



UNIVERSIDADE FEDERAL DO CEARÁ

CAMPUS SOBRAL

DEPARTAMENTO DE ENGENHARIA DE COMPUTAÇÃO

CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

THAÍS FÉLIX DOS SANTOS

**ANÁLISE EXPLORATÓRIA DOS DADOS DA CENTRAL DE ATENDIMENTO À
MULHER**

SOBRAL

2024

THAÍS FÉLIX DOS SANTOS

ANÁLISE EXPLORATÓRIA DOS DADOS DA CENTRAL DE ATENDIMENTO À MULHER

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação do *Campus* Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Iális Cavalcante de Paula Júnior.

SOBRAL

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- F36a Félix dos Santos, Thaís.
Análise Exploratória dos Dados da Central de Atendimento à Mulher / Thaís Félix dos Santos. – 2024.
53 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Sobral,
Curso de Engenharia da Computação, Sobral, 2024.
Orientação: Prof. Dr. Iális Cavalcante de Paula Júnior.
1. violência contra a mulher. 2. análise exploratória de dados. 3. Ligue 180. 4. ETL. 5. Python. I. Título.
CDD 621.39
-

THAÍS FÉLIX DOS SANTOS

ANÁLISE EXPLORATÓRIA DOS DADOS DA CENTRAL DE ATENDIMENTO À MULHER

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação do *Campus* Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Computação.

Aprovada em: 16 de Julho de 2024

BANCA EXAMINADORA

Prof. Dr. Iális Cavalcante de Paula
Júnior (Orientador)
Universidade Federal do Ceará (UFC)

Profª. Dra. Jermana Lopes de Moraes
Universidade Federal do Ceará (UFC)

Prof. Dr. Thiago Iachiley Araújo de Souza
Universidade Federal do Ceará (UFC)

AGRADECIMENTOS

Primeiramente a Deus, por ser a fonte de sabedoria e força que me sustentou ao longo desta jornada acadêmica.

À minha amada mãe, Suely, que sempre foi meu exemplo de perseverança e dedicação. Sua infinita paciência e amor incondicional me inspiraram a buscar meus sonhos e superar os desafios. Obrigada por tudo, minha querida mãe.

Ao meu amado esposo, Hugo. Você tem sido meu porto seguro, meu confidente e minha maior fonte de motivação. Seu apoio incansável e paciência foram essenciais para a conclusão deste projeto. Agradeço por estar ao meu lado em todas as etapas desta jornada.

À minha querida irmã, Jéssica, companheira de todas as horas e minha maior incentivadora. Agradeço por todo o apoio e por ser a melhor irmã que alguém poderia ter.

Às minhas amigas verdadeiras, Lucinara e Beatriz, que me acompanharam nesta jornada e tornaram os momentos difíceis mais suportáveis. Agradeço por todo o amor, apoio e compreensão.

À minha querida psicóloga, Lívia, que foi fundamental em minha jornada de crescimento pessoal. Com a sua orientação e apoio, pude superar obstáculos e descobrir meu potencial.

Ao professor Iális pela orientação e paciência que foram fundamentais para o meu crescimento acadêmico e profissional.

A todos que mencionei, vocês são a razão pela qual cheguei até aqui. Sou grata por cada um de vocês fazer parte da minha vida e por terem contribuído de maneira significativa para o meu crescimento pessoal e acadêmico.

"Se eu vi mais longe, foi por estar sobre ombros de gigantes." (ISAAC NEWTON, 1676.)

RESUMO

Nos últimos anos, tem acontecido um preocupante aumento da violência contra as mulheres. Diante dessa problemática, este trabalho tem como objetivo coletar, tratar e analisar os dados provenientes da Central de Atendimento à Mulher (Ligue 180) com o propósito de criar um *dashboard* contendo as principais informações das vítimas, suspeitos e denúncias. Para isso, foram utilizadas técnicas de ETL (Extração, Transformação e Carregamento de Dados). Foi usada a linguagem Python para a extração dos dados, estes dados foram armazenados no *BigQuery* e processados com SQL, por fim, o *Power BI* foi utilizado para construção da visualização dos dados. A partir das percepções e informações obtidas, nota-se que há um problema recorrente no registro das informações das denúncias, com muitos dados preenchidos como 'Não Informado'. Isso impede a elaboração de perfis completos sobre os grupos mais afetados. Para mitigar essa questão, seria importante implementar medidas para garantir a coleta mais completa e precisa de dados, como treinamentos para quem registra as denúncias e a revisão dos formulários para incluir campos obrigatórios, para direcionar medidas específicas para atender as populações mais vulneráveis. Apesar deste problema, foi possível gerar informações relevantes. A pesquisa revelou que dos dados informados 78% das denúncias ocorreram na própria casa da vítima. Além disso, observou-se que 91% das vítimas utilizam canais de atendimento telefônico, evidenciando a importância de manter e expandir o serviço "Ligue 180". Os *insights* fornecidos pelos *dashboards* podem ser utilizados por políticas públicas que atuam na defesa das mulheres para direcionar recursos e esforços de maneira mais eficaz. Por exemplo, os dados sobre a alta frequência de agressões em ambientes domésticos podem ajudar a formular campanhas de conscientização focadas em ambientes familiares.

Palavras-chave: violência contra a mulher; análise exploratória de dados; Ligue 180; ETL; Python; *BigQuery* *Power BI*.

ABSTRACT

In recent years, there has been a worrying increase in violence against women. Faced with this issue, this work aims to collect, process, and analyze data from the Central de Atendimento à Mulher (Ligue 180) in order to create a dashboard containing key information about the victims, suspects, and reports. To achieve this, ETL techniques (Extraction, Transformation, and Loading of Data) were used. Python language was employed for data extraction, with the data being stored in BigQuery and processed using SQL. Finally, Power BI was used to construct data visualizations. Based on the insights and information obtained, it is noted that there is a recurring issue in the recording of complaint information, with many entries filled as 'Not Informed.' This hinders the creation of complete profiles about the most affected groups. To mitigate this issue, it would be important to implement measures to ensure more complete and accurate data collection, such as training for those who record complaints and reviewing forms to include mandatory fields, so that specific measures can be directed to assist the most vulnerable populations. Despite this problem, it was possible to generate relevant information. The research revealed that of the reported data, 78% of complaints occurred in the victim's own home. Additionally, it was observed that 91% of victims use telephone support channels, highlighting the importance of maintaining and expanding the "Ligue 180" service. The insights provided by the dashboards can be used by public policies that work in the defense of women to direct resources and efforts more effectively. For example, data on the high frequency of aggressions in domestic environments can help formulate awareness campaigns focused on family settings.

Keywords: violence against women; exploratory data analysis; Ligue 180; ETL; Python; BigQuery; Power BI.

LISTA DE FIGURAS

Figura 1 – Fluxograma Extração, Transformação e Carregamento (ETL)	21
Figura 2 – Fluxograma Extração	22
Figura 3 – Fluxograma Transformação	25
Figura 4 – Modelo Dimensional	27
Figura 5 – Página do Perfil das Denúncias	28
Figura 6 – Página do Perfil das Vítimas	29
Figura 7 – Página do Perfil dos Suspeitos	30
Figura 8 – Filtros do Perfil das Vítimas	30
Figura 9 – Filtros do Perfil dos Suspeitos	31
Figura 10 – Filtros do Perfil das Denúncias	31
Figura 11 – Indicadores do Perfil das Denúncias	32
Figura 12 – <i>Top 50</i> - Violações mais citadas	33
Figura 13 – <i>Top 5</i> UF Denúncias	33
Figura 14 – <i>Top 5</i> Canal de atendimento a vítima	33
Figura 15 – <i>Top 5</i> Cenário de violação	34
Figura 16 – Frequência das agressões	34
Figura 17 – <i>Top 3</i> - Tipos de denunciantes	34
Figura 18 – Tabela de grupo vulnerável	35
Figura 19 – Data da denúncia	35
Figura 20 – <i>Top 5</i> - Faixa Etárias das Vítimas	36
Figura 21 – <i>Top 5</i> - Raça/Cor das Vítimas	36
Figura 22 – <i>Top 5</i> - País das Vítimas	36
Figura 23 – Tabela Grau de Instrução das Vítimas	37
Figura 24 – <i>Top 5</i> - UF das Vítimas	37
Figura 25 – <i>Top 5</i> - Deficiência das Vítimas	38
Figura 26 – Faixa de Renda das Vítimas	38
Figura 27 – <i>Top 3</i> - Sexo das Vítimas	38
Figura 28 – <i>Top 5</i> - Orientação Sexual da Vítima	39
Figura 29 – <i>Top 5</i> - Doença Rara da Vítima	39
Figura 30 – Relação Vítima X Suspeito	39
Figura 31 – Relação Vítima X Suspeito - Faixa Etária: 0 à 9 anos	40

Figura 32 – Relação Vítima X Suspeito - Faixa Etária: 10 à 19 anos	40
Figura 33 – Relação Vítima X Suspeito - Faixa Etária: 30 à 39 anos	41
Figura 34 – Profissão da Vítima	41
Figura 35 – <i>Top 5</i> - Faixa Etárias do Suspeito	41
Figura 36 – <i>Top 5</i> - Raça/Cor do Suspeito	42
Figura 37 – Tabela Grau de Instrução do Suspeito	42
Figura 38 – <i>Top 5</i> - País do Suspeito	42
Figura 39 – <i>Top 5</i> - UF do Suspeito	43
Figura 40 – <i>Top 5</i> - Suspeito Preso	43
Figura 41 – Faixa de Renda do Suspeito	43
Figura 42 – <i>Top 3</i> - Sexo do Suspeito	44
Figura 43 – <i>Top 5</i> - Orientação Sexual do Suspeito	44
Figura 44 – <i>Top 5</i> - Doença Rara do Suspeito	45
Figura 45 – Profissão do Suspeito	45

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
BI	<i>Business Intelligence</i>
CSV	<i>Comma-Separated Values</i>
ETL	Extração, Transformação e Carregamento
ID	Identificação
LGBTQIA+	Lésbicas, Gays, Bissexuais, Transexual, <i>Queer</i> , Intersexo, Assexuais e mais
MDHC	Ministério dos Direitos Humanos e da Cidadania
PCD	Pessoa com Deficiência
SQL	<i>Structured Query Language</i>
URL	<i>Uniform Resource Locator</i>

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Justificativa	13
1.2	Objetivos	13
1.2.1	<i>Objetivos Específicos</i>	13
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Violência contra a mulher	14
2.1.1	<i>Fontes dos Dados</i>	14
2.2	Análise de Dados	15
2.2.1	<i>Tipos de Análise de Dados</i>	15
2.2.1.1	<i>Análise Descritiva</i>	15
2.2.1.2	<i>Análise Exploratória</i>	16
2.2.1.3	<i>Análise Preditiva</i>	16
2.2.1.4	<i>Análise Prescritiva</i>	16
2.2.2	<i>Processo de Análise de Dados</i>	16
2.3	Bancos de Dados	17
2.3.1	<i>Modelagem de Dados</i>	18
2.3.1.1	<i>Modelagem Relacional</i>	18
2.3.1.2	<i>Modelagem Dimensional</i>	18
2.3.2	<i>SQL</i>	19
2.4	Visualização de Dados	19
2.5	ETL (Extração, Transformação e Carregamento)	20
3	METODOLOGIA	21
3.1	Dados	21
3.2	ETL - Extração com Python	22
3.2.1	<i>Estrutura de pastas e arquivos do projeto</i>	23
3.2.2	<i>Etapa - Download dos Dados</i>	23
3.2.3	<i>Etapa - Preparação dos Dados</i>	23
3.2.4	<i>Etapa - Upload dos Dados</i>	24
3.3	ETL - Transformação com BigQuery	24
3.3.1	<i>União das tabelas</i>	24

