Desafio do Módulo 4

Entrega 8 jul em 17:00 Limite de tempo Nenhum Pontos 40

Perguntas 15

Disponível até 8 jul em 17:00

Instruções

O Desafio do Módulo 4 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".

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.

O gabarito será disponibilizado partir de sexta-feira, 10/07/2020, às 21h.

Bons estudos!

2. O arquivo abaixo contém o enunciado do desafio

<u> Enunciado do Desafio - Módulo 3 - Bootcamp Analista de Dados.pdf</u> 🗟

Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	Tentativa 1	4 minutos	40 de 40

(!) As respostas corretas estarão disponíveis em 10 jul em 21:00.

Pontuação deste teste: 40 de 40

Enviado 2 jul em 10:01

Esta tentativa levou 4 minutos.

Pergunta 1	2,67 / 2,67 pts

Quantas instâncias (linhas) e características (colunas) existem, respectivamente, no dataset?

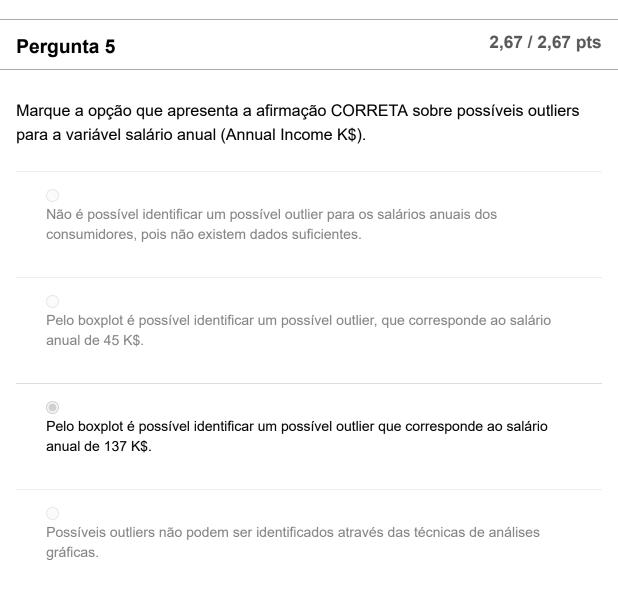
(5, 200).			
(200, 5).			
(500, 7).			
(7, 500).			

Pergunta 2	2,67 / 2,67 pts
Quantas variáveis do tipo "string" estão presentes no dataset ?	
4.	
1.	
O 3.	
O 2.	

Pergunta 3	2,67 / 2,67 pts
Qual é a idade (age) média dos consumidores?	
50,20 anos.	
○ 64,28 anos.	
38,85 anos.	
○ 57,87 anos.	

Г

Pergunta 4	2,67 / 2,67 pts
Qual é o desvio padrão para os salários anuais (Annual Income consumidores em K\$?	(K\$)) dos
© 26,26 K\$.	
○ 60,56 K\$.	
○ 13,86 K\$.	
○ 38,85 K\$.	



Marque a afirmação CORRETA acerca da distribuição salarial anual (Annual Income K\$) e o sexo (Genre) dos consumidores, presentes no dataset Mall_Custumers.csv.

Os homens (male) ganham, em média, menos que as mulheres (female).

Homens (male) e mulheres (female) recebem, em média, os mesmos salários anuais.

Os homens (male) ganham, em média, mais que as mulheres (female).

Nada pode ser afirmado, pois os dados são insuficientes para esse tipo de análise.

Pergunta 7

2,67 / 2,67 pts

A variável "Spending Score (1-100)" indica o quanto o consumidor é "lucrativo" para o shopping. Assim, quanto mais próximo de 100, mais "lucrativo" é o consumidor. Analisando a relação entre o sexo (Genre), o salário anual (Annual Income) e o "Spending Score", marque a opção CORRETA.

Essas variáveis não podem ser analisadas em conjunto, pois os clientes não foram segmentados.

Os dados mostram que o salário anual (Annual Income K\$) é o único fator que influencia o "Spending Score", pois quem ganha mais gasta mais.

Homens (male) e mulheres (female) são poucos "lucrativos" para os shoppings, pois possuem baixos salários anuais.

Apesar de terem maiores salários (Annual Income), os homens (male) são menos "lucrativos" (Spending Score) para o shopping.

Pergunta 8 2,67 / 2,67 pts

Mesmo com o coeficiente de "Pearson" não sendo muito alto, ainda é possível identificar algum tipo de relacionamento linear. Comparando o coeficiente de correlação de "Pearson" entre as variáveis idade ("Age") e a pontuação de consumo ("Spending Score"), é CORRETO afirmar:

O coeficiente de correlação de "Pearson" positivo indica que as duas variáveis possuem um relacionamento de "causalidade".

