

## Recrutamento Data Scientist

### Teste Janeiro/2021

#### Questão 1

##### Questão de lógica

Na ilha da Tristeza todas as ruas são paralelas. O fundador decidiu que os nomes das N ruas da ilha da Tristeza seriam números que vão de 1 a N.

Com o aumento de casos do coronavírus na ilha, o prefeito resolveu fazer lockdown nas ruas que sejam múltiplas de A e/ou de B por 20 dias para diminuir a disseminação do vírus.

Para os epidemiologistas estimarem o impacto da medida, você deve ajudá-los calculando o número de ruas afetadas pela medida.

Escreva em Python a função:

```
def conta_ruas_lockdown(N, A, B):
```

```
    """
```

```
        Devolve o número de ruas afetadas pelo lockdown.
```

```
    """
```

Exemplos:

```
>>> print(conta_ruas_lockdown(10, 2, 3))
```

```
7
```

```
>>> print(conta_ruas_lockdown(50, 5, 7))
```

```
16
```

```
>>> print(conta_ruas_lockdown(1000000, 28, 32))
```

```
62500
```

**Obs:** Busque desenvolver o algoritmo que tenha a melhor performance possível

## Questão 2

Para responder esta questão você utilizará o dataset no formato CSV recebido juntamente com este teste.

O dataset contém medições de TCH (Toneladas de Cana de açúcar por Hectare) para um conjunto de talhões de plantação de cana de açúcar e para diferentes safras de colheita desses talhões.

Para o desenvolvimento das análises que serão propostas, descrevemos abaixo o significado de cada coluna do dataset:

- *bloco*: identificador da região de plantação e colheita de cana de açúcar
- *talhao*: identificador da sub-região de plantação e colheita de cana de açúcar (um bloco contém vários talhões)
- *safra*: ano que a cana de açúcar foi colhida
- *area*: tamanho da área colhida em hectares
- *data\_colheita*: data em que a cana de açúcar foi colhida
- *tch\_bloco*: Toneladas de Cana por Hectare colhida referente ao bloco
- *NDVI\_b01*: NDVI (Normalized De Vegetation Index), e o “b01” corresponde ao índice médio relativo ao primeiro mês antes da colheita
- *NDVI\_bN*: NDVI médio relativo ao N-ésimo mês antes da colheita

Análises propostas:

- a) Faça uma rápida análise exploratória desse dataset, por exemplo análises estatísticas das variáveis, gráficos de distribuição, etc.
- b) Desenvolva um modelo capaz de prever o TCH dos blocos do dataset para a safra de 2020 (atenção que a previsão deve ser no nível de blocos e não de talhões)
- c) Qual o erro esperado do modelo para a safra de 2020? Utilize a métrica de erro que você julgar necessária, podendo ser utilizada mais de uma