

158 dias restantes. [Assine agora!](https://checkout.igti.com.br/assine/?=nome_completo_conforme_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=(11) 96871-9014) (https://checkout.igti.com.br/assine/?=nome\_completo\_conforme\_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=(11) 96871-9014)

# Desafio do Módulo 2

**Entrega** 26 jun em 21:00      **Pontos** 40      **Perguntas** 15  
**Disponível** até 26 jun em 21:00      **Limite de tempo** Nenhum

## Instruções

O Desafio do Módulo 2 está disponível!

### 1. Instruções para realizar o desafio

Consulte a data de entrega no teste e em seu calendário.

Reserve um tempo para realizar a atividade, leia as orientações e enunciados com atenção. Em caso de dúvidas utilize o "Fórum de dúvidas do Desafio do Módulo 2".

Para iniciá-lo clique em "Fazer teste". Você tem somente **uma** tentativa e não há limite de tempo definido para realizá-lo. Caso precise interromper a atividade, apenas deixe a página e, ao retornar, clique em "Retomar teste".

Clique em "Enviar teste" **somente** quando você concluí-lo. Antes de enviar confira todas as questões.

O gabarito será disponibilizado partir de segunda-feira, **26/06/2022**, às 23h59.

Bons estudos!

### 2. O arquivo abaixo contém o enunciado do desafio

[Enunciado do Desafio - Modulo 2 - Bootcamp Cientista de Dados.pdf](#)

## Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	<a href="#">Tentativa 1</a>	7 minutos	34,68 de 40

⚠ As respostas corretas estarão disponíveis em 26 jun em 23:59.

Pontuação deste teste: **34,68** de 40

Enviado 25 jun em 13:02

Esta tentativa levou 7 minutos.

158 dias restantes. **Assine agora!** ([https://checkout.igti.com.br/assine/?=nome\\_completo\\_conforme\\_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=\(11\) 96871-9014](https://checkout.igti.com.br/assine/?=nome_completo_conforme_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=(11) 96871-9014))

Quantos registros existem no arquivo?

- ☐ 91
- ☐ 7410
- ☐ 5110
- ☒ 67135

## Pergunta 2

2,66 / 2,66 pts

Quantas colunas existem no arquivo? Quantas são numéricas? Ao ler o arquivo com `spark.read.csv`, habilite `inferSchema=True`. Use a função `printSchema()` da API de Dataframes.

- ☐ 10 e 5
- ☐ 10 e 6
- ☐ 12 e 5
- ☒ 12 e 7

## Pergunta 3

2,66 / 2,66 pts

No conjunto de dados, quantos pacientes sofreram e não sofreram derrame (stroke), respectivamente?

- ☐ 26848 e 40287

158 dias restantes. **Assine agora!** ([https://checkout.igti.com.br/assine/?=nome\\_completo\\_conforme\\_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=\(11\) 96871-9014](https://checkout.igti.com.br/assine/?=nome_completo_conforme_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=(11) 96871-9014))

☐ 40200 e 26841

☒ 40287 e 26848

#### Pergunta 4

2,66 / 2,66 pts

A partir do dataframe, crie uma tabela temporária usando `df.createOrReplaceTempView('table')` e a seguir use `spark.sql` para escrever uma consulta SQL que obtenha quantos pacientes tiveram derrame por tipo de trabalho (`work_type`). Quantos pacientes sofreram derrame e trabalhavam respectivamente, no setor privado, de forma independente, no governo e quantas são crianças?

☐ 23711, 10807, 520, 5164.

☒ 23711, 10807, 5164, 520.

☐ 520, 10807, 5164, 23711.

☐ 10807, 23711, 5164, 520.

#### Pergunta 5

2,66 / 2,66 pts

Escreva uma consulta com `spark.sql` para determinar a proporção, por gênero, de participantes do estudo. A maioria dos participantes é:

☐ outro gênero

☐ masculina

☐ homens e mulheres foram estudados na mesma proporção.

