

Análise, construção e visualização de dados

Prof. Dr. Álvaro Campos
Ferreira

alvaro.ferreira@idp.edu.br

Operadores

Operadores de comparação

<code>==</code>	Igual	<code>x == y</code>
<code>!=</code>	Diferente	<code>x != y</code>
<code>></code>	Maior que	<code>x > y</code>
<code><</code>	Menor que	<code>x < y</code>
<code>>=</code>	Maior ou igual	<code>x >= y</code>
<code><=</code>	Menor ou igual	<code>x <= 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

is Verdadeiro se forem o mesmo objeto

$x \text{ is } y$

is not Verdadeiro se não forem o mesmo objeto

$x \text{ is not } y$

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

$x \text{ not in } y$

Pandas e DataFrames

Pandas e Dataframes

Vamos usar o Pandas para acessar os nossos dados sobre Pokemons.

```
import pandas as pd  
file = "pokemon_data.csv"  
df = pd.read_csv(file)  
print(df)
```

Análise exploratória

**Quais perguntas
responder?**

Quais perguntas responder?

A partir de agora, a análise vai acontecer através de uma série de perguntas cujas respostas podem ser obtidas através da análise e interpretação dos dados.

Definir bem as perguntas é útil para guiar a análise exploratória.

Quais perguntas responder?

Algumas perguntas que podemos responder de nosso conjunto de dados sobre Pokemons:

- Quantos pokemons estão no conjunto?
- Quantos pokemons únicos?
- Quantos de cada tipo?
- Quantos por geração?

Quais perguntas responder?

Outras perguntas:

- Qual é o Pokemon com o maior ataque?
- Qual é o com a maior defesa?

Indexação de DataFrames

O DataFrames é organizado em colunas que podem ser acessadas a partir de seu nome.

Para determinar os nomes das colunas, pode-se escrever na tela os primeiros valores com a função `head()`

```
df.head()
```

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]
```

Funções de DataFrames

Funções de DataFrames

DataFrames possuem diversas funções para facilitar a análise exploratória e de estatística descritiva. Algumas das funções que vamos utilizar são:

- `describe()`
- `plot()`



INSTITUTO BRASILEIRO DE ENSINO,
DESENVOLVIMENTO E PESQUISA