


# DESM1 - Desafio do Módulo 1

**Entrega** 7 ago em 23:59      **Pontos** 40      **Perguntas** 15  
**Disponível** até 7 ago em 23:59      **Limite de tempo** Nenhum

## 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 segunda-feira, **07/08/2023**, às 23h59.

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

### Enunciado do Desafio - Módulo 1 - Bootcamp Arquiteto(a) de Big Data

(<https://online.igti.com.br/courses/6613/files/442481?wrap=1>)\_ ↓

([https://online.igti.com.br/courses/6613/files/442481/download?download\\_frd=1](https://online.igti.com.br/courses/6613/files/442481/download?download_frd=1))

**Bons estudos!**

**Atenciosamente,**

**Equipe XP Educação**

# Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	<u>Tentativa 1</u>	107 minutos	37,33 de 40

⚠ As respostas corretas estarão disponíveis em 7 ago em 23:59.

Pontuação deste teste: **37,33** de 40

Enviado 6 ago em 18:37

Esta tentativa levou 107 minutos.

Pergunta 1

2,67 / 2,67 pts

Analizando a base de clientes no momento da coleta dos dados, sem nenhum tratamento, qual é a média da variável colesterol?

☐

143.44.

☐

39.14.

☐

139.14.

☒

170.43.

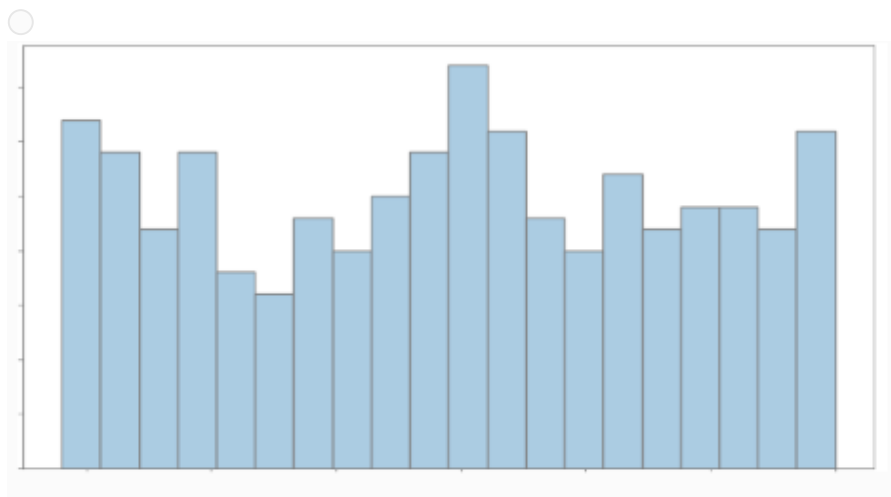
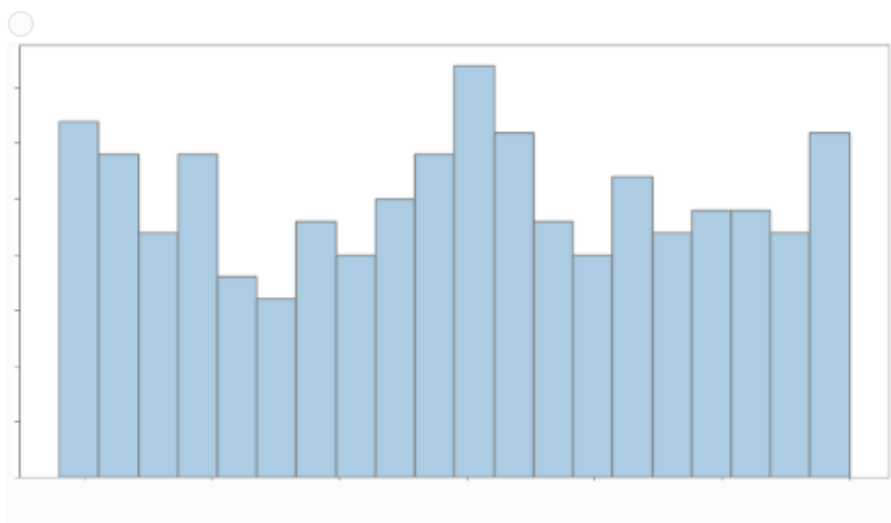
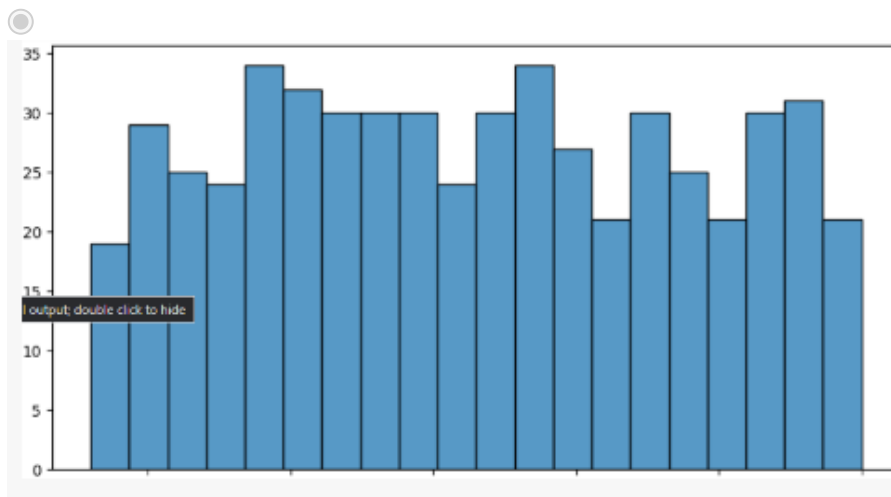
Pergunta 2

