

# AV I

**Entrega** 7 mai em 23:59**Pontos** 40**Perguntas** 10**Disponível** 29 abr em 0:00 - 7 mai em 23:59**Limite de tempo** 120 Minutos

## Instruções

### IMPORTANTE:

- A tentativa será finalizada após clicar no botão ENVIAR
- Programe se para realizar suas avaliações com tranquilidade, pois você terá 120 minutos cronometrados (por tentativa) para conclusão e envio das respostas. Após este prazo, a avaliação será automaticamente finalizada.
- Será permitida apenas UMA tentativa para realizar esta avaliação.
- Ao iniciar a avaliação o cronômetro não para, independentemente da plataforma estar aberta ou não.
- Durante a realização da prova, será exibido uma questão por vez, podendo AVANÇAR ou RETORNAR quando necessário, dentro do período da tentativa.
- Após abrir a avaliação mesmo não marcando nenhuma opção de resposta, será contabilizada uma tentativa.
- Somente finalize a avaliação após conferir as alternativas marcadas por você.



## Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	<a href="#">Tentativa 1</a>	54 minutos	21 de 40 *

\* Algumas perguntas ainda não avaliadas

⚠ As respostas corretas estarão disponíveis em 8 mai em 0:00.

Pontuação deste teste: **21** de 40 \*

Enviado 1 mai em 23:02

Esta tentativa levou 54 minutos.

### Pergunta 1

**3 / 3 pts**

Uma empresa vendedora de produtos de praia deseja analisar a quantidade de vendas do produto bola de praia na cidade de Florianópolis. A empresa deseja analisar as vendas do mês de fevereiro do ano passado realizadas pelo vendedor José Camilo, comparada com

os números da receita com o mesmo produto em fevereiro deste ano, também na cidade de Florianópolis. Após esta análise, a empresa fará uma comparação de outras vendas do mesmo produto em Florianópolis no mesmo período com o objetivo de realizar uma previsão de vendas.

Considerando o caso apresentado e o conteúdo estudado, assinale a alternativa que indica a melhor ferramenta a ser utilizada por essa empresa para realizar essa ação:

- ☒ Ferramentas OLAP.
- ☐ Mineração de dados.
- ☐ Big Data.
- ☐ Monodimensional Data Analytic.
- ☐ Business Intelligence.

#### Feedback:

O OLAP (Processamento Analítico Online) é um método de computação que permite aos usuários extrair e consultar dados de maneira fácil e seletiva a fim de analisá-los sob diferentes pontos de vista. As consultas de OLAP geralmente ajudam na análise de tendências, relatórios financeiros, previsão de vendas, orçamento e outros fins de planejamento. Você pode ler mais na unidade 1, tópico 3.

## Pergunta 2

3 / 3 pts

Segundo o Fórum Econômico Mundial, o mundo produz 2,5 quintilhões de bytes de dados todos os dias. Com tantos dados, tornou-se cada vez mais difícil gerenciar e entender tudo. Seria impossível para uma única pessoa percorrer os dados linha por linha, analisando os dados, buscando por padrões, distinguindo dados e fazer observações, tudo para uma tomada de ação estratégica.



Forum Economico Mundial. The Value of Data. Disponível em:<  
<https://www.weforum.org/agenda/2017/09/the-value-of-data/>>,  
Acesso em: 18/06/2020.

A proliferação de dados pode ser gerenciada como parte do processo de ciência de dados. De acordo com o seu conhecimento adquirido durante os estudos e os conceitos de ciência de dados, marque a alternativa que, corretamente apresenta a técnica de procurar padrões ocultos em um conjunto de dados:

☒ Mineração de Dados.

☐ Big Data.

☐ Data Warehouse.

☐ Ferramentas OLAP.

☐ Algoritmos genéticos.

#### Feedback:

A mineração de dados é o processo de analisar grandes bancos de informações para gerar novas informações. A mineração de dados trata de buscar por padrões e novos conhecimentos dos dados que você já coletou, porém que estão ocultos.

Você pode ler mais na unidade 1, tópico 3.

