



**Centro Paula Souza**

**Faculdade de Tecnologia de Votorantim**

Curso Superior de Tecnologia em Ciência de Dados para Negócios

**PROJETO INTEGRADOR IV**

**Análise de Ocorrências de Violência Contra a Mulher em Sorocaba e  
Votorantim**

DÉBORA KOCKS NOGUEIRA  
GIOVANA PERUGINI GUENKA  
STEFANIE MAYUMI INACIO KOBAYASHI  
VITÓRIA LISAUSKAZ FERRAZ DA SILVA

NOVEMBRO/2025

# SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>3</b>
<b>2. APRESENTAÇÃO DA EMPRESA .....</b>	<b>4</b>
<b>3. DESCRIÇÃO E ANÁLISE DO PROBLEMA.....</b>	<b>5</b>
3.1 ANÁLISE DO CENÁRIO .....	5
3.2 DIAGNÓSTICO E IDENTIFICAÇÃO DE PROBLEMAS .....	7
3.3 ALINHAMENTO COM OBJETIVOS ESTRATÉGICOS DA EMPRESA .....	8
3.4 ESTRUTURAÇÃO DAS AÇÕES E PROJETOS PROPOSTOS.....	9
<b>4. PLANEJAMENTO ESTRATÉGICO DA SOLUÇÃO (OBJETIVOS E PLANO DE AÇÃO) .....</b>	<b>12</b>
4.1 ESTRUTURA ESTRATÉGICA: OBJETIVOS SMART E PRIORIDADES .....	13
4.2 MAPEAMENTO ESTRATÉGICO (BALANCED SCORECARD) E METODOLOGIA PDCA .....	14
<b>5. METODOLOGIA DE DESENVOLVIMENTO E FONTE DE DADOS .....</b>	<b>18</b>
5.1 PIPELINE DE DADOS (PROPOSTA DE ETL).....	18
5.2 ANÁLISE EXPLORATÓRIA E COMPREENSÃO DOS DADOS.....	21
<b>6. DESENVOLVIMENTO DO PROTÓTIPO E IMPLEMENTAÇÃO DA SOLUÇÃO.....</b>	<b>25</b>
6.1 ARQUITETURA DO SISTEMA PROPOSTO .....	25
6.2 FUNCIONALIDADES IMPLEMENTADAS.....	26
6.3 FLUXO DO USUÁRIO E EXPERIÊNCIA (UX/UI).....	31
6.5 DESAFIOS E AJUSTES DURANTE O DESENVOLVIMENTO .....	37
<b>7. TESTES E VALIDAÇÕES.....</b>	<b>38</b>
7.1 TESTES FUNCIONAIS E TÉCNICOS .....	39
7.2 TESTES COM USUÁRIOS E STAKEHOLDERS.....	40
7.3 INDICADORES DE SUCESSO E PRÓXIMOS PASSOS.....	40
<b>8. CONCLUSÃO .....</b>	<b>41</b>
<b>9. REFERÊNCIAS.....</b>	<b>43</b>

## 1. Introdução

Diante da persistente e preocupante realidade da violência contra a mulher, o presente trabalho foi desenvolvido no âmbito da disciplina Projeto Integrador 4, com o objetivo de aplicar técnicas de ETL (*Extract, Transform, Load*), Extrair, Transformar e Carregar, para o tratamento e a análise de dados públicos provenientes da Secretaria de Segurança Pública do Estado de São Paulo.

O projeto tem como foco as ocorrências registradas pelas Delegacias de Defesa da Mulher (DDM) nos municípios de Sorocaba e Votorantim, buscando organizar e estruturar as informações disponíveis de forma a viabilizar análises consistentes e interpretações fundamentadas sobre o tema.

A proposta consiste no desenvolvimento de um painel público de dados interativo, voltado à conversão de dados brutos em uma ferramenta de monitoramento acessível, dinâmica e informativa. Tal iniciativa visa subsidiar os órgãos de administração pública no direcionamento estratégico e no aprimoramento de políticas públicas, bem como promover a conscientização e a educação da população acerca da gravidade e da complexidade da violência de gênero.

Embora a atenção dedicada ao tema e a disponibilidade de dados sobre a violência contra a mulher no ambiente virtual tenham aumentado ao longo dos anos, a gestão e a formulação de políticas públicas eficazes no Brasil frequentemente esbarram no desafio da dispersão e da complexidade dos dados oficiais. Diversas bases de dados e relatórios não são apresentadas de maneira acessível ou visualmente intuitiva, dificultando a rápida e profunda compreensão da dinâmica da violência por parte dos gestores públicos e da população em geral. Nesse sentido, o desenvolvimento deste painel justifica-se como uma aplicação prática da Ciência de Dados, utilizando métodos de coleta, tratamento e visualização para gerar insights (percepções) claros e aplicáveis. A ferramenta se estabelece como uma solução tecnológica que centraliza, organiza e permite a análise das informações específicas de Sorocaba e Votorantim, oferecendo suporte para a tomada de decisão e o monitoramento contínuo.

## 2. Apresentação da Empresa

A Secretaria da Mulher (Semul) é um órgão da Administração Municipal de Sorocaba, instituída em 2025, com a finalidade de desenvolver e implementar políticas públicas voltadas à valorização, proteção e emancipação das mulheres. Sua criação representa um marco no fortalecimento da rede municipal de apoio e promoção da equidade de gênero, reafirmando o compromisso da administração pública com o combate à violência contra a mulher e a garantia dos direitos fundamentais.

Entre suas principais atribuições estão o desenvolvimento de políticas de equidade, a coordenação de programas de prevenção e enfrentamento à violência de gênero e a oferta de suporte psicossocial, jurídico e econômico a mulheres em situação de vulnerabilidade.

A parceria estabelecida com a Semul tem como propósito disponibilizar dados tratados e organizados por meio deste projeto, possibilitando que a secretaria utilize informações estruturadas e acessíveis para subsidiar o planejamento de ações, a formulação de políticas públicas e o monitoramento de indicadores relacionados à violência contra a mulher.

### 3. Descrição e Análise do Problema

Em 2025, a cidade de Sorocaba inaugurou a Secretaria da Mulher (Semul), órgão municipal destinado a promover o apoio integral às mulheres e a coordenar políticas de prevenção e combate à violência de gênero. Essa criação representa um avanço significativo no fortalecimento da rede de proteção e promoção da equidade de gênero no município, contextualizando a relevância desta iniciativa.

Apesar da amplitude de sua atuação, identificou-se uma carência de informações estruturadas, especialmente no que diz respeito à disponibilização de dados unificados e atualizados sobre a incidência de casos de violência de gênero em Sorocaba. Essa lacuna limita a capacidade da secretaria de planejar, monitorar e avaliar de forma eficiente suas políticas públicas.

Diante desse contexto, este projeto propõe o desenvolvimento de um dashboard interativo, reunindo informações provenientes dos Boletins de Ocorrência das Delegacias de Defesa da Mulher (DDM) do município. O painel incluirá variáveis essenciais, como município, delegacia, data, hora, período do dia, dia da semana, mês, ano, tipo e local da ocorrência, bairro, logradouro, coordenadas geográficas, rubrica, descrição da conduta e natureza apurada.

A ferramenta permitirá identificar padrões e tendências, fornecendo à Semul uma base analítica sólida para subsidiar a formulação de estratégias mais eficazes de prevenção, proteção e acolhimento das mulheres. Dessa forma, o projeto contribui diretamente para o fortalecimento da gestão pública baseada em evidências, promovendo a transparência, a eficiência e o impacto social das ações voltadas à defesa dos direitos das mulheres em Sorocaba.

#### 3.1 Análise do Cenário

De acordo com a Organização das Nações Unidas, a violência contra as mulheres é definida como qualquer ato de agressão baseado no gênero que cause ou possa causar danos físicos, sexuais ou psicológicos, incluindo ameaças,

coerção ou privação de liberdade, tanto em contextos públicos quanto privados (ONU, s.a.).

O cenário da violência contra a mulher no Brasil revela uma crescente e preocupante escalada. De acordo com o Relatório Anual Socioeconômico da Mulher (RASEAM) de 2025, foram registradas no Sistema de Informação de Agravos de Notificação (SINAN) 302.856 notificações de violência doméstica, sexual e outras formas de violência contra mulheres no ano de 2023, o que representa um aumento significativo em comparação ao ano de 2022, no qual foi observado um registro de 216.024 notificações.

Entre 2015 e 2023, o *Fórum Brasileiro de Segurança Pública* registrou mais de dez mil casos de feminicídio no país, evidenciando um aumento de 1,4% nas ocorrências entre 2022 e 2023. Em 2023 ano, foram contabilizadas 1.463 vítimas, o que corresponde a uma média de mais de quatro mulheres assassinadas por dia (RIBEIRO *et al.*, 2024).

Um estudo publicado na Revista do SUS (2025), demonstra com base nos registros do Sistema de Informação de Agravos de Notificação de Violência Interpessoal e Autoprovocada (SINAN-VIVA), que foram realizadas mais de 220 mil notificações de violência contra mulheres em 2014, número que triplicou em uma década, alcançando aproximadamente 670 mil registros em 2023. As formas de violência mais prevalentes foram a física e a psicológica/moral, sendo o domicílio o principal local das ocorrências e o cônjuge o principal agressor, o que evidencia a persistência da violência de gênero em contextos íntimos e familiares. Esse crescimento evidencia o agravamento da violência de gênero no Brasil, além de possíveis avanços na notificação e visibilidade dos casos junto aos serviços de saúde e segurança pública.