2,67 / 2,67 pts

Crie um histograma baseado na idade dos clientes e escolha alternativa que melhor a representa. Utilize 20 barras na visualização \*\* bins=20

☐

Nenhuma das demais alternativas.



### Pergunta 3

2,67 / 2,67 pts

Qual comando abaixo realiza a correção de dados ausentes e aplica a alteração no mesmo dataset?

- ☐ `df.altura.dropna(round(media,2), inplace=False)`
- ☐ `df.altura.replace(subset='altura', drop=true)`
- ☐ `df.altura.fillna(round(media,2), inplace=False)`
- ☒ `df.altura.fillna(round(media,2), inplace=True)`

#### Pergunta 4

2,67 / 2,67 pts

Crie o **gráfico** de WCSS para um conjunto de 10 clusters e, após análise dos dados no gráfico, responda: qual é o valor de WCSS para cluster de número 6? Considere apenas os 3 primeiros dígitos apresentados no gráfico.

- ☒ 103.
- ☐ 838
- ☐ 298.

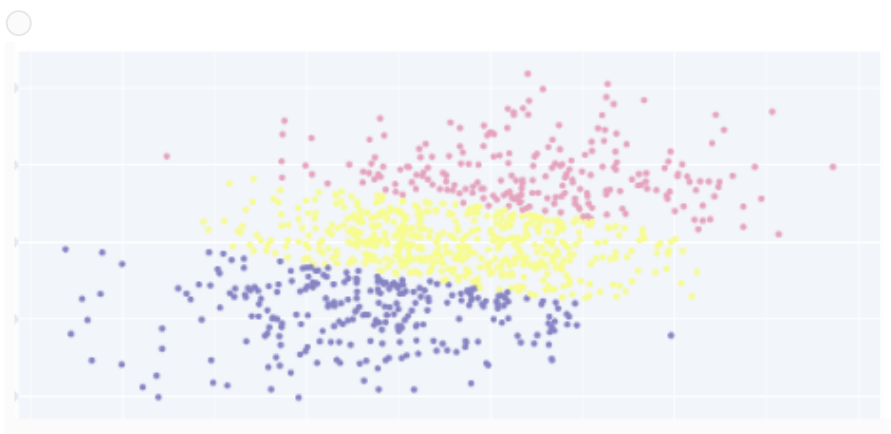
