

PROGRAMA DE RESIDÊNCIA EM TECNOLOGIA DA INFORMAÇÃO - TRE-RN
EDITAL 001/2024 - PROVA DE CONHECIMENTOS ESPECÍFICOS
ESPECIALISTA CONVIDADO

Candidato: _____

CPF: _____

Telefone: _____

QUESTÕES OBJETIVAS

01. Dada a seguinte base de dados:

#	x1	x2	x3	x4	x5	x6	x7	x8	x9	Class
1	5	2	3	4	2	7	3	6	1	malignant
2	5	4	4	5	7	10	3	2	1	benign
3	3	1	1	1	2	2	3	1	1	benign
4	6	8	8	1	3	4	3	7	1	benign
5	4	1	1	3	2	1	3	1	1	benign
6	8	10	10	8	7	10	9	7	1	malignant
7	1	1	1	1	2	10	3	1	1	benign
8	2	1	2	1	2	1	3	1	1	benign
9	8	7	5	10	7	9	5	5	4	malignant
10	7	4	6	4	6	1	4	3	1	malignant

Utilizando o classificador k-NN (*k-nearest neighbors* ou “k-vizinhos mais próximos”) com os valores 1, 2 e 3 para k , e a distância Euclidiana como métrica de distância, classifique esta instância de teste: (10, 10, 10, 4, 8, 1, 8, 10, 1). Quais serão os resultados para os diferentes valores de k ?

- A) benign, malignant, benign
- B) benign, malignant, malignant
- C) benign, benign, malignant
- D) malignant, malignant, benign

02. De acordo com o processamento do trecho de código abaixo, escrito em Python (versão 3.11.3), é correto afirmar que:

```
class Funcionario:
    titulos = [ 'Doutor', 'Mestre', 'Especialista']

    def __init__(self, nome, titulo, salario=2949):
        self.nome = nome
        self.titulo = titulo
        self.salario = salario

    @property
    def exibe_salario(self):
        return self.salario

    def exibe_titulo(self):
        if self.titulo not in self.titulos:
            raise ValueError('Titulo inválido')
        return self.titulo

funcionario = Funcionario( 'Francisco', 'Especialista')
print (funcionario.exibe_salario)
print (funcionario.exibe_titulo())
```

- A) funciona corretamente e o valor do salário e o título do funcionário são exibidos ao fim da execução.
- B) lança um erro devido ao problema de sintaxe na chamada “funcionario.exibe_salario”.
- C) lança um erro devido ao problema de sintaxe na chamada “funcionario.exibe_titulo()”.
- D) lança um erro por faltar o parâmetro “salario” na instancição da classe.

03. Frameworks são criados para linguagens de programação com o intuito de facilitar o desenvolvimento de uma determinada aplicação. Dentre os *frameworks* abaixo, qual deles é voltado para a linguagem Python no que se refere ao desenvolvimento *web*?

- A) Laravel.
- B) Scikit-learn.
- C) Django.
- D) Pandas.

04. Pentaho Data Integration (PDI) é um conjunto de ferramentas *open source* responsável pelos processos de ETL da Pentaho Business Intelligence Suite. No PDI, os componentes de ligação entre os *steps* que definem a direção da execução da transformação são conhecidos como:

- A) Hops.
- B) Kettle.
- C) Spoon.
- D) Table input.

05. Qual é a diferença entre a cláusula WHERE e a cláusula HAVING em uma consulta SQL?

- A) WHERE é usada para filtrar registros com base em critérios especificados, enquanto HAVING é usada para filtrar grupos agregados com base em critérios especificados.
- B) WHERE é usada para agrupar registros com base em critérios especificados, enquanto HAVING é usada para ordenar registros com base em critérios especificados.
- C) WHERE e HAVING são usadas de forma intercambiável em consultas SQL.
- D) WHERE é usada para agrupar registros com base em critérios especificados, enquanto HAVING é usada para filtrar registros com base em critérios especificados.

06. Considere o fluxo de transformação do *Pentaho Data Integration* apresentado a seguir:



Os passos S1, S2 e S3 realizam, respectivamente, as operações de:

- A) Leitura de Arquivo JSON, Split de Colunas e Saída em Formato de Tabela.
- B) Leitura de Banco de Dados, Join de Colunas e Log.
- C) Leitura de Arquivo CSV, Split de Colunas e Calculadora de Valores.
- D) Leitura de Arquivo de Propriedades, Seleção de Linhas Únicas e Calculadora de Valores.

