

# Dados da Violência contra a Mulher no Município de Duque de Caxias

## Resumo

## Introdução

Uma em cada 3 mulheres no mundo possui experiência de violência física e/ou sexual (WORLD HEALTH ORGANIZATION *et al.*, 2020). Em tempos de emergência, a violência contra as mulheres continua ameaçando seriamente a saúde da população, especialmente das mulheres, em todo mundo, sendo que o tipo mais recorrente de violência é aquela realizada pelo seu próprio parceiro íntimo (WORLD HEALTH ORGANIZATION *et al.*, 2020).

**Segundo a Organização Mundial de Saúde, violência é** o uso intencional de poder ou força física, em forma de ameaça ou indo às vias de fato, contra si mesmo, outra pessoa, ou contra um grupo ou comunidade, que resulta em ou tem alta probabilidade de resultar em ferimentos, morte, dano psicológico, desenvolvimento deficiente ou privação (WORLD HEALTH ORGANIZATION *et al.*, 2020). **Violência contra a mulher é,** segundo (BURELOMOVA *et al.*, 2018), o mau uso do poder pelo parceiro íntimo (homem ou mulher), que resulta em perda de dignidade, controle e segurança, bem como sentimento de impotência e aprisionamento experimentado pela mulher que é vítima direta de problemas físicos, psicológicos contínuos ou repetidos, abuso econômico, sexual, verbal e/ou espiritual. Violência contra a mulher também inclui ameaças ou forçar mulheres a testemunharem violência contra seus filhos, parentes, amigos, animais de estimação e/ou bens queridos de seus maridos, parceiros, ex-maridos ou ex-sócios.

No Brasil, a Lei Maria da Penha (Lei nº 11.340 de 2006) prevê a criação de um Sistema Nacional de Dados e Estatísticas sobre a Violência Doméstica e Familiar contra a Mulher, que [deveria ter sido] implementado nos quatro anos seguintes a sua publicação pela Secretaria Especial de Políticas para as Mulheres (SPM) em articulação com outros Ministérios e órgãos da Administração Pública. **O problema é** que a ausência de bases de dados detalhadas sobre a violência contra a mulher ou até mesmo a sua existência porém espalhada não favorece o reconhecimento da gravidade e alcance dessa situação para a população, além das suas sérias consequências para a saúde e segurança das mulheres do nosso país (xxxx). Nota-se ainda que existe um agravamento do problema em regiões de maior vulnerabilidade econômica dos habitantes (referência).

Compreende-se que é urgente reunir dados e aperfeiçoar sua coleta de maneira que, de fato, reflita a realidade vivenciada e oriente ações públicas capazes de garantir não apenas qualidade de vida, mas ela própria às mulheres brasileiras.

Neste contexto, reconhecendo a pulverização das bases de dados, este trabalho tem como objetivo localizar e conhecer a base de dados referente ao registro de atendimentos a mulheres vítimas de violência doméstica do município de Duque de Caxias

para a realização de aplicações de *Data Science*. A escolha deste município se deu por ser estar em região de vulnerabilidade econômica e pela possibilidade de acesso às informações. Este trabalho contribui para a disponibilização de uma base de dados consistente possibilitando que outras organizações, pesquisadores e gestores produzam ou utilizem modelos para automatização da visualização dos dados para geração de conhecimento e publicações atualizadas sobre os dados de violência doméstica, fundamentando, inclusive, a intensificação de ações de combate a este tipo de crime.

Este artigo está organizado da seguinte forma.....

## Fundamentação teórica

### 2.1 Anonimização como forma de proteção de dados

Instaurada em 2018, a Lei Nº 13.709, de 14 de Agosto de 2018, Lei Geral de Proteção de Dados Pessoais (LGPD) prevê o tratamento dos dados pessoais de pessoa física ou jurídica de direito público ou privado para a preservação dos direitos fundamentais de liberdade, privacidade e o livre desenvolvimento da personalidade. Entre as normas gerais contidas nesta lei objetiva-se a preservação do direito de pessoa, incluindo a anonimização de dados que possam identificar direta ou indiretamente um indivíduo.