Embora o debate nacional seja robusto, a disponibilidade pública e acessível desses dados frequentemente se restringe a contextos regionais, estaduais ou nacionais, falhando em prover a granularidade necessária para intervenções localizadas.

Nesse panorama de dados dispersos em nível federal, observa-se que, ao se concentrar nos municípios de Sorocaba e Votorantim, o acervo informacional sobre a violência contra a mulher evidencia a ausência de uma plataforma unificada e acessível. Essa segmentação de dados dificulta que gestores municipais e redes de apoio tenham uma visão holística e atualizada do problema, comprometendo a rápida identificação de padrões de risco e a otimização dos recursos na ponta.

A problemática central deste projeto, portanto, estabelece-se na seguinte questão:

De que forma o desenvolvimento de um painel de monitoramento interativo, baseado nos conceitos de Ciência de Dados, pode transformar a dispersão de dados sobre violência contra a mulher em Sorocaba e Votorantim em um recurso estratégico, capaz de subsidiar políticas públicas mais eficazes e promover a conscientização da sociedade?

Por meio da resposta a essa questão, busca-se demonstrar a viabilidade e o impacto da aplicação tecnológica na gestão social e na segurança pública no contexto local.

### 3.2 Diagnóstico e Identificação de Problemas

Conforme mencionado anteriormente, este trabalho busca auxiliar os órgãos responsáveis pela segurança e gestão pública dos municípios de Sorocaba e Votorantim no monitoramento dos dados de violência local, na criação e direcionamento de políticas públicas de enfrentamento da violência contra a mulher e aumentar a conscientização da população sobre o presente tema.

Os municípios enfrentam desafios semelhantes quanto à fragmentação e falta de padronização das informações, o que dificulta a análise integrada e o acompanhamento dos casos.

Entre os pontos fortes, destacam-se o foco territorial, que permite análises mais detalhadas da realidade local, e o uso de um pipeline automatizado em Python e Pandas, que garante atualização contínua dos dados provenientes da Secretaria

de Segurança Pública do Estado de São Paulo (SSP-SP). A visualização interativa dos resultados também torna o acesso às informações mais intuitivo e útil para gestores e cidadãos.

Por outro lado, o projeto enfrenta fragilidades, como a dependência da qualidade dos dados oficiais, a necessidade de manutenção técnica constante e a escassez de profissionais especializados em análise de dados nos municípios.

No contexto externo, há oportunidades ligadas à crescente demanda por transparência, apoio institucional e engajamento social em torno da violência de gênero. Contudo, ameaças como a subnotificação dos casos e possíveis restrições no acesso a dados públicos podem comprometer a continuidade e a precisão das análises.

Conclui-se que o principal problema está na falta de integração e padronização dos dados de violência contra a mulher em nível municipal, o que impede uma resposta mais rápida e eficaz dos órgãos públicos. O projeto, portanto, busca transformar dados dispersos em informações estratégicas para apoiar decisões e fortalecer o enfrentamento à violência de gênero em Sorocaba e Votorantim.

### 3.3 Alinhamento com Objetivos Estratégicos da Empresa

Através da reunião com a Sra. Aparecida Gutierrez, representante do Conselho Municipal dos Direitos da Mulher (CMESO) e da Secretaria da Mulher, evidencia-se que a Secretaria da Mulher (Semul) enfrenta um desafio central: a inexistência de dados integrados, atualizados e padronizados sobre violência de gênero nos municípios de Sorocaba e Votorantim. Esse contato inicial foi breve e limitado à apresentação dessa dificuldade, sem que outros aspectos adicionais fossem levantados posteriormente.

A ausência de informações compromete a capacidade do órgão de planejar, monitorar e avaliar políticas públicas de forma eficiente. Nesse cenário, um dashboard interativo é a solução estratégica para transformar dados dispersos



em informações acionáveis, fortalecendo a tomada de decisão e a formulação de políticas públicas.

O projeto se apoia em pontos fortes identificados, como o foco territorial que permite análises detalhadas da realidade local, o uso de pipelines automatizados em Python, através da biblioteca Pandas, garantindo atualização contínua dos dados, e a visualização interativa, que torna as informações acessíveis e intuitivas para gestores e cidadãos. Ao mesmo tempo, reconhece fragilidades, incluindo a dependência da qualidade dos dados oficiais, a necessidade de manutenção técnica constante e a escassez de profissionais especializados em análise de dados nos municípios.

Para potencializar o impacto da solução, foram definidas estratégias específicas:

- Aproveitar o foco territorial e a automação de pipelines de dados para desenvolver um dashboard interativo que facilite a visualização das informações, promovendo transparência, engajamento social e suporte à tomada de decisão.
- Garantir documentação detalhada do processo de análise e manutenção contínua do pipeline de dados, assegurando a confiabilidade e a atualização das informações disponíveis.

Embora a reunião com a Aparecida tenha sido única, os pontos levantados contribuíram para insights que orientaram as decisões sobre as funcionalidades e a estrutura do dashboard, aspectos que foram discutidos e validados no encontro, assegurando que a solução proposta atende às necessidades da Semul.

### 3.4 Estruturação das Ações e Projetos Propostos

Para suprir a ausência de um sistema integrado de dados sobre violência de gênero nos municípios de Sorocaba e Votorantim, este projeto propõe o desenvolvimento de uma solução automatizada para coleta, tratamento,

armazenamento e visualização das informações disponibilizadas pela Secretaria de Segurança Pública do Estado de São Paulo (SSP-SP). A iniciativa visa facilitar o acesso a indicadores atualizados e confiáveis, apoiando a tomada de decisões estratégicas voltadas ao enfrentamento da violência contra a mulher.

#### Visão Geral das Soluções

As soluções propostas concentram-se no desenvolvimento de um pipeline de dados automatizado, na estruturação de um ambiente de armazenamento em nuvem e na criação de painéis interativos de visualização, conforme descrito a seguir:

- **ETL Automatizado (Extract, Transform, Load)**

Será desenvolvido um pipeline em linguagem Python, utilizando a biblioteca Pandas, responsável por coletar, processar e consolidar os dados das planilhas disponibilizadas pela Secretaria de Segurança Pública de São Paulo SSP-SP e Sistema de Informação de Agravos de Notificação - Sinan. O sistema executará automaticamente o download, leitura e unificação das planilhas referentes às DDM de Sorocaba e Votorantim.

Justificativa: Essa automatização elimina a necessidade de processamento manual, reduz erros e assegura atualizações contínuas e consistentes.

- **Padronização e Enriquecimento dos Dados**

Após a coleta, os dados passarão por tratamento e formatação, incluindo a exclusão de registros inválidos, a padronização de nomes de colunas e a criação de novas variáveis derivadas, como mês, dia da semana e perfil das pessoas envolvidas.

Justificativa: A padronização garante a qualidade e a integridade das informações, possibilitando análises mais precisas.

- **Armazenamento em Nuvem (Google BigQuery)**

Os dados tratados serão armazenados no Google BigQuery, que atuará como repositório central para o projeto, possibilitando consultas otimizadas e integrações diretas com ferramentas de análise e visualização.

Justificativa: A utilização do BigQuery garante alto desempenho e segurança no gerenciamento das informações, além de proporcionar acesso ágil e confiável aos dados, favorecendo a sustentabilidade e expansão futura da solução.

- **Definição de Indicadores-chave de desempenho**

Serão definidos indicadores como número de ocorrências por bairro, dia da semana, tipo de violência e perfil de agressores e das vítimas, permitindo o acompanhamento de tendências e a avaliação das políticas públicas locais.

Justificativa: Os KPIs auxiliam no monitoramento contínuo dos resultados e na identificação de áreas que demandam maior atenção ou recursos.

- **Dashboard interativo (Google Looker Studio)**

Serão desenvolvidos dashboards interativos no Google Looker Studio, permitindo a exploração visual dos indicadores por meio de filtros de tempo, localidade e tipo de ocorrência.

Justificativa: A interface interativa facilita a interpretação dos dados e apoia a tomada de decisão pela Secretaria da Mulher e demais órgãos públicos.

### **Priorização e Etapas de Implementação**

O plano de execução foi estruturado em três fases sequenciais, organizadas conforme a ordem de implementação e o grau de complexidade técnica de cada etapa:

#### **Fase 1 – Coleta e Padronização dos Dados:**

Implementação do pipeline de ETL e integração com o Google BigQuery, consolidando as informações históricas das DDM de Sorocaba e Votorantim. Essa fase é fundamental para garantir uma base de dados estruturada, consistente e confiável, que servirá de alicerce para as etapas seguintes.

## Fase 2 – Desenvolvimento dos Dashboards Interativos:

Criação dos painéis no Google Looker Studio, com definição dos indicadores, métricas e filtros de análise. Essa fase permitirá a disponibilização visual das informações, facilitando a exploração dos dados e apoiando a interpretação dos resultados por gestores e tomadores de decisão.

## Fase 3 – Validação e Expansão:

Realização de testes de consistência e qualidade, além da integração de novas fontes de dados e expansão do sistema para outros municípios ou unidades especializadas. Essa etapa garantirá a consolidação da solução e sua capacidade de evolução contínua.

Ao final das três fases, espera-se disponibilizar uma solução robusta e automatizada de monitoramento da violência de gênero, capaz de apoiar a formulação e avaliação de políticas públicas baseadas em dados confiáveis e atualizados.

