Carregando e armazenando dados em vários formatos de arquivos

Trabalhando com JSON

Para trabalhar com JSON temos uma biblioteca padrão chamada json. Vamos iniciar importando pandas, numpy e json.

```
import pandas as pd
import numpy as np
import json
```

Trabalhando com JSON

Vamos agora criar uma string contendo um conteúdo em

formato json.

```
string_json = """
  "marca":"Renault",
  "veiculos":
    {"modelo": "Sandero", "lancamento": "2007"},
    {"modelo": "Kwid", "lancamento": "2017"},
    {"modelo": "Duster", "lancamento": "2011"},
    {"modelo": "Captur", "lancamento": "2017"},
    {"modelo": "Logan", "lancamento": "2007"}
111111
```

Trabalhando com JSON

Nós podemos criar um objeto Python (dicionário) através do método loads da biblioteca json informando a string JSON.



Trabalhando com JSON

Podemos também converter um dicionário Python em JSON usando o método dumps da biblioteca json.

```
In [5]: meu_json = json.dumps(obj_python)
meu_json

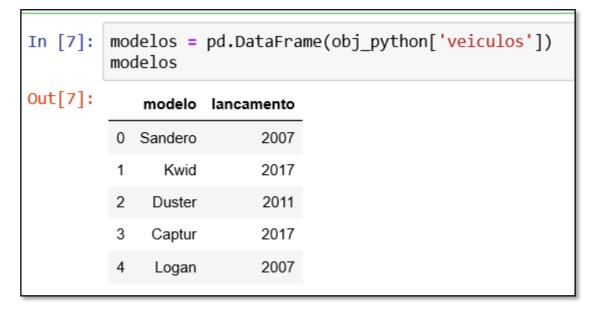
Out[5]: '{"marca": "Renault", "veiculos": [{"modelo": "Sandero", "lancamento": "2007"}, {"modelo": "Kwid", "lancamento": "2017"}, {"modelo": "Duster", "lancamento": "2011"}, {"modelo": "Captur", "lancamento": "2017"}, {"modelo": "Logan", "lancamento": "2007"}]}'

In [6]: type(meu_json)
Out[6]: str
```



Trabalhando com JSON

Após ter carregado o JSON para um dicionário, podemos criar um DataFrame à partir deste dicionário.





Trabalhando com JSON

Podemos carregar um arquivo JSON diretamente para um DataFrame ou Series usando o método read json.





Trabalhando com JSON

Se não for informado o tipo, por padrão o objeto será um DataFrame.

```
In [10]: type(dados)
Out[10]: pandas.core.frame.DataFrame
```

Trabalhando com JSON

Se quisermos carregar para um objeto Series, teremos que especificar.

Trabalhando com JSON

Se cada linha do arquivo for um JSON diferente, temos que informar o parâmetro lines como verdadeiro.

```
Veja o conteúdo do arquivo exemplo2.json: {"modelo": "Sandero", "lancamento": "2007"} {"modelo": "Kwid", "lancamento": "2017"} {"modelo": "Duster", "lancamento": "2011"} {"modelo": "Captur", "lancamento": "2017"} {"modelo": "Logan", "lancamento": "2007"}
```



```
dados = pd.read json('exemplo2.json', lines=True)
In [13]:
In [14]:
          dados
Out[14]:
              modelo
                     lancamento
           0 Sandero
                           2007
                Kwid
                           2017
                           2011
               Duster
           3
               Captur
                           2017
               Logan
                           2007
```



Trabalhando com JSON

Para exportar os dados de um objeto Series ou DataFrame, podemos usar o método to_json() destes objetos.

Primeiro vamos criar um dicionário com a população dos estados brasileiros.

```
população estados = {'São Paulo':45538936,
'Minas Gerais':21040662, 'Rio de Janeiro':17159960, 'Bahia':14812617,
'Parana':11348937.
'Rio Grande do Sul':11329605, 'Pernambuco':9496294,
'Ceara':9075649, 'Para':8513497, 'Santa Catarina':7075494,
'Maranhao':7035055, 'Goias':6921161, 'Amazonas':4080611, 'Paraiba':3996496, 'Espirito Santo':3972388,
'Rio Grande do Norte':3479010, 'Mato Grosso':3441998,
'Alagoas':3322820, 'Piaui':3264531, 'Distrito Federal':2974703,
'Mato Grosso do Sul':2748023, 'Sergipe':2278308, 'Rondonia':1757589,
'Tocantins':1555229, 'Acre':869265, 'Amapa':829494,
'Roraima':576568}
```



Trabalhando com JSON

Agora vamos criar uma série com este dicionário.

In [28]:	<pre>populacao = pd.Series(populacao_estados) populacao</pre>	
Out[28]:	Sao Paulo	45538936
	Minas Gerais	21040662
	Rio de Janeiro	17159960
	Bahia	14812617
	Parana	11348937
	Rio Grande do Sul	11329605
	Pernambuco	9496294
	Ceara	9075649
	Para	8513497
	Santa Catarina	7075494



Trabalhando com JSON

Agora vamos exportar a série para JSON.

Trabalhando com JSON

Vamos agora criar um DataFrame com o dicionário e exportar os dados para JSON com o método to_json().

```
populacao_estados = {'estado': ['Minas Gerais', 'Espirito Santo', 'Rio de Janeiro', 'Sao Paulo'], 'populacao_2000': [17891494, 3097232, 14391282, 37032403], 'populacao_2010': [19595309, 3512672, 15993583, 41252160]}
```



```
In [38]:
           populacao = pd.DataFrame(populacao estados)
           populacao
Out[38]:
                    estado populacao_2000 populacao_2010
               Minas Gerais
                                                 19595309
                                 17891494
                                  3097232
                                                  3512672
               Espirito Santo
           2 Rio de Janeiro
                                 14391282
                                                 15993583
                 Sao Paulo
                                                 41252160
           3
                                 37032403
```



FIM