3.3.2	<i>Exclusão de colunas</i>	25
3.3.3	<i>União de colunas redundantes</i>	26
3.3.4	<i>Padronização dos Dados</i>	26
3.3.5	<i>Colunas de Identificação (ID)</i>	26
3.3.6	<i>Modelo Dimensional</i>	26
3.4	ETL - Carregamento para o Power BI	27
4	RESULTADOS E DISCUSSÕES	32
4.1	Perfil das Denúncias	32
4.2	Perfil das Vítimas	35
4.3	Perfil dos Suspeitos	40
4.4	Problemas Encontrados	45
5	CONCLUSÃO	47
5.1	Trabalhos Futuros	47
	REFERÊNCIAS	49
	APÊNDICE A – LINKS ÚTEIS	51

1 INTRODUÇÃO

Ao longo dos anos, as mulheres sempre foram restritas a seguir suas obrigações, como serem devotas e se submeterem aos homens. Caso elas desafiassem essa posição, os mecanismos sociais buscavam corrigir seu comportamento. Ao longo da história, a sociedade continuou a perpetuar a ideia de que as mulheres eram inferiores aos homens, utilizando tradições e costumes como justificativas. Isso resultou na banalização e naturalização da opressão que elas sofreram por décadas, refletindo atualmente em diversos setores sociais nos quais as mulheres estão presentes (ESSY, 2017).

Com a ruptura dos papéis preestabelecidos para cada gênero, surge a violência como meio de reprimir a mulher que anteriormente limitava-se ao lar, desempenhando seu papel de mãe e esposa. Com isso, nota-se que a violência doméstica é resultante de um processo histórico que legitima a diminuição social da mulher, além disso, é observado uma incapacidade ou recusa masculina de adaptar-se à nova esfera social na qual as mulheres detêm o poder sobre si mesmas (ESSY, 2017).

O início do ano de 2020 foi marcado pelo surgimento da COVID-19, consequentemente, por sua rápida disseminação foi necessário conter o avanço desta doença, para isso diversos países, incluindo o Brasil, tiveram que optar pelo isolamento social, mantendo as famílias em suas residências. Diante disso, muitas mulheres foram submetidas a ficarem com seus agressores por um longo tempo em isolamento, resultando assim em um aumento dos índices de violência contra a mulher (OKABAYASHI *et al.*, 2020).

Diante desse contexto, o presente trabalho tem como objetivo realizar um levantamento e uma análise de bancos de dados abertos da Central de Atendimento à Mulher (Ligue 180), o qual contém informações sobre violência contra a mulher e violência doméstica contra a mulher, com a pretensão de estabelecer e descrever as características das vítimas, suspeitos e denúncias, a fim de identificar o perfil social que tem sido mais afetado nos últimos anos.

O trabalho é organizado nas seguintes seções: a seção 2 aborda a fundamentação teórica, fornecendo os principais conceitos e teorias relacionados ao tema. A seção 3 descreve a metodologia utilizada na pesquisa, detalhando os procedimentos, fontes de dados e técnicas empregadas. Os resultados obtidos são apresentados e discutidos na seção 4. Por fim, a seção 5 apresenta a conclusão do trabalho, sintetizando as informações e apontando possíveis trabalhos futuros.

1.1 Justificativa

A violência contra as mulheres é uma realidade cotidiana no Brasil que requer ações efetivas para garantir a segurança e o bem-estar das mulheres em todas as esferas da sociedade. Essas violências envolvem desde o assédio moral e sexual até o feminicídio. A gravidade do problema é reconhecida por conquistas legais, como a Lei do Feminicídio (Lei nº 13.104/2015), a qual se refere ao assassinato de mulheres por razões relacionadas ao seu gênero, que busca reconhecer e combater a especificidade dessa violência. Para reduzir esses índices, é fundamental produzir informações, pesquisas e análises que mensuram as características da violência de gênero (Fórum Brasileiro de Segurança Pública, 2023).

Dado o aumento alarmante da violência contra as mulheres, é crucial coletar, tratar e analisar os dados disponíveis para compreendermos melhor o perfil das vítimas, suspeitos e denúncias, pois além de dimensionar o problema, as estatísticas auxiliam a conhecer melhor o cenário da violência contra a mulher, incluindo a violência doméstica. Com isso, esse trabalho propõe a coleta, o tratamento e a análise de bancos de dados públicos da Central de Atendimento à Mulher (Ligue 180) com o intuito de avaliar e descrever as características mais relevantes para definição do perfil de vítimas, suspeitos e denúncias. Logo, com os perfis traçados, as políticas públicas podem focar em medidas específicas que possam atender aos públicos mais afetados.

1.2 Objetivos

Este trabalho tem como objetivo geral a coleta, o tratamento e a análise dos dados abertos da Central de Atendimento à Mulher (Ligue 180) para definição dos perfis das vítimas, suspeitos e denúncias de violência contra a mulher no período de 2019 a 2023.

1.2.1 *Objetivos Específicos*

- Elaborar a metodologia de coleta, tratamento e análise de bancos de dados abertos da Central de Atendimento à Mulher (Ligue 180), seguindo o ETL adaptado para a problemática;
- Elaboração de um *dashboard* que reúne os dados dos anos de 2019 a 2023, sendo possível avaliar e descrever as características mais relevantes para definição dos perfis das vítimas, suspeitos e denúncias.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados os conceitos essenciais para a compreensão do tema, incluindo a violência contra a mulher e as fontes de dados associadas, os fundamentos da análise de dados, bancos de dados e visualização. Além disso, será abordado o conceito de ETL (Extração, Transformação e Carregamento).

2.1 Violência contra a mulher

Violência contra a mulher é uma realidade alarmante que se manifesta por meio de atos ou condutas baseados no gênero, resultando em morte, danos físicos, sexuais ou psicológicos à mulher, tanto em ambientes públicos quanto privados (SILVA; OLIVEIRA, 2015). No intuito de estabelecer punições adequadas e coibir a violência doméstica contra as mulheres, foi promulgada a Lei nº 11.340 em 7 de agosto de 2006, popularmente conhecida como Lei Maria da Penha, em homenagem a uma mulher que sobreviveu a duas tentativas de homicídio cometidas por seu marido e, desde então, se dedica à luta contra a violência feminina (FONSECA *et al.*, 2012).

Segundo a Lei Maria da Penha, a violência doméstica e familiar contra a mulher é uma forma de violação dos direitos humanos. É definida como qualquer ação ou omissão baseada no gênero que resulte em morte, lesões, sofrimentos físicos, sexuais, psicológicos, morais ou patrimoniais (BRASIL, 2006). Dado o contexto alarmante de violência contra as mulheres, é de extrema importância que sejam priorizadas iniciativas que busquem produzir informações, realizar pesquisas e análises para mensurar as diferentes características da violência de gênero (Fórum Brasileiro de Segurança Pública, 2023)

Além da violência contra a mulher, é importante ressaltar que a violência doméstica também afeta outros grupos vulneráveis, como idosos, crianças, Pessoa com Deficiência (PCD) e membros da comunidade de Lésbicas, Gays, Bissexuais, Transexual, *Queer*, Intersexo, Assexuais e mais (LGBTQIA+). Esses indivíduos enfrentam desafios específicos em relação à violência doméstica, demandando uma atenção especial para garantir a proteção de seus direitos e bem-estar (MARCOLINO *et al.*, 2021).

2.1.1 Fontes dos Dados

O site do Governo Federal (GOV.BR, 2024) reúne, em um só lugar, serviços para o cidadão e informações sobre a atuação de todas as áreas do governo. Uma dessas informações é

do Ministério dos Direitos Humanos e da Cidadania (MDHC, 2024), este é um dos ministérios que compõem o gabinete executivo do Governo Federal do Brasil.

O Ministério dos Direitos Humanos e da Cidadania (MDHC) formula políticas e diretrizes para promover os direitos da cidadania. Além disso, serve como ouvidoria nacional das mulheres, da igualdade racial e dos direitos humanos, promovendo ações contra a discriminação e pela igualdade entre mulheres e homens (GOV.BR, 2024). No site é possível encontrar os dados abertos da Central de Atendimento à Mulher e dados dos canais de atendimento da Ouvidoria Nacional de Direitos Humanos (ONDH): Disque 100, Ligue 180 e do aplicativo Direitos Humanos Brasil.

2.2 Análise de Dados

A análise de dados é o processo de examinar, limpar, transformar e modelar dados com o objetivo de descobrir informações úteis, obter *insights* e tomada de decisões. Essas técnicas são amplamente utilizadas em diversas áreas, como ciência, negócios, saúde e tecnologia (MILTON, 2010)

2.2.1 Tipos de Análise de Dados

Existem diversas formas de aplicação da análise de dados, que dependem dos objetivos da análise. Os principais tipos são citados abaixo.

2.2.1.1 Análise Descritiva

A análise descritiva foca na descrição, resumo e apresentação dos dados de forma clara e objetiva. A visão geral das características dos dados é fornecida por este tipo de análise, como métricas de estatística descritiva. Essa análise é fundamental para compreensão da distribuição e das características principais dos dados (FÁVERO; BELFIORE, 2017).

As técnicas mais comuns incluem medidas de tendência central (média, mediana, moda) e de dispersão (variância, desvio padrão). Também podem ser usadas visualizações como histogramas, que ajudam a identificar padrões nos dados, tornando-os mais compreensíveis.

2.2.1.2 *Análise Exploratória*

A análise exploratória é formada pela exploração detalhada dos dados, buscando identificar padrões e tendências. Ela utiliza técnicas visuais, como gráficos, diagramas e tabelas, além de incluir técnicas de estatística descritiva. Com essa análise é possível gerar hipóteses e *insights* sobre os dados (FÁVERO; BELFIORE, 2017).

Entre as técnicas mais utilizadas na análise exploratória estão a construção de gráficos para visualizar a distribuição dos dados, *box plots* para identificar possíveis *outliers*, e *scatter plots* para analisar relações entre variáveis. Além disso, mapas de calor são frequentemente empregados para explorar correlações e interações complexas. Essas ferramentas gráficas permitem uma compreensão visual imediata das características e padrões dos dados, facilitando a geração de *insights* iniciais.

2.2.1.3 *Análise Preditiva*

A análise preditiva visa fazer previsões e estimativas utilizando padrões históricos dos dados existentes. Esta análise usa técnicas estatísticas avançadas para criação de modelos preditivos que podem antecipar demandas futuras (MCKINNEY, 2018).