## 4. Planejamento Estratégico da Solução (Objetivos e Plano de Ação)

Esta seção apresenta o planejamento estratégico que orienta a execução técnica e operacional do projeto, com ênfase na estruturação do pipeline de dados e nas etapas do processo de ETL (Extract, Transform, Load). Essas etapas viabilizam a coleta, o tratamento e a disponibilização das informações referentes à violência contra a mulher nos municípios de Sorocaba e Votorantim. O plano contempla ainda as metas específicas do projeto, a organização das fases de implementação e as diretrizes necessárias para assegurar a manutenção, a confiabilidade e a expansão futura da solução desenvolvida.

## 4.1 Estrutura Estratégica: Objetivos SMART e Prioridades

O objetivo principal deste projeto é implementar, até novembro de 2025, uma solução automatizada para extração, tratamento, armazenamento e visualização de dados sobre violência contra a mulher nos municípios de Sorocaba e Votorantim, utilizando informações provenientes das Delegacias de Defesa da Mulher (DDM) da Secretaria de Segurança Pública do Estado de São Paulo (SSP-SP).

A seguir, os objetivos são detalhados conforme a metodologia SMART, que orienta o planejamento de metas Específicas, Mensuráveis, Alcançáveis, Relevantes e Temporais:

### **Específico:**

- Desenvolver um pipeline de ETL (Extract, Transform, Load) em Python utilizando Pandas.
- Realizar a extração automática das planilhas da SSP-SP, processando apenas abas e dados das DDM.
- Efetuar o tratamento, padronização e enriquecimento dos dados, removendo inconsistências e criando variáveis derivadas.
- Consolidar os registros válidos em uma base unificada no Google BigQuery.
- Integrar a base tratada a dashboards interativos no Google Looker Studio, permitindo análises dinâmicas e filtragens por período, localidade e tipo de ocorrência.

### **Mensurável:**

- Processar 100% das planilhas e abas relevantes das DDM.
- Garantir execução automática mensal do pipeline.
- Assegurar acurácia mínima de 98% em relação aos dados originais da SSP-SP.
- Disponibilizar todos os indicadores-chave (número de ocorrências, tipo, bairro, período do dia, perfil de vítimas e agressores).

**Alcançável:**

- Utilizar ferramentas acessíveis e integradas (Python, Pandas, Google BigQuery e Looker Studio).
- Executar o pipeline em ambiente Google Colab, permitindo colaboração e testes contínuos.

**Relevante:**

- A solução contribuirá para a gestão pública baseada em evidências, apoiando a Secretaria da Mulher (Semul) e demais órgãos na formulação de políticas de prevenção e combate à violência de gênero.
- Promoverá transparência, acesso público e conscientização social sobre o tema, transformando dados dispersos em informações estratégicas.

**Temporal:**

- Fase 1 (ago–set/2025): Desenvolvimento e testes do pipeline ETL.
- Fase 2 (out/2025): Integração dos dados ao Google BigQuery e criação dos dashboards no Looker Studio.
- Fase 3 (nov/2025): Validação final, ajustes de performance, documentação e entrega da versão funcional da solução.

## 4.2 Mapeamento Estratégico (Balanced Scorecard) e Metodologia PDCA

**Balanced Scorecard (BSC)**

Para garantir uma execução estruturada e orientada a resultados, foi elaborado um Balanced Scorecard (BSC) que traduz a estratégia do projeto em objetivos mensuráveis e interconectados. O modelo permite acompanhar o desempenho do pipeline de ETL e dos dashboards interativos sob diferentes perspectivas.

Essa abordagem possibilita uma visão integrada do progresso do projeto, assegurando que as ações técnicas e operacionais estejam alinhadas aos propósitos institucionais de promover transparência, eficiência e apoio à tomada de decisão no enfrentamento à violência contra a mulher.

Perspectiva	Objetivo Estratégico	Indicadores de Desempenho (KPIs)	Metas	Iniciativas / Ações Principais
<b>Financeira / Recursos</b>	Garantir execução do projeto sem custos financeiros, utilizando ferramentas gratuitas e acessíveis.	Nº de ferramentas pagas utilizadas	0 ferramentas pagas	Utilizar Python, Google Colab, BigQuery (versão gratuita) e Looker Studio.
<b>Processos Internos</b>	Automatizar a coleta, tratamento e integração dos dados das DDM.	% de execução do pipeline	100% do processo ETL concluído e funcional	Desenvolver pipeline ETL em Python e consolidar base no Google BigQuery.
<b>Aprendizado e Desenvolvimento</b>	Aprimorar as habilidades do grupo do Projeto Integrador 4, no âmbito de extração, tratamento e carregamento de dados	Aplicação prática das tecnologias do projeto	Implementação completa até nov/2025	Desenvolvimento colaborativo do pipeline e dos dashboards interativos.
<b>Sociedade / Impacto Social</b>	Promover o acesso público e a compreensão dos dados sobre violência contra a mulher.	Nº de visualizações / consultas ao dashboard	Dashboard disponível publicamente até nov/2025	Compartilhar dashboards e relatórios com gestores públicos de Sorocaba e Votorantim.
<b>Usuário Final / Público-Alvo</b>	Fornecer uma ferramenta confiável e intuitiva para a Secretaria da Mulher e demais órgãos de gestão pública.	Grau de usabilidade e clareza das visualizações	100% dos indicadores com filtros funcionais e linguagem acessível	Criar dashboards no Looker Studio com indicadores sobre tipo de violência, local e perfil das vítimas e agressores.

Tabela 1: Balanced Score Card

O Balanced Scorecard apresentado na Figura 1 sintetiza a estratégia do projeto em cinco perspectivas complementares, garantindo equilíbrio entre viabilidade técnica, operacional e social. A dimensão de recursos assegura a execução sem custos financeiros, enquanto os processos internos priorizam a automatização e confiabilidade do pipeline ETL. A perspectiva de aprendizado evidencia o desenvolvimento técnico da equipe, e as dimensões social e do público-alvo reforçam o compromisso com a transparência, o acesso à informação e a criação de ferramentas interativas e acessíveis para os órgãos públicos e a população. Em conjunto, essas diretrizes fortalecem a efetividade e o impacto do projeto na gestão e análise dos dados sobre violência de gênero.

## **Metodologia PDCA**

Durante a execução do projeto, foram realizadas reuniões semanais com o professor orientador do Projeto Integrador, com o objetivo de obter feedbacks, esclarecer dúvidas técnicas e ajustar as entregas de acordo com o modelo e as diretrizes estabelecidas. Esse processo de acompanhamento constante permitiu manter a coerência entre as etapas de planejamento, desenvolvimento e validação dos resultados.

## **Mecanismos de Acompanhamento**

As etapas de coleta, tratamento e armazenamento dos dados das Delegacias de Defesa da Mulher (DDM) de Sorocaba e Votorantim, foram acompanhadas através de indicadores de desempenho, como tempo de execução, padronização dos dados e confiabilidade das integrações dos dados.

### **Ações Corretivas e Melhoria Contínua:**

Sempre que foram identificadas dificuldades, como falhas de leitura em determinados conjuntos de dados ou lentidão em partes do fluxo, a equipe aplicou ações corretivas imediatas, revisando scripts, ajustando parâmetros e otimizando rotinas de automação. Esse processo constante de revisão e aperfeiçoamento fortaleceu a confiabilidade do pipeline e consolidou o aprendizado prático.

### **Sustentabilidade e Continuidade após o PI4**

Ao término do projeto, recomenda-se que a Secretaria da Mulher incorpore o pipeline e os dashboards ao processo regular de Tecnologia da Informação, designando um responsável técnico pela manutenção e atualização periódica das ferramentas. Dessa forma, o sistema poderá continuar sendo utilizado para monitorar e analisar dados das DDM de maneira padronizada e automatizada, garantindo a sustentabilidade e o impacto duradouro dos resultados alcançados.



## Ciclo PDCA

A Tabela a seguir apresenta a aplicação prática do ciclo PDCA (Planejar, Executar, Verificar e Agir) no contexto do Projeto Integrador, detalhando as principais etapas, atividades realizadas e mecanismos de acompanhamento adotados para garantir uma execução adequada do objetivo.

<b>P - Planejar (Plan)</b>	<ul style="list-style-type: none"><li>• Definição dos objetivos estratégicos e técnicos do projeto;</li><li>• Planejamento das etapas do pipeline de dados e elaboração do cronograma de entregas;</li><li>• Escolha das ferramentas tecnológicas (Python, Pandas, BigQuery e Looker Studio);</li><li>• Estruturação do modelo de dados e definição dos critérios de padronização.</li></ul>
<b>D - Executar (Do)</b>	<ul style="list-style-type: none"><li>• Desenvolvimento e implementação das etapas do ETL (Extração, Transformação e Carga);</li><li>• Construção e integração dos dashboards interativos;</li><li>• Ajustes contínuos baseados nos feedbacks das reuniões semanais de acompanhamento;</li><li>• Testes de automação e verificação de integridade dos dados.</li></ul>
<b>C - Verificar (Check)</b>	<ul style="list-style-type: none"><li>• Monitoramento do desempenho e confiabilidade do pipeline;</li><li>• Avaliação dos indicadores de qualidade e eficiência das execuções;</li><li>• Validação da integridade e consistência dos dados carregados;</li><li>• Comparação dos resultados obtidos com as metas e cronograma estabelecidos.</li></ul>
<b>A - Agir (Act)</b>	<ul style="list-style-type: none"><li>• Aplicação de melhorias contínuas com base nas análises e feedbacks obtidos;</li><li>• Revisão e otimização dos scripts e parâmetros do pipeline;</li><li>• Documentação das boas práticas e lições aprendidas;</li><li>• Preparação para manutenção e evolução futura da solução.</li></ul>

Tabela 2: Ciclo PDCA

O processo favoreceu a integração entre planejamento e execução, consolidando boas práticas de gestão e garantindo que o projeto seja entregue com qualidade, eficiência e potencial de continuidade para uso futuro.

