Desafio do Módulo 2

Entrega 16 nov em 21:00 **Pontos** 40 **Perguntas** 15
Disponível até 16 nov em 21:00 **Limite de tempo** Nenhum

Instruções

O Desafio do Módulo 2 está disponível!

1. Instruções para realizar o desafio

Consulte a data de entrega no teste e em seu calendário.

Reserve um tempo para realizar a atividade, leia as orientações e enunciados com atenção. Em caso de dúvidas utilize o "Fórum de dúvidas do Desafio do Módulo 2".

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.

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

O gabarito será disponibilizado partir de sexta-feira, **19/11/2021,** às 23h59.

Bons estudos!

2. O arquivo abaixo contém o enunciado do desafio

Enunciado do Desafio – Módulo 2 – Bootcamp Engenheiro(a) de dados.pdf

Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	Tentativa 1	5.055 minutos	34,66 de 40

ⓘ As respostas corretas estarão disponíveis em 19 nov em 23:59.

Pontuação deste teste: **34,66** de 40
Enviado 16 nov em 0:10
Esta tentativa levou 5.055 minutos.

Pergunta 1

2,67 / 2,67 pts

Analise as asserções a seguir:

I) A limpeza dos dados é um processo obrigatório na análise de dados

PORQUE

II) Os dados brutos são naturalmente sujos.

- As assertivas são falsas.
- As assertivas I e II são verdadeiras, mas a II não justifica a I.
- **As assertivas I e II são verdadeiras, sendo que a II justifica a I.**
- A assertiva I é falsa, mas a II é verdadeira..

Pergunta 2

2,67 / 2,67 pts

A partir da tabela abaixo, um analista de dados fez as seguintes asserções:

Sexo	Idade	Renda	Cidade	Classificação
M	50	abaixo de 5000	BH	Adimplente
M	50	abaixo de 5000	BH	Inadimplente

I) Existe um ruído de classe na base de dados

PORQUE

II) O atributo alvo da base é contraditório e os valores dos atributos explicativos são idênticos.

A assertiva I é falsa, mas a II é verdadeira.

As assertivas I e II são verdadeiras, sendo que a II justifica a I.

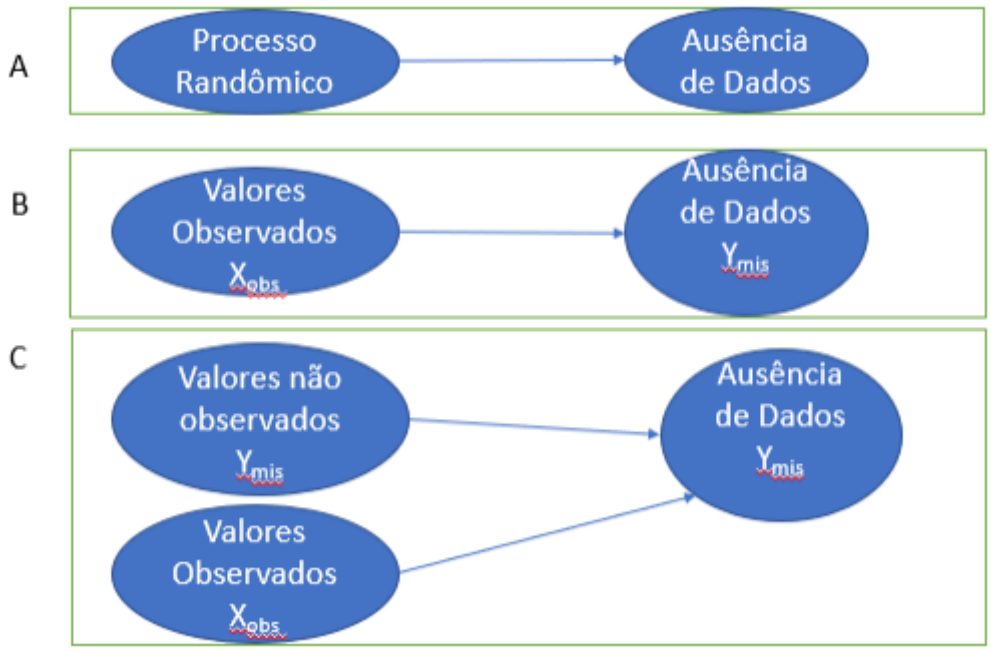
As assertivas I e II são verdadeiras, mas a II não justifica a I.