Técnicas comuns incluem regressões (linear, logística) e algoritmos de aprendizado de máquina, como árvores de decisão e redes neurais, que modelam padrões nos dados para prever resultados futuros.

2.2.1.4 *Análise Prescritiva*

A análise prescritiva identifica a melhor ação ou decisão a ser tomada com base nos dados disponíveis. Ela mescla técnicas de análise preditiva, análise exploratória, e outras abordagens para recomendar melhores ações e estratégias para tomada de decisões. Essa análise é crucial para resolução de problemas complexos, como situações que envolvem múltiplas variáveis e restrições (MCKINNEY, 2018).

2.2.2 *Processo de Análise de Dados*

O processo de análise de dados pode seguir as etapas (SILVESTRE, 2007):

1. Definir o problema: identificar o objetivo da análise e quais questões devem ser respondidas;

2. Coleta de dados: extrair os dados relevantes das fontes disponíveis, como bancos de dados, arquivos, pesquisas etc;
3. Limpeza e transformação dos dados: realizar o pré-processamento dos dados, corrigindo valores errados, removendo valores ausentes ou não relevantes, e padronizando os dados, como a conversão de valores para formatos adequados para análise;
4. Exploração dos dados: buscar nos dados padrões, tendências, e relações entre os dados;
5. Modelagem e análise dos dados: usar técnicas estatísticas e de modelagem de dados para analisar as informações contida nos dados;
6. Interpretação dos resultados: interpretar os resultados da análise buscando responder às hipóteses da pesquisa;
7. Comunicação dos resultados: apresentar os resultados de forma clara, utilizando gráficos, tabelas e métricas, construindo *dashboards* e relatórios.

A análise de dados é uma abordagem essencial na tomada de decisões em diversos domínios. Ao utilizar técnicas e conceitos básicos da análise de dados, é possível transformar dados brutos em informação, além de conseguir extrair *insights* significativos.

Neste trabalho, o foco é a análise exploratória dos dados, seguindo os passos do processo de análise de dados adaptados para a temática em questão. Essa abordagem foi escolhida por sua capacidade de proporcionar uma compreensão inicial e aprofundada dos dados, permitindo identificar padrões e tendências. A análise exploratória é indicada quando se trabalha com conjuntos de dados complexos ou desconhecidos, pois oferece flexibilidade e facilita a geração de hipóteses, fornecendo uma base sólida para investigações posteriores. Com essa abordagem é possível explorar diversas perspectivas dos dados, promovendo uma visão mais informada sobre o tema estudado.

2.3 Bancos de Dados

Bancos de Dados são sistemas que organizam e armazenam informações de forma estruturada, possibilitando a recuperação e manipulação dos dados (DATE, 2004). Na modelagem de dados, é necessário compreender certos princípios:

- Entidades: são objetos ou conceitos do mundo real, como clientes, usuários, pedidos, transações, produtos etc.
- Atributos: informações das entidades. Exemplo: usuários podem ter atributos como nome, tipo de usuário, data de nascimento e telefone.

- Relacionamentos: conexões entre as entidades. Exemplo: uma transação pode estar relacionada a um usuário.
- Chaves primárias e estrangeiras: são identificadores únicos para cada amostra em uma tabela e são usados no relacionamento entre tabelas.

2.3.1 Modelagem de Dados

As duas principais modelagens de dados são a relacional e a dimensional, elas são explicados a seguir.

2.3.1.1 Modelagem Relacional

A modelagem relacional é amplamente utilizada na indústria para organizar dados em tabelas compostas por linhas e colunas. Essas tabelas representam entidades relevantes do sistema, onde cada tabela possui um nome e suas colunas representam atributos.

Os relacionamentos entre as entidades são construídos por meio de chaves primárias e estrangeiras, conectando as tabelas. Restrições são usadas para garantir a integridade dos dados, como restrições de chave primária garantindo que os registros são únicos e restrições de chave estrangeira para manter as relações entre tabelas (DATE, 2004).

2.3.1.2 Modelagem Dimensional

A modelagem dimensional é amplamente utilizada em ambientes de *data warehouse* e sistemas de *Business Intelligence* (BI). Ela é diferente da modelagem relacional, na estrutura e em como os dados são organizados.

Nesta modelagem, os dados são estruturados em torno de tabelas de fatos e dimensões. Os fatos vão representar os eventos que ocorrem na problemática, como vendas ou lucros. As dimensões são os fatores relacionados aos fatos, como data, tipo de venda, cliente, produto, etc (KIMBALL; ROSS, 2011).

A tabela de fatos é constituída de chaves estrangeiras que se relacionam com as tabelas de dimensões, além de conter outras métricas que representam os fatos. As tabelas de dimensões são formadas pelos atributos que descrevem essas entidades.

A principal vantagem da modelagem dimensional é a facilidade da realização da análise dos dados. O formato de hierarquia das dimensões permite a manipulação e a união

de dados de forma eficiente, facilita também o uso de consultas complexas para responder às hipóteses da problemática (KIMBALL; ROSS, 2011).

Cada tipo de modelagem, relacional e dimensional, possui suas aplicações específicas, e devem ser escolhidos com base nas necessidades do problema. Para este trabalho, os dados foram estruturados com a modelagem dimensional. Foi escolhida essa abordagem devido à sua eficácia em trabalhar com grandes volumes de dados, facilitando a análise e o entendimento das informações. A modelagem dimensional permite a criação de um esquema intuitivo que melhora o desempenho das consultas e facilita a geração de relatórios, através de dimensões e fatos, tornando-a ideal para o tipo de análise exigido neste estudo.

2.3.2 SQL

Structured Query Language (SQL) é uma linguagem usada para interagir com bancos de dados, permitindo consultas e atualizações. Os comandos básicos incluem *SELECT* (recuperar dados), *INSERT* (adicionar dados), *UPDATE* (atualizar dados) e *DELETE* (remover dados). Além disso, existem operadores, funções de agregação, junções, subconsultas e técnicas de otimização para melhorar a busca de informações no banco de dados (BEAULIEU, 2009).

2.4 Visualização de Dados

A Visualização de Dados engloba técnicas que visam facilitar a compreensão e acessibilidade de informações, por meio da representação gráfica dos dados. Ela possibilita analisar os dados de forma mais intuitiva, e que possam ser usados na tomada de decisões (KNAFLIC, 2015).

As principais formas de visualização de dados são os *Dashboards* e os Relatórios. O *dashboard* é uma um formato utilizado para monitorar e apresentar os principais indicadores de uma problemática. Ele permite visualizar de forma clara e direta os dados mais importantes, em um só lugar, desta forma facilita a identificação de padrões e mudanças significativas (YAU, 2011). O relatório é um documento que apresenta informações de maneira estruturada, além de análises dos dados de uma problemática. Ele é uma abordagem mais detalhada de apresentar dados e é útil para análises mais profundas.

Neste trabalho, o formato do *dashboard* foi escolhido, pois além de ser uma ferramenta interativa e dinâmica, permite que o usuário altere parâmetros e filtros, obtendo *insights*

diferentes e específicos. Os *dashboards* também têm a capacidade de se adaptar facilmente para inclusão de mais dados.

2.5 ETL (Extração, Transformação e Carregamento)

O processo de ETL é essencial para a análise de dados. Ele envolve a extração de dados de diversas fontes, a transformação desses dados para um formato adequado e o carregamento dos dados transformados em um local de destino (CASERTA; KIMBALL, 2013).

A etapa de Extração (E) envolve a obtenção dos dados em sua forma original de diversas fontes, como bancos de dados, planilhas, arquivos e *Application Programming Interface* (API). Durante esse processo, é importante assegurar a integridade dos dados, selecionando fontes confiáveis e realizando a coleta correta das informações necessárias.

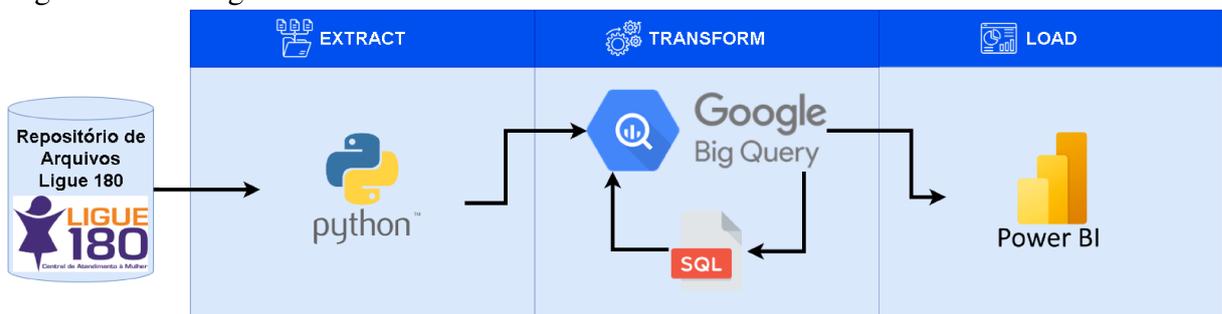
A etapa de Transformação (T) é responsável por converter os dados brutos em um formato uniforme e apropriado para análise. Nessa etapa, são executadas atividades como limpeza dos dados, padronização, enriquecimento com informações adicionais e agregação. Também é possível aplicar regras de negócio para gerar métricas e indicadores relevantes, que facilitam a compreensão e a interpretação dos dados.

Finalmente, a etapa de Carregamento (L) envolve a inserção dos dados transformados em um local de destino, como um *data warehouse*, *dashboard*, banco de dados. Durante o carregamento, é crucial levar em consideração a estrutura de destino, garantindo a eficiência e a integridade dos dados.

3 METODOLOGIA

O capítulo de Metodologia explora os passos seguidos para a coleta e análise dos dados. A Figura 1 mostra o fluxograma principal do trabalho: os dados do repositório de arquivos do Ligue 180 passam por um processo de ETL. Na etapa de Extração (*Extract*), os dados são coletados com *Python* e inseridos no banco *BigQuery*. Na etapa de Transformação (*Transform*), os dados são processados com *SQL*. Na etapa de Carregamento (*Load*), os dados são enviados para o *Power BI*, onde foi criado um *dashboard* para explorar os dados.

Figura 1 – Fluxograma ETL



Fonte: Autoria própria.

3.1 Dados

Os dados utilizados são provenientes da Central de Atendimento à Mulher (LIGUE180, 2024), um serviço que registra e encaminha denúncias de violência contra a mulher aos órgãos competentes. Na seção Dados Abertos do Ligue 180, estão disponíveis os relatórios com os dados de 2014 a 1º Trimestre de 2023, em arquivos *Comma-Separated Values (CSV)*, sobre violência contra a mulher e violência doméstica e familiar contra a mulher.

No entanto, para este trabalho, foram utilizados apenas os dados mais recentes, compreendendo o período de 2019 até o primeiro trimestre de 2023. Essa escolha se deu devido às alterações significativas na estrutura de coleta dos dados mais recentes em relação aos mais antigos, tornando os dados atuais mais padronizados e contendo informações relevantes para traçar o perfil das vítimas, suspeitos e denúncias.

