# **Construindo uma Base de Dados sobre Violência contra a Mulher no Município de Duque de Caxias**

# Alayne Duarte Amorim, Luciana Sá Brito

[alayneduarte@cp2.g12.br](mailto:alayneduarte@cp2.g12.br), lbrito@cecierj.edu.br

**Abstract.** Violence against women is a serious public health and safety problem in Brazil. However, there is a lack of reliable data about the problem, even as unified and open national databases. This research reports the steps for the generation of a domestic violence database committed in the municipality of Duque de Caxias to provide a suitable database for data mining procedures. The data are from women attended by a specialized center. Following the same steps, it is possible to reproduce the base creation process. This research aims to recognize the seriousness of the problem by population and public management.

**Resumo.** A violência contra a mulher é um grave problema de segurança e saúde pública no Brasil. Entretanto, os números divulgados pela mídia e sítios do governo parecem não retratar a realidade e a ausência de base unificada de dados abertos não favorece o reconhecimento da gravidade e alcance da situação pela população e gestão pública. Esta pesquisa relata a construção de uma base de dados através da aquisição e tratamento de dados de vítimas de violência doméstica do município de Duque de Caxias atendidas em um centro especializado. A partir da descrição das etapas é possível a reprodutibilidade do processo de criação da base.

# **Introdução**

Uma em cada 3 mulheres no mundo possui histórico 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).

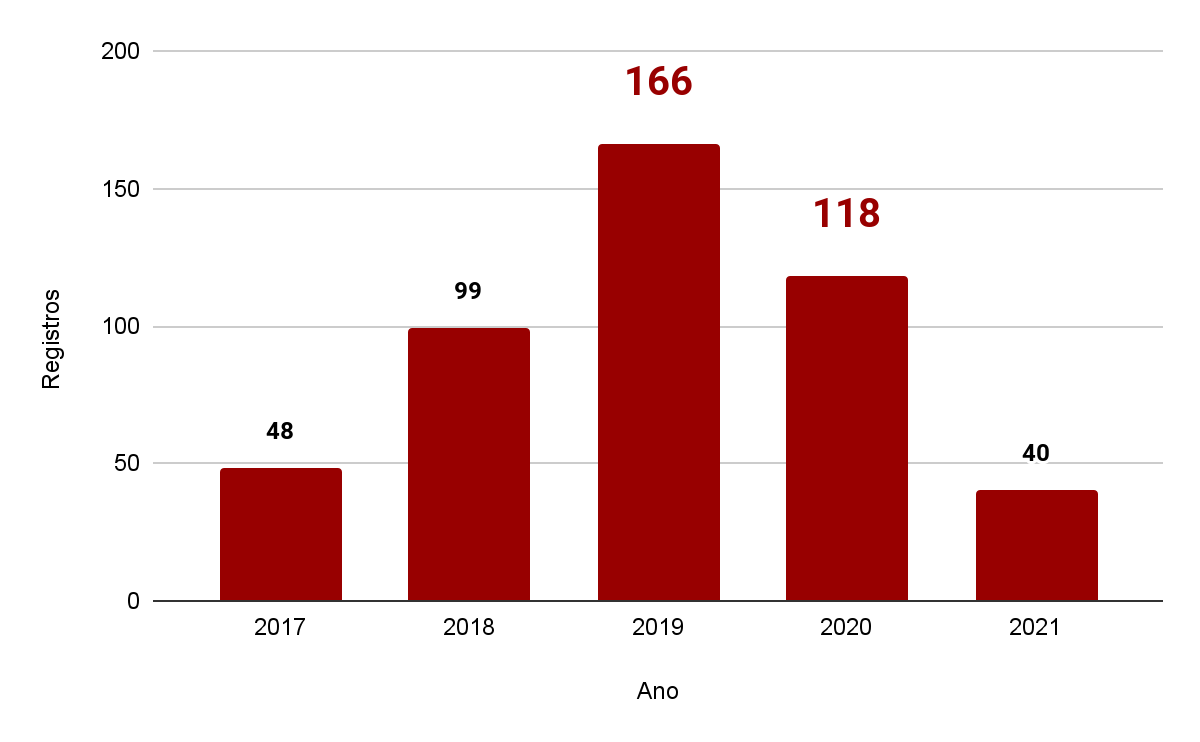
Segundo a Organização Mundial de Saúde, violência é o uso intencional de poder ou força física, em forma de ameaça ou indo às vias de fato, contra si mesmo, outra pessoa, ou contra um grupo ou comunidade, que resulta em ou tem alta probabilidade de resultar em ferimentos, morte, dano psicológico, desenvolvimento 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 à 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 de forma dispersa não favorece o reconhecimento da gravidade e alcance dessa situação para a população, nem das suas sérias consequências para a saúde e segurança das mulheres do nosso país (ALVES; DUMARESQ; SILVA, 2016). Nota-se ainda que existe um agravamento do problema em regiões de maior vulnerabilidade social das habitantes (BORBUREMA *et al*., 2017).