A anonimização, segundo a referida Lei, trata-se de “utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo”. Durante a análise de uma base de dados é possível haver dados pessoais que identifiquem um indivíduo e essa vinculação não seja desejada. Para fins de preservação, esses dados se tornam em dados anonimizados (BRASIL, 2018) e, para isso, técnicas específicas são aplicadas para a desvinculação dos dados sem que prejudique o objetivo de utilização da base de dados.

Segundo a própria definição, anônimo é aquele que não apresenta nome ou assinatura, indivíduo desconhecido (MICHAELIS, 2021). A definição do conceito de dados pessoais pode seguir uma orientação expansionista (a partir da delimitação de “pessoa identificável”) ou reducionista (“pessoa identificada”), respectivamente alargando ou restringindo o escopo de aplicação da lei (BIONI, 2019). A princípio, o processo de anonimização pode estar atrelado aos dados pessoais numa orientação reducionista, contudo, é necessário uma análise contextual para verificar se na combinação de outros dados, como num quebra-cabeça, a identificação seria possível, tornando a pessoa identificável, segundo a orientação expansionista.

Neste sentido, há técnicas que buscam eliminar elementos identificadores de uma base de dados. Biondi (2019) esclarece que o processo de anonimizar um dado é composto por técnicas que buscam eliminar elementos identificadores ou identificáveis de uma base de dados. O autor aponta quatro técnicas para nortear o processo: (i) a supressão; (ii) generalização; (iii) randomização e; (iv) pseudoanonimização. A escolha da técnica ou da combinação delas precisa considerar os dados que devem ter seus vínculos quebrados com seus respectivos titulares.

A **supressão** se trata da exclusão de campos da tabela de dados ou na substituição de parte de caracteres do campo de identificação. Em alguns casos, a substituição de caracteres pode não ser suficiente para anonimizar. É o caso do CPF. Esse é um campo

que trata a identificação exclusiva de um indivíduo e mesmo com a substituição de muitos caracteres pode não ser suficiente. Nesta situação, vale a pena considerar a exclusão do campo.

Outras situações podem considerar a supressão apontando para a **generalização**. Um exemplo é a supressão do CEP. Os últimos dígitos de um CEP podem ser excluídos e ainda assim não ser suficiente para gerar a anonimização, ainda mais se combinados com outros dados parte do conjunto da base. Assim, a generalização propõe a substituição de um dado por outro que traz significado mais geral. e modo mais prático, neste exemplo, o CEP poderia ser substituído pela nomenclatura de uma região, ou uma data de nascimento poderia ser substituída por uma faixa etária.

A **randomização** é uma técnica que busca mascarar uma informação, misturando os valores da tabela, sem afetar as análises, apenas visando não identificar seus titulares. Considerando uma tabela com a abordagem relacional, os dados de uma coluna (um campo) são alternados entre suas linhas em suas ordens, sem alterar seus valores.

O **pseudoanonimização** é uma técnica baseada na definição do termo pela LGPD que diz respeito ao “tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro”. Nesta técnica a base de dados é dividida em duas partes, a primeira fica com os dados genéricos enquanto na outra parte fica com os dados sensíveis.

Não há uma técnica melhor ou combinação perfeita para direcionar o processo de anonimização, sendo essencial analisar cada contexto onde o tratamento será implementado para que os indivíduos titulares dos dados anonimizados não sejam reidentificados.

## 2.2. Dados sobre a violência doméstica

A Lei Maria da Penha classifica as formas de violência doméstica e familiar contra a mulher, como: violência física, violência psicológica, violência sexual, violência patrimonial e a violência moral. No artigo 38 a referida lei indica que “as estatísticas sobre a violência doméstica e familiar contra a mulher serão incluídas nas bases de dados dos órgãos oficiais do Sistema de Justiça e Segurança a fim de subsidiar o sistema nacional de dados e informações relativas às mulheres”. Contudo, após anos da publicação desta lei, não foi possível identificar um sistema nacional de dados como indicado.

Em 2016, o Núcleo de Estudos e Pesquisas da Consultoria Legislativa do Senado nacional apontou as lacunas existentes nas políticas de enfrentamento da violência contra a mulher, expressando a relevância de um banco de dados unificado acerca desse tema (Alves, Dumaresq e Silva, 2016). O estudo indicou a existência de bases de dados pulverizadas e como os órgãos utilizam-se dessas bases parciais para políticas e análises dos dados sobre o assunto.

