

Bootcamp: Análise de Dados

Trabalho Prático

Módulo 3	CDD – Coleta e Obtenção de Dados
-----------------	---

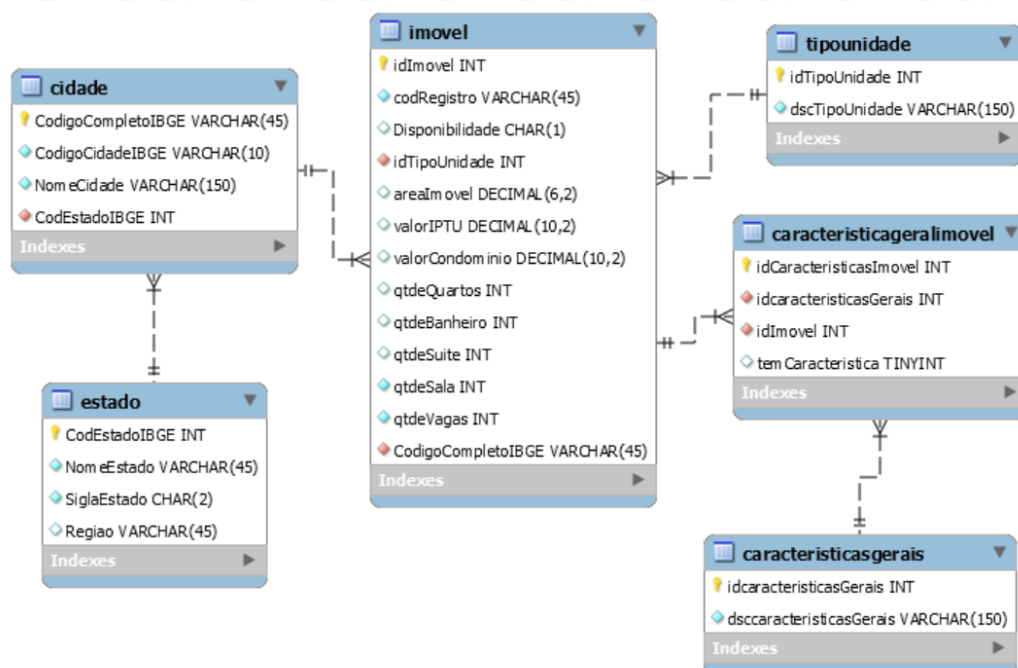
Objetivos

Exercitar os seguintes conceitos trabalhados no Módulo:

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

Enunciado

Para esta atividade, o aluno deverá assistir atentamente as seguintes aulas, disponíveis no ambiente virtual de aprendizagem, **exatamente na ordem indicada nas atividades**. Considere o esquema de banco de dados “imobiliária”, criado no SGBD MySQL, conforme tutorial e alterado nas videoaulas demonstrativas:



Considere ainda os arquivos em anexo com o trabalho prático:

- *bootcamp.mwb*
- *caracteristicasgerais.csv*
- *imóveis.txt*
- *caracteristicaimovel.txt*
- *caracteristicaimovel.csv*

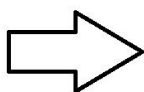
Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Deve-se assistir ao vídeo do tutorial MySQL Server e MySQL Workbench, anexo ao final do capítulo 2, e criar o banco de dados “bootcamp” com as tabelas conforme o modelo a seguir. Para isso, utilize o arquivo “*bootcamp.mwb*” disponível em conjunto com o trabalho, e siga os passos descritos no vídeo para realizar a engenharia reversa do modelo de dados e criar seu banco de dados. Durante a criação, salve o script gerado pois você vai precisar dele para responder as questões.
2. Deve-se assistir aos vídeos do capítulo 1, unidade 1.7, e seguir o tutorial de instalação do ambiente Anaconda (Aula 1.7.2) e validar seu ambiente Python no Anaconda replicando o exemplo “Aló Mundo!” da aula. Deve-se ainda configurar o ambiente Anaconda para trabalhar com a linguagem R e o R Studio (Aula 1.7.3) e validar seu ambiente replicando o exemplo “Aló Mundo!” da aula.
3. Deve-se assistir aos vídeos do capítulo 1, unidade 1.7, e seguir o tutorial de instalação da Plataforma Knime Analytics (Aulas 1.7.4 e 1.7.5).
4. O modelo abaixo é uma evolução do modelo anterior. Para alcançar este modelo, execute os passos a seguir:
 - a) Crie um script em linguagem R para inserir novos registros na tabela *caracteristicasgerais*, a imagem abaixo apresenta a tabela como deve ficar ao final da execução de seu script. Explore pelo menos 2 maneiras distintas para inserir os registros de id de 4 a 13. Para apoiar a execução das atividades, foi fornecido o arquivo csv chamado “*caracteristicasgerais.csv*”

De

idcaracteristicasGerais	dscaracteristicasGerais
1	Portaria 24 horas
2	Elevador
3	Piscina
NULL	NULL



Para

idcaracteristicasGerais	dscaracteristicasGerais
1	Portaria 24 horas
2	Elevador
3	Piscina
4	Salão de festas
5	Área gourmet
6	Água individual
7	Gás canalizado
8	Aquecimento solar
9	Vaga coberta
10	Vaga livre
11	Armários na cozinha
12	Closet
13	Armários no quarto

- b) Crie um script em linguagem Python para inserir registros na tabela *caracteristicasimovel*. Para isso, utilize os arquivos *caracteristicasimovel.txt* e *caracteristicasimovel.csv*. Os arquivos são diferentes, o primeiro vai inserir 104 linhas de dados e o segundo vai inserir 129 linhas. Lembre-se que para inserir dados na tabela *caracteristicasimovel* você deve ter inserido os dados na tabela imóvel e na tabela *caracteristicasgerais*.
- c) Agora utilizando a plataforma Knime crie um workflow para ler os dados do esquema bootcamp. Cada arquivo vai Explore pelo menos 2 maneiras distintas para inserir os registros de id de 4 a 13. Para apoiar a execução das atividades, foi fornecido o arquivo csv chamado “*caracteristicasgerais.csv*”

Respostas Finais

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