## 5. Metodologia de Desenvolvimento e Fonte de Dados

O pipeline desenvolvido neste projeto tem como finalidade automatizar o fluxo de dados desde a coleta até o armazenamento, garantindo que as informações registradas nas Delegacias de Defesa da Mulher (DDM) de Sorocaba e Votorantim sejam processadas de forma padronizada para futuras análises e visualizações.

O script adotado segue as três etapas clássicas do ETL (Extract, Transform, Load), estruturado em três etapas principais, implementadas em linguagem Python, utilizando a biblioteca Pandas para manipulação dos dados e o Google BigQuery como ambiente de destino para armazenamento.

### 5.1 Pipeline de Dados (Proposta de ETL)

#### **Extração (Extract)**

A etapa de extração obtém os dados brutos diretamente das fontes oficiais da Secretaria de Segurança Pública do Estado de São Paulo (SSP-SP). O pipeline automatiza a coleta das planilhas Excel publicadas pela SSP-SP, contendo os registros de ocorrências criminais e dados de produtividade policial entre 2022 e 2025.

No primeiro script, o sistema realiza o download e leitura automática das planilhas `SPDadosCriminais_2022.xlsx` a `SPDadosCriminais_2025.xlsx`, consolidando todas as abas e registros relevantes em um único `DataFrame`. Já no segundo script, são extraídas as planilhas `DadosProdutividade_2024.xlsx` e `DadosProdutividade_2025.xlsx`, com foco nas abas que contêm os registros de “Presos e Apreendidos”, correspondentes às DDMs de Sorocaba e Votorantim.

Os arquivos são baixados e armazenados temporariamente no ambiente do Google Colab, onde ocorre a consolidação dos dados antes do início do tratamento. Em seguida, os registros válidos são integrados em uma base única, que servirá de entrada para as próximas etapas do pipeline.

Essa estratégia assegura automação e facilidade de atualização, permitindo que o sistema incorpore novos arquivos assim que disponibilizados pela SSP-SP, mantendo o processo reproduzível e escalável.

### **Transformação (Transform)**

A etapa de transformação é responsável por tratar, organizar e padronizar os dados brutos, assegurando qualidade, consistência e conformidade com o modelo de dados do BigQuery.

As principais operações realizadas incluem:

- Filtragem geográfica: mantém apenas registros dos municípios de Sorocaba e Votorantim, limitando-se às Delegacias de Defesa da Mulher (DDM) dessas localidades.
- Limpeza e formatação: exclusão de registros incompletos, conversão das colunas de data para o tipo *datetime* e substituição de valores nulos por “Não Informado”.
- Padronização textual:
  - uniformização dos textos para o formato *Title Case*;
  - adequação dos nomes das colunas para convenções legíveis e consistentes (ex.: NUM\_BO → *codigo\_bo*);
  - tratamento dos nomes dos bairros, com correção de acentuação, remoção de variações incorretas e padronização de abreviações — por exemplo, “Jd.” convertido para “Jardim”.
- Conversão de tipos de dados: transformação de colunas numéricas, como *ano\_ocorrendia*, *mes\_ocorrendia*, *latitude* e *longitude*, para tipos inteiros ou decimais.

- Criação de variáveis derivadas: novas colunas são geradas a partir das informações existentes, tais como:
  - mes\_ocorrendia, extraído da data original;
  - dia\_semana, obtido a partir da data da ocorrência (ex.: “Segunda-feira”).

No ETL de produtividade, além das operações anteriores, são realizadas transformações complementares para descrever o perfil do agressor, incluindo a renomeação e padronização das seguintes colunas:

SEXO\_PESSOA → sexo\_autor

IDADE\_PESSOA → idade\_autor

COR\_CURTIS → raca\_autor

GRAU\_INSTRUCAO → escolaridade\_autor

PROFISSAO → profissao\_autor

Essas variáveis possibilitam uma análise detalhada sobre as características demográficas e sociais dos autores das ocorrências, permitindo futuras correlações com os tipos de crimes registrados nas DDMs.

Ao final dessa etapa, os dados estão estruturados, limpos e consistentes, prontos para serem carregados no ambiente de armazenamento.

### **Carga (Load)**

A etapa de carga é responsável por transferir os dados transformados para o ambiente de armazenamento definitivo, garantindo sua disponibilidade para consultas, cruzamentos e visualizações.

No projeto, essa fase é implementada pela função `carregar_dados_bigquery`, que recebe o DataFrame final e realiza a inserção no Google BigQuery, dentro do projeto `projetointegrador4-473718`.

As principais características da carga são:

Validação e consistência: antes do envio, o DataFrame é verificado quanto à integridade e compatibilidade de tipos com o schema da tabela de destino.

Especificação do destino: os dados são carregados no dataset dados\_ssp, sendo armazenados nas tabelas:

dados\_ddm – contendo os registros das ocorrências criminais das Delegacias de Defesa da Mulher (DDM) de Sorocaba e Votorantim;

dados\_produtividade – contendo os dados referentes ao perfil dos agressores.

Atualização da base: o modo de escrita WRITE\_TRUNCATE, é utilizado para substituir a tabela anterior a cada nova execução para manter a base sempre atualizada.

Integração com visualização: após o carregamento, os dados são disponibilizados automaticamente para o Google Looker Studio, possibilitando a construção de dashboards interativos e relatórios analíticos.

Além disso, a tabela dados\_ssp.dados\_produtividade serve como base para a criação da view derivada dados\_ssp.perfil\_agressor, projetada para facilitar as consultas e análises relacionadas às características demográficas e sociais dos autores das ocorrências.

Essa abordagem assegura que o pipeline seja automatizado, confiável e reproduzível, permitindo que novas execuções atualizem os dados de forma consistente e sem necessidade de intervenção manual.

## 5.2 Análise Exploratória e Compreensão dos Dados

Após o tratamento e a padronização das bases de dados, foi conduzida uma Análise Exploratória de Dados (Exploratory Data Analysis – EDA) com o objetivo de compreender a distribuição das variáveis, identificar padrões relevantes e avaliar a qualidade das informações.

Durante o EDA, foram aplicadas diferentes abordagens analíticas, incluindo:

- **Análise de distribuições**
  - Cálculo de medidas de tendência central (média) e proporções para variáveis como idade, sexo, raça, escolaridade e tipo de prisão.
  - Identificação de valores ausentes e categorias com subnotificação (ex.: “Não informado”).
- **Segmentações por município e bairro**
  - Comparação entre Sorocaba e Votorantim, destacando diferenças e similaridades nos perfis dos envolvidos.
  - Mapeamento dos bairros com maior volume de ocorrências, considerando densidade populacional e zonas de concentração.
- **Análises temporais**
  - Avaliação da distribuição das ocorrências ao longo do tempo (anos, meses e horários).
  - Identificação de padrões sazonais ou recorrentes.
- **Relações entre variáveis**
  - Comparação entre tipo de crime, circunstância da prisão, perfil demográfico do agressor e perfil da vítima.
  - Cruzamento entre local da ocorrência e tipo de violência.

- **Principais Insights Identificados**

**Insight 1 – Sorocaba concentra mais ocorrências, especialmente em bairros populosos da Zona Norte**

A análise territorial mostrou que Sorocaba apresenta maior volume geral de denúncias, o que é esperado pela diferença populacional. O EDA evidencia que a Zona Norte concentra a maior parte dos casos, especialmente no bairro Parque São Bento, indicando áreas críticas para intervenção.

## **Insight 2 – A maior parte das ocorrências acontece dentro da residência e envolve violência física**

O EDA confirma que:

- Lesão Corporal Dolosa é o tipo de violência mais registrado em ambos os municípios.
- O domicílio é o principal local das ocorrências, reforçando o padrão característico da violência doméstica.

## **Insight 3 – O perfil do agressor é majoritariamente masculino, adulto jovem**

Com base nos painéis:

- Média de idade próxima a 34 anos.
- Predominância de autuações em flagrante.
- Predomínio absoluto de agressores do sexo masculino.
- Alta frequência de escolaridade até o 2º grau completo
- Profissões variadas, mas grande parcela não informada ou associada a ocupações informais.

## **Insight 4 – Perfil da vítima e vulnerabilidade:**

A análise da base SINAN mostrou:

- Média de idade em torno de 26 anos.
- A maioria das vítimas são mulheres jovens.
- Raças branca e parda aparecem com maior frequência.
- Mais de 1.700 registros sem informação de escolaridade, reforçando a limitação da base.
- Percentual de gestantes relevante (cerca de 6%), indicando risco ampliado e necessidade de atenção específica.

### **Implicações para o projeto:**

A Análise Exploratória de Dados forneceu insights essenciais sobre a base, enriquecendo a interpretação e a contextualização dos indicadores presentes no dashboard.

### **Estruturação de filtros por município e bairro**

A segmentação territorial observada na EDA orientou a criação de filtros geográficos, permitindo análises locais mais precisas e apoiando a formulação de políticas públicas direcionadas.

### **Tratamento e destaque para dados faltantes**

A elevada proporção de campos incompletos levou à inclusão de alertas e da categoria “Não informado”, garantindo transparência e evitando interpretações equivocadas sobre a base.

### **Reforço da importância do monitoramento contínuo**

As variações temporais identificadas, embora não lineares, evidenciam a necessidade de acompanhamento constante, fator que justifica a automação dos dashboards e a atualização recorrente.