O Observatório da Mulher Contra a Violência—órgão instituído pelo Senado Federal em 2016 para funcionar em conjunto com o Instituto DataSenado com a função de reunir e sistematizar as estatísticas oficiais sobre a violência contra a mulher— inaugurou, em março de 2019 o recurso Painel de Violência Contra Mulheres que relaciona dados de diversas fontes para oferecer um panorama da violência contra a mulher no Brasil. O painel se obtém de dados coletados através do Ministério da Saúde e Conselho Nacional de Justiça, mais especificamente do Sistema de Mortalidade do Ministério da Saúde, Sistema de Informação de

Agravos de Notificação do Ministério da Saúde, Pesquisa Nacional por Amostra de Domicílios do IBGE, Secretarias de Segurança Pública estaduais, Conselho Nacional de Justiça e Disque 180 da Secretaria de Políticas para Mulheres.

A navegação por este painel contempla a visualização de dados de 2011 a 2018 com aplicação de filtros por ocorrência: homicídios de mulheres, notificações de saúde, boletins de ocorrência e poder judiciário. Adicionalmente, podem ser aplicados filtros como ano, UF, raça, escolaridade entre outros. Contudo, ao realizar testes de visualizações para este estudo, foram constatadas algumas inconsistências como a ilustrada na Figura 1, fonte de questionamento quanto a sua discrepância frente à realidade.

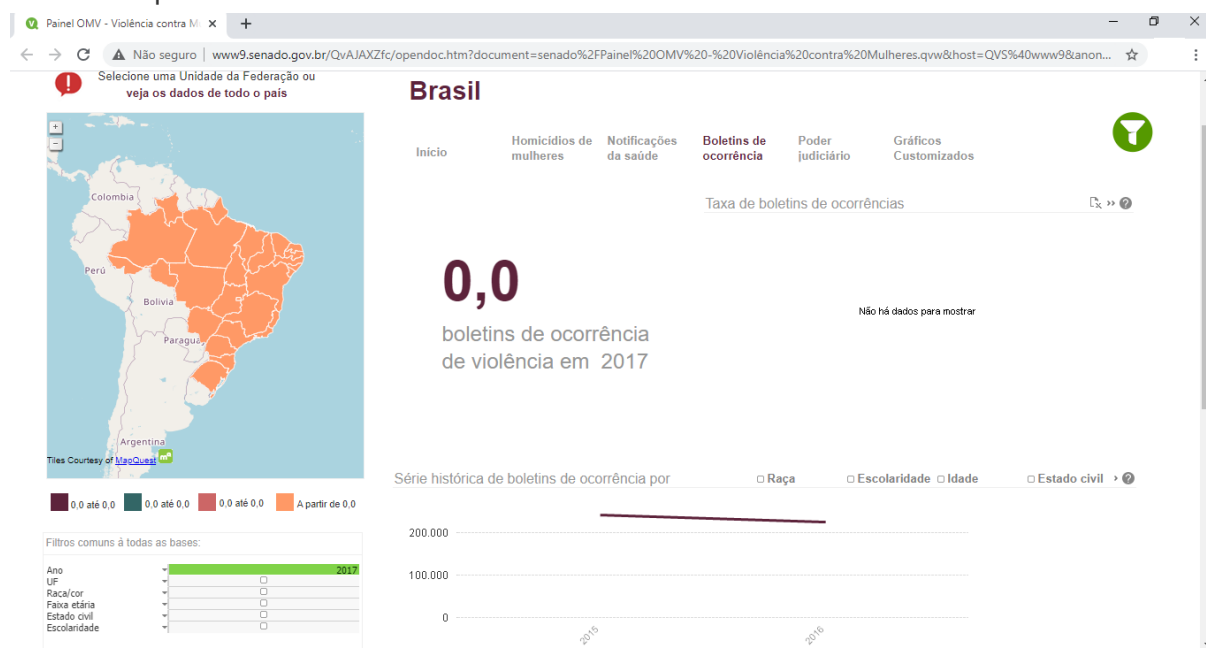


Figura 1: Painel Violência contra Mulher, aplicação de filtro por ano 2017 para Boletins de Ocorrência

Parece intrigante não haver registros de boletins de ocorrência de violência em 2017 para todo o território nacional. Isso traz a inquietação sobre a reunião dos dados como indicado nos textos lidos, refletindo a percepção de que as bases de dados não estão unificadas.

