

SUMÁRIO

EXECUTIVO

Planejamento e desenho do Sistema de Dados

Amanda Pavanelli
Dirceu André Gerardi
Rafaela Selem Moreira

2025

INTRODUÇÃO

O Observatório do Femicídio do Estado do Rio de Janeiro (OFRJ) é uma iniciativa estratégica criada para monitorar, analisar e divulgar dados sobre feminicídios. A proposta nasce da constatação de que, embora haja diversas bases e registros sobre violência contra a mulher, esses dados estão dispersos, possuem diferentes formatos e metodologias e, muitas vezes, não permitem recortes detalhados por município ou região.

A Fase I do projeto, realizada entre maio e julho de 2025, teve como objetivo principal **planejar e estruturar o desenho do sistema do Portal de Dados** do Observatório.

Este trabalho incluiu o mapeamento de fontes de dados, a definição de critérios de priorização de bases, a elaboração de mapeamentos técnicos e a produção de protótipos que pudessem demonstrar o potencial analítico do Observatório.



METODOLOGIA

A metodologia adotada baseou-se no modelo CRISP-DM (*Cross Industry Standard Process for Data Mining*), frequentemente utilizado em projetos de Ciência de Dados. Nesta etapa inicial, a equipe concentrou esforços nas fases de entendimento do projeto e de análise das fontes de dados públicos.

O trabalho incluiu, entre outras ações:

- ✓ **Mapeamento e análise de bases públicas;**
- ✓ **Criação e validação de critérios técnicos para classificar a relevância dessas bases;**
- ✓ **Criação de um formulário para levantamento dos sistemas utilizados internamente pelas Secretarias;**
- ✓ **Análise das reuniões dos Grupos de Trabalho, a fim de identificar oportunidades e limitações na coleta e integração de dados;**
- ✓ **Desenvolvimento de um modelo conceitual para a produção e o tratamento dos dados; e**
- ✓ **Elaboração de protótipos visuais, como *dashboards* e organogramas, destinados a orientar as próximas fases do projeto.**

O trabalho começou com o mapeamento das bases nos eixos **Executivo** (Segurança Pública, Saúde, Assistência social e Educação) e **Judiciário**. Com o objetivo de padronizar o levantamento das informações, desenvolvemos uma planilha de mapeamento de dados contendo campos relacionados a formato, origem, acesso, atualização, variáveis disponíveis, desafios de integração e outros aspectos, conforme apresentamos a seguir:

- Nome do Banco de Dados
- Formato dos dados (painel, banco de dados...)
- É uma base de dados?
- Origem dos dados
- Instituição de Vinculação
- Os dados podem ser baixados?
- Dados públicos ou fechados
- Alimentação contínua ou encerrada
- Abrangência Nacional, Estadual ou Municipal
- Há dados sobre o Estado do RJ?
- O dado é desagregável por município?
- Se Municipal, indique o Município
- Eixo Executivo ou Judiciário?
- Indique o Eixo Executivo (segurança pública, saúde, assistência social, educação)
- Descreva o propósito da base
- Indique os marcadores/as variáveis
- O termo feminicídio aparece como marcador?
- Quantas linhas tem a base de dados, ou quantas observações existem, se for possível identificar
- Indique as categorias de georreferenciamento identificadas (endereço da vítima, endereço do local de registro do dado, endereço do local de consumação da violência)
- Indique as categorias de dados de agressor cadastradas (relação afetiva com a vítima, gênero, idade, raça, religião, escolaridade, ocupação)
- Indique informações de encaminhamento identificadas (tipo de atendimento fornecido)
- Indique se há possibilidade de mapeamento de reincidência
- Há informação sobre letalidade nos eixos? (saúde, segurança pública, assistência social)
- Nível de relevância (1-5)
- Observações
- Link para a Base de Dados



MAPEAMENTO DOS DADOS

Foram identificadas **mais de vinte fontes de dados relevantes**, incluindo o Sistema de Informação de Agravos de Notificação (SINAN), o Sistema de Informações sobre Mortalidade (SIM), o Cadastro de Feminicídio do CNMP, o DataSenado, a Plataforma EVA, o Dossiê Mulher e bases de institutos especializados, como Patrícia Galvão.

