

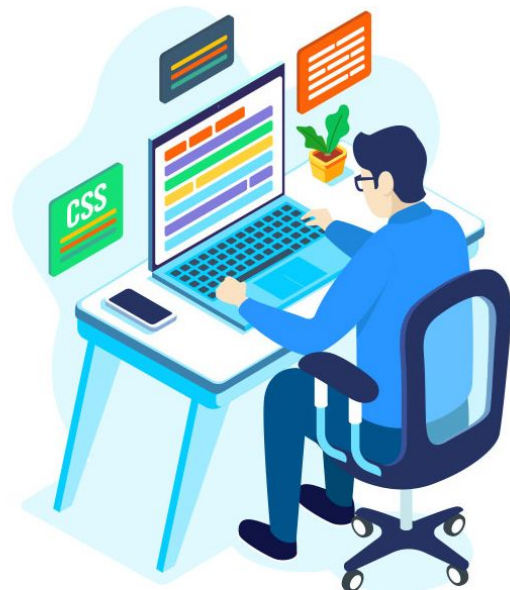


Uma Biblioteca Python para Análise de Dados

Prof. José Ferdinandy Silva Chagas

Vamos aprender..

- Representação de Dados
- Criando uma tabela de dados
- Estatísticas dos Dados
- Subconjuntos de Dados
- Agrupamentos de Dados
- Construção de Gráficos
- Exportando dados para arquivos
- Importando dados de arquivos
- Introdução a EDA



Representação de Dados

Representação dos Dados

23

Representação dos Dados



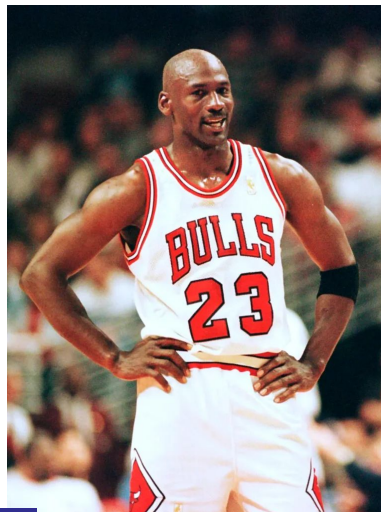
23 de Maio

Avenida 23 de Maio

Bela Vista

CEP 01316-060

Sé
12



MELHORES
ZAGUEIROS
DO FIFA 23



Representação de Dados

- DataFrame

Diagram illustrating the structure of a DataFrame, showing a grid of data organized into rows and columns. The first column is highlighted in dark blue, and the second row is highlighted with a red border. The text "linha (row)" is written in red next to the second row.

linha (row)				

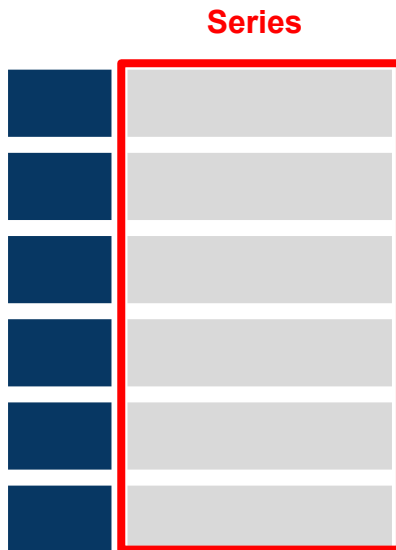
Representação de Dados

- DataFrame

coluna (column)

Representação de Dados

- Series



Manipulando DataFrames

Criando um DataFrame

```
dados = pd.DataFrame({
    "Name": [
        "Pedro Henrique",
        "Alan Malagueta",
        "Maria Eunalia",
    ],
    "Age": [22, 35, 58],
    "Sex": ["male", "male", "female"],
})
```

