



Prova Online – Automação do Fluxo de Dados

PROF.: Leandro Lessa

Objetivos de Ensino

Exercitar os seguintes conceitos:

1. Coleta de dados
2. Analisar e explorar os conjuntos de dados
3. Analisar e realizar tratamento de dados se necessário
4. Realizar agrupamentos de dados
5. Analisar resultados obtidos
6. Responder as questões objetivas

Enunciado da Atividade Prática: Análise de Dados

Você foi contratado para criar uma automação de fluxo de dados completa para dados clínicos dos pacientes em um projeto de pesquisa de saúde. Seus dados incluem três conjuntos principais: `dados_clinicos`, `dados_pacientes`, e `dados_estados`. Cada conjunto de dados possui as seguintes colunas:

- I. **`dados_clinicos`**: contém informações sobre os pacientes como `id_cliente`, `peso`, `colesterol` e `genero`.
- II. **`dados_pacientes`**: fornece detalhes sobre os pacientes como `id_cliente`, `idade`, `classe_trabalho`, `escolaridade`, `id_estado`, `estado_civil`, `raca`, `qtde_filhos`, e `salario`.
- III. **`dados_estados`**: oferece informações sobre os estados, incluindo `id_estado`, `sigla`, `estado`, `regiao` e `pais`.

Os datasets utilizados no trabalho podem ser obtidos nesse [link](#).

Tarefas:

1. Coleta de Dados:

Realize a coleta de dados dos pacientes disponibilizados.

2. Integração dos Dados:

Realize um inner join entre `dados_clinicos`, `dados_pacientes` e `dados_estados` para obter uma visão completa dos dados dos pacientes.

3. Correção dos Dados:

Identifique e corrija dados inconsistentes, como valores ausentes, duplicados ou incorretos.

4. Armazenamento de Dados:

Desenvolva o diagrama de entidade-relacionamento entre as tabelas e persista os dados no banco de dados MySQL.

5. Consultas SQL:

Após o armazenamento dos dados, responda às perguntas do questionário utilizando SQL.

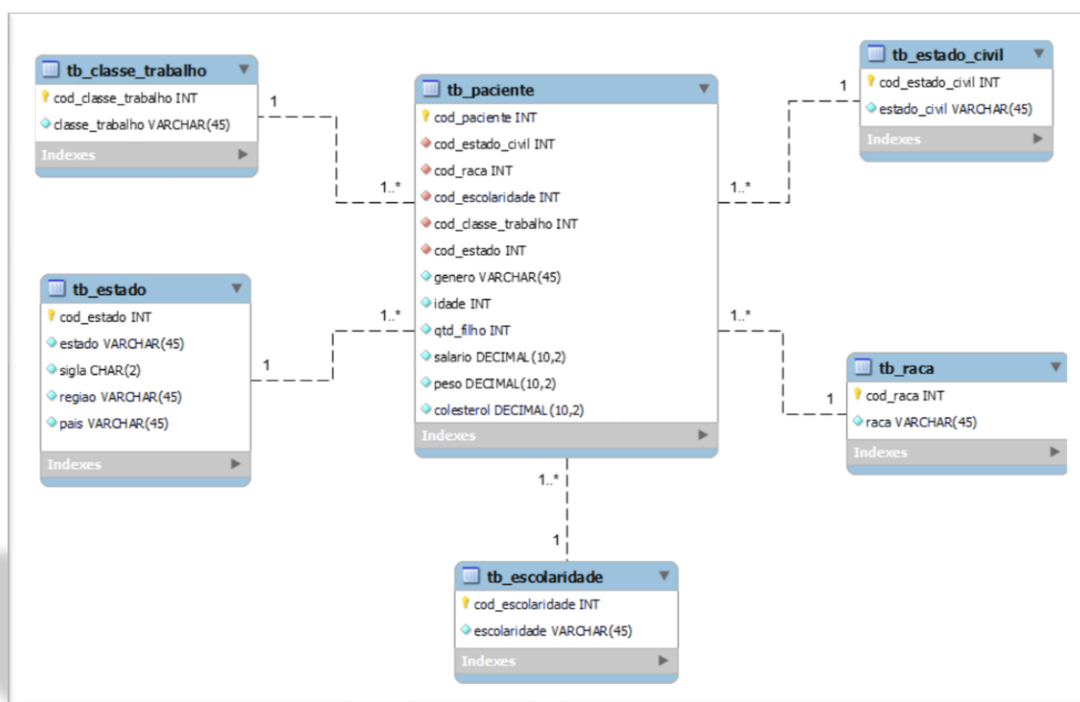
ATENÇÃO PARA TRATAMENTO DE DADOS

Avaliem se será necessário realizar tratamento de dados ausentes no datasets disponibilizados.

Instruções para correção de dados ausentes

- Moda para Variáveis Categóricas Preencher valores ausentes nas colunas categóricas com a moda.
- Mediana para Variáveis Numéricas: Preencher valores ausentes nas colunas numéricas com a mediana.

Utilize o diagrama de entidade e relacionamento



Dicas do professor:

1. Antes de finalizar suas respostas, certifique-se de verificar se o gabarito está corretamente preenchido.
2. Preste atenção aos requisitos específicos de cada questão para garantir uma resposta adequada.
3. Elimine dados duplicados.
4. Lembre-se de que os dados fornecidos nos datasets são fictícios e não refletem dados do mundo real.
5. Siga rigorosamente todas as etapas descritas no enunciado das questões para uma análise completa e precisa.