

Análise, construção e visualização de dados

Prof. Dr. Álvaro Campos
Ferreira

alvaro.ferreira@idp.edu.br

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

Análise exploratória

Obtendo um sumário descritivo estatístico de um DataFrame inteiro e de colunas selecionadas.

Usando a função `describe()`

É possível utilizar as funções de estatística descritiva diretamente no DataFrame.

Análise exploratória

O Pandas possui funções para facilitar a descrição de dados.

```
df.describe()
```

```
df.plot()
```

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]
```

Onde encontrar dados?

Onde encontrar dados?

- 1) Google Data Set Search
- 2) OpenDataSUS
- 3) Kaggle
- 4) Webscrapping



INSTITUTO BRASILEIRO DE ENSINO,
DESENVOLVIMENTO E PESQUISA