

Bootcamp: Cientista de Dados

Trabalho Prático

Módulo 2	CDD – Coleta e Obtenção de Dados
-----------------	---

Objetivos

Exercitar os seguintes conceitos trabalhados no Módulo:

- ✓ Realizar coleta de dados em arquivos utilizando uma das seguintes opções de tecnologia: Linguagem R ou Python;
- ✓ Realizar operações de criação de bases de dados relacional utilizando uma das seguintes opções: Linguagem R ou Python.
- ✓ Realizar carga e coleta de dados em SGBD relacional utilizando uma das seguintes opções de tecnologia: Linguagem R ou Python.

Enunciado

Para esta atividade, o aluno deverá assistir atentamente as aulas disponíveis no ambiente virtual de aprendizagem.

Considere ainda os arquivos complementares anexo ao enunciado do Trabalho Prático:

- Script para criação do banco de dados: *script_BD_bootcamp.sql*.
- Arquivos de carga para as tabelas do banco de dados:
 - *estado.xlsx*
 - *cidade.csv*
 - *caracteristicasgerais.csv*
 - *imoveis.csv*
 - *caracteristicageralimovel.txt*

– *caracteristicageralimovel.csv*

- Arquivos com os notebooks R e Python:

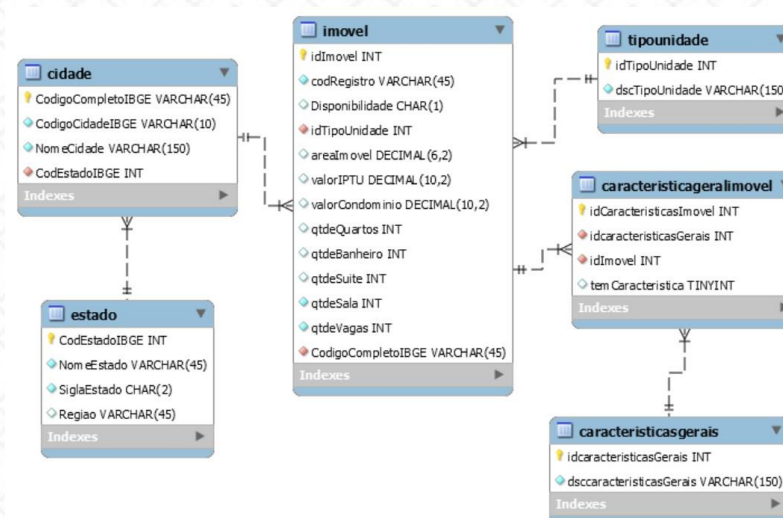
– *coletaDadosMySQL_R.ipynb*

– *coletaDadosMySQL_Python.ipynb*

Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Criar o banco de dados “*bootcamp*” com as tabelas do esquema, conforme o modelo a seguir. Para isso, utilize o arquivo “*bootcamp.sql*”, disponível em conjunto com o enunciado deste trabalho. Durante a criação, analise o script gerado, pois você vai precisar dele para responder as questões.



2. Execute o notebook “*coletaDadosMySQL_R.ipynb*”, disponível em conjunto com o enunciado deste trabalho. Analise e compreenda o código do notebook e sua respectiva execução, pois você vai responder questões relacionadas a ele.
 - Para esta atividade, você vai precisar do arquivo “*estados.xlsx*”, salvo na pasta “*C:\Bootcamp\Datasets\XLS*”. Se deseja alterar a pasta, certifique-se de fazer a mesma alteração no notebook.

3. Alterar o notebook “coletaDadosMySQL_R.ipynb”, utilizando a linguagem R, para inserir novos registros na tabela *caracteristicasgerais*. Para apoiar a execução desta atividade, foi fornecido o arquivo csv chamado “*caracteristicasgerais.csv*”.
4. Executar o notebook “coletaDadosMySQL_Python.ipynb”, disponível em conjunto com o enunciado deste trabalho. Analise e compreenda o código do notebook e sua respectiva execução, pois você vai responder questões relacionadas a ele.
 - Para esta atividade, você vai precisar do arquivo “*idades.csv*” salvo em na pasta “C:\Bootcamp\Datasets\CSV”. Se deseja alterar esta pasta, certifique-se de fazer a mesma alteração no notebook.
5. Alterar o notebook “coletaDadosMySQL_Python.ipynb”, utilizando a linguagem Python, para inserir registros na tabela *imovel*. Para apoiar a execução desta atividade, foi fornecido o arquivo csv chamado “*imoveis.csv*”.
 - Identifique as colunas existente na tabela do banco de dados e no arquivo. Verifique se a chave é ou não auto incremento e realize a inserção de todos os registros do arquivo no banco de dados.
6. Alterar o notebook “coletaDadosMySQL_Python.ipynb”, utilizando a linguagem Python, para inserir registros na tabela *caracteristicageralimovel*. Para apoiar a execução desta atividade, foi fornecido o arquivo txt chamado “*caracteristicageralimovel.txt*”.
 - O arquivo *caracteristicageralimovel.txt* possui apenas 3 colunas: *idcaracteristicasGerais*, *idImovel* e *temCaracteristica*. A coluna *idCaracteristicasImovel* é definida como auto incrementável, e por isso vai gerar seu valor automaticamente no momento da inserção de cada registro.

Respostas Finais

Os alunos deverão desenvolver a prática e, depois, responder às questões objetivas.