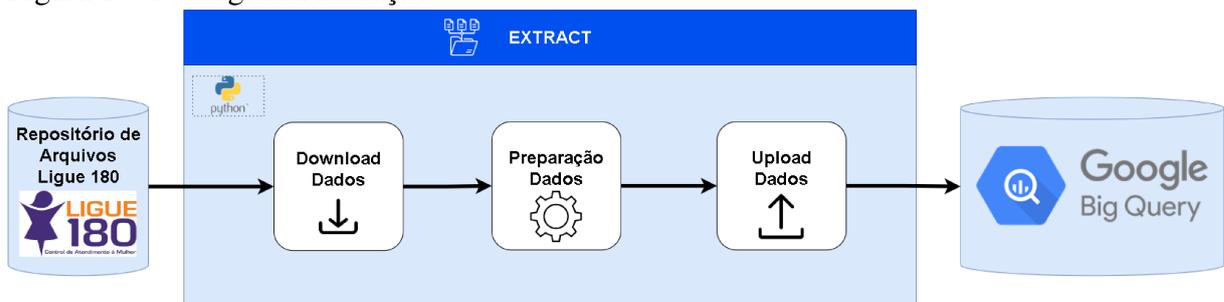
Após uma análise dos dados, constatou-se que as informações não se limitam apenas a vítimas do sexo feminino, abrangendo também outros grupos vulneráveis, como idosos, crianças, pessoas com deficiência, a população LGBTQIA+ e indivíduos em situação de rua.

3.2 ETL - Extração com *Python*

Durante a etapa de Extração de Dados, foi desenvolvido um projeto com a linguagem *Python*, com o objetivo de coletar informações da Central de Atendimento à Mulher (Ligue 180) e inseri-las automaticamente em um banco de dados.

O fluxograma apresentado na Figura 2 representa as etapas do processo de Extração. Inicialmente, a etapa de Download dos Dados é responsável por coletar os arquivos do repositório do Ligue 180. Em seguida, a etapa de Preparação dos Dados processa os arquivos coletados, realizando as devidas preparações necessárias. Por fim, a etapa de *Upload* dos Dados insere os arquivos preparados em um banco de dados *BigQuery*. Estas etapas serão mais detalhadas nas subseções a seguir.

Figura 2 – Fluxograma Extração



Fonte: Autoria própria.

O projeto foi desenvolvido com o objetivo de automatizar e estruturar o processo de extração de dados do Ligue 180. Os códigos do projeto estão disponíveis em um repositório do *GitHub*, que pode ser acessado por meio do link fornecido no Apêndice A.

A seguir são apresentadas as principais bibliotecas utilizadas nesse projeto. O *chardet* é empregado para verificar o formato da estrutura interna dos arquivos, garantindo que a leitura de dados seja correta. Para acessar o *BigQuery*, é utilizado a biblioteca *google-cloud-bigquery*, que facilita a integração com o banco de dados para consultas e análises. O *requests* é responsável pelo download dos arquivos necessários, enquanto o *pandas* é utilizado para a manipulação dos dados.

Nesse repositório, além dos códigos, é possível acessar a documentação e outros materiais relevantes para entender e utilizar o projeto. A disponibilidade do projeto no *GitHub* também permite a contribuição e acompanhamento de atualizações por parte da comunidade interessada.

3.2.1 Estrutura de pastas e arquivos do projeto

A estrutura é explicada abaixo:

- Pasta "dados" contém:
 - Arquivo "fontes.json": contém os nomes e links de cada arquivo que serão baixados.
 - Pasta "1_raw": usada para armazenar os arquivos baixados.
 - Pasta "2_preparados": usada para armazenar os arquivos preparados para o *upload*.
- Pasta "local" contém:
 - Arquivo "key.json": possui as credenciais do banco *BigQuery*.
- Pasta "src" contém:
 - Arquivo "download_dos_dados.py": códigos relacionados ao download.
 - Arquivo "preparacao_dos_dados.py": códigos relacionados à preparação.
 - Arquivo "upload_dos_dados.py": códigos relacionados ao upload.
 - Arquivo "main.py": código principal, que faz uso dos códigos presentes nos outros arquivos.
 - Arquivo "requirements.txt": dependências do projeto.
- Arquivo "README.md": fornece uma visão geral do projeto, bem como instruções de instalação e uso.

3.2.2 Etapa - Download dos Dados

A etapa de Download dos Dados é responsável por percorrer o arquivo "fontes.json" para obter a *Uniform Resource Locator* (URL) das fontes, esse processo é feito de forma automatizada e estrutura no código *Python*. Em seguida, é realizado o download dos arquivos do tipo CSV correspondentes a cada URL. Esses arquivos são salvos na pasta "1_raw" com o seguinte formato de nome: "ligue180_" + data da coleta dos dados + ".csv". Por exemplo, um arquivo pode ser chamado de "ligue180_2019.csv". A data da coleta dos dados pode ser representada apenas pelo ano, pelo semestre mais o ano ou pelo trimestre mais o ano, dependendo do formato fornecido pelo repositório do Ligue 180.

3.2.3 Etapa - Preparação dos Dados

Na etapa de Preparação dos Dados, é necessário padronizar os arquivos CSV baixados, pois há diferentes formatos de estruturação e problemas com os nomes das colunas. Durante

essa etapa, os arquivos CSV são ajustados para seguir um formato de estruturação consistente, e os nomes das colunas são padronizados para garantir a conformidade dos dados com as restrições de nomenclatura do *BigQuery*. Não são permitidas colunas que começam com números, ou que contém letras com acentos, pontuação, “ç” e outros caracteres especiais. Os arquivos ajustados são salvos na pasta "2_preparados".

3.2.4 Etapa - Upload dos Dados

Por fim, na etapa de *Upload* de Dados, os arquivos preparados na etapa anterior são carregados em um banco de dados *BigQuery*. Tabelas são criadas utilizando o código correspondente a cada arquivo CSV. Essa etapa garante que os dados estejam disponíveis de forma estruturada e otimizada para consultas e análises posteriores.

Ao concluir todo o processo de Extração dos Dados, é possível prosseguir para a etapa de Transformação dos Dados.

3.3 ETL - Transformação com *BigQuery*

A transformação dos dados no *BigQuery* tem como objetivos pré-processar os dados removendo inconsistências e levá-los para um modelo dimensional.

O *BigQuery* é um serviço de análise de dados em nuvem fornecido pelo *Google Cloud Platform*. É amplamente utilizado em projetos de dados, por causa da sua escalabilidade, desempenho rápido, integração com outras ferramentas do *Google Cloud* e recursos avançados de segurança e governança de dados.

Na Figura 3, o fluxograma da etapa de transformação é apresentado, os dados são recebidos da etapa anterior, estes passam por uma série de sub etapas de processamento, que serão explicadas detalhadamente nos próximos tópicos. Por fim, após serem processados, os dados são enviados para a próxima etapa do fluxo. A seguir serão apresentadas as etapas realizadas para a transformação dos dados.

3.3.1 União das tabelas

Após a extração das tabelas e sua inclusão automática no *BigQuery*, é viável proceder com o tratamento dos dados. As tabelas em questão, que abrangem informações sobre violência contra a mulher e violência doméstica e familiar contra a mulher, estão contidas no conjunto de

Figura 3 – Fluxograma Transformação



Fonte: Autoria própria.

dados denominado "raw_files", sendo elas:

- 2019
- Primeiro semestre 2020
- Segundo semestre 2020
- Primeiro semestre 2021
- Segundo semestre 2021
- Primeiro semestre 2022
- Segundo semestre 2022
- Primeiro trimestre 2023

Para isso, inicialmente foi verificado quais as colunas divergem entre as tabelas. Após o mapeamento das colunas faltantes foi possível fazer a junção de todas as tabelas, somando 61 colunas com 5.099.150 amostras.

3.3.2 Exclusão de colunas

Foram identificadas diversas colunas que não apresentavam um padrão consistente na inserção dos dados, contendo grande parte das informações como nulas. Um exemplo claro disso é a coluna "Religião do Suspeito", na qual apenas 12% das informações estavam preenchidas. Diante dessa situação, optou-se por excluir um total de 18 colunas, como as colunas Deficiência relacionada a doença rara do suspeito, Vínculo Órgão PJ do suspeito e Natureza Jurídica do Suspeito. Essa abordagem de exclusão das colunas com dados inconsistentes permite uma análise mais precisa e confiável dos dados remanescentes. Ao remover as colunas problemáticas, evitamos qualquer viés ou distorção nos resultados que poderiam ser causados pela presença de informações incompletas ou inconsistentes.

3.3.3 União de colunas redundantes

Durante a análise dos dados, observou-se a presença de colunas redundantes, como é o caso das colunas "País Da Vítima" e "Nacionalidade Vítima", que continham informações semelhantes, por exemplo, o país da vítima é o Brasil e a nacionalidade Brasileira, podendo nesse caso usar apenas uma dessas duas informações. Com o objetivo de simplificar e padronizar o conjunto de dados, decidiu-se mesclar essas colunas, adotando um nome padronizado, a fim de manter apenas uma única coluna.

3.3.4 Padronização dos Dados

Durante o processo de união das tabelas, foi constatado que existiam divergências nos nomes atribuídos a uma mesma categoria da coluna, como é o caso do canal de atendimento, que apresentava variações como "DENÚNCIA - ATENDIMENTO TELEFÔNICO", "TELEFÔNICO" e "Atendimento Telefônico". Com o intuito de padronizar essas informações e evitar duplicações, todos os valores semelhantes foram consolidados em um único nome. Outra padronização foi a substituição dos valores "Null" ou "N/D" por "NÃO INFORMADO". Ao evitar a presença de múltiplos nomes para a mesma categoria, evitamos confusões e ambiguidades na interpretação dos resultados.

3.3.5 Colunas de Identificação (ID)

Foi necessário criar a Coluna de Identificação (ID) para as vítimas e para os suspeitos como forma de identificação da pessoa. Todas as denúncias foram consideradas feitas por uma vítima e um suspeito diferentes, já que a base de dados não apresenta informações que verificam quem é a vítima e quem é o suspeito da denúncia, ou seja, não é possível ter o número real de quantas vezes a mesma vítima relatou alguma denúncia e nem quantas vezes um determinado suspeito foi denunciado por uma ou mais vítimas.

