

# **Análise, construção e visualização de dados**

Prof. Dr. Álvaro Campos  
Ferreira

[alvaro.ferreira@idp.edu.br](mailto:alvaro.ferreira@idp.edu.br)

# Aulas de Reforço e o PythonLab

# Aulas de reforço

Se tivermos quórum, vou organizar aulas de reforço a partir da semana que vem durante a tarde.



# PythonLab

Se tivermos  
quórum, o grupo  
PythonLab irá ser  
realizado toda  
semana quarta à  
tarde.



<https://forms.gle/XCQUWKZrDZmeHi8U6>

# Gerando QR Codes para links

```
import qrcode  
data = 'https://forms.gle/iruMGtr419EDhAQ5G6'  
img = qrcode.make(data)  
img.save('meu_codigo_qr.png')
```

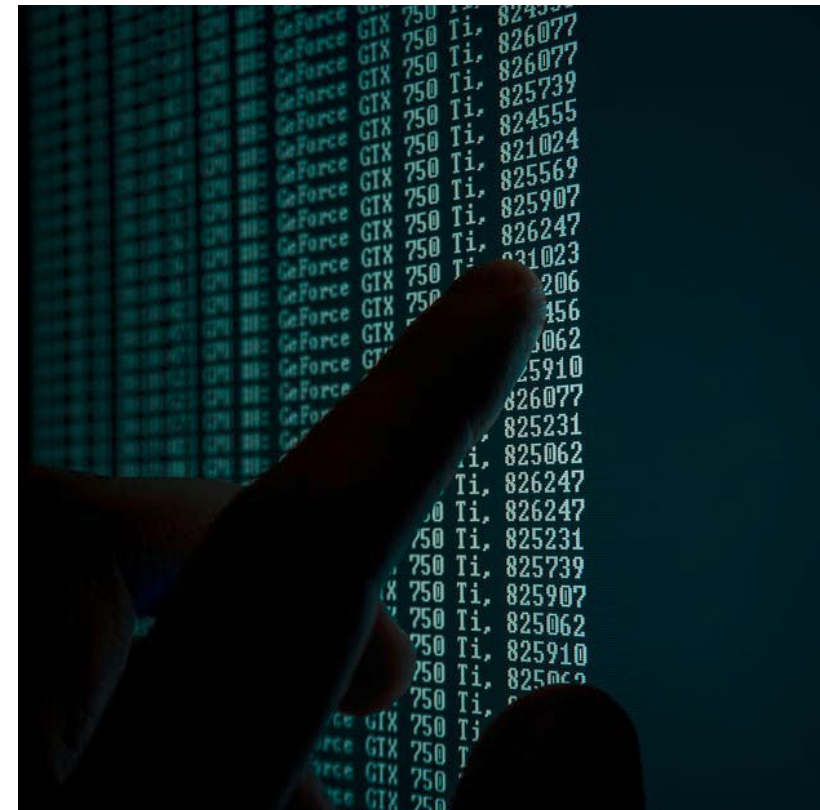


# Revisão

# Pandas e DataFrames

# O que são dados?

Fatos e números  
coletados, analisados e  
sintetizados para a  
apresentação e  
interpretação.



# Tipos de variáveis

Para cada tipo de variável, as funções tem comportamento diferente.

- Texto (str)
- Número inteiro (int)
- Número real (float)
- Booleano (bool), True ou False



# Análise exploratória

Etapas da estatística descritiva.



# Análise descritiva dos dados

Uma descrição óbvia de um conjunto de dados está nos próprios dados.

- Número de observações
- Valores mínimos e máximos
- Ordenação dos valores

# Análise descritiva dos dados

Quando temos muitos dados para analisar cada observação, precisamos de novas ferramentas.

- Tendências centrais
- Dispersão
- Correlação

# Análise descritiva dos dados

## Tendências centrais

- Média
- Mediana

# Análise descritiva dos dados

Dispersão

- Amplitude
- Variância
- Desvio-padrão

# Análise descritiva dos dados

## Correlação

- Covariância
- Correlação

A correlação varia entre -1 (anticorrelação perfeita) e 1 (correlação perfeita)

# Indexação de DataFrames

Para selecionar uma coluna de um DataFrame, a sintaxe é a mesma de um dicionário, com o nome da coluna no lugar da chave:

```
df["Name"]
```

Para selecionar elementos, use a indexação de listas:

```
df["Name"][0]
```

# Indexação de DataFrames

Slices ou fatias são seleções de mais de um elemento de um objeto ao mesmo tempo.

A faixa de índices é indicada entre o sinal “:”.

Ou seja, para selecionar os dez primeiros elementos, usa-se:

```
df[“Name”][0:10]
```



# Indexação de DataFrames

Outras duas formas de se selecionar elementos de DataFrames é através das funções `loc` e `iloc`:

- Essas funções permitem acessar os elementos pelos índices das linhas e colunas diretamente

# Operadores

# Operadores de comparação

<code>==</code>	Igual	<code>x == y</code>
<code>!=</code>	Diferente	<code>x != y</code>
<code>&gt;</code>	Maior que	<code>x &gt; y</code>
<code>&lt;</code>	Menor que	<code>x &lt; y</code>
<code>&gt;=</code>	Maior ou igual	<code>x &gt;= y</code>
<code>&lt;=</code>	Menor ou igual	<code>x &lt;= y</code>

# Operadores lógicos

**and** Verdadeiro se ambas as condições forem verdadeiras

$x < 5$  and  $x < 10$

**or** Verdadeiro se alguma das condições for verdadeira

$x < 5$  or  $x < 4$

# Operadores lógicos

**not** Inverte o resultado

`not(x < 5 and x < 10)`

# Operadores de identidade e pertencimento

**in** Verdadeiro se o valor está presente no objeto

$x \text{ in } y$

**not in** Verdadeiro se o valor não está presente no objeto

$x \text{ not in } y$

# Selecionando elementos através de operações

# Indexação de DataFrames

Para acessar um elemento ou mais a partir de uma operação, pode-se utilizar a seguinte sintaxe:

```
df[df["Nome"] == "Xerxes"]
```

Isso vai retornar todas as linhas em que o valor da coluna "Nome" seja "Xerxes".





INSTITUTO BRASILEIRO DE ENSINO,  
DESENVOLVIMENTO E PESQUISA