



DOMINGO DE MANHÃ

**AGÊNCIA ESTADUAL DE REGULAÇÃO DOS SERVIÇOS PÚBLICOS
DELEGADOS DO RIO GRANDE DO SUL – AGERGS
CONCURSO PÚBLICO Nº 01/2022**

TÉCNICO SUPERIOR ENGENHEIRO DE DADOS

INSTRUÇÕES

Leia atentamente e cumpra rigorosamente as instruções que seguem, pois elas são parte integrante das provas e das normas que regem esse certame.

1. Atente-se aos avisos contidos no quadro da sala.
2. Seus pertences deverão estar armazenados dentro de embalagem específica fornecida pelo fiscal, permanecendo em sua posse somente caneta esferográfica de ponta grossa, de material transparente, com tinta preta, documento de identificação, lanche e água, se houver. A utilização de qualquer material não permitido em edital é expressamente proibida, acarretando a sua imediata eliminação do certame.
3. Certifique-se de que este caderno:
 - contém 70 (setenta) questões;
 - refere-se ao cargo para o qual realizou a inscrição.
4. Cada questão oferece 5 (cinco) alternativas de respostas, representadas pelas letras A, B, C, D e E, sendo apenas 1 (uma) a resposta correta.
5. Será respeitado o tempo para realização da prova conforme previsto em edital, incluindo o preenchimento da grade de respostas.
6. Os três últimos candidatos deverão retirar-se da sala de prova ao mesmo tempo, devendo assinar a Ata de Prova.
7. A responsabilidade referente à interpretação dos conteúdos das questões é exclusiva do candidato.
8. No caderno de prova, você poderá rabiscar, riscar e calcular.
9. Os gabaritos preliminares da prova objetiva serão divulgados na data descrita no Cronograma de Execução desse certame.



V1_27/06/2022 13:06:06



Instrução: As questões de números 01 a 20 referem-se ao texto abaixo.

Dialogar ou cancelar, eis a questão

Por Nilson Souza

01 Já não sei mais se a frase é de Confúcio ou de Bernard Shaw, mas acredito que qualquer
02 um deles – gênios que foram – poderia ser o pai da criança. Se você tem uma maçã e eu tenho
03 outra, e nós trocamos as maçãs, cada um fica com uma. Mas se você tem uma ideia e eu tenho
04 outra, e nós as trocamos, então cada um terá duas ideias.

05 Adoro essa mensagem e me esforço para colocá-la em prática com amigos e amigas que
06 eventualmente encontro fora da clau__ura tecnológica. Quando o celular está presente, fica mais
07 difícil, pois o brinquedinho, como bem sabem os usuários, aproxima quem está longe e afasta
08 quem está próximo. Além disso, o ambiente das redes sociais está se transformando numa arena
09 repleta de gladiadores e leões furiosos espera de vítimas incautas.

10 Claro, sempre é possível dar um drible no Grande Irmão e tentar uma conversa à moda
11 antiga, totalmente analógica, com a__ionamento de cordas vocais e tímpanos. Costumo fazer
12 isso nas minhas caminhadas matinais, na companhia de um velho parceiro de exercício e ofício.
13 Trocamos ideias como as crianças que fomos trocavam gibis na porta do cinema – a maioria,
14 histórias já lidas e conhecidas, mas sempre divertidas. Não é raro, porém, que voltemos aos
15 nossos afazeres diários com alguma nova reflexão.

16 Nem sempre concordamos um com o outro. Mas é exatamente quando temos visões
17 conflitantes sobre algum tema que tiramos maior proveito do debate. Tudo porque nossa
18 amizade, longa e sólida, nos torna impermeáveis cultura do cancelamento – hoje uma das
19 maiores ameaças à liberdade de expressão. Sim, pois o verbo cancelar já não se resume
20 tornar algo sem efeito ou a excluir das nossas relações aqueles que pensam de forma diferente
21 da nossa. No tribunal sumário do mundo digital, qualquer divergência de opinião tende a derivar
22 para banimentos, boicotes, intimidações e, não raro, linchamentos públicos.

23 Nesse contexto, o ódio flore__e, a autocensura cresce e o debate público empobrece. Mas
24 sempre pode haver espaço para o diálogo, para o debate construtivo e para a convivência de
25 contrários. Precisamos conversar sobre isso, antes que o fungo da intolerância contamine
26 irremediavelmente as nossas maçãs.

(Disponível em: Zero Hora, Porto Alegre, ano 58, n. 20.292, 29 mar. 2022. p. 3 – texto adaptado especialmente para esta prova).

QUESTÃO 01 – Assinale a alternativa cuja informação encontra respaldo no texto.

- A) É necessário dialogar sobre o uso incorreto do celular.
- B) Ao invés de aproximar as pessoas, o celular as afasta.
- C) Não há mais a possibilidade de um diálogo saudável entre os opostos.
- D) A intolerância não tem solução, pois ela já está enraizada na sociedade.
- E) A liberdade de expressão é ameaçada pela cultura do cancelamento.

QUESTÃO 02 – De acordo com a leitura do texto, é correto afirmar que a cultura do cancelamento:

- A) Utiliza um novo sentido para o verbo “cancelar”.
- B) Favorece a liberdade de expormos a nossa opinião na internet.
- C) Aproxima quem está perto e afasta quem está longe.
- D) Ocorre fora do mundo digital, quando há divergências entre as pessoas.
- E) Resume-se apenas como a exclusão de um indivíduo de nosso convívio social.

QUESTÃO 03 – Considerando o emprego do acento indicativo de crase, assinale a alternativa que preenche, correta e respectivamente, as lacunas pontilhadas das linhas 09, 18 e 19.

