

# Respostas - Questionário do Projeto Final

## Introdução à Ciência de Dados

### **Integrantes do Grupo:**

Leonardo de Faria Sales -  $N_{USP}$ : 11815973

Nathan Mayer Hunhoff -  $N_{USP}$ : 13686664

Pedro Calligaris Delbem -  $N_{USP}$ : 5255417

Prof. Francisco Rodrigues

24 de outubro de 2025

## **1 - Descreva o tema do projeto**

Este projeto investiga a correlação entre a cobertura vacinal infantil e as taxas de mortalidade nos Estados Unidos. Adicionalmente, o estudo busca contextualizar essa relação frente ao fenômeno do movimento antivacina no país.

## **2 - Descreva a base de dados que será usada**

Utilizaremos a base de dados do governo dos Estados Unidos, que abrange a cobertura vacinal infantil no país para as vacinas HEPB3, POL3, HIB3, MCV1 e DTP3, bem como a base de dados da UNICEF que registra as taxas de mortalidade infantil para o mesmo período de 2002 a 2022.

## **3 - Quais os passos serão adotados na análise dos dados**

O fluxo de trabalho adotado está detalhado no fluxograma da Figura 1, na página seguinte.

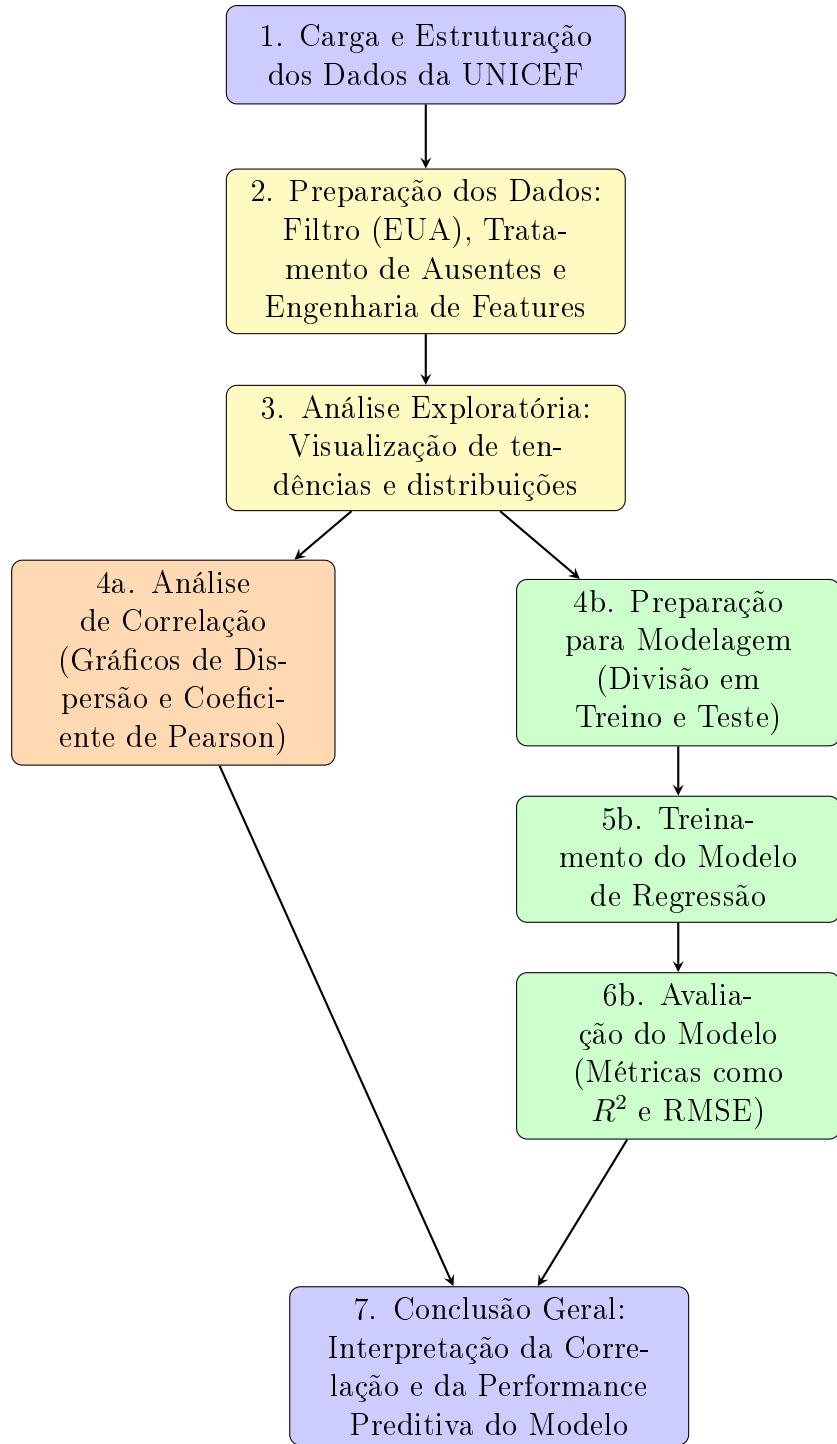


Figura 1: Fluxograma do projeto com ramos de Análise de Correlação e Modelagem Preditiva.

## 4 - Quais os desafios esperados?

Durante o desenvolvimento do projeto, esperamos encontrar os seguintes desafios:

- **Irregularidade dos Dados:** A base de dados da UNICEF, embora confiável, apresenta dados para crianças de 0-23 meses durante o período de 2012 a 2014, enquanto para os demais períodos os dados se referem a crianças de 0-35 meses.

- **Complexidade da Causalidade:** O maior desafio conceitual será o de discutir a correlação sem afirmar causalidade. A mortalidade infantil é um fenômeno multifatorial, influenciado por saneamento, acesso à saúde, fatores socioeconômicos, etc. Isolar o impacto da vacinação e discutir a máxima “correlação não implica causalidade” de forma rigorosa será um ponto crítico do trabalho.
