
DATA ENGINEER

OBJETIVO

Este teste tem o objetivo de avaliar a capacidade do candidato em realizar tratamentos em bases de dados, como cruzamento, consolidação e ordenação. Fica a critério do candidato escolher a linguagem de programação de sua preferência para realização do teste (Python, Go, Java, C#, C++, C, Ruby, etc), sendo os critérios de avaliação a funcionalidade da solução, a clareza e facilidade de manutenção do código e a performance de execução, nesta ordem.

Para a entrega deverão ser enviadas as respostas das perguntas, bem como os arquivos e códigos utilizados para a resolução do mesmo.

BASES

Para este teste utilize o seguinte dataset <https://www.kaggle.com/sp1thas/book-depository-dataset>, considerando os arquivos listados a seguir, enviados junto a este teste:

- authors.csv
 - Contém o nome dos autores e o id que o representa no arquivo “dataset”;
- categories.csv
 - Contém as categorias e o id que a representa no arquivo “dataset”;
- formats.csv
 - Contém o formato em que os livros são disponibilizados (papel, para download, áudio, etc) e o id que o representa no arquivo “dataset”;
- dataset.csv
 - Contém a lista de livros e suas informações;

PREPARAÇÃO DOS DADOS

- Cruze os dados do arquivo “authors” com o “dataset”, de forma a incluir o nome dos autores de acordo com os ids dos autores para cada livro da base “dataset”;
- Cruze os dados do arquivo “categories” com o “dataset”, de forma a incluir o nome de todas as categorias a que o livro pertence de acordo com os ids das categorias da base “dataset”;
- Cruze os dados do arquivo “formats” com o “dataset”, de forma a incluir o nome do formato do livro de acordo com id do formato da base “dataset”;

PERGUNTAS

- Qual a quantidade total de livros da base?
- Qual a quantidade de livros que possui apenas 1 autor?
- Quais os 5 autores com a maior quantidade de livros?
- Qual a quantidade de livros por categoria?
- Quais as 5 categorias com a maior quantidade de livros?
- Qual o formato com a maior quantidade de livros?
- Considerando a coluna “bestsellers-rank”, quais os 10 livros mais bem posicionados?
- Considerando a coluna “rating-avg”, quais os 10 livros mais bem posicionados?
- Quantos livros possuem “rating-avg” maior do que 3,5?
- Quantos livros tem data de publicação (publication-date) maior do que 01-01-2020?