

## **Exercício 12: Análise de dados do Bolsa Família**

Com base na necessidade de compreender melhor o impacto e a distribuição dos recursos do programa do novo Bolsa Família, de novembro de 2023, solicito o desenvolvimento de uma análise abrangente dos dados disponíveis.

Esta análise se concentrará na identificação de padrões pelas regiões (Norte, Nordeste, Centro-Oeste, Sudeste e Sul) nos valores das parcelas distribuídas, permitindo uma compreensão mais profunda da distribuição dos recursos.

A análise também precisa considerar outros indicadores para auxiliar nessa compreensão mais profunda acerca do impacto. Sendo assim é importante apresentar o total recebido por região, a média por região, a dispersão e o valor dessa dispersão, ambos por região, além de uma comparação entre o total recebido por cada região e o total geral distribuído pelo programa, tomando como base o seguinte cálculo:

Participação por Região = total recebido por região / total geral (valor pago para todas as regiões)

Os resultados devem ser apresentados em um formato claro e informativo, facilitando a interpretação por parte do Ministro de Desenvolvimento Social e de outros interessados da pasta.

Esperamos que essa análise forneça insights valiosos para o aprimoramento contínuo do programa Bolsa Família e apoie as decisões políticas relacionadas à distribuição de recursos.

Os dados podem ser "baixados" em:

<https://portal.datatransparencia.gov.br/download-de-dados/novo-bolsa-familia>

Prazo para entrega: 16/02/2024

**Dúvidas sobre a solicitação? Fale comigo!**

## Dúvidas técnicas? Utilize os seguintes insumos

1. Estatística e visualização de dados
  - a. Consulte os materiais das aulas 2, 3, 6, 7, 8, 9
2. Análise de dados em larga escala
  - a. Consulte os materias das aulas 9 e 10
  - b. Assista aos vídeos
    - i. Computação e pseudo-computação distribuída (pseudo-paralelismo): <https://youtu.be/SkWzwA1w6R0>
    - ii. Processamento de dados em larga escala: [https://youtu.be/VC9XGcJw\\_Ts](https://youtu.be/VC9XGcJw_Ts)
  - c. PySpark:  
<https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/80/346/605-1>
3. Materias extras de apoio:
  - a. Visualização de dados:
    - i. [https://plotly.com/python-api-reference/generated/plotly.subplots.make\\_subplots.html](https://plotly.com/python-api-reference/generated/plotly.subplots.make_subplots.html)
    - ii. <https://plotly.com/python/graph-objects/>
    - iii. [https://plotly.com/python-api-reference/generated/plotly.graph\\_objects.Figure.html](https://plotly.com/python-api-reference/generated/plotly.graph_objects.Figure.html)
    - iv. <https://plotly.com/python/box-plots/>
    - v. <https://plotly.com/python/table/>
  - b. Estatística
    - i. Biblioteca digital: Matemática e estatísticas (Angela Mitie Otta Kinoshita)
    - ii. Estatística descritiva: <https://youtu.be/ftptOzF1YXU>
  - c. Python e Dataframes
    - i. Relacionamento entre Dataframes:  
<https://youtu.be/s0RnRJzG5Vs?feature=shared>

- ii. Biblioteca digital: Introdução a programação em python  
(Sérgio Furgeri)