

Case Técnico - Analista de Dados

Contexto

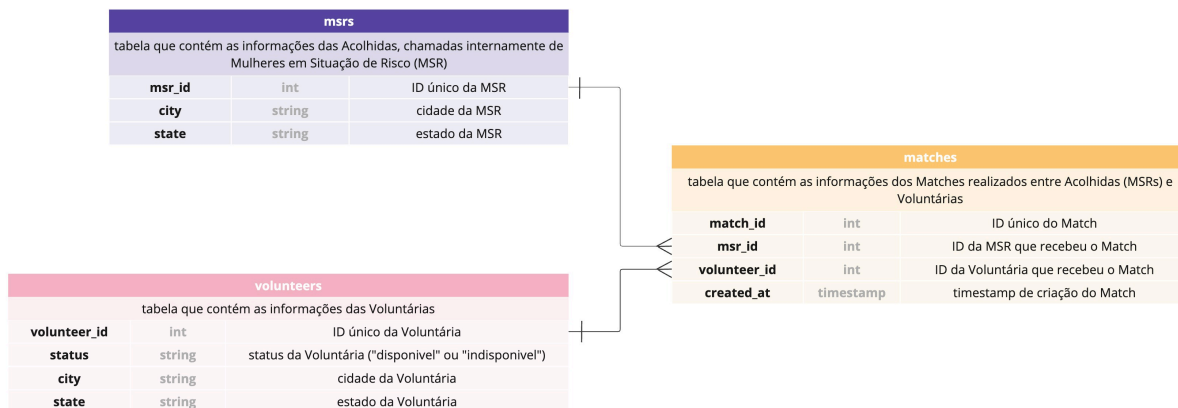


No **Mapa do Acolhimento**, trabalhamos para garantir que mulheres em situação de violência encontrem acolhimento eficaz para romper o ciclo da violência e prevenir o feminicídio.

Uma das principais formas de atuação da organização é através da tecnologia, conectando sobreviventes de violência de gênero (as **acolhidas**) a uma rede de **voluntárias**, composta por psicólogas e advogadas. Essa conexão é viabilizada por um algoritmo de **match**, que identifica automaticamente a voluntária mais adequada para atender às demandas de cada acolhida cadastrada na plataforma.

Banco de Dados

Os dados das **acolhidas**, **voluntárias** e **matches** são armazenados em nosso banco de dados. Para este case, disponibilizamos três arquivos **.csv** contendo amostras anonimizadas de algumas dessas tabelas. A imagem abaixo apresenta a estrutura e as relações entre essas tabelas:



msrs.csv

volunteers.csv

matches.csv

Análise e Visualização de Dados

No Mapa do Acolhimento, acreditamos que os dados têm um papel fundamental na tomada de decisões, definição de prioridades e avaliação do impacto das nossas iniciativas. Com base nos dados fornecidos, gostaríamos que você realizasse uma análise exploratória e nos ajudasse a responder às perguntas a seguir.

1. Temos uma meta de longo prazo de realizar um total 10 mil matches até 2030

- Quantos matches foram realizados apenas em 2024? Como esse número se compara com os anos anteriores?

- Crie uma visualização que nos ajude a entender se estamos próximas de alcançar a meta final.

2. Um dos desafios da operação é manter nossas voluntárias engajadas. Para isso, acompanhamos de perto o número de voluntárias ociosas, definido da seguinte forma:

Voluntária ociosa é aquela que está com status de "disponível", mas não possui nenhum match.

- Elabore uma análise que nos permita entender em quais territórios temos mais voluntárias ociosas. Estamos interessadas em entender:
 - O número absoluto de voluntárias ociosas por UF
 - A distribuição percentual de voluntárias ociosas por UF, ou seja, quais estados concentram a maior proporção de voluntárias ociosas, em relação ao total observado

Orientações técnicas:

- Utilize **SQL** para elaborar as consultas que respondem às perguntas acima. Você pode usar a engine com a qual tiver mais familiaridade (PostgreSQL, MySQL, BigQuery, etc.).
- Para as visualizações, utilize a ferramenta de sua preferência (ex: Metabase, Tableau, Jupyter Notebook, etc.).
- Sinta-se livre para complementar suas análises com breves interpretações que ajudem pessoas não técnicas a entender os dados e tomar decisões com base neles.



O objetivo desta etapa é entender como você estrutura seu código, conduz análises e traduz dados em insights acionáveis e acessíveis para diferentes públicos.

