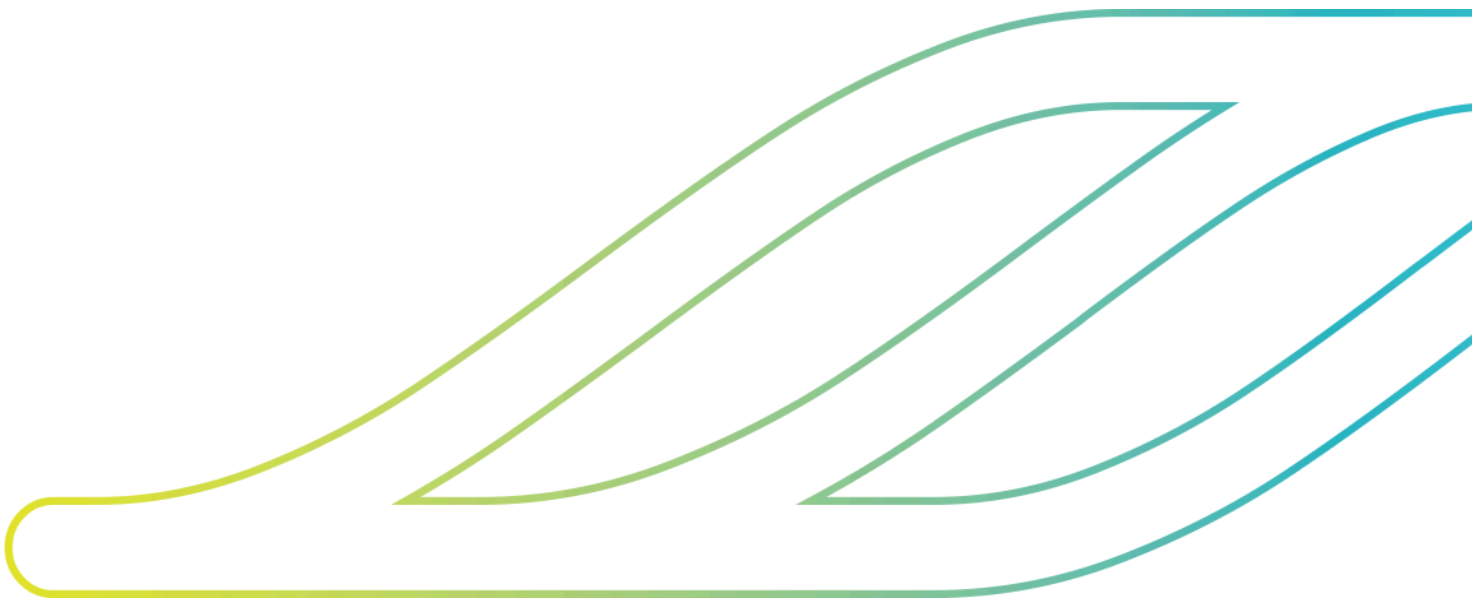


Processo Seletivo - Analytics - Case



Sumário

1. Objetivo	3
2. Introdução	3
3. Data Extract	4
4. Data Processing	4
5. Data Storage	4
6. Data Visualization	4
7. Data Analysis	4
8. Entregas	5

1. Objetivo

Este case tem como objetivo avaliar as habilidades do candidato em Data Engineering, Data Visualization e Data Analysis, áreas essenciais para o sucesso da equipe de Data Analytics da nossa empresa. Através de desafios práticos envolvendo um conjunto de dados real e relevante, o candidato terá a oportunidade de demonstrar sua capacidade de:

- **Preparar dados para análise:** Extrair, transformar e preparar dados para serem utilizados na análise.
- **Criar visualizações de dados impactantes:** Desenvolver visualizações de dados que comuniquem informações valiosas de forma clara, concisa e eficaz.
- **Dominar ferramentas de Data Engineering e Data Visualization:** Utilizar ferramentas como SQL, Python e Tableau para solucionar problemas e gerar insights acionáveis.
- **Analisar dados criticamente:** Identificar padrões, tendências e anomalias nos dados.
- **Comunicar resultados de forma clara e concisa:** Apresentar resultados de forma convincente e engajadora para stakeholders de todos os níveis.

