# Estudo de Caso – PGFN

Este relatório documenta o processo de resolução do estudo de caso proposto pela Procuradoria-Geral da Fazenda Nacional (PGFN), com foco em análise e manipulação de dados utilizando Python e Excel. A proposta foi desenvolvida por Shaka F. Nken, estudante de Análise e Desenvolvimento de Sistemas, com forte interesse em Ciência de Dados e proatividade comprovada em entrevistas e projetos.

## Item 1 – Total de Servidores por Região

Optei por usar Python nessa etapa para lidar com múltiplas abas em uma mesma planilha, algo que ainda não havia experimentado. Meu objetivo era identificar quantas vezes cada região se repete, ou seja, quantos servidores estão alocados por região. O desafio incluiu lidar com:  
- A aba 'Unidade' iniciando a partir da 4ª linha (resolvido com o parâmetro 'skiprows')  
- Espaços nos nomes de colunas (solucionado com renomeação)  
- Valores nulos que atrapalhavam o merge  
- Conversão de tipos de dados para possibilitar o cruzamento entre os DataFrames  
Durante o processo, consultei a documentação oficial do pandas, fóruns como Stack Overflow e usei apoio de IA como ferramenta de aprendizado. Isso demonstra minha curiosidade natural, capacidade de pesquisa e foco em resolver problemas de forma independente.

## Item 2 – Total de Mulheres na PGFN

Neste item, a lógica era simples: contar quantas vezes 'F' aparece na coluna de sexo. Usei Excel com a fórmula CONT.SE e também resolvi em Python. Escolhi o método shape por considerar uma solução mais sólida e flexível em projetos futuros.

=CONT.SE(F2:F4438;"F")

## Item 3 – Filtro de Servidores AGU, EXERC, DESCENT CARREI

Já tinha familiaridade com filtros básicos, mas precisei aprender sobre filtragem de strings em pandas. Utilizei str.contains para buscar padrões textuais em colunas, filtrando os servidores da AGU em exercício com cargo 'DESCENT'. Essa parte demonstrou minha facilidade em aprender com a documentação e aplicar rapidamente.

## Item 4 – Aposentadoria por Idade

Esse item exigiu conhecimento sobre datas no formato serial do Excel. Estudei sobre o padrão de datas (origin='1899-12-30') e utilizei pandas para converter e calcular a idade. Com base na idade e no sexo, criei uma nova coluna indicando se o servidor está aposentável. Esse desafio mostrou minha capacidade de interpretar regras e automatizá-las em código limpo e reutilizável..