



Importação de Diferentes Tipos de Dados

O pandas disponibiliza um conjunto bastante amplo de funções que possibilitam a obtenção de informações a partir de diversos formatos de arquivos. Estas funções importam dados tabulares a partir de arquivos em formatos diversos (csv, xlsx, txt, json, etc) para dentro das estruturas de dados do pandas (Series e DataFrames).

Lendo um arquivo json

```
json = open('../dados/aluguel.json')
print(json.read())
```

Lendo um arquivo json com o pandas

```
df_json = pd.read_json('../dados/aluguel.json')
# mostra as 10 primeiras observações
# o pandas organiza as colunas de forma alfabética
```

Lendo um arquivo txt

```
txt = open('../dados/aluguel.txt')
print(txt.read())
```

Lendo um arquivo do excel xlsx com o pandas

```
# arquivos xlsx não possuem uma leitura nativa como os outros tipos
df_xlsx = pd.read_excel('../dados/aluguel.xlsx')
df_xlsx
```

- foi necessário a instalação do módulo **openpyxl**

```
!pip3 install openpyxl
```

Lendo um Data Frame em um endereço html

```
# pt.wikipedia.org/wiki/Lista_das_maiores_empresas_do_Brasil_por_faturamento
df_html = pd.read_html('https://pt.wikipedia.org/wiki/Lista_das_maiores_empresas_do_Brasil_por_faturamento')
len(df_html) # mostra o nº de df presente no endereço
df_html[1]
df_html[2]
```

Resumo

- Como ler fontes de dados diferentes:
 - Uma base CSV (`pd.read_csv(...)`)
 - Uma base JSON (`pd.read_json(...)`)
 - Uma base TXT (`pd.read_table(...)`)
 - Um arquivo EXCEL (`pd.read_excel(...)`)
 - Uma página HTML (`pd.read_html(...)`)
- Vários métodos e atributos úteis de *dataframes*, como:
 - `info()`
 - `head()`
 - `dtypes`
 - `columns`
 - `shape`