

Análise Preditiva de Mortalidade por Infarto (IAM) no Brasil

Uma abordagem de Big Data e Machine Learning no SUS

Nome: André Felipe Baretta, Henry Meneguini Farias

RA: 2551713, 2551810

Problema

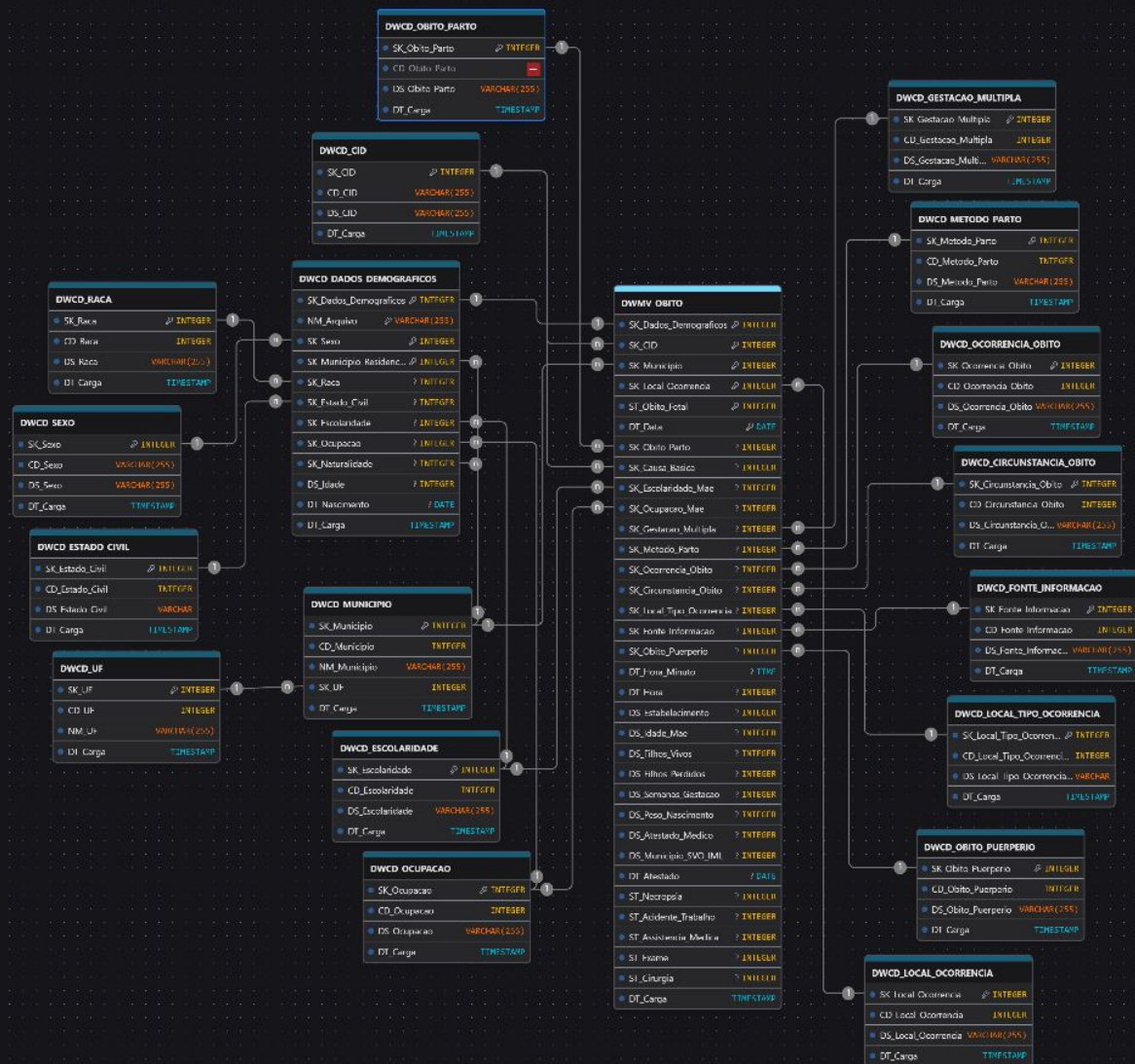
- Doenças Cardiovasculares: Causa nº 1 de mortes no mundo.
- O paradoxo do SUS: Temos muitos dados (DataSUS), mas pouca informação preditiva.
- Desafio: Como identificar padrões de risco em 1,5 milhão de registros heterogêneos?

Perguntas da pesquisa

- Como Idade e Sexo impactam a mortalidade? (Quantificar o risco).
- Quais Estados têm a maior mortalidade proporcional? (Geografia).

