

Iniciado em quinta, 16 abr 2020, 01:46

Estado Finalizada

Concluída em quinta, 16 abr 2020, 02:28

Tempo empregado 42 minutos 8 segundos

Questão **1**

Completo

Vale 2,00 ponto(s).

Considere a base de dados disponível em [TACTD-01-Base.csv](#), e analise os dados faltantes no atributo 'Sex', selecionando as linhas que não sejam 'F' nem 'M'. O tipo de dados faltantes é:

Escolha uma:

- ☐ a. (Condicionalmente) aleatório (MAR)
- ☒ b. Completamente aleatório (MCAR)
- ☐ c. Não aleatório
- ☐ d. Uma mistura de não aleatórios com MAR

Questão **2**

Completo

Vale 2,00 ponto(s).

Considere a base de dados disponível em [TACTD-01-Base.csv](#), e analise os dados no atributo 'Religion'. Esse atributo possui os seguintes problemas:

Escolha uma:

- ☐ a. 5 outliers
- ☐ b. 5 dados errôneos, sendo desses 2 outliers
- ☐ c. 5 dados faltantes, sendo desses 2 dados errôneos
- ☒ d. 3 dados faltantes devido a omissão do usuário

Questão **3**

Completo

Vale 1,00 ponto(s).

Considere a base de dados disponível em [TACTD-01-Base.csv](#) . Encontre e remova todos os dados errôneos no atributo 'Religion'. Após remover os dados, analise todos as linhas relativas a dados faltantes nesse mesmo atributo. Esses dados faltantes:

Escolha uma:

- ☒ a. Podem ser tanto do tipo completamente aleatório (MCAR) quanto não-aleatórios, portanto é mais seguro considerá-los não aleatórios
- ☐ b. Podem ser tanto (condicionalmente) aleatórios (MAR) quanto completamente aleatórios, portanto é mais seguro considerá-los do tipo MAR
- ☐ c. São seguramente completamente aleatórios (MCAR)
- ☐ d. São seguramente não aleatórios

Questão **4**

Completo

Vale 1,00 ponto(s).

Dados estruturados, quando comparados aos não-estruturados, são caracterizados por:

Escolha uma:

- ☐ a. estar armazenado em formato binário ou de texto livre, e não permitir controle de versões
- ☐ b. possuir formato bem definido para armazenamento, sendo inviável executar validação dos dados
- ☐ c. possibilitar maior liberdade/flexibilidade na coleta dos dados, sendo ideais para coleta com baixo controle
- ☒ d. possuir campos e atributos que permitam endereçar os dados, e maior custo de manutenção

Questão **5**

Completo

Vale 2,00 ponto(s).

Considere a base de dados disponível em [TACTD-01-Base.csv](#) , e analise os dados no atributo 'Years of experience' por meio do gráfico de caixa e depois visualizando as linhas contendo outliers mostrados no gráfico de caixa. O atributo possui:

Escolha uma:

- ☐ a. nenhum outlier, pois todos os dados são plausíveis
- ☐ b. 2 outliers globais, sendo um deles também um outlier contextual
- ☒ c. 1 outlier global, e 1 outlier contextual
- ☐ d. 2 outliers contextuais, sendo um deles também um outlier global

Gere o gráfico de caixa do atributo 'Year of Birth' e depois selecione e imprima na tela todas as linhas em que o atributo 'Year of Birth' possui valores fora das barras de mínimo e máximo do gráfico de caixa. Como interpretar as linhas retornadas e quais ações são mais indicadas, em termos de manter ou remover os dados suspeitos?

Escolha uma:

- ☐ a. a linha 67 e a linha 22 possuem valores outliers, mas não é recomendado removê-los
- ☐ b. a linha 22 e a linha 67 possuem outliers e dados errôneos, todos esses valores devem ser removidos
- ☒ c. a linha 22 possui valores outliers, mas são plausíveis e podem ser mantidos; a linha 67 possui outliers e dados errôneos, os quais devem ser removidos
- ☐ d. a linha 22 possui outliers e dados errôneos, os quais devem ser removidos; a linha 67 possui outliers, mas são plausíveis e podem ser mantidos

[◀ Exercícios - solucoes.pdf](#)

Seguir para...



[Avaliacao.ipynb ▶](#)