

# Comandos úteis em Pandas

<https://s3-us-west-2.amazonaws.com/secure.notion-static.com/eded06cd-02ee-40eb-8324-2a72efba5747/PandasPythonForDataScience.pdf>

Pandas é a minha biblioteca favorita do Python. Seja para Data Visualization ou para Data Analysis, a praticidade e funcionalidade que essa ferramenta oferece não é encontrada em nenhum outro módulo.

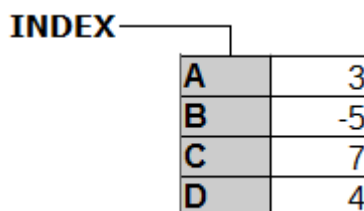
## Importando Pandas

```
import pandas as pd
```

## Series (Coluna de uma tabela)

Series nada mais é que um array de 1 dimensão. Você pode considerar um Series também como uma coluna de uma tabela.

```
s = pd.Series([3, -5, 7, 4],  
index=['a', 'b', 'c', 'd'])
```



INDEX	
A	3
B	-5
C	7
D	4

## DataFrame

Um DataFrame é simplesmente um conjunto de series (colunas). Trata-se de uma estrutura de dados de 2 dimensões – colunas e linhas – que transforma os dados em uma bela tabela.

Cada chave do dicionário representa uma coluna

```
DataFrame >>> data = {  
    'País': ['Bélgica', 'Índia', 'Brasil'],  
    'Capital': ['Bruxelas', 'Nova Delhi', 'Brasília'],  
    'População': [123465, 456789, 987654]  
}
```

## Criando um DataFrame:

```
df = pd.DataFrame(data, columns=['País', 'Capital', 'População'])
```

## Abrindo e escrevendo arquivos CSV:

Para ler arquivos CSV codificados em ISO

```
pd.read_csv('nome_do_arquivo.csv', encoding='ISO-8859-1')
```

## Salvando arquivos CSV:

```
pd.to_csv('nome_do_arquivo_para_salvar.csv')
```

## Abrindo arquivos do Excel:

```
xlsx = pd.ExcelFile('arquivo_excel.xlsx')  
  
df = pd.read_excel(xlsx, 'Planilha_1')
```

## Removendo linhas e colunas:

Removendo linhas pelo index

```
s.drop([0, 1])
```

Removendo coluna completa utilizando o argumento axis=1

```
df.drop('País', axis=1)
```

## Coletando informações básicas do DataFrame

Quantidade de linhas e colunas do DataFrame

```
df.shape
```

Descrição do index

```
df.index
```

Colunas do DataFrame

```
df.columns
```

Contagem dados não nulos

```
df.count()
```

## Criando coluna em um DataFrame:

```
df['Nova Coluna'] = 0
```

## Renomeando colunas do DataFrame:

Se o DataFrame possui 3 colunas, passar uma lista com os 3 novos nomes.

```
df.columns = ['Coluna 1', 'Coluna 2', 'Coluna 3']
```

## Resumo dos dados

Resumo estatístico

```
df.describe()
```

Index de menor valor

```
df.idmin()
```

Index de maior valor

```
df.idmax()
```

## Aplicando funções:

Aplicando uma função que substitui A : B

```
df.apply(lambda x: x.replace('a', 'b'))
```

## Ordenando valores:

Ordenando em ordem crescente

```
df.sort_values
```

Ordenando em ordem decrescente

```
df.sort_values(ascending=False)
```

## Operações aritméticas em Series:

```
s = pd.Series([1, 2, 3, 4, 5],  
index=['a', 'b', 'c', 'd', 'e'])
```

Somando todos os valores presentes na Series por 2

```
s.add(2)
```

Subtraindo todos os valores presentes na Series por 2

```
s.sub(2)
```

Multiplicando todos os valores da Series por 2

```
s.mul(2)
```

Dividindo todos os valores da Series por 2

```
s.div(2)
```

## Indexação por Boolean:

Filtrando DataFrame para mostrar apenas valores pares

```
df[df['População'] % 2 == 0]
```

## Selecionando valores

Selecionando primeira linha da coluna país

```
df.loc[0, 'País']
```