- **Definição de ”Movimento Antivacina”:** Quantificar a força do ”movimento antivacina” é complexo. Como não temos uma variável direta para isso no dataset, teremos que tratar a cobertura vacinal como uma *proxy* indireta, um desafio que exige uma interpretação cuidadosa e a menção dessa limitação nos resultados.

## 5 - Quais os resultados esperados?

Com a conclusão deste projeto, esperamos obter os seguintes resultados:

- **Análise Correlacional Quantitativa:** Esperamos encontrar e quantificar uma correlação negativa estatisticamente significativa entre a cobertura das principais vacinas (como DTP3 e MCV1) e a taxa de mortalidade infantil nos EUA para o período analisado. O resultado principal será o valor do coeficiente de Pearson acompanhado de gráficos de dispersão.
- **Modelo Preditivo Funcional:** Almejamos construir um modelo de regressão linear capaz de prever a taxa de mortalidade infantil com base na cobertura vacinal. O resultado esperado é um modelo validado, com métricas de performance (como  $R^2$  e RMSE) que demonstrem sua capacidade preditiva e suas limitações.
- **Visualizações Claras e Informativas:** Produzir uma série de gráficos (séries temporais, gráficos de dispersão) que não apenas suportem nossas conclusões, mas também comuniquem de forma clara a evolução das variáveis e a relação entre elas ao longo das duas décadas analisadas.

## 6 - Por que vocês escolheram esse tema?

A escolha deste tema foi motivada pela crescente relevância do debate sobre vacinação na saúde pública global. Nos últimos anos, observamos um aumento da hesitação vacinal e a disseminação de desinformação, fenômenos que coincidem com o ressurgimento de doenças antes controladas. O projeto nos permite aplicar as ferramentas de Ciência de Dados para investigar, com base em evidências, um problema real e de alto impacto social. A oportunidade de analisar dados concretos da UNICEF para explorar a relação entre a cobertura vacinal e um indicador tão crítico quanto a mortalidade infantil nos pareceu uma aplicação prática e significativa dos conceitos aprendidos na disciplina.