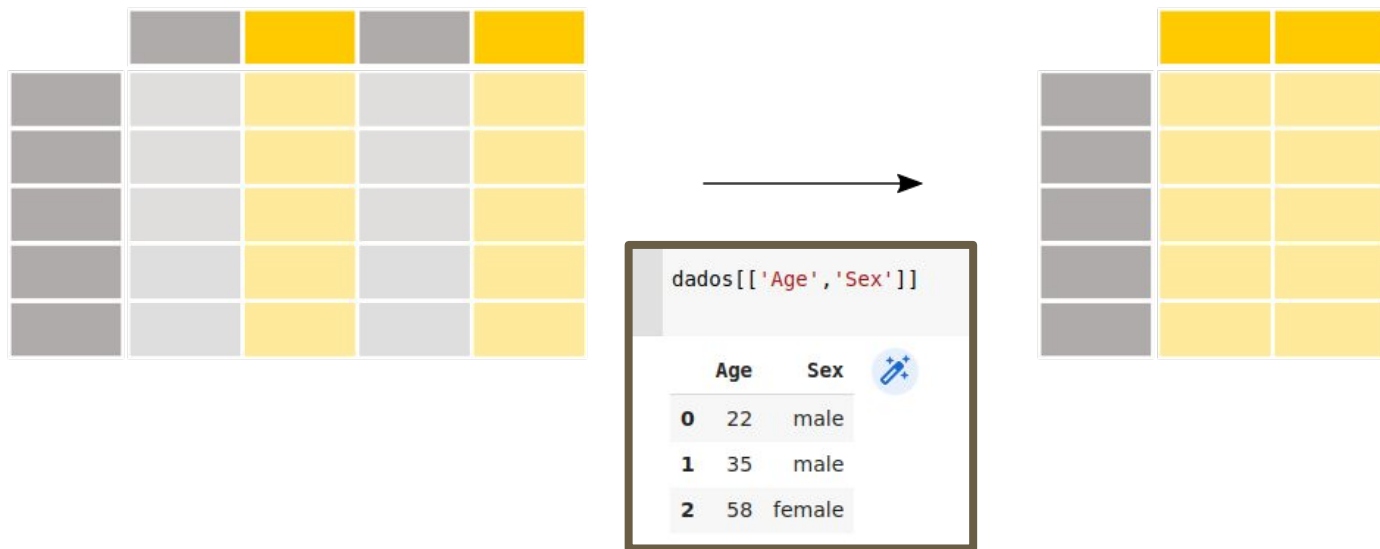
	Name	Age	Sex
0	Pedro Henrique	22	male
1	Alan Malagueta	35	male
2	Maria Eunalia	58	female

```
# ## Series
dados['Age']

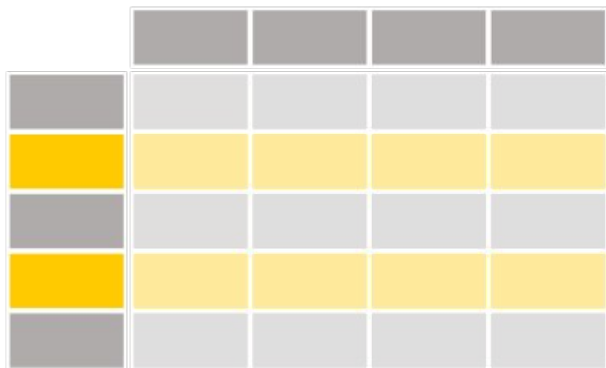
# ## Maior valor
dados['Age'].max()

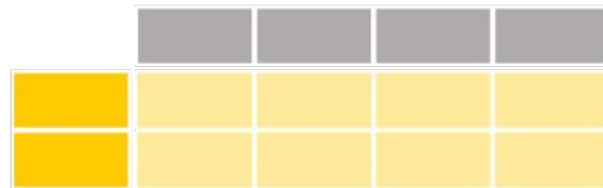
# ## Atribuindo a uma variável
ages = dados['Age']
```

Seleccionando Subconjuntos de Datos - Columnas



Selecionando Subconjunto de Dados - Linhas

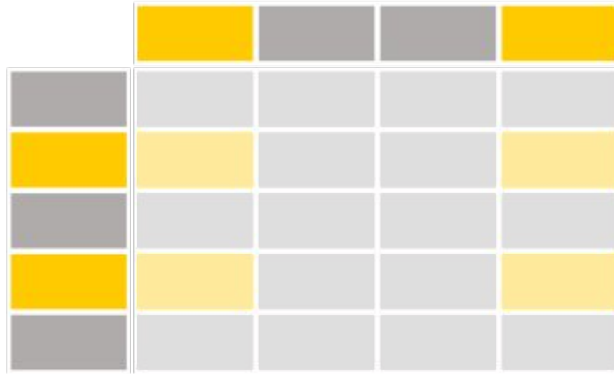





```
## Subconjunto de dados (Linhas)  
dados[dados['Sex']=='male']
```

	Name	Age	Sex
0	Pedro Henrique	22	male
1	Alan Malagueta	35	male

Seleccionando Subconjunto de Datos - Subset



	Yellow	Gray	Gray	Yellow
Gray				
Yellow	Yellow			Yellow
Gray				
Yellow	Yellow			Yellow
Gray				

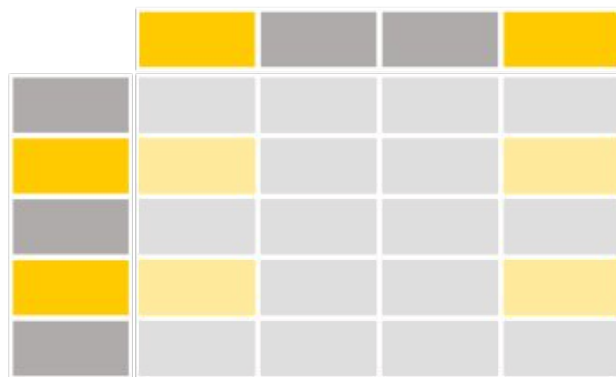


	Yellow	Yellow
Yellow	Yellow	Yellow
Yellow	Yellow	Yellow

```
dados.loc[dados['Age']<50, 'Name']
```

```
0    Pedro Henrique  
1    Alan Malagueta  
Name: Name, dtype: object
```

