

Projeto Prático

Luan P. Barbosa

¹Universidade do Estado de Minas Gerais (UEMG)
Passos, Brasil

Resumo. *Este projeto busca identificar a existência ou não de relação entre vulnerabilidade social, renda, trabalho e outros indicadores econômicos e suicídio, utilizando dados de uma década dos estados brasileiros e aplicando o algoritmo de aprendizagem de máquina Random Forest.*

1. Introdução

De acordo com a Organização Mundial da Saúde o Brasil ocupa a oitava posição em casos de suicídio e dados recentes indicam que o número de casos no país tem aumentado nos últimos anos. Isto reforça a importância de centros de atenção psicossocial (CAPS) e a necessidade de entender o cenário brasileiro, identificando causas para a organização e execução de medidas preventivas. Este projeto busca verificar se há alguma relação entre a situação socioeconômica de estados brasileiros e as taxas de suicídios, utilizando aprendizagem de máquina para verificar se é possível prever ou identificar um padrão a partir dos dados.

2. Metodologia

Para a realização do projeto foram coletados dados de diferentes órgãos públicos, como:

DataSus, para a obtenção dos dados de registro de óbitos por suicídio, cid X60 - X84. Também a quantidade de CAPS.

No IPEA que agrupa dados próprios e também de outros órgãos como o Instituto Brasileiro de Geografia e Estatística (IBGE) foram encontrados dados socioeconômicos como, Índice de vulnerabilidade social (IVS) dos estados; índice de Gini; quantidade de pessoas vivendo abaixo da linha de pobreza, com paridade de poder de até 6,85 dólares por dia; renda média domiciliar per capita, taxa de desemprego, média de anos de estudo e população total.

Utilizando os dados da população total e quantidade de óbitos, foi obtido a taxa de suicídio a cada 100.000 habitantes.

Estas diferentes tabelas foram unificadas e tratadas utilizando python no Google Colab.

Algumas tabelas tiveram colunas redundantes removidas manualmente após o download, como remoção de nome e sigla dos estados, deixando apenas o código.

2.1. Análise

Para a análise exploratória do dataset os 27 estados brasileiros foram divididos entre os 10 com maiores taxas de suicídio e os 10 com as menores taxas.

Com isso foi gerado os seguintes gráficos.

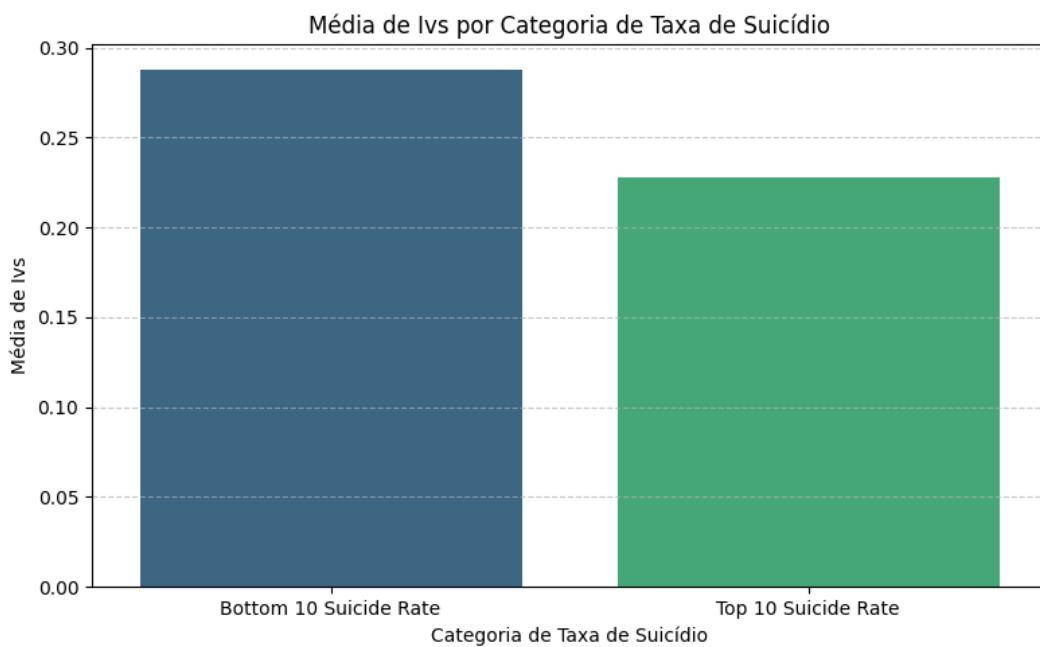


Figura 1. Média do IVS.

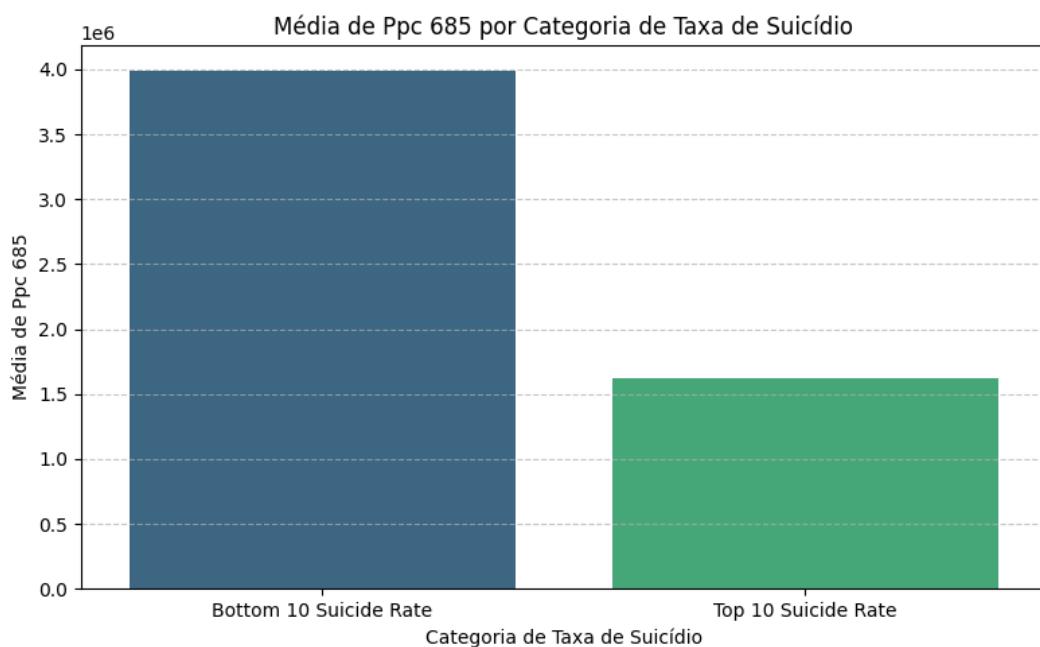


Figura 2. População abaixo da linha de pobreza. Paridade do poder de compra de 6,85 dólares diários

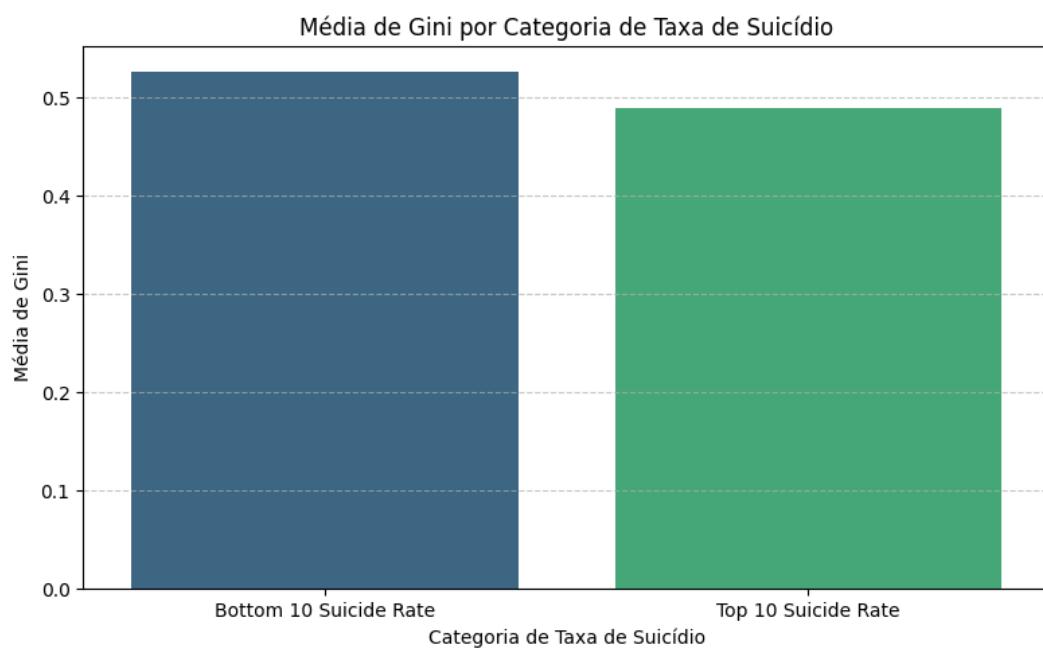


Figura 3. Média do índice de gini.