07. Dados ausentes ou missing values representam uma das maiores dificuldades da etapa exploratória dos dados em Mineração de Dados. Dentre as opções abaixo, escolha apenas aquela que apresenta abordagens que preencham os dados ausentes.

- A) `df.dropna(axis=0); df.describe(); df.fillna()`
- B) `df.isnull().sum(); df.dropna(axis=1); df.value_counts()`
- C) `df.fillna(); df.fillna(value_counts()); df.fillna(median())`
- D) nenhuma das opções.

08. Assinale a consulta SQL que recupera exatamente os 10 primeiros registros de uma tabela.

- A) `SELECT * FROM usuarios WHERE codigo = 10`
- B) `SELECT * FROM usuarios WHERE LIMIT BETWEEN 1 AND 10`
- C) `SELECT * FROM usuarios ORDER BY codigo ASC LIMIT 10`
- D) `SELECT * FROM usuarios ORDER BY codigo DESC LIMIT 10`

09. CRISP-DM (*Cross Industry Standard Process for Data Mining*) é uma metodologia utilizada em projetos orientados a dados (*data driven*). De acordo com esta metodologia, a definição do problema ocorre na seguinte etapa:

- A) data understanding.
- B) business understanding.
- C) evaluation.
- D) data preparation.

10. No Power BI, qual o nome da linguagem de transformação de dados do Power Query?

- A) M
- B) VB
- C) BAX
- D) R

11. Vários operadores são binários em Álgebra Relacional. Dessa forma, é possível que uma operação binária entre duas relações crie uma terceira. Dentre os operadores listados abaixo, assinale o único que **NÃO** é binário:

- A) divisão
- B) projeção
- C) junção natural
- D) produto cartesiano

12. Considere um conjunto de dados que possui o atributo do tipo contínuo TEMPERATURA com os valores {38.0, 39.5, 36.0, 35.5}. Para alimentar uma tarefa de Mineração de Dados, é necessário transformar os dados desse atributo em um número finito de intervalos, como: {35-37, 38-39, 40-41}. A tarefa de preparação de dados utilizada para tal seria:

- A) partição de dados
- B) normalização
- C) discretização
- D) padronização

13. Considere a seguinte estrutura de dados.

```
{
  titulo: "Reforma tributária",
  corpo: "A aprovação foi realizada ontem na Câmara dos Deputados.",
  categoria: "Notícias",
  vistas: 800,
  assuntos: ["lei", "impostos"],
  data: 23/01/2024
}
```

No que se refere aos Bancos de Dados NoSQL, esse tipo de estrutura é considerado(a):

- A) uma *hash table*.
- B) um grafo.
- C) um documento
- D) um par chave/valor

14. Modelagem dimensional é uma técnica normalmente utilizada para modelar um Data Warehouse. Sobre esta técnica é correto afirmar que:

- I. Proposta por R. Kimball em 1996;
- II. Possui três modelos de dados, sendo: estrela, floco de neve e híbrido;
- III. O modelo floco de neve reproduz o Modelo Relacional no Data Warehouse.

- A) somente as afirmativas I e III estão corretas.
- B) somente as afirmativas II e II estão corretas.
- C) somente a afirmativa II está correta.
- D) somente a afirmativa III está correta.

15. Sobre a linguagem de programação Python, é correto afirmar que:

- I. São exemplos de palavras reservadas do Python: `lambda`, `raise`, `finally`.
- II. A linguagem Python diferencia caracteres minúsculos de maiúsculos, ou seja, é *Case Sensitive*.
- III. A função `sqrt()` é uma função de exponenciação do Python.

- A) Somente os itens I e III estão corretos.
- B) Somente os itens I e II estão corretos.
- C) Os itens I, II e III estão corretos.
- D) Somente os itens II e III estão corretos.

16. Na literatura, é possível encontrar os termos *data warehouse* e *data mart*. Acerca desses termos, selecione a alternativa correta:

- A) *Data mart* é um sinônimo de *data warehouse*.
- B) Um *data mart* não tem nenhuma relação com um *data warehouse*.
- C) Um *data mart* pode ser entendido como uma combinação de *data warehouses*.
- D) Um *data warehouse* pode ser visto como uma combinação de *data marts*.