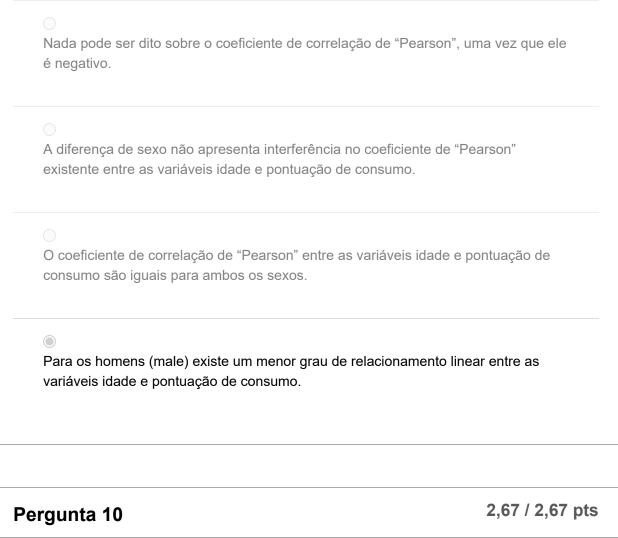
Mesmo possuindo um valor, em módulo inferior a 0,5, é possível dizer que possuem um relacionamento negativo.

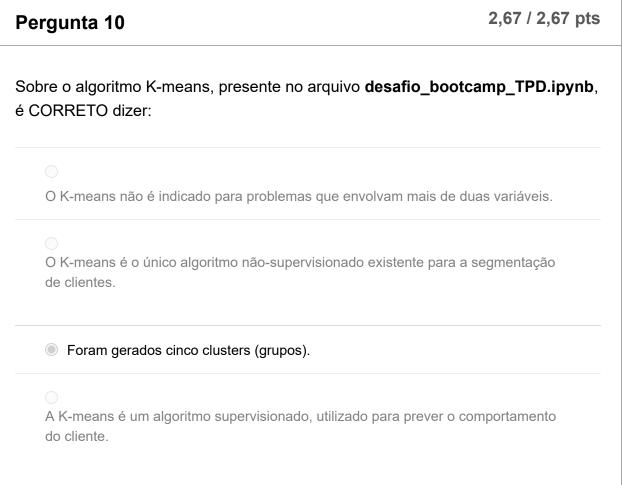
Um coeficiente de correlação de "Pearson" negativo indica que, se uma variável aumenta, a outra tende a aumentar em uma mesma proporção.

Quando a variável "Age" aumenta, a variável "Spending Score" também aumenta.

Pergunta 9 2,67 / 2,67 pts

Após separar o dataset entre homens (male) e mulheres (female), e aplicar, novamente, a análise de correlação de "Pearson" entre as variáveis idade ("Age") e pontuação de consumo ("Spending Score"), é CORRETO afirmar:





2,67 / 2,67 pts

2,67 / 2,67 pts Pergunta 12

Marque a opção INCORRETA sobre a utilização do K-means para a segmentação de clientes.

Com mais de duas dimensões (duas variáveis), torna-se impossível realizar a clusterização de clientes.

A segmentação auxilia em identificar grupos de clientes e oferecer produtos e serviços de forma mais assertiva.

Algoritmos como K-means e hierárquicos, por exemplo, representam uma poderosa ferramenta para estratégias de marketing e avaliação de cenários.

Através dos grupos é possível identificar, por exemplo, padrões de consumo através de gênero e idade.

Modifique o código do K-means presente no arquivo, para a geração de três grupos, utilizando as mesmas variáveis "Spending Score" e "Annual Income" para o processo de clusterização. Analisando esses três clusters, é CORRETO afirmar:

A identificação das particularidades dos clientes torna-se ainda mais complexa, pois os grupos (clusters) são maiores.

Não ocorreu alteração nos clusters.

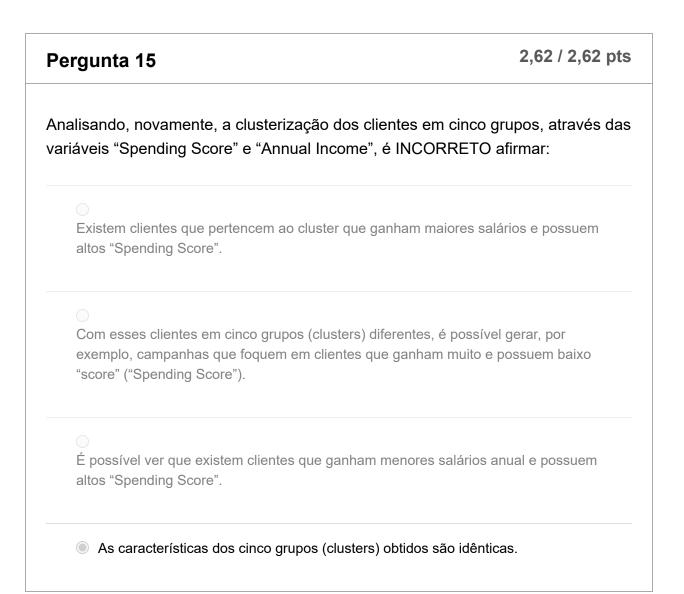
Alterar a quantidade de clusters não influencia nas análises a serem realizadas.

Com uma menor quantidade de clusters, temos que as distâncias entre as amostras e o centróide do cluster não é alterada.

Pergunta 14 Sobre o processo de clusterização, é INCORRETO afirmar: O processo de clusterização pode envolver algoritmos não-supervisionados. Através do processo de clusterização é possível acompanhar o comportamento dos clientes e aumentar a fidelização. Diferentes ramos de negócio podem empregar a clusterização como forma de conhecer melhor as características dos clientes.



A utilização de métodos de segmentação não pode ser empregada para visualizar estratégias de curto prazo.



Pontuação do teste: 40 de 40