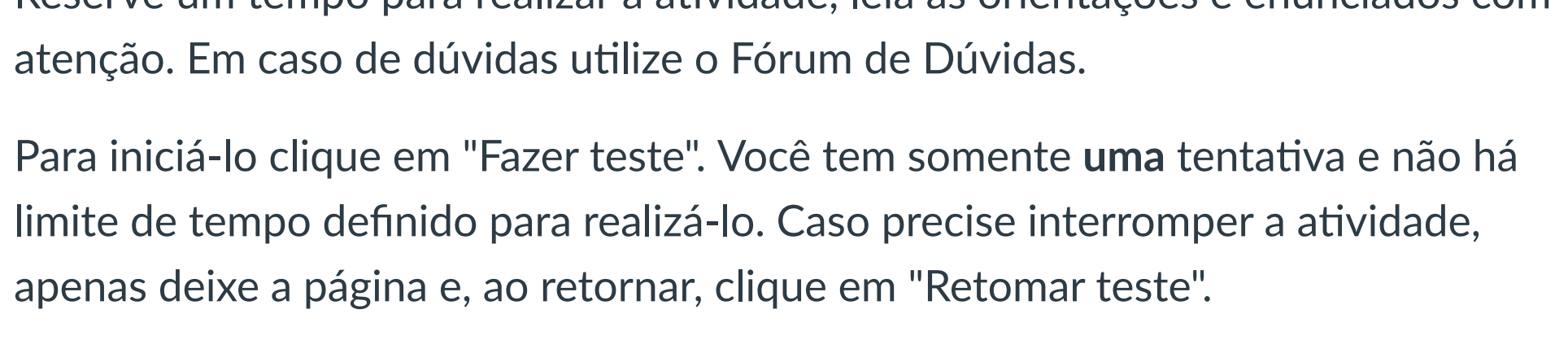


Entrega	22 jun em 19:00	Pontos	45	P
---------	-----------------	--------	----	---

Instruções


É hora de mostrar
O Desafio é um te



Clique em "Enviar teste" somente quando você concluir 10. 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,

O gabarito será disponibilizado a partir de sexta-feira, **21/06/2024**, às 23h59.

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

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

Bons estudos:

Atenciosamente

- Fazer o teste novamente

	Tentativa	Tempo	Pontuação
MAIS RECENTE	Tentativa 1	438 minutos	42 de 42

5

Pontuação des

Esta tentativa levou

Quantos registros ex	

Pergunta 2

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

☐ 12 e 5

☐ 10 e 6

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

☐ 40200 e 26841

☒ 40287 e 26848

☐ 26841 e 40287

☐ 26848 e 40287

Pergunta 4 3 / 3 pontos

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?

☐ 520, 10807, 5164, 23711.

☐ 10807, 23711, 5164, 520.

☐ 23711, 10807, 520, 5164.

Pergunta 5 3 / 3 pts

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

☒ feminina

☐ homens e mulheres foram estudados na mesma proporção.

☐ masculina

☐ outro gênero

Pergunta 6

3 / 3 pontos

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

☐ A hipertensão, neste conjunto de dados, diminui a probabilidade de derrame.

☐ As probabilidades de derrame não são afetadas por este atributo.

Pergunta 7 3 / 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

☒ 79

☐ 81

☐ 78

Pergunta 8

3 / 3 pts

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

☒ 28938

☐ 26121

☐ 30189

☐ 31091

Pergunta 9 3 / 3 pts

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

☐ 136 e 130

120 e 91

119 e 103

Pergunta 10

3 / 3 pts

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

29,94 e 27,99

28,90 e 28,62

30,93 e 29,31

28,10 e 27,18

Pergunta 11

3 / 3 pts

Crie um modelo de árvore de decisão que prevê a chance de derrame (stroke) a partir das variáveis contínuas/catóricas: 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 50%

☒ Menor que 75%

☐ Menor que 30%

Pergunta 12 0 / 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 95%

☐ Acima de 80%

☐ Acima de 60%

☒ Acima de 70%

Pergunta 13

3 / 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 hipertensão.

☐ Ter doença cardíaca.

☒ Status sobre fumo.

ção?

☐ Menor que 1.