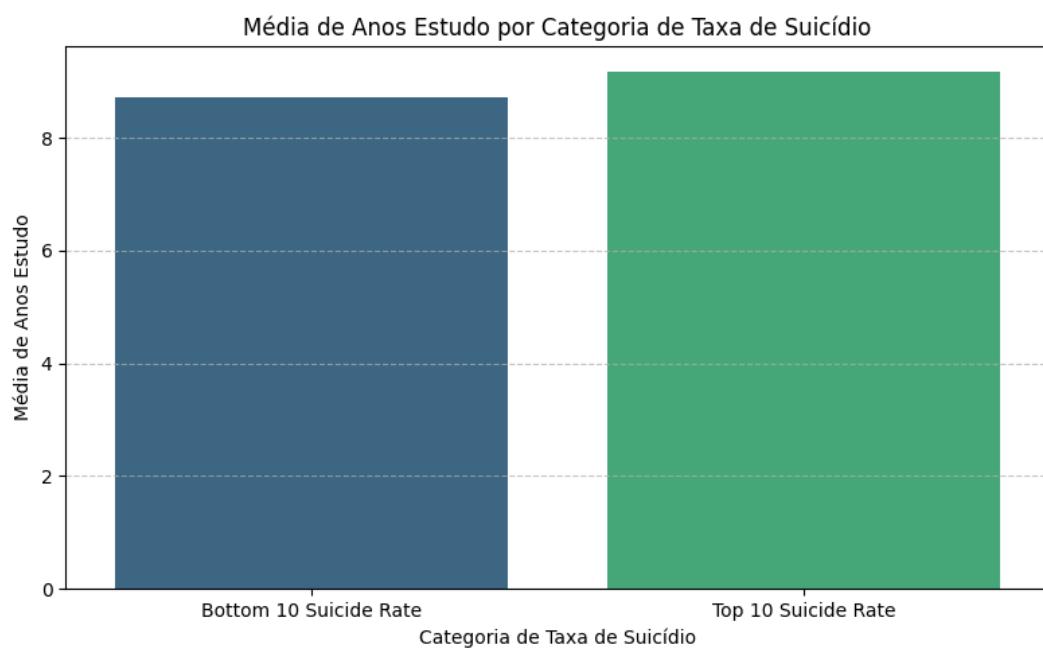


Figura 4. Média dos anos de estudo.

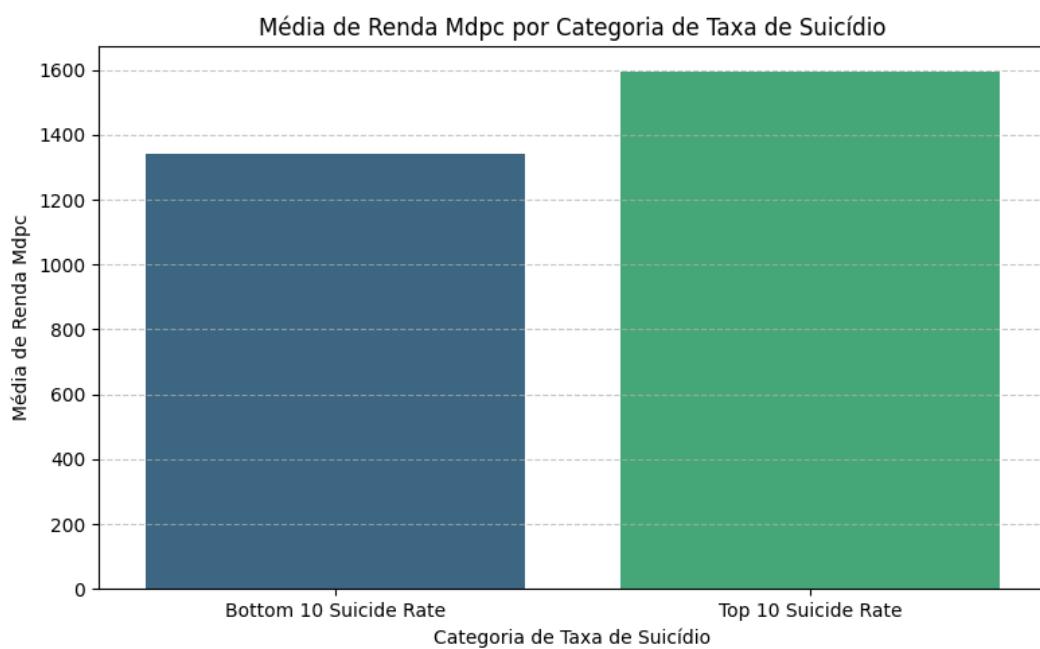


Figura 5. Média da renda domiciliar per capita.

