



Atividade técnica

Engenharia de Dados

Essa atividade tem o objetivo de **mensurar o conhecimento técnico** para as habilidades desejadas para o trabalho de Engenharia de Dados, como extrair dados de diferentes fontes, lidar com APIs e transformar as bases de dados coletadas.

Você deverá entregar essa atividade para o e-mail que recebeu no formato Jupyter Notebook ou RMarkdown e, a partir dele, gerar um HTML e hospedá-lo no [GitHub Pages](#). Todo o código usado nas respostas precisa estar visível, bem como a tabela resultante, quando o exercício pedir. Se precisar criar pastas no Google Drive, deixe o link visível e dê permissão pública para acesso.

Se você tiver alguma dúvida durante a elaboração da atividade, não hesite em responder ao e-mail.

Exercício 1

Nesse primeiro exercício, vamos importar e transformar uma tabela do **conjunto de dados públicos do BigQuery**. Antes de partir para o código, você precisa ativá-los dentro da sua conta do BigQuery seguindo as instruções [desse guia](#). Esse passo é importante para 1) liberar os datasets no BigQuery; e 2) conectar a eles usando o R usando o pacote [bigrquery](#), ou no Python usando [python-bigquery](#) se aplicável. Fique tranquilo(a), pois usar esses dados não gera cobrança na sua conta.

1. Nos **conjuntos de dados públicos do BigQuery**, dentro do dataset [fda_food](#), faça a importação da tabela **food_events**. Essa tabela reúne mais de 90 mil observações com reações adversas a alimentos. Mostre que a tabela foi importada com sucesso.
2. Transforme a tabela para descobrir as informações abaixo. Em cada uma das perguntas, deixe visível o dataset resultante da transformação necessária.
 1. Qual o **tipo de reação** mais reportado na tabela? Lembre-se que uma linha pode ter mais de uma reação separada por “;”.
 2. Qual o **nome da indústria** que tem **mais mortes** reportadas?
 3. Quais as **3 principais reações adversas** de pessoas entre 18-25 anos na **indústria de cosméticos**?

Exercício 2

A equipe de **relacionamento com doador** vai enviar um novo “mimo” para os nossos apoiadores! Vamos trabalhar com os dados de doadores para construir uma estrutura que permita simplificar o envio desses “mimos”.

Sua tarefa é:

- Usar a API do [Random User Generator](#) para gerar uma lista de **1.000 usuários brasileiros**
 - Faça uma tabela com todas as informações desses usuários

GREENPEACE

- Crie uma coluna que indique o Estado do usuário, baseado em seu telefone celular
 - Se o telefone celular tiver um DDD inexistente remova o usuário da tabela
- Baixe os Pokémons iniciais de todas as gerações por meio da [PokeAPI](#)
 - Atribua um [Pokémon inicial](#) aleatório para cada um dos usuários
- Conecte com o Google Drive e crie uma pasta para cada elemento dos Pokémons
 - Adicione em cada uma das pastas dois arquivos em CSV
 - Cada pasta deve ter dois arquivos um com todos os homens com Pokémons daquele tipo, e outro com todas as mulheres com Pokémons daquele tipo
- Por fim, crie um Google Sheets com um resumo indicando:
 - O número de pessoas por região do Brasil (Norte, Sul, Sudeste, Nordeste, Centro-Oeste) e o elemento de seus Pokemons

Exercício 3

O Greenpeace Brasil utiliza um *data warehouse* baseado em Redshift para reunir nossos bancos de dados. Frequentemente, precisamos revisar e alterar as permissões dos datasets existentes. Partindo da [documentação do Redshift](#), descreva qual o código em SQL seria utilizado para esses diferentes cenários:

1. Cada analista da equipe de Inteligência de Dados tem um perfil próprio e temos uma conta genérica para conseguir lidar com bases de dados coletivas. No cenário em que o usuário `gp_user` gostaria de ter acesso para a tabela `press_data`, qual seria o código para disponibilizá-la?
2. Em algumas vezes, mesmo após dar permissão uma vez, a tabela é reescrita com o parâmetro `drop` e perde as permissões. Qual o código que deve ser feito para que a tabela sempre herde os privilégios de forma padrão?
3. Em um cenário que alguém não é mais responsável por uma tabela específica, qual o código para mudar a pessoa proprietária dessa tabela?