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 relatório tem como objetivo localizar o dataset referente ao registro de atendimentos a mulheres vítimas de violência doméstica do município de Duque de Caxias e criar uma base de dados consolidada para a realização de aplicações de Ciência de Dados.

A escolha do município de Duque de Caxias se deu por ser uma região historicamente marcada pela vulnerabilidade social e pela possibilidade de acesso às informações através do Centro Especializado de Atendimento à Mulher (CEAM) Caxiense. Esta pesquisa 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.

Gráfico 1: Relação do número de registros de violência doméstica entre os anos 2017 e 2021, com base nos dados do CEAM. A baixa quantidade de registros nos anos de 2017 e 2021 revelam o início da contagem e contagem ainda em curso.



Este artigo está organizado da seguinte forma: na seção 2 está a fundamentação teórica que norteia as decisões tomadas no processo de aquisição do *dataset*, no tratamento dos dados e geração da base consolidada. Na seção 3 está a descrição da metodologia adotada no desenvolvimento deste trabalho. A seção 4 mostra os resultados obtidos. A seção 5 traz os próximos passos.

# **Fundamentação teórica**

Para possibilitar o estudo envolvido neste relatório, foi realizado estudo da Lei Geral de Proteção de Dados Pessoais (LGPD) e da Lei Maria da Penha, de literatura técnica sobre anonimização de dados, além da investigação de sítios com dados sobre violência doméstica mantidos pelo Governo Federal do Brasil.

# **Anonimização como forma de proteção de dados**

Instaurada em 2018, a [Lei Nº 13.709, de 14 de Agosto de 2018](http://legislacao.planalto.gov.br/legisla/legislacao.nsf/Viw_Identificacao/lei%2013.709-2018?OpenDocument), a 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 da pessoa seria possível, 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 a 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 em que o tratamento será implementado para que os indivíduos titulares dos dados anonimizados não sejam reindentificados.

# **Cenário: 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[[1]](#footnote-1) que relaciona dados de diversas fontes para oferecer um panorama da violência contra a mulher no Brasil. O painel é carregado com 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](http://sinan.saude.gov.br/sinan/) 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, que aponta a inexistência de boletins de ocorrência de violência contra mulheres no ano de 2017 ─ 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

É intrigante não haver registros de boletins de ocorrência de violência doméstica em 2017 para todo o território nacional. Isso corrobora com a inquietação acerca da urgência da reunião dos dados como indicado na literatura, porque reflete a percepção de que as bases de dados nacionais sobre violência contra a mulher 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.

**3.1. Tipo de Pesquisa**

Para responder à questão de pesquisa a análise foi conduzida em abordagem qualitativa a partir do estudo exploratório da literatura e pesquisas 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, realizada nas seguintes etapas: (i) identificação e download do *dataset* de violência doméstica do CEAM em formato XLSX; (ii) instruções computacionais para carga dos dados, tratamento e estruturação do conteúdo dos arquivos de dados;

**3.2. 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, os *datasets* possuíam a estrutura de campos listados na Tabela 1 que dizem respeito ao atendimento à vítima.

