# Relatório de Segurança – Tópico 3.3

## Objetivo

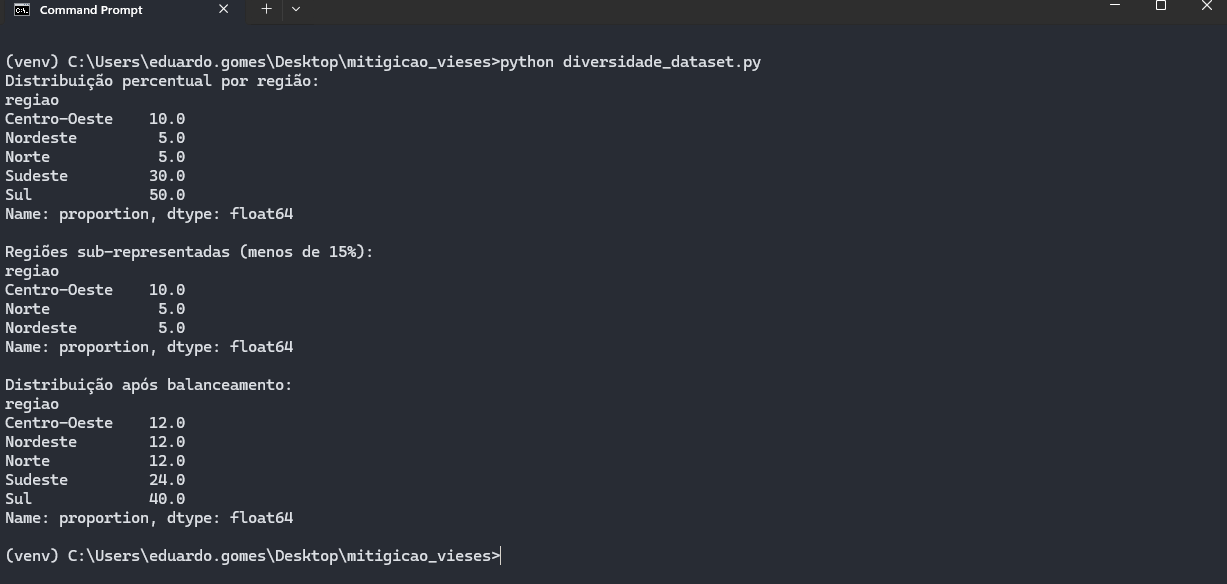
Garantir que o conjunto de dados utilizado no treinamento de modelos represente adequadamente todos os grupos populacionais, evitando decisões enviesadas por super ou sub-representação de regiões ou perfis demográficos.

## Ações Realizadas

- Simulado um dataset com desigualdade regional (Sul e Sudeste super-representados).  
- Calculado o percentual de representatividade de cada região.  
- Identificadas regiões sub-representadas com menos de 15% de presença no conjunto.  
- Aplicada técnica de reamostragem para aumentar a presença das regiões menos representadas.  
- Verificada nova distribuição após balanceamento.

## Evidência

Execução do script com análise de representatividade e resultado após balanceamento:



## Conclusão

A diversificação de datasets é essencial para garantir que sistemas de IA operem de forma justa com todas as regiões ou grupos representados.   
A análise demonstrou que é possível aplicar balanceamento simples para corrigir desproporções e melhorar a equidade na base de dados.