

- Nas questões da prova objetiva a seguir, marque, para cada uma, a única opção correta, de acordo com o respectivo comando. Para as devidas marcações, use a **Folha de Respostas**, único documento válido para a correção das suas respostas.
- Caso haja questões que avaliem **conhecimentos de informática** e(ou) **tecnologia da informação**, a menos que seja explicitamente informado o contrário, considere que todos os programas mencionados estão em configuração-padrão e que não há restrições de proteção, de funcionamento e de uso em relação aos programas, arquivos, diretórios, recursos e equipamentos mencionados.
- Eventuais espaços livres — identificados ou não pela expressão “Espaço livre” — que constarem deste caderno de provas poderão ser utilizados para rascunho.

-- PROVA DE CONHECIMENTOS --

-- PARTE I – PROVA OBJETIVA --

LÍNGUA PORTUGUESA

Texto 1A8

Já se foi o tempo em que acordar tarde e tomar um café da manhã reforçado na cobertura de um *duplex* era símbolo de *status* e ascensão social. Hoje em dia, emergente de verdade toma é um *brunch* no *roof top* de seu *residence club*.

A classe média, quando chega ao paraíso, adora *gourmetizar* o idioma, recheando o bom e velho português tupiniquim com um punhado de expressões de outras línguas para dar um ar de sofisticação e requinte a boa parte daquilo que diz, faz e pensa. Alguns exemplos são *crush*, *delivery*, *fashion*, *haters*, *job*, *like*, e por aí vai.

O discurso empresarial e corporativo também está repleto desses enunciados carregados de expressões tomadas de empréstimo a outras línguas. Nesse quesito, o inglês é o campeão de audiência nas interações verbais de uma parte considerável das profissões urbanas. Palavras e expressões como *briefing*, *workshop*, *call*, *mindset* e *deadline* já se tornaram corriqueiras no vocabulário de muitos profissionais brazucas.

Não que isso seja um problema ou uma ameaça à integridade do português brasileiro. Nosso país, desde a sua origem, sempre teve um caráter plurilinguístico. O fundo lexical do nosso português está repleto de vocábulos e expressões vindos de uma porção de idiomas que, ao longo do tempo, foram incorporados ao falar brasileiro.

A questão a ser abordada não é o quanto o uso de estrangeirismo pode ser nocivo à língua, mas até que ponto a hegemonia político-social de determinado grupo de falantes pode exercer um domínio ideológico e cultural sobre as dinâmicas linguísticas praticadas por boa parte de nossa população. A censura linguística de comunidades subalternizadas pode ser muito mais nociva à língua do que o uso de palavras e(ou) expressões de origem estrangeira.

Nesse contexto, a negação dos acréscimos linguísticos vindos de grupos étnicos e sociais diversos não se justifica, porque propõe o apagamento intencional e deliberado de nossas raízes históricas, linguísticas e culturais.

A presença de estrangeirismos nas vitrines das lojas e nos hábitos linguísticos de determinada comunidade de falantes, embora provoque reações afetivas e passionais tão variadas quanto contraditórias, é perfeitamente compreensível, porque as línguas são espaços de comunicação nos quais os processos de interlocução não se esgotam no simples ato de dizer.

Questão 1

Entende-se da leitura do texto 1A8 que a principal preocupação do autor em relação ao uso de estrangeirismos no Brasil diz respeito

- A à submissão dos brasileiros ao modo de agir e pensar dos estrangeiros, sobretudo os estadunidenses.
- B ao risco de perda da pureza e da identidade original da língua portuguesa.
- C às falhas de comunicação decorrentes do desconhecimento do significado de determinadas expressões estrangeiras.
- D ao nível de dominação de grupos sociais hegemônicos sobre as demais parcelas da população.

Questão 2

No último parágrafo do texto 1A8, o autor qualifica como “perfeitamente compreensível”

- A os hábitos linguísticos de determinados grupos de falantes.
- B o fato de as reações aos estrangeirismos serem, ao mesmo tempo, variadas e contraditórias.
- C a presença de estrangeirismos em vitrines de lojas.
- D o fato de estrangeirismos provocarem reações afetivas e passionais.

Questão 3

Cada uma das opções a seguir apresenta uma proposta de reescrita em que se unem o primeiro e o segundo períodos do quarto parágrafo do texto 1A8. Assinale a opção em que a proposta apresentada mantém a correção gramatical, a coerência e os sentidos do texto.

- A Não que isso seja um problema ou uma ameaça à integridade do português brasileiro, tendo em vista que, nosso país, desde a sua origem, sempre teve um caráter plurilinguístico.
- B Não que isso seja um problema ou uma ameaça à integridade do português brasileiro, no entanto nosso país, desde sua origem, sempre teve um caráter plurilinguístico.
- C Não que isso seja um problema ou uma ameaça à integridade do português brasileiro, pois nosso país, desde a sua origem, sempre teve um caráter plurilinguístico.
- D Não que isso seja um problema ou uma ameaça à integridade do português brasileiro, enquanto nosso país, desde a sua origem, sempre teve um caráter plurilinguístico.

Questão 4

No último período do quinto parágrafo do texto 1A8, é estabelecida uma relação de

- A proporcionalidade.
- B comparação.
- C contraste.
- D causa e consequência.

Questão 5

No último período do terceiro parágrafo do texto 1A8, o vocábulo “corriqueiras” é empregado com o mesmo sentido de

- A hodiernas.
- B esporádicas.
- C prescindíveis.
- D triviais.