- A) a – à – a
- B) à – a – à
- C) a – a – à
- D) à – à – a
- E) à – à – à

QUESTÃO 04 – Assinale a alternativa que preenche, correta e respectivamente, as lacunas tracejadas das palavras nas linhas 06, 11 e 23.

- A) s – c – c
- B) ç – sc – ss
- C) s – c – sc
- D) z – ss – c
- E) z – c – sc

QUESTÃO 05 – Os vocábulos “incautas” (l. 09) e “afazeres” (l. 15) podem ser substituídos, respectivamente, sem alteração de sentido ou de outras palavras nos trechos nos quais se encontram, por:

- A) descuidadas – encargos.
- B) enganadas – obrigações.
- C) desprevenidas – tarefas.
- D) enganadas – encargos.
- E) descuidadas – obrigações.

QUESTÃO 06 – Os vocábulos “pois” (l. 07) e “mas” (l. 14) podem ser substituídos, respectivamente, sem alteração de sentido ou de outras palavras nas frases, por:

- A) afinal – logo.
- B) porque – entretanto.
- C) uma vez que – portanto.
- D) já que – assim.
- E) contudo – porque.

QUESTÃO 07 – Observe o trecho a seguir retirado do texto: “Quando o celular está presente, fica mais difícil, pois o brinquedinho, como bem sabem os usuários, aproxima quem está longe e afasta quem está próximo”.

Se as palavras “celular” e “brinquedinho” fossem flexionadas no plural, quantas outras palavras precisariam ser modificadas para garantir a correta concordância verbo-nominal?

- A) 3.
- B) 4.
- C) 5.
- D) 6.
- E) 7.

QUESTÃO 08 – Os vocábulos “la” (l. 05) e “brinquedinho” (l. 07) referem-se, respectivamente, às palavras:

- A) mensagem – ambiente.
- B) prática – celular.
- C) ideia – ambiente.
- D) prática – difícil.
- E) mensagem – celular.

QUESTÃO 09 – Tendo em vista a oração “o ódio flore__e” (l. 23), a qual teve a lacuna tracejada mantida, assinale a alternativa em que apenas advérbios foram acrescentados à frase.

- A) Hoje, o terrível ódio flore__e.
- B) O teu ódio flore__e continuamente.
- C) Aqui o enorme ódio flore__e.
- D) O atual ódio flore__e.
- E) Hodiernamente, o ódio não flore__e.

QUESTÃO 10 – Analise as assertivas abaixo e assinale V, se verdadeiras, ou F, se falsas, relacionadas ao trecho “Mas sempre pode haver espaço para o diálogo, para o debate construtivo e para a convivência de contrários”.

- () O vocábulo “sempre” é um advérbio.
- () O termo “para” é uma preposição formada pelo processo de combinação.
- () O trecho possui uma locução verbal.
- () “Construtivo” é um substantivo simples.

A ordem correta de preenchimento dos parênteses, de cima para baixo, é:

- A) F – F – F – F.
- B) V – F – V – F.
- C) F – V – V – V.
- D) V – F – F – V.
- E) F – V – F – V.

QUESTÃO 11 – Assinale a alternativa que apresenta um pronome indefinido.

- A) “Se você tem uma maçã e eu tenho outra (...)” (l. 02–03).
- B) “(...) sempre é possível dar um drible no Grande Irmão (...)” (l. 10).
- C) “Mas sempre pode haver espaço para o diálogo (...)” (l. 23–24).
- D) “Costumo fazer isso nas minhas caminhadas matinais (...)” (l. 11–12).
- E) “Já não sei mais se a frase é de Confúcio ou de Bernard Shaw (...)” (l. 01).

QUESTÃO 12 – Em relação às classes gramaticais, assinale a alternativa correta.

- A) Em “(...) E nós as trocamos (...)” (l. 04), o termo em destaque é um artigo definido.
- B) Em “(...) E me esforço para colocá-la em prática (...)” (l. 05), o vocábulo em destaque é um substantivo.
- C) Em “costumo fazer isso nas minhas caminhadas matinais (...)” (l. 11–12), a palavra em destaque é um adjetivo.
- D) Em “não é raro, porém, que voltemos aos nossos afazeres diários (...)” (l. 14–15), o termo em destaque é um pronome pessoal oblíquo.
- E) Em “(...) E o debate público empobrece” (l. 23), o vocábulo em destaque é um substantivo.

QUESTÃO 13 – Em relação à oração “Adoro essa mensagem”, retirada do texto, analise as assertivas abaixo e assinale V, se verdadeiras, ou F, se falsas.

- () O tipo de sujeito da oração é indeterminado.
- () O verbo “adoro” é um verbo transitivo indireto.
- () O tipo de predicado é verbal.
- () A expressão “essa mensagem” é complemento verbal.

A ordem correta de preenchimento dos parênteses, de cima para baixo, é:

- A) F – V – F – V.
- B) V – V – F – V.
- C) F – F – V – V.
- D) F – V – F – F.
- E) V – F – V – F.

QUESTÃO 14 – Assinale a alternativa que apresenta um sujeito oculto.

- A) “Se você tem uma maçã e eu tenho outra, (...)” (l. 02–03).
- B) “Costumo fazer isso nas minhas caminhadas matinais, (...)” (l. 11–12).
- C) “qualquer divergência de opinião tende a derivar para banimentos, (...)” (l. 21–22).
- D) “Nesse contexto, o ódio flore___e, (...)” (l. 23).
- E) “(...) o ambiente das redes sociais está se transformando numa arena repleta de gladiadores e leões furiosos (...)” (l. 08–09).

QUESTÃO 15 – Assinale a alternativa que apresenta um verbo intransitivo.