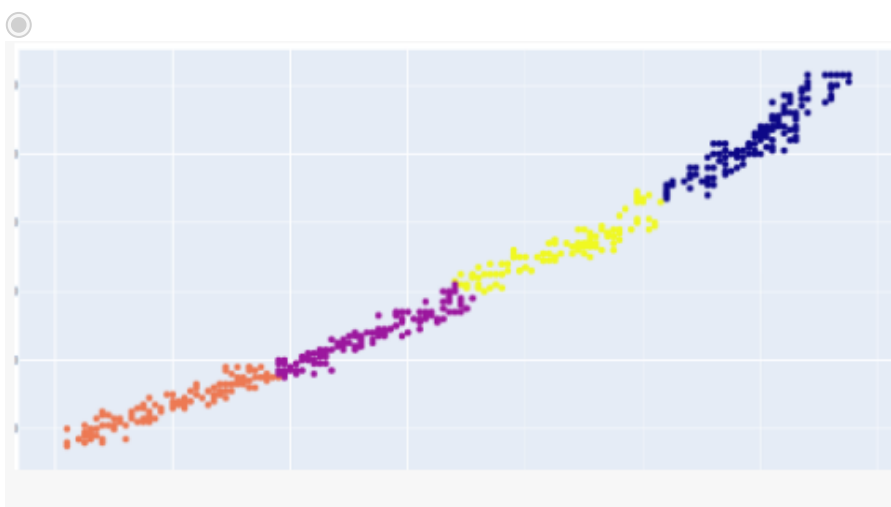
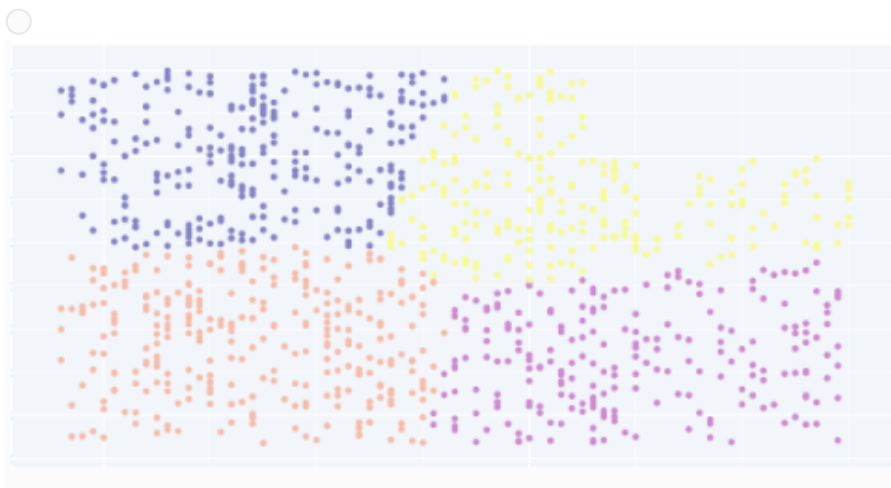
Incorreta

#### Pergunta 5

0 / 2,67 pts

Crie uma representação gráfica da formação do cluster. Após isso, escolha a imagem correta para esse agrupamento. Considere no eixo x o peso dos clientes e no eixo y o colesterol.

- ☐ Nenhuma das alternativas.

**Pergunta 6****2,67 / 2,67 pts**

Após aplicar o algoritmo de agrupamento, gere uma estatística com a função `describe` da variável `peso`, agrupando os resultados por cluster, e responda a alternativa CORRETA.



A média do peso do grupo de risco moderado é maior em comparação ao grupo de risco alto.



O desvio padrão do grupo de risco alto é maior que os demais grupos.



O desvio padrão do grupo de risco baixo é maior que os demais grupos.

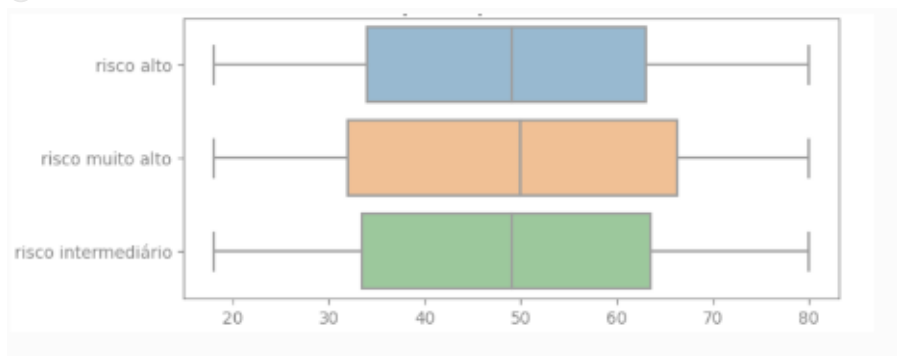
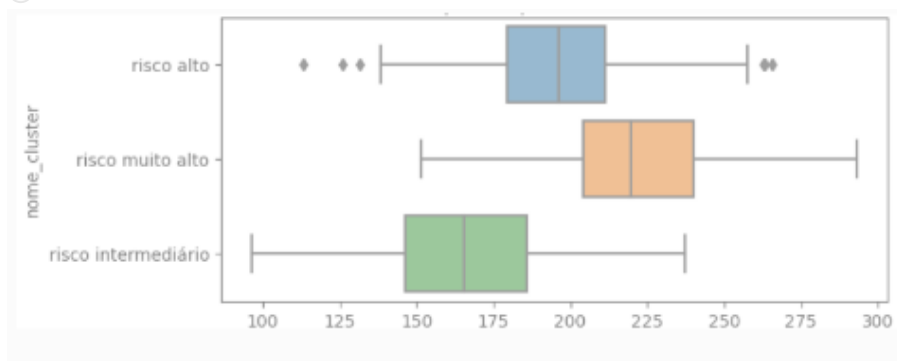


