

Questão 1

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Leia os dados com erro (iris-with-errors.csv), faça a limpeza e remova as duas últimas colunas. Quantas e colunas sobraram nos dados?

- ☐ a. 12 linhas e 5 colunas
- ☐ b. 12 linhas e 4 colunas
- ☐ c. 25 linhas e 5 colunas
- ☐ d. 14 linhas e 3 colunas
- ☐ e. 10 linhas e 4 colunas

Questão 2

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Leia os dados com erro novamente e insira a media de cada atributo onde for encontrada NaN e '?'. Quais os valores das medianas obtidas para os atributos?

- ☐ a. [5,05; 5,45; 1,40; 2,2]
- ☐ b. [4,05; 2,45; 1,40; 0,2]
- ☐ c. [5,05; 2,45; 0,40; 0,2]
- ☐ d. [5,05; 3,45; 1,40; 0,2]
- ☐ e. [4,05; 1,25; 2,40; 0,2]

Questão 3

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Monte um boxplot para cada variável dos dados da Iris. Quais atributos apresentam outliers?

- ☐ a. sepal_length, sepal_width
- ☐ b. sepal_length, sepal_width, petal_length
- ☐ c. sepal_width, petal_length
- ☐ d. sepal_length, petal_length
- ☐ e. Todos.

Questão 4

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Considere a base de dados do Titanic, disponível na plataforma Kaggle. Calcule a correlação entre as variáveis. Quais são as três variáveis mais correlacionadas?

Para ler os dados, considere o código a seguir:

```
import pandas as pd
import numpy as np
import requests
from io import StringIO
```

```
# URL da base de dados no Kaggle
url = "https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv"
```

```
# Fazer a requisição à URL e obter o conteúdo
response = requests.get(url)
data = response.text
```

```
# Criar um DataFrame a partir dos dados
df = pd.read_csv(StringIO(data))
```

- ☐ a. Fare e Sex; Cabin e SibSp
- ☐ b. SibSp e Parch; Fare e Survived
- ☐ c. Parch e Pclass; Sex e Name
- ☐ d. Fare e Ticket; Age e SibSp
- ☐ e. Fare e Pclass; Embarked e Name

Questão **5**

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙️ Editar questão

Qual das seguintes alternativas descreve melhor o objetivo principal da limpeza de dados em um processo de análise ou modelagem?

- ☐ a. Introduzir ruído nos dados para torná-los mais realistas.
- ☐ b. Reduzir a dimensionalidade dos dados para economizar espaço de armazenamento.
- ☐ c. Aumentar artificialmente o tamanho do conjunto de dados.
- ☐ d. Garantir que os dados estejam livres de erros, inconsistências e valores inválidos.
- ☐ e. Ocultar informações sensíveis para proteger a privacidade dos usuários.