Texto 1A9

Em uma sociedade em que a imagem pública e a privada constituem fatores preponderantes de prestígio, credibilidade e liderança, esses elementos estão, cada vez mais, permeando nossas vidas, nossas formas de agir e de decidir, que estão sob a vigilância de olhares sociais atentos e fiscalizadores. Os ritos e as cerimônias são elementos estratégicos a serviço da construção e da consolidação das imagens das organizações, que se apoiam na credibilidade e na aceitação social das ações e realizações desenvolvidas.

O rito consiste em um conjunto de atos formalizados, expressivos e portadores de uma dimensão simbólica. É caracterizado por uma configuração espaço-temporal específica, por sistemas de linguagem e comportamentos específicos e por signos emblemáticos cujo sentido codificado constitui um dos bens comuns do grupo. O uso do rito é paralelo ao aparecimento da humanidade.

Em todas as sociedades, os grupos sociais participam de acontecimentos ou eventos especiais e únicos. Porém, para cada um desses grupos, os eventos têm um significado diferente. No Brasil, por exemplo, a Copa do Mundo e uma formatura são eventos com rituais reconhecidos por diferentes classes sociais e culturais. Um rito bem executado é mais que uma mera apresentação teatral. Usa elementos e símbolos e evoca a cultura e as crenças dos povos envolvidos.

Os ritos podem ser vistos como algo que não se resume a repetições das coisas reais e concretas do mundo rotineiro. Como coisa real e concreta, consistem no que pode ser materializado e simbolizado. Exemplo disso é a troca de presentes entre personalidades de diferentes culturas ou o aperto de mão entre duas pessoas que se saúdam. Este protocolo é uma forma de comunicação na qual os participantes do processo denotam uma mensagem diplomática. Remete, ainda, ao protocolo elementar significativo de boas relações entre povos, governos ou grupos.

Internet: <uel.br> (com adaptações).

Questão 6

De acordo com as ideias do texto 1A9, os ritos

- A são tão antigos quanto a humanidade.
- B são a base das imagens das organizações.
- C são os fatores que mais interferem na consolidação do prestígio das organizações.
- D não repetem as coisas reais e concretas do mundo rotineiro.

Questão 7

A expressão “em que” (primeiro período do primeiro parágrafo) poderia ser substituída, mantendo-se a correção e o sentido do texto 1A9, por

- A cuja.
- B que.
- C na qual.
- D pela qual.

Questão 8

Assinale a opção em que é corretamente apresentado o significado da palavra “preponderantes” (primeiro período do primeiro parágrafo).

- A relevantes
- B indispensáveis
- C mais conhecidos
- D de maior importância

Questão 9

Assinale a opção em que é apresentada proposta de reescrita gramaticalmente correta e coerente para o último período do terceiro parágrafo do texto 1A9.

- A A partir dele, é usado elementos e símbolos e é evocado a cultura e as crenças dos povos envolvidos.
- B Nele, usam-se elementos e símbolos e evocam-se a cultura e as crenças dos povos envolvidos.
- C Nele se usa elementos e símbolos e evocam a cultura e as crenças dos povos envolvidos.
- D Dele se usa elementos e símbolos e se evocam a cultura e as crenças dos povos envolvidos.

Questão 10

No que se refere ao emprego do sinal indicativo de crase, a correção gramatical e a coerência do texto 1A9 seriam mantidas caso se substituisse

- A “a repetições” (primeiro período do último parágrafo) por **à repetições**.
- B “sob a vigilância” (primeiro período do primeiro parágrafo) por **ante à vigília**.
- C “ao aparecimento” (último período do segundo parágrafo) por **à aparição**.
- D “a serviço” (segundo período do primeiro parágrafo) por **à serviço**.

CONHECIMENTOS DO SEBRAE**Questão 11**

Em relação à história do SEBRAE, assinale a opção correta.

- A Atualmente, o SEBRAE alcançou a posição de órgão público autônomo, sem fins lucrativos, mantido por repasses de grandes empresas, o que possibilitou a ampliação de sua estrutura de atendimento para todos os estados do país.
- B Nos governos Sarney e Collor (1985-1990), o SEBRAE, então denominado Centro Brasileiro de Assistência Gerencial à Pequena Empresa (CEBRAE), passou a ser vinculado ao Ministério da Indústria e Comércio, período em que enfrentou crises por instabilidade orçamentária.
- C A história do SEBRAE inicia-se com a sua criação institucional em 1972, por iniciativa do Banco Nacional de Desenvolvimento Econômico e do Ministério do Planejamento.
- D A atuação política do SEBRAE surge a partir da promulgação da Constituição Federal de 1988, que inseriu em seu texto o dever de o Estado promover a criação e o desenvolvimento das microempresas e pequenas empresas.

Questão 12

Integra(m) o público-alvo do SEBRAE

- A a pessoa física que ainda não possua um negócio, desde que esteja efetivamente envolvida na sua estruturação.
- B a microempresa e a empresa de pequeno porte, excluindo-se, por consequência, as empresas de médio e grande porte.
- C a pessoa física com negócio próprio, desde que possua registro no Cadastro Nacional de Pessoa Jurídica (CNPJ).
- D o microempreendedor individual, o produtor rural e o artesão.

Questão 13

No que concerne aos produtos e serviços oferecidos pelo SEBRAE, assinale a opção correta.

- Ⓐ O aplicativo Up Digital SEBRAE é uma plataforma voltada à criação de planos de negócios completos, totalmente digital, podendo ser acessada de qualquer lugar e em qualquer horário.
- Ⓑ O programa Respostas e Dossiês Técnicos gera documentos a partir de buscas e análise de informações, para capacitação e inovação tecnológica.
- Ⓒ O Emissor de NF-e, programa versão *web* para emissão de nota fiscal eletrônica, é disponibilizado pelo SEBRAE ao seu público mediante o pagamento de uma taxa módica de utilização.
- Ⓓ O PNBOX SEBRAE é um programa de aceleração digital, que busca potencializar os negócios, a partir da organização das finanças e da melhoria dos canais digitais para a conquista de novos clientes.

