial ofensivo podem estar inseridas em ciclos de repetição. A maioria dos subgrupos identificados apresentou mais de 99% de ocorrências reincidentes, com valores de *WRAcc* elevados, como evidenciado na Tabela 3. Esses achados sugerem que certos tipos de violência — especialmente os mais frequentes e de caráter físico ou psicológico leve — tendem a evoluir em ciclos sucessivos, muitas vezes sem uma interrupção efetiva por parte do Estado.

A descoberta de subgrupos reincidentes traz implicações diretas para o desenvolvimento de políticas públicas mais eficazes. Por exemplo, a detecção precoce de perfis

com alta reincidência pode subsidiar estratégias de monitoramento contínuo de vítimas de AMEAÇA, evitando sua escalada para agressões físicas mais graves. Também se destaca a necessidade de priorizar o atendimento a ocorrências reincidentes no interior do estado, onde os recursos de proteção e apoio podem ser mais limitados. Além disso, os padrões temporais sugerem que campanhas educativas e ações de prevenção voltadas para os dias úteis podem alcançar públicos em situação de maior vulnerabilidade. Dessa forma, a análise da reincidência não apenas complementa os achados da associação de padrões, mas também fortalece o caráter preventivo da investigação, ao permitir antecipar quais perfis estão mais propensos a múltiplas ocorrências.

4.2.2. Perfis contextuais com maior probabilidade de consumação

A análise de descoberta de subgrupos foi aplicada com o objetivo de identificar padrões contextuais associados a maiores taxas de consumação da violência doméstica. Utilizando como variável-alvo o campo *consumado*, foram geradas regras com alta proporção de casos consumados, avaliadas por métricas como *WRAcc*, *lift* e *target share*.

Table 4. Subgrupos com alta taxa de consumação em casos de violência doméstica.

Subgrupo		Tamanho	Consumados	Proporção
fim_de_semana=='Sim'	AND	17.387	17.274	99,35%
rmbh=='interior'				
natureza_delito=='LESAO CORPO-RAL'	AND rmbh=='interior'	8.596	8.566	99,65%
natureza_delito=='LESAO CORPO-RAL'		9.780	9.736	99,55%
natureza_delito=='DESCUMPRIMENTO DE MEDIDA PROTETIVA'		4.029	4.029	100,00%
fim_de_semana=='Sim'		20.876	20.710	99,20%
fim_de_semana=='Não'	AND	4.869	4.854	99,69%
natureza_delito=='LESAO CORPO-RAL'				
natureza_delito=='DESCUMPRIMENTO DE MEDIDA PROTETIVA'	AND	3.163	3.163	100,00%
rmbh=='interior'				
rmbh=='interior'		50.484	50.033	99,11%
dia_da_semana=='Sunday'	AND	9.383	9.321	99,34%
rmbh=='interior'				

Os resultados, demonstrados na tabela 4 destacam que finais de semana no interior de Minas Gerais formam o subgrupo com maior *quality* (0,00085), abrangendo 28,3% dos registros e apresentando uma taxa de consumação de 99,35%. Esse padrão sugere uma possível vulnerabilidade institucional nesses contextos, com menor capacidade de resposta imediata.

Delitos como *lesão corporal* também se destacam: quando ocorrem no interior

(cobertura de 14,1%), a taxa de consumação ultrapassa 99,6%, mesmo valor observado em dias úteis. Isso evidencia a gravidade ou a baixa capacidade de interrupção dos casos.

Outro resultado crítico refere-se ao descumprimento de medidas protetivas: em todos os registros encontrados (100%), o caso foi consumado, principalmente no interior, onde há indicação de fragilidade no acompanhamento e fiscalização dessas medidas.

O interior de Minas Gerais, de maneira geral, concentra 82% da base analisada e apresenta taxa de consumação de 99,1%, reforçando um padrão de risco geográfico elevado. Além disso, domingos no interior compõem outro cenário de vulnerabilidade, com 99,34% de consumação.

Esses achados apontam que tanto o tempo (como os fins de semana) quanto o local (interior do estado) são fatores determinantes para a consumação da violência. Além disso, certos tipos de delito, como lesão corporal e descumprimento de medida protetiva, mantêm alto grau de consumação independentemente do contexto, exigindo atenção especial de políticas públicas e equipes de proteção social.

5. Discussão

Os resultados obtidos neste estudo evidenciam padrões contextuais robustos na dinâmica da violência doméstica contra mulheres em Minas Gerais em 2023. A partir de técnicas de associação e descoberta de subgrupos, identificaram-se fatores territoriais, temporais e de tipificação criminal associados tanto à reincidência quanto à consumação das ocorrências. Esses achados oferecem subsídios concretos para a formulação de políticas públicas mais eficazes, permitindo intervenções baseadas em evidências.

Do ponto de vista da aplicação social, as análises sugerem que medidas de proteção devem ser intensificadas nos municípios do interior, especialmente durante os fins de semana e também nos municípios da RMBH fora da Capital, mesmo nos dias úteis, pois este pode estar sobrecarregado pelos grandes municípios que atende, como Betim, Contagem e Ribeirão das Neves. A alta taxa de consumação em regiões com menor cobertura institucional aponta para uma possível fragilidade na atuação preventiva do Estado, como menor presença policial, serviços de acolhimento ou monitoramento eletrônico de agressores. Para além das políticas de segurança, isso reforça a importância de ampliar o acesso a serviços sociais e a redes de apoio comunitário, especialmente nos territórios de maior vulnerabilidade.