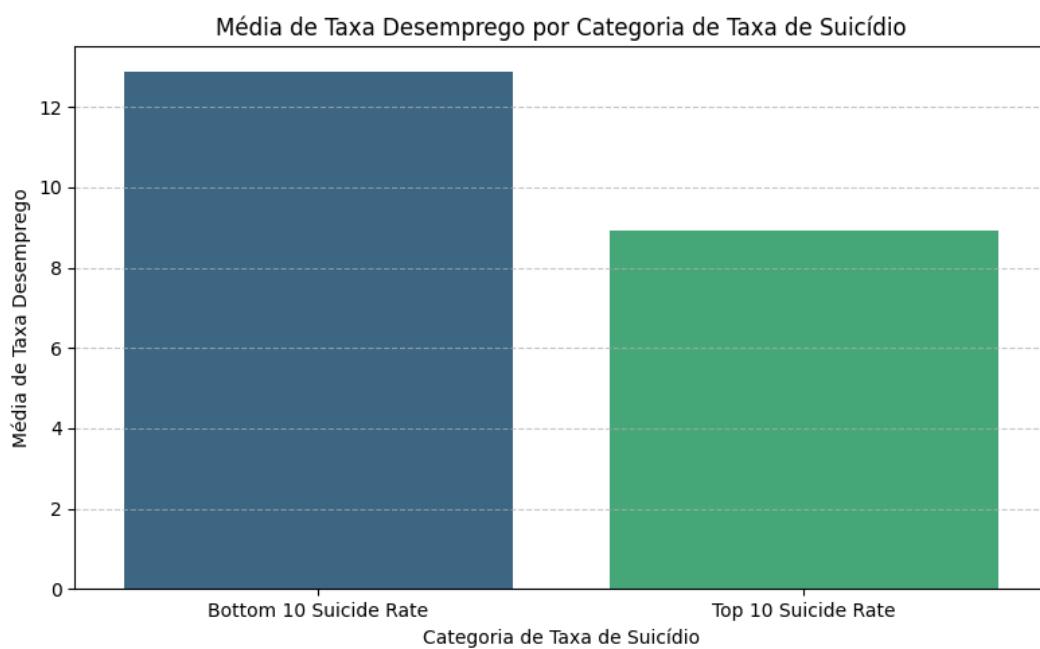


Figura 6. Média da taxa de desemprego.