Questão 14

À luz da Lei Geral de Proteção de Dados Pessoais (Lei n.º 13.709/2018), assinale a opção correta.

- Ⓐ É vedado o tratamento de dados pessoais como condição para o fornecimento de produto ou de serviço ou para o exercício de direito.
- Ⓑ O tratamento de dados pessoais não poderá ocorrer à revelia do consentimento do titular dos dados.
- Ⓒ As atividades de tratamento de dados pessoais devem observar o princípio da finalidade, segundo o qual há a necessidade de o tratamento ser compatível com a finalidade informada ao titular.
- Ⓓ O SEBRAE e as microempresas e empresas de pequeno porte que com ele se relacionam não são titulares de dados pessoais sensíveis.

Questão 15

Suponha que determinada microempresa esteja promovendo algumas mudanças e, nesse ínterim, tenha adotado (i) práticas de controle dos seus processos; (ii) práticas de valorização da saúde e segurança no ambiente de trabalho; (iii) práticas para promover a transparência na política de remuneração dos seus diretores; e (iv) práticas para promover redução na emissão de poluentes. Considerando essa situação hipotética e as práticas de ESG (*environmental, social and governance*) utilizadas pelo SEBRAE, assinale a opção correta acerca das ações promovidas pela referida microempresa.

- Ⓐ As ações (i) e (iii) são relativas à governança, a ação (ii) é de caráter social, e a ação (iv), de caráter ambiental.
- Ⓑ As ações (i) e (ii) são relativas à governança, a ação (iii) é de caráter social, e a ação (iv), de caráter ambiental.
- Ⓒ A ação (i) é de caráter ambiental, a ação (ii) é relativa à governança, e as ações (iii) e (iv) são de caráter social.
- Ⓓ As ações (i) e (iii) são de caráter social, a ação (ii) é relativa à governança, e a ação (iv) é de caráter ambiental.

CONHECIMENTOS ESPECÍFICOS**Questão 16**

Nos seguintes modelos de séries temporais, X_t representa uma observação e w_t denota um ruído branco no instante $t \in \mathbb{Z}$.

I $X_t = X_{t-1} - 0,25X_{t-2} + w_t - 0,5w_{t-1}$

II $X_t = w_t - 0,5w_{t-1}$

III $X_t = 0,8X_{t-1} - 0,5w_t$

IV $X_t = 0,09X_{t-2} + w_t - 0,3w_{t-1}$

Considerando os modelos de séries temporais apresentados, assinale a opção em que é corretamente indicada a quantidade de modelos cuja função de autocorrelação apresenta a forma de um processo autorregressivo de primeira ordem.

- Ⓐ 0
- Ⓑ 1
- Ⓒ 2
- Ⓓ 3

Questão 17

Se $\{y_t\}$ for uma série temporal fracamente estacionária descrita pelo modelo na forma $y_t = 3 + 0,7y_{t-1} + \varepsilon_t$, no qual ε_t é um ruído branco com média nula e $t \in \mathbb{Z}$, então o valor esperado da variável y_t será igual a

- Ⓐ 2,1.
- Ⓑ 3.
- Ⓒ 3,7.
- Ⓓ 10.

Questão 18

A respeito do modelo de séries temporais

$$S_t = \varepsilon_t + \varepsilon_{t-12} + \varepsilon_{t-24} + \varepsilon_{t-36} + \dots = \sum_{k=0}^{+\infty} \varepsilon_{t-12k},$$

no qual $t \in \mathbb{Z}$ representa um índice temporal e ε_t denota um erro aleatório no instante t , que segue uma distribuição normal com média zero e desvio padrão 5, assinale a opção correta.

- Ⓐ A média da série temporal $\{S_t\}$ descreve um padrão sazonal com período igual a 12.
- Ⓑ A série temporal diferenciada $\{S_t - S_{t-12}\}$ apresenta um padrão sazonal.
- Ⓒ A série temporal $\{S_t\}$ não é estacionária.
- Ⓓ A variância da variável S_t é igual a 25.

Questão 19

Um analista pretende ajustar um modelo de regressão linear simples com um intercepto e um coeficiente angular β , utilizando uma amostra de tamanho igual a 402.

Nessa situação, se a razão t correspondente à estimativa de β a ser obtida pelo método de mínimos quadrados ordinários for igual a 20, então o coeficiente de explicação (ou determinação) R^2 proporcionado pelo modelo em tela será igual a

- Ⓐ 0,44.
- Ⓑ 0,50.
- Ⓒ 0,95.
- Ⓓ 0,99.

Questão 20

Em um modelo de regressão linear simples na forma $y = ax + b + \epsilon$, x representa a variável regressora, y denota a variável resposta e ϵ é um erro aleatório com média zero e variância 100.

Nessa hipótese, considerando-se que \hat{a} denote o estimador de mínimos quadrados ordinários do coeficiente a produzido por uma amostra aleatória de tamanho igual a 101 e que o desvio padrão amostral da variável regressora seja igual a 2, é correto afirmar que o desvio padrão de \hat{a} será igual a

- A 0,25.
- B 0,5.
- C 1.
- D 4.

Questão 21

Um modelo de regressão linear múltipla com dez coeficientes foi ajustado pelo método de mínimos quadrados ordinários, tendo produzido um coeficiente de determinação (R^2) igual a 80%.