A Jornada do Dado (Pipeline)

OpenDataSUS (CSV) → Limpeza Python → SQLite (Data Warehouse) →
Amostragem (50k) → Machine Learning.



Estratégia de Modelagem

Target: CID I21 (Infarto) = 1 | Outros = 0.

Features: Idade, Sexo, Raça, Região.

Pré-processamento: Tratamento de Nulos, Normalização (Scaler), One-Hot Encoding.

Os Algoritmos Comparados

Regressão Logística: O clássico (Baseline).

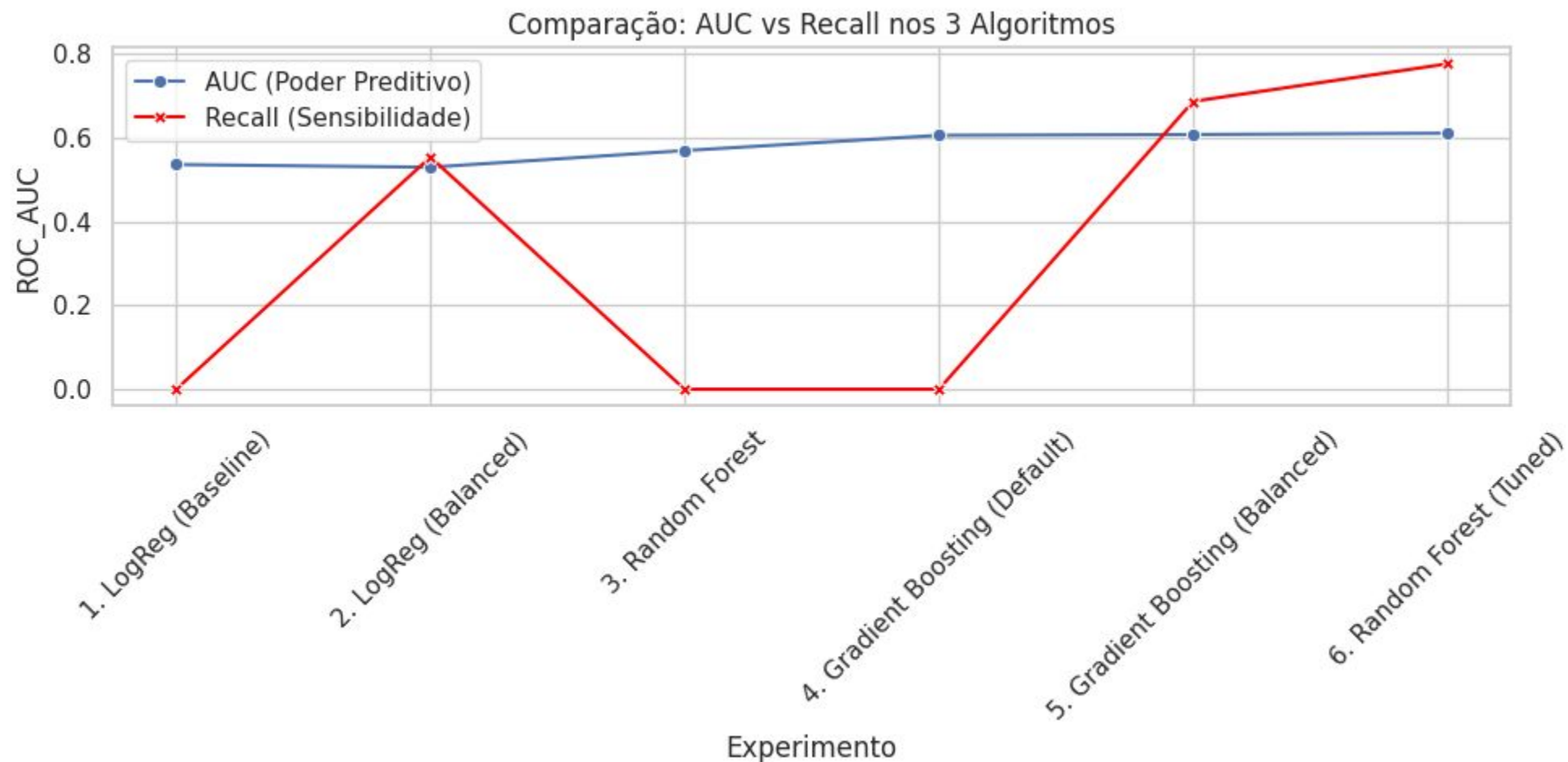
Random Forest: O robusto (Árvores de Decisão).

Gradient Boosting: O estado da arte.