As assertivas são falsas.

Pergunta 3

2,67 / 2,67 pts

A imagem abaixo representa os três mecanismos de produção de dados ausentes em uma base de dados. As letras A, B e C representam respectivamente os mecanismos:



MCAR, MNAR e MAR.

MAR, MCAR e MNAR.

MAR, MNAR e MCAR.

MCAR, MAR e MNAR.

Pergunta 4

2,67 / 2,67 pts

Sobre a imputação de valores em dados ausentes, considere as seguintes afirmações:

I) A imputação pode ser realizada através de algoritmos de aprendizado de máquinas.

II) A imputação simples pode ser realizada através da média, mediana ou moda, dependendo do tipo de atributo.

III) A imputação EM é baseada na média das múltiplas imputações.

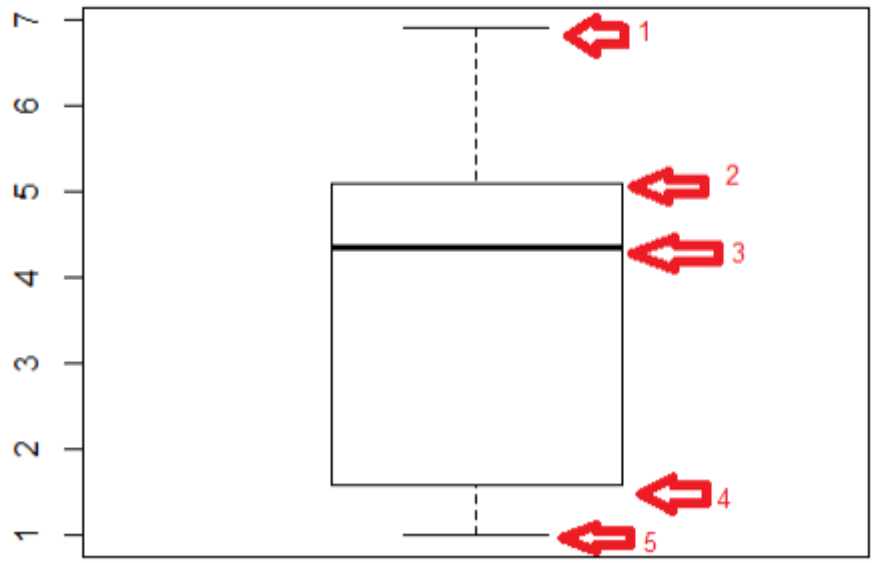
Estão CORRETAS as afirmativas:

- Todas estão corretas.
- Apenas I.
- Apenas II e III.
- (selected) Apenas I e II.

Pergunta 5

2,67 / 2,67 pts

Um analista de dados gerou o seguinte boxplot de um atributo de base de dados qualquer. Qual das alternativas representa uma análise CORRETA realizada pelo analista de dados?



Essa base de dados possui outilers que estão entre as setas 4 e 5, e entre as setas 1 e 2.

A seta 2 representa a média da base de dados.

A maioria das observações dessa base de dados estão no primeiro quartil.

A distribuição da base de dados é simétrica.

Pergunta 6

2,67 / 2,67 pts

Considere uma base de dados de alunos em que o atributo “nota” varia de 0 a 100. A média da turma é 65 e o desvio padrão 5. A partir da análise inicial, um analista de dados optou por transformar esse atributo da base usando a normalização z-score.

Assim, um aluno com nota 80 ficará com qual valor após a transformação?

65.00.

- 5.00.
- 0.82.
- 3.00.

Pergunta 7

2,67 / 2,67 pts

A tabela abaixo exibe o resultado de uma discretização do atributo nota, contendo 35 observações. As notas foram transformadas em conceitos A, B ou C. Sobre esta discretização, análise as seguintes afirmativas:

A	B	C
11	12	12

I) A técnica utilizada foi Equal Width.

II) A discretização aplicada foi a supervisionada.

III) A discretização simplifica a análise dos dados.