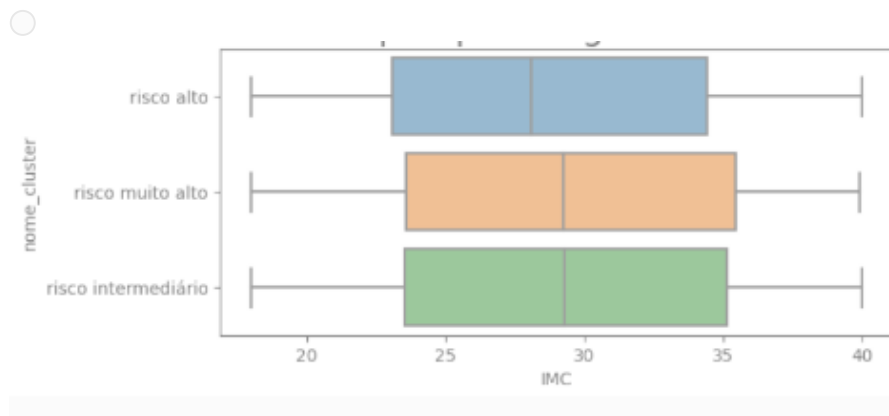
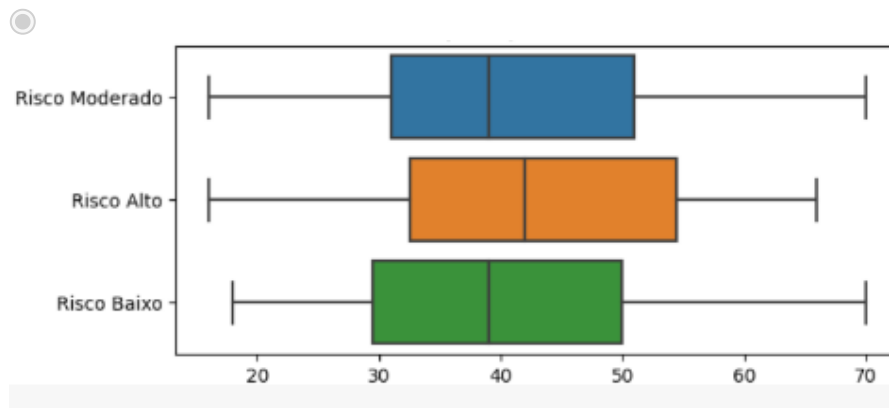
O grupo de risco muito alto possui o menor número de pessoas.

## Pergunta 7

2,67 / 2,67 pts

Gera um diagrama de boxplot filtrando todos os gêneros feminino da base criada e assinale a alternativa que apresenta o resultado das variáveis idade e nome do cluster.





### Pergunta 8

2,67 / 2,67 pts

Qual a quantidade de pessoas do gênero feminino presentes no cluster de risco baixo?

☐ 96.☐ 79.☐ 91.☒ 71.

### Pergunta 9

2,67 / 2,67 pts

Qual estado possui maior número de pessoas no grupo de risco moderado?