3.3.6 Modelo Dimensional

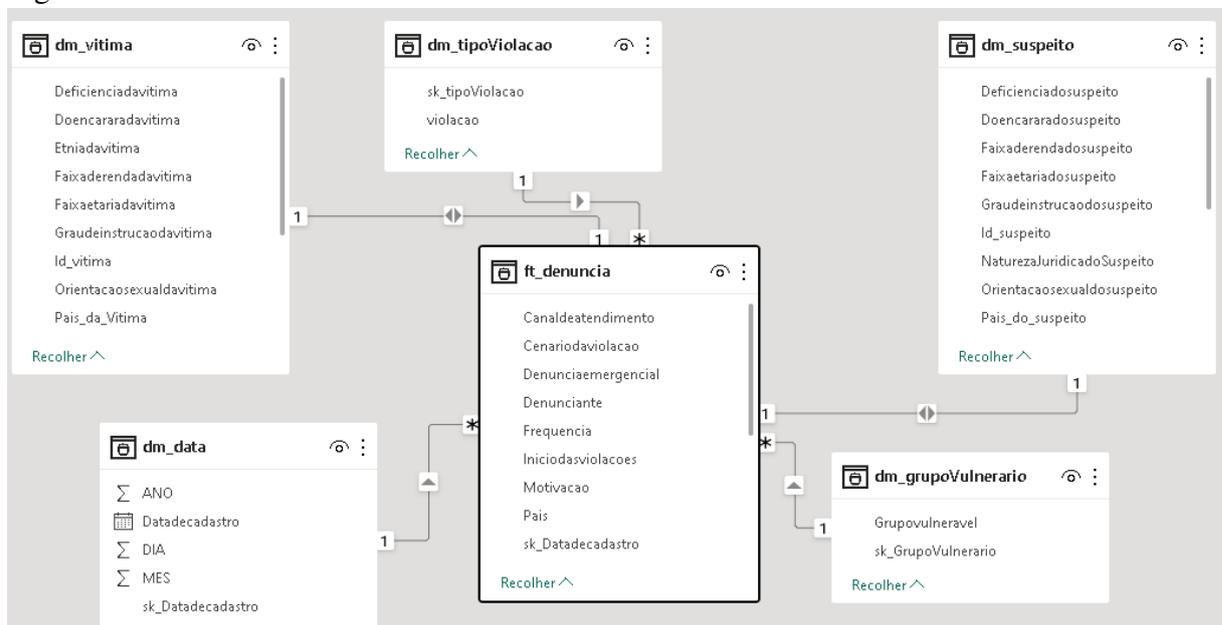
Ao concluir esses processos de transformação dos dados, é possível prosseguir para a etapa de transformar a tabela final, com 44 colunas e 5.099.150 amostras, em um modelo dimensional. A escolha dessa metodologia se deve ao fato de que a modelagem dimensional é altamente adequada para análise e consulta eficientes em grandes conjuntos de dados, pois

é uma técnica de projeto de banco de dados que visa facilitar consultas e análises de grandes volumes de dados, organizando os dados em dimensões e fatos e simplificando a compreensão da estrutura dos dados (SANTANA; SILVA, 2015).

A Figura 4 mostra o resultado final da transformação da tabela resultante para o modelo dimensional. Apresentando as tabelas:

- Fato: "ft_denuncia"
- Dimensões: "dm_data", "dm_grupoVulnerario", "dm_suspeito", "dm_tipoViolacao" e "dm_vitima"

Figura 4 – Modelo Dimensional



Fonte: Autoria própria.

Os códigos das transformações foram feitos em SQL e estão disponíveis em um repositório do GitHub, que pode ser acessado por meio do link fornecido no Apêndice A. Como no projeto de extração, a disponibilidade dos códigos no GitHub permite a contribuição por parte da comunidade interessada.

3.4 ETL - Carregamento para o Power BI

Após concluir a etapa de transformação dos dados, foi feita a conexão do *BigQuery* ao *Power BI* para a visualização e análise exploratória dos dados. Com essa conexão foi possível extrair *insights* significativos dos dados. O *Power BI* é uma plataforma de análise de dados e geração de *dashboards* da *Microsoft*, que permite conectar, transformar e visualizar dados de

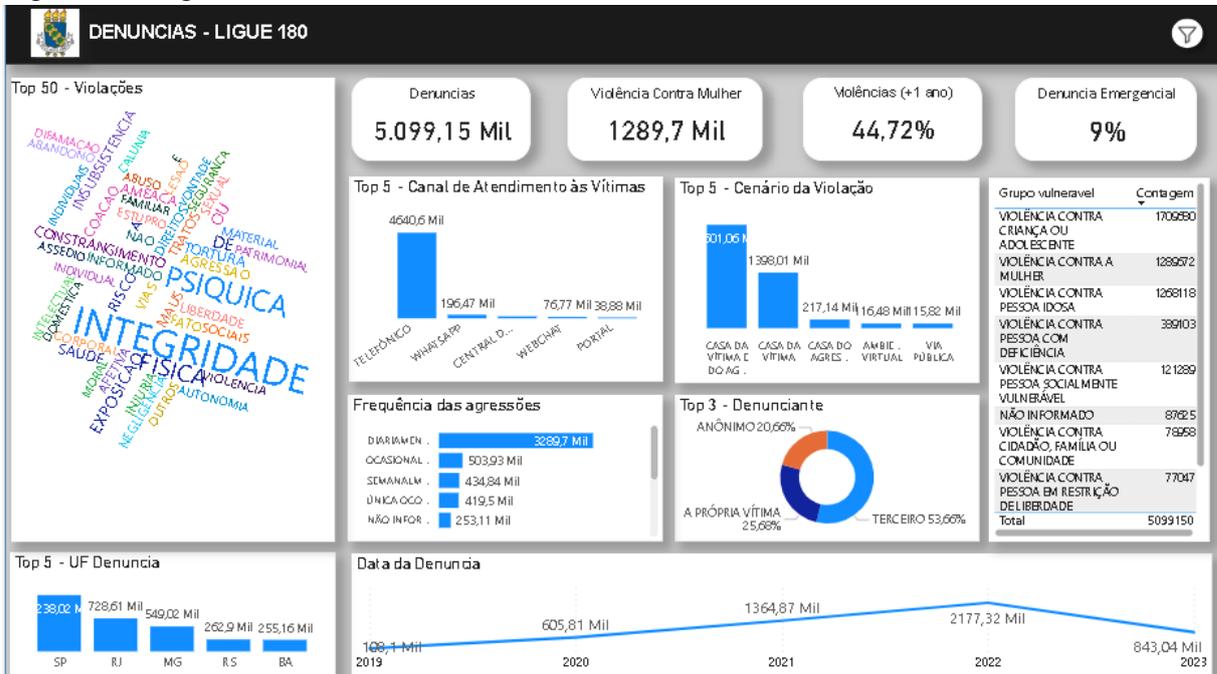
várias fontes em painéis interativos e intuitivos.

O *dashboard* foi dividido em 3 páginas:

- Perfil das Denúncias
- Perfil das Vítimas
- Perfil dos Suspeitos

A página de perfil das denúncias, como mostra a Figura 5, apresenta gráficos de barras com os 5 principais cenários que houveram as violações, os 5 canais de atendimento mais usados, a frequência das agressões, e os estados que mais foram feitas as denúncias. Além disso, um gráfico de pizza com o top 3 tipo de denunciante, um gráfico de linha mostrando a quantidade de denúncias ao longo dos últimos anos, uma tabela com a quantidade de denúncias por grupo vulnerável e uma nuvem com as palavras mais citadas nas descrições das violações. Ademais, foram incluídos 4 indicadores, sendo eles a quantidade de denúncias realizadas no período de 2019 ao primeiro trimestre de 2023, a quantidade de denúncias do grupo vulnerável “violência contra a mulher”, porcentagem das vítimas que sofrem violência a mais de 1 ano e porcentagem das denúncias emergenciais. O link do dashboard está disponível no apêndice A.

Figura 5 – Página do Perfil das Denúncias



Fonte: Autoria própria.

A página de perfil das vítimas, como mostra a Figura 6, apresenta as características mais comuns entre as vítimas de violência. A página conta com gráficos de barras, destacando o top 5 dos aspectos: faixa etária, país, UF, deficiência, doença rara, orientação sexual e a relação

entre vítima e suspeito. Além disso, possui dois gráficos de pizza, sendo eles, a raça/cor e o sexo mais comum das vítimas. Ademais, três tabelas com a quantidade de denúncias por grupo vulnerável, faixa de renda e profissão das vítimas.

Figura 6 – Página do Perfil das Vítimas



Fonte: Autoria própria.

O perfil do suspeito, apresentado na Figura 7, é formado por gráficos semelhantes ao perfil das vítimas, destacando as características mais comuns entre os suspeitos de violência. Além disso, fornece uma visualização adicional indicando se o suspeito já possui histórico de prisões.

As páginas estão equipadas com um botão de filtro que permite realizar pesquisas mais específicas nos perfis de denúncia, vítimas e suspeitos. Isso possibilita a realização de buscas mais direcionadas e a obtenção de resultados mais precisos com base em critérios específicos.

As Figuras 8 e 9 apresentam as janelas de filtros das páginas de vítimas e de suspeitos, respectivamente, sendo possível manipular informações de:

- Faixa etária;
- Sexo;
- Doença rara;
- Orientação sexual;
- Raça/cor;
- UF;

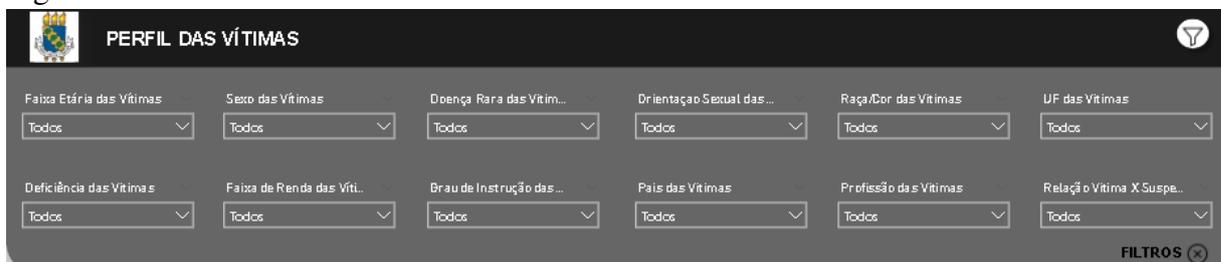
Figura 7 – Página do Perfil dos Suspeitos



Fonte: Autoria própria.

- Deficiência;
- Faixa de renda;
- Grau de instrução;
- País;
- Profissão;
- Suspeito Preso (Apenas para suspeitos)
- Relação Vítima X Suspeito (Apenas para vítimas)

Figura 8 – Filtros do Perfil das Vítimas



Fonte: Autoria própria.

A Figura 10 apresenta a janela de filtros da página de denúncias, é possível manipular informações de:

- Data de cadastro da denúncia;
- Canal de atendimento;

Figura 9 – Filtros do Perfil dos Suspeitos

Fonte: Autoria própria.

- Cenário da violência;
- Denúncia emergencial;
- Denunciante;
- Frequência;
- Início das violações.

Figura 10 – Filtros do Perfil das Denúncias

Fonte: Autoria própria.

O *dashboard* completo está disponível por meio do link fornecido no Apêndice A.

4 RESULTADOS E DISCUSSÕES

Para realizar a análise exploratória dos dados da Central de Atendimento à Mulher (Ligue 180), foram coletados, tratados e carregados os dados no *Power BI*. Durante o período analisado, que abrange de 2019 ao primeiro trimestre de 2023, foram registradas um total de 5.099.150 denúncias de violações. A análise desses dados permite uma compreensão mais aprofundada do problema. Para isso, o *Dashboard* foi dividido em três páginas que traçam o perfil das denúncias, vítimas e suspeitos.

