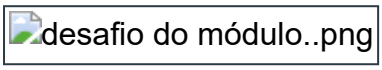
DESF5 - Desafio Final

Entrega 20 set em 23:59 **Pontos** 100 **Perguntas** 15
Disponível até 20 set em 23:59 **Limite de tempo** Nenhum
Tentativas permitidas 2

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 domingo-feira, **20/09/2024** 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/7642/files/614575?wrap=1)
(https://online.igti.com.br/courses/7642/files/614575/download?download_frd=1)

Bons estudos!

Atenciosamente,

Equipe XP Educação

Histórico de tentativas

	Tentativa	Tempo	Pontuação
MANTIDO	**Tentativa 2**	2 minutos	100 de 100
MAIS RECENTE	**Tentativa 2**	2 minutos	100 de 100
	Tentativa 1	106 minutos	93,33 de 100

As respostas corretas estarão disponíveis em 20 set em 23:59.

Pontuação desta tentativa: **100** de 100
Enviado 14 set em 17:38
Esta tentativa levou 2 minutos.

Pergunta 1

6,67 / 6,67 pts

Durante a análise inicial dos dados, foram identificados dados duplicados no dataset?

- Não. O dataset não possui dados duplicados.
- (●) Sim. Existem dados duplicados no dataset de preço de imóveis.
- Sim. Apenas no dataset de dados de estados.
- Sim. Apenas no dataset de dados de estados e de preço de imóveis.

Pergunta 2

6,67 / 6,67 pts

Após todos os tratamentos e integração de dados, responda: qual é a média da área construída para os imóveis da região sudeste?

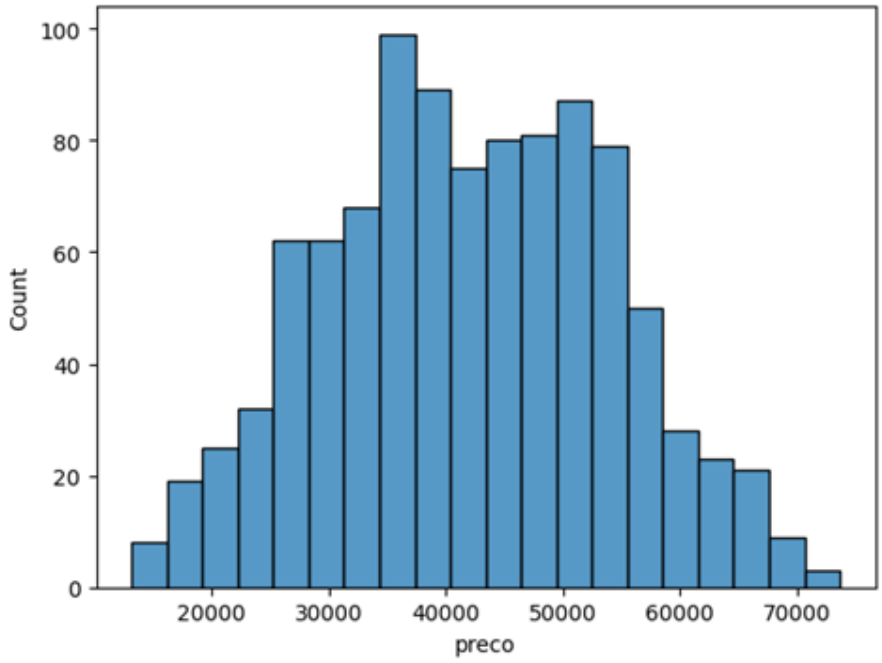
- 135.97.
- 204.16.
- 147.52.