Estão CORRETAS as afirmativas:

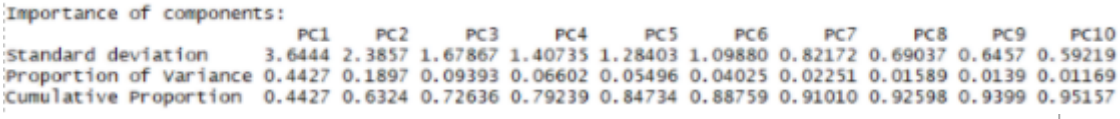
- Todas estão corretas.
- Apenas II e III.
- Apenas I.
- Apenas II.

Incorreta

Pergunta 8

0 / 2,67 pts

A imagem abaixo exibe o resultado do comando summary da lista resultante da função prcomp.



```
Importance of components:
                          PC1    PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10
Standard deviation     3.6444 2.3857 1.67867 1.40735 1.28403 1.09880 0.82172 0.69037 0.6457 0.59219
Proportion of Variance 0.4427 0.1897 0.09393 0.06602 0.05496 0.04025 0.02251 0.01589 0.0139 0.01169
Cumulative Proportion  0.4427 0.6324 0.72636 0.79239 0.84734 0.88759 0.91010 0.92598 0.9399 0.95157
```

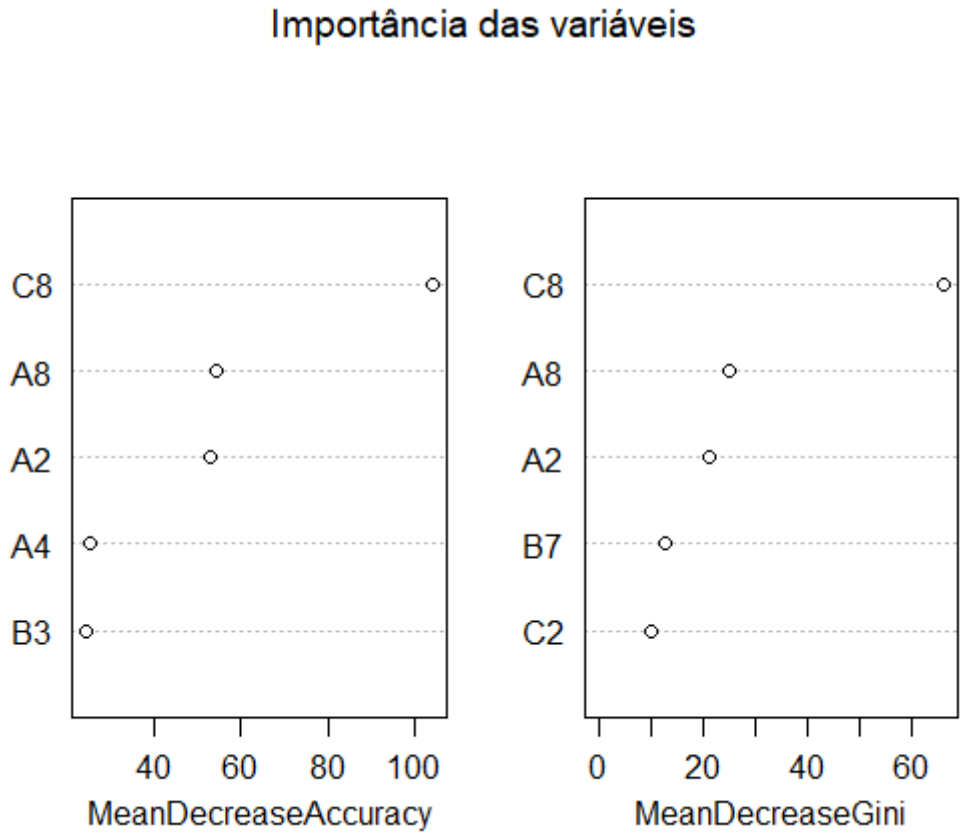
Sobre o resultado exibido na imagem pode-se afirmar que:

- Caso o analista opte por quatro componentes, ele terá mais de 80% de variância acumulada.
- .A segunda componente, PC2, acumulada com a primeira, representa 63,24% de variância.
- A primeira componente, PC1, representa 3.64 de variância.
- Os nove primeiros componentes representam 100% da variância.

Pergunta 9

2,67 / 2,67 pts

A imagem abaixo exibe o ranqueamento da importância de atributos para explicar uma determinada variável alvo, usando dois métodos distintos de ranqueamento. No caso do exemplo, a busca foi pela explicação dos fatores que determinam a longeividade de uma pessoa.