- A) "(...) e nós trocamos as maçãs (...)" (l. 03).
- B) "Costumo fazer isso nas minhas caminhadas matinais (...)" (l. 11-12).
- C) "Trocamos ideias (...)" (l. 13).
- D) "(...) a autocensura cresce (...)" (l. 23).
- E) "(...) antes que o fungo da intolerância contamine irremediavelmente as nossas maçãs" (l. 25-26).

QUESTÃO 16 – O fragmento "Se você tem uma maçã e eu tenho outra, e nós trocamos as maçãs, cada um fica com uma" apresenta:

- A) Uma oração.
- B) Duas orações.
- C) Três orações.
- D) Quatro orações.
- E) Cinco orações.

QUESTÃO 17 – Assinale a alternativa em que o advérbio destacado NÃO está se referindo a um verbo:

- A) "Já não sei mais se a frase é de Confúcio ou de Bernard Shaw (...)" (l. 01).
- B) "(...) e tentar uma conversa à moda antiga, totalmente analógica (...)" (l. 10-11).
- C) "(...) antes que o fungo da intolerância contamine irremediavelmente as nossas maçãs" (l. 25-26).
- D) "Mas é exatamente quando temos visões conflitantes sobre algum tema (...)" (l. 16-17).
- E) "(...) e me esforço para colocá-la em prática com amigos e amigas que eventualmente encontro (...)" (l. 05-06).

QUESTÃO 18 – Analise o seguinte fragmento: "No tribunal sumário do mundo digital, qualquer divergência de opinião tende a derivar para banimentos, boicotes, intimidações e, não raro, linchamentos públicos".

Assinale a alternativa que indica o número correto de preposições presentes no trecho, incluindo as que possam aparecer combinadas ou contraídas com outras palavras.

- A) 2.
- B) 3.
- C) 4.
- D) 5.
- E) 6.

QUESTÃO 19 – Assinale a alternativa que apresenta uma palavra com a mesma regra de sílaba tônica que o termo em destaque no trecho "(...) como as crianças que fomos trocavam gibis na porta do cinema (...)".

- A) Geleia.
- B) Árvore.
- C) Jiboia.
- D) Voo.
- E) Avental.

QUESTÃO 20 – Assinale a alternativa que apresenta uma palavra com grafia correta conforme o Acordo Ortográfico vigente, assim como ocorre no vocábulo destacado no seguinte trecho: "(...) a autocensura cresce (...)".

- A) Autohipnose.
- B) Autoobservação.
- C) Autopeças.
- D) Autoretrato.
- E) Autoserviço.

CONHECIMENTOS ESPECÍFICOS

QUESTÃO 21 – Três estruturas de dados fundamentais em Python são as listas ("list"), tuplas ("tuple") e dicionários ("dict"). A respeito dessas estruturas, é correto afirmar que:

- A) Dicionários não podem ser modificados depois de criados, ao passo que listas e tuplas podem.
- B) Dicionários e tuplas são indexados por inteiros, ao passo que listas podem ser indexadas por "strings".
- C) Listas podem ser modificadas, mas seu tamanho não pode ser modificado após a criação, ao passo que tuplas e dicionários não têm essa limitação.
- D) Tuplas e listas são indexadas por inteiros, ao passo que dicionários podem ser indexados por "strings".
- E) Listas não podem ser modificadas depois de criadas, ao passo que tuplas e dicionários podem.

QUESTÃO 22 – Sobre Linguagem de Programação Python, analise as assertivas abaixo e assinale a alternativa correta.

- I. Uma tupla é uma sequência de valores. Os valores podem ser de qualquer tipo, e podem ser indexados por números inteiros, portanto, nesse sentido, as tuplas são muito parecidas com as listas. A diferença importante é que as tuplas são imutáveis.
 - II. Sintaticamente, uma tupla é uma lista de valores separados por vírgulas:
a) `>>> t = 'a', 'b', 'c', 'd', 'e'`
 - III. Embora não seja sempre necessário, é comum colocar tuplas entre parênteses:
a. `>>> t = ('a', 'b', 'c', 'd', 'e')`
 - IV. Para criar uma tupla com um único elemento, é preciso incluir uma vírgula final:
a. `>>> t1 = 'a',`
b. `>>> type(t1)`
c. `<class 'tuple'>`
- A) Todas estão corretas.
 - B) Todas estão incorretas.
 - C) Apenas I está correta.
 - D) Apenas I e II estão corretas.
 - E) Apenas III e IV estão corretas.

QUESTÃO 23 – No código a seguir, n é um ----- ? n = '5'

- A) int
- B) complex
- C) str
- D) float
- E) list

QUESTÃO 24 – Analise o código Python a seguir:

```
import numpy as np
series = [[23,45,12,679], [14,48,69,38]]
new_series = np.array(series)
print(new_series.ndim)
print(new_series.shape)
```

Qual é a saída correta?

- A) 2 (2, 4)
- B) 2 (4, 2)
- C) 4 (2)
- D) 2 (4)
- E) 4 (2,4)

QUESTÃO 25 – Zip é uma função integrada que recebe duas ou mais sequências e devolve uma lista de tuplas, onde cada tupla contém um elemento de cada sequência. O nome da função tem a ver com o zíper, que se junta e encaixa duas carreiras de dentes. Após a execução do Código, qual é o resultado?

```
>>> s = 'abc'
>>> t = [0, 1, 2]
>>> zip(s, t)
<zip object at 0x7f7d0a9e7c48>

>>> for pair in zip(s, t):
...     print(pair)
...
```

- A) (a,b,c,d)
- B) ('a', 0)
 - a. ('b', 1)
 - b. ('c', 2)
- C) (zip,s,t,object)
- D) (0)
 - a. (1)
 - b. (2)
- E) ('Error for execution')

QUESTÃO 26 – Suponha que o DataFrame df tenha sido criado conforme o código a seguir:

```
df = pd.read_csv("dados.csv")
```

Assinale as instruções que podemos utilizar para retornar as primeiras 5 linhas do DataFrame df.

