

Bootcamp IGTI: Cientista de Dados

Desafio

| | |
|-----------------|-----------------------------------|
| Módulo 2 | Coleta e Obtenção de Dados |
|-----------------|-----------------------------------|

Objetivos

Exercitar os seguintes conceitos trabalhados no Módulo:

- ✓ Coletar dados de API e armazenar em um banco de dados relacional (MySQL).
- ✓ Coletar e tratar dados de arquivos.
- ✓ Criar e manipular banco de dados não relacional (MongoDB).

Enunciado

Para realizar o desafio, utilize os arquivos que estão em anexo. São eles:

- “der.png”: Diagrama entidade relacionamento para a base de dados de pessoas.
- “videogame_sales_mongo”: Documentos para importar no MongoDB.
- “videogame_sales_readme”: Descrição dos atributos dos documentos utilizados no MongoDB.

Os dados das vendas de jogos de videogame foram extraídos do site <https://www.vgchartz.com/> através de um Web scraping utilizando o pacote BeautifulSoup em Python. O código está disponível publicamente através do gitHub: <https://github.com/GregorUT/vgchartzScrape>.

Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Para responder às questões de 1 a 6, siga as seguintes instruções:
 - a. Você deverá coletar todos os dados disponível no endpoint da seguinte API:
<https://api-coleta-dados.herokuapp.com/api/pessoas>.
 - b. Após a coleta, você deverá verificar quais pessoas estão com o campo “idade” nulos e deverá calcular a idade a partir da data de nascimento e preenchê-los.
 - c. Verificar também as pessoas que não possuem o ano de nascimento, apenas o dia e o mês. Você deverá calcular o ano de nascimento a partir da idade e preencher corretamente a data de nascimento.
 - d. Você deverá salvar todos os dados em um banco de dados MySQL, para isso, siga os passos a seguir.
 - e. Crie um banco de dados com o nome “db_desafio” e crie as tabelas conforme a imagem em anexo do diagrama de entidade e relacionamento “der.png”. Fique atento aos nomes das tabelas, nome dos atributos e os seus tipos de dados, siga o DER.
 - f. Após criar o banco, insira todos os registros. Fique atento, pois, os dados deverão ser inseridos conforme à estrutura do banco, logo você deverá povoar todas as tabelas do banco e ficar atento aos Ids atribuídos para manter o relacionamento entre as tabelas.
 - g. Exemplo: para o seguinte retorno, você deverá inserir a sigla do estado na tabela estado e se ela já tiver sido inserida não realizar a inserção. Inserir a cidade na tabela de cidade e o id do respectivo estado cadastrado na tabela estado. Inserir a cor na tabela de cor e se ela já tiver sido inserida não deverá ser inserida novamente, fazer o mesmo para o tipo sanguíneo. Após, inserir os dados da pessoa na tabela pessoa com os respectivos ID das tabelas relacionadas a ela.

```
{  
  "nome": "Thales Kaique",  
  "idade": 27,  
  "data_nasc": "10/12/1993",  
  "sexo": "Masculino",  
  "signo": "Sagitário",  
  "cidade": "Campo Grande",  
  "estado": "MS",  
  "altura": "1,63",  
  "peso": 110,  
  "tipo_sanguineo": "B+",  
  "cor": "laranja"  
},
```

- h. Após realizar todos os passos, reponda as questões de 1 a 6.
2. Para responder às questões de 7 a 15, siga as seguintes instruções:
- Crie uma coleção no mongoDB e importe o arquivo "videogame_sales_mongo".
 - Após a importação dos documentos para mongo, crie consultas que respondam as questões de 7 a 15.
 - Para ajudar nas consultas realizados com o mongo, acesse os sites:
<https://felipetoscano.com.br/consultar-documento-no-mongodb/> e o
<https://pymongo.readthedocs.io/en/stable/tutorial.html>.

Respostas Finais

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