A partir da análise dos gráficos exibidos na imagem, pode-se afirmar:

I) A importância do atributo depende do método selecionado.

II) O Atributo C8 é o mais importante, independente do método de ranqueamento utilizado.

III) Apenas a partir dos gráficos é possível identificar que o método usado pelo gráfico da esquerda é melhor.

Estão CORRETAS as afirmativas:

- Apenas II.
- Apenas I.
- Todas estão corretas.
- Apenas I e II.

Incorreta

Pergunta 10

0 / 2,67 pts

A coluna da esquerda exibe alguns desafios na Integração de Dados e a coluna da direita exibe alguns exemplos desses desafios.

Desafios	Exemplos
a) Integração do Esquema.	( ) Atributo com nomes diferentes, mas que possuem o mesmo conteúdo.
b) Redundância de atributos.	( ) O atributo nome do cliente em duas bases distintas.
c) Redundância de tuplas.	( ) Os dados dos clientes em duas bases distintas.

A ordem que representa corretamente o desafio e seu respectivo exemplo, é:

- (●) b, a, c.
- ( ) c, b, a.
- ( ) a, b, c.
- ( ) a, c, b.

Pergunta 11

2,67 / 2,67 pts

A modelagem multidimensional proposta por Kimball pode ser feita atráves das seguintes abordagens:

- (●) Estrela e Floco de Neve.
- ( ) Anel e Floco de Neve.
- ( ) Estrela e Workflow.
- ( ) Anel e Estrela.

Pergunta 12

2,67 / 2,67 pts

(Questão adaptada do concurso **FCC - 2011 - INFRAERO - Analista - Banco de Dados (https://www.qconcursos.com/questoes-de-concursos/provas/fcc-2011-infraero-analista-banco-de-dados)**).

Considere as seguintes afirmativas:

I) No *Data Warehouse,* o dado tem um valor histórico, por referir-se aà algum momento específico do tempo, portanto, ele não é atualizável. A cada ocorrência de uma mudança, uma nova entrada é criada para sinalizar esta mudança.

II) O estágio de transformação no processo ETL deve ser capaz de selecionar determinadas colunas (ou nenhuma) para carregar; transformar múltiplas colunas em múltiplas linhas; traduzir e unificar códigos heterogêneos de um mesmo atributo, oriundos de diversas fontes de dados (tabelas).

III) No Snow Flake, as subdimensões, por não serem normalizadas, geram aumento significativo no número de registros e, como consequência, aumentam também a quantidade de joins necessários à exibição de uma consulta.

- (●) Apenas as afirmativas I e II estão corretas.
- ( ) Apenas as afirmativas II e III estão corretas.
- ( ) Todas as afirmativas estão corretas.
- ( ) Apenas as afirmativas I e III estão corretas.

Pergunta 13

2,67 / 2,67 pts

Sobre Data Lake, análise as seguintes afirmativas:

I) Data Lake pode ser definido como um repositório massivo de dados.

II) Uma tecnologia adotada para a implementação de Data Lake é o Hadoop desenvolvido pela Apache.

Os dados em um repositório Data Lake são transformados antes de serem ingeridos.

- Apenas II e III.
- Apenas I.
- Apenas I e II.
- Todas estão corretas.

Pergunta 14

2,67 / 2,67 pts

Na coluna da esquerda, temos as camadas do pipeline de dados de uma forma genérica. Na coluna da direita temos as descrições dessas fases.

Desafios	Exemplos
a) Ingestão de dados.	( ) Transformação dos dados.
b) Armazenamento e Processamento.	( ) Dados são coletadas de diversas fontes.
c) Serving Layer.	( ) Disponibiliza os serviços do Pipeline.

A ordem que representa corretamente o desafio e seu respectivo exemplo é:

- a, c, a.
- a, b, c.
- b, a, c.

a, c, b.

Pergunta 15

2,62 / 2,62 pts

Sobre a arquitetura Kappa, analise as seguintes afirmações:

I) É uma ampliação da arquitetura Lambda.

II) É uma arquitetura estruturada em duas camadas.

III) Trabalha apenas com Streaming de dados.

Está CORRETO o que se afirma:

- Apenas em I e II.
- Apenas em I e II.
- Todas afirmativas estão corretas.
- (selected) Apenas em II e III.

Pontuação do teste: **34,66** de 40