



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

Arthur Y. R. Codama,
Caique B. Fortunato,
Fabio C. M. Filho,
Victor G. M. Oliveira,
Wesley M. D. Chaves

INTRODUÇÃO E CONTEXTUALIZAÇÃO

Introdução



- Segundo a OMS, 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. Em 2013 eram 100,5 milhões de mulheres dando uma taxa de **4825** por ano. Cerca de **30% das agressões no ambiente doméstico**.
- Segundo a Fundação Perseu Abramo **a cada dois minutos, cinco mulheres sofrem agressões violentas no país**
- Foram **4.181 vítimas registradas em 2024**, o que representa um aumento de 12,4% em relação a 2023

Nesse contexto a ideia do trabalho é aplicar os métodos FP-Growth e Descoberta de Subgrupos (SD) para identificar padrões frequentes e excepcionais relacionados à reincidência e à consumação dos delitos.

METODOLOGIA

Metodologia



- **Coleta de dados:** Dados públicos sobre violência doméstica contra mulheres em Minas Gerais, disponibilizados pelo Governo do Estado através do Portal de Dados Abertos. O conjunto de dados contém 61.536 registros referentes ao ano de 2023.
 - Link: <https://dados.mg.gov.br/dataset/violencia-contramulher>
- **Pré-processamento:** Os dados foram processados no Google Colab.
 - A coluna de datas (data fato) foi corrigida e então 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.

Metodologia



- 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).
 - Os dados foram transformados no formato transacional (atributo:valor) para viabilizar a mineração de padrões.
- Toda a análise foi realizada em notebooks Jupyter no Google Colab, utilizando Python e as bibliotecas pandas, mlxtend, pysubgroup, matplotlib e seaborn.

Metodologia - Base de Dados

Tipo de dado	Nome da Coluna	Descrição	Opções
Int	municipio_cod	Identificador único da ocorrência policial	
String	municipio_fato	Nome do município onde a ocorrência foi registrada	853 Municípios de MG
DateTime	data_fato	Data completa da ocorrência	
Int	mes	Mês da ocorrência	
Int	ano	Ano da ocorrência	
String	risp	Região Integrada de Segurança Pública (1 a 19)	19 Depto.
String	rmbh	Região Integrada de Segurança Pública	Belo Horizonte; RMBH (Sem BH); Interior de MG
String	natureza_delito	Categorização do crime cometido	167 tipos de delitos
String	tentado_consumado	status do crime em relação à sua execução	Consumado ou tentado
Int	qtde_vitimas	Quantidade de vítimas envolvidas na ocorrência	
String	dia_da_semana	Qual dia da semana ocorreu	Sunday; Monday;Tuesday ...
String	fim_de_semana	Se ocorreu no fim de semana ou não	Sim; Não
String	reincidência	Se houve repetição do delito (Município e Natureza)	Booleano

Em verde estão as variáveis que foram adicionadas após derivações

*RMBH já existia mas não categorizada como mostrado

Técnicas aplicadas: FP-Growth



- Algoritmo de mineração de padrões frequentes sem geração explícita de candidatos.
- Os dados foram convertidos para o formato transacional binário (atributo:valor)
- Parâmetro ajustado iterativamente:
 - Suporte mínimo $\geq 5\%$ \rightarrow filtra padrões muito raros.
- Geração de regras de associação usando:
 - **Suporte**: frequência relativa do padrão na base.
 - **Confiança**: probabilidade de ocorrência do consequente, dado o antecedente.
 - **Lift**: fator de independência entre antecedente e consequente (valores >1 indicam associação positiva).