4.1 Perfil das Denúncias

Os indicadores mostrados na Figura 11 revelam que foram registradas 5.099.150 denúncias e a quantidade de denúncias do grupo vulnerável “violência contra a mulher” foram 1.289.700, sendo 25% do total de denúncias, isso significa, que o Ligue 180 também é utilizado para denunciar casos de violências contra outros grupos vulneráveis. Ademais, outro indicador mostra que apenas 9% das denúncias são emergenciais, analisando este indicador nota-se que o Ligue 180 quase não é utilizado para denúncias urgentes, outra informação coletada é que mais de 44% das vítimas sofrem a violência a mais de 1 ano.

Figura 11 – Indicadores do Perfil das Denúncias



Fonte: Autoria própria.

Top 50 - Violações mais citadas: é uma nuvem de palavras, como mostra a Figura 12, que mais foram usadas na descrição das agressões, ou seja, as violações mais relatadas nas denúncias foram a de integridade, psíquica e física.

Top 5 UF Denúncias: Os 5 estados que mais foram registradas denúncias no Ligue 180 foram São Paulo, Rio de Janeiro, Minas Gerais, Rio Grande do Sul e Bahia, como é exibido na Figura 13. Esse número está ligado aos estados mais populosos do país, já que segundo o Censo 2022 do IBGE apontaram que São Paulo, Minas Gerais e Rio de Janeiro seguem sendo os estados mais populosos do país. Juntos, os três concentravam 39,9% da população brasileira (Instituto Brasileiro de Geografia e Estatística (IBGE), 2022).

Top 5 Canal de atendimento a vítima: O canal mais usado para denúncia foi o

Figura 15 – Top 5 Cenário de violação

Top 5 - Cenário da Violação



Fonte: Autoria própria.

que aumentasse o número das vítimas de violência já que mais de 50% das vítimas dividiam a casa com seus agressores por tempo integral.

Frequência das agressões: 64% das denúncias relatam que a violência é diária, como mostra a Figura 16. Segundo o site Atitude (2013), não denunciar é apontado como um caminho inseguro, pois cerca de 92% dos casos de violências frequentes, podem terminar em assassinato. Ou seja, o risco de morte por violência doméstica é iminente.

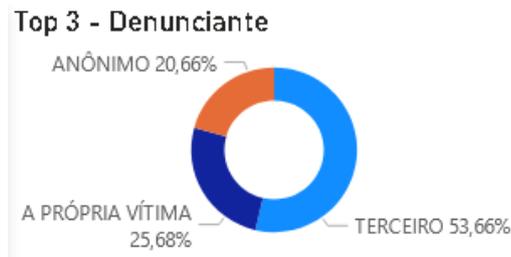
Figura 16 – Frequência das agressões



Fonte: Autoria própria.

Top 3 - Tipos de denunciante: Apenas 25% das vítimas de violência fizeram a denúncia contra o agressor, como mostra a Figura 17. Sendo a maioria dos casos feitas por terceiros ou em anonimato.

Figura 17 – Top 3 - Tipos de denunciante



Fonte: Autoria própria.

Tabela de grupo vulnerável: é possível observar na Figura 18, que os grupos vulneráveis mais atingidos são “violência contra a criança e o adolescente”, logo em seguida, “violência contra a mulher” e “violência contra o idoso”.

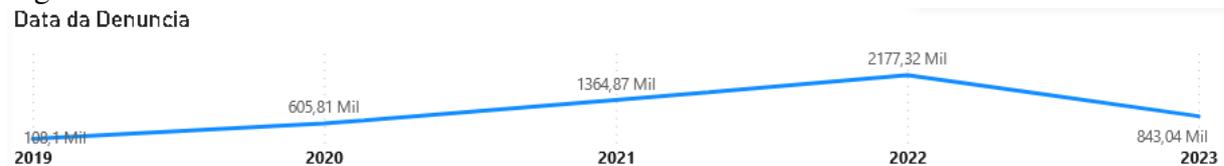
Figura 18 – Tabela de grupo vulnerável

Grupo vulneravel	Contagem
VIOLÊNCIA CONTRA CRIANÇA OU ADOLESCENTE	1709680
VIOLÊNCIA CONTRA A MULHER	1289672
VIOLÊNCIA CONTRA PESSOA IDOSA	1268118
VIOLÊNCIA CONTRA PESSOA COM DEFICIÊNCIA	389103
VIOLÊNCIA CONTRA PESSOA SOCIALMENTE VULNERÁVEL	121289
NÃO INFORMADO	87625
VIOLÊNCIA CONTRA CIDADÃO, FAMÍLIA OU COMUNIDADE	78958
VIOLÊNCIA CONTRA PESSOA EM RESTRIÇÃO	77047
Total	5099150

Fonte: Autoria própria.

Data da denúncia: Observa-se, na Figura 19, o crescimento durante o período de isolamento social por conta do Coronavírus, reforçando que o cenário onde mais ocorre a violação é na casa da própria vítima.

Figura 19 – Data da denúncia



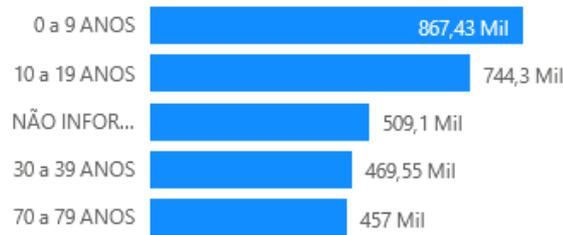
Fonte: Autoria própria.

4.2 Perfil das Vítimas

Top 5 - Faixa Etária das Vítimas: Ao analisar a faixa etária das vítimas, Figura 20, identificamos que os grupos mais afetados correspondem às pessoas com idade entre 0 e 19 anos, representando 31% do total de denúncias. Esses dados ressaltam a vulnerabilidade das mulheres jovens diante da violência de gênero.

Figura 20 – Top 5 - Faixa Etárias das Vítimas

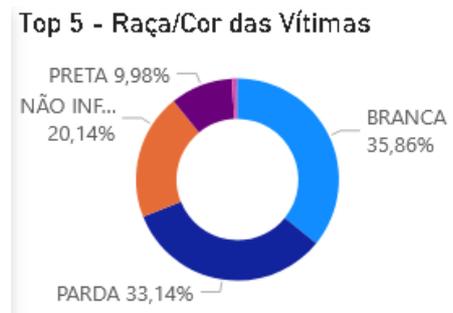
Top 5 - Faixa Etária das Vítimas



Fonte: Autoria própria.

Top 5 - Raça/Cor das Vítimas: Ao avaliar a raça/cor das vítimas, Figura 21, destacaram-se os seguintes grupos: de cor branca e parda, somando 69% das denúncias. Essa análise é importante para compreender a relação entre violência de gênero e questões raciais, permitindo direcionar ações específicas para os grupos mais afetados.

Figura 21 – Top 5 - Raça/Cor das Vítimas

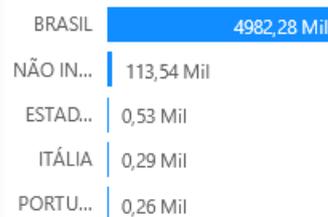


Fonte: Autoria própria.

Top 5 - País das Vítimas: A análise dos dados revelou que o Brasil é o país que mais frequentemente é mencionado nas denúncias, sendo responsável por mais de 97% das denúncias, Figura 22. O resultado é esperado já que é um serviço criado pelo Governo Federal do Brasil.

Figura 22 – Top 5 - País das Vítimas

Top 5 - País das Vítimas



Fonte: Autoria própria.

Tabela Grau de Instrução das Vítimas: A tabela de grau de instrução das vítimas forneceu *insights* sobre o nível educacional delas. É possível observar na Figura 23, que mais

57% dos valores não foram informados. Essa informação é fundamental para compreender o impacto da educação na prevenção da violência de gênero. Analisando a parcela dos dados informados a porcentagem de vítimas com escolaridade de nível superior é apenas 3,9%.

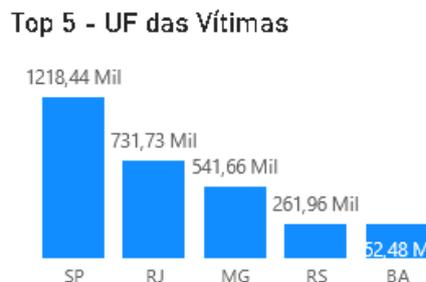
Figura 23 – Tabela Grau de Instrução das Vítimas

Grau de Instrução das Vítimas	Contagem
NÃO INFORMADO	2932286
FUNDAMENTAL INCOMPLETO	847533
ENSINO MÉDIO COMPLETO	447354
ANALFABETO(A)	244543
ENSINO MÉDIO INCOMPLETO	191461
SUPERIOR COMPLETO	164800
FUNDAMENTAL COMPLETO	126380
Total	5099150

Fonte: Autoria própria.

Top 5 - UF das Vítimas: Os 5 estados que mais foram registradas denúncias no Ligue 180 foram São Paulo, Rio de Janeiro, Minas Gerais, Rio Grande do Sul e Bahia, como é exibido na Figura 24. Comportamento similar ao UF das denúncias.

Figura 24 – *Top 5 - UF das Vítimas*



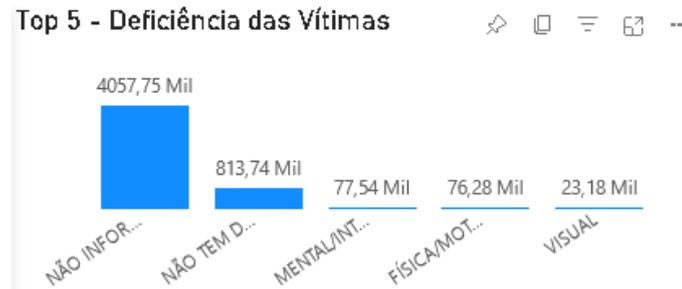
Fonte: Autoria própria.

Top 5 - Deficiência das Vítimas: Ao analisar a deficiência das vítimas, observamos na Figura 25, que mais de 79% das informações estão como “Não informado”. Essa informação deveria ser preenchida dada a importância de considerar as necessidades específicas das mulheres com deficiência. Além disso, de acordo com os dados, é possível observar que mais de 15% das vítimas não possuem nenhuma deficiência.

Faixa de Renda das Vítimas: A análise da faixa de renda das vítimas, Figura 26, revela que existem mais de 71% dos dados faltando, sendo que esses resultados são cruciais para entender como fatores socioeconômicos podem influenciar a exposição das mulheres à violência. Mas, analisando a parcela dos dados informados, apenas 12% das vítimas tem uma renda acima de 1 salário mínimo.

Top 3 - Sexo das Vítimas: Apesar da Central de Atendimento (Ligue 180) ser voltado

Figura 25 – Top 5 - Deficiência das Vítimas



Fonte: Autoria própria.