**Tabela 1: Estrutura original dos datasets CEAM**

|  |  |
| --- | --- |
| **Campos** | **Descrição** |
| 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 |

**3.3. Coleta de dados**

Por conta da impossibilidade de identificação de uma base de dados unificada acerca do objeto de pesquisa, este estudo foi realizado sobre os dados do CEAM - Duque de Caxias. O dataset contém registros dos atendimentos a mulheres vítimas de violência doméstica e está dividido em 5 planilhas eletrônicas que refletem as características da população atendida, bem como dados contextuais sobre a violência sofrida no período de outubro de 2017 a março de 2021, época da realização desta pesquisa.

Em entrevista com os especialistas, foi identificado que as planilhas de dados fazem parte de uma iniciativa independente, uma vez que não há nenhuma indicação de que esses dados alimentarão uma base integrada no município ou em outras instâncias governamentais. Contudo, os especialistas relataram que remetem relatórios periódicos aos gestores a partir dos dados do CEAM.

Os dados utilizados neste estudo foram cedidos pela diretoria do CEAM. Devido à natureza sensível dos dados, o centro não os disponibiliza como dados abertos nem autoriza seu acesso ao público. Por esse motivo, 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.

**3.4. Anonimização dos dados**

O processo de anonimização não garante total eficiência na desvinculação dos dados da identificação das suas titulares. 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 esta pesquisa.

Inicialmente o dataset estava organizado em 5 tabelas .XLS, correspondentes aos anos de coleta de dados (cadastro\_ceam\_2017, cadastro\_ceam\_2018, cadastro\_ceam\_2019, cadastro\_ceam\_2020 e cadastro\_ceam\_2021). As 5 tabelas foram importadas para um banco de dados MySql, através do software Navicat Premium versão 10.0.3 e unidas em uma tabela única, para que fosse iniciado o procedimento para a anonimização dos dados sensíveis.

A aplicação das técnicas de supressão e generalização para a anonimização não parece eficaz na tarefa de 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 geração de uma base de dados dividida em duas tabelas (cadastro\_ceam\_usuarias e cadastro\_ceam\_atual) em um banco de dados MySql relacionadas através de um campo id. Ainda assim, tais medidas não são suficientes para inviabilizar a identificação das usuárias cadastradas na base.

Na tabela cadastro\_ceam\_usuarias foram reunidas 5 colunas relativas aos dados que representam as informações julgadas como as mais sensíveis da base ― Ficha/ano, Usuária, Contato e Idade, mais a coluna id. Na tabela cadastro\_ceam\_atual foram reunidas as demais colunas constantes na tabela 1, de modo que qualquer tarefa de mineração de dados pode ser realizada sem a necessidade de compartilhamento dos dados contidos na tabela cadastro\_ceam\_usuarias. A estrutura de dados das tabelas da base estão representadas nas tabelas 2 e 3.

**Tabela 2: Estrutura de dados da tabela cadastro\_ceam\_usuarias, com os dados pessoais das usuárias.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Type** | **Length** | **Decimals** | **Allow Null** |
| id | int | 11 | 0 |  |
| ficha\_ano | varchar | 255 | 0 | x |
| ano | varchar | 255 | 0 | x |
| usuaria | varchar | 255 | 0 | x |
| contato | varchar | 255 | 0 | x |
| idade | varchar | 255 | 0 | x |

**Tabela 3: Estrutura de dados da tabela cadastro\_ceam\_atual. Com os dados efetivamente utilizados na análise.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Type** | **Length** | **Decimals** | **Allow Null** |
| id | int | 11 | 0 |  |
| ficha\_ano | varchar | 255 | 0 | x |
| ano | varchar | 255 | 0 | x |
| tecnica\_de\_ref | varchar | 255 | 0 | x |
| idade | varchar | 255 | 0 | x |
| bairro | varchar | 255 | 0 | x |
| municipio | varchar | 255 | 0 | x |
| distrito | varchar | 255 | 0 | x |
| origem | varchar | 255 | 0 | x |
| tipos\_de\_vd | varchar | 255 | 0 | x |
| encam | varchar | 255 | 0 | x |
| dia | varchar | 255 | 0 | x |
| religiao\_SN | varchar | 255 | 0 | x |
| grau\_escolaridade | varchar | 255 | 0 | x |
| cor | varchar | 255 | 0 | x |
| com\_renda | varchar | 255 | 0 | x |
| filhos | varchar | 255 | 0 | x |

A partir dessa organização, iniciou-se a tarefa de tratamento dos dados existentes na tabela cadastro\_ceam\_atual. Durante a fase de tratamento dos dados, foram realizadas ações para a sua limpeza e transformação.