Esperamos que este case seja uma oportunidade para que o candidato demonstre suas habilidades e potencial para se tornar um membro essencial da nossa equipe de Data Analytics.

Lembre-se: O sucesso neste case depende da sua capacidade de navegar pelos dados com maestria, transformar informações em insights acionáveis e comunicar seus resultados de forma impactante.

2. Introdução

Esse case será dividido em 5 partes:

1. **Data Extract:** Você irá realizar uma coleta de dados com python de aerogeradores do ArcGIS do SIGEL/ANEEL com **Web Scraping** com **Python**.
2. **Data Processing:** Você irá realizar o processamento de dados com **geopandas** (python).
3. **Data Storage:** Você salvará os dados em um formato de dados que você preferir para ser utilizado no Tableau Public
4. **Data Visualization:** Você realizará uma conexão de dados com o Tableau Public para criar dashboards, fazer análises e obter insights.
5. **Data Analysis:** Você irá montar uma apresentação rápida no Google Slides (máximo 5 slides) dos principais insights encontrados

Serão levados em consideração principalmente os seguintes aspectos:

- Boas práticas nos Scripts Python:
 - Limpeza
 - Legibilidade
 - Modularidade/Escalabilidade
 - Tipagem
 - Performance
- Tableau (no máximo 1 dashboard/painel)
 - Layout
 - Responsividade
- Apresentação (até 5 slides)
 - Insights retirados no dashboard (evidenciados por prints)
 - Layout dos slides

Deixe os métodos/funções definidos em um ou mais .py e deixar o código executado em um **.ipynb**

3. Data Extract

Você realizará o web scraping de dados a partir da seguinte url :
<https://sigel.aneel.gov.br/arcgis/rest/services/PORTAL/WFS/MapServer/0/query>.

Algumas observações sobre os parâmetros de entrada:

- **Where** : o valor **1=1** é necessário para que a query possa rodar;
- **Out Fields** : o valor ***** é importante para ter a saída de outras colunas além da coluna de geometria
- **Format** : Ajustar o valor que você considerar mais adequado para esse Web Scraping

Para esse passo é necessário usar apenas a biblioteca requests ou httpx.

4. Data Processing

A partir da resposta da request

- Obtenha um GeoDataFrame
- Gere o lat e long da geometria adicionando em novas colunas.
- Faça outros tratamentos que achar necessário.
- Faça validação dos dados

Para esse passo é necessário usar apenas a biblioteca geopandas.

5. Data Storage

A partir do GeoDataFrame, gere um .csv em uma pasta chamada outputs.

O GeoDataFrame já possui um método nativo que realiza essa operação.

6. Data Visualization

Entre no [tableau public](https://public.tableau.com/), e crie um dashboard utilizando o .csv gerado no passo anterior com:

- Filtros na aba superior
- Mapa com os pontos dos aerogeradores
- Gráfico de Scatter Plot
- Gráfico de boxplot
- Gráfico de série temporal com a data de atualização
- Outras sheets/planilhas que achar conveniente

Você pode e deve gerar novamente o .csv caso seja necessário.

7. Data Analysis

Crie um Google Slides com 1 capa simples e no máximo 4 slides com título, com principais insights obtidos e sempre evidenciados por prints

8. Entregas

Deve ser enviado um email com o assunto “PS Analytics - Solução Case” para gabriel.carvalho@casadosventos.com.br, henrique.diogenes@casadosventos.com.br com as seguintes links:

- Repositório git com:
 - os arquivos .py
 - o arquivo .ipynb
 - o arquivo .csv gerado
- Tableau dashboard
- Google Presentation