# $\mathbb{Z}_{8}$

Página inicial **Notas** 

## DESM2 - Desafio do Módulo 2

Entrega 22 jun em 19:00 Pontos 45 Perguntas 15 Disponível até 22 jun em 19:00 Limite de tempo Nenhum Tentativas permitidas 2 Instruções

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 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, **Equipe XP Educação** 

Fazer o teste novamente

### Histórico de tentativas **Tentativa** Tempo Pontuação

**MAIS RECENTE** Tentativa 1 438 minutos 42 de 45 (!) As respostas corretas estarão disponíveis em 21 jun em 23:59.

Pontuação desta tentativa: 42 de 45 Enviado 13 jun em 16:02 Esta tentativa levou 438 minutos.

Quantos registros existem no arquivo?

Pergunta 1

Dataframes.

12 e 7

0 10 e 5

12 e 5

0 10 e 6

Pergunta 4

são crianças?

proporção.

masculina

outro gênero

Pergunta 6

67135 7410

3 / 3 pts

91 5110

3 / 3 pts 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

3 / 3 pts Pergunta 3 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

A partir do dataframe, crie uma tabela temporária

use spark.sql para escrever uma consulta SQL que

trabalho (work\_type). Quantos pacientes sofreram

derrame e trabalhavam respectivamente, no setor

520, 10807, 5164, 23711.

23711, 10807, 5164, 520.

usando df.createOrReplaceTempView('table') e a seguir

obtenha quantos pacientes tiveram derrame por tipo de

privado, de forma independente, no governo e quantas

3 / 3 pts

Próximo ▶

438 minutos

42 de 45

42 de 45

0 10807, 23711, 5164, 520. 23711, 10807, 520, 5164. 3 / 3 pts Pergunta 5 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

A hipertensão, neste conjunto de dados, aumenta a probabilidade de derrame. Não é possível extrair esta informação do conjunto de dados. A hipertensão, neste conjunto de dados, diminui a probabilidade de derrame. As probabilidades de derrame não são afetadas por este atributo. 3 / 3 pts Pergunta 7

Escreva uma consulta com spark.sql que determine o

número de pessoas que sofreram derrame por idade.

Usando a API de dataframes, determine quantas

pessoas sofreram derrames após os 50 anos.

de dados sofreu derrame?

08

• 79

81

78

Pergunta 8

28938

O 26121

30189

31091

Pergunta 10

29,94 e 27,99

28,90 e 28,62

30,93 e 29,31

28,10 e 27,18

Pergunta 11

Com qual idade o maior número de pessoas do conjunto

Escreva uma consulta com spark.sql para determinar

hipertensos ou não-hipertensos. Você pode escrever

probabilidades que você obteve, você conclui que:

quem tem mais probabilidade de sofrer derrame:

uma consulta para cada grupo. A partir das

3 / 3 pts Pergunta 9 Usando spark.sql, determine qual o nível médio de glicose para pessoas que, respectivamente, sofreram e não sofreram derrame. 136 e 130 128 e 100 120 e 91 • 119 e 103

Qual é o BMI (IMC = índice de massa corpórea) médio

Crie um modelo de árvore de decisão que prevê a

chance de derrame (stroke) a partir das variáveis

Qual a acurácia de um modelo construído?

Menor que 50%

Menor que 10%

Acima de 70%

decisão que você construiu?

Nível médio de glicose.

Ter hipertensão.

Ter doença cardíaca.

Status sobre fumo.

Pergunta 14

Pergunta 15

2.

Pergunta 13

Incorreta

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.

de quem sofreu e não sofreu derrame?

• Menor que 75% Menor que 30% 0 / 3 pts Pergunta 12 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%

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

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? Maior que 5. Menor que 2. Menor que 1. • Entre 2 e 7.

0 1. 3. Mais que 5. Pontuação do teste: 42 de 45

Quantos nodos a árvore de decisão possui?

Pontuação mantida: Mais 1 tentativa disponível Fazer o teste novamente (Será mantida sua pontuação mais alta)

Anterior

Tempo:

Pontuação atual:

Detalhes da última tentativa: