

Recrutamento Data Scientist – DataViz

Teste – Março/2024

Em busca de elementos para tomar decisões estratégicas, um investidor do ramo agrícola quer entender melhor a evolução histórica da produtividade das safras de soja nos estados brasileiros. Além do mais, considerando o ritmo acelerado das mudanças climáticas, ele gostaria de avaliar a influência de fatores climáticos na produtividade da soja. Você ficou responsável por todas as etapas do processo: coleta, tratamento, análise e apresentação dos dados, que devem ser obtidos de fontes públicas.

Considerando o cenário apresentado, pede-se:

1) Faça uma análise histórica da evolução anual dos valores de área plantada, produção e produtividade por estado brasileiro.

Fonte sugerida: arquivo SojaSerieHist.xls em <https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras/itemlist/category/911-soja>

2) Faça um cruzamento dos dados históricos de produtividade de soja com dados históricos climáticos (principalmente temperatura e precipitação). Avalie a existência de correlações entre fatores climáticos e produtividade de soja.

Fonte sugerida: dados históricos de estações do INMET em <https://portal.inmet.gov.br/dadoshistoricos> (os nomes dos arquivos CSV contêm a informação de em qual UF a estação fica)

Dica: alguns arquivos CSV são armazenados utilizando formatos diferentes dos padrões de funções como `read_csv` do pandas. Caso tenha problemas ao abrir os arquivos, experimente alterar as opções de encoding, sep e decimal.

3) Elabore um ou mais painéis demonstrando, de forma resumida, as análises e conclusões dos itens anteriores. Recomenda-se que os painéis contenham elementos variados, como, por exemplo:

- mapas coropléticos dos estados brasileiros, representando algumas das grandezas analisadas
- gráficos de séries temporais
- gráficos de distribuição

Fonte sugerida para os shapefiles dos mapas:

https://geofpt.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/municipio_2022/Brasil/BR/BR_UF_2022.zip

Observação: É desejável que os painéis sejam interativos. Eles podem ser elaborados utilizando a ferramenta ou framework de sua escolha. Sugerimos Tableau, PowerBI ou Plotly-Dash. No entanto, este ponto não é obrigatório. Os resumos podem ser desenvolvidos e apresentados, também, em Jupyter Notebook.

As análises detalhadas devem ser feitas, preferencialmente, utilizando **Python** com **Jupyter Notebook**.