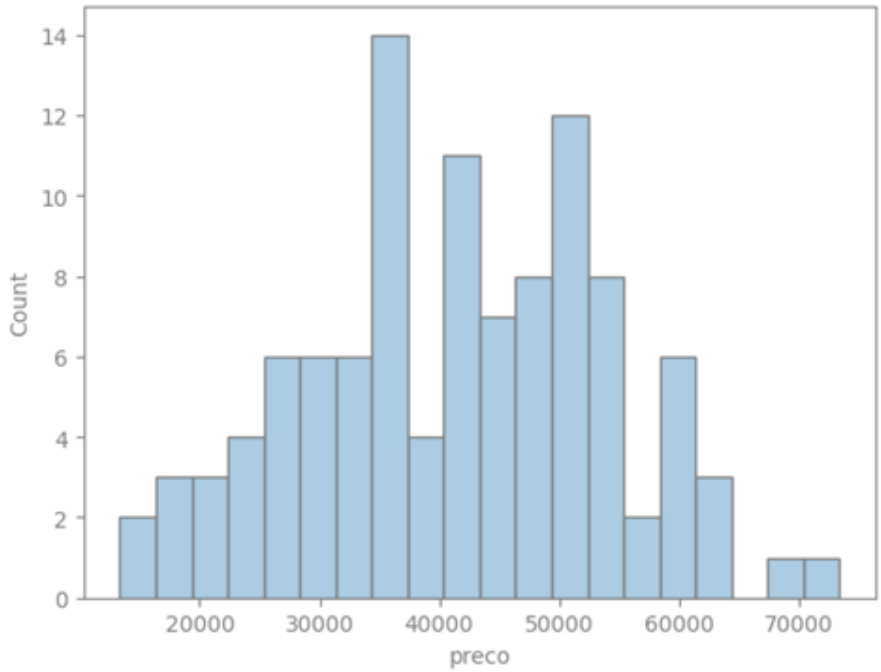
124.14.

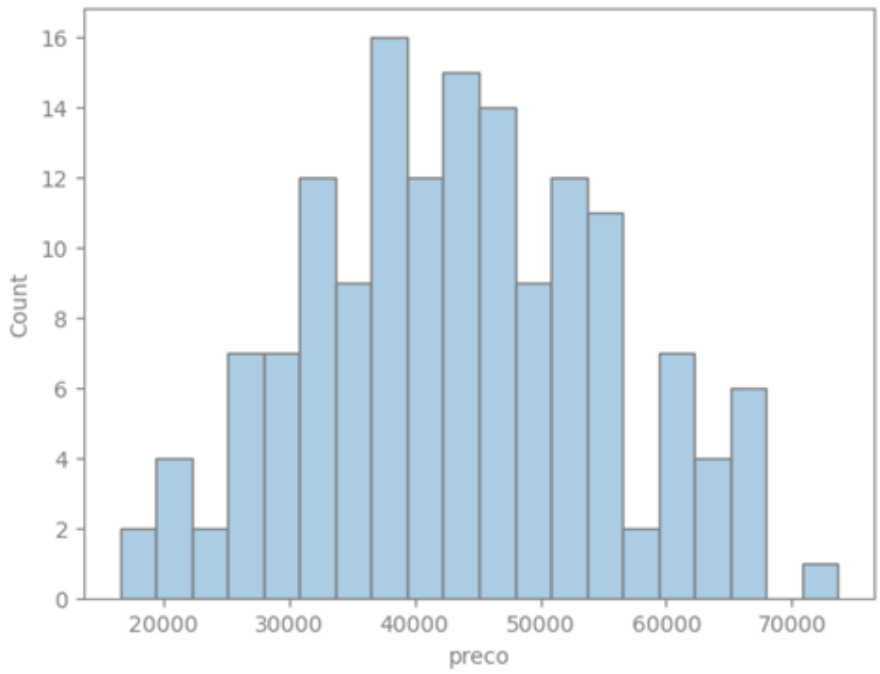
Pergunta 3

6,67 / 6,67 pts

Crie um histograma que ilustre a distribuição do consumo dos clientes ao longo do tempo, empregando 20 barras para a análise. Posteriormente, selecione a opção que melhor se adequa aos dados apresentados.





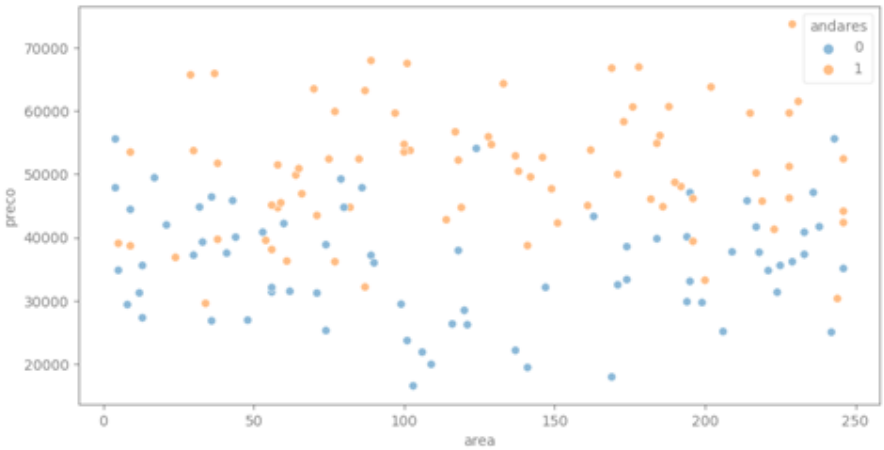


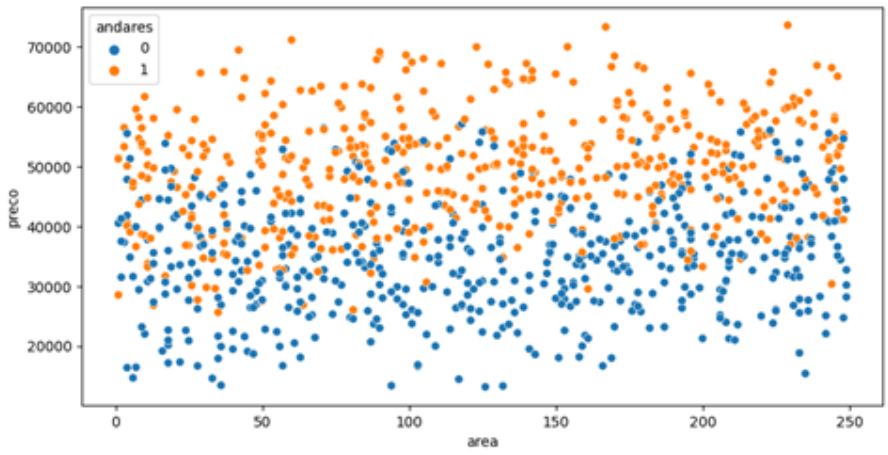
Nenhuma das alternativas está correta.

Pergunta 4

6,67 / 6,67 pts

Elabore um gráfico de dispersão que represente a relação entre a área e o preço, agrupados por andares. No eixo x, represente a área, enquanto no eixo y, represente os valores. Posteriormente, analise os resultados obtidos e selecione o gráfico que melhor ilustra essa relação.





Nenhuma das alternativas.

Pergunta 5

6,67 / 6,67 pts

Qual variável possui maior correlação com o valor do imóvel?

- Região.
- andares.
- Garagem.
- marmore.

Pergunta 6

6,67 / 6,67 pts

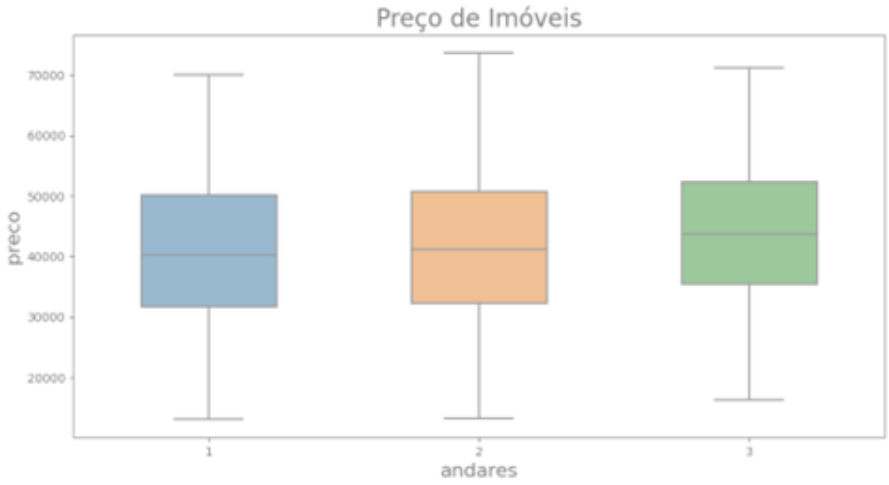
Qual região possui a maior média de preço de imóveis?

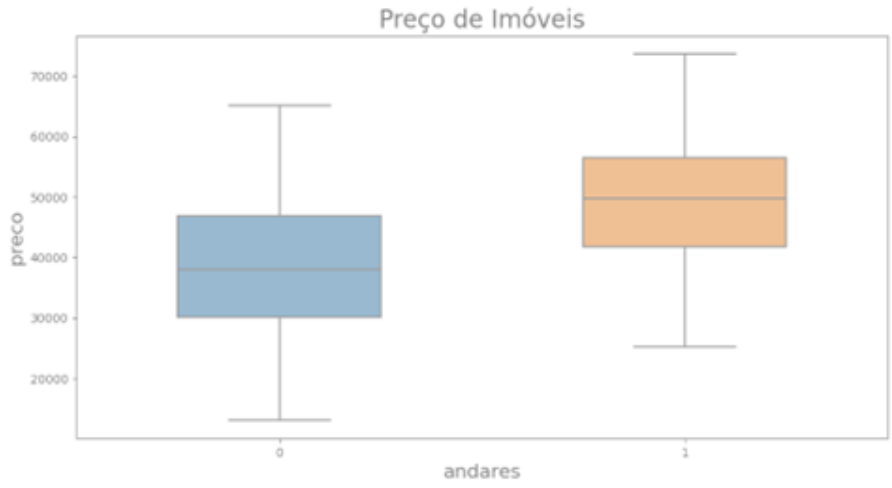
- Nordeste.
- Nordeste.
- Sul.
- Centro-oeste.

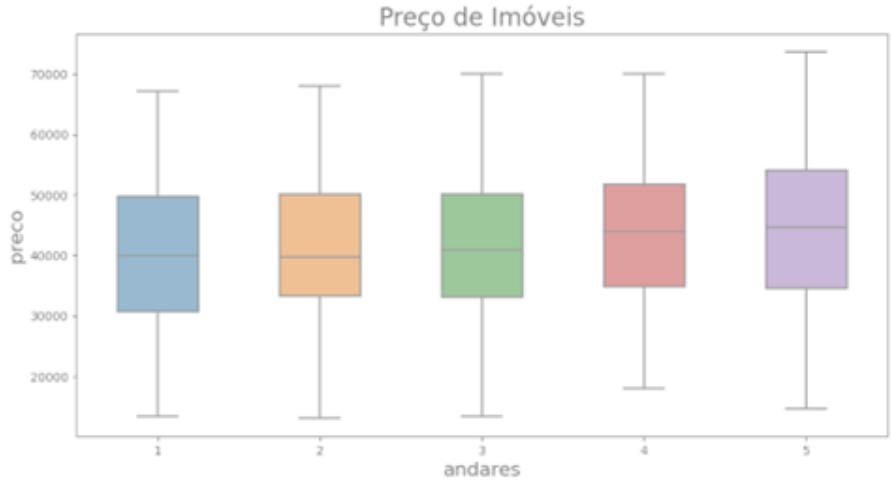
Pergunta 7

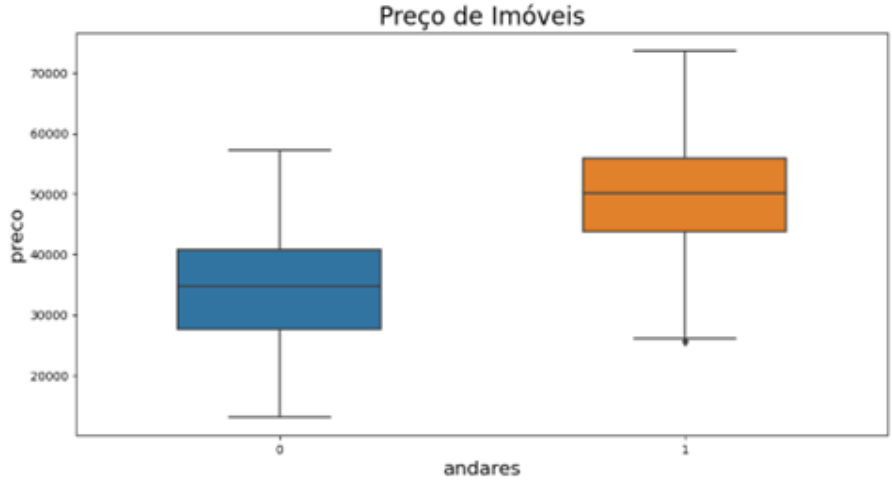
6,67 / 6,67 pts

Crie um gráfico de boxplot, onde o eixo y representa a variável preço e o eixo x representa o número de andares. Em seguida, escolha a melhor opção.









Pergunta 8

6,67 / 6,67 pts

Qual estado possui a maior média de preço de imóveis?

- Pernambuco.
- Minas Gerais.
- (selected) Amapá.
- São Paulo.

Pergunta 9

6,67 / 6,67 pts

Ao aplicar um modelo de regressão linear usando machine learning, responda à seguinte pergunta: qual é o preço predito para os dados de entrada abaixo?

- 50858.02
- (selected) 40419.12
- 72022.98
- 39661.74

Pergunta 10

6,67 / 6,67 pts

Ao aplicar um modelo de regressão linear usando machine learning, responda à seguinte pergunta: qual é o valor predito para os dados de entrada abaixo?

- 138450.00
- 204048.49
- 146054.65
- 205657.31

Pergunta 11

6,67 / 6,67 pts

Qual é o valor do coeficiente angular da reta para variável andares?

- 1034.09
- 803.06
- 23.19
- 15801.92

Pergunta 12

6,67 / 6,67 pts

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

- 0.80
- 0.91

- (●) 0.61
- ( ) 0.72

Pergunta 13

6,67 / 6,67 pts

Qual o valor do mean absolute error do modelo criado?

- ( ) 3.74
- (●) 6098.97
- ( ) 2.51
- ( ) 20.19

Pergunta 14

6,67 / 6,67 pts

Os algoritmos de aprendizado não supervisionado podem ser divididos em duas classes: Associação e Clusterização, que se definem respectivamente como:

- (●) Os algoritmos de associação permitem o descobrimento de regras e correlação em uma base de dados, identificando conjuntos de itens que ocorrem juntos dentro de uma determinada frequência, e os algoritmos de clusterização ou agrupamento permitem que seja feito agrupamento de grupos com base nas semelhanças encontradas.
- ( ) Ambos os algoritmos de associação não permitem o descobrimento de regras e correlação em uma base de dados. Eles apenas definem os conjuntos de itens que ocorrem juntos dentro de uma determinada frequência.

Os algoritmos de clusterização ou agrupamento permitem que seja feito agrupamento de grupos com base nas semelhanças não encontradas e os algoritmos de associação permitem o descobrimento de regras e associação em uma base de dados, identificando conjuntos de itens que ocorrem separadamente dentro de uma determinada frequência.

Nenhuma das alternativas está correta.

Pergunta 15

6,62 / 6,62 pts

Qual é o propósito principal do Apache Spark em processamento de dados distribuído?

Implementar algoritmos de aprendizado de máquina para análise de dados não estruturados.

Agilizar a execução de tarefas de processamento de dados em grandes clusters, através de computação distribuída em memória.

Todas as alternativas estão corretas.

Reduzir a complexidade do código Java para processamento de grandes conjuntos de dados.

Pontuação do teste: **100** de 100