Nessa hipótese, caso o tamanho da amostra utilizado para esse ajuste tenha sido igual a 46, então o valor correspondente do coeficiente conhecido como “ R^2 ajustado” deve ter sido igual a

- A 0,68.
- B 0,72.
- C 0,75.
- D 0,80.

Questão 22

Supondo-se que a variável aleatória X possa assumir valores 0, 1, 2 ou 3 conforme a função de distribuição de probabilidade $P(X = h) = 6 \times \frac{0,2^h}{h!} \times \frac{0,8^{3-h}}{(3-h)!}$, na qual $h \in \{0, 1, 2, 3\}$, é correto afirmar que o valor esperado de X seja igual a

- A 2,4.
- B 0,6.
- C 1,2.
- D 1,5.

Questão 23

Se N for uma variável aleatória que siga uma distribuição normal com média igual a 10 e desvio padrão igual a 5 e se $Z = \frac{N-10}{5}$, então a probabilidade de ocorrência do evento “ $Z = 1,96$ ” será igual a

- A 0.
- B 0,025.
- C 0,5.
- D 0,975.

Questão 24

O conjunto de dados $\{0, 4, 3, 3, 0\}$ é uma realização de uma amostra aleatória simples retirada de uma população binomial com parâmetros n e p , sendo $n=4$ e p uma probabilidade desconhecida.

Com base nessas informações, é correto afirmar que a estimativa de máxima verossimilhança para a probabilidade de ocorrência do valor 2 na população em questão é igual a

- A 0.
- B 0,0625.
- C 0,20.
- D 0,375.

Questão 25

10 ± 4 representa a estimativa intervalar de 95% de confiança para a média de uma população normal, tendo sido obtida a partir de uma amostra aleatória de tamanho n . Para a obtenção dessa estimativa, considerou-se que a variância populacional fosse conhecida. Em novo levantamento feito sobre essa mesma população, mas, dessa vez, tendo-se quadruplicado o tamanho da amostra ($4n$), foi obtida média amostral igual a 8.

Nesse caso, se $8 \pm \epsilon$ representar a nova estimativa intervalar de 95% de confiança para a média dessa população, é correto afirmar que ϵ deverá ser igual a

- A 0,5.
- B 1.
- C 2.
- D 2,5.

Questão 26

O seguinte código Python utiliza o algoritmo KNN (*k-nearest neighbors*) para classificação, em que o parâmetro `n_neighbors` define o número de vizinhos que o classificador KNN irá considerar para realizar a previsão.

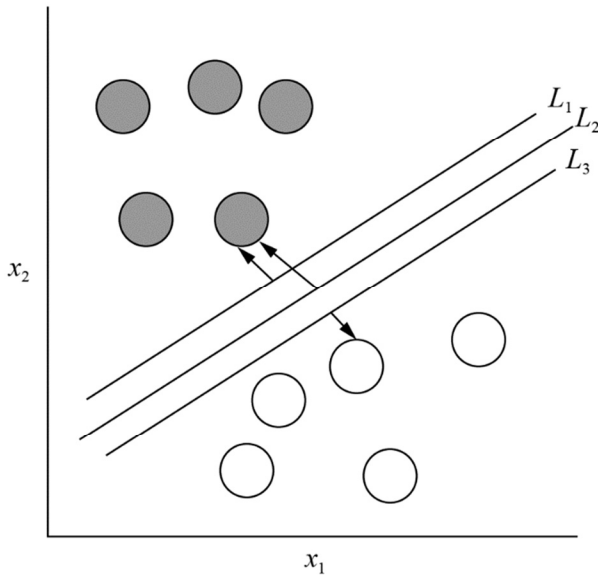
```
from sklearn.neighbors import KNeighborsClassifier
X = [[0], [1], [2], [3]]
y = [0, 0, 1, 1]
neigh = KNeighborsClassifier(n_neighbors=3)
neigh.fit(X, y)
print(neigh.predict([[1]]))
print(neigh.predict_proba([[0.9]]))
```

Com base no código precedente, é correto afirmar que, caso o valor de `n_neighbors` fosse alterado de 3 para 4, o modelo

- A não seria capaz de classificar novos pontos de dados e retornaria `ValueError: Expected n_neighbors <= n_samples_fit`.
- B poderia resultar em empates mais frequentes, levando a previsões que dependam mais do critério de desempate do algoritmo.
- C consideraria todos os pontos de treinamento para qualquer nova previsão, o que levaria a resultados mais precisos.
- D poderia fornecer previsões mais sensíveis a mudanças pequenas no dado, com menos probabilidade de empate.

Espaço livre

Questão 27



Considerando a figura precedente, assinale a opção correta em relação ao algoritmo de SVM (*support vector machine*).

- Ⓐ Uma margem mais ampla geralmente indica um melhor desempenho na classificação, pois sugere que o modelo tem uma separação mais clara entre as classes.
- Ⓑ O hiperplano ilustrado na figura é caracterizado por um plano, pois estão sendo considerados dois recursos de entrada x_1 e x_2 .
- Ⓒ Uma escolha razoável como melhor hiperplano é aquele que representa a menor separação ou margem entre as duas classes.
- Ⓓ O hiperplano de margem suave separa perfeitamente os pontos de diferentes classes, sem erros de classificação e com uma margem fixa.

Questão 28

Em aprendizado de máquina, especialmente em algoritmos de árvores de decisão, é fundamental avaliar como os dados são organizados e classificados em diferentes níveis da árvore. Três conceitos-chave que auxiliam na construção e otimização dessas árvores são o *gini impurity*, a *entropy* e o *information gain*. A respeito desses conceitos, julgue os itens a seguir.