## 6. Desenvolvimento do Protótipo e Implementação da Solução

O desenvolvimento da solução concentrou-se na automatização da coleta, tratamento e visualização de dados das Delegacias de Defesa da Mulher (DDM) de Sorocaba e Votorantim. A arquitetura integrou Python e Pandas para o processamento e tratamento das informações, Google BigQuery para o armazenamento em nuvem e Google Looker Studio para a visualização interativa dos resultados.

### 6.1 Arquitetura do Sistema Proposto

A arquitetura do sistema foi estruturada em três principais esferas:

- **Extração e Transformação:** implementada em Python, utilizando a biblioteca Pandas para leitura e limpeza dos dados brutos fornecidos pela Secretaria de Segurança Pública (SSP-SP).
- **Armazenamento:** os dados tratados foram enviados e organizados no Google BigQuery, otimizando consultas e garantindo escalabilidade para análises futuras.
- **Visualização:** construída no Google Looker Studio, com painéis dinâmicos e filtros interativos que permitem explorar os dados de forma intuitiva.

#### **Diagrama de arquitetura:**

O diagrama a seguir apresenta a arquitetura geral do processo ETL implementado, descrevendo o fluxo completo desde a extração automatizada das planilhas até a transformação, carga e visualização dos dados

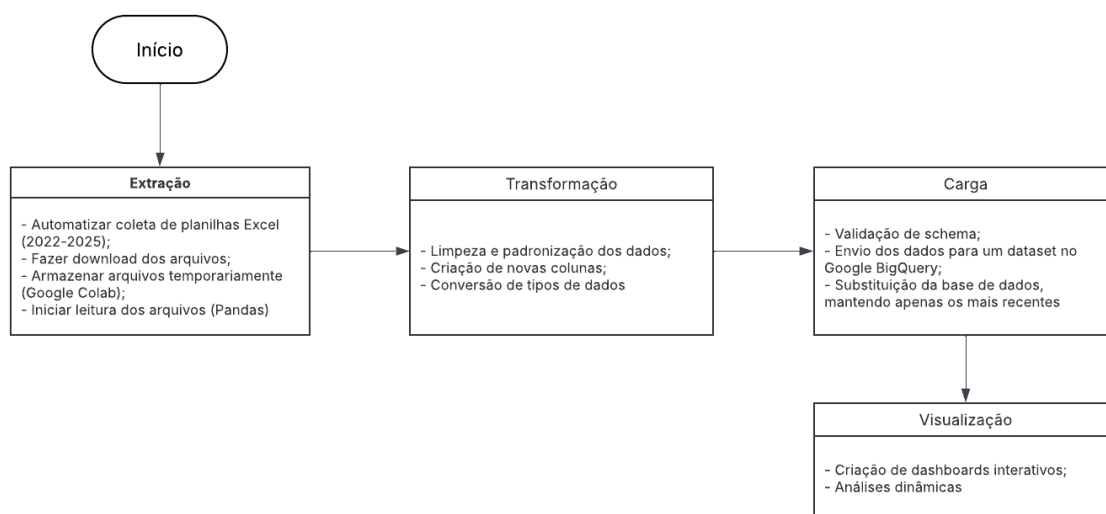


Figura 1: Diagrama de Arquitetura

Conforme ilustrado na Figura 1, o sistema foi estruturado segundo um fluxo ETL integrado à visualização de dados, automatizando o processamento das informações das Delegacias de Defesa da Mulher (DDM) de Sorocaba e Votorantim.

Na etapa de Extração, as planilhas são coletadas e consolidadas no Google Colab com linguagem Python através da biblioteca Pandas. A Transformação aplica limpeza, padronização e criação de colunas derivadas, garantindo consistência. Durante a Carga, os dados são enviados ao Google BigQuery mantendo apenas a versão mais recente. A Visualização ocorre no Google Looker Studio, permitindo dashboards interativos para análise das ocorrências. A arquitetura em nuvem assegura portabilidade, escalabilidade e reprodutibilidade, com possibilidade de futuras expansões.

## 6.2 Funcionalidades Implementadas

O protótipo desenvolvido automatiza o ciclo de coleta, tratamento, armazenamento e visualização dos dados sobre violência contra a mulher, tornando as informações mais acessíveis e confiáveis. Cada funcionalidade foi

criada para atender aos objetivos SMART do projeto, apoiando a análise e a tomada de decisão baseada em dados.

### **Funcionalidade 1 – Extração Automática de Dados**

O sistema realiza automaticamente o download e a leitura das planilhas disponibilizadas pela SSP-SP, filtrando apenas as abas referentes às Delegacias de Defesa da Mulher (DDM) de Sorocaba e Votorantim. Objetivo atendido: Automatizar a coleta dos dados e eliminar processos manuais, garantindo atualizações mensais conforme o item “Mensurável” do objetivo SMART.

### **Funcionalidade 2 – Tratamento e Padronização dos Dados**

Os dados extraídos passam por limpeza e padronização no Python com a biblioteca Pandas, incluindo a remoção de inconsistências, normalização de colunas e criação de variáveis derivadas, como data, tipo de violência e perfil das vítimas.

Objetivo atendido: Garantir qualidade e consistência dos dados, atendendo ao critério “Específico” e assegurando acurácia mínima de 98%.

### **Funcionalidade 3 – Armazenamento Centralizado no BigQuery**

Após o tratamento, os dados são consolidados em uma base única no Google BigQuery, permitindo consultas rápidas e integração direta com ferramentas de visualização. Objetivo atendido: Centralizar e otimizar o acesso às informações, atendendo ao requisito “Alcançável” e “Temporal” do plano de execução.

### **Funcionalidade 4 – Integração com Looker Studio (Dashboard Interativo)**

A solução disponibiliza dashboards interativos que apresentam indicadores como número de ocorrências, tipos de violência, bairros e perfis das vítimas e agressores, com filtros dinâmicos por período e localidade. Objetivo atendido: Transformar os dados em informações estratégicas, apoiando a gestão pública e a formulação de políticas baseadas em evidências).

## **Funcionalidade 5 – Execução Automatizada e Reprodutível**

O pipeline foi estruturado para execução periódica e pode ser adaptado para incluir novos municípios ou fontes de dados no futuro.

Objetivo atendido: Garantir a sustentabilidade e expansão da solução a longo prazo, conforme a meta de continuidade e impacto duradouro do projeto.).

O protótipo entrega um pipeline completo e funcional de ETL automatizado, integrando Python, BigQuery e Looker Studio, com dashboards interativos que transformam dados da SSP-SP em insights práticos para políticas públicas.

O relatório pode ser acessado através [deste hiperlink](https://lookerstudio.google.com/s/p4VwL9Uz6Dk) (https://lookerstudio.google.com/s/p4VwL9Uz6Dk) ou visualizado a partir das figuras a seguir.

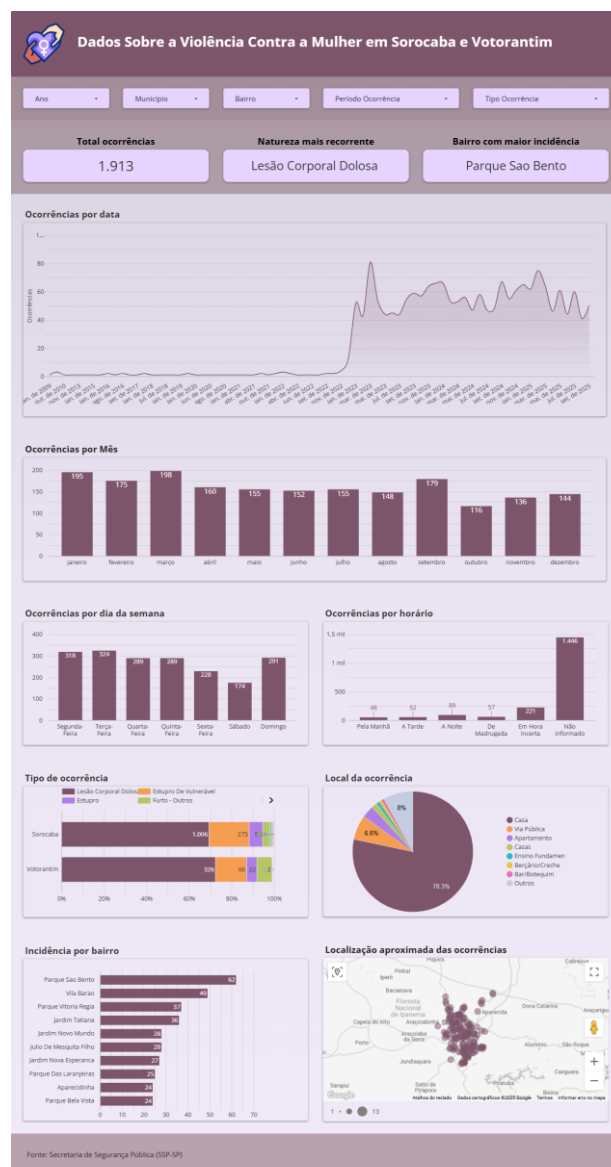


Figura 2: Dashboard – Análise de Ocorrências de Violência Contra a Mulher - Página 1 - Ocorrências DDM