**3.5. Limpeza dos dados**

O processo de limpeza de dados é uma parte integrante da análise de dados como um todo e representa o processo de fixar ou remover dados formatados incorretamente, corrompidos, duplicados ou incompletos.

Durante a análise exploratória dos dados contidos na tabela cadastro\_ceam\_atual, constatou-se o uso de múltiplos valores para a representação de uma mesma informação dentro de uma mesma coluna da tabela, como podemos ver na tabela 4. Esse problema deveu-se ao fato de que a anotação das informações contidas no dataset originário do CEAM foi realizada de forma manual, ocasionando problemas de digitação e ausência de padronização dos valores. Para resolver esse problema e garantir que a base pudesse retornar consultas SQL de modo eficiente, foi necessário realizar procedimentos para a normalização dos valores das colunas da tabela cadastro\_ceam\_atual.



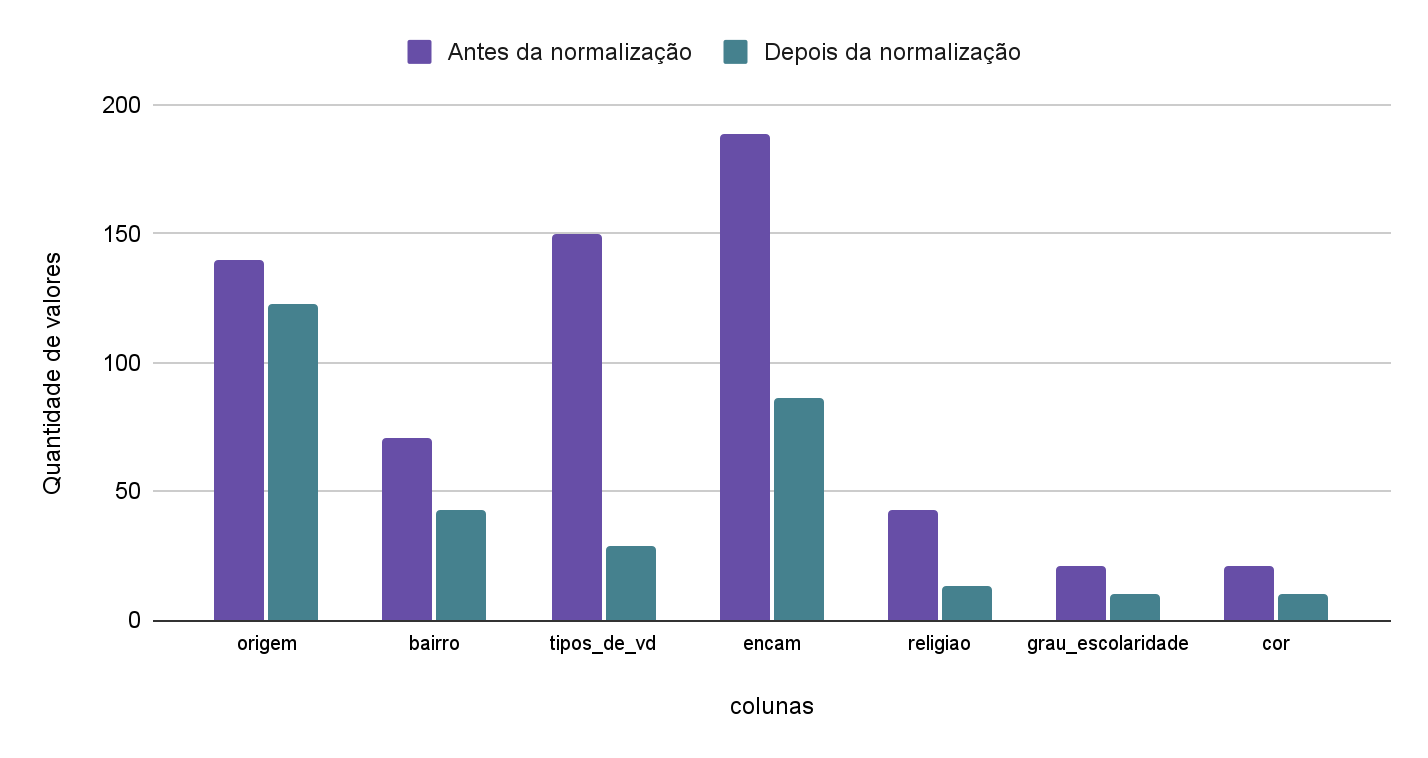
Figura 2: Distintos valores encontrados inicialmente no *dataset* CEAM, coluna “religião” para o ano de 2019 antes (à esquerda) e depois da limpeza de dados (à direita).