A História da Evolução

	ROC_AUC	Recall	Precision
Experimento			
1. LogReg (Baseline)	0.536209	0.000000	0.000000
2. LogReg (Balanced)	0.529805	0.554068	0.067716
3. Random Forest	0.569567	0.000000	0.000000
4. Gradient Boosting (Default)	0.605946	0.000000	0.000000
5. Gradient Boosting (Balanced)	0.607826	0.686921	0.083001
6. Random Forest (Tuned)	0.610940	0.776519	0.080711

A História da Evolução



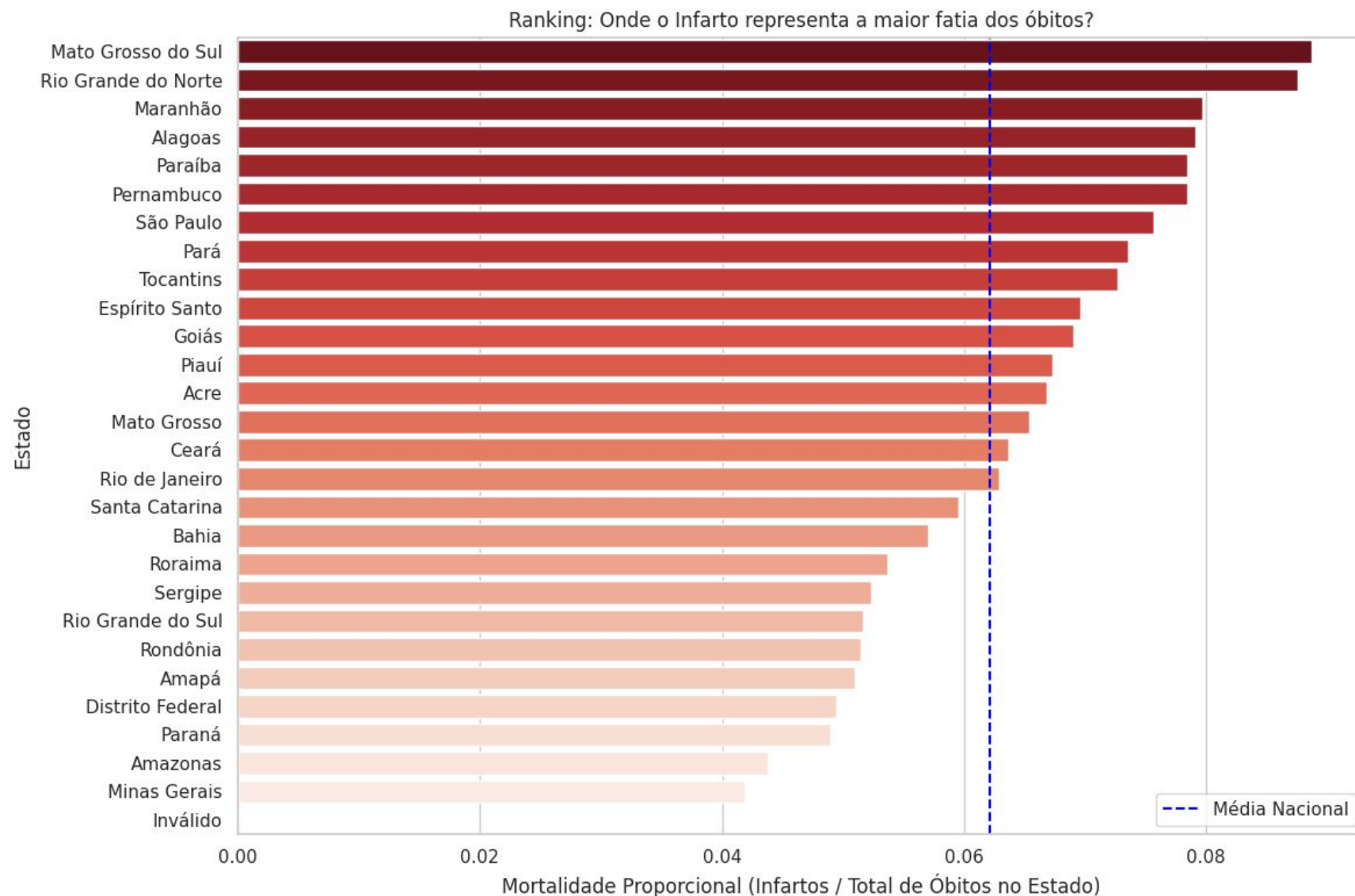
A História da Evolução

Matriz de Confusão (Recall ~90%)

