

Vaga: Analista de BI ([6524586](#) | [Analista de BI SR \(gupy.io\)](#))

Desafio de Integração de Dados do IBGE

Objetivo:

Realizar uma análise integrada de dados provenientes das API's do IBGE, abrangendo informações sobre localidade, população, PIB e IDH. Utilizando PySpark, SQL e PowerBI, demonstrar habilidades na análise, modelagem e apresentação visual dos dados. A escolha dos atributos a serem capturados é responsabilidade do candidato, com ênfase na análise da relevância das informações extraídas, transformações, agrupamentos e correlações.

Tarefas do Desafio:

1. Extração de Dados:

- Utilizar as API's do IBGE para obter informações de localidade, população, PIB e IDH.
- Armazenar os dados em um modelo snowflake ou outro sugerido, garantindo integridade e consistência.

2. Transformação de Dados com PySpark e SQL:

- Utilizar PySpark para transformações, assegurando qualidade e relevância.
- Realizar consultas e análises específicas com SQL no conjunto de dados integrado.

3. Modelagem e Documentação:

- Apresentar a modelagem do conjunto de dados, explicando relações entre tabelas e decisões de design adotadas.
- Entregar um notebook PySpark documentando etapas de transformação e análise.

4. Desenvolvimento de Dashboard em PowerBI:

- Criar um dashboard interativo no PowerBI, destacando insights relevantes dos dados integrados.
- O dashboard deve ser intuitivo, com visualizações claras e informativas.

Entregas Esperadas:

- Notebook Python/PySpark documentando extração, transformação e análise.
- Documentação do modelo snowflake ou outro sugerido, explicando relações entre tabelas.
- Arquivo do PowerBI com o dashboard desenvolvido.
- Esquema do fluxo de dados (opcional).
- As entregas devem ser compartilhadas no GitHub do candidato.

Critérios de Avaliação:

- Qualidade e consistência dos dados integrados.
- Eficiência nas transformações com Python, PySpark e SQL.
- Documentação clara e concisa da modelagem e análise.
- Criatividade e relevância das visualizações no PowerBI.

Regras do Desafio:

- Prazo de entrega: 2 dias a partir do recebimento do teste.
- Espera-se um MVP que demonstre valor e resultados para tomada de decisão.
- Se o candidato utilizar ferramentas de inteligência artificial como ChatGPT, deverá explicar como e por que foi usado.