

DESAFIO - 03

ME315 - Manipulação de Banco de Dados

NOME: Amanda Rehder Cedro

RA: 253507

- **INTRODUÇÃO:**

Para a realização desse desafio foram utilizados dois bancos de dados retirados do site Kaggle, para isso foi aplicado um filtro em "**DATASETS**" e escolhidos os tipos de arquivos necessários. Após isso, foram selecionados bancos que seriam interessantes, um deles sendo no formato Parquet e o outro no formato JSON.

O banco de dados do formato **Parquet** é sobre droids do universo de Star Wars, especificando seu nome, suas características, sua função e o filme no qual aparece. Esse banco de dados pode ser acessado através do link: [\[banco de dados: Droids em Parquet\]](#)

O banco de dados do formato **JSON** é sobre a venda de produtos de uma boutique, especificando o tipo de produto, a marca, suas características, preços, data de compra, quantidade de peças no estoque, avaliação do cliente, se foi trocada e por qual motivo. Esse banco de dados pode ser acessado através do link: [\[banco de dados: Vendas da boutique em JSON\]](#).

- **PASSO 1:**

Instalando e carregando os pacotes que serão utilizados para podermos carregar os bancos de dados dentro do R. O pacote arrow será utilizado para carregar o banco de dados no formato parquet e o pacote "**jsonlite**" será utilizado para carregar o banco de dados no formato JSON.

```
```{r}
#baixando
install.packages("arrow")
install.packages("jsonlite")
```

```{r}
#carregando
library(arrow)
library(jsonlite)
```
```

- **PASSO 2:**

Importando e lendo o banco de dados no formato Parquet e o nomeando como um objeto "**droids_parquet**". Foi utilizada a função "**read_parquet**" do pacote "**arrow**" para poder importar e ler o banco como um data frame.

```
```{r}
#lendo o arquivo parquet
droids_parquet <- read_parquet("dados/droids.parquet")
```
```

- **PASSO 3:**

Importando e lendo o banco de dados no formato JSON e o nomeando como um objeto "**loja_json**".

Inicialmente foi utilizada a função "**read_json**" do pacote "**jsonlite**" para poder importar e ler o banco, porém, com essa função, o banco de dados não é lido da forma correta, já que sai em um formato de várias listas e teria que ser convertido depois para um data frame, o que gastaria mais tempo e, conseqüentemente, sairia mais caro.

Por esse motivo, foi utilizado a função "**fromJSON**", dessa forma o banco é lido e importado como um data frame.

```
```{r}
#lendo o arquivo JSON
loja_json <- fromJSON("dados/fashion_boutique_dataset.json")
```
```

- **CONSIDERAÇÕES SOBRE OS FORMATOS UTILIZADOS:**

Um arquivo **Parquet** é um formato de armazenamento de dados de código aberto, orientado por colunas, projetado para otimizar o armazenamento e a recuperação eficientes de grandes conjuntos de dados, como os encontrados em cenários de Big Data.

Diferente de formatos tradicionais baseados em linha, o Parquet organiza os dados em colunas, o que permite a leitura apenas das colunas necessárias para uma consulta, aumentando a velocidade da consulta e a eficiência do uso de recursos.

JSON é um acrônimo para JavaScript Object Notation, um formato leve e de fácil leitura para a organização e troca de dados. É amplamente usado em aplicações web e móveis para a comunicação entre servidores e clientes, sendo independente de linguagem de programação e de fácil compreensão para humanos e máquinas.

JSON organiza dados em pares de chave-valor e é uma alternativa mais eficiente ao XML para a transmissão de informações estruturadas.