Figura 26 – Faixa de Renda das Vítimas

Faixa de Renda das Vítimas	Contagem
NÃO INFORMADO	3626040
ATÉ 1 SALÁRIO MÍNIMO	755903
MAIS DE 1 A 2 SALÁRIOS MÍNIMOS	453680
MAIS DE 2 A 5 SALÁRIOS	125056
Total	5094652

Fonte: Autoria própria.

para violência contra mulher, é possível verificar que mais de 26% das denúncias foram de vítimas do sexo masculino, Figura 27, fato que engloba as vítimas de grupos vulneráveis como idosos e crianças.

Figura 27 – Top 3 - Sexo das Vítimas



Fonte: Autoria própria.

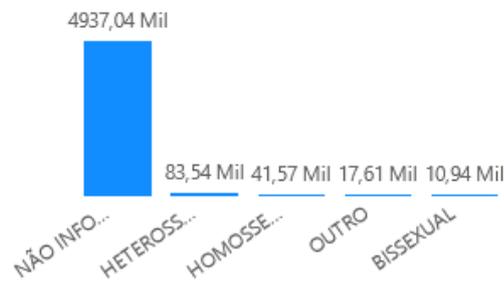
Top 5 - Orientação Sexual da Vítima: A análise da orientação sexual das vítimas revelou que mais de 96% dos dados não foram informados, Figura 28, sendo difícil traçar um perfil com a pouca quantidade de dados sobre esse tema.

Top 5 - Doença Rara da Vítima: Ao analisar a presença de doenças raras entre as vítimas, Figura 29, observou-se que 87% das vítimas não possuem nenhum tipo de doença rara.

Relação Vítima X Suspeito: A análise das relações entre vítimas e suspeitos revelou que a relação que mais reincidiu nas denúncias foi Pai/Mãe, sendo 24% do total das denúncias,

Figura 28 – Top 5 - Orientação Sexual da Vítima

Top 5 - Orientação Sexual das Vítimas



Fonte: Autoria própria.

Figura 29 – Top 5 - Doença Rara da Vítima

Top 5 - Doença Rara das Vítimas



Fonte: Autoria própria.

Figura 30. Essa análise permite identificar padrões e dinâmicas relacionais que podem contribuir para a compreensão e prevenção da violência de gênero.

Figura 30 – Relação Vítima X Suspeito

Relação Vítima X Suspeito



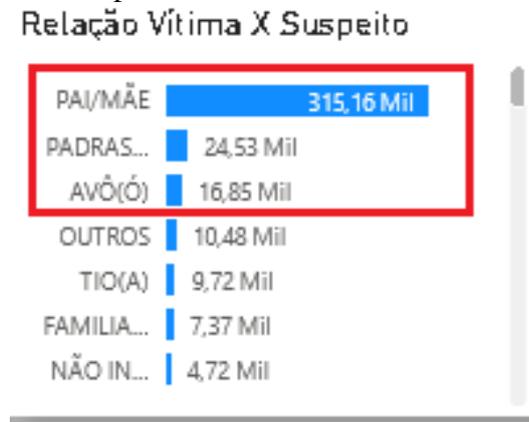
Fonte: Autoria própria.

Análise Complementar Relação Vítima X Suspeito: Filtrando o sexo da vítima para feminino e analisando as 3 faixas etárias que mais reincidiram, temos que as relações entre vítimas e suspeitos mais presentes foram:

- Faixa Etária: 0 à 9 anos. Como mostra a Figura 31, Pai/Mãe, Padrasto e Avô(ó);
- Faixa Etária: 10 à 19 anos. Como mostra a Figura 32, Pai/Mãe, Padrasto, Outros e Tio(a);
- Faixa Etária: 30 à 39 anos. Como mostra a Figura 33, Companheiro(a), Esposo(a),

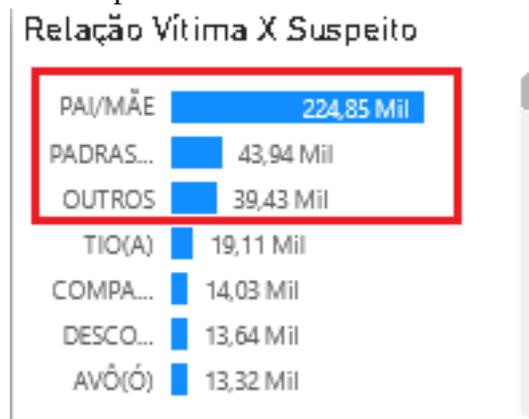
Ex-Companheiro(a) e Marido;

Figura 31 – Relação Vítima X Suspeito - Faixa Etária: 0 à 9 anos



Fonte: Autoria própria.

Figura 32 – Relação Vítima X Suspeito - Faixa Etária: 10 à 19 anos



Fonte: Autoria própria.

Profissão da Vítima: Os dados informados das profissões das vítimas tem um total de mais de 80% sem informação, como mostra a Figura 34. A segunda categoria de profissão que mais reincide é de vítimas estudantes com mais de 5%. Como a maioria dos dados estão faltando, acaba dificultando o direcionamento das ações de conscientização e suporte específicas para setores com maior incidência de casos.

4.3 Perfil dos Suspeitos

Top 5 - Faixa Etárias do Suspeito: Ao analisar a faixa etária dos suspeitos, foi identificado que o grupo com maior incidência não foi informado, sendo 1.250.000 faltantes, como mostra a Figura 35. Logo em seguida, a segunda faixa etária com maior incidência é entre 30 a 49, representando 40% do total de denúncias. Os dados informados ressaltam que os

Figura 33 – Relação Víctima X Suspeito - Faixa Etária: 30 à 39 anos
Relação Víctima X Suspeito



Fonte: Autoria própria.

Figura 34 – Profissão da Víctima

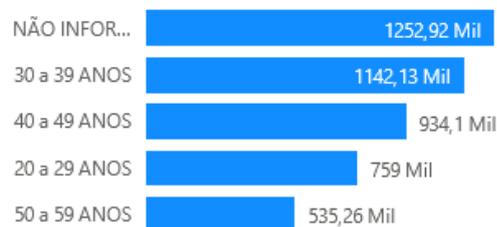
Profissão da Víctima	Contagem
NÃO INFORMADO	4121307
ESTUDANTE	303338
NÃO SABE INFORMAR	141487
OUTROS NÃO LISTADOS	139681
DOMÉSTICO/FAXINEIRO	19625
OUTROS	18319
TRABALHADORES DO	
Total	5099150

Fonte: Autoria própria.

suspeitos são mais maduros quando comparados com as vítimas, como mostrado no gráfico da Figura 20.

Figura 35 – Top 5 - Faixa Etárias do Suspeito

Faixa Etária dos Suspeitos

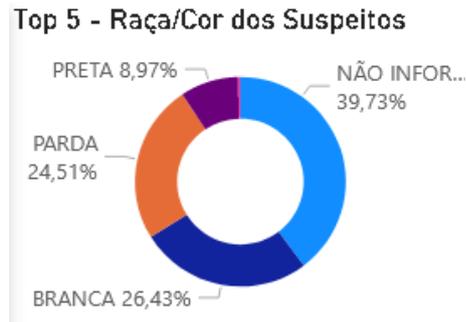


Fonte: Autoria própria.

Top 5 - Raça/Cor do Suspeito: Ao avaliar a raça/cor dos suspeitos, observa-se que mais de 39% dos dados não foram informados, Figura 36. Em segundo lugar o valor que mais incidu foi de suspeitos brancos com mais de 26% e em terceiro suspeitos pardos com mais de 24%.

Tabela Grau de Instrução do Suspeito: com a tabela da Figura 37 é possível observar

Figura 36 – Top 5 - Raça/Cor do Suspeito



Fonte: Autoria própria.

que mais de 80% dos valores não foram informados, e apenas 2,4% possuem grau de instrução superior.

Figura 37 – Tabela Grau de Instrução do Suspeito

Grau de Instrução dos Suspeitos	Contagem de sk_Suspeito
NÃO INFORMADO	4109488
FUNDAMENTAL INCOMPLETO	317235
ENSINO MÉDIO COMPLETO	289107
SUPERIOR COMPLETO	114000
Total	5099150

Fonte: Autoria própria.

Top 5 - País do Suspeito: A análise dos dados revelou que o Brasil é o país que a maioria suspeitos dos residem, sendo mais de 96% como é exibido na Figura 38.

Figura 38 – Top 5 - País do Suspeito

Top 5 - Pais dos Suspeitos



Fonte: Autoria própria.

Top 5 - UF do Suspeito: Os 5 dados de UF que mais frequentemente apareceram foram São Paulo, Rio de Janeiro, Minas Gerais, Não informado e Rio Grande do Sul, como é exibido na Figura 39.

Top 5 - Suspeito Preso: Analisando o gráfico se o suspeito já foi preso, foi observado

Figura 39 – Top 5 - UF do Suspeito



Fonte: Autoria própria.

na Figura 40 que mais de 55% dos dados não foram preenchidos e 40% dos suspeitos não têm antecedentes criminais.

Figura 40 – Top 5 - Suspeito Preso



Fonte: Autoria própria.

Faixa de Renda do Suspeito: A análise da faixa de renda dos suspeitos, Figura 41, revela que há mais de 87% dos dados faltando, sendo que essas informações são cruciais para entender como fatores socioeconômicos podem influenciar na violência. Mas, analisando os dados informados, apenas 6,3% dos suspeitos tem uma renda acima de 1 salário mínimo.

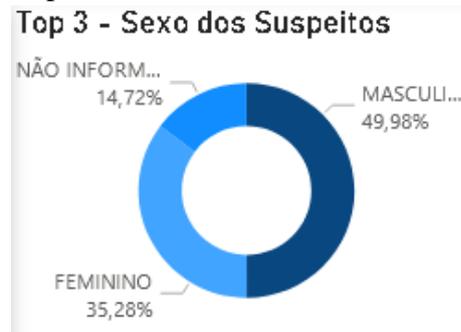
Figura 41 – Faixa de Renda do Suspeito

Faixa de renda do suspeito	Contagem
NÃO INFORMADO	4457529
ATÉ 1 SALÁRIO MÍNIMO	279191
MAIS DE 1 A 2 SALÁRIOS MÍNIMOS	243427
MAIS DE 2 A 5 SALÁRIOS MÍNIMOS	63081
MAIS DE 5 A 15	38734
Total	5094652

Fonte: Autoria própria.

Top 3 - Sexo do Suspeito: Ao analisar o sexo dos suspeitos, identificamos que os três principais grupos que são masculino com quase 50% dos dados, feminino com mais de 35%, e mais de 14% não foram informados, como mostra a Figura 42. Essa análise fornece *insights* sobre a distribuição de gênero dos suspeitos envolvidos nas denúncias do Ligue 180.

Figura 42 – Top 3 - Sexo do Suspeito



Fonte: Autoria própria.

Top 5 - Orientação Sexual do Suspeito: Ao analisar foi observado, Figura 43, que a grande maioria dos dados referentes a essa variável estava ausente. Cerca de 98% dos registros informados não continham informações sobre a orientação sexual dos suspeitos, o que limita a capacidade de traçar um perfil abrangente nesse aspecto.

Figura 43 – Top 5 - Orientação Sexual do Suspeito

Top 5 - Orientação Sexual dos Suspeitos



Fonte: Autoria própria.

