

DESF5 - Desafio Final

Entrega 1 abr em 23:59

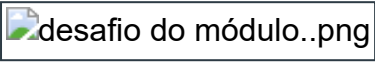
Pontos 100

Perguntas 15

Disponível até 1 abr em 23:59

Limite de tempo Nenhum

Instruções



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 Final".

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 a partir de sexta-feira, **05/04/2024** às 23h59.

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

Enunciado do Desafio Final – Bootcamp Cientista de

Dados.pdf (<https://online.igti.com.br/courses/6910/files/513895?wrap=1>)_ ↓
(https://online.igti.com.br/courses/6910/files/513895/download?download_frd=1)

Bons estudos!

Atenciosamente,

Equipe XP Educação

Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	<u>Tentativa 1</u>	4.170 minutos	100 de 100

⚠ As respostas corretas estarão disponíveis em 5 abr em 23:59.

Pontuação deste teste: **100** de 100

Enviado 24 mar em 18:39

Esta tentativa levou 4.170 minutos.

Pergunta 1

6,67 / 6,67 pts

Assinale a alternativa CORRETA.



Big Data é um termo que se refere a enormes volumes de dados gerados por diferentes fontes, que excedem a capacidade de processamento e armazenamento convencionais, e requerem técnicas e tecnologias avançadas para serem capturados, armazenados, processados e analisados.



O MongoDB é um banco de dados que não suporta transações ACID, o que torna difícil a garantia da consistência dos dados em operações críticas. Além disso, o MongoDB não oferece suporte a JOINS, o que torna complicado o gerenciamento de relacionamentos entre dados em diferentes coleções. Essas limitações tornam o MongoDB inadequado para aplicações que exigem alta confiabilidade e complexidade de relacionamentos entre dados.



O KDD (Knowledge Discovery in Databases) é um método de análise de dados que não requer nenhuma técnica especializada e pode ser realizado por qualquer pessoa sem conhecimento em tecnologia de informação ou estatística.



Hadoop é um banco de dados relacional altamente flexível, escalável e com recursos de recuperação de falhas integrados. Além disso, Hadoop é capaz de executar transações ACID em um grande volume de dados sem sacrificar o desempenho. No entanto, o Hadoop não oferece suporte nativo a consultas em tempo real e não possui um esquema de dados fixo, o que dificulta a execução de consultas complexas e a garantia da consistência dos dados. Como resultado, Hadoop pode não ser a escolha ideal para aplicativos que exigem alta performance em consultas complexas e precisam de uma garantia estrita da consistência dos dados.

Pergunta 2

6,67 / 6,67 pts

Qual é a idade média das pessoas que residem no estado de Minas Gerais?

☐ 43.8

☒ 42.7

☐ 14.7

☐ 15.6

Pergunta 3

6,67 / 6,67 pts

Você foi contratado como analista de dados por uma empresa de mídia para analisar o desempenho de campanhas de marketing em várias plataformas. A empresa deseja entender como o público interage com as campanhas em diferentes canais e como isso impacta as taxas de conversão. Você coletou dados detalhados de várias campanhas e canais, incluindo redes sociais, e-mail marketing e anúncios on-line.

Para criar um relatório de Data Storytelling que forneça insights acionáveis à equipe de marketing, quais das seguintes etapas você consideraria ao projetar sua abordagem de análise e comunicação?



Segmentar o público-alvo com base em dados demográficos e de comportamento, destacando as diferenças nas taxas de conversão entre grupos.



Apresentar apenas gráficos detalhados com todos os dados brutos para que a equipe possa fazer sua própria análise.



Fornecer um resumo geral dos resultados sem aprofundar em detalhes específicos.



Utilizar uma narrativa única e linear, ignorando variações nos dados que possam ser relevantes para a equipe de marketing.

Pergunta 4

6,67 / 6,67 pts

Crie um gráfico de caixa (boxplot) para representar a distribuição da renda por gênero das pessoas que residem no estado de Minas Gerais e, em seguida, responda:

Qual é o gênero que possui maior mediana da renda representado no gráfico de caixa?



Outro



Feminino



Masculino



Masculino e feminino

Pergunta 5

6,67 / 6,67 pts

Considere um ambiente empresarial que utiliza sistemas de gerenciamento de banco de dados para suas operações diárias e para análises de negócios. Nesse contexto, é essencial compreender as diferenças entre OLTP (Processamento de Transações Online) e OLAP (Processamento Analítico Online). Analise as afirmações a seguir:

I- OLTP é projetado para suportar transações de rotina e é otimizado para inserções, atualizações e exclusões de registros em tempo D-1.