- A) df.head()
- B) df.head(5)
- C) df.tail()
- D) df.tail(5)
- E) df.sample(5)

QUESTÃO 27 – Estamos interessados em ajustar um modelo de regressão linear simples no pacote estatístico R. O comando a ser utilizado para esse fim é:

- A) ts
- B) glm
- C) lm
- D) cor
- E) road

QUESTÃO 28 – Quando usamos o símbolo "=" ou "<" seguido de "-" estamos criando objetos com um nome que aparece à esquerda e que contém alguns elementos (o que vem à direita do símbolo). Quando criamos um objeto, ele fica guardado na memória do R até que se feche o programa. Porém, os objetos criados ficam ocultos. Para ver a lista de arquivos ocultos, basta dar o seguinte comando:

- A) tuple()
- B) console()
- C) rain()
- D) table_rain()
- E) ls()

QUESTÃO 29 – A técnica que vetoriza as palavras com base na importância que elas compartilham é:

- A) Nenhuma das alternativas.
- B) BERT.
- C) Stemming e Lemmatization.
- D) Hashtagsteam.
- E) Tf-idf.

QUESTÃO 30 – Sobre os comandos R, analise as assertivas abaixo e assinale a alternativa correta.

- I. `license` – detalha as condições de distribuição do R.
- II. `contributors` – lista a equipe do *time-cerne* de desenvolvimento.
- III. `Citation` – ensina como citar o R em trabalhos acadêmicos.
- IV. `Demo` – inicia uma sessão interativa de demonstração do R.

- A) Todas estão corretas.
- B) Todas estão incorretas.
- C) Apenas I está correta.
- D) Apenas I e II estão corretas.
- E) Apenas III e IV estão corretas.

QUESTÃO 31 – Análise descritiva ou exploratória de dados é uma forma de utilizar métricas e técnicas estatísticas simples ou avançadas para entender e explicar como os dados são. Sobre os tipos de dados, analise as assertivas abaixo e assinale a alternativa correta.

- I. Univariada.
- II. Bivariada.
- III. Multivariada.

- A) Todas estão corretas.
- B) Todas estão incorretas.
- C) Apenas I está correta.
- D) Apenas I e II estão corretas.
- E) Apenas II e III estão corretas.

QUESTÃO 32 – Em uma pesquisa estatística, foi construída uma tabela com o perfil dos pesquisados a partir das seguintes variáveis: Sexo; raça; cor dos olhos; cor do cabelo; altura; idade (anos); peso; estado civil; salário mensal (R\$); número de dependentes. Considerando as variáveis apresentadas, assinale a alternativa que apresenta apenas as variáveis qualitativas.

- A) Altura; peso; salário mensal (em reais); número de dependentes.
- B) Altura; idade (anos); peso; salário mensal (em reais); número de dependentes.
- C) Cor dos olhos; número de dependentes; sexo; raça; cor do cabelo; estado civil.
- D) Sexo; raça; cor dos olhos; cor do cabelo; estado civil.
- E) Altura; peso; idade (anos); raça.

QUESTÃO 33 – Antes mesmo de entrar na parte da estatística descritiva, após a coleta dos dados, é importante saber qual o tipo de dados que irá trabalhar. Em uma pesquisa, foram coletadas as seguintes variáveis de um total de 200 pessoas:

- Idade.
- Renda.
- Estado Civil.
- Escolaridade.
- Número de Aparelhos de TV na Residência.
- Grau de Satisfação sobre a TV a Cabo (1 – Péssimo a 5 – Ótimo).

A classificação correta dessas seis variáveis é:

- A) Qualitativa, qualitativa, quantitativa, quantitativa, qualitativa e quantitativa.
- B) Quantitativa, quantitativa, qualitativa, quantitativa, quantitativa e qualitativa.
- C) Qualitativa, qualitativa, quantitativa, qualitativa, qualitativa e quantitativa.
- D) Quantitativa, quantitativa, qualitativa, qualitativa, quantitativa e qualitativa.
- E) Nenhuma das alternativas.

QUESTÃO 34 – Sobre os modelos de aprendizagem de máquina supervisionada, analise as assertivas abaixo e assinale a alternativa correta.

- I. Em modelos de aprendizado de máquina do tipo classificação a ideia é prever variáveis categóricas, e numéricas.
- II. Um exemplo básico de aprendizado de máquina supervisionado por classificação é o uso da regressão logística.
- III. Os modelos de regressão não buscam encontrar como uma variável se comporta na medida em que outra variável sofre oscilações.
- IV. Nos modelos de aprendizagem de máquina supervisionado, não temos uma variável específica a ser respondida, pois estamos apenas buscando encontrar os indivíduos, itens ou elementos semelhantes.

- A) Todas estão corretas.
- B) Todas estão incorretas.
- C) Apenas II está correta.
- D) Apenas I e II estão corretas.
- E) Apenas III e IV estão corretas.

QUESTÃO 35 – Podemos afirmar que uma árvore de decisão sofreu overfitting, quando ocorrer:

- A) erro validação (h) > erro treinamento (h)
- B) erro validação (h) - erro treinamento (h) > limiar determinado
- C) erro treinamento (h) = 0
- D) erro validação (h)/erro treinamento (h) < 1
- E) erro do treinamento (h) / erro de validação (h) / 100

QUESTÃO 36 – Sobre Machine Learning, é correto afirmar que:

- A) Podemos considerar correlação como sendo positiva, negativa e atemporal.
- B) Na correlação atemporal, não há relação entre todas as variáveis.
- C) Na correlação negativa, quando o valor de uma variável aumenta, as outras variáveis também sobem.
- D) A causalidade de um modelo de Machine Learning está diretamente relacionada ao volume interpretado do arquivo delta.
- E) Um dos motivos pelo qual é difícil interpretar os coeficientes angulares produzidos pelo processo de regressão linear é porque correlação não implica em causalidade.

