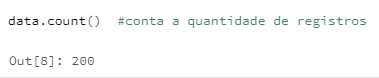
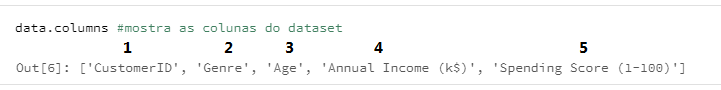
**Relatório Desafio**

**Pergunta 1**

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

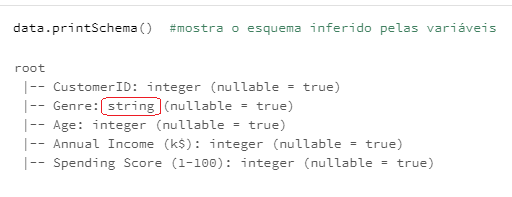




**R.: (200,5)**

**Pergunta 2**

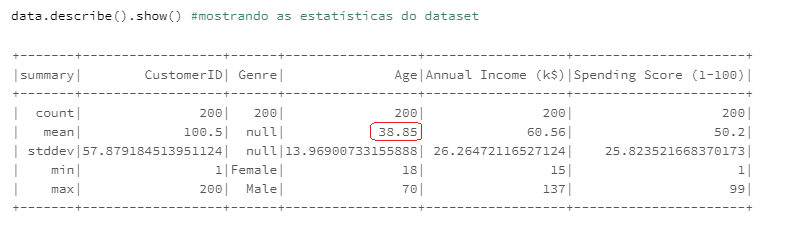
Quantas variáveis do tipo “string” estão presentes no dataset ?



**R.: 1**

**Pergunta 3**

Qual é a idade (age) média dos consumidores?

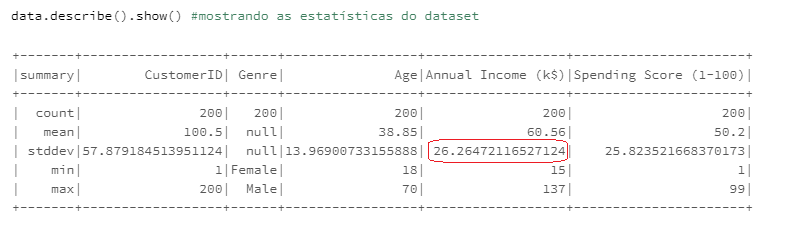


**R.: 38,85**

**Pergunta 4**

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

consumidores em K$?

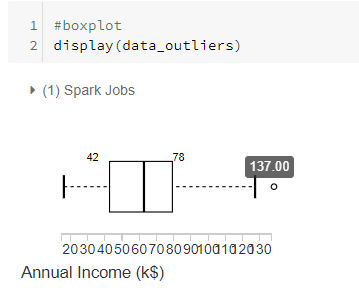


**R.: 26,26**

**Pergunta 5**

Marque a opção que apresenta a afirmação CORRETA sobre possíveis outliers

para a variável salário anual (Annual Income K$).



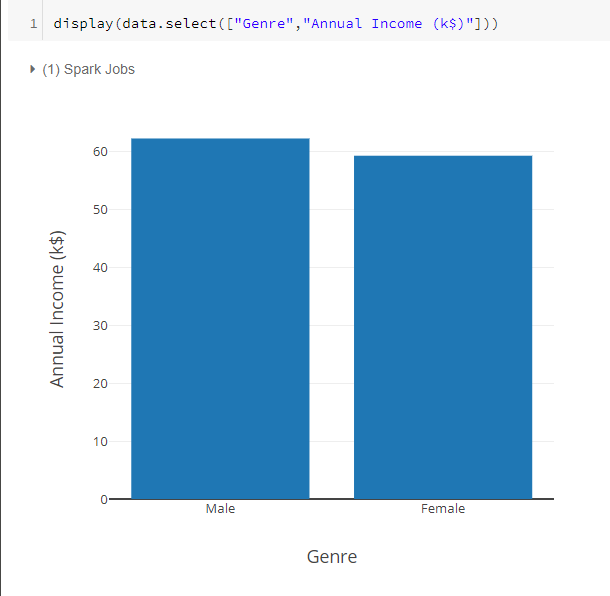
**R.: Pelo boxplot é possível identificar um possível outlier que corresponde ao salário anual de 137 K$.**