## Metodologia

A metodologia empregada na realização deste estudo envolveu, inicialmente, a revisão da literatura para contextualizar o cenário de pesquisa atual, buscando identificar inconsistências e gerar inquietações que direcionassem este trabalho. Também foram realizadas buscas por bases de dados abertos governamentais relacionados ao tema de pesquisa, considerando os últimos 10 anos que, em teoria, corresponde ao tempo máximo indicado na lei Maria da Penha para a implantação do Sistema Nacional de Dados e Estatísticas sobre a Violência Doméstica e Familiar contra a Mulher.

A partir da constatação da dificuldade de encontrar uma base de dados unificada sobre o tema e a improbabilidade de acesso público, este estudo empenhou esforços em localizar os dados referentes aos registros de violência doméstica contra mulher do município de Duque de Caxias. Para isso, a primeira forma de busca aconteceu através de

contato telefônico com a Secretaria de Assistência Social do município que direcionou os pesquisadores para o Centro Especializado de Atendimento à Mulher (CEAM) Caxiense.

#### 4.1 - Tipo de Pesquisa

Para responder à questão de pesquisa a análise foi conduzida em abordagem qualitativa a partir do estudo exploratório da literatura e trabalhos de pesquisa anteriores, levantamentos de documentação e comparações entre eles, incluindo leis federais relacionadas ao tema e identificação de respectivas bases de dados abertos. Há a proposta de implementação de uma metodologia de Data Science para análise, estruturação e consistência para a base de dados definida realizado nas seguintes etapas: (i) identificação e download da base de dados em formato XLSX; (ii) instruções computacionais para carga dos dados, tratamento e estruturação do conteúdo dos arquivos de dados; (iii) seleção, análise, testes de hipóteses e validação das informações oriundas dos dados estruturados.

#### 4.2 - Seleção de *DataSets*

Para a realização deste trabalho foram utilizados os dados coletados desde 2017 até o início desta pesquisa, em março de 2021. No total foram obtidos 5 (cinco) *datasets* organizados por ano e com campos equivalentes ao indicado na ficha de acompanhamento à mulher vítima de violência doméstica. Originalmente, o *dataset* possui a estrutura de campos listados na Tabela 1 que dizem respeito ao atendimento à vítima.

Tabela 1: Estrutura original do dataset CEAM

<b>Campos</b>	<b>Descrição</b>
Ficha/Ano	Identificador da ficha de acompanhamento com número de ordem e o ano do atendimento
Usuária	Nome da pessoa atendida
Contato	Número de telefone para contato
Técnica De Ref	Nome do profissional que realizou o atendimento à vítima
Religião	Religião da vítima atendida
Grau de Escolaridade	Grau de formação escolar
Cor	Cor da pele
Idade	Idade em anos
Município	Município de residência
Bairro	Bairro de residência
Distrito	Número de identificação do distrito onde o bairro está localizado dentro do município de residência
Origem	De onde a vítima foi encaminhada ao CEAM
Tipos de VD	Tipos de violência doméstica sofrida pela vítima

Encam.	Encaminhamento dado à vítima a partir do atendimento no CEAM
Data	Data do atendimento da vítima no CEAM
Com Renda	Indicação se a vítima possui ou não renda
Filhos	Quantidade de filhos
Rel. Agressor	Parentesco ou relação do agressor face à vítima

#### 4.3 - Proveniência de Dados

Por conta da impossibilidade de identificação de uma base de dados unificada acerca do objeto de pesquisa, este estudo foi realizado sobre a base de dados do CEAM - Duque de Caxias. A base contém registros dos atendimentos a mulheres vítimas de violência doméstica e está em planilha eletrônica que reflete a população atendida no período de outubro de 2017 a março de 2021, época da realização desta pesquisa.

Em entrevista com os especialistas, foi identificado que a planilha de dados é uma iniciativa independente. Não há a indicação para alimentar uma base integrada no município ou em instâncias governamentais superiores. Contudo, os especialistas relataram que remetem relatórios periódicos aos gestores.