Exemplo: Se `rmbh = Interior de MG` e `fim_de_semana = Não`, então `natureza_delito = Ameaça`

Técnicas aplicadas: Descoberta de Subgrupos



- Técnica de mineração descritiva orientada a um alvo específico (variável-alvo booleana).
- Foram definidas duas variáveis-alvo:
 - reincidência = 1 → repetição por município e tipo de delito
 - consumado = 1 → ocorrência do crime consumado
- Métrica de avaliação: WRAcc (Weighted Relative Accuracy)
 - Mede o ganho de qualidade estatística da regra em relação à média global, ponderando pela cobertura.
- Espaço de busca: combinações de até 3 atributos categóricos (ex.: rmbh, risp, dia_categoria)

Exemplo: Se `risp = 2`, `rmbh = RMBH sem BH`, `fim_de_semana = Não`

→ WRAcc alto, consumo > 99%, cobertura significativa

ALGORITMOS E RESULTADOS

FP-Growth

FP - Growth (Overview)




1. Pré-processamento
2. Preparação das transações
 - a. Seleciona as colunas categóricas relevantes: `municipio_fato`, `mes`, `resp`, `rmbh`, `natureza_delito`, `tentado_consumado`, `dia_da_semana`, `fim_de_semana`
 - b. Transforma cada linha em uma transação no formato `atributo:valor`.
3. Codificação One-Hot utilizando o `TransactionEncoder` para transformar as transações em uma matriz binária
4. Execução do FP-Growth com suporte mínimo de 5%
 - a. Regras de associação com `support`, `confidence`, `lift`

Metodologia - FP Growth



- A etapa de pré-processamento incluiu a **seleção e transformação de variáveis categóricas, conversão do conjunto de dados para o formato binário transacional**. Posteriormente foi aplicado o algoritmo FP-Growth, com parametrização dos valores mínimos de suporte e confiança. As variáveis analisadas foram:
 - **municipio_fato, mes, natureza_delito, tentado_consumado, dia_da_semana, fim_de_semana**
- Foi utilizada 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, confiança e lift**
- **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.

Resultados - FP growth - Consequente Ameaça



- **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**.
- A variável **rmbh:Interior de MG** já é suficiente para compor uma regra com **suporte elevado (24,3%) e confiança próxima a 29,6%**, o que reforça a importância dessa variável na **composição do perfil territorial desse tipo de violência**.

Metodologia - FP Growth

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

Resultados - FP Growth - Lift superiores a 7



- O resultados 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 (Contagem, Betim e Ribeirão das Neves)**.
- Regras complementares revelaram que **ocorrências fora do fim de semana e nos municípios da RMBH (sem incluir a capital)**, possuem uma **probabilidade superior a 58%** de estarem sob responsabilidade do **2º Departamento da RISP** e resultar em **crime consumado**.

Resultados - FP Growth - Lift superiores a 7

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

SD - Descoberta de Subgrupos

Foco: Reincidência de violência doméstica

SD (Overview)



1. Pré-processamento
2. Geração da variável alvo: reincidência
 - a. Repetição da combinação de município e tipo de delito
3. Configuração para SD
 - a. Atributos considerados: `municipio_fato`, `mes`, `resp`, `rmbh`, `natureza_delito`, `dia_da_semana`, `fim_de_semana`
 - b. Medida de qualidade: WRAcc
 - c. Profundidade máxima da condição: 3
4. Geração dos subgrupos
 - a. Qualidade (WRAcc)
 - b. Tamanho do subgrupo (`size_sg`)
 - c. Casos positivos dentro do subgrupo (`positives_sg`)

Resultados - SD - Maior Reincidência



- 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.
- Os subgrupo **AMEAÇA no interior de MG**, que apresentou uma taxa de reincidência próxima de 100%, e **AMEAÇA durante os dias úteis**, reforçam a hipótese de que esse tipo de delito ocorre de maneira recorrente e **previsível em certos contextos temporais e territoriais**.
- Delitos como **LESÃO CORPORAL** e **VIAS DE FATO / AGRESSÃO** 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. 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**

Metodologia - SD



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.
VIAS DE FATO / AGRESSAO	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.
LESAO CORPO- RAL	9.780	9.679	0,0100	Lesões corporais também se repetem com frequência.