II- OLAP é projetado para consultas complexas e análises de dados históricos, sendo otimizado para agregações e consultas multidimensionais.

III- Em sistemas OLTP, as consultas geralmente envolvem grandes volumes de dados históricos e análises de tendências de longo prazo.

IV- Sistemas OLAP são usados para suportar o processamento de pedidos, a manutenção de registros de funcionários e a atualização de inventários em tempo real.

V- O principal objetivo de sistemas OLAP é fornecer insights de negócios por meio de análises de dados para melhorar a tomada de decisões.

Identifique as afirmações CORRETAS:

☐ I, III e IV.

☐ I, III.

☐ I, III e IV.

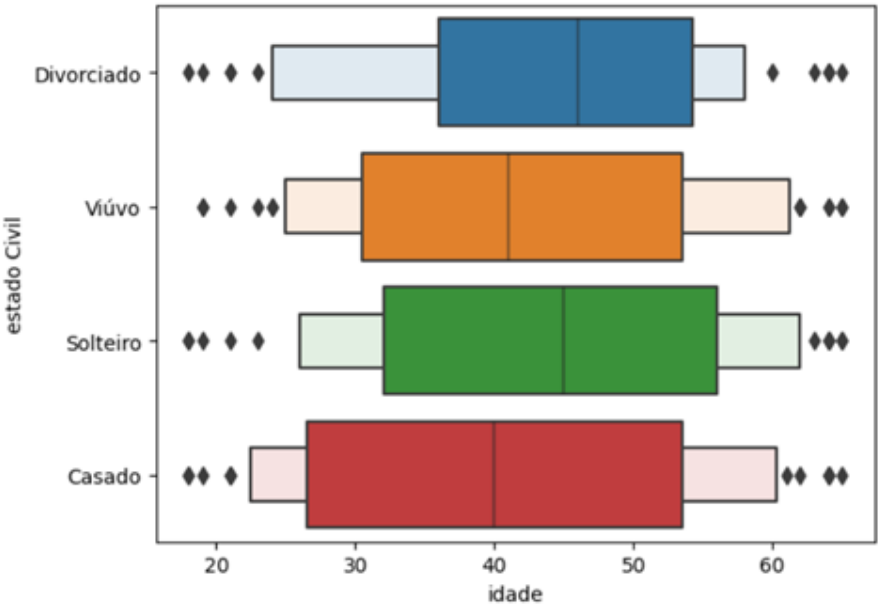
☒ II e V.

Pergunta 6

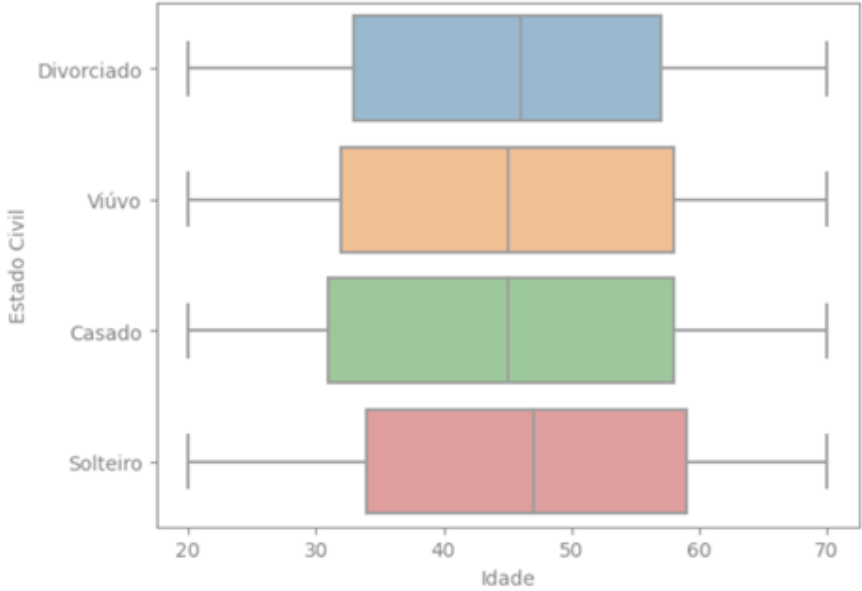
6,67 / 6,67 pts

Crie um gráfico de caixa (boxplot) para representar as idades das pessoas que possuem níveis de educação superior, categorizadas por seu estado civil. Em seguida, escolha o gráfico que melhor representa essa distribuição.