QUESTÃO 37 – São hiperparâmetros para a configuração de uma rede neural, EXCETO:

- A) Taxa de aprendizado.
 - B) Número de neurônios na camada oculta.
 - C) Taxa de ajuste do erro.
 - D) Número de camadas ocultas da rede.
 - E) Número de características aleatórias que são exemplificadas na análise.
-

QUESTÃO 38 – São formas de separação entre base de treinamento e base de testes, EXCETO:

- A) Amostragem estratificada.
 - B) Lista de tuplas simplificadas.
 - C) Percentage Split.
 - D) Cross Validation.
 - E) Amostragem com reposição.
-

QUESTÃO 39 – Para minimizar o erro de estimação de um valor contínuo baseado em um conjunto de atributos, o algoritmo mais adequado para o problema é:

- A) AGNES – Agglomerative Nesting.
 - B) Naïve Bayes.
 - C) Árvore de sinergia.
 - D) Regressão linear.
 - E) K-means.
-

QUESTÃO 40 – Segundo Inmon, assinale a alternativa que apresenta características de um Data Warehouse.

- A) Orientado por assunto, integrado, volátil, variável no tempo.
 - B) Orientado por assunto, integrado, volátil, invariante no tempo.
 - C) Orientado por assunto, integrado, não volátil, variável no tempo.
 - D) Orientado por departamento, integrado, volátil, invariante no tempo.
 - E) Orientado por departamento, integrado, volátil, variável no tempo.
-

QUESTÃO 41 – Sobre o conceito de Business Intelligence, é correto afirmar que é:

- A) O processo de coleta, organização, análise, compartilhamento e monitoramento de informações para a gestão de negócios.
 - B) Um software.
 - C) O mesmo que inteligência artificial.
 - D) O nome dado a um algoritmo de pesquisa.
 - E) Um padrão de projetos.
-

QUESTÃO 42 – Sobre arquitetura das ferramentas OLAP, qual das alternativas é utilizada para se conectar à base de dados multidimensionais?

- A) DOLAP
 - B) WOLAP
 - C) MOLAP
 - D) SOLAP
 - E) ROLAP
-

QUESTÃO 43 – Sobre conceitos de Business Intelligence, analise as assertivas abaixo e assinale a alternativa correta.

- I. As ferramentas de Self Service BI não se enquadram dentro das ferramentas OLAP.
 - II. Não podemos utilizar ferramentas OLAP para tomada de decisão das organizações, uma vez que são consideradas ferramentas estáticas e de baixa performance.
 - III. É possível efetuar as operações de Drill Across, Drill Down e Drill Up em ferramentas OLAP.
 - IV. OLAP e MOLAP tem a mesma função e característica.
- A) Todas estão corretas.
 - B) Todas estão incorretas.
 - C) Apenas III está correta.
 - D) Apenas I e II estão corretas.
 - E) Apenas III e IV estão corretas.

QUESTÃO 44 – Sobre o conceito do Self-Service BI, assinale a alternativa correta.

- A) Na visão dos usuários é um bom negócio, pois eles podem se dedicar na segurança e governança dos dados.
- B) Compartilhamento das informações depende da TI.
- C) Disponibilidade de informações dinâmicas e interativas, permitindo que o usuário final crie/modifique seus relatórios.
- D) Maior dependência da TI.
- E) Ferramenta de banco de dados com características NOSQL com características de ferramentas de ETL.

QUESTÃO 45 – Sobre a representação do desafio da construção de uma solução de business intelligence, analise as assertivas abaixo e assinale a alternativa correta.

- I. Integração de dados de diferentes fontes.
 - II. Qualidade dos dados.
 - III. Criação de uma cultura organizacional baseada em dados.
 - IV. Falta de conhecimento técnico em ferramentas de business intelligence.
- A) Todas estão corretas.
 - B) Todas estão incorretas.
 - C) Apenas I está correta.
 - D) Apenas I e II estão corretas.
 - E) Apenas III e IV estão corretas.

QUESTÃO 46 – Para evitar a polarização dos modelos de aprendizado de máquina, as técnicas de balanceamento buscam equilibrar a quantidade de instâncias de cada classe do conjunto de dados. Dentre as diversas técnicas existentes, podemos citar: Seleção aleatória pela menor classe, Seleção por agrupamento pela menor classe e Replicação de instâncias. Sobre o assunto, analise as assertivas a seguir:

- I. Dado dois conjuntos de registros com N e M registros (onde $N < M$) vinculados a duas classes, o balanceamento por seleção aleatória ocorre selecionando de forma aleatória N registros dentro do conjunto contendo M registros.
- II. Dado dois conjuntos de registros com N e M registros (onde $N < M$) vinculados a duas classes, o balanceamento por seleção de grupo ocorre selecionando por meio de uma técnica de agrupamento os N registros mais representativos dentro do conjunto contendo M registros.
- III. Dado dois conjuntos de registros com N e M registros (onde $N < M$) vinculados a duas classes, o balanceamento ocorre gerando artificialmente instâncias a partir das instâncias do conjunto contendo M registros (classe majoritária).

Quais estão INCORRETAS?

- A) Apenas I.
- B) Apenas III.
- C) Apenas I e II.
- D) Apenas II e III.
- E) I, II e III.

QUESTÃO 47 – Durante a etapa do pré-processamento da base de dados, a análise de Outliers é uma tarefa comum e relevante para obter modelos de aprendizado de máquina consistentes. A presença de outliers pode levar a modelos imprecisos quando o modelo é testado ou colocado em produção. Analise as assertivas a seguir e assinale a alternativa correta.