SD - Descoberta de Subgrupos

Foco: Casos consumados de violência doméstica

SD (Overview)



1. Pré-processamento
2. Geração da variável alvo: reincidência
 - a. `target = 1` se `tentado_consumado == 'CONSUMADO'`, 0 caso contrário.
3. Configuração para SD
 - a. Atributos considerados: `municipio_fato`, `mes`, `resp`, `rmbh`, `natureza_delito`, `dia_tipo`, `dia_categoria`
 - b. Alvo: `target = 1` (casos consumados).
 - c. Medida de qualidade: WRAcc
4. Geração dos subgrupos
 - a. Qualidade (WRAcc)
 - b. Tamanho do subgrupo (`size_sg`)
 - c. Casos positivos dentro do subgrupo (`positives_sg`)

Resultados - SD - Delitos Consumados



Subgrupo	Cobertura (%)	Taxa de Consumo (%)	WRAcc
fim de semana	34.0%	99.20%	0.00053
descumprimento medida protetiva	6.6%	100.00%	0.00062
lesão corporal	16.0%	99.55%	0.0008
lesão corporal Λ Interior de MG	14.1%	99.65%	0.00084
fim de semana Λ Interior de MG	28.3%	99.35%	0.00085

Resultados - SD - Delitos Consumados

Foram geradas regras com **alta proporção de casos consumados**, avaliadas por métricas como **WRAcc, lift e target share**:

Subgrupo	Cobertura (%)	Taxa de Consumo (%)	WRAcc
fim de semana	34.0%	99.20%	0.00053
descumprimento medida protetiva	6.6%	100.00%	0.00062
lesão corporal	16.0%	99.55%	0.0008
lesão corporal \wedge Interior de MG	14.1%	99.65%	0.00084
fim de semana \wedge Interior de MG	28.3%	99.35%	0.00085

- Delitos nos **finais de semana no interior de Minas Gerais** formam o subgrupo com **maior quality**, abrangendo **28,3% dos registros e apresentando uma taxa de consumo 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 **ocorrendo no interior com 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**

Resultados - SD - Delitos Consumados



Continuando...