- ☐ São Paulo
- 
- ☒ Mato Grosso do Sul.
- 
- ☐ Pará.
- 
- ☐ Ceará

**Pergunta 10****2,67 / 2,67 pts**

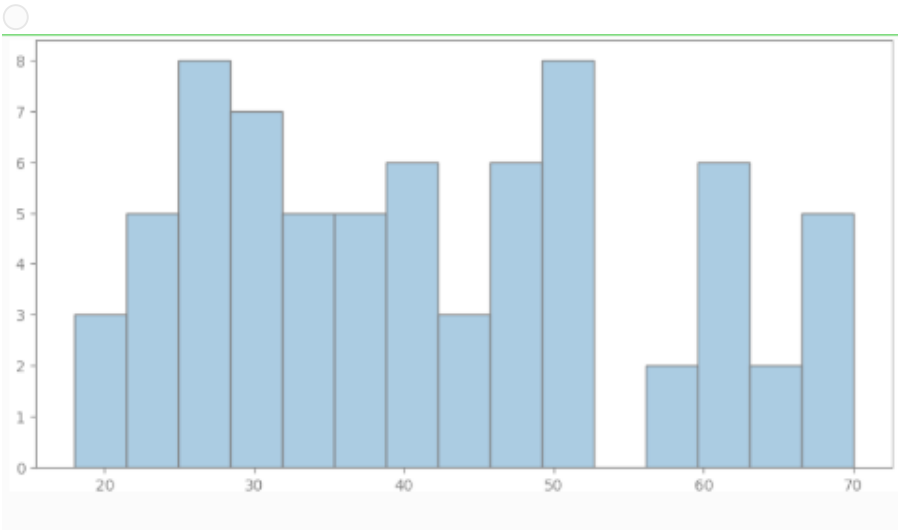
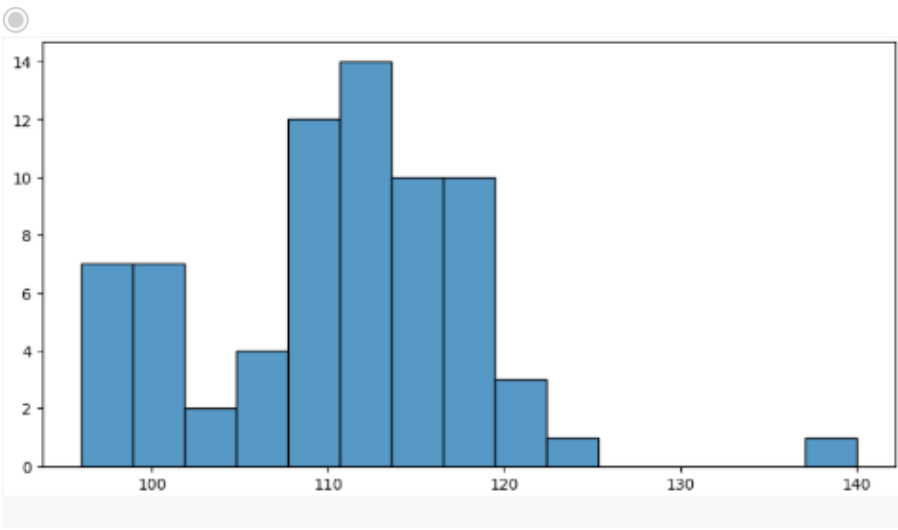
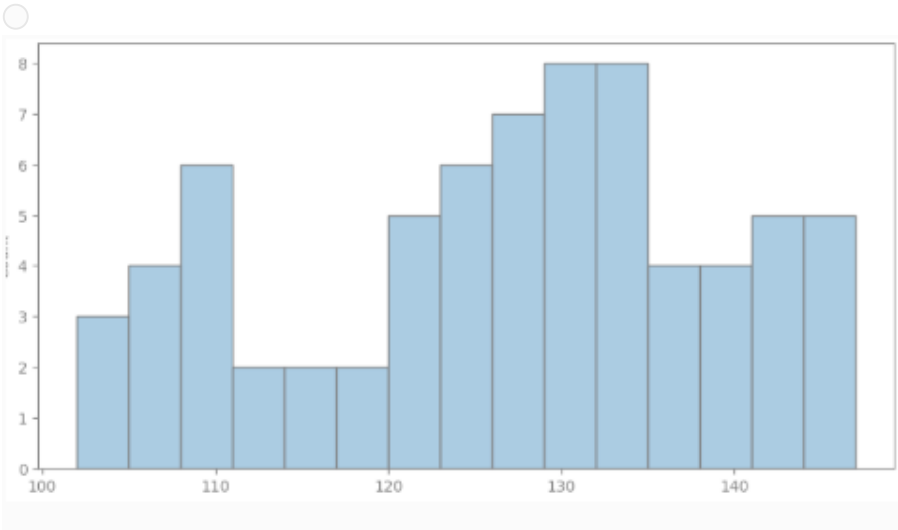
Qual é o terceiro estado que possui a maior média de colesterol?