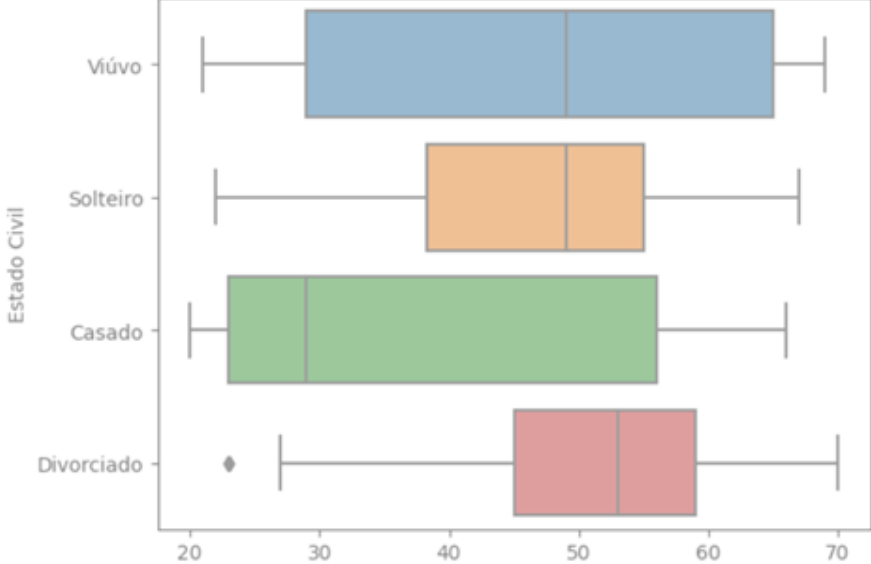
●

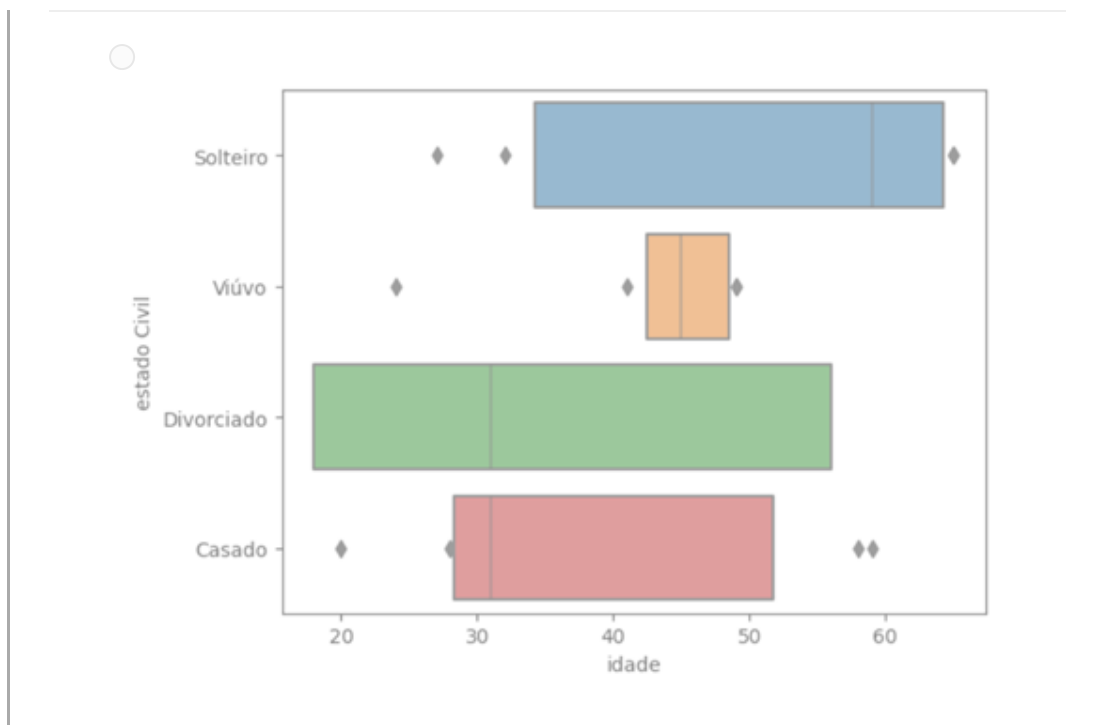


○



○





Pergunta 7

6,67 / 6,67 pts

Calcule estatísticas descritivas para a variável "colesterol" agrupadas por nível de educação. Em seguida, com base nas estatísticas calculadas, escolha a alternativa CORRETA.



