

Desafio para o processo seletivo de estágio em análise de dados

Sobre os dados

Os dados utilizados nesse desafio são provenientes de duas fontes públicas mantidas pela Prefeitura de São Paulo: o ObservaSampa e o Geosampa.

O Observatório de Indicadores da Cidade de São Paulo (ObservaSampa) é uma política pública de Estado que tem por objetivo apresentar dados, indicadores e evidências sobre a cidade de São Paulo e suas políticas públicas, por meio de estudos, publicações, dados abertos, representações visuais e outros meios.

O ObservaSampa é gerido pela Coordenadoria de Avaliação e Gestão da Informação (CAGI) da Secretaria Executiva de Planejamento e Entregas Prioritárias (SEPEP), que atua na articulação, promoção e disseminação de informações (dados, indicadores, diagnósticos e avaliações) sobre a realidade da cidade, seus cidadãos e suas políticas públicas, fomentando a cultura de políticas públicas baseadas em evidências na Prefeitura de São Paulo.

O GeoSampa é um portal que agrega as informações geoespaciais da Cidade de São Paulo, disponibilizando-as em mapa, em arquivos para download e com serviços web.

Outras informações como ferramentas de geo, normatizações e os metadados correlatos também são acessadas via portal.

As principais informações cadastrais e territoriais da Cidade, oriundas de diversas Secretarias e órgãos públicos, podem ser acessadas on line em ambiente intranet, auxiliando as atividades de fiscalização, licenciamento, planejamento e gestão.

O GeoSampa é administrado pelo Departamento de Produção e Análise de Informação - DEINFO e foi desenvolvido pela Prodam com tecnologia livre GeoServer/Open Layers. É uma das ferramentas previstas no Projeto SIG-SP, iniciado em dezembro/2013, que visa a gestão corporativa dos dados geoespaciais do município.

Sobre os arquivos

O arquivo `observasampa_familias_extrema_pobreza.csv` contém os dados do indicador "01.01.01 Quantidade de famílias em situação de extrema pobreza (até 1/4 salário mínimo)", disponível no ObservaSampa. Ele possui 4 colunas:

- **distrito**, que contém o nome amigável dos Distritos do município de São Paulo;
- **ano**, que contém o ano de referência do indicador;
- **qtd_famílias**, que contém o valor do indicador para o ano de referência; e

- **ds_nome**, que contém o nome do Distrito no mesmo padrão utilizado pelo cadastro de distritos do Geosampa.

E o arquivo `geosampa_distritos.csv` contém os dados sobre a divisão político-administrativa do município de São Paulo obtidos do portal Geosampa. Esse arquivo possui 3 colunas:

- **ds_nome**, que contém o nome de cada Distrito de São Paulo;
- **ds_sigla**, que contém a sigla de cada Distrito de São Paulo; e
- **ds_subpref**, que contém o nome da Subprefeitura à qual o Distrito pertence.

Instruções

- As soluções devem ser desenvolvidas utilizando Python ou R, nas ferramentas de sua preferência, e as respostas devem ser enviadas para o e-mail babalzano@prefeitura.sp.gov.br.
- Sugerimos fortemente o uso de repositórios públicos (como GitHub ou GitLab) e o envio da URL do repositório. Caso prefiram apenas enviar os arquivos utilizados para solucionar as questões, estes podem ser anexados ao e-mail com as respostas, desde que o e-mail contenha as devidas instruções de uso.
- O teste possui algumas questões com maior dificuldade, então, relaxe, deixar algumas questões em branco não te desclassifica do processo seletivo. Responda o máximo de questões que conseguir.

Questões

1. Quantas famílias em situação de extrema pobreza existiam em São Paulo em 2023?
2. Qual o percentual (em relação ao total da cidade) de famílias em situação de extrema pobreza por Distrito em 2023?
3. Qual distrito apresentou o maior aumento de famílias em situação de pobreza entre 2022 e 2023?
4. Quantas famílias em situação de extrema pobreza existiam em cada Subprefeitura de São Paulo em 2023?
5. Como variou o número de famílias em situação de pobreza entre 2013 e 2023 nos Distritos de Grajaú, Jardim Ângela, Cidade Ademar? Exiba graficamente o resultado, escolhendo a visualização que achar mais adequada.
6. Complete a seguinte frase: "De acordo com os dados do ObservaSampa, em 2023 metade dos Distritos de São Paulo, no máximo, ___ famílias em situação de extrema pobreza". Qual informação você utilizou para completar a frase?