### Pergunta 3

3 / 3 pts

Alan Turing desenvolveu uma técnica para determinar se um computador poderia ou não demonstrar a Inteligência Artificial. Essa técnica ficou conhecida como teste de Turing. Pouco tempo depois, o filósofo John Searle apresentou o argumento do Quarto chinês.



Considerando essas informações e o conteúdo estudado sobre Inteligência Artificial, assinale a alternativa que melhor descreve o argumento do Quarto chinês:



A conclusão do argumento é que se uma máquina pode entrar em uma conversa com um humano sem ser detectada como uma máquina, ela demonstra ser inteligente como um humano.



A conclusão do argumento é que a programação de um computador digital pode fazer com que pareça que ele entende a linguagem, mas não pode produzir uma compreensão real.



A conclusão desse argumento é que, se um computador pode passar por humano no bate-papo online, devemos garantir que esse computador é realmente inteligente.



A conclusão do argumento é que, se submetidos as mesmas variáveis de entrada, tanto o teste de Turing quanto o Quarto chinês terão as mesmas variáveis de saída.



A conclusão do argumento do Quarto chinês possibilitou a John simular todos os tipos de algoritmos de computador, por mais complicado que esse algoritmo seja.



**Feedback:**

Searle argumenta que o experimento mental enfatiza o fato de que os computadores simplesmente usam regras sintáticas para manipular cadeias de símbolos, mas não têm entendimento de significado ou semântica. A conclusão mais ampla do argumento é que a teoria de que as mentes humanas são sistemas computacionais ou de processamento de informações é semelhante ao computador. Em vez disso, as mentes devem resultar de processos biológicos; os computadores podem, na melhor das hipóteses, simular esses processos biológicos.

Você pode ler mais na unidade 1, tópico 1.

**Incorreta****Pergunta 4****0 / 3 pts**

Analise com atenção o código-fonte a seguir:

```
1  a = 41
2  b = 9
3
4  if a > b:
5      aux = a
6      a = b
7      b = aux
8
9  print ("O valor de a é %d" % a)
10 print ("O valor de b é %d" % b)
```

De acordo com o conteúdo estudado sobre Python e após analisar o código-fonte, marque a alternativa correta:

☐

O código apresentará um erro quando o valor de a for igual ao valor de b, uma vez que não existe uma condição que contemple essa igualdade.

☐

A linha 5 e 7 poderiam ser suprimidas, ficando apenas os comandos a = b e b = a dentro do escopo do if.



O valor de a após a execução do comando da linha 6 é 9 e o valor de b após a execução do mesmo comando é 41.



O código apresentará um erro na linha 4 uma vez que existe um sinal desconhecido na linguagem que são os dois pontos :



O objetivo desse código é apresentar uma sequência ordenada nas linhas 9 e 10, e a linha 5 é essencial nesse processo.

**Feedback:**

O algoritmo realiza a troca dos valores entre as variáveis a e b sempre que a variável a for maior que a variável b. Entretanto, se fizer uma troca direta, ou seja, a receber b e b receber a, perderemos o valor da variável a. Desse modo, precisamos de um auxiliar que guarde para nós o valor de a, então isso é feito na linha 5. A única condição de entrarmos nesse if é quando a variável a for maior que a variável b, logo estamos ordenando os valores.

Você pode ver mais sobre variáveis no tópico 1 e sobre estrutura condicional no tópico 2.

**Incorreta****Pergunta 5****0 / 3 pts**

As variáveis, em linguagem de programação, são representações textuais de espaços na memória onde podemos guardar valores. Temos as variáveis de tipo inteiro, responsável por guardar números inteiros, e variáveis do tipo real, que guardam números reais. Temos variáveis do tipo cadeia de caracteres, que armazenam palavras e frases.