- I *Gini impurity* mede a redução da *entropy* após a divisão de um conjunto de dados com base em um atributo.
- II *Entropy* mede a quantidade de incerteza ou impureza no conjunto de dados.
- III *Information gain* mede a probabilidade de uma nova instância ser classificada incorretamente, com base na distribuição de classes no conjunto de dados.

Assinale a opção correta.

- Ⓐ Apenas o item II está certo.
- Ⓑ Apenas o item III está certo.
- Ⓒ Apenas os itens I e II estão certos.
- Ⓓ Apenas os itens I e III estão certos.

Questão 29

A respeito dos diferentes tipos de algoritmos *naive* Bayes, assinale a opção correta.

- Ⓐ No *multinomial naive* Bayes a média e a variância dos atributos são estimadas para cada classe a partir dos dados de treinamento.
- Ⓑ O *bernoulli naive* Bayes representa as frequências com as quais certos eventos foram gerados por uma distribuição normal.
- Ⓒ Os *feature vectors* do *Bernoulli naive* Bayes são geralmente contagens de eventos ou frequências.
- Ⓓ No *gaussian naive* Bayes, assume-se que os valores contínuos associados a cada recurso são distribuídos de acordo com uma distribuição normal.

Questão 30

Assinale a opção em que é apresentado exemplo no qual o uso prático do *gaussian naive* Bayes é mais apropriado.

- Ⓐ classificação de flores com base em características como comprimento e largura das pétalas, em que os atributos são contínuos
- Ⓑ detecção de fraudes em transações bancárias, utilizando-se apenas indicadores binários de eventos suspeitos
- Ⓒ classificação de documentos com base na frequência de palavras em textos
- Ⓓ análise de sentimentos em redes sociais, onde os dados de entrada são contagens de palavras positivas e negativas

Questão 31

Em relação aos conceitos do algoritmo *k-means*, julgue os itens a seguir.

- I É importante continuar as iterações do algoritmo *k-means* até que a mudança na posição dos centroides entre as iterações seja menor que um limite predefinido.
- II No coeficiente de silhueta, quanto mais próximo o coeficiente estiver de 1, menor a distância entre os *clusters*; 0 indica que os dados podem estar no *cluster* errado; valores negativos sugerem que o ponto está na borda.
- III Apesar de um maior número *clusters* sempre reduzir o SSE (*sum of squared errors*), isso não significa que mais *clusters* sempre sejam melhores, pois um número muito grande de *clusters* pode levar a *overfitting* do modelo.

Assinale a opção correta.

- Ⓐ Apenas o item I está certo.
- Ⓑ Apenas o item II está certo.
- Ⓒ Apenas os itens I e III estão certos.
- Ⓓ Apenas os itens II e III estão certos.

Questão 32

A respeito da linguagem R, é correto afirmar que

- Ⓐ a função `lm()` em R é usada para realizar regressão linear.
- Ⓑ o operador `==` é utilizado para atribuir valores a variáveis.
- Ⓒ a função `plot()` cria tabelas de frequência.
- Ⓓ o operador `%>%` é utilizado para definir novas variáveis em R.

Questão 33

```
import re
text = "Inteligencia artificial generativa&#10;Promoc"
clean_text = re.sub(r'\d', ' ', text)
print(clean_text)
```

Considerando o código precedente, escrito em Python, assinale a opção que corresponde à sua correta execução.

- Ⓐ `Inteligencia artificial generativa&# ;Promoc`
- Ⓑ `Inteligencia artificial generativaPromoc`
- Ⓒ `Inteligencia artificial generativa&#;Promoc`
- Ⓓ `Inteligencia artificial generativa Promoc`

Questão 34

```
import nltk
from nltk.stem import PorterStemmer, WordNetLemmatizer
nltk.download("punkt")
nltk.download('wordnet')

lemmatizer = WordNetLemmatizer()
ps = PorterStemmer()
example_words = ["program", "programming", "programer", "programs", "programmed"]

print("".format("--Word--", "--Stem--", "--Lemma"))

for word in example_words:
    print ("{0:20}{1:20}{2:20}".format(word, ps.stem(word), lemmatizer.lemmatize(word)))
```

Assinale a opção que corresponde à correta execução do código precedente, escrito em Python.

- A** --Word-- --Stem-- --Lemma
 program program program
 programming programming program
 programer programer program
 programs program program
 programmed programmed program
- B** --Word-- --Stem-- --Lemma
 program program program
 programming program programming
 programer program programer
 programs program program
 programmed program programmed
- C** program program program
 programming program programming
 programer program programer
 programs program program
 programmed program programmed
- D** --Word-- --Stem-- --Lemma
 program program program
 --Word-- --Stem-- --Lemma
 programming program programming
 --Word-- --Stem-- --Lemma
 programer program programer
 --Word-- --Stem-- --Lemma
 programs program program
 --Word-- --Stem-- --Lemma
 programmed program programmed

Questão 35

Assinale a opção correta a respeito da implantação de um modelo de classificação de árvore de decisão em Python.

- A** Para treinar um modelo, deve-se utilizar o método `fit(X_train, y_train)`, em que a variável `X_train` representa os rótulos verdadeiros, e `y_train` as características dos dados.
- B** Após treinar o modelo por meio do comando `fit()`, o método `score(X_test, y_test)` deve ser utilizado para avaliar o desempenho do modelo com novos dados de teste.
- C** O método `model.predict(X_train)` é utilizado para avaliar o desempenho do modelo de árvore de decisão aos dados de treinamento.
- D** Após treinar o modelo por meio do comando `fit()`, deve-se utilizar `model.predict(Y_test)` para realizar previsões em novos dados de teste.