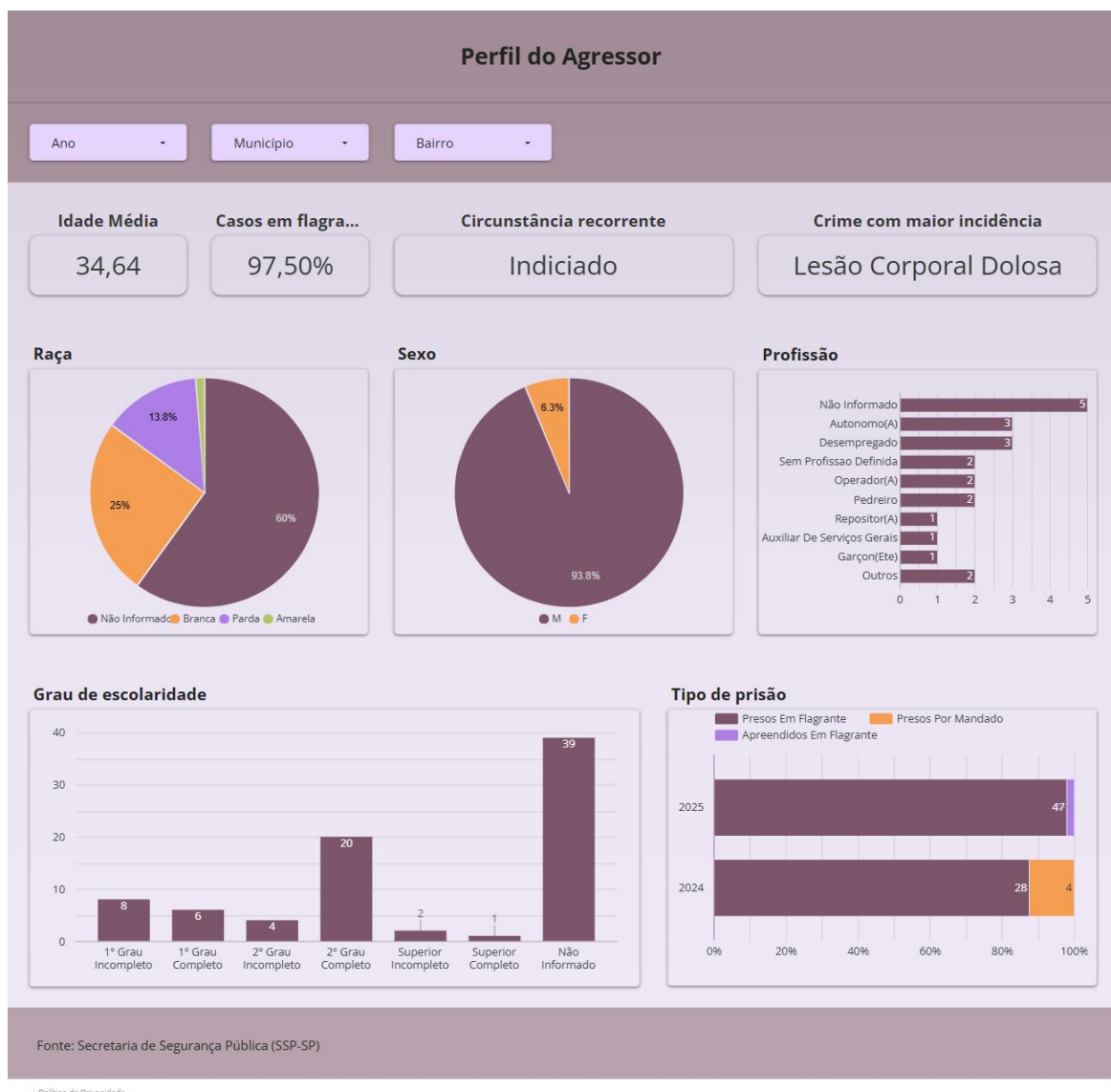


Figura 3: Dashboard – Análise de Ocorrências de Violência Contra a Mulher - Página 2 - Perfil Agressor

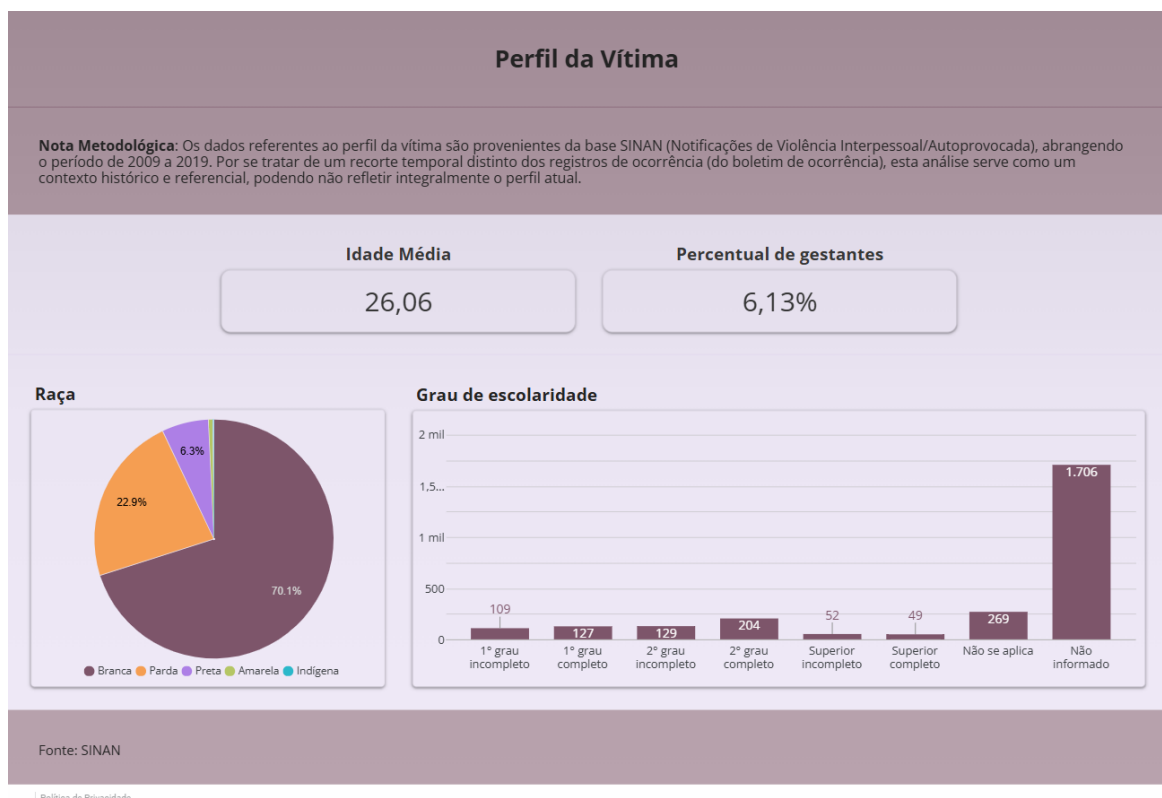


Figura 4: Dashboard – Análise de Ocorrências de Violência Contra a Mulher - Página 2 - Perfil da Vítima

### 6.3 Fluxo do Usuário e Experiência (UX/UI)

A interface do projeto consiste em um dashboard interativo desenvolvido no Google Looker Studio, projetado para facilitar a exploração e interpretação dos dados das Delegacias de Defesa da Mulher (DDM) de Sorocaba e Votorantim. O dashboard permite o acompanhamento visual e dinâmico de indicadores, consolidando informações de forma organizada e acessível. A solução se conecta diretamente aos dados armazenados no Google BigQuery, garantindo que as visualizações reflitam sempre as informações mais recentes e mantenham consistência com o pipeline de extração, limpeza e carga.

- **Fluxo de navegação:** O acesso ao dashboard é feito por meio de um link compartilhado no Google Looker Studio, o que elimina a necessidade de instalação local e permite sua visualização em diferentes dispositivos. Além disso, a plataforma é totalmente online e gratuita, não exigindo assinatura para uso.

Ao abrir o painel, o usuário visualiza uma página inicial com filtros interativos como ano, município, bairro, período de ocorrência e tipo de ocorrência que permitem personalizar a visualização conforme o interesse da análise.

- **Jornadas de usuário:**

O gráfico “Ocorrências por data” possibilita observar variações sazonais e períodos de aumento significativo, subsidiando análises sobre o impacto de políticas públicas, campanhas de conscientização ou eventos sociais que possam influenciar o comportamento das ocorrências.

Aplicando o filtro “Município: Sorocaba”, a equipe examina o painel “Incidência por bairro”, identificando o Vila Barão como a região com maior número de ocorrências. Com base nesse mapeamento, são definidas ações prioritárias de prevenção e suporte, incluindo a expansão de serviços especializados, a realização de campanhas educativas e a implementação de programas de sensibilização dirigidos à comunidade.

Por meio dos gráficos “Ocorrências por dia da semana” e “Ocorrências por horário”, a Secretaria pode identificar padrões temporais de maior incidência, subsidiando o planejamento estratégico das ações de resposta e prevenção. Essas informações permitem o dimensionamento adequado das equipes de plantão policial e atendimento emergencial, bem como a definição de campanhas de comunicação e conscientização voltadas aos períodos de maior vulnerabilidade observados.

Um analista de segurança pública pode cruzar informações sobre o perfil do agressor e da vítima a fim de identificar padrões e fatores associados à reincidência, subsidiando a definição de estratégias preventivas e de intervenção mais eficazes.

- **Design e usabilidade:** A paleta de cores foi composta por tons de roxo e lilás, que são tradicionalmente associados à luta pelos direitos das mulheres e frequentemente utilizados em campanhas de enfrentamento à violência de gênero.



## **Detalhamento dos gráficos**

### **Página 1: Ocorrências DDM**

Filtros de Navegação - painel disponibiliza filtros dinâmicos que permitem segmentar e refinar as informações apresentadas. Os filtros implementados foram: Ano, Município, Bairro, Período da Ocorrência e Tipo da Ocorrência. Os filtros permitem ao usuário personalizar a visualização para análises direcionadas, atualizando automaticamente os gráficos de forma consistente. A navegação é intuitiva, guiando o usuário de indicadores gerais para análises mais específicas (como temporais ou territoriais), facilitando a compreensão do panorama geral.

