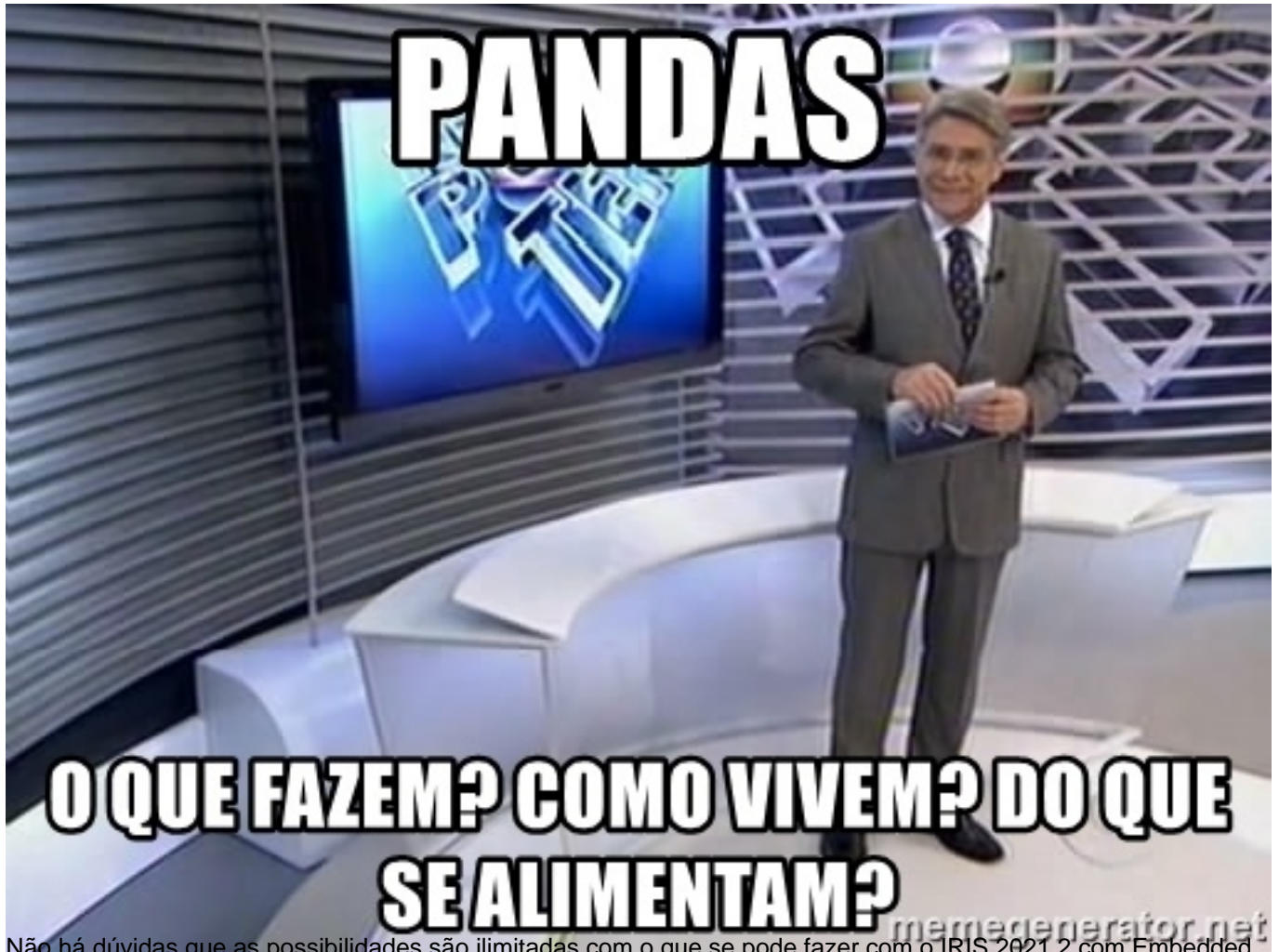


Artigo

[Henry Pereira](#) · Abr. 4, 2022 3min de leitura

## Introdução à análise de dados com IRIS e Pandas



Não há dúvidas que as possibilidades são ilimitadas com o que se pode fazer com o IRIS 2021.2 com Embedded Python, desde visão computacional, automações, blockchain e principalmente a inteligência artificial.

Python é umas das linguagens que mais tem crescido nos últimos anos e existe uma infinidade de bibliotecas para facilitar a vida dos desenvolvedores.

Meu objetivo é escrever de forma simples, uma introdução ao Pandas para desenvolvedores de ObjectScript.

### Mas o que é Pandas?

Pandas é uma biblioteca criada para análise, manipulação e tratamento de dados.

A título de curiosidade, Pandas vem de "panel data" (dados em painel), um termo usado em estatística e econometria para conjunto de dados que incluem várias unidades amostrais (indivíduos, empresas, etc) acompanhadas ao longo do tempo. (fonte: [Wikipedia](#))

No IRIS 2021.2 para utilizar o Pandas, primeiro é necessário importar.

```
import pandas as pd
```

Existem dois tipos principais de estruturas de dados no Pandas, que são Series e DataFrames.

Uma Serie nada mais é que um vetor, um array unidimensional.

```
selecaoPenta = pd.Series(["Lúcio", "Edmílson", "Roque Junior", "Gilberto Silva", "Marcos", "Kaká", "Vampeta", "Anderson Polga", "Dida", "Rogério Ceni", "Belletti", "Ronaldo", "Ronaldo", "Roberto Carlos", "Kléberson", "Rivaldo", "Cafu", "Júnior", "Ricar-dinho", "Luizão", "Edílson", "Denílson", "Juninho Paulista"])
```

Já o DataFrame é uma estrutura de dados bidimensional. Um DataFrame é constituído de Series.

Vamos criar uma variável chamada artilheiros, com conjuntos de chaves e valores, com nome e quantidades de gols

```
artilheiros = {'Jogador': ['Miroslav Klose', 'Ronaldo', 'Gerd Muller', 'Just Fontaine', 'Pele', 'Sandor Kocsis'], 'Gols': [16, 15, 14, 13, 12, 11]} df = pd.DataFrame(artilheiros)
```

O método info() podemos obter um resumo do DataFrame

Com o método head() é possível visualizar as 5 primeiras linhas de um DataFrame e com tail() as 5 últimas. Podemos utilizar sortvalues() para ordenar um DataFrame

No Pandas podemos utilizar condicionais como índice para filtrar dados de maneira simples.

Com isnull() filtra apenas os que são nulos e notnull() é o oposto, apenas os que não são nulos.

O mais incrível é que é possível executar Queries SQL no IRIS e retornar diretamente em um DataFrame.

```
import iris
import pandas as pd
rs = iris.sql.exec("select Name, Super, TimeCreated from %Dictionary.ClassDefinition WHERE Name %STARTSWITH %Net.")
mydataframe = rs.dataframe()
```

É possível importar e exportar dados nos mais diversos formatos, podemos exportar um DataFrame para csv, xml, json ou html simplesmente executando to\_csv(), to\_xml(), to\_json() ou to\_html()

Para saber mais do Pandas aqui está a [documentação oficial](#)

Como sempre, se tiver dúvidas, comentários ou sugestões não hesite em me escrever! Obrigado por ler.

[#Embedded Python](#) [#InterSystems IRIS](#)

---

URL de origem: <https://pt.community.intersystems.com/post/introdu%C3%A7%C3%A3o-%C3%A0-an%C3%A1lise-de-dados-com-iris-e-pandas>

---