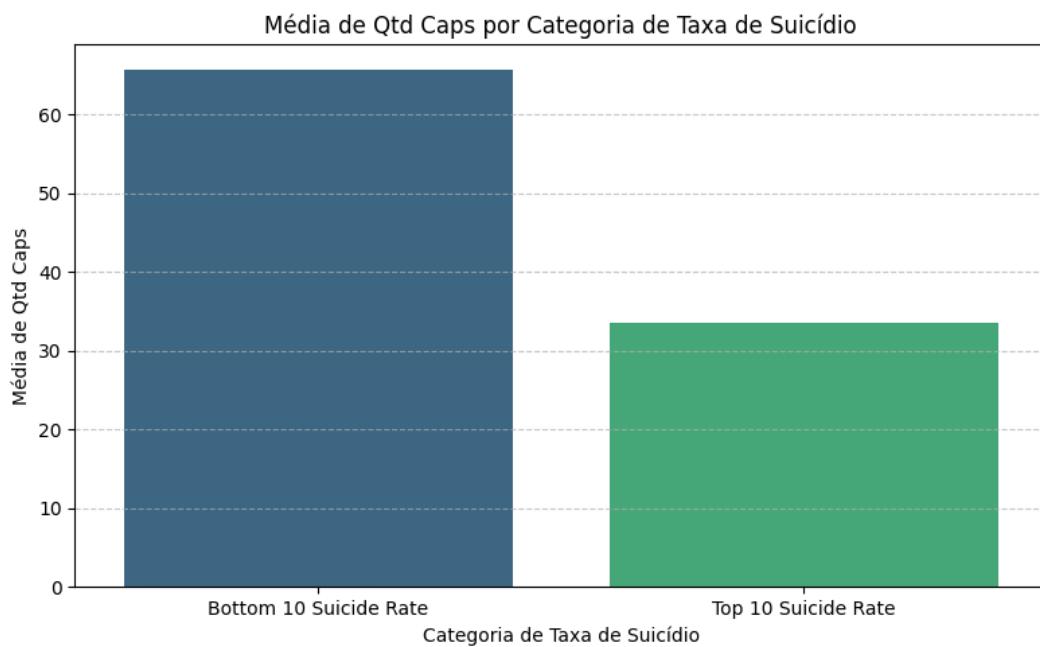


Figura 7. Média da quantidade de CAPS.

Com essas visualizações é possível perceber que estados com melhores condições socioeconômicas apresentam maiores taxas de suicídio, infelizmente não é possível entender de fato o motivo para isso, muitos fatores podem influenciar nos dados e algumas métricas podem ser difíceis de analisar apenas quantitativamente. Um importante indicativo é a quantidade de CAPS presentes nos estados, os 10 estados com menor taxa de suicídio apresentam uma média superior de unidades CAPS que os 10 com maiores taxas.

2.2. Random Forest

Antes de aplicar o modelo foi removida a coluna ano, o dataset foi normalizado, e foram eliminadas as linhas com dados faltantes.

O dataset final contou com 157 linhas e 7 colunas.

A matriz de correlação encontrada foi:

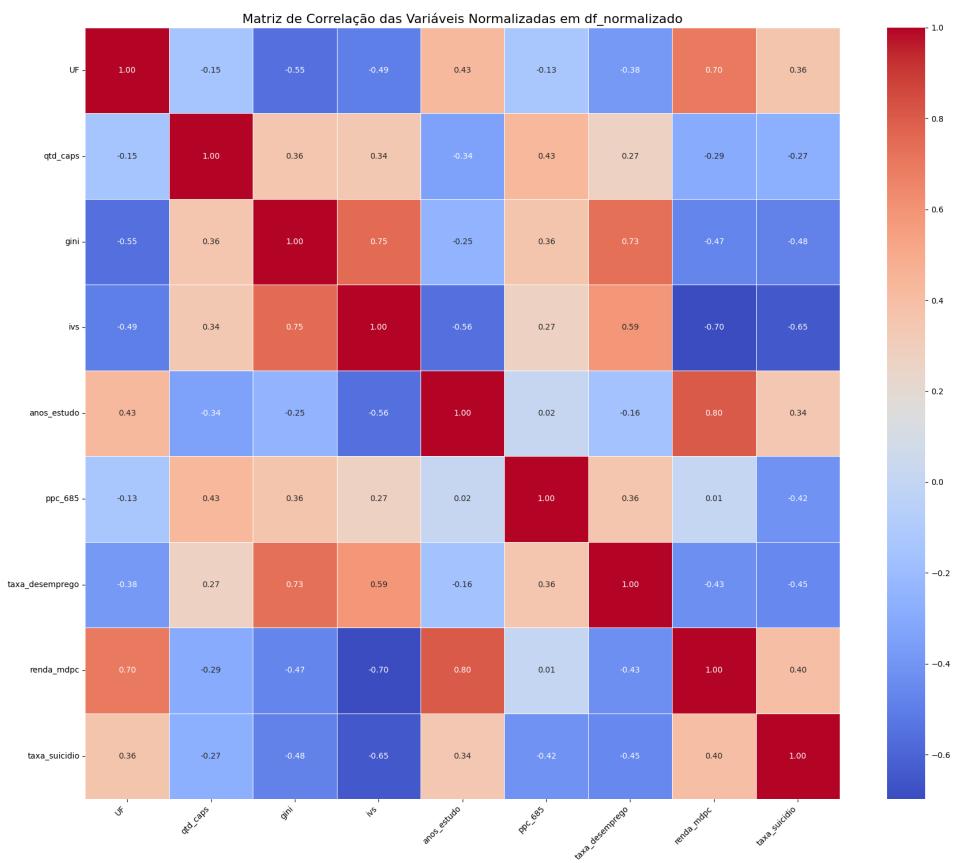


Figura 8. Matriz de correlação.