Também temos os tipos booleanos, os quais retornam os valores VERDADEIRO ou FALSO. Assim, analise com atenção o código a seguir:

```
1 cond1 = (True and (True and not False)) and not (not False)
2 cond2 = not ((not (not (not False)) and True))
3 cond3 = (not (2 != 4) and (3 < 6)) or True
4 cond4 = (4 > (2 < 3)) or not True
5 cond5 = ((True or True) or True) and False
6
7 if cond1:
8     print ("Brasil")
9 if cond2 == True:
10    print ("Estados Unidos")
11    if cond3:
12        print ("México")
13 if cond4:
14    print ("Itália")
15    if cond5:
16        print ("Inglaterra")
```

Assim, considerando as informações apresentadas e os conteúdos estudados sobre estruturas condicionais, o que é correto falar sobre a saída desse código?

- ☐ Será apresentada a mensagem "Brasil Itália".
- ☐ Será apresentada na tela as mensagens "México Itália".
- ☐ Será apresentada na tela a mensagem "Inglaterra".
- ☒ Será apresentada na tela a mensagem "México".
- ☐ Será apresentada na tela a mensagem "Itália".



**Feedback:**

A variável cond3 receberá na linha 3 e a variável cond4 receberá na linha 4 o valor igual a True. Entretanto, na linha 11 a condição if é uma condição interna a outro if, o da linha 9. Ou seja, a única possibilidade de realizar o teste da linha 11 é se o if da linha 9 for verdadeiro, o que não é, por isso não é impressa a mensagem México. Se retirarmos a tabulação do if da linha 11, dessa forma sim será apresentada a palavra México. Por outro lado, na linha 13, o if verificará se cond4 é verdadeiro e, como é verdadeiro, então mostra a mensagem Itália.

Veja mais sobre estrutura condicional no tópico 2.

**Pergunta 6****3 / 3 pts**

O proprietário de um restaurante, vendo a baixa nas vendas diárias, decidiu realizar uma pesquisa com o objetivo de saber a opinião dos clientes sobre os seus pratos. Durante 7 dias ele pediu aos clientes para avaliarem o serviço com notas que iam de zero (muito insatisfeito) até dez (muito satisfeito).

No findar dos 7 dias, o proprietário estava com os seguintes dados: 20 clientes atribuíram a nota 0 ao restaurante; 180 dos entrevistados atribuíram a nota 10; e 200 clientes deram a nota 5 ao restaurante, totalizando 400 clientes. Como forma de incentivo, o proprietário ofereceu R\$200,00 aos empregados se acaso a média, a moda ou a mediana estivessem acima de 7,5 pontos. Por outro lado, caso os clientes tenham avaliado muito mal e a média for menor ou igual a 5, o proprietário dará desconto de 10% aos clientes.

A partir dessas informações e do conteúdo estudado no livro da disciplina, podemos considerar que o que acontecerá no restaurante será:



A média foi 4, a mediana foi 4 e a moda que foi obtida pelo método mode() resultou em 8. Dessa forma, os funcionários receberão R\$200,00 e os clientes terão desconto de 10%.





Com o comando `mean()` foi obtida a média 7, com o comando `median()` foi obtida a mediana 5 e a moda obtida pelo comando `mode()` foi 5. Dessa forma, não mudará nada no restaurante



A média obtida com o comando `median()` foi 4, a mediana foi 6 e a moda foi 5. Dessa forma, o proprietário dará desconto de 10% aos clientes e não mudará nada em relação aos funcionários



A média obtida com o comando `median()` foi 8, a mediana foi 5 e a moda obtida pelo comando `mode()` foi 5, logo não mudará nada no restaurante.

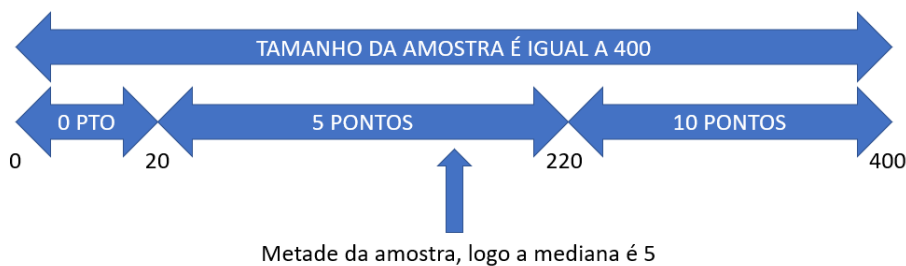


Com o comando `mean()` obteve-se média igual a 7,5, a mediana foi 10 e a moda foi 5. Dessa forma, o proprietário dará R\$200,00 aos funcionários e não mudará nada em relação aos clientes