Testes e experimentação

Após a realização do match entre acolhida e voluntária, enviamos uma pesquisa para que a acolhida compartilhe sua experiência com o atendimento recebido por meio do Mapa do Acolhimento.

Essas respostas são fundamentais para que possamos avaliar e melhorar continuamente nossos processos, garantindo que o acolhimento oferecido seja realmente transformador na vida das mulheres.

Atualmente, essa pesquisa é enviada por **e-mail**, mas temos enfrentado uma **baixa taxa de resposta**. No último mês, por exemplo, enviamos **577 pesquisas** e recebemos apenas **151 respostas**.

Diante desse desafio, estamos avaliando a possibilidade de mudar o canal de comunicação e passar a enviar as pesquisas via **WhatsApp**. Acreditamos que, por ser um canal mais direto e acessível, o WhatsApp pode aumentar o engajamento das acolhidas com as pesquisas.

No entanto, antes de adotar essa mudança em larga escala, queremos entender se ela será de fato **efetiva**.

Como você abordaria esse problema? Que tipo de teste ou experimento faria para avaliar se o envio por WhatsApp é mais eficiente?



Queremos entender como você estrutura hipóteses, define experimentos e avalia resultados para orientar mudanças operacionais com base em dados.

Modelagem e Engenharia de Dados

No Mapa do Acolhimento, todos os dados utilizados em nossas análises são gerados pelos nossos próprios serviços. Isso significa que, com frequência,

precisamos **modelar novas tabelas** e **estruturar pipelines** que alimentem essas tabelas de forma consistente, segura e confiável.

A seguir, apresentamos um exemplo de **payload** que recebemos no momento do **cadastro de uma acolhida**:

```
{
  "email": "acolhida.exemplo@gmail.com",
  "nome": "Acolhida Exemplo",
  "data_de_nascimento": "01/01/1990",
  "telefone": "(00) 00000-0000",
  "cep": "0000-000",
  "bairro": "Sé",
  "cidade": "São Paulo",
  "uf": "SP",
  "possui_deficiencia": "não",
  "genero": "mulher cis",
  "cor": "parda",
  "latitude": -23.551447,
  "longitude": -46.634321,
  "timestamp_cadastro": "2025-04-10T18:31:41+00:00",
  "pode_receber_atendimento_online": "sim",
  "violencia": {
    "tipos_de_violencia": "['psicológica', 'patrimonial', 'física']",
    "quanto_tempo_esta_sofrendo_violencia": "menos de 1 ano",
    "ja_procurou_ajuda": "não",
    "e_a_primeira_vez_que_sofre_violencia": "sim",
    "onde_ocorreu_a_violencia": "ambiente domestico",
    "realizou_registro_da_violencia": "não"
  }
}
```

Com base nesse payload, responda às seguintes perguntas:

1. **Como você estruturaria esses dados em tabelas no banco de dados?**
 - Quais entidades e relacionamentos você criaria?
2. **Quais práticas de governança de dados você consideraria essenciais?**

3. Transformação em formato estruturado:

Considerando o modelo de dados que você propôs, escreva um código em **Python** que transforme o payload acima em um formato estruturado.

- Pode ser um arquivo `.csv`, um `pandas.DataFrame`, ou uma carga em um banco de dados relacional.
- O objetivo é mostrar como você transforma dados semi-estruturados em algo que pode ser analisado com facilidade.



Queremos entender sua familiaridade com conceitos de engenharia de dados e modelagem, além da sua capacidade de lidar com dados não estruturados e prepará-los para análises futuras.

Instruções para a entrega

- Não se preocupe em apresentar uma solução perfeita, mas sim em elaborar uma proposta que demonstre **sua abordagem e experiência**.
- Estamos interessadas em conhecer sua experiência com **análise de dados, experimentação, modelagem e governança**.
- A documentação entregue pode ser no formato em que preferir (.pdf, .ppt, etc). Não é necessário se preocupar em estilizar a sua documentação.
- Não se esqueça de incluir o **código gerado** durante a resolução. Pode ser em arquivos `.sql`, `.py` ou em um **notebook** com as consultas e scripts utilizados.
- Você deverá enviar sua solução como resposta a este e-mail em até **7 dias corridos** após o recebimento deste Case. Utilize esse padrão se necessário para o assunto do e-mail:
 - **[Analista de Dados] Case Técnico - Seu Nome**

Se tiver dúvidas ou precisar de mais informações sobre o problema, entre em contato conosco e estaremos à disposição para ajudar da melhor forma possível.