Top 5 - Doença Rara do Suspeito: Analisando a relação de presença de doença rara entre os suspeitos, Figura 44, observou-se que a maioria dos registros (83%) não mencionava a existência de qualquer condição de doença rara entre os suspeitos. Além disso, em aproximadamente 16% dos casos, as informações sobre a presença ou ausência de doença rara não foram fornecidas.

Profissão do Suspeito: observou-se que uma parcela significativa dos registros não continha informações sobre a ocupação dos suspeitos. Cerca de 64% dos casos não forneciam dados sobre a profissão dos suspeitos, mais de 19% não puderam informar essa informação e mais de 6% foram categorizados como "outro" ou não listados, Figura 45.

Figura 44 – Top 5 - Doença Rara do Suspeito



Fonte: Autoria própria.

Figura 45 – Profissão do Suspeito

Profissão do suspeito	Contagem
NÃO INFORMADO	3269466
NÃO SABE INFORMAR	710162
OUTROS NÃO LISTADOS	321981
PEDREIRO	46432
PEDREIRO, SERVENTE DE OBRAS, PINTOR DE OBRAS, ELETRICISTA OU PROFISSIONAL DA CONSTRUÇÃO CIVIL	38070
Total	5099150

Fonte: Autoria própria.

4.4 Problemas Encontrados

A presença de uma quantidade significativa de informações faltantes nos dados da Central de Atendimento à Mulher (Ligue 180) apresenta alguns problemas, como:

- Limitação na análise: A falta de dados sobre variáveis importantes, como orientação sexual da vítima e do suspeito, doenças raras e profissão, limita a capacidade de realizar análises abrangentes e compreensivas. Essas lacunas dificultam a obtenção de *insights* completos e precisos sobre os fatores associados à violência contra as mulheres.
- Dificuldade na identificação de padrões: A ausência de informações prejudica a identificação de possíveis padrões, tendências ou correlações relevantes nos dados. Isso dificulta a compreensão dos fatores de risco e a implementação de estratégias efetivas de prevenção e combate à violência de gênero.

Os dados da Central de atendimento à mulher possui muitas informações e estrutura não padronizadas. Além disso, é explanado que o Ligue 180 é voltado para denúncias de violência contra a mulher e violência doméstica contra a mulher, mesmo assim foram cadastrados 27% de vítimas do sexo masculino. Ademais, existem outros grupos vulneráveis nos dados como violência contra criança, idoso, pessoa em situação de rua, entre outros. Outra dificuldade é

não ter um identificador de vítima e de suspeito, com isso torna-se impossível calcular e fazer análises sobre reincidência de violência com a mesma vítima e suspeito

Para superar esses problemas, é fundamental investir em melhorias no registro de dados, garantindo a inclusão de informações abrangentes, consistentes e de qualidade. Isso permitirá uma análise mais robusta e precisa, contribuindo para uma compreensão mais completa da violência contra as mulheres e o desenvolvimento de estratégias efetivas para combatê-la.

5 CONCLUSÃO

Este trabalho apresentou uma proposta de metodologia de coleta, transformação e análise de dados da Central de Atendimento à Mulher (Ligue 180). Os objetivos propostos neste trabalho foram parcialmente alcançados, pois não foi possível definir completamente os perfis das vítimas, suspeitos e denúncias de violência contra a mulher, devido às inconsistências apresentadas no registro das informações. Para diminuir este problema é importante a implementação de medidas que garantam a coleta mais completa e precisa de dados, com a construção de treinamentos para quem registra as denúncias das vítimas e a revisão dos formulários, incluindo campos obrigatórios, para direcionar medidas específicas para atender as populações mais vulneráveis em futuras pesquisas sobre o tema.

A metodologia elaborada para este trabalho se mostrou eficiente, permitindo a estruturação dos dados coletados, e possibilitando contribuições externas ao estudo, com a disponibilização dos códigos utilizados. O desenvolvimento de um *dashboard* que reúne os dados dos anos analisados proporcionou a avaliação e descrição das características relevantes sobre as vítimas, suspeitos e denúncias. Essa ferramenta permitiu uma visualização detalhada das informações, facilitando a compreensão do cenário de violência contra a mulher nos últimos anos. Os *dashboards* podem ser utilizados por políticas públicas para direcionar recursos e esforços de maneira mais eficiente, como os dados sobre a alta frequência de agressões em ambientes domésticos podem ajudar a formular campanhas de conscientização focadas em ambientes familiares.

A análise dos dados evidenciou a importância desse tipo de estudo no combate à violência de gênero. Os resultados obtidos fornecem informações relevantes para a formulação de políticas públicas mais eficazes, como, a faixa etária mais afetada, os estados que mais registram a violência e a relação entre vítimas e suspeitos, além de evidenciar a necessidade de melhoria do registro de informações. Além disso, os resultados contribuem para ampliar o conhecimento sobre o tema, conscientizar a sociedade e incentivar ações concretas em prol de uma realidade onde todas as mulheres possam viver livres de violência e opressão.

5.1 Trabalhos Futuros

Em relação aos trabalhos futuros, existem oportunidades promissoras para expandir o conhecimento e a compreensão da violência contra a mulher. Um possível caminho seria

a adição dos anos iniciais do Ligue 180 à análise, realizando uma adaptação nos processos de transformação e visualização, tendo como objetivo proporcionar uma visão histórica mais completa. Outra possibilidade é a exploração e integração de outras bases de dados relevantes, como o Anuário Brasileiro de Segurança Pública e o DATASUS, para fornecer uma perspectiva mais abrangente sobre o fenômeno da violência de gênero.

Para aprimorar a eficiência e escalabilidade do processo de coleta, tratamento e análise de dados, é possível considerar a automação do processo de transformação de dados utilizando ferramentas como o *dbt*. Essa abordagem facilitaria a atualização contínua dos dados e a execução de análises em tempo real.

Além disso, é possível usar algoritmos de aprendizado de máquina para a realização de análises preditivas de violência contra a mulher, sendo possível desenvolver modelos capazes de fazer previsões sobre a ocorrência de casos, levando em consideração diversas variáveis e características importantes. Essa análise preditiva poderia contribuir para a criação de estratégias preventivas mais eficazes e auxiliar ações de intervenção direcionadas.

REFERÊNCIAS

- ATITUDE, C. e. **Feminicídio: Desafios e recomendações para enfrentar a mais extrema violência contra as mulheres**. 2013. Disponível em: <https://bit.ly/3rcd1wI>. Acesso em: 02/07/2023.
- BEAULIEU, A. **Learning SQL: master SQL fundamentals**. [S. l.]: "O'Reilly Media, Inc.", 2009.
- BRASIL. **Lei Nº 11.340. Lei Maria da Penha**. Brasília, DF, 2006. Art. 5.
- CASERTA, J.; KIMBALL, R. **The Data Warehouseetl Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data**. [S. l.]: Wiley, 2013.
- DATE, C. J. **Introdução a sistemas de bancos de dados**. [S. l.]: Elsevier Brasil, 2004.
- ESSY, D. B. **A evolução histórica da violência contra a mulher no cenário brasileiro: do patriarcado à busca pela efetivação dos direitos humanos feminino**. 2017. Disponível em: <https://bit.ly/2XMPFz5>. Acesso em: 20/08/2021).
- FÁVERO, L. P.; BELFIORE, P. **Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®**. [S. l.]: Elsevier Brasil, 2017.
- FONSECA, D. H. d.; RIBEIRO, C. G.; LEAL, N. S. B. Violência doméstica contra a mulher: realidades e representações sociais. **Psicologia & Sociedade**, SciELO Brasil, v. 24, p. 307–314, 2012.
- Fórum Brasileiro de Segurança Pública. **Violência contra Meninas e Mulheres**. 2023. Disponível em: <https://forumseguranca.org.br/publicacoes/violencia-contra-meninas-e-mulheres/>. Acesso em: 02/07/2023.
- GOV.BR. **Serviços e Informações do Brasil**. 2024. Disponível em: <https://www.gov.br/pt-br>.
- Instituto Brasileiro de Geografia e Estatística (IBGE). **Censo Demográfico 2022: Panorama**. 2022. Disponível em: <https://censo2022.ibge.gov.br/panorama/>.
- KIMBALL, R.; ROSS, M. **The data warehouse toolkit: the complete guide to dimensional modeling**. [S. l.]: John Wiley & Sons, 2011.
- KNAFLIC, C. N. **Storytelling with data: A data visualization guide for business professionals**. [S. l.]: John Wiley & Sons, 2015.
- LIGUE180. 2024. Disponível em: <https://dados.gov.br/dados/conjuntos-dados/central-de-atendimento-a-mulher--ligue-180>. Acesso em: 02/07/2023.
- MARCOLINO, E. d. C.; SANTOS, R. C. d.; CLEMENTINO, F. d. S.; LEAL, C. Q. A. M.; SOARES, M. C. d. S.; MIRANDA, F. A. N. d.; SOUTO, R. Q. O distanciamento social em tempos de covid-19: uma análise de seus rebatimentos em torno da violência doméstica. **Interface-Comunicação, Saúde, Educação**, SciELO Brasil, v. 25, 2021.
- MCKINNEY, W. **Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython**. [S. l.]: Novatec Editora, 2018.
- MDHC. 2024. Disponível em: <https://www.gov.br/mdh/pt-br>. Acesso em: 02/07/2023.

MILTON, M. **Use a Cabeça! Análise de Dados**. Rio de Janeiro: Alta Books, 2010.

OKABAYASHI, N. Y. T.; TASSARA, I. G.; CASACA, M. C. G.; FALCÃO, A. de A.; BELLINI, M. Z. Violência contra a mulher e feminicídio no brasil-impacto do isolamento social pela covid-19. **Brazilian Journal of Health Review**, v. 3, n. 3, p. 4511–4531, 2020.

SANTANA, L. T.; SILVA, E. d. O. da. Análise de modelos dimensionais: um estudo de caso com dados do sistema de processo seletivo do ifet. **Caderno de Estudos em Sistemas de Informação**, v. 2, n. 2, 2015.

SILVA, L. E. L. d.; OLIVEIRA, M. L. C. d. Violência contra a mulher: revisão sistemática da produção científica nacional no período de 2009 a 2013. **Ciência & Saúde Coletiva**, SciELO Brasil, v. 20, p. 3523–3532, 2015.

SILVESTRE, A. **Análise de dados e estatística descritiva**. [S. l.]: Escolar editora, 2007.

YAU, N. **Visualize this: the FlowingData guide to design, visualization, and statistics**. [S. l.]: John Wiley & Sons, 2011.

APÊNDICE A – LINKS ÚTEIS

O projeto em *Python* para Extração de Dados está disponível no seguinte link: <https://github.com/thaisfelix20/extracao_ligue180>.

Os códigos da fase de Transformação de Dados, e as tabelas finais transformadas estão disponíveis no link <https://github.com/thaisfelix20/transformacao_ligue180>.

O *dashboard* construído com o *Power BI* para exploração dos resultados está disponível no link <<https://bit.ly/3zrtjGf>>