


# DESM2 - Desafio do Módulo 2

Iniciado: 12 ago em 20:31

## Instruções do teste



**Desafio do Módulo**

É hora de mostrar que você aprendeu!  
O Desafio é um teste com 15 questões objetivas e **deve ser realizado com atenção!**

Reserve um tempo para realizar a atividade, leia as orientações e enunciados com atenção. Em caso de dúvidas utilize o Fórum de Dúvidas.

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. E Caso o teste seja iniciado, e não enviado até o final do prazo de entrega, a plataforma enviará a tentativa não finalizada automaticamente, independentemente do progresso no teste. Fique atento(a) ao seu teste e ao prazo final, pois novas tentativas só serão concedidas devido às questões médicas.

O gabarito será disponibilizado a partir de sábado, **17/08/2024**, às 23h59.

- O arquivo abaixo contém o enunciado do Desafio. Confira agora:

**Enunciado do Desafio - Módulo 2 - Bootcamp Cientista de Dados.pdf**

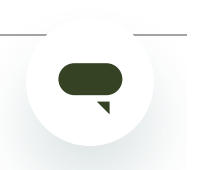
(<https://online.igti.com.br/courses/7642/files/582860?wrap=1>)\_ ↓

([https://online.igti.com.br/courses/7642/files/582860/download?download\\_frd=1](https://online.igti.com.br/courses/7642/files/582860/download?download_frd=1))

**Bons estudos!**

**Atenciosamente,**

**Equipe XP Educação**



**Pergunta 1****3 pts**

Quantos registros existem no arquivo?

- ☐ 91
- ☐ 5110
- ☒ 67135
- ☐ 7410

**Pergunta 2****3 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.

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

**Pergunta 3****3 pts**

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

- ☐ 40200 e 26841
- ☐ 26848 e 40287
- ☐ 26841 e 40287

☒ 40287 e 26848

#### Pergunta 4

3 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, 5164, 520.
- ☐ 23711, 10807, 520, 5164.
- ☐ 520, 10807, 5164, 23711.
- ☐ 10807, 23711, 5164, 520.

#### Pergunta 5

3 pts

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

- ☒ feminina
- ☐ homens e mulheres foram estudados na mesma proporção.
- ☐ masculina
- ☐ outro gênero

#### Pergunta 6

3 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:

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

### Pergunta 7

3 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
- ☐ 78
- ☒ 79
- ☐ 81

### Pergunta 8

3 pts

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

- ☐ 26121
- ☐ 31091
- ☒ 28938

☐ 30189

### Pergunta 9

3 pts

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

☐ 136 e 130

☒ 119 e 103

☐ 120 e 91

☐ 128 e 100

### Pergunta 10

3 pts

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

☐ 30,93 e 29,31

☐ 28,90 e 28,62

☐ 28,10 e 27,18

☒ 29,94 e 27,99

### Pergunta 11

3 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 30%
- ☒ Menor que 75%
- ☐ Menor que 50%
- ☐ Menor que 10%

### Pergunta 12

3 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 60%
- ☐ Acima de 70%
- ☐ Acima de 95%
- ☒ Acima de 80%

### Pergunta 13

3 pts

Adicione ao modelo as variáveis categóricas: gênero e status de fumante. Use o conteúdo da aula interativa para lidar com as variáveis categóricas. Qual dessas variáveis é mais importante no modelo de árvore de decisão que você construiu?

- ☐ Ter doença cardíaca.
- ☒ Nível médio de glicose.

- ☐ Ter hipertensão.
- ☐ Status sobre fumo.

**Pergunta 14****3 pts**

Adicione ao modelo as variáveis categóricas: gênero e status de fumante. Use o conteúdo da aula interativa para lidar com as variáveis categóricas. Qual a profundidade da árvore de decisão?

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

**Pergunta 15****3 pts**

Quantos nodos a árvore de decisão possui?

- ☒ Mais que 5.
- ☐ 1.
- ☐ 3.
- ☐ 2.

Salvo em 7:51

Enviar teste