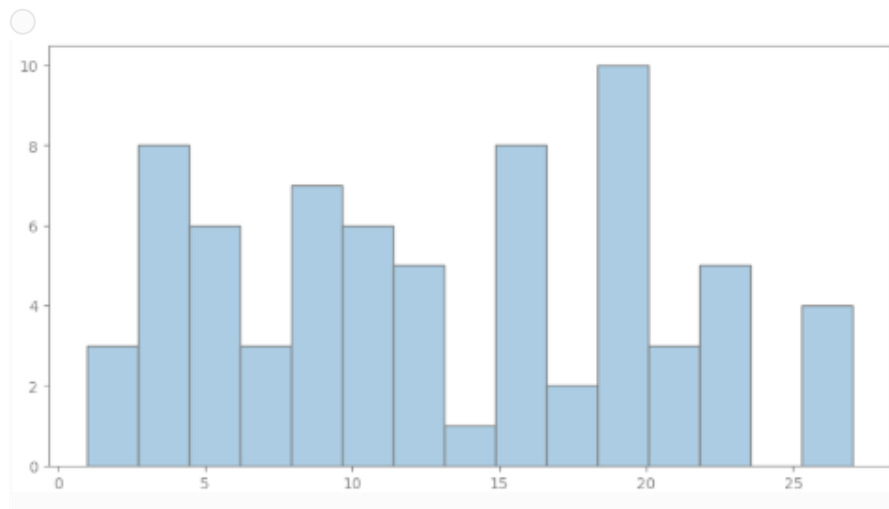
- ☐ Minas Gerais
- 
- ☒ Pernambuco.
- 
- ☐ Santa Catarina
- 
- ☐ Rio de Janeiro

**Pergunta 11****2,67 / 2,67 pts**

Filtre o dataframe para todas as pessoas do gênero feminino que estão no cluster de baixo risco e em seguida crie um histograma com a variável peso para esse conjunto de dados. Utilize 15 barras na representação do gráfico







### Pergunta 12

2,67 / 2,67 pts

São exemplos de aprendizado supervisionado:

- ☒ Regressão e Classificação.
- ☐ Clusterização e Associação.
- ☐ Associação e Classificação.
- ☐ Todas as anteriores.

### Pergunta 13

2,67 / 2,67 pts

Assinale a alternativa CORRETA:

- ☒ A clusterização k-means é uma técnica poderosa e eficiente para segmentar dados e identificar padrões relevantes em conjuntos de informações complexas.
- ☐ MAE (Mean Absolute Error) é uma métrica de avaliação para os algoritmos de clusterização.



O algoritmo k-means é uma técnica de aprendizado supervisionado que permite agrupar dados similares em clusters distintos com base em suas características quantitativas.



Regressão linear é um algoritmo de aprendizado não supervisionado, que tem como objetivo encontrar a relação linear entre uma variável independente (x) e uma variável dependente (y), a fim de prever valores de y para novos valores de x. É amplamente utilizado em diversas áreas, como economia, finanças e ciência de dados.

### Pergunta 14

2,67 / 2,67 pts

Os algoritmos de aprendizado não supervisionado podem ser divididos em duas classes: Associação e Clusterização, que se definem respectivamente como:



Ambos os algoritmos de associação não permitem o descobrimento de regras e correlação em uma base de dados. Eles apenas definem os conjuntos de itens que ocorrem juntos dentro de uma determinada frequência.



Os algoritmos de clusterização ou agrupamento permitem que seja feito agrupamento de grupos com base nas semelhanças não encontradas e os algoritmos de associação permitem o descobrimento de regras e associação em uma base de dados, identificando conjuntos de itens que ocorrem separadamente dentro de uma determinada frequência.



Os algoritmos de associação permitem o descobrimento de regras e correlação em uma base de dados, identificando conjuntos de itens que ocorrem juntos dentro de uma determinada frequência e os algoritmos de clusterização ou agrupamento permitem que seja feito agrupamento de grupos com base nas semelhanças encontradas.



Os algoritmos de clusterização ou agrupamento permitem que seja feito agrupamento de grupos com base nas semelhanças encontradas no algoritmo de regras de associação.

**Pergunta 15****2,62 / 2,62 pts**

Sobre a MAE, indique a opção que contém sua equação.



$$\frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$



$$\sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$



Nenhuma das alternativas.



$$\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$

Pontuação do teste: **37,33** de 40