A mediana do colesterol é a mesma em todos os grupos de nível de educação.



A moda do colesterol é a mesma em todos os grupos de nível de educação.



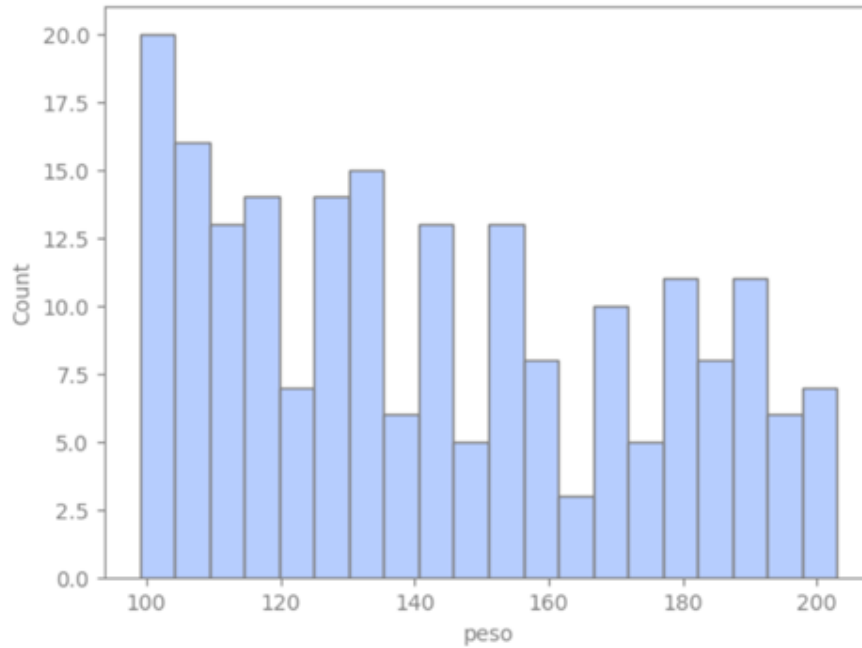
A média do colesterol é mais elevada no grupo com nível de Superior em comparação com os outros grupos.



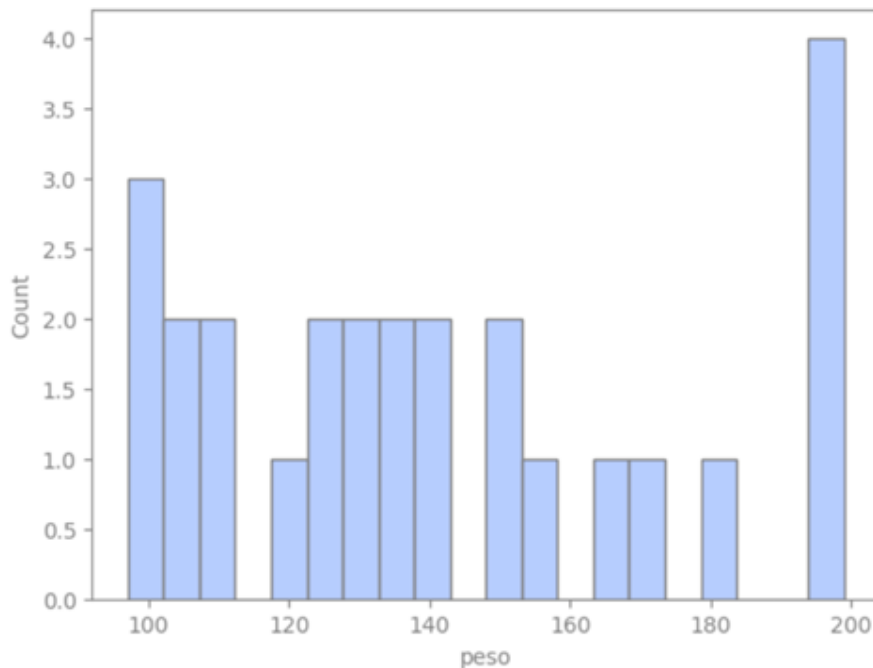
A média do colesterol é mais elevada no grupo com nível de educação fundamental em comparação com os outros grupos.

