



**4intelligence**

Scalable  
Decision  
Science

# PROVA DE CIÊNCIA DE DADOS

---

## Instruções

1. O candidato deve nos enviar o(s) código(s) via github, bem como todas as bases adicionais utilizadas para que possamos reproduzir os resultados.
2. A reprodução completa do código faz parte da avaliação. Devemos ser capazes de reproduzir qualquer resultado que você queira nos mostrar.
3. A qualidade visual do documento gerado, bem como o formato geral da apresentação dos códigos, também serão levados em consideração.
4. Os arquivos de input enviados por nós não podem conter nenhuma alteração. Todas as modificações têm que ser feitas em código. Assim, não serão aceitos scripts que operem sobre uma base diferente da versão enviada ao candidato.
5. O código tem que ser auto-contido, ou seja, ao ser executado deve chegar aos resultados a que se propôs.
6. Dê preferência para entregar os resultados em Rmarkdown ou Jupyter (via git), mas sinta-se à vontade para desenvolver o trabalho em qualquer linguagem.
7. O teste deve ser feito individualmente, sem auxílio de terceiros.
8. Fique à vontade para exercer a sua criatividade! Ir além e nos surpreender será considerado um diferencial!
9. As provas selecionadas devem ser apresentadas na entrevista técnica, quando avaliaremos a capacidade de síntese, organização e comunicação.



### Questão 1

Análise Descritiva: Em anexo, você recebeu uma base de dados (Bases Final ADS Jun2021) com o consumo de energia residencial, comercial e industrial de cada região brasileira. Faça uma análise descritiva das variáveis e, eventualmente, da relação entre elas.

### Questão 2

Modelagem: Utilizando-se das variáveis fornecidas na base de dados Bases Final ADS Jun2021.xlsx, forneça um modelo que projete, com a melhor acurácia possível, o consumo de energia industrial da região Sudeste para os próximos 24 meses.

1. Explique o método e a razão de utilizar a abordagem escolhida na sua projeção. Quais “insights” podem ser obtidos da modelagem?
2. Forneça medidas para avaliar a qualidade da projeção do modelo
3. Justifique a escolha das variáveis explicativas e avalie o poder explicativo delas.

### Questão 3

Levando em consideração a modelagem apresentada acima, escolha os 5 melhores modelos em termos de acurácia e argumente a razão de tê-los escolhido.

### Questão 4

O que é possível tirar de conclusões a partir dos exercícios 1, 2, e 3?