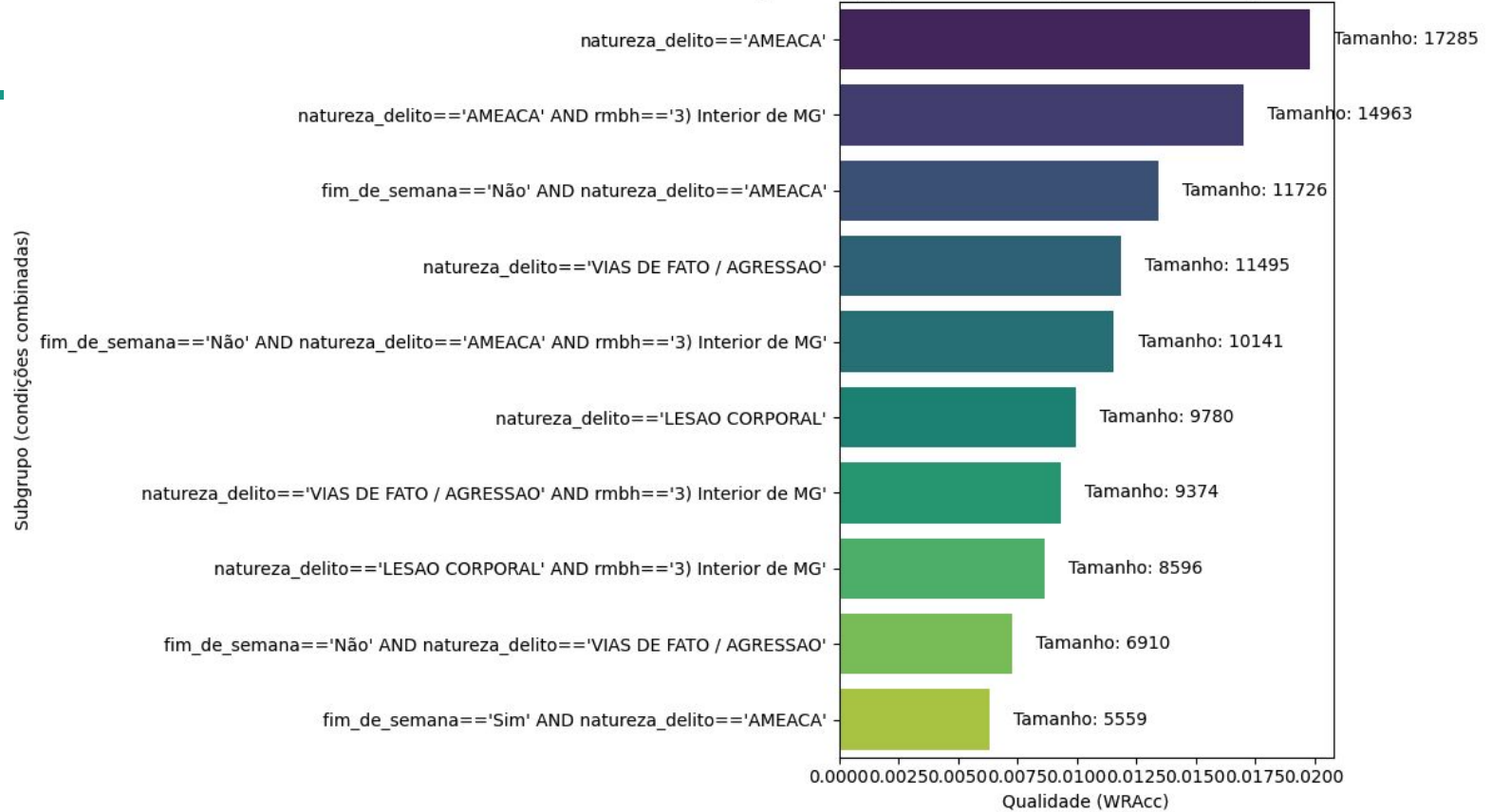
- O interior de Minas Gerais concentra 82% da base analisada e apresenta taxa de **consumação de 99,1%**, reforçando um padrão de **risco geográfico elevado**. Além disso, **domingos no interior** compõem outro cenário de vulnerabilidade, com **99,34% de consumação**
- **Descumprimento de medidas protetivas foi consumado** em todos os registros encontrados (**100%**), isso se deve à natureza do crime (nesse caso não há tentativa).

Resultados - SD

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

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

Top 10 Subgrupos com maior qualidade (WRAcc) para reincidência



Conclusão



- 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
- 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
- O uso desses métodos poderiam auxiliar na **priorização de casos com maior risco de reincidência ou consumação**, contribuindo para a otimização dos recursos disponíveis. Contudo é necessário avaliar **vieses estruturais dos dados**, como discriminação geográfica ou social. 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

Conclusão - Limitações e Comentários



- A análise se baseia em **dados** administrativos, que **podem conter subnotificações ou inconsistências**.
- A definição de **reincidência** 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.
- A **análise não considerou variáveis socioeconômicas ou de perfil demográfico**, o que restringe a compreensão dos fatores estruturais associados à violência.
- 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.
- As interpretações propostas **devem ser vistas como indícios estatísticos que devem ser analisados em conjunto com o conhecimento de especialistas da área**.

Conclusão - Trabalhos futuros



Trabalhos futuros podem ampliar esta abordagem incluindo:

- Dados com variáveis contextuais mais ricas e métodos híbridos que combinem estatística descritiva com aprendizado de máquina interpretável.
- Parcerias com instituições públicas e organizações da sociedade civil podem validar os achados e promover sua aplicação ética e responsável.

Obrigado!
Dúvidas?