Seleccionando Subconjunto de Datos - Subset

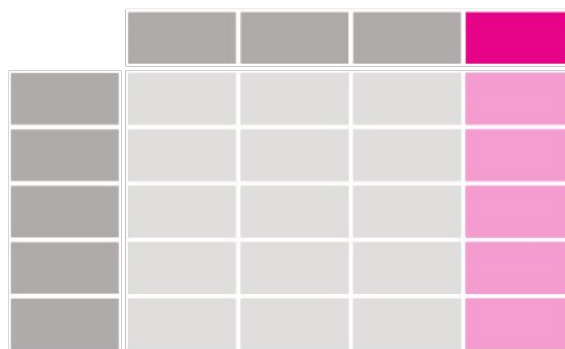
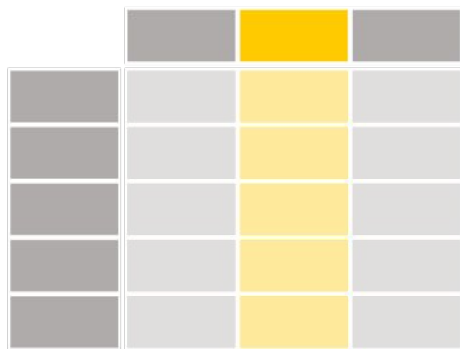


```
dados.iloc[0:2, 1:3]
```

	Age	Sex
0	22	male
1	35	male

Gerando Novos Dados na Tabela

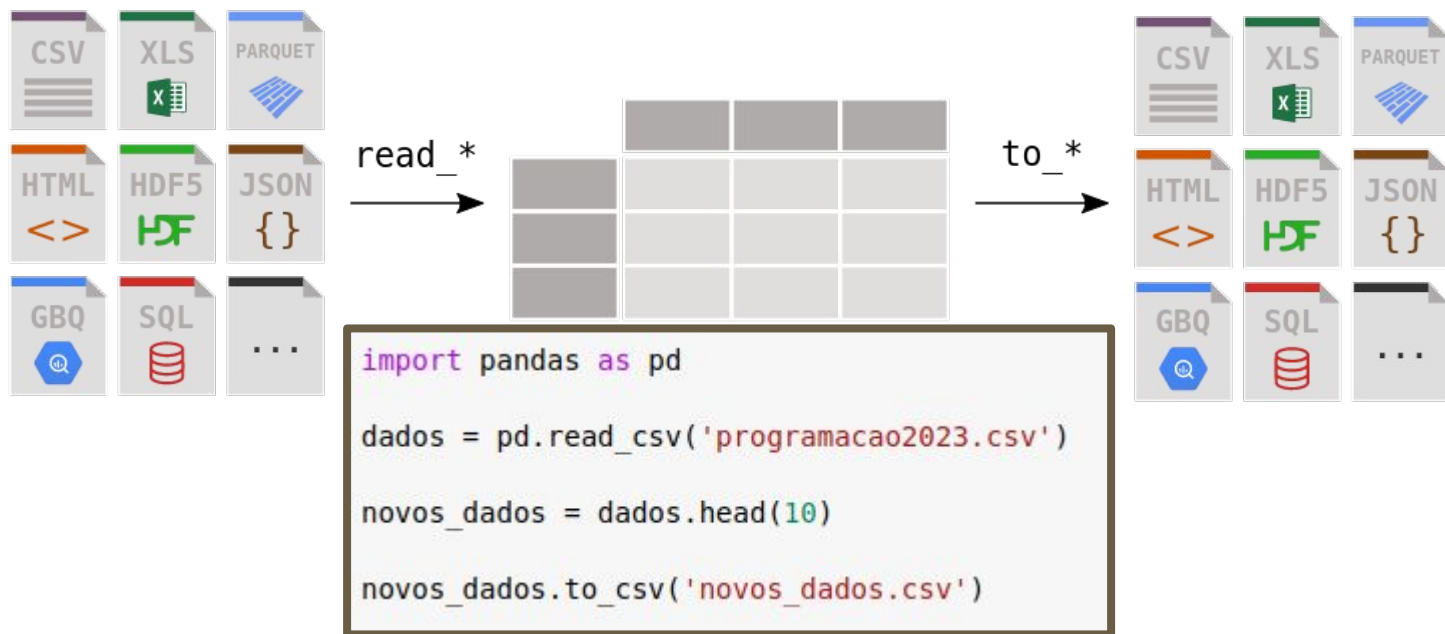
Gerando Novos Dados na Tabela



```
dados['maior_idade'] = dados['Age'] > 18
```

	Name	Age	Sex	maior_idade
0	Pedro Henrique	22	male	True
1	Alan Malagueta	35	male	True
2	Maria Eunalia	58	female	True

Gerando DataFrames de Arquivos



Referências

- https://pandas.pydata.org/docs/getting_started/index.html
- <https://mlcontests.com/state-of-competitive-machine-learning-2022/>