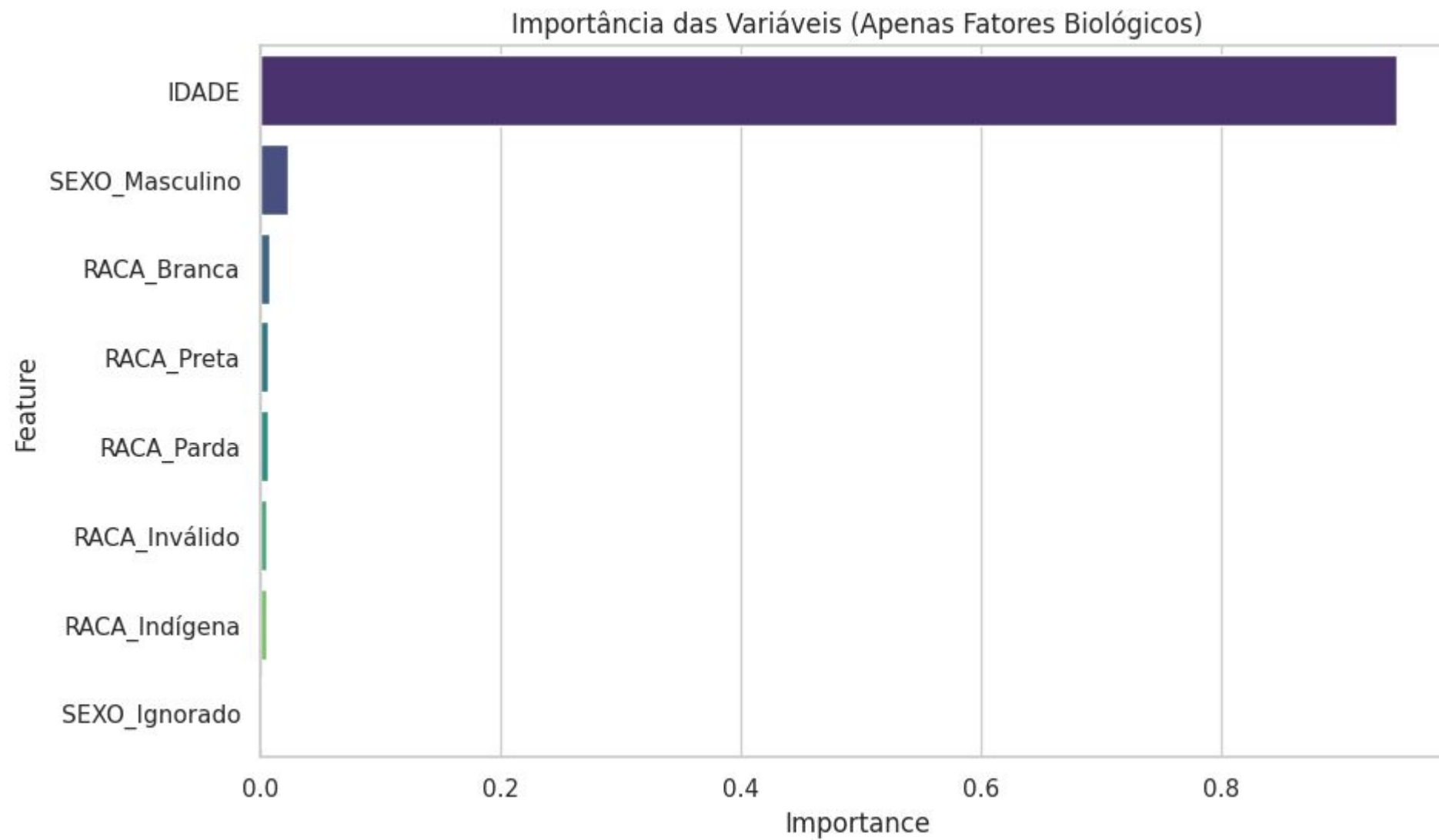
Real	0	1
	0	1
0	3405	10624
1	97	874

Predito (0=Outros, 1=Infarto)

Resultados Geográficos



Fatores de Risco Biológicos



Os Algoritmos Comparados

- Alta Sensibilidade = Segurança (Não perde casos).
- Baixa Precisão = Custo de Triagem (Muitos alertas).

Limitações Identificadas

- O "Teto Demográfico".
- Variáveis demográficas servem para **Triagem**, não para **Diagnóstico**.

Recomendações Práticas

Uso Sugerido: Ferramenta de Gestão Pública (Vigilância).

Não usar para: Diagnóstico clínico individual.

Ação: Direcionar verbas para as regiões apontadas no Slide 8.

Próximos Passos

- Cruzar com Dados Climáticos.
- Testar aprendizados de máquina mais complexos.
- Variáveis de Causa Associada