**Pergunta 6**

Marque a afirmação CORRETA acerca da distribuição salarial anual (Annual

Income K$) e o sexo (Genre) dos consumidores, presentes no dataset Mall\_Custumers.csv.

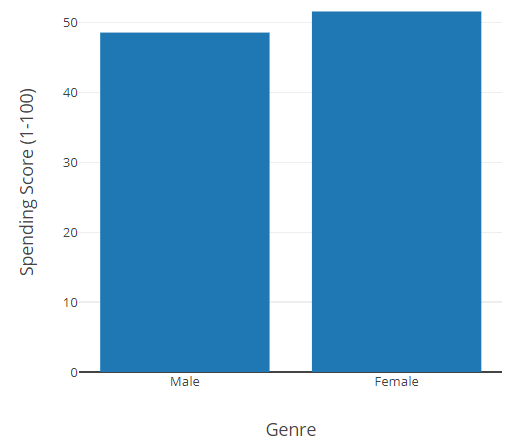


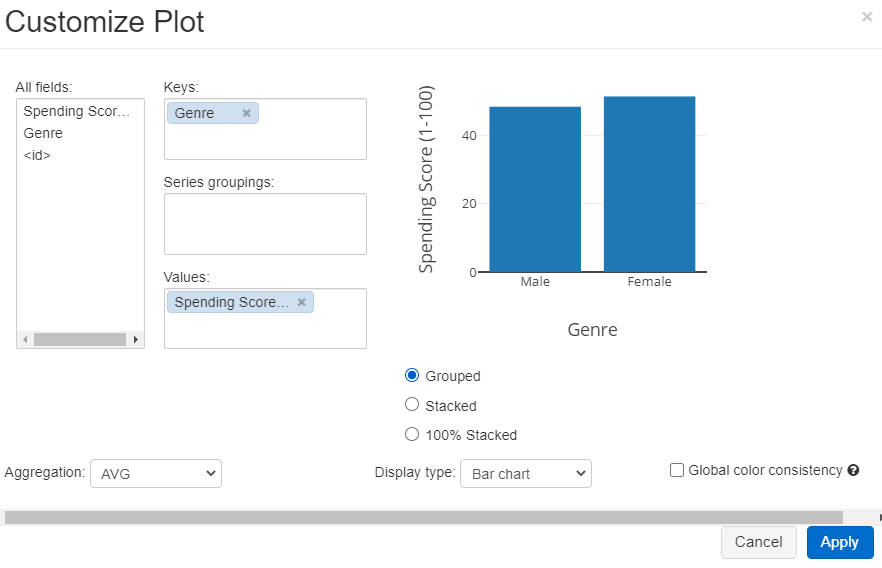


**R.: Os homens (male) ganham, em média, mais que as mulheres (female).**

**Pergunta 7**

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.





**R.: Apesar de terem maiores salários (Annual Income), os homens (male) são menos “lucrativos” (Spending Score) para o shopping.**