158 dias restantes. **Assine agora!** ([https://checkout.igti.com.br/assine/?=nome\\_completo\\_conforme\\_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=\(11\) 96871-9014](https://checkout.igti.com.br/assine/?=nome_completo_conforme_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=(11) 96871-9014))

Incorreta

### Pergunta 6

0 / 2,66 pts

Escreva uma consulta com *spark.sql* para determinar quem tem mais probabilidade de sofrer derrame: hipertensos ou não-hipertensos. Você pode escrever uma consulta para cada grupo. A partir das probabilidades que você obteve, você conclui que:

- ☐ Não é possível extrair esta informação do conjunto de dados.
- ☒ A hipertensão, neste conjunto de dados, diminui a probabilidade de derrame.
- ☐ A hipertensão, neste conjunto de dados, aumenta a probabilidade de derrame.
- ☐ As probabilidades de derrame não são afetadas por este atributo.

### Pergunta 7

2,66 / 2,66 pts

Escreva uma consulta com *spark.sql* que determine o número de pessoas que sofreram derrame por idade. Com qual idade o maior número de pessoas do conjunto de dados sofreu derrame?

- ☐ 80
- ☐ 81
- ☒ 79
- ☐ 78

**Pergunta 8****2,66 / 2,66 pts**

Usando a API de dataframes, determine quantas pessoas sofreram derrames após os 50 anos.

☐ 30189☒ 28938☐ 26121☐ 31091**Pergunta 9****2,66 / 2,66 pts**

Usando *spark.sql*, determine qual o nível médio de glicose para pessoas que, respectivamente, sofreram e não sofreram derrame.

☐ 136 e 130☐ 128 e 100☒ 119 e 103☐ 120 e 91**Pergunta 10****2,66 / 2,66 pts**

Qual é o BMI (IMC = índice de massa corpórea) médio de quem sofreu e não sofreu derrame?

158 dias restantes. **Assine agora!** ([https://checkout.igti.com.br/assine/?=nome\\_completo\\_conforme\\_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=\(11\) 96871-9014](https://checkout.igti.com.br/assine/?=nome_completo_conforme_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=(11) 96871-9014))

☐ 30,93 e 29,31

☒ 29,94 e 27,99

☐ 28,10 e 27,18

### Pergunta 11

2,66 / 2,66 pts

Crie um modelo de árvore de decisão que prevê a chance de derrame (stroke) a partir das variáveis contínuas/categóricas: idade, BMI, hipertensão, doença do coração, nível médio de glicose. Use o conteúdo da segunda aula interativa para criar e avaliar o modelo.

Qual a acurácia de um modelo construído?

☐ Menor que 10%

☒ Menor que 75%

☐ Menor que 30%

☐ Menor que 50%

### Pergunta 12

2,66 / 2,66 pts

Adicione ao modelo as variáveis categóricas: gênero e status de fumante. Use o conteúdo da aula interativa para lidar com as variáveis categóricas. A acurácia (qualidade) do modelo aumentou para:

☒ Acima de 80%

158 dias restantes. **Assine agora!** ([https://checkout.igti.com.br/assine/?=nome\\_completo\\_conforme\\_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=\(11\) 96871-9014](https://checkout.igti.com.br/assine/?=nome_completo_conforme_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=(11) 96871-9014))

☐ Acima de 60%

☐ Acima de 95%

Incorreta

### Pergunta 13

0 / 2,66 pts

Qual dessas variáveis é mais importante no modelo de árvore de decisão que você construiu na questão (12)?

- ☒ Ter hipertensão.
- ☐ Status sobre fumo.
- ☐ Nível médio de glicose.
- ☐ Ter doença cardíaca.

### Pergunta 14

2,66 / 2,66 pts

Qual a profundidade da árvore de decisão da questão (12)?

- ☐ Maior que 5.
- ☐ Menor que 2.
- ☒ Entre 2 e 7.
- ☐ Menor que 1.

### Pergunta 15

2,76 / 2,76 pts

158 dias restantes. [Assine agora!](https://checkout.igti.com.br/assine/?=nome_completo_conforme_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=(11) 96871-9014) (https://checkout.igti.com.br/assine/?=nome\_completo\_conforme\_cpf=Priscila&email=priscila.palazzo@centralnacionalunimed.com.br&cpf=076.570.396-39&telefone=(11) 96871-9014)

Quantos nodos a árvore de decisão possui?

- ☐ 1.
- ☒ Mais que 5.
- ☐ 3.
- ☐ 2.

Pontuação do teste: **34,68** de 40