Questão 36

Caso um cientista de dados precise extrair, transformar e analisar grandes volumes de dados em um banco de dados relacional, usando SQL, a prática mais recomendável para garantir que as operações de análise sejam eficientes e precisas consiste em

- A** aplicar funções de agregação como `SUM()`, `AVG()` e `COUNT()` com cláusulas `GROUP BY` para resumir e agregar os dados de forma eficiente.
- B** ignorar o uso de índices ao executar consultas, já que a criação e a manutenção de índices são responsabilidades do administrador do banco de dados.
- C** usar comandos `SELECT *` em todas as consultas para garantir que todas as colunas sejam retornadas, facilitando a análise posterior.
- D** utilizar subconsultas (*subqueries*) complexas em vez de operações de junção (`JOIN`) para simplificar a lógica das consultas.

Texto 14A3

Em certa base de dados de *e-commerce*, as tabelas pedidos e clientes possuem as seguintes estruturas:

```
pedidos(pedido_id, cliente_id, data_pedido, e
valor_pedido)
clientes(cliente_id, nome_cliente, email,
cidade)
```

Questão 37

Um cientista de dados precisa analisar o comportamento de compra dos clientes na base de dados descrita no texto 14A3. O objetivo dessa análise é calcular o valor total gasto por cliente em pedidos feitos no mês anterior ao atual.

Com base nessa situação hipotética, assinale a opção em que consta o comando que executará corretamente o cálculo requerido.

- A** `SELECT nome, valor_total FROM clientes WHERE data_pedido > CURRENT_DATE - INTERVAL '1 month';`
- B** `SELECT nome, SUM(valor_total) AS total_gasto FROM clientes LEFT JOIN pedidos ON clientes.cliente_id = pedidos.cliente_id GROUP BY nome;`
- C** `SELECT nome, SUM(valor_pedido) AS total_pedido FROM clientes INNER JOIN pedidos ON clientes.cliente_id = pedidos.cliente_id WHERE data_pedido > CURRENT_DATE - INTERVAL '1 month' GROUP BY nome;`
- D** `SELECT nome, valor_pedido FROM clientes INNER JOIN pedidos ON clientes.cliente_id = pedidos.cliente_id WHERE data_pedido > CURRENT_DATE - INTERVAL '1 month';`

Questão 38

Assinale a opção em que é indicada a consulta SQL que, na base de dados descrita no texto 14A3, retorna o nome dos clientes e o valor total de seus pedidos, apenas para aqueles que tenham realizado mais de 5 pedidos no ano de 2023.

- A** `SELECT c.nome_cliente, SUM(v.valor_pedido) AS total_pedidos
FROM vendas v
RIGHT JOIN clientes c ON v.id_cliente = c.id_cliente
WHERE v.data_pedido BETWEEN '2023-01-01' AND '2023-12-31'
GROUP BY c.nome_cliente
HAVING COUNT(v.id_pedido) >= 5;`
- B** `SELECT c.nome_cliente, SUM(v.valor_pedido) AS total_pedidos
FROM vendas v
INNER JOIN pedidos c ON v.id_cliente = c.id_cliente
WHERE YEAR(v.data_pedido) = 2023
GROUP BY c.nome_cliente
HAVING SUM(v.id_pedido) > 5;`
- C** `SELECT c.nome_cliente,
SUM(v.valor_pedido) AS total_pedidos
FROM vendas v
LEFT JOIN clientes c ON v.id_cliente = c.id_cliente
WHERE YEAR(v.data_pedido) = 2023
GROUP BY c.nome_cliente
HAVING SUM(v.valor_pedido) > 5;`
- D** `SELECT c.nome_cliente,
SUM(v.valor_pedido) AS total_pedidos
FROM vendas v
INNER JOIN clientes c ON v.id_cliente = c.id_cliente
WHERE v.data_pedido BETWEEN '2023-01-01' AND '2023-12-31'
GROUP BY c.nome_cliente
HAVING COUNT(v.id_pedido) > 5;`

Questão 39

Um analista está criando um *dashboard* no Power BI para visualizar as vendas mensais de uma empresa e necessita criar uma medida que calcule a média móvel de 3 meses de vendas.

Nessa situação, a fórmula DAX mais adequada para a tarefa mencionada é

- A** `MédiaMóvel3Meses =
CALCULATE(
AVERAGE('Vendas'[ValorVenda]),
DATESINPERIOD('Vendas'[DataVenda],
LASTDATE('Vendas'[DataVenda]), -3, MONTH)
)`
- B** `MédiaMóvel3Meses =
CALCULATE(
AVERAGE('Vendas'[ValorVenda]),
DATESINPERIOD('Vendas'[DataVenda],
LASTDATE('Vendas'[DataVenda]), -3, MONTH)
) / 3`
- C** `MédiaMóvel3Meses =
CALCULATE(
SUM('Vendas'[ValorVenda]),
DATESINPERIOD('Vendas'[DataVenda],
LASTDATE('Vendas'[DataVenda]), -3, MONTH)
) / 3`
- D** `MédiaMóvel3Meses =
CALCULATE(
AVERAGE('Vendas'[ValorVenda]),
DATESINPERIOD('Vendas'[DataVenda],
LASTDATE('Vendas'[DataVenda]), -2, MONTH)
)`

Questão 40

A respeito das semelhanças e diferenças entre o Power BI e o Qlik Sense na construção de *dashboards*, assinale a opção correta.

- A** O Qlik Sense permite a criação de relatórios paginados nativamente, enquanto o Power BI requer o uso de uma ferramenta adicional para essa funcionalidade.
- B** O Power BI permite a criação de *dashboards* interativos, enquanto o Qlik Sense não oferece essa funcionalidade.
- C** O Power BI possui uma integração mais robusta com o Microsoft Excel, enquanto o Qlik Sense é mais conhecido por sua capacidade de associar dados de diferentes fontes de forma automática.
- D** O Qlik Sense utiliza a mesma linguagem que o Power BI para a criação de medidas e cálculos.