Indicadores Sintéticos (cards informativos) - Apresenta, de maneira direta, informações resumidas, como o total de ocorrências, o tipo de ocorrência mais registrado e o bairro com maior incidência, fornecendo ao usuário um ponto de partida para a análise detalhada dos demais componentes do dashboard.

Gráfico de Série Temporal: Ocorrências por Data - O gráfico de linha é utilizado para acompanhar a evolução de dados ao longo do tempo, permitindo identificar tendências, oscilações e períodos de maior ou menor incidência. Essa visualização apresenta a quantidade diária de ocorrências registradas nas DDMs, possibilitando observar o comportamento temporal dos registros ao longo do período analisado.

Gráfico de Barras Verticais: Ocorrências por Mês - Esse tipo de gráfico é adequado para comparar categorias discretas, permitindo identificar diferenças claras entre os dias da semana. Essa visualização permite identificar o comportamento mensal dos registros de violência atendidos pelas Delegacias de Defesa da Mulher, evidenciando padrões de incidência e possíveis variações sazonais ao longo do ano.

Gráfico de Barras Verticais: Ocorrências por Horário - O gráfico de barras é apropriado para mostrar concentrações dentro de faixas horárias. Ele apresenta a distribuição das ocorrências registradas ao longo do dia, destacando horários com maior incidência e permitindo compreender a concentração dos registros dentro do ciclo diário.

Gráfico de Barras Empilhadas: Tipo de Ocorrência - O gráfico de barras empilhadas permite comparar categorias entre diferentes grupos. No dashboard, ele apresenta a distribuição dos tipos de ocorrências registradas em Sorocaba e Votorantim, permitindo visualizar diferenças na composição dos registros entre os municípios.

Gráfico de Pizza: Local da Ocorrência – Esse tipo de visualização é utilizado para representar proporções de categorias dentro de um todo. No relatório, ele mostra a proporção dos locais informados nos registros de ocorrência, permitindo identificar os ambientes mais recorrentes entre as ocorrências registradas.

Gráfico de Barras Horizontais: Ocorrências por Bairro - Esse tipo de gráfico é adequado para categorias com textos longos, como nomes é o caso. No painel, ele apresenta a distribuição das ocorrências registradas entre os bairros, permitindo identificar regiões com maior concentração de casos.

Mapa Georreferenciado: Localização Aproximada das Ocorrências – Esses mapas são utilizados para representar dados distribuídos no território, permitindo a identificação de padrões espaciais. Na visualização, são exibidas as localizações aproximadas das ocorrências registradas nas DDMs, destacando áreas com maior concentração por meio de “bolhas” que variam conforme a densidade dos registros.

## **Página 2: Perfil do Agressor**

Filtros – Na segunda página também há filtros dinâmicos, como Ano, Município e Bairro, que permitem segmentar os dados sobre o perfil dos agressores nas DDMs de Sorocaba e Votorantim. Assim como na primeira página, esses filtros atualizam automaticamente todos os gráficos, mantendo uma navegação intuitiva que leva o usuário dos indicadores gerais a análises mais detalhadas.

Indicadores Sintéticos: (cards informativos) - Os cards fornecem uma visão inicial e objetiva sobre características centrais dos agressores registrados, apresentando informações como idade média, percentual de casos em flagrante, circunstância mais recorrente e crime com maior incidência. Esses indicadores de

síntese facilitam a compreensão imediata do perfil predominante, preparando o usuário para a análise mais detalhada dos demais gráficos da aba.

**Gráfico de Pizza: Raça** - O gráfico de pizza é ideal para representar a proporção de cada categoria dentro de um conjunto. Nesta visualização, ele mostra a distribuição dos agressores por raça/cor registrada nas ocorrências, permitindo identificar rapidamente o perfil predominante e comparar a participação relativa de cada grupo.

**Gráfico de Pizza: Sexo** - Facilita a leitura da proporção de cada categoria dentro do total de registros. Neste painel, apresenta a distribuição do sexo dos agressores registrados, evidenciando a predominância das categorias informadas nos boletins de ocorrência.

**Gráfico de Barras Horizontais: Profissão** - A utilização de barras horizontais favorece a leitura de categorias mais extensas, como descrições de ocupações. No painel, essa visualização mostra a distribuição das profissões informadas pelos agressores nos registros de ocorrência, permitindo identificar quais atividades profissionais aparecem com maior frequência.

**Gráfico de Barras Verticais: Grau de Escolaridade** - O gráfico de barras é apropriado para comparar diferentes níveis de escolaridade entre grupos. No painel, exibe a quantidade de agressores registrada para cada nível educacional informado, permitindo observar a distribuição educacional dentro do conjunto de ocorrências.

**Gráfico de Barras Empilhadas: Tipo de Prisão** - O gráfico de barras empilhadas permite comparar categorias sobrepostas dentro de um mesmo período. A visualização apresenta a distribuição dos tipos de prisão registrados (presos em flagrante, presos por mandado e apreendidos em flagrante), permitindo observar a composição das modalidades de prisão por ano e compreender a prevalência de cada tipo entre os casos analisados.

### **Página 3: Perfil da Vítima**

**Indicadores Sintéticos (cards informativos)** - Os cards apresentam, de forma objetiva, informações essenciais sobre o perfil das vítimas, como a idade

média e o percentual de gestantes registrados. Eles oferecem um panorama inicial das principais características observadas, facilitando a contextualização antes da exploração dos gráficos mais detalhados.

**Gráfico de Pizza: Raça** - A visualização em formato de pizza facilita a compreensão das proporções entre categorias que compõem um conjunto total. No painel, o gráfico apresenta a distribuição das vítimas conforme a raça/cor informada nos registros de ocorrência, permitindo identificar com clareza a representatividade percentual de cada grupo declarado.

**Gráfico de Barras Verticais: Grau de Escolaridade** - A disposição em barras verticais permite comparar rapidamente categorias educacionais distintas, destacando variações entre os grupos. No painel, o gráfico apresenta a quantidade de vítimas por nível de escolaridade informado nos registros de ocorrência, possibilitando identificar quais níveis são mais frequentes dentro da base de dados.

#### **6.4 Metodologia de Desenvolvimento e Ferramentas Utilizadas**

O desenvolvimento do projeto ocorreu de forma colaborativa, sustentado por comunicação contínua entre os integrantes e por reuniões semanais dedicadas à revisão das etapas concluídas. Mesmo sem o uso de ferramentas formais de gestão, o grupo manteve um fluxo organizado por meio da divisão clara de responsabilidades, da validação periódica das atividades e do uso de plataformas digitais que facilitaram a coordenação do trabalho.

O GitHub foi adotado como repositório oficial para versionamento e armazenamento dos materiais técnicos, reunindo códigos, notebooks e documentos produzidos ao longo do projeto. As contribuições foram registradas por meio de *commits* individuais, permitindo acompanhar a evolução das entregas.

Para orientar a navegação, foi elaborado um arquivo README.md contendo o objetivo do projeto, a estrutura das pastas, as dependências necessárias e as instruções de execução do pipeline e acesso ao dashboard. Esse documento atuou como guia técnico inicial e reforçou a organização da solução. Durante o processo, surgiram limitações relacionadas à atualização de arquivos

Word diretamente no GitHub. Para contornar essa restrição, optou-se pelo uso de um documento Word online, que permitiu edição simultânea e maior estabilidade no registro do relatório. Além disso, ferramentas colaborativas foram essenciais ao fluxo técnico: o Google Colab viabilizou o desenvolvimento compartilhado dos notebooks de tratamento e análise de dados, enquanto o Looker Studio, também em ambiente online, foi utilizado para construção do dashboard interativo, facilitando acesso, atualização e compartilhamento das visualizações.

As reuniões semanais tiveram papel central na organização do processo, permitindo acompanhar o andamento das tarefas, revisar cada etapa do pipeline e garantir alinhamento entre indicadores e objetivos do projeto. Para a elaboração dos materiais de apresentação, utilizou-se o Canva, que possibilitou a criação de slides padronizados e favoreceu a colaboração visual entre os membros do grupo.

## 6.5 Desafios e Ajustes durante o Desenvolvimento

O desenvolvimento do projeto não ocorreu de forma linear e apresentou desafios que exigiram adaptação, reorganização das etapas e tomada de decisão por parte da equipe.

**Ajustes Decorrentes de Feedback:** As reuniões semanais revelaram pontos de melhoria nas visualizações e indicadores, evidenciando a necessidade de readequar algumas informações, como o perfil da vítima. Por se tratar de uma base de dados distinta das ocorrências, identificou-se a importância de incluir uma nota metodológica para contextualizar o uso desses dados, esclarecendo que seu período e origem diferem dos boletins de ocorrência utilizados no restante do projeto. Esse ajuste foi incorporado juntamente com a separação do dashboard por abas, para evitar interpretações tendenciosas e garantir maior transparência

**Limitação de Contato:** Durante o desenvolvimento do projeto, a equipe também enfrentou dificuldades para estabelecer contato com a representante institucional, responsável pelo alinhamento inicial e pela validação de algumas etapas. A indisponibilidade de retorno em determinados momentos prejudicou o

esclarecimento de dúvidas e atrasou ajustes esperados no escopo. Diante disso, o grupo reorganizou seu planejamento, avançando com as etapas possíveis e tomando decisões baseadas nas diretrizes já fornecidas, de forma a não comprometer o cronograma geral.