- I. Outliers são dados com padrões muito diferentes aos demais, que fogem ao padrão dos dados. Estes dados precisam ser identificados e analisados.
- II. Outliers podem ser produzidos por erros de medição, valores default assumidos durante o preenchimento de uma base de dados ou podem corresponder a valores corretos, mas pertencentes a uma base de dados desbalanceada.
- III. Na prática, os outliers comumente são eliminados. Porém, pode-se estar negligenciando um conjunto de instâncias que podem trazer novos conhecimentos acerca do domínio de problema.
- IV. A detecção de outliers pode ser feita por meio de técnicas univariadas, que consistem em explorar cada atributo e variabilidade dos valores em torno da média. Quando a variabilidade é grande, pode indicar registros, potenciais outliers.

- A) Todas estão corretas.
- B) Todas estão incorretas.
- C) Apenas II e IV estão corretas.
- D) Apenas I, II e III estão corretas.
- E) Apenas II, III e IV estão corretas.

QUESTÃO 48 – Considere o seguinte esquema relacional e responda às questões apresentando o respectivo SQL:

Cliente - (cod_cliente, cliente, profissao, localidade)
Agencia - (cod_agencia, agencia, localidade)
Conta - (num_conta, tipo_conta, cod_cliente, cod_agencia, saldo)
Emprestimo - (num_emprestimo, cod_cliente, cod_agencia, valor)

Qual comando representa a quantidade de clientes cadastrados na base?

- A) Select cod_cliente from cliente;
- B) Select count (*) conta Where cod_cliente >1;
- C) Select count (*) from conta;
- D) Select * from cliente a inner join agencia b on a.cod_cliente = b.cod_agencia.
- E) Select count (*) from cliente;

QUESTÃO 49 – Considere o seguinte esquema relacional e resolva em SQL as questões propostas:

Pessoa - (Código Interno , Nome , Endereco , Tipo)
 Cliente -(Código Interno , Atividade)
 Empregado (Código Interno , Número Sequencial , Vencimento Mensal, Responsavel , Sigla , Data de Admissao)
 Departamento - (Sigla , Designação , Localização , Chefe, Data de Posse)

Quais os empregados, considerando o código de registro no sistema, que entraram no serviço da empresa no ano de 1995?

- SELECT Nome
 A) FROM Empregado
 WHERE year (Data de Admissao) = 1995
- SELECT Código Interno
 B) FROM Empregado
 WHERE Data de Admissao BETWEEN '01-01-1995' and '31-12-1995'
- SELECT *
 C) FROM Empregado
 WHERE Data de Admissao = 1995
- SELECT Nome
 D) FROM Pessoa
 WHERE year (Data de Admissao) = 1995
- SELECT Nome
 E) FROM Admissão
 WHERE year (Data de Admissao) = 1995

QUESTÃO 50 – Sobre a Linguagem SQL, é correto afirmar que:

- A) DML (Data Manipulation Language) é utilizado para criação, alteração e remoção de linhas de um banco de dados.
 B) DML é uma linguagem de escrita python para extração de soluções da descritiva de tuplas.
 C) DML é considerada por Ralph Kimball como uma linguagem para criação de um Data Warehouse.
 D) DML, linguagem para configuração de estrutura de banco de dados, é utilizada para remoção e alteração.
 E) DML (Data Manager Language) é utilizada para gerenciar um banco de dados.

QUESTÃO 51 – MDM (Master Data Management) é responsável por:

- A) Misturar os dados para criar um panorama virtual.
 B) Monitorar e processar fluxos de dados e ajudar a tomar decisões mais rapidamente.
 C) Fornecer tanto agendamento em lote quanto capacidades em tempo real.
 D) Unir os dados para criar uma visão única deles, através de múltiplas fontes. Ela inclui tanto o etl quanto capacidades de data integration, para misturar as informações e criar o "melhor registro".
 E) Criar um ambiente de testes onde os dados possam ser integrados, limpos e padronizados.

QUESTÃO 52 – Sobre o Fluxo de eficiente de um processo de ETL executar funções, analise as assertivas abaixo e assinale a alternativa correta.

- I. Extração de Dados de Fontes Diversas.
 II. Transformação de dados.
 III. Cargas dos dados.
 IV. Multiplicação das informações.
- A) Todas estão corretas.
 B) Todas estão incorretas.
 C) Apenas I e II estão corretas.
 D) Apenas III e IV estão corretas.
 E) Apenas I, II e III estão corretas.

QUESTÃO 53 – No processo de ETL, qual é o nome dados para a área intermediária criada, antes de serem processados?

- A) cubo OLAP.
- B) dicionário de dados.
- C) staging.
- D) data vault.
- E) data mart.

QUESTÃO 54 – Sobre os métodos que podemos considerar em um processo de construção de um Data Warehouse, analise as assertivas abaixo:

- I. Star Schema.
- II. Bifurcação entre tabelas.
- III. Snowflake.
- IV. Fullflake.
- V. Dimension Only.

Quais estão corretas?

- A) Apenas I.
- B) Apenas V.
- C) Apenas I e III.
- D) Apenas II, III e IV.
- E) I, II, III, IV e V.

QUESTÃO 55 – Sobre granularidade de um modelo dimensional, assinale a alternativa correta.

- A) Grão é o menor nível de informação e é definido de acordo com as necessidades do modelo de dados.
- B) Grão é o nome dados a uma tabela de um modelo Starschema.
- C) Grão é considerado um prefixo de nomenclatura de tabela de banco de dados na construção de um modelo dimensional.
- D) Grão é a sumarização dos caracteres de um resultado de uma consulta de select no banco de dados PostgreSQL.
- E) O Grão de uma tabela é a soma de todas as colunas de um banco de dados.