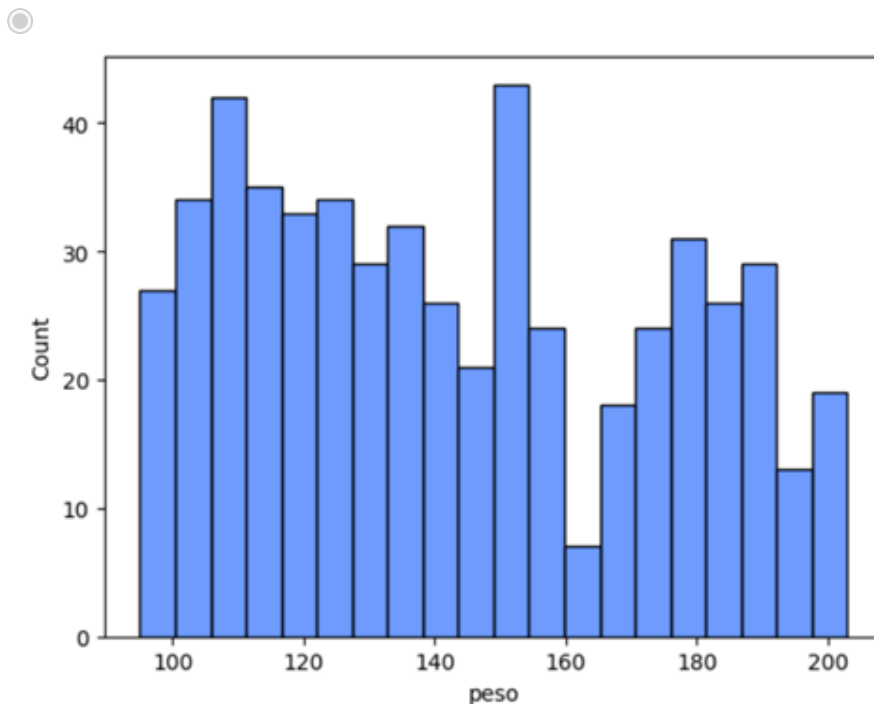
Pergunta 8**6,67 / 6,67 pts**

Crie um histograma com 20 barras representando o peso das pessoas. Em seguida, com base no histograma gerado, escolha a alternativa que melhor descreve a distribuição dos pesos.

☐☐

Nenhuma das alternativas.