Após a execução do modelo as métricas de desempenho foram:

- Mean Absolute Error (MAE): 0.0781
- Mean Squared Error (MSE): 0.0097
- R-squared (R2): 0.8000

O modelo teve um desempenho satisfatório considerando a base de dados, testes foram realizados preenchendo a base de dados com a média ou mediana dos dados, nestes testes o Random Forest teve um desempenho pior, na casa dos 0.50 a 0.60 para o R2.

As features mais impactantes foram:

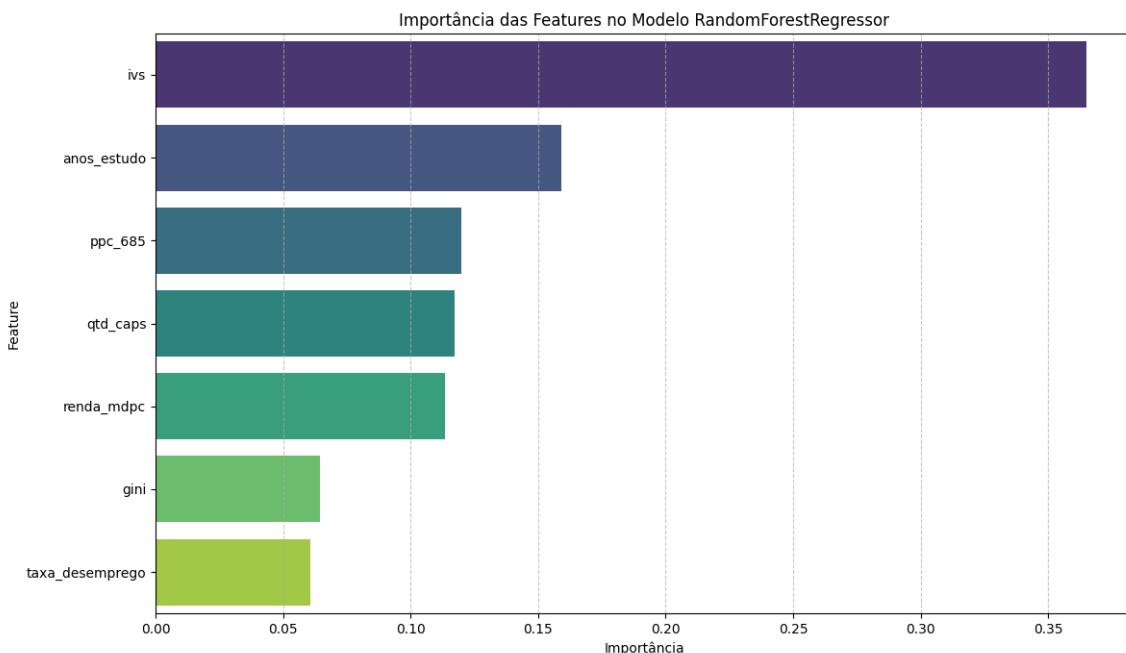


Figura 9. Features mais importantes.

3. Conclusão

Este trabalho dá indícios de que há um recorte socioeconômico relacionado as taxas de suicídio, entretanto um estudo mais aprofundado é necessário para entender as causas e a extensão deste recorte, pois se trata de um tema bastante delicado e com muitos fatores capazes de o influenciar.

Trabalhos futuros podem incluir estudos sobre o tema para identificar melhores variáveis, ampliar o período analisado, neste caso foi de 2014 a 2023. Os dados disponibilizados publicamente contém inconsistências no período da pandemia do COVID-19, prejudicando a base de dados. Uma alteração no escopo do projeto, alterando para nível municipal, também pode ser interessante.

Referências

- [1] **IMPACT OF INCOME INEQUALITY AND OTHER SOCIAL DETERMINANTS ON SUICIDE RATE IN BRAZIL.** Machado, D. B.; Rassella, D.; Santos, D. N.. PLOS ONE, v.10, n.4, 2015. Disponível em: <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0124934&type=printable>. Acesso em: 2 dez. 2025.
- [2] **IPEADATA – Portal de Dados do Ipea.** Ipea, 2025. Disponível em: <http://www.ipeadata.gov.br/Default.aspx>. Acesso em: 2 dez. 2025.
- [3] **DATASUS – Informações de Saúde.** Ministério da Saúde, 2025. Disponível em: <https://datasus.saude.gov.br/>. Acesso em: 2 dez. 2025.
- [4] **PORTAL IBGE. Instituto Brasileiro de Geografia e Estatística,** 2025. Disponível em: <https://www.ibge.gov.br/>. Acesso em: 2 dez. 2025.