**Definição dos Insights Relevantes:** Outro desafio esteve na escolha dos insights que seriam apresentados no dashboard. A equipe precisou analisar quais indicadores realmente contribuíam para responder aos objetivos do projeto, evitando excesso de informações ou visualizações redundantes. A seleção dos dados mais significativos demandou revisões sucessivas, reestruturação das categorias e ajustes nas representações gráficas.

**Coleta de Dados:** A equipe enfrentou dificuldades iniciais pela falta de bases completas e consolidadas, o que dificultou inclusive a identificação de informações essenciais, como o município de ocorrência e elementos do perfil do agressor. Isso exigiu ampliar a busca em diferentes fontes públicas, validar a confiabilidade dos dados encontrados e selecionar apenas os conjuntos compatíveis com os objetivos do projeto.

Apesar dos desafios técnicos e organizacionais, a equipe manteve o trabalho alinhado aos objetivos, adaptando-se continuamente e tomando decisões fundamentadas. Esse processo fortaleceu o domínio das ferramentas, aprimorou as práticas de análise de dados e consolidou a colaboração entre os integrantes, resultando em uma solução estruturada e coerente com as exigências do projeto.

## 7. Testes e Validações

Para garantir o correto funcionamento da solução desenvolvida, foram realizados diferentes tipos de testes ao longo do desenvolvimento do pipeline e da construção do dashboard. Esses testes tiveram como objetivo verificar se as funcionalidades implementadas atendiam ao planejado, identificar possíveis falhas e assegurar estabilidade, coerência e confiabilidade dos resultados apresentados.

## 7.1 Testes Funcionais e Técnicos

Testes de Integração do Pipeline: Para verificar a integridade e funcionamento do pipeline, foram realizados testes de integração completos, garantindo que todo o fluxo, desde a leitura dos dados tratados no Google Colab até a atualização das tabelas no BigQuery, ocorresse sem interrupções. Esses testes asseguraram a consistência dos dados exportados e confirmaram que o Looker Studio integrava corretamente as fontes de dados. Durante o processo, foi identificado um problema causado por caracteres especiais, que gerava duplicava determinadas linhas. A correção foi implementada por meio da inclusão de uma etapa de normalização e padronização textual no tratamento dos dados, eliminando os erros e permitindo a execução estável do pipeline.

Testes Funcionais no Dashboard: Após a integração da fonte de dados com o Looker Studio, foram realizados testes funcionais diretamente no dashboard para verificar a estabilidade das visualizações e a confiabilidade dos resultados apresentados. Entre os principais testes funcionais realizados, destacam-se:

- Verificação de resposta dos gráficos ao aplicar combinações diferentes de filtros (ano, município, bairro, período da ocorrência e tipo de ocorrência);
- Conferência da coerência dos valores exibidos após cada filtragem;
- Testes de comportamento de cada aba para garantir que todas as visualizações carregassem corretamente sem erros.

Durante esses testes, identificou-se que algumas visualizações não respondiam corretamente quando determinados filtros eram combinados, impedindo o cruzamento adequado das informações. Essa situação foi revisada e ajustada, com correções na configuração dos filtros e na relação dos campos utilizados, garantindo o funcionamento consistente das visualizações.

Testes de Consistência dos Indicadores: Os indicadores calculados nos notebooks foram validados por meio de comparações manuais com amostras da base de dados, assegurando que os valores apresentados nos cartões informativos do dashboard refletiam corretamente as regras definidas. Esse procedimento

incluiu a conferência dos totais de ocorrências e a verificação pontual de categorias específicas. Esses testes garantiram que as transformações estavam sendo reproduzidas de forma fiel nas visualizações do Dashboard.

## 7.2 Testes com Usuários e Stakeholders

A validação da solução foi realizada internamente pela equipe, com apoio do professor orientador, por meio de testes que avaliaram a navegação, o funcionamento dos filtros e a consistência dos indicadores em relação ao Dashboard. Como não houve continuidade na participação da empresa parceira ou usuários externos, os próprios membros da equipe assumiram o papel de usuários para testar a navegação, a aplicabilidade dos filtros e a coerência das visualizações. Durante esse processo, foram identificados ajustes necessários, como a correção de filtros que não se cruzavam e a reorganização de categorias.

Nas discussões com o professor orientador, a equipe identificou a necessidade de incluir uma nota metodológica para contextualizar o uso dos dados do SINAN no perfil da vítima, já que essa base difere dos boletins de ocorrência utilizados no restante da análise. Além disso, a equipe identificou a importância de reorganizar o dashboard em abas temáticas para melhorar a clareza e a usabilidade. Após esses ajustes, o dashboard e o pipeline apresentaram estabilidade e coerência, resultando em uma solução funcional e alinhada aos objetivos do projeto.

## 7.3 Indicadores de Sucesso e Próximos Passos

A avaliação final do projeto demonstrou que os principais objetivos foram atingidos. O pipeline desenvolvido mostrou-se funcional e capaz de tratar, padronizar e disponibilizar os dados de forma consistente, garantindo que o



dashboard no Looker Studio recebesse as atualizações corretamente. As visualizações apresentaram estabilidade, coerência e alinhamento com o propósito analítico definido, assegurando que os usuários possam explorar filtros e indicadores com clareza e precisão. O pleno funcionamento do fluxo de dados e a qualidade das análises produzidas configuram os principais indicadores de sucesso alcançados pelo grupo.

Apesar dos resultados positivos, algumas metas não foram totalmente atingidas. A ausência de uma empresa parceira ao longo do projeto limitou a possibilidade de validação com usuários reais do contexto institucional. Além disso, determinadas análises, como o perfil da vítima que tinha períodos e estruturas distintas, o que restringiu a comparabilidade direta entre conjuntos de dados. Esses fatores representam oportunidades de melhoria para versões futuras da solução.

Também é recomendável documentar e registrar rotinas de manutenção contínua para assegurar a longevidade da solução ao longo do tempo.

Em perspectiva de longo prazo, a continuidade desse projeto pode fortalecer a cultura de tomada de decisão orientada por dados, contribuindo para que gestores e equipes técnicas tenham acesso a informações mais estruturadas e precisas. A evolução da ferramenta pode ampliar seu impacto social, apoiando políticas públicas e ações estratégicas voltadas à prevenção e ao enfrentamento da violência contra a mulher.

## 8. Conclusão

Este Projeto Integrador teve como propósito desenvolver uma solução automatizada para coleta, tratamento, armazenamento e visualização dos dados referentes às ocorrências registradas nas Delegacias de Defesa da Mulher (DDM) de Sorocaba e Votorantim. A iniciativa buscou transformar dados públicos dispersos em um painel analítico estruturado, capaz de apoiar análises consistentes e facilitar a interpretação das informações.

Ao longo do desenvolvimento, foi implementado um pipeline de ETL em Python integrado ao Google BigQuery, garantindo padronização, qualidade e

atualização contínua dos dados. A partir desse fluxo, foram construídos dashboards interativos no Looker Studio, possibilitando a exploração visual de indicadores temporais, territoriais e demográficos. Os resultados demonstraram que a solução atende ao objetivo central do projeto ao disponibilizar uma ferramenta acessível, organizada e alinhada às boas práticas de análise de dados.

O projeto também promoveu o aprimoramento das habilidades dos integrantes, especialmente no tratamento de dados, na construção de visualizações e no uso de ferramentas amplamente aplicadas no mercado. Como continuidade, recomenda-se a ampliação das fontes de dados e o aprofundamento das análises, acompanhados de validação técnica e institucional, fatores que contribuirão para o fortalecimento e a evolução da solução desenvolvida.

Conclui-se, portanto, que o projeto cumpriu os objetivos propostos, entregando uma solução coerente, funcional e tecnicamente estruturada, contribuindo para o avanço do conhecimento e para o desenvolvimento das competências exigidas no âmbito do Projeto Integrador.

## 9. Referências

**Base dos Dados - Conjuntos de dados abertos.** Disponível em: <<https://basedosdados.org/dataset/f51134c2-5ab9-4bbc-882f-f1034603147a?table=7d2d29c1-cfa8-470b-ab13-a83cf522c5ac>>. Acesso em: 12 out. 2025.

**LIMA, G. C. C.; PASSOS, C. M.; PINHEIRO, A. L. S.; RIBEIRO, I. J. S.; MAIA, E. G.** Temporal trend and epidemiological profile of notifications of violence against women in Brazil: 2014–2023. **Epidemiologia e Serviços de Saúde, Brasília, v. 34, 2025.** DOI: 10.1590/S2237-96222025v34e20240475.

**Portal SSP.** Disponível em: <<https://www.ssp.sp.gov.br/estatistica/consultas>>. Acesso em: ago. 2025.

**RASEAM 2025 Relatório Anual Socioeconômico da Mulher.** : [2025]. Disponível em: <<https://www.gov.br/mulheres/pt-br/central-de-conteudos/publicacoes/raseam-2025.pdf>>. Acesso em: 15 nov. 2025.

VASCONCELOS, Naina Machado de; BARNABÉ, Regina Tomie Ivata; SANTOS, Juliana Botene dos; FONSECA, Polyanna Hudson Coelho; LIMA, Carolina Silva; VARGAS, Caroline dos; MURPHY, Joseph; CAVALHEIRO, Deborah Carvalho Malta. *Subnotificação de violência contra as mulheres: uma análise de duas fontes de dados.* **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 29, n. 7, p. 2437–2447, 2024. DOI: 10.1590/1413-81232024297.07872023

**Violência contra as mulheres.** Disponível em: <<https://www.paho.org/pt/topics/violence-against-women>>. Acesso em: 9 nov. 2025.