As fontes apresentaram formatos variados, como *dashboards*, planilhas, PDFs, além de desafios como ausência de recorte municipal, dados agregados e falta de mecanismos de exportação.

CRITÉRIOS DE RELEVÂNCIA

A priorização das fontes foi feita a partir de cinco critérios principais: cobertura e origem, presença de indicadores de violência e letalidade, acesso e transparência, detalhamento das informações sobre vítimas e agressores e possibilidade de integração com outras bases.

Cada base recebeu uma nota de 1 a 5, resultando em uma classificação que orientará a coleta e análise nas próximas fases.

Critério	Item Avaliado
1. Cobertura e origem	Contém dados específicos do Estado do RJ desagregáveis por município
	Está vinculado ao eixo executivo (segurança, saúde, assistência, educação) ou judiciário
2. Violência e letalidade	Violência: sim/não (grau de letalidade)
3. Acesso e transparência	Acesso público e gratuito
	Dados exportáveis (CSV, Excel, API, etc.)
4. Detalhamento da informação	Contém perfil da vítima (gênero, idade, raça)
	Relação de afetividade com o agressor
5. Integração	Atualização periódica (mensal/trimestral/anual)
	*Disponibiliza dados históricos (anos anteriores) - atualização automática
	*Permite cruzamento com outras bases públicas do RJ

Esses critérios serviram como norteadores para a priorização das fontes primárias, permitindo avaliar de forma padronizada os aspectos gerais dos dados disponíveis publicamente.

É importante destacar que essa classificação das bases prioritárias não é definitiva: há bases de dados que, embora não tratem diretamente do feminicídio, serão utilizadas pela sua importância estratégica, como é o caso do Censo Demográfico, que, apesar de ser de extrema relevância, recebeu um nível de classificação inferior ao definido.

BASES DE DADOS PRIORITÁRIAS

A seleção das bases prioritárias para o primeiro ciclo de coleta considerou a pontuação atribuída conforme os critérios de relevância definidos na etapa metodológica, bem como a disponibilidade e a qualidade das informações oferecidas por cada fonte.

Foram priorizadas as bases com maior potencial de análise direta sobre feminicídios e violência contra a mulher, sem desconsiderar aquelas que, embora apresentem relação indireta com o tema, oferecem dados estratégicos para análises complementares e contextuais.

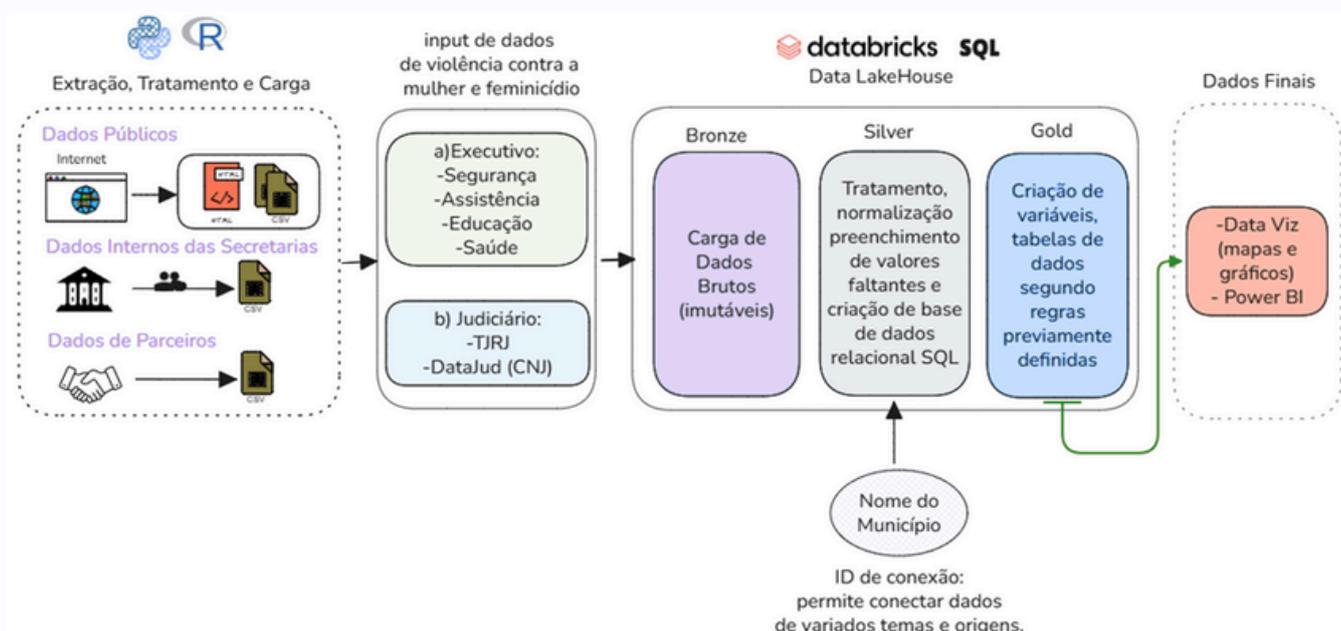
As bases de dados mapeadas foram avaliadas segundo níveis de relevância, variando de 1 a 5. Para a etapa seguinte, foram selecionadas prioritariamente aquelas com classificação entre os níveis 3 a 5.

