Desafio do Módulo 4

Entrega 13 jul em 23:59 Pontos 40 Perguntas 15

Disponível até 13 jul em 23:59 **Limite de tempo** Nenhum

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 segunda-feira, 13/07/2020, às 21h.

Bons estudos!

2. O arquivo abaixo contém o enunciado do desafio

Enunciado do Desafio - Módulo 3 - Bootcamp Analista de Dados.pdf

Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	Tentativa 1	3.245 minutos	31,99 de 40

(!) As respostas corretas estarão disponíveis em 13 jul em 23:59.

Pontuação deste teste: 31,99 de 40

Enviado 9 jul em 19:50

Esta tentativa levou 3.245 minutos.

Pergunta 1	2,67 / 2,67 pts

Quantas instâncias (linhas) e características (colunas) existem, respectivamente, no dataset?
(5, 200).
© (200, 5).
(7, 500).
(500, 7).

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

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

38,85 anos.

Incorreta

Pergunta 4

Qual é o desvio padrão para os salários anuais (Annual Income (K\$)) dos consumidores em K\$?

- 0 60,56 K\$.
- 26,26 K\$.
- 13,86 K\$.
- 38,85 K\$.

Pergunta 5

2,67 / 2,67 pts

0 / 2,67 pts

Marque a opção que apresenta a afirmação CORRETA sobre possíveis outliers para a variável salário anual (Annual Income K\$).

Pelo boxplot é possível identificar um possível outlier que corresponde ao salário anual de 137 K\$.

Não é possível identificar um possível outlier para os salários anuais dos consumidores, pois não existem dados suficientes.

Possíveis outliers não podem ser identificados através das técnicas de análises gráficas.

Pelo boxplot é possível identificar um possível outlier, que corresponde ao salário anual de 45 K\$.

Pergunta 6

2,67 / 2,67 pts

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.

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

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

Incorreta

Pergunta 7

0 / 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.

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

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

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

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.

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:

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

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.

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:

Nada pode ser dito sobre o coeficiente de correlação de "Pearson", uma vez que ele é negativo.

O coeficiente de correlação de "Pearson" entre as variáveis idade e pontuação de consumo são iguais para ambos os sexos.

Para os homens (male) existe um menor grau de relacionamento linear entre as variáveis idade e pontuação de consumo.

A diferença de sexo não apresenta interferência no coeficiente de "Pearson" existente entre as variáveis idade e pontuação de consumo.

Pergunta 10

2,67 / 2,67 pts

Sobre o algoritmo K-means, presente no arquivo **desafio_bootcamp_TPD.ipynb**, é CORRETO dizer:

means é um algoritmo supervisionado, utilizado para prever o portamento do cliente.
means é o único algoritmo não-supervisionado existente para a nentação de clientes.
means não é indicado para problemas que envolvam mais de duas liveis.
Foram gerados cinco clusters (grupos).

Pergunta 11

2,67 / 2,67 pts

Marque a afirmativa CORRETA sobre o K-means presente no arquivo desafio_bootcamp_TPD.ipynb.

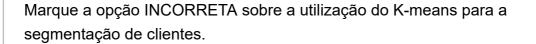
Clientes que estão em um mesmo cluster (grupo) possuem características similares de salário anual e pontuação de consumo.

A clusterização só deve ser empregada para a análise de correlação.

- O Todos os clusters apresentam a mesma quantidade de clientes.
- Não é possível realizar a divisão em clusters.

Pergunta 12

2,67 / 2,67 pts



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.

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

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

Incorreta

Pergunta 13

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:

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

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

0 / 2,67 pts

Não ocorreu alteração nos clusters.
Alterar a quantidade de clusters não influencia nas análises a serer realizadas.

Pergunta 14

2,67 / 2,67 pts

Sobre o processo de clusterização, é INCORRETO afirmar:

O processo de clusterização pode envolver algoritmos nãosupervisionados.

Diferentes ramos de negócio podem empregar a clusterização como forma de conhecer melhor as características dos clientes.

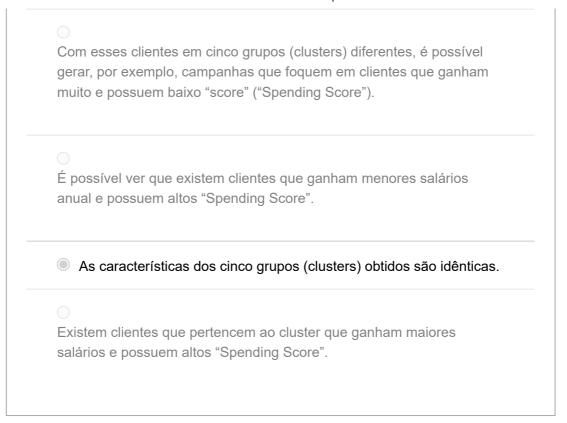
Através do processo de clusterização é possível acompanhar o comportamento dos clientes e aumentar a fidelização.

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

Pergunta 15

2,62 / 2,62 pts

Analisando, novamente, a clusterização dos clientes em cinco grupos, através das variáveis "Spending Score" e "Annual Income", é INCORRETO afirmar:



Pontuação do teste: 31,99 de 40