**Feedback:**

A figura abaixo ilustra o processo de análise para obtermos a moda e a mediana. Note que o tamanho da amostra é igual a 400. Vamos ordenar os dados do menor para o maior (ou seja, de zero a dez). A primeira seta da segunda linha indica que, do intervalo de 0 a 20, essas pessoas deram nota zero. A segunda seta, ainda na mesma linha, indica que, no intervalo de 20 a 220, os entrevistados deram 5 pontos. Por fim, a última seta indica que, do intervalo entre 220 e 400, os entrevistados deram 10 pontos.



Note que, se colocarmos uma marcação bem no centro da segunda linha, essa marcação cairá exatamente sobre o intervalo de 20 a 220, ou seja, intervalo de 5 pontos. Logo, a mediana é 5. A moda é o valor que mais aparece que, nesse caso, também é o 5. Por outro lado, a média é a soma de todos os valores dividida pelo tamanho da amostra, que é  $400 - 2800/400 = 7$ , logo a média é 7. Assim temos: média 7, mediana e moda 5. Os respectivos métodos para encontrarmos esses valores são: para média é `mean()`, para mediana é `median()` e para moda é o `mode()`. Dessa forma, de acordo com os dados do enunciado, não haverá mudança no restaurante.

Veja mais no tópico 2 da unidade 3.

**Pergunta 7****3 / 3 pts**

Leia o trecho a seguir:

Em uma pesquisa, o entrevistador fez a seguinte pergunta: “A caridade é o melhor meio de ajudar as pessoas que mais necessitam?”. Foi elaborado um questionário com 10 pontuações em uma escala, em que

1 (valor mínimo) é “Discordo totalmente” e 10 (valor máximo) é “Concordo totalmente”. A seguir encontram-se os dados coletados.

6	4	7	2	9
1	4	4	6	4
1	6	9	8	4
2	7	1	8	3

Considerando esses dados, bem como as informações fornecidas, pode-se considerar como informações corretas a respeito dos dados da amostra:



Pode-se dizer que pelo menos 60% das pessoas deram nota menor do que a média.



O valor modal (moda) é o que aparece com maior frequência e, nesse caso, foi o número 1.



O desvio padrão para essa amostra é igual ao valor da média e é obtido pelo método `desvpad()`.



A mediana e a média possuem o mesmo valor. Nesse caso, podem ser obtidas pelo método `median()`.



O valor modal (moda) é 4 e corresponde a 25% da amostra. Pode ser obtido com o método `mode()`.



**Feedback:**

Para esse exercício, coloque todos os dados em ordem crescente. Como a amostra possui um número par de dados, haverá 2 números centrais cujo valor é 4. Ache a média desses dois valores e, então, teremos a mediana, que é 4. Na amostra temos 5 entradas para o número 4, logo  $5/20$  ou  $1/4$  ou 25% da amostra corresponde ao valor 4. Por fim, para obtermos o valor modal, utilizamos o método `mode()`.

Veja mais no tópico 2 da unidade 3.

**Pergunta 8****Não avaliado ainda / 13 pts**

aprendizado de máquina (ou machine learning) tem como premissa principal a aprendizagem a partir dos dados (MITCHELL, 1997). Dessa forma, procura-se inferir um conceito mais geral a partir da observação de dados e exemplos passados. Quando se fala em Ciência de dados, há basicamente três paradigmas do aprendizado de máquina que são importantes conhecer. Explique brevemente cada uma delas.

Sua Resposta:

Os três tipos de aprendizado de máquina são: Supervisionado, não supervisionado e por reforço.

No aprendizado supervisionado, usamos exemplos rotulados para ensinar o modelo a fazer previsões precisas. Por exemplo, ensinamos um modelo a identificar se uma imagem é de um cachorro ou um gato mostrando várias imagens rotuladas para ele.

