

Desafio 3

AUTHOR
Henry Honda

Introdução

Neste documento, apresentamos como carregar e visualizar datasets nos formatos **JSON** e **Parquet** usando R.

1. Dados JSON

Os dados JSON contêm informações sobre **doenças cardíacas**, com 3.000 linhas de pacientes. Cada linha representa um paciente com variáveis como idade, sexo, pressão arterial, colesterol, etc.

Fonte: [Heart Disease Dataset no Kaggle](#)

```
#Instalando biblioteca que permite ler arquivos JSON.
library(jsonlite)

# Lendo o arquivo JSON.
json_arquivo <- fromJSON("Dados/heart_disease_dataset.json")

# Visualizando os dados.
head(json_arquivo)
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
1	67	1	2	111	536	0	2	88	0	1.3	3	2	3
2	57	1	3	109	107	0	2	119	0	5.4	2	0	3
3	43	1	4	171	508	0	1	113	0	3.7	3	0	7
4	71	0	4	90	523	0	2	152	0	4.7	2	1	3
5	36	1	2	119	131	0	2	128	0	5.9	3	1	3
6	49	1	1	186	571	0	0	176	0	4.0	3	0	3
smoking diabetes bmi heart_disease													
1		1		0	23.4		0						
2		0		1	35.4		0						
3		1		1	29.9		0						
4		1		0	15.2		1						
5		1		0	16.7		1						
6		1		0	33.8		0						

2. Dados Parquet

O Parquet é um formato binário eficiente para **armazenamento e leitura de grandes volumes de dados**, com compressão e otimização para análise em colunas.

O exemplo abaixo usa dados sobre os **sobreviventes do Titanic**.

fonte: <https://www.timestored.com/data/sample/parquet>

```
# Instalando e carregando o pacote arrow, como foi desenvolvido pela APACHE devemos instalar bibl

install.packages("arrow", repos = c("https://apache.r-universe.dev", "https://cloud.r-project.org/"))
```

Instalando pacote em 'C:/Users/rseit/AppData/Local/R/win-library/4.4'
(como 'lib' não foi especificado)

pacote 'arrow' desempacotado com sucesso e somas MD5 verificadas

Warning: não foi possível remover a instalação prévia do pacote 'arrow'

```
Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
C:\Users\rseit\AppData\Local\R\win-library\4.4\00LOCK\arrow\libs\x64\arrow.dll
to C:\Users\rseit\AppData\Local\R\win-library\4.4\arrow\libs\x64\arrow.dll:
Permission denied
```

Warning: 'arrow' restaurado

Os pacotes binários baixados estão em
C:\Users\rseit\AppData\Local\Temp\RtmpUzG15n\downloaded_packages

```
library(arrow)
```

Warning: pacote 'arrow' foi compilado no R versão 4.4.3

```
Anexando pacote: 'arrow'
```

O seguinte objeto é mascarado por 'package:utils':

timestamp

```
# Lendo o arquivo Parquet
parquet_dados <- read_parquet("Dados/titanic.parquet")

# Visualizando os dados
head(parquet_dados)
```

PassengerId	Survived	Pclass
1	1	0
2	2	1
3	3	1
4	4	1
5	5	0
6	6	0

Name	Sex	Age	SibSp	Parch
Braund, Mr. Owen Harris	male	22	1	0
Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0
Heikkinen, Miss. Laina	female	26	0	0
Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0
Allen, Mr. William Henry	male	35	0	0
Moran, Mr. James	male	NA	0	0

Ticket	Fare	Cabin	Embarked
A/5 21171	7.2500	<NA>	S
PC 17599	71.2833	C85	C
STON/O2. 3101282	7.9250	<NA>	S
113803	53.1000	C123	S
373450	8.0500	<NA>	S
330877	8.4583	<NA>	Q