O procedimento de normalização consistiu em identificar quais os valores distintos presentes em cada uma das colunas da tabela e padronizar formas variadas que estivessem representando um mesmo valor. Esse procedimento foi repetido para todas as colunas da base, resultando em diminuição substancial dos valores distintos em cada uma das colunas, como se pode observar na tabela 4.

**Tabela 4: Uma amostra das colunas da tabela cadastro\_ceam\_atual que tiveram seus valores normalizados durante o estudo.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Coluna** | **origem** | **bairro** | **tipos\_de\_vd** | **encam** | **religiao** | **grau\_escolaridade** | **cor** |
| Antes | 140 | 71 | 150 | 189 | 43 | 21 | 21 |
| Depois | 123 | 43 | 29 | 86 | 13 | 10 | 10 |
| **Redução** | **12%** | **39%** | **81%** | **54%** | **70%** | **52%** | **52%** |

Gráfico 2: Relação entre os valores das colunas antes e depois dos procedimentos adotados para a normalização.



**3.6. Transformação dos dados**

Após o procedimento de normalização obteve-se uma tabela mais apropriada para a realização de análises e dados e criação de visualizações. Entretanto, ainda havia a necessidade de obter dados monovalorados para as colunas que possuíam dados multivalorados, como por exemplo, a coluna tipos\_de\_vd, que podia possuir até 5 valores dentro de uma mesma célula, desta forma, por exemplo: FIS/PSI/MOR/PAT/SEX, na coluna tipos\_de\_vd.

O processo de transformação dos dados representa a tarefa de modificar o formato de dados para um mais interessante ou apropriado para a mineração de dados. Para tornar as colunas multivaloradas em monovaloradas, foi adotada a estratégia de transposição de uma coluna para várias, convertendo os dados para 0 ou 1 (type int). As colunas transformadas foram tipos\_de\_vd e encam, que representam, respectivamente os tipos de violência doméstica praticados e os locais para onde as vítimas foram encaminhadas após a denúncia. A coluna tipos\_de\_vd deu origem a 9 novas colunas na base, enquanto a coluna encam, deu origem a 36 novas colunas.

1. **Resultados**

Como resultado do processo de aquisição e transformação dos dados, obteve-se a base de dados nomeada violencia\_mulher, com as tabelas cadastro\_ceam\_usuarias e cadastro\_ceam\_atual (tabelas 5 e 6), com 471 registros e 65 colunas, relacionadas através da coluna id. A base consolidada permite a última etapa da integração de dados, que é o carregamento de dados, que tem como objetivo gerar valor sobre os dados para a formação de relatórios e análises capazes de dar suporte à tomada de decisão.