**Pergunta 8**

Mesmo com o coeficiente de “Pearson” não sendo muito alto, ainda é possível

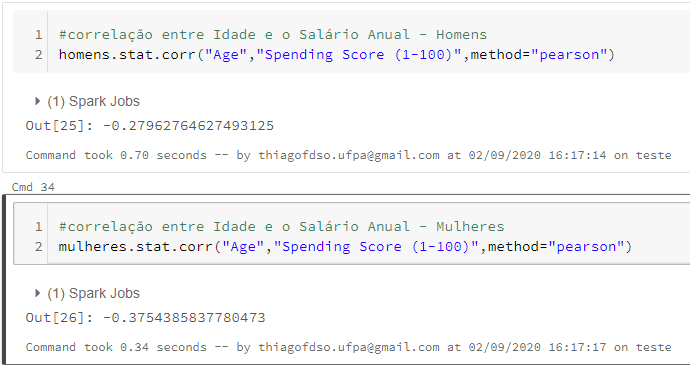
identificar algum tipo de relacionamento linear. Comparando o coeficiente de correlação de “Pearson” entre as variáveis idade (“Age”) e a pontuação de consumo (“Spending Score”), é CORRETO afirmar:

****

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

**Pergunta 9**

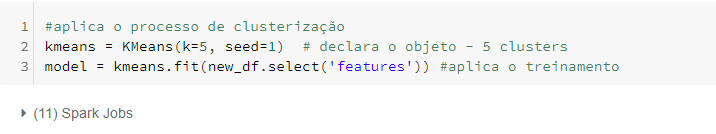
Após separar o dataset entre homens (male) e mulheres (female), e aplicar, novamente, a análise de correlação de “Pearson” entre as variáveis idade (“Age”) e pontuação de consumo (“Spending Score”), é CORRETO afirmar:

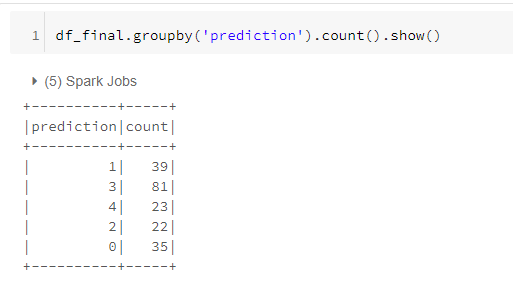


**R.: Para os homens (male) existe um menor grau de relacionamento linear entre as variáveis idade e pontuação de consumo.**