A base de dados utilizada para esse estudo foi cedida pela diretoria do CEAM aqui citado. Não há permissão de acesso público a ela, bem como sua disponibilidade como base de dados abertos. A partir da Tabela 1 é possível perceber a sensibilidade na natureza dos dados contidos no *dataset*. Foi requerido aos pesquisadores deste estudo um termo de uso de dados e anonimato dos indivíduos listados nos registros.

Esta pesquisa se reserva na obrigação de não publicar a base de dados e seu conteúdo.

#### 4.5 - Anonimização dos dados

O processo de anonimização não garante total eficiência na desvinculação do dado a identificação do seu titular. Para (NARAYANAN; SHMATIKOV, 2010) afirmar o anonimato das pessoas é um mito. Por meio da combinação dos campos do *dataset* o sujeito (titular do dado) pode ser identificado (BIONI, 2020). Essas afirmações fazem muito sentido para o *dataset* obtido para esse trabalho.

Com a aplicação de maior grau de técnicas de anonimização - supressão e generalização - não parece eficaz tornar o indivíduo indetectável e preservar os dados relevantes para análise. Para esse estudo, foram consideradas as técnicas de supressão para os campos identificadores explícitos (Nome e Contato) e o pseudoanonimato para a divisão deste *dataset* em partes. Ainda assim, tais medidas não são suficientes para inviabilizar a identificação.

#### 4.4 - Análise dos dados

Durante a análise da base de dados foi constatado haver a identificação de indivíduos através do registro de seus dados pessoais e também a inconsistência no preenchimento de outros campos quanto a falta de padronização para os possíveis valores. Percebeu-se que para um mesmo possível valor de campo, foi registrado dois ou até mais nomenclaturas distintas. Como exemplo, os termos “*psi*”, “*psico*”, “*psic*” ou até mesmo

“*psicologics*” foram adotados como forma de registrar o termo “Psicológico”. A falta de padronização prejudica a organização dos dados na base e, consequentemente, as possíveis consultas que podem ser geradas por ela.

## Resultados obtidos

## Conclusão

## Referência

ALVES, M. C. L.; DUMARESQ, M. L.; SILVA, R. V. As Lacunas no Enfrentamento à Violência contra a Mulher: análise dos bancos de dados existentes acerca da vigilância doméstica e familiar. Brasília: Núcleo de Estudos e Pesquisas/CONLEG/Senado, abril/2016 (Texto para Discussão nº 196). Disponível em: [www.senado.leg.br/estudos](http://www.senado.leg.br/estudos). Acesso em 25 de março de 2021.

BIONI, B. R. Proteção de dados pessoais: a função e os limites do consentimento. Rio de Janeiro: Forense, 2019.

BIONI, B. R. Compreendo o conceito de anonimização e dado anonimizado. Cadernos Jurídicos, São Paulo, ano 21, nº 53, p. 191-201, Janeiro-Março/2020

BRASIL, Lei Maria da Penha .Lei nº. 11.340, de 7 de agosto de 2006.

BRASIL. Lei Geral de Proteção de Dados Pessoais. Lei nº13.709, de 14 de agosto de 2018.

Buneman, P., Khanna, S., and Wang-Chiew, T. (2001). Why and where: A characterization of data provenance. International Conference on Database Theory, 1:316–330.

BURELOMOVA, Anastasia S.; GULINA, Marina A.; TIKHOMANDRITSKAYA, Olga A. Intimate partner violence: An overview of the existing theories, conceptual frameworks, and definitions. **Psychology in Russia: State of the art**, v. 11, n. 3, p. 128-144, 2018.

MICHAELIS. Dicionário Brasileiro da Língua Portuguesa. Disponível em <https://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=an%C3%B4nimo> Acesso em Abril de 2021

WORLD HEALTH ORGANIZATION et al. **COVID-19 and violence against women: what the health sector/system can do, 7 April 2020**. World Health Organization, 2020.

[https://app.diagrams.net/#G15qOcRGukcSi2Oyi4bLAZY08V\\_JIMLL6y](https://app.diagrams.net/#G15qOcRGukcSi2Oyi4bLAZY08V_JIMLL6y)