- Relevância máxima (5): SINAN – Sistema de Informação de Agravos de Notificação.
- Alta relevância (4): SIM, Cadastro de Feminicídio (CNMP), DataSenado, Plataforma EVA e Dossiê Mulher.
- Relevância média (3): PNAD, Estatísticas de Gênero (IBGE), DATASUS, Painel de Violência Interpessoal e Autoprovocada e SINESP.

MODELO DE PRODUÇÃO DOS DADOS

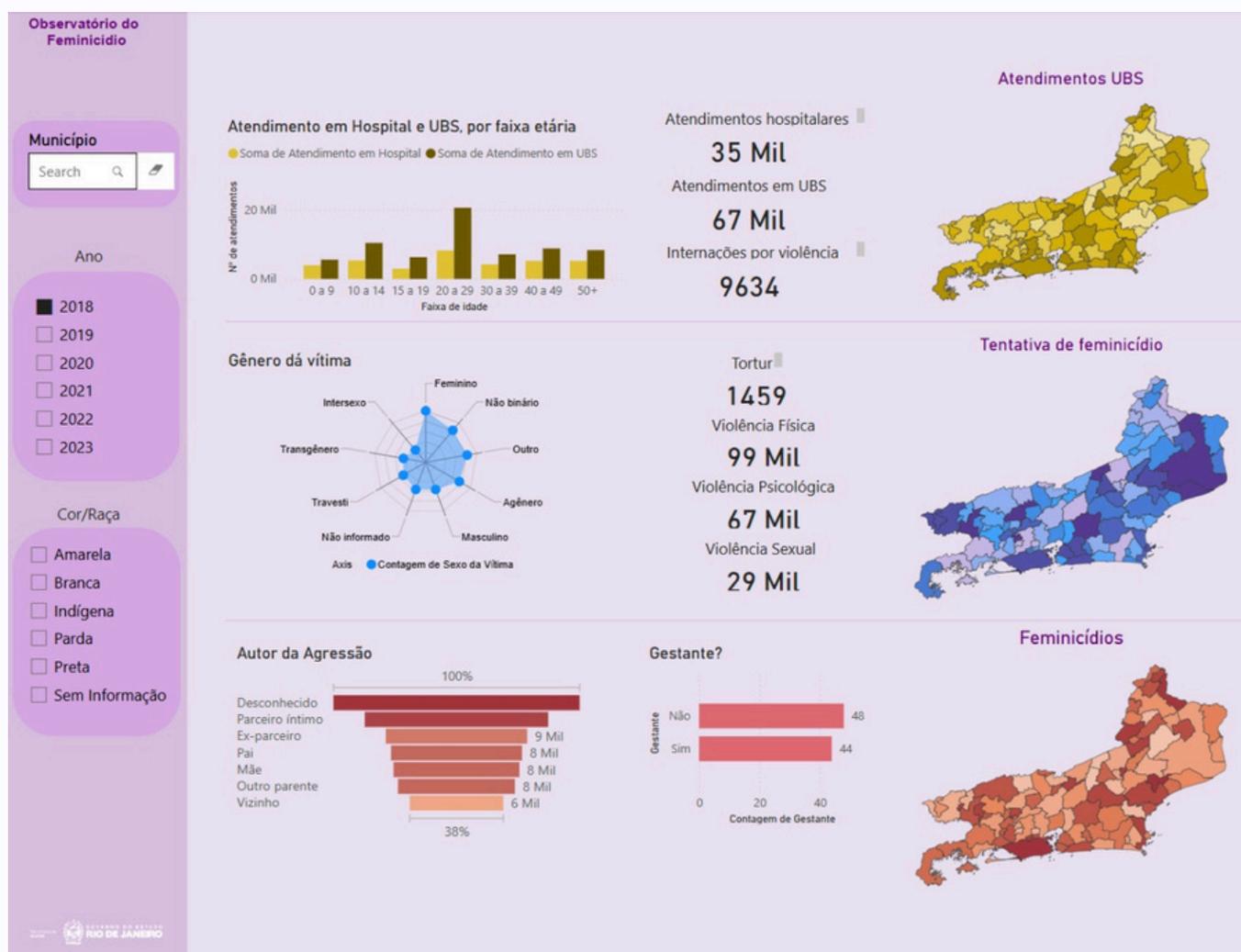
Para trabalharmos com os dados, foi proposto um modelo de produção baseado em três camadas no ambiente Databricks: Bronze, para armazenamento de dados brutos; Silver, para tratamento e normalização; e Gold, para geração de variáveis e visualizações finais.

