

Anonimização de dados

O QUE É ANONIMIZAÇÃO DE DADOS?

Você tem aprendido sobre a importância da privacidade na análise de dados. Agora, é hora de falar sobre **anonimização de dados** e quais tipos de dados devem ser anônimos. As **Informações de identificação pessoal**, ou **PII**, são informações que podem ser usadas sozinhas ou com outros dados para rastrear a identidade de uma pessoa.

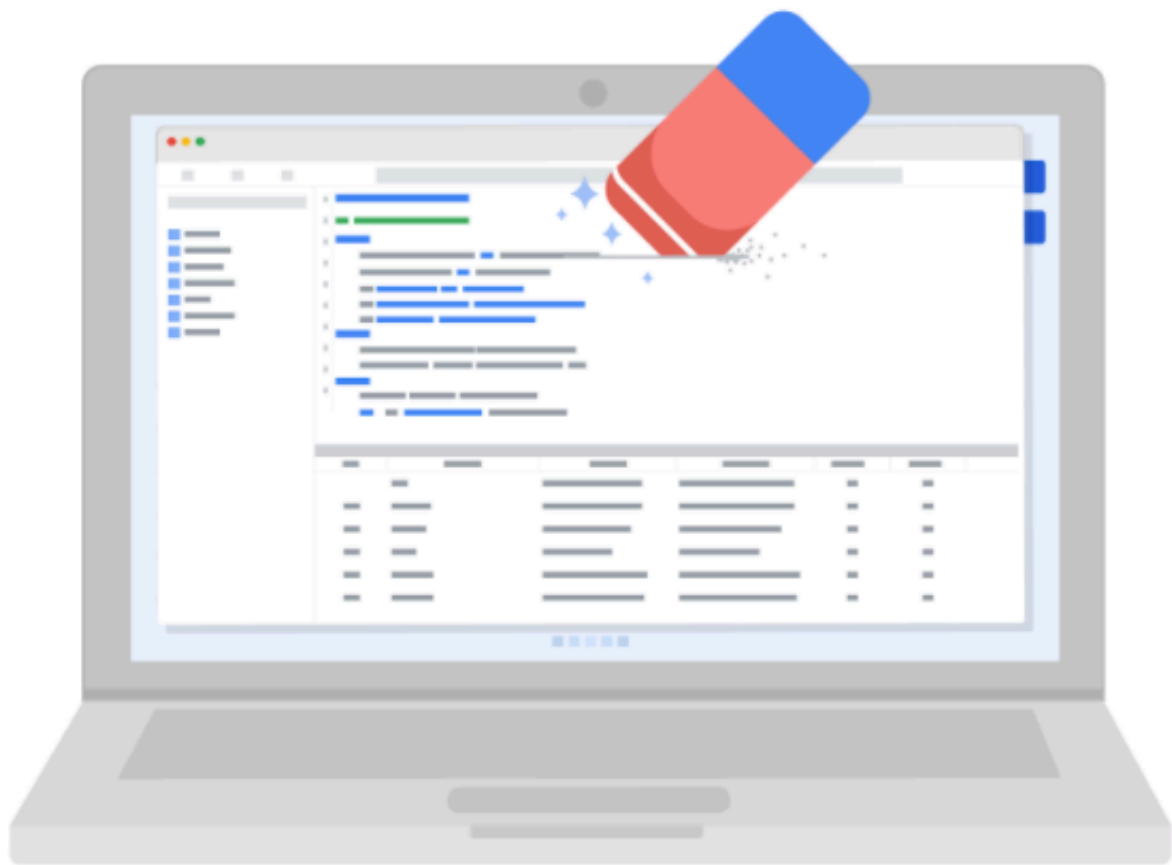
A anonimização de dados é o processo de proteger os dados privados ou confidenciais das pessoas, eliminando esse tipo de informação. Normalmente, a anonimização de dados envolve anulação, hashing ou mascaramento de informações pessoais, muitas vezes usando códigos de comprimento fixo para representar colunas de dados ou ocultando dados com valores alterados.

SEU PAPEL NA ANONIMIZAÇÃO DE DADOS

As organizações têm a responsabilidade de proteger seus dados e as informações pessoais que eles possam conter. Como analista de dados, espera-se que você entenda quais dados precisam ser anônimos, mas geralmente você não seria responsável pela sua própria anonimização de dados. Uma rara exceção pode ser se você trabalhar com uma cópia dos dados para fins de teste ou desenvolvimento. Nesse caso, pode ser necessário tornar os dados anônimos antes de trabalhar com eles.

QUAIS TIPOS DE DADOS DEVEM SER ANÔNIMOS?

Os dados de saúde e financeiros são dois dos tipos de dados mais confidenciais. Esses setores dependem muito de técnicas de anonimização de dados. Afinal, os riscos são muito altos. É por isso que os dados nesses dois setores geralmente passam pela **desidentificação**, que é **um processo usado para limpar os dados de todas as informações de identificação pessoal**.



A anonimização de dados é usada em quase todos os setores. É por isso que é tão importante que os analistas de dados entendam o básico. Aqui está uma lista de dados que muitas vezes são anônimos:

- **Números de telefone**
- **Nomes**
- **Placas e números de licença**
- **Números de previdência social**
- **Endereços de IP**
- **Registros médicos**
- **Endereços de e-mail**
- **Fotos**
- **Números de conta**

Para algumas pessoas, faz sentido que esse tipo de dados seja anônimo. Para outros, temos que ser muito específicos sobre o que precisa ser anônimo. Imagine um mundo onde todos nós tivéssemos acesso aos endereços, números de contas e outras informações identificáveis uns dos outros. Isso invadiria a privacidade de muitas pessoas e tornaria o mundo menos seguro. A anonimização de dados é uma das maneiras de manter os dados privados e seguros!