**Tabela 5: Estrutura da tabela cadastro\_ceam\_usuarias da base de dados consolidada. Em cinza, coluna adicionada.**

|  |  |
| --- | --- |
| **Campos** | **Descrição** |
| id |  |
| ficha\_ano | Identificador da ficha de acompanhamento com número de ordem e o ano do atendimento |
| usuaria | Nome da pessoa atendida |
| contato | Número de telefone para contato |
| idade | Idade em anos |

**Tabela 6: Estrutura da tabela cadastro\_ceam\_atual da base de dados consolidada. Em cinza, colunas adicionadas**.

|  |  |
| --- | --- |
| **Campos** | **Descrição** |
| id |  |
| ficha\_ano | Identificador da ficha de acompanhamento com número de ordem e o ano do atendimento |
| tecnica\_de\_ref | Nome do profissional que realizou o atendimento à vítima |
| idade | Idade em anos |
| bairro | Bairro de residência |
| municipio | Município 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 |
| religiao\_SN | Informa se a vítima possui religião ou não |
| religiao | Religião da vítima atendida |
| grau\_ecolaridade | Grau de formação escolar |
| cor | Cor da pele |
| com\_renda | Indicação se a vítima possui ou não renda |
| filhos | Quantidade de filhos |
| reincidencia\*\*\* | Informa o número de entradas da usuária na base |
| rel\_agressor | Parentesco ou relação do agressor face à vítima |
| vd\_fis | Informa se a vítima sofreu violência física |
| vd\_psi | Informa se a vítima sofreu violência psicológica |
| vd\_mor | Informa se a vítima sofreu violência moral |
| vd\_pat | Informa se a vítima sofreu violência patrimonial |
| vd\_sex | Informa se a vítima sofreu violência sexual |
| vd\_n\_informou | Informa se o tipo de violência sofrida não foi informado |
| vd\_vazio | Informa se o campo violência sofrida está vazio |
| vd\_null | Informa se violência sofrida não foi um item pesquisado |
| vd\_nao | Informa se a vítima não sofreu nenhuma violência |
| encam\_n\_informou | Informa se o tipo de encaminhamento não foi informado |
| encam\_vazio | Informa se o campo encaminhamento não foi preenchido |
| encam\_null | Informa se encaminhamento não foi um item pesquisado |
| encam\_outros | Informa se houveram outros tipos de encaminhamento |
| encam\_psi | Informa se a vítima foi encaminhada para a psicologia |
| encam\_social | Informa se a vítima foi encaminhada para a assistência social |
| encam\_jurid | Informa se a vítima foi encaminhada para a assistência jurídica |
| encam\_grupo | Informa se a vítima foi encaminhada para o grupo de reflexão |
| encam\_un\_saude | Informa se a vítima foi encaminhada para uma unidade de saúde |
| encam\_def | Informa se a vítima foi encaminhada para a Defensoria Pública |
| encam\_deam | Informa se a vítima foi encaminhada para a Delegacia Especializada de Atendimento à Mulher |
| encam\_c\_tut | Informa se a vítima foi encaminhada pelo conselho tutelar |
| encam\_gineco | Informa se a vítima foi encaminhada para a ginecologia |
| encam\_sme | Informa se a vítima foi encaminhada para a Secretaria Municipal de Educação |
| encam\_jvd | Informa se a vítima foi encaminhada para o Juizado de Violência Doméstica contra a Mulher |
| encam\_b\_emprego | Informa se a vítima foi encaminhada para o balcão de emprego |
| encam\_sms | Informa se a vítima foi encaminhada para a Secretaria Municipal de Saúde |
| encam\_cras | Informa se a vítima foi encaminhada para o Centro de Referência de Assistência Social |
| encam\_craesm | Informa se a vítima foi encaminhada para o Centro de Referência e Atenção Especializada à  Saúde da Mulher |
| encam\_creas | Informa se a vítima foi encaminhada para o Centro de Referência Especializado de Assistência Social / curso de artesanato |
| encam\_crepop | Informa se a vítima foi encaminhada para o Centro de Referências Técnicas em Psicologia e Políticas Públicas |
| encam\_def\_m | Informa se a vítima foi encaminhada para a Defensoria da Mulher |
| encam\_smas | Informa se a vítima foi encaminhada para a Secretaria Municipal de Assistência Social |
| encam\_vf | Informa se a vítima foi encaminhada para a vara da família |
| encam\_edu | Informa se a vítima foi encaminhada para a |
| encam\_aa | Informa se a vítima foi encaminhada para o Alcoólicos Anônimos |
| encam\_ca\_ap\_fam | Informa se a vítima foi encaminhada para a Casa de Apoio à Família |
| encam\_cejuvida | Informa se a vítima foi encaminhada para a Central de Acolhimento da Mulher Vítima de Violência Doméstica |
| encam\_cr\_homem | Informa se a vítima foi encaminhada para o Centro de Referência do homem |
| encam\_abrigo | Informa se a vítima foi encaminhada para alocação em um abrigo |
| encam\_ativ\_casa | Informa se a vítima foi encaminhada para atividades da Casa da Mulher (artesanato, ginástica, etc) |
| encam\_inss | Informa se a vítima foi encaminhada para o INSS |
| encam\_forum | Informa se a vítima foi encaminhada para o Fórum |
| encam\_def\_crh | Informa se a vítima foi encaminhada para |
| encam\_dm | Informa se a vítima foi encaminhada para a Delegacia da Mulher |
| encam\_dp | Informa se a vítima foi encaminhada para a Delegacia de Polícia |
| origem\_nova | Informa por qual meio a vítima chegou ao CEAM |