O modelo utiliza um identificador único de município para permitir cruzamentos entre fontes e garantir coerência territorial.



DASHBOARD DEMONSTRATIVO

Para validar o que deverá ser produzido nas próximas fases, criamos um *dashboard* com dados sintéticos gerados em Python, simulando as variáveis das bases oficiais para os 92 municípios do estado. O protótipo permite demonstrar como os dados reais poderão ser visualizados e explorados, conforme imagem abaixo:



ROTA CRÍTICA DA VIOLÊNCIA

Além disso, desenvolvemos um organograma que ilustra o percurso institucional e social percorrido pelas mulheres, reunindo tipos de violência, pontos de entrada da rede de atendimento, perfis de vítimas e agressores e encaminhamentos disponíveis. O material é preliminar e será refinado ao longo do projeto. Abaixo, apresentamos um recorte do organograma em construção:

VIOLÊNCIA CONTRA AS MULHERES

Quadro comparativo das tipificações

CATEGORIA	LEI MARIA DA PENHA	SINAN (NOTIFICAÇÃO COMPULSÓRIA)
Violência Física	Sim	Sim
Violência Sexual	Sim	Sim
Violência Psicológica	Sim	Sim
Violência Moral	Sim	Não
Violência Patrimonial	Sim	Sim
Violência Obstétrica / Reprodutiva	Não explicitamente (Parcialmente incluída como violência Institucional)	Parcial (em contexto de violência sexual ou física)
Violência Política de Gênero	Não	Não
Violência Institucional	Não explicitamente (reconhecida por interpretações ampliadas)	Parcial
Violência Digital	Não	Não
Consequências graves	Não (mas reconhece o risco de morte como agravante)	Não (registradas como desfechos, não como categorias de violência)

RESULTADOS ALCANÇADOS

O trabalho desta fase resultou em um mapeamento estruturado das fontes de dados, na criação de um instrumento padronizado para coleta de informações sobre os sistemas internos das Secretarias, na definição de critérios técnicos validados para priorização de bases de dados públicas e disponíveis, no desenvolvimento de um protótipo de *dashboard* e na elaboração de representações gráficas da rota crítica das mulheres em situação de violência.

PRÓXIMOS PASSOS

A próxima etapa do projeto contemplará a coleta e o tratamento das bases priorizadas, bem como a incorporação de dados provenientes de parceiros institucionais, a integração de dados compatíveis entre si, e o desenvolvimento de *dashboards* interativos com os dados reais. Também será iniciada a implementação do sistema no ambiente do Observatório.

IMPACTO ESPERADO

Com a implantação do sistema, esperamos viabilizar um monitoramento territorializado dos feminicídios no estado, apoiar políticas públicas baseadas em dados, fortalecer a rede de proteção e colaborar para a transparência desses dados de violência contra as mulheres.