No aprendizado não supervisionado, usamos dados não rotulados para encontrar padrões e estruturas nos dados. Por exemplo, podemos usar aprendizado não supervisionado para agrupar clientes em diferentes grupos de acordo com seu comportamento de compra

No aprendizado por reforço, o modelo aprende a tomar decisões com base em recompensas e penalidades. Por exemplo, podemos ensinar um robô a jogar futebol usando aprendizado por reforço, onde ele aprende a tomar ações para maximizar a pontuação.

**PAG 77**

No aprendizado supervisionado, tendo como base um conjunto de dados, o algoritmo procura inferir o valor da variável dependente a partir de um conjunto de variáveis independentes.

No aprendizado não supervisionado, existem somente as variáveis independentes que descrevem os dados,

O aprendizado por reforço o elemento-chave é o estímulo-resposta, no qual um agente produz respostas comportamentais a depender de estímulos positivos ou negativos, os quais, por sua vez, vão gerar novas respostas em um ciclo.

**Pergunta 9****3 / 3 pts**

Nos conjuntos de dados, é preciso observar, além da distribuição dos dados, a ordem de grandeza das variáveis (colunas) que os contêm, em especial quando houver discrepância entre as próprias variáveis de um mesmo conjunto de dados. Para remediar isso, é possível empregar técnicas de padronização e normalização.

Se você utilizar uma técnica de normalização baseada em Min-Max, em uma determinada coluna, o que se pode afirmar em relação aos valores da média e do desvio padrão mínimo e máximo?



O mínimo será o valor da média menos o desvio padrão e o máximo será o valor da média mais o desvio padrão.



A média será igual a 0 e o desvio padrão será igual a 1.



O mínimo será 0 e o máximo será 1.



A média será igual a 1 e o desvio padrão será igual a 0.



Nada se pode afirmar, porque não se conhece a distribuição dos dados.

**Feedback:**

Pela própria característica da fórmula do Min-Max, a normalização é feita de forma que os novos valores são calculados dentro de um regime intervalar entre o mínimo e o máximo da distribuição. Isso faz com que os novos valores fiquem sempre dentro do intervalo  $[0, 1]$ , ou seja, mínimo = 0 e máximo = 1.

Reveja esses conceitos na unidade 5, tópico 2.

**Pergunta 10****3 / 3 pts**

A presença de dados ausentes em conjuntos de dados é algo presente na maior parte das informações provenientes do mundo real. É possível remediar a existência de dados ausentes utilizando técnicas estatísticas, como a média, a mediana e a moda.

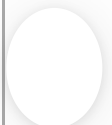
Marque a alternativa INCORRETA a respeito das razões que propiciam a ocorrência de dados ausentes:



Campos não obrigatórios que, ao serem preenchidos, são deixados em branco ou zerados, como idade, endereço, renda etc.



Campos obrigatórios, mas sem a correta consistência que obrigue a correta captura do preenchimento (nome completo, CPF, RG etc.).





Campos que possuam cálculo automático dependente de outros campos do conjunto de dados e que o sistema aleatoriamente possa estar deixando de executar essa tarefa, que fica pré-agendada no sistema.



Campos com opção correta faltante, por exemplo, uma lista de cidades onde não consta uma delas em particular. A pessoa, ao preencher e notar que não tem a sua cidade, provavelmente irá deixar em branco.



Campos que dependem de uma ação temporal, por exemplo, os dados de navegação em um e-commerce, em que o nome do produto adquirido e o valor somente existirão no conjunto de dados se o cliente fizer a compra, senão eles estarão em branco até que o referido evento ocorra.



### Feedback:

A existência de dados ausentes sempre envolve uma questão de modelagem do sistema para fazer a aquisição das informações, seja por negligência de validação dos campos e/ou pela coleta esparsa em tabelas pré-formatadas. Os sistemas de informação simplesmente vão fazer o que o programador mandou e não o que ele desejaria que pudesse ser feito, já que os sistemas não agem por conta própria e de forma aleatória.

Reveja esses conceitos na unidade 5, tópico 1.1.

Pontuação do teste: **21** de 40