Questão 41

Assinale a opção correta no que diz respeito à personalização de visualizações no Power BI e no Qlik Sense.

- A** No Power BI, é possível importar visualizações personalizadas da AppSource, enquanto no Qlik Sense as visualizações personalizadas devem ser criadas do zero.
- B** O Qlik Sense oferece suporte nativo para visualizações em Python, enquanto o Power BI não possui essa funcionalidade.
- C** No Power BI, é possível criar visualizações personalizadas com a utilização da linguagem R, enquanto no Qlik Sense não há suporte para integração com R.
- D** O Qlik Sense permite a criação de visualizações personalizadas com a utilização do JavaScript, enquanto o Power BI não oferece suporte para visualizações personalizadas com JavaScript.

Questão 42

Um cientista de dados é responsável por criar *dashboards* interativos para uma empresa que pretende monitorar suas vendas e seu desempenho financeiro. A empresa utiliza tanto o Power BI quanto o Qlik Sense para diferentes departamentos. O cientista de dados precisa criar um *dashboard* que permita aos usuários filtrar dados por região, produto e período de tempo, além de incluir gráficos de linha, barras e mapas interativos.

Considerando a situação hipotética apresentada, assinale a opção correta em relação às capacidades do Power BI e do Qlik Sense para atender aos requisitos mencionados.

- A** O Power BI oferece suporte nativo para gráficos de linha e barras, mas mapas interativos só podem ser criados no Qlik Sense.
- B** O Power BI permite a criação de gráficos de linha, barras e mapas interativos, mas não oferece suporte para filtros dinâmicos por região, produto e período de tempo.
- C** Tanto o Power BI quanto o Qlik Sense permitem a criação de gráficos de linha, barras e mapas interativos, além de oferecerem suporte para filtros dinâmicos por região, produto e período de tempo.
- D** O Qlik Sense permite a criação de gráficos de linha, barras e mapas interativos, mas não oferece suporte para filtros dinâmicos por região, produto e período de tempo.

Questão 43

Um analista foi encarregado de criar um *dashboard* que mostre a evolução das vendas trimestrais de uma empresa que utiliza tanto Power BI quanto Qlik Sense e armazena seus dados em um banco de dados SQL. O analista precisa extrair os dados trimestrais de 2023 e criar um *dashboard* interativo que permita aos usuários filtrar por categoria de produto e região.

Nessa situação hipotética, o referido analista deverá escrever uma consulta SQL para extrair os dados trimestrais de 2023 e criar um *dashboard* no

- A Qlik Sense, utilizando a função `DATESQTD` para calcular os dados trimestrais e aplicar filtros dinâmicos por categoria de produto e região.
- B Power BI, utilizando a função `DATESQTD` para calcular os dados trimestrais e aplicar filtros dinâmicos por categoria de produto e região.
- C Power BI, utilizando a função `QUARTER` para calcular os dados trimestrais e aplicar filtros dinâmicos por categoria de produto e região.
- D Qlik Sense, utilizando as funções `DATESQTD` ou `QUARTER` para calcular os dados trimestrais e aplicar filtros dinâmicos por categoria de produto e região.

Questão 44

Um profissional de dados está analisando um conjunto de dados com informações sobre as vendas de produtos em diferentes regiões ao longo do tempo. Seu objetivo é visualizar a tendência das vendas para cada região de forma clara e comparativa. Para isso, ele decidiu utilizar o `ggplot2` em R.

Com base nessa situação hipotética, assinale a opção que apresenta o código que, se executado, gerará um gráfico de linhas que mostra a tendência de vendas ao longo do tempo para cada região, utilizando cores diferentes para cada região.

- A `ggplot(data = vendas, aes(x = tempo, y = vendas, fill = regioao)) + geom_area()`
- B `ggplot(data = vendas, aes(x = tempo, y = vendas, group = regioao)) + geom_line()`
- C `ggplot(data = vendas, aes(x = tempo, y = vendas)) + geom_point(aes(color = regioao))`
- D `ggplot(data = vendas, aes(x = tempo, y = vendas, color = regioao)) + geom_line()`

Questão 45

Assinale a opção em que é apresentado o código escrito em Python ou em R que, se executado, gerará um *scatter plot* para a visualização da relação entre quantidade de vendas e receita gerada por determinado produto, destacando a tendência linear entre essas variáveis.

- A

```
import matplotlib.pyplot as plt

plt.scatter(vendas['quantidade'],
vendas['receita'])
plt.plot(vendas['quantidade'],
vendas['receita'])
plt.show()
```
- B

```
import matplotlib.pyplot as plt
import numpy as np

plt.scatter(vendas['quantidade'],
vendas['receita'])
m,b=np.polyfit(vendas['quantidade'],vendas['re
ceita'],1)
plt.plot(vendas['quantidade'], m*vendas['quant
idade'] + b)
plt.show()
```
- C

```
ggplot(data = vendas, aes(x = quantidade,
y = receita)) + geom_point() +
geom_smooth(method = "lm")
```
- D

```
ggplot(data = vendas, aes(x = quantidade,
y = receita)) + geom_point() + geom_line()
```

Questão 46

Um cientista de dados foi encarregado de criar um *dashboard* para a equipe de vendas da empresa cujo objetivo é monitorar em tempo real as métricas de desempenho, como volume de vendas, receita e número de clientes novos por região. Os dados estão armazenados em um banco de dados SQL, e esse profissional deverá integrar essas consultas SQL no processo de criação do *dashboard*.