☐



Pergunta 9

6,67 / 6,67 pts

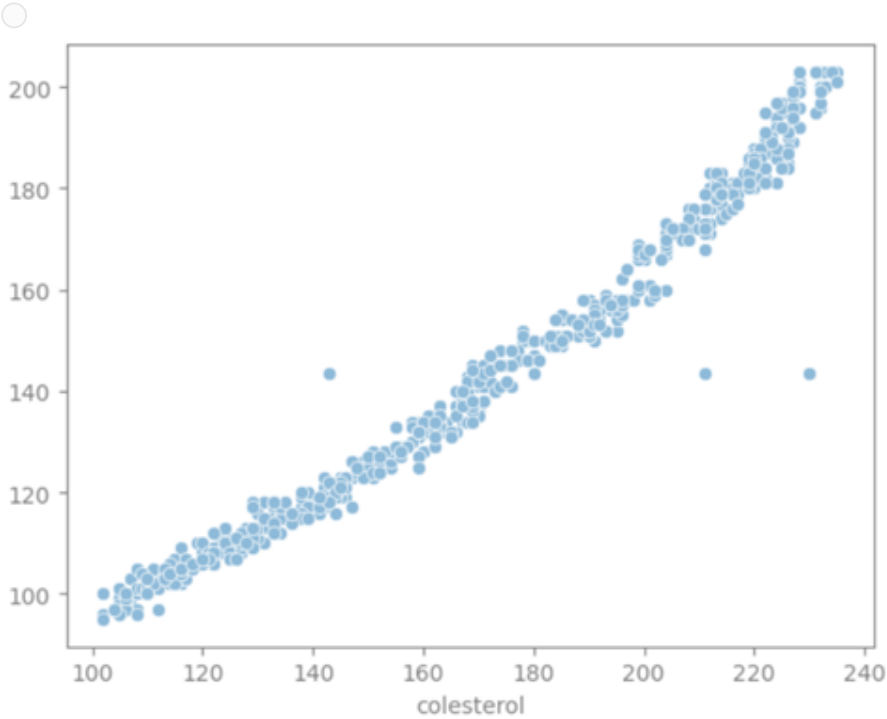
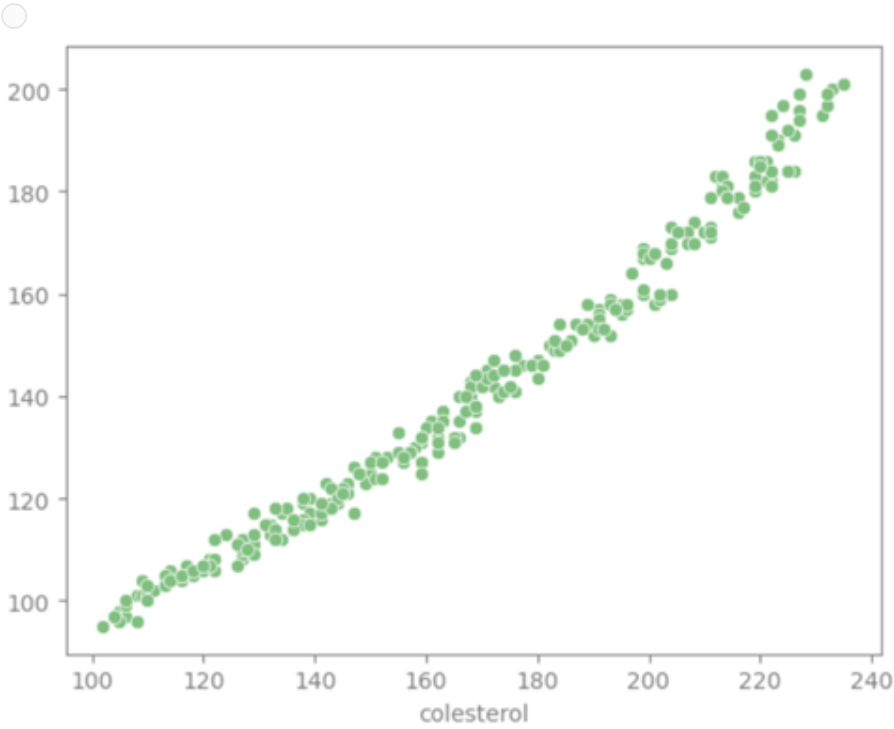
Qual é o estado cujo residente possui a maior renda, e qual é o estado civil dessa pessoa?

