

# Desafio do Módulo 4

---

<b>Entrega</b> 29 abr em 21:00	<b>Pontos</b> 40	<b>Perguntas</b> 15
<b>Disponível</b> até 29 abr em 21:00	<b>Limite de tempo</b> 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 do Módulo 4".

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, **30/04/2021**, às 23h59.

Bons estudos!

### 2. O arquivo abaixo contém o enunciado do desafio

[Enunciado do Desafio – Módulo 4 – Bootcamp Cientista de Dados.pdf](#)

## Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	<a href="#">Tentativa 1</a>	30 minutos	40 de 40

---

⚠ As respostas corretas estarão disponíveis em 30 abr em 23:59.

Pontuação deste teste: **40** de 40

Enviado 29 abr em 18:19

Esta tentativa levou 30 minutos.

### Pergunta 1

2,67 / 2,67 pts

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

☒ (200, 5).

☐ (500, 7).

☐ (5, 200).

☐ (7, 500).

### Pergunta 2

2,67 / 2,67 pts

Quantas variáveis do tipo “string” estão presentes no dataset?

☒ 1.

☐ 3.

☐ 2.

☐ 4.

### Pergunta 3

2,67 / 2,67 pts

Qual é a idade (age) média dos consumidores?

☐ 50,20 anos.

☐ 57,87 anos.

☐ 64,28 anos.

☒ 38,85 anos.

#### Pergunta 4

2,67 / 2,67 pts

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

☐ 38,85 K\$.

☐ 13,86 K\$.

☐ 60,56 K\$.

☒ 26,26 K\$.

#### Pergunta 5

2,67 / 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 45 K\$.



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



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



Pelo boxplot é possível identificar um possível outlier, que corresponde ao salário anual de 137 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\_Customers.csv.



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



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



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.

☐

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.

☐

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.

### 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:



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



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.



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:



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



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.



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

### Pergunta 10

2,67 / 2,67 pts

Sobre o algoritmo K-means, presente no arquivo **desafio\_bootcamp\_TPD.ipynb**, é CORRETO dizer:



O K-means não é indicado para problemas que envolvam mais de duas variáveis.



A K-means é um algoritmo supervisionado, utilizado para prever o comportamento do cliente.



Foram gerados cinco clusters (grupos).



O K-means é o único algoritmo não-supervisionado existente para a segmentação de clientes.

### Pergunta 11

2,67 / 2,67 pts

Marque a afirmativa CORRETA sobre o K-means presente no arquivo **desafio\_bootcamp\_TPD.ipynb**.

☐ Não é possível realizar a divisão em clusters.



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.

☐ Todos os clusters apresentam a mesma quantidade de clientes.

## Pergunta 12

2,67 / 2,67 pts

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



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.



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





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

### Pergunta 13

2,67 / 2,67 pts

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.



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



Não ocorreu alteração nos clusters.



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

### Pergunta 14

2,67 / 2,67 pts

Sobre o processo de clusterização, é INCORRETO afirmar:



O processo de clusterização pode envolver algoritmos não-supervisionados.



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:



Com esses clientes em cinco grupos (clusters) diferentes, é possível gerar, por exemplo, campanhas que foquem em clientes que ganham muito e possuem baixo “score” (“Spending Score”).



É possível ver que existem clientes que ganham menores salários anual e possuem altos “Spending Score”.



As características dos cinco grupos (clusters) obtidos são idênticas.



Existem clientes que pertencem ao cluster que ganham maiores salários e possuem altos “Spending Score”.

Pontuação do teste: **40** de 40