O uso de métodos baseados em padrões frequentes e descobertas de subgrupos também pode fundamentar o desenvolvimento de ferramentas digitais de apoio à tomada de decisão, como dashboards preditivos para centros de apoio à mulher, ou sistemas de alerta precoce em delegacias especializadas. Tais soluções tecnológicas poderiam auxiliar na priorização de casos com maior risco de reincidência ou consumação, contribuindo para a otimização dos recursos disponíveis.

Contudo, a adoção prática desses métodos impõe desafios éticos significativos. Sistemas automatizados baseados em padrões históricos podem reproduzir vieses estruturais dos dados, como discriminação geográfica ou social, reforçando estigmas sobre determinados territórios ou populações. Além disso, o uso de perfis de risco pode levantar preocupações relacionadas à privacidade e à liberdade individual, especialmente se for empregado de forma invasiva ou sem consentimento.

Em suma, os padrões identificados não apenas revelam vulnerabilidades estruturais no enfrentamento à violência doméstica, mas também indicam caminhos possíveis para a intervenção social. Cabe às instituições e agentes envolvidos avaliar, com responsabilidade, o melhor modo de traduzir essas evidências em ação concreta e ética.

6. Conclusão

Este estudo aplicou técnicas de mineração de dados para analisar padrões contextuais em ocorrências de violência doméstica contra mulheres em Minas Gerais no ano de 2023. Os resultados obtidos evidenciaram a existência de perfis específicos associados a maior reincidência e maior probabilidade de consumação dos delitos, com destaque para o interior do estado e os fins de semana como fatores críticos.

A análise permitiu caracterizar padrões frequentes e fornecer insumos para ações preventivas e políticas públicas mais direcionadas. A identificação de perfis territoriais, temporais e criminais de maior risco pode subsidiar a priorização de recursos, o desenvolvimento de ferramentas tecnológicas de apoio à decisão e a formulação de campanhas educativas com foco em contextos específicos.

Todavia, este trabalho apresenta limitações. A análise se baseia em dados administrativos, que podem conter subnotificações ou inconsistências. A definição de reincidência, por exemplo, foi limitada à repetição de registros no mesmo município com a mesma natureza de delito, não permitindo rastrear a trajetória de vítimas ou agressores individualmente. Além disso, a análise não considerou variáveis socioeconômicas ou de perfil demográfico, o que restringe a compreensão dos fatores estruturais associados à violência.

Outro porém é que o FP-Growth privilegia padrões frequentes, podendo negligenciar eventos raros mas relevantes, enquanto a descoberta de subgrupos depende fortemente da definição da variável-alvo e pode sofrer influência de correlações ruins. Por fim, as interpretações propostas devem ser vistas como indícios estatísticos que devem ser analisados em conjunto com o conhecimento de especialistas da área.

Trabalhos futuros podem ampliar esta abordagem incluindo dados longitudinais, variáveis contextuais mais ricas (como renda, escolaridade e raça/cor), e métodos híbridos que combinem estatística descritiva com aprendizado de máquina interpretável. Também se recomenda a construção de parcerias com instituições públicas e organizações da sociedade civil para validar os achados e promover sua aplicação ética e responsável.

References

- Abramo, F. P. and do Comércio (SESC), S. S. (2010). Mulheres brasileiras e gênero nos espaços público e privado. <https://fpabramo.org.br/2011/02/11/violencia-domestica/>. Pesquisa nacional sobre violência contra a mulher.
- Brasil (2006). Lei n.º 11.340, de 7 de agosto de 2006. https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/111340.htm. Cria mecanismos para coibir a violência doméstica e familiar contra a mulher, nos termos do § 8º do art. 226 da Constituição Federal. Diário Oficial da União, 8 ago. 2006.
- DataSenado, S. F. I. (2013). Pesquisa violência doméstica e familiar contra a mulher. <https://www12.senado.leg.br/institucional/datasenado/>

arquivos/a-violencia-domestica-contr-a-mulher. Pesquisa realizada em fevereiro de 2013, quinta da série histórica iniciada em 2005.

Han, J., Kamber, M., and Pei, J. (2012). *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco, CA, 3 edition.

Han, J., Pei, J., and Yin, Y. (2000). Mining frequent patterns without candidate generation. *ACM SIGMOD Record*, 29(2):1–12.

Herrera, F., Carmona, C. J., González, P., and del Jesus, M. J. (2011). An overview on subgroup discovery: foundations and applications. *Knowledge and Information Systems*, 29(3):495–525.

Langenberg, B. (2023). subgroupsem: Subgroup discovery algorithms for structural equation models. <https://rdr.io/github/langenberg/subgroupsem/>. Accessed: 2025-06-29.

Lavrač, N., Flach, P., and Zupan, B. (1999). Rule evaluation measures: A unifying view. In *Proceedings of the 9th International Workshop on Inductive Logic Programming*, pages 174–185. Springer.

Waiselfisz, J. J. (2015). Mapa da violência 2015: Homicídio de mulheres no brasil. https://www.onumulheres.org.br/wp-content/uploads/2016/04/MapaViolencia_2015_mulheres.pdf. Relatório produzido com base em dados da OMS e do SIM/Ministério da Saúde.