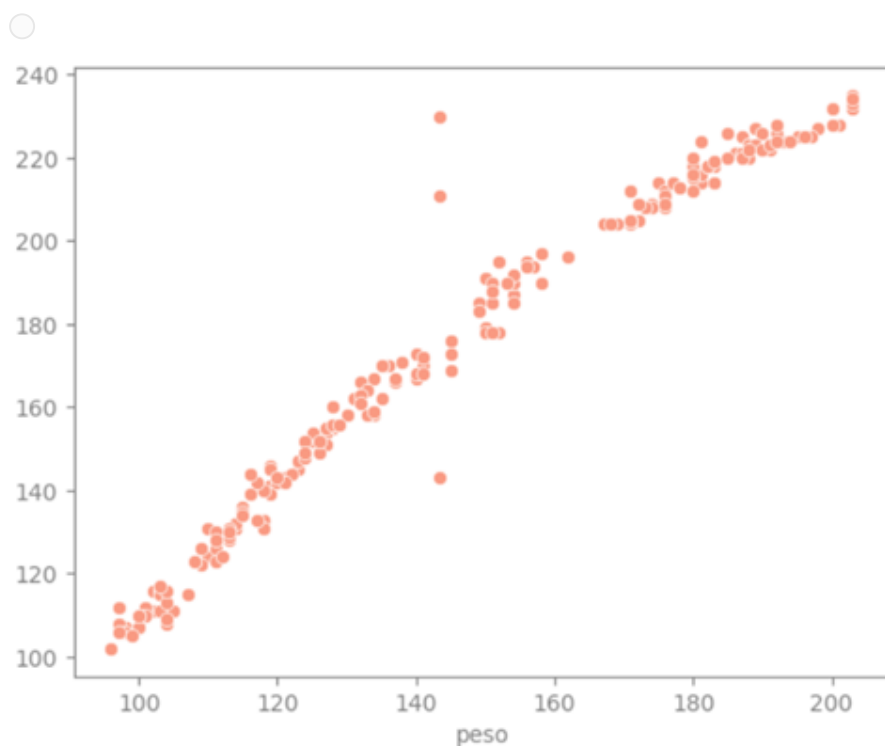
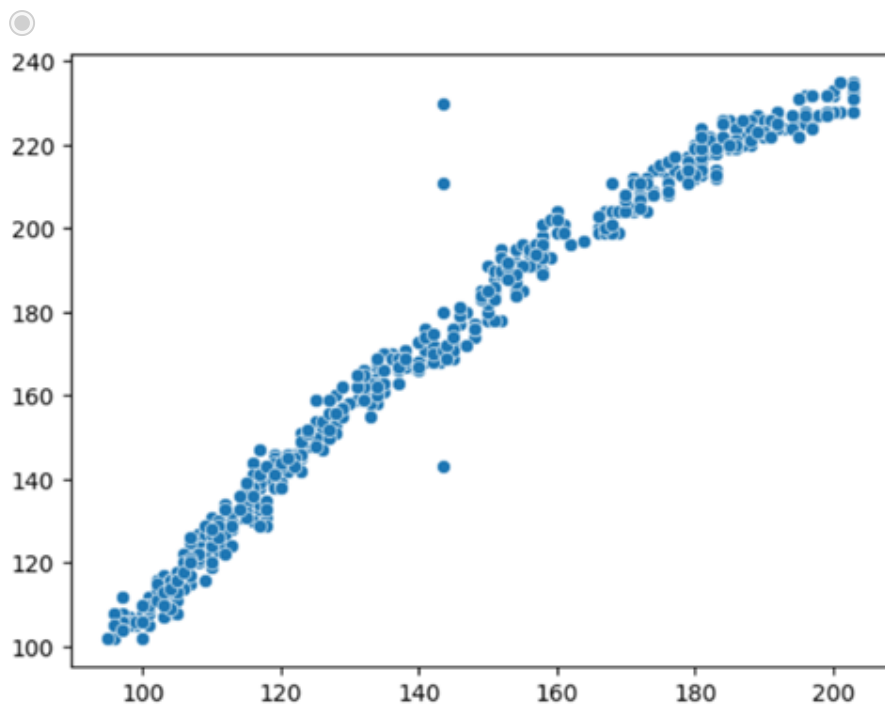
- ☐ Mato Grosso do Sul - Viúvo
- ☐ Rio Grande do Norte - Solteiro
- ☐ Rio Grande do Sul - Viúvo
- ☒ Mato Grosso do Sul - Casado

Pergunta 10

6,67 / 6,67 pts

Crie um gráfico de dispersão que represente as relações entre as variáveis peso (eixo x) e colesterol (eixo y) e selecione o gráfico que melhor ilustra essa relação.





Pergunta 11

6,67 / 6,67 pts

Ao aplicar um modelo de regressão linear usando machine learning, responda à seguinte pergunta: Qual é o valor estimado de colesterol para uma pessoa que possui um peso de 128 quilos, com base no modelo?

113

☒ 151.13☐ 149.27☐ 160**Pergunta 12****6,67 / 6,67 pts**

Qual é o valor do coeficiente angular da reta?

☐ 12.5☒ 1.25☐ 4.68☐ 9.08**Pergunta 13****6,67 / 6,67 pts**

Qual foi o coeficiente de determinação R^2 do modelo?

☐ 4.68☒ 0.97☐ 1☐ 0.87**Pergunta 14****6,67 / 6,67 pts**

Qual o valor do mean absolute error do modelo criado?

☐ 3.5☒ 4.6☐ 35.8☐ 1.25**Pergunta 15****6,62 / 6,62 pts**

Qual o valor do mean squared error do modelo criado?

☐ 3.58☒ 35.8☐ 27.94☐ 4.6Pontuação do teste: **100** de 100