17. *Dashboards* são soluções que auxiliam os usuários do negócio a visualizarem de forma rápida os dados. Nesse contexto, marque o item que **NÃO** possua uma característica desejável dos *dashboards*:

- A) Informação em tempo real.
- B) Apresentar os chamados indicadores chave de desempenho.
- C) Informar em detalhes todos os dados do negócio.
- D) Transparecer metas, prioridades e níveis de desempenho.

18. No momento de construir um *dashboard*, a escolha dos elementos visuais é uma tarefa muito importante para o sucesso da interpretação do que se quer visualizar. Dessa forma, podemos dizer que:

- A) Os gráficos de pizza e de barras devem ser os únicos utilizados dentro de um *dashboard*.
- B) Os gráficos de linhas são úteis quando queremos visualizar a evolução de uma determinada entidade (produto, evento, etc) ao longo do tempo.
- C) Os gráficos de pizza ou rosca são indicados quando temos um número muito grande de categorias.
- D) Os gráficos de dispersão são mais indicados quando cada entidade a ser representada possui muitas características ou dimensões.

19. Dentre as opções a seguir, marque a alternativa que contém pelo menos uma ferramenta que **NÃO** pode ser utilizadas para construir dashboards no contexto de *Business Intelligence*:

- A) QlikView e Looker Studio
- B) Pentaho e Metabase
- C) Kibana e Tableau
- D) Power BI e Hadoop

20. A seguir é apresentado um trecho de código escrito na linguagem Python.

```
x = [2,3,1,5,4]
for i in range(len(x)-1):
    x[i],x[i+1] = x[i+1],x[i]
print(x)
```

O valor da variável **x** após a execução desse trecho de código será:

- A) [2, 3, 1, 5, 4]
- B) [3, 1, 5, 4, 2]
- C) [3, 4, 1, 2, 5]
- D) [4, 2, 5, 3, 1]

21. O Power BI é uma coleção de serviços de *software*, aplicativos e conectores que trabalham juntos para transformar suas fontes de dados não relacionadas em informações coerentes. Sobre o Power BI, assinale a afirmativa correta.

- A) *Deployment pipeline* é um recurso disponível no Power BI Desktop que permite testar relatórios antes do lançamento para os usuários.
- B) *Paginated Reports* são relatórios criados no Power Bi Service para serem exibidos em dispositivos com limitação de memória, como celulares e tablets.
- C) *Power Query Editor* permite conectar a uma ampla variedade de tipos de fontes de dados, porém, é necessário usar a linguagem BAX para acesso aos dados
- D) *Direct Query* permite criar visualizações de conjuntos de dados muito grandes, nos casos em que seria impraticável importar todos os dados com pré-agregação.

22. Foi implementado, nas versões mais recentes do Python (3.10 e posteriores), uma estrutura de seleção múltipla, qual das estruturas abaixo representa essa implementação?

- A) CASE-OF
- B) SWITCH-CASE
- C) CASE-IN
- D) MATCH-CASE

23. Por meio da ferramenta *Pentaho Data Integration* (PDI), é possível realizar processos de ETL (*Extraction, Transformation and Load*). A ferramenta gráfica que faz parte do PDI, usada para modelar as transformations e os jobs, é denominada de:

- A) spoon
- B) kettle
- C) kitchen
- D) mondrian

24. Qual é o resultado da seguinte consulta SQL, que usa a função RANK()?

```
SELECT nome, salario,
RANK() OVER (ORDER BY salario DESC) AS ranking
FROM funcionarios;
```

- A) Retorna o nome e o salário de cada funcionário, mas não inclui informações de classificação.
- B) Retorna um erro, pois as funções de janela não são suportadas em banco de dados relacionais.
- C) Retorna o nome, o salário e a classificação de cada funcionário com base no salário, em ordem decrescente.
- D) Retorna o nome e o salário de cada funcionário, em ordem alfabética decrescente.

25. O que é consistência eventual em bancos de dados NoSQL?

- A) Todos os dados são consistentes em tempo real.
- B) A consistência é alcançada eventualmente, mas não é garantida em tempo real.
- C) Os dados são consistentes apenas quando não há falhas no sistema.
- D) A consistência é garantida por meio de transações.