**\*\*\***Optou-se por manter a coluna ‘reincidencia’ (em verde), que ainda não possui dados, como um indicativo de que a *feature* ‘reincidencia’ começará a ser avaliada em próximas denúncias de violência.

1. **Trabalhos Futuros**

Diante da inexistência de uma base de dados unificada e detalhada que favorecesse a tarefa de minneração de dados para a geração de informações úteis para o desenvolvimento de políticas públicas voltadas para a defesa da integridade da mulher, optou-se pela realização de um estudo do *dataset* do CEAM de Duque de Caxias, com o objetivo de criação de uma base de dados consolidada.

Para a consolidação da base de dados violecia\_mulher, foi necessário primeiro obter os *datasets* presencialmente no CEAM e na sequência realizar etapas referentes à anonimização das usuárias, limpeza e transformação dos dados. Através da adoção de técnicas de supressão e pseudoanonimato, foi possível anonimizar os dados separando-os em duas tabelas relacionadas em um banco de dados MySql. Na fase de limpeza dos dados as pesquisadoras empenharam-se na tarefa de normalização dos valores das colunas da tabela principal de trabalho ─ cadastro\_ceam\_atual, reduzindo de 12% a 81% a quantidade de valores distintos em cada uma das colunas. Na fase de transformação dos dados, realizou-se a transposição das colunas tipos\_de\_vd e encam, convertendo dados multivalorados para univalorados tipo inteiro, com o objetivo de facilitar procedimentos de análise levando em conta os tipos de violência ocorridos e os locais para onde as vítimas foram encaminhadas.

Esta pesquisa será continuada para refinar os dados obtidos através deste mapeamento preliminar e próximos passos levarão em conta a submissão desta pesquisa em um uma conferência voltada para a área de visualização de dados ou de interfaces.

**Referências**

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. 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., WANG-CHIEW, T. (2001). Why and where: A characterization of data provenance. International Conference on Database Theory, 1:316–330.

BURELOMOVA, A. S.; GULINA, M. A.; TIKHOMANDRITSKAYA, O. 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.

BORBUREMA T. L. R.; PACHECO, A. P. ; NUNES, A.A. ; MORÉ C. L. O. O. ; KRENKEL, S. Violência contra mulher em contexto de vulnerabilidade social na Atenção Primária: registro de violência em prontuários. Rev Bras Med Fam Comunidade. 2017;12(39):1-13. http://dx.doi.org/10.5712/rbmfc12(39)1460

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

NARAYANAN, A.; SHMATIKOV, V. Myths and fallacies of “personally identifiable information”. Communications of the ACM, Nova York, v. 53, n. 6, p. 24-26, 2010. Disponível em: [http://bit.ly/30G9CVq](https://bit.ly/30G9CVq). Acesso em: março 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.

1. <https://bit.ly/3vst86w>. Acesso em 27/05/2021. [↑](#footnote-ref-1)