QUESTÃO 56 – Podemos afirmar que uma tabela factless é:

- A) São tabelas fato comuns, com métricas a atributos que referenciam determinadas dimensões.
- B) As tabelas de fato sem fato são encontradas na modelagem de um data warehouse, elas não possuem nenhuma medida e contém apenas chaves estrangeiras para as dimensões.
- C) São tabelas snapshots, que acumulam e descrevem as atividades de negócio do início ao fim.
- D) Tabela Factless são conhecidas como aggregate table, armazenam e pré-calculam as informações, são alimentadas por uma tabela intermediária.
- E) São tabelas excluídas de um modelo multidimensional e armazenada como backup em uma stagin área.

QUESTÃO 57 – MongoDB é um banco de dados não relacional. Esses bancos de dados também são chamados de NOSQL. Permitem controlar com precisão os dados armazenados globalmente, para garantir desempenho rápido em qualquer lugar. Um desenvolvedor está trabalhando com o modelo deste banco de dados não relacional, e encontra-se com o prompt de comando do Windows aberto como administrador, localizado na pasta c:\mongo\bin, o comando: mongod --dbpath D:\BD\Data\ é executado na linha de comando. Considerando as informações acima, a execução deste comando é realizada para:

- A) Executar o Mongo Cliente para executar as queries na coleção data.
- B) Executar o servidor do Mongo, pois, no processo de instalação, o caminho não foi configurado nas variáveis de ambiente do Windows.
- C) Criar os arquivos de log e executar o Mongo cliente na pasta BD/Data.
- D) Executar o Mongo e criar o banco de dados através do comando mongod.
- E) Criar documentos em JSON, utilizando a estrutura Binary JSON.

QUESTÃO 58 – Classificado como um programa de banco de dados NoSQL, o MongoDB é um banco de dados de código aberto e de alta performance, que se caracteriza por não possuir esquemas, ser escrito em C++, ser multiplataforma e ser formado por um conjunto de aplicativos JSON. O MongoDB é um banco de dados NoSQL:

- A) Orientado a colunas, semelhante ao banco de dados BigTable do Google.
- B) Orientado a documentos, com estruturas flexíveis que podem ser obtidas por meio de dados semiestruturados, como o formato XML e JSON.
- C) Orientado a objetos e possui conceitos similares ao do modelo relacional, como a criação de linhas e colunas.
- D) Do tipo chave-valor, que usa uma tabela hash na qual há uma chave única e um indicador de um dado ou de um item em particular.
- E) Orientado a grafos, semelhante ao banco de dados Neo4J.

QUESTÃO 59 – Sobre os tipos de banco de dados NOSQL, analise as assertivas abaixo:

- I. Colunar.
- II. Grafos.
- III. Chave-valor.
- IV. Documento.

Quais estão corretas?

- A) Apenas II.
- B) Apenas III.
- C) Apenas I, II e III.
- D) Apenas II, III e IV.
- E) I, II, III e IV.

QUESTÃO 60 – Uma das características da linguagem HIVEQL é ter uma extensão que permita distribuir consultas entre reducers em um script do tipo map-reduce. Qual extensão deve-se usar para distribuir as consultas e garantir a ordenação ou o agrupamento das chaves distribuídas?

- A) DISTRIBUTE BY
- B) CLUSTER BY
- C) GROUP BY
- D) REDUCE BY
- E) USING

QUESTÃO 61 – Sobre BIG Data, analise as assertivas abaixo e assinale a alternativa correta.

- I. Hadoop e Elasticsearch são exemplos de tecnologias que permitem a computação em nuvem.
 - II. BIG Data é o termo em tecnologia da informação que tem a capacidade de trabalhar com grandes conjuntos de dados.
 - III. O Conceito de BIG DATA pode ser subdividido em 5 categorias ou "5 v's", VOLUME, VALOR, VELOCIDADE, VERACIDADE e VARIEDADE.
 - IV. Um dos maiores desafios em trabalhar com soluções de BIG DATA é o volume dos dados.
- A) Todas estão corretas.
 - B) Todas estão incorretas.
 - C) Apenas I está correta.
 - D) Apenas I e II estão corretas.
 - E) Apenas III e IV estão corretas.

QUESTÃO 62 – Recentemente, encontram-se muitas referências na literatura e na mídia em geral ao uso de aprendizagem profunda (ou “*deep learning*”). Sobre o assunto, assinale a alternativa INCORRETA.

- A) Na aprendizagem profunda, conceitos complexos (de níveis mais altos ou abstratos) são definidos em termos de sua relação com conceitos mais simples (em níveis mais baixos).
- B) A aprendizagem profunda não apresenta nenhuma relação com o aprendizado de máquina, podendo ser vista como uma abordagem alternativa para tarefas em que o aprendizado de máquina não consegue obter bons resultados.
- C) A aprendizagem profunda pode ser entendida como um tipo de aprendizagem de representação em que características mais abstratas são inferidas a partir de características mais simples.
- D) Uma das razões determinantes para o recente sucesso da aprendizagem profunda se encontra na maior disponibilidade de grandes conjuntos de dados para as mais diversas tarefas.
- E) Deep Learning é um ramo de aprendizado de máquina.

QUESTÃO 63 – Uma função de perda mede a diferença entre uma predição do valor alvo e o valor disponível no conjunto de treinamento. Sobre o assunto, assinale a alternativa INCORRETA.

- A) A função de perda permite que se busque por um modelo que minimize o risco empírico.
- B) O uso de uma função de perda garante a minimização do risco.
- C) Existem inúmeras funções de perda e, portanto, deve-se selecionar a mais adequada para se avaliar a proximidade entre predição e alvo esperado.
- D) O uso de uma função de perda é uma alternativa interessante para se avaliar a proximidade entre predição e alvo esperado.
- E) Uma função de perda mapeia um evento ou valores de uma ou mais variáveis.