Considerando a situação hipotética apresentada, assinale a opção em que é descrita a abordagem mais eficiente para garantir que o *dashboard* seja atualizado em tempo real e que as consultas SQL sejam otimizadas para melhor *performance*.

- A carregar para a memória todos os dados brutos do banco de dados SQL no Qlik Sense e, depois, utilizar as ferramentas de transformação de dados dentro do Qlik Sense para calcular as métricas no momento da visualização
- B criar no Power BI um *script* SQL que agregue todos os dados necessários em uma única consulta e utilizar o Power BI para atualizar os dados manualmente se necessário
- C utilizar várias consultas SQL no Power BI, separadas para cada métrica individual, e configurar o Power BI para fazer a junção desses dados na camada de visualização
- D utilizar uma conexão direta ao banco de dados SQL no Qlik Sense, em que as consultas sejam executadas diretamente na fonte de dados, e configurar o *dashboard* para atualizações em intervalos definidos

Questão 47

Um profissional de dados foi encarregado de criar um *dashboard* em Qlik Sense que deverá exibir métricas de vendas por região e por produto em tempo real. Os dados estão armazenados em um banco de dados relacional SQL. Para garantir a melhor *performance* e a integridade das informações, esse profissional deve considerar como o Qlik Sense se conecta ao banco de dados e como as consultas SQL são estruturadas.

Tendo como referência a situação hipotética apresentada, assinale a opção em que é descrita a abordagem mais eficiente e correta para a consulta dos dados na construção desse *dashboard*.

- A configurar uma conexão direta ao banco de dados SQL e utilizar a linguagem SQL para realizar todas as agregações e cálculos na origem, carregando-se apenas os resultados finais para o Qlik Sense
- B carregar todos os dados brutos diretamente do banco de dados SQL para o Qlik Sense sem qualquer pré-processamento e utilizar *scripts* de transformação dentro do Qlik Sense para realizar todas as operações de agregação e cálculo
- C escrever uma consulta SQL complexa para carregar todos os dados necessários no Qlik Sense de uma só vez, e realizar todas as agregações e cálculos dentro do Qlik Sense com a utilização de expressões Qlik
- D exportar os dados necessários do banco de dados SQL para arquivos CSV e carregá-los no Qlik Sense, realizando-se as transformações e cálculos diretamente no Qlik Sense

Questão 48

Assinale a opção que apresenta corretamente a consulta SQL que retornará a média de salários por departamento de uma empresa, considerada uma tabela `funcionarios` com colunas `departamento` e `salario`.

- A** `SELECT departamento, COUNT(salario)`
`FROM funcionarios`
`GROUP BY departamento;`
- B** `SELECT departamento, AVG(salario)`
`FROM funcionarios`
`GROUP BY departamento;`
- C** `SELECT departamento, SUM(salario)`
`FROM funcionarios`
`GROUP BY departamento;`
- D** `SELECT departamento, MAX(salario)`
`FROM funcionarios`
`GROUP BY departamento;`

Questão 49

Um analista está trabalhando em um projeto que envolve a análise de dados sobre micro e pequenas empresas no Brasil. Seu objetivo é construir um *dashboard* no Power BI que mostre a evolução dos negócios em diferentes setores, com métricas como faturamento, número de empregados, e taxa de sobrevivência das empresas ao longo dos anos. Os dados estão armazenados em um banco de dados relacional SQL.

Com base nessa situação hipotética, assinale a opção em que é descrita a abordagem mais eficiente para configurar o Power BI, com a utilização de SQL, de forma que o *dashboard* seja otimizado para desempenho e atualizado regularmente com novos dados.

- A** escrever uma consulta SQL complexa que una todas as tabelas necessárias e calcule as métricas diretamente na consulta, carregando os dados agregados para o Power BI
- B** exportar os dados do banco de dados SQL para um arquivo `.csv`, realizar os cálculos em Excel e carregar os resultados no Power BI
- C** criar várias consultas SQL pequenas e independentes, cada uma carregando diferentes partes dos dados para o Power BI, e juntar esses dados na camada de visualização do Power BI
- D** carregar todas as tabelas do banco de dados para o Power BI separadamente e utilizar o Power Query para realizar as junções e cálculos de métricas dentro do ambiente do Power BI

Questão 50

Um analista está trabalhando em um projeto sobre o desempenho de micro e pequenas empresas no Brasil. Os dados utilizados no projeto incluem informações como faturamento anual, número de empregados e distribuição geográfica por estado. O objetivo desse projeto é criar visualizações que facilitem a análise dessas informações e identifiquem tendências regionais e setoriais.

Considerando a situação hipotética apresentada e o volume e a complexidade dos dados utilizados, assinale a opção em que é descrita a abordagem mais apropriada para a criação de visualizações eficientes e informativas, utilizando-se `ggplot` em R ou `matplotlib` em Python.

- A** utilizar `ggplot` ou `matplotlib` para criar gráfico de dispersão (*scatter plot*) que mostre a relação entre faturamento e número de empregados para todas as empresas em um único gráfico
- B** gerar gráficos de barras empilhadas (*stacked bar charts*) com `ggplot` ou `matplotlib` para mostrar a distribuição de empresas por estado, utilizando-se dados brutos diretamente carregados do banco de dados para R ou Python
- C** criar gráficos de linha (*line charts*) com `ggplot` ou `matplotlib` para visualizar a evolução temporal do faturamento médio por setor, com dados já agregados e filtrados por setor e ano no banco de dados antes de carregá-los para R ou Python
- D** criar gráficos *pizza* (*pie charts*) com `ggplot` ou `matplotlib` para representar a proporção de empresas por setor, após carregar todos os dados brutos do banco de dados para R ou Python

Espaço livre