



dev@meantrix.com

Desafio

Sugestões para o desafio:

- Responda ao desafio num prazo máximo de 7 dias (uma semana);
- Envie sua resposta para dev@meantrix.com com o título: **desafio-seu nome** ;
- Sua resposta deverá conter um arquivo .py com os códigos e comentários (se necessários) dessa linguagem, um arquivo .Rmd com os códigos, funções gráficas e conclusões feitos a partir da linguagem R e demais subrotinas criadas todas compactadas em um único arquivo .zip;
- Procure utilizar as melhores práticas de programação e justificar sua análise se achar necessário;

Contextualização

O Base de dados em formato CSV representa um conjunto de empresas do Canadá com a respectiva descrição de seus produtos, dados econômicos e localização.

Assim, podemos caracterizar cada variável:

- name: nome da empresa;
- description: descrição do produto da empresa;
- employees: número de empregados da empresa;
- total_funding: Total de investimento já recebido pela empresa;
- city: cidade;
- subcountry: estado;
- lat: latitude da cidade;
- lng: Longitude da cidade.

O problema:

Deseja-se prospectar empresas que possuam soluções em **tratamento de água** , principalmente, relativas à : **solutions on waste and water, Improve water quality and water efficiency use, water contamination, water for human consumption, water resources** .

Em Python:

Aplice um algoritmo de ML (ou um conjunto deles) capaz de selecionar as principais empresas indicadas para desenvolver a solução de acordo com seu alinhamento com o tema (Justifique a escolha do algoritmo).

Em R:

Faça uma análise exploratória dos resultados acrescentando as demais variáveis contidas no dataset. Quais insights você pode obter a partir desses dados? Quais são as principais cidades (pólos de desenvolvimento) para essa solução? Desenvolva sua resposta em um arquivo .Rmd(Rmarkdown) criando um "mini-relatório".

Hardcore-extra:

Não obrigatório. Se você terminou o desafio de forma rápida, temos mais algumas perguntas para serem respondidas. Elas, como dito, não são obrigatórias, então sinta-se à vontade em não as responder ou até mesmo respondê-las parcialmente. Essa parte visa observar seu entendimento de um ambiente real de produção.

- a) Organize seus códigos em pacotes garantindo seu versionamento e documentação (bibliotecas auxiliares,etc).
- b) Construa testes automatizados para validação do seu pacote.
- c) Crie uma imagem docker capaz de executar suas análises em um ambiente de produção.

Boa Sorte!