QUESTÃO 64 – A ideia principal da técnica de “*dropout*” é descartar aleatoriamente unidades da rede neural (junto com suas conexões) durante a etapa de treinamento. Sobre a técnica de “*dropout*”, assinale a alternativa INCORRETA.

- A) O uso da técnica de “*dropout*” força a rede a construir uma representação distribuída e redundante.
- B) O uso da técnica de “*dropout*” pode ser visto como equivalente a treinar uma grande coleção (ou “*ensemble*”) de modelos que compartilham parâmetros.
- C) Para a realização de “*dropout*”, basta multiplicar-se as ativações por variáveis aleatórias de Bernoulli com uma dada probabilidade.
- D) Durante a predição, todas as unidades são mantidas ativas, não sendo necessário se realizar nenhum outro ajuste das ativações.
- E) Ao contrário de regularização L1 e L2, o dropout não depende da modificação da função de custo.

QUESTÃO 65 – Sobre Apache Spark, analise as assertivas abaixo e assinale a alternativa correta.

- I. De forma geral, o Spark é uma engine rápida, escrita em Scala, para processamento de grandes volumes de dados em um cluster de computadores.
 - II. Scala é uma linguagem funcional que roda na JVM.
 - III. O Spark, assim como o Hadoop, também foi pensado para ser escalável.
 - IV. Uma das grandes vantagens do Spark em relação ao Hadoop são as High Level API's de programação. Enquanto que o Hadoop é nativamente apenas Java, no Spark, temos a disposição API's em Scala, Java e Python.
- A) Todas estão corretas.
 - B) Todas estão incorretas.
 - C) Apenas I está correta.
 - D) Apenas I e II estão corretas.
 - E) Apenas III e IV estão corretas.

QUESTÃO 66 – A programação em Spark se baseia no conceito de RDD's. Os RDD's são a unidade fundamental de dados no Spark e têm como principal característica a propriedade de ser imutável. Sobre os RDD's, é correto afirmar que:

- A) Os RDD's não podem ser particionados ou divididos no cluster.
- B) As operações em RDD são divididas em 5 tipos, transformação, ação, segmentação, logarítmica e semântica.
- C) As operações logarítmicas do RDD, são responsáveis pelas ações de classificação dos modelos desenvolvidos em SPARK.
- D) RDD's são algoritmos desenvolvidos em JAVA para o cálculo de espaço em disco para armazenamento das informações distribuídas em SPARK.
- E) Os RDD's podem armazenar qualquer tipo de elemento, como tipos primitivos (inteiros, caracteres, booleanos, etc.), sequências (strings, listas, arrays, tuples, etc.), assim como tipos mistos e objetos Scala ou Java.

QUESTÃO 67 – São exemplos de operações de transformação em Spark, EXCETO:

- A) map (func) – retorna um novo RDD aplicando a função func em cada elemento.
- B) filter (func) – retorna um novo RDD aplicando o filtro func.
- C) flatMap (func) – similar ao map, mas retornando mais itens ao invés de apenas um.
- D) Joinstep – responsável pelo relacionamento de objetos.
- E) distinct() – retorna um novo dataset contendo os valores distintos do RDD original.

QUESTÃO 68 – Sobre Apache Spark, analise as assertivas abaixo e assinale a alternativa correta.

- I. Similar a linguagem SQL, Spark SQL funciona de maneira similar ao Hive, que converte códigos SQL para Map Reduce Java.
 - II. Spark Streaming é uma extensão do Spark, voltada para processamento de dados em tempo real, apresentando diversas propriedades interessantes, entre elas, podemos destacar a escalabilidade e a tolerância a falhas, processamento único e a possível integração entre processos batch e em tempo real.
 - III. O MLlib consiste em uma biblioteca de códigos de machine learning prontos e disponíveis para uso, funcionando de forma muito parecida aos pacotes do R ou ao numpy e ao scikit-learn do python.
 - IV. O GraphX, foi desenvolvido com o intuito de substituir os sistemas especializados de grafos que foram feitos para Hadoop, permitindo a análise e processamento de grafos em paralelo.
- A) Todas estão corretas.
 - B) Todas estão incorretas.
 - C) Apenas I está correta.
 - D) Apenas I e II estão corretas.
 - E) Apenas III e IV estão corretas.

QUESTÃO 69 – Sobre NLP, é INCORRETO afirmar que:

- A) NLP trata a possibilidade de humanos se comunicarem com máquinas da forma mais natural possível, sem ser preciso aprender línguas artificiais muito específicas, que são os comandos nos quais manipula-se um computador.
- B) Linguagem natural destina-se à comunicação dos seres humanos sobre o mundo, sendo que a maior parte da comunicação linguística ocorre através da fala. A linguagem escrita ainda é muito recente se comparado à fala, e para um computador, mais fácil de ser interpretada.
- C) NLP pode ser entendido como aplicações computacionais que compreendem linguagens naturais (compreensão e geração de textos), tanto na fonética quando na escrita, sendo que a última é mais fácil de um computador conseguir interpretar de maneira correta, pois não tem que lidar com problemas do sinal do áudio, como o ruído por exemplo.
- D) É possível através da NLP interpretar na língua portuguesa a interpretação.
- E) NLP é uma área dentro da Inteligência Artificial.

QUESTÃO 70 – Sobre os termos das características de processo de processamento de texto NLP, analise as assertivas abaixo e assinale a alternativa correta.

- I. Stopwords.
- II. Tf-id+f (Term Frequency-inverse document frequency).
- III. Word embedding.
- IV. Word2vec.

- A) Todas estão corretas.
- B) Todas estão incorretas.
- C) Apenas I está correta.
- D) Apenas I e II estão corretas.
- E) Apenas III e IV estão corretas.