**Pergunta 10**

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

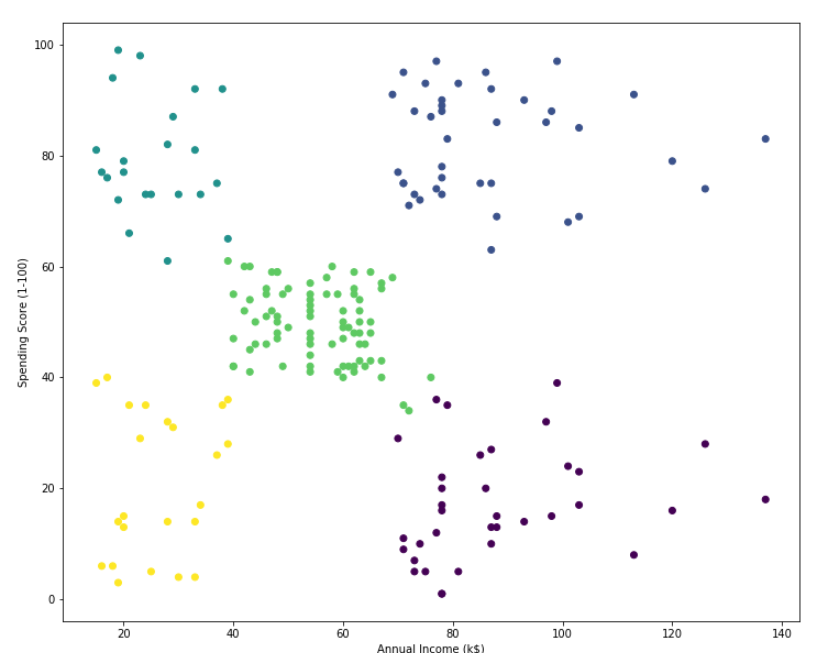




**R.: Foram gerados cinco clusters (grupos).**

**Pergunta 11**

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



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

**Pergunta 12**

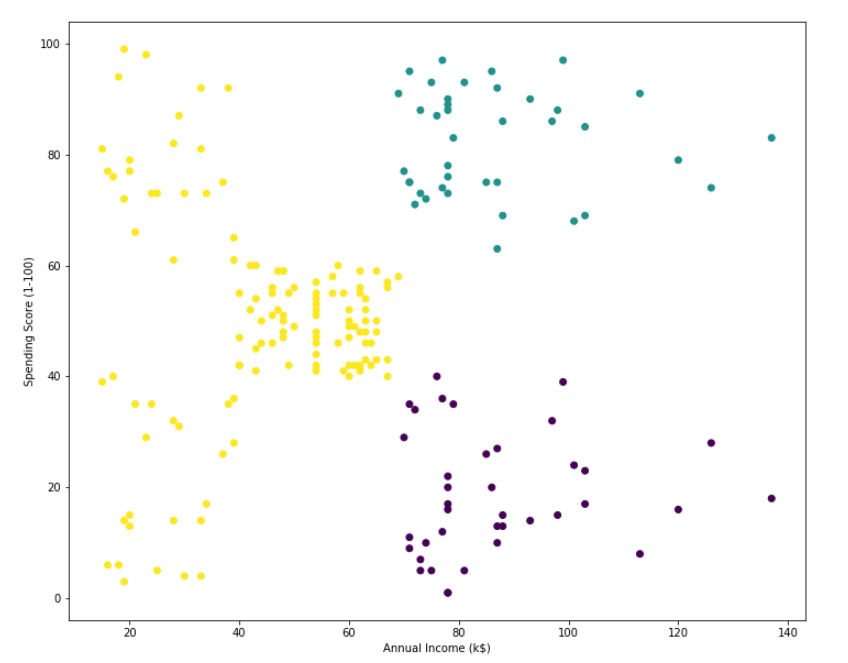
Marque a opção INCORRETA sobre a utilização do K-means para a segmentação de clientes.

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

**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:



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

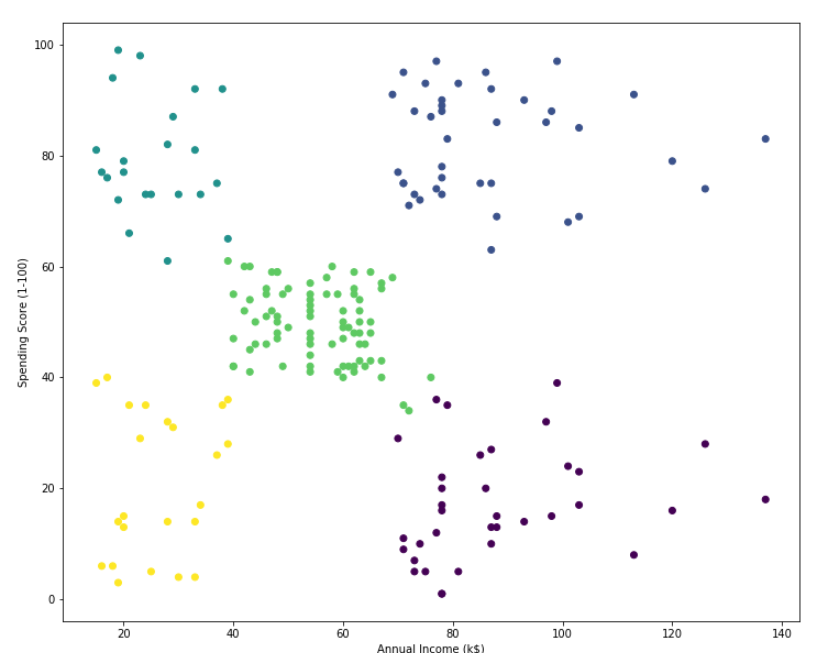
**Pergunta 14**

Sobre o processo de clusterização, é INCORRETO afirmar:

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

**Pergunta 15**

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



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