26. O que é o teorema CAP (Consistência, Disponibilidade e Tolerância a Partições) em relação a bancos de dados NoSQL?

- A) É um teorema que descreve a estrutura de tabelas em bancos de dados NoSQL.
- B) É um teorema que define os tipos de dados que podem ser armazenados em bancos de dados NoSQL.
- C) É um teorema que descreve as limitações na capacidade de um sistema distribuído em garantir simultaneamente consistência, disponibilidade e tolerância a partições.
- D) É um teorema que descreve a escalabilidade de bancos de dados NoSQL.

27. O que é um DAG (*Directed Acyclic Graph*) no contexto do Apache Airflow?

- A) Uma estrutura de dados para armazenar registros de log.
- B) Uma representação visual de um fluxo de trabalho com tarefas e suas dependências.
- C) Um tipo de banco de dados NoSQL.
- D) Um formato de arquivo usado para armazenar configurações.

28. Qual das seguintes opções inclui apenas órgãos que fazem parte do Poder Judiciário?

- A) Supremo Tribunal Federal (STF), Tribunal de Contas da União (TCU).
- B) Superior Tribunal de Justiça (STJ), Tribunal Marítimo (TM).
- C) Ministério Público Federal (MPF), Tribunal Superior do Trabalho (TST).
- D) Conselho Nacional de Justiça (CNJ), Tribunal Superior Eleitoral (TSE).

29. Qual o órgão que representa o Poder Judiciário Municipal?

- A) Tribunal Municipal.
- B) Juntas Municipais.
- C) Auditorias Municipais.
- D) Não existe Poder Judiciário Municipal.

30. O Tribunal Regional Eleitoral do Rio Grande do Norte é um órgão:

- A) Municipal.
- B) Estadual.
- C) Federal.
- D) Misto.

QUESTÕES DISCURSIVAS

31. Imagine que você é o responsável pela parte de visualização de dados de uma empresa qualquer. Sabendo disso, o engenheiro de dados disponibilizou um arquivo CSV com os dados que serão visualizados através do painel (*dashboard*) que será desenvolvido por você. Veja abaixo uma pequena amostra dos dados (VendasProdutos.csv).

idProduto	Produto	Categoria	Segmento	Fabricante	idLoja	Cidade	Estado	Vendedor	idVendedor	DataVenda	ValorVenda
SKU-0000001	LG K10 TV Power	Celulares	Corporativo	LG	SP8821	São Paulo	São Paulo	Ana Teixeira	1009	04/10/2012	R\$ 679,00
SKU-0000002	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Josias Silva	1006	01/01/2012	R\$ 832,00
SKU-0000003	Lavadora 11 Kg	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Josias Silva	1006	02/02/2012	R\$ 790,00
SKU-0000004	Lavadora 11 Kg	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Mateus Gonçalves	1003	03/03/2012	R\$ 765,32
SKU-0000005	Lavadora 11 Kg	Eletrodomésticos	Doméstico	Electrolux	SP8821	São Paulo	São Paulo	Artur Moreira	1004	04/04/2012	R\$ 459,89
SKU-0000006	Lavadora 11 Kg	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Rodrigo Fagundes	1005	04/05/2012	R\$ 590,98
SKU-0000007	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Josias Silva	1006	04/06/2012	R\$ 1.000,91

Porém, antes de implementar o painel você precisará aplicar os conceitos de modelagem dimensional, e para tal, precisará construir um modelo estrela. Dessa forma, mostre graficamente (diagrama) como ficaria o seu modelo.

32. Imagine que você passou a ser o engenheiro de dados responsável pela construção dos modelos dimensionais de uma empresa, onde todos os modelos são do tipo estrela. Com base no arquivo CSV, da questão anterior, faça uma função denominada “*cria_dimensao_csv*”, utilizando preferencialmente Python, para criação das dimensões, onde o parâmetro dessa função é o arquivo CSV e o retorno dessa função é uma estrutura de dados do tipo dicionário onde a chave é o nome da dimensão e o valor é a própria dimensão. (Considere que cada coluna é uma dimensão, que o nome da dimensão é o título da coluna e todas as colunas são do tipo *string*).