

2

Lab02 (fonte de dados)

- Os exemplos de código não estão diretamente relacionadas à construção de modelos de machine learning, mas essas etapas de aquisição e pré-processamento de dados são fundamentais. Antes de aplicar técnicas de machine learning, é necessário ter acesso aos dados relevantes e prepará-los de maneira adequada.
- Pandas é uma biblioteca muito utilizada em Python para manipulação e análise de dados. Ela fornece estruturas de dados flexíveis e eficientes, como DataFrame, que facilitam a organização e manipulação de conjuntos de dados

```
# Importação de Bibliotecas
import pandas as pd
```

- O arquivo CSV contendo dados sobre Pesquisa de Preços de Carros é lido e carregado em um DataFrame do Pandas chamado `df`. O parâmetro `sep=","` especifica que o separador no arquivo CSV é uma vírgula. Esta é uma forma comum de carregar dados de um arquivo local para análise

```
# Fonte de Dados - Opção 1: Arquivo CSV
# Dados sobre Pesquisa de Preços de Carros
```

```
df = pd.read_csv("C:/Users/Walmir/Documents/_WORK/US/_IAML/2023,
```

- A biblioteca `requests` é utilizada para fazer uma requisição HTTP para uma API do Banco Central do Brasil que fornece dados sobre inflação. Os dados são então convertidos para o formato JSON usando o método `.json()`.

```
# Fonte de Dados - Opção 2: API
# Dados sobre inflação do BACEN (Banco Central do Brasil)

import requests

url = 'https://api.bcb.gov.br/dados/serie/bcdata.sgs.433/dados?'
response = requests.get(url).json()
response
```

```
[{'data': '01/01/1980', 'valor': '6.62'},
 {'data': '01/02/1980', 'valor': '4.62'},
 {'data': '01/03/1980', 'valor': '6.04'},
 {'data': '01/04/1980', 'valor': '5.29'},
 {'data': '01/05/1980', 'valor': '5.70'},
 {'data': '01/06/1980', 'valor': '5.31'},
 {'data': '01/07/1980', 'valor': '5.55'},
 {'data': '01/08/1980', 'valor': '4.95'},
 {'data': '01/09/1980', 'valor': '4.23'},
 {'data': '01/10/1980', 'valor': '9.48'},
 {'data': '01/11/1980', 'valor': '6.67'},
 {'data': '01/12/1980', 'valor': '6.61'},
 {'data': '01/01/1981', 'valor': '6.84'},
 {'data': '01/02/1981', 'valor': '6.40'},
 {'data': '01/03/1981', 'valor': '4.97'},
 {'data': '01/04/1981', 'valor': '6.46'},
 {'data': '01/05/1981', 'valor': '5.56'},
 {'data': '01/06/1981', 'valor': '5.52'},
 {'data': '01/07/1981', 'valor': '6.26'},
 {'data': '01/08/1981', 'valor': '5.50'},
 {'data': '01/09/1981', 'valor': '5.26'},
 {'data': '01/10/1981', 'valor': '5.08'},
 {'data': '01/11/1981', 'valor': '5.27'},
 {'data': '01/12/1981', 'valor': '5.93'},
 {'data': '01/01/1982', 'valor': '6.97'},
 ...
 {'data': '01/11/2023', 'valor': '0.28'},
 {'data': '01/12/2023', 'valor': '0.56'},
 {'data': '01/01/2024', 'valor': '0.42'},
 {'data': '01/02/2024', 'valor': '0.83'},
 {'data': '01/03/2024', 'valor': '0.16'}]
```

- A biblioteca `requests` é usada novamente para fazer uma solicitação HTTP para o site oficial do National Art Gallery. A biblioteca `BeautifulSoup` é usada para analisar o HTML da página e extrair informações relevantes.

`soup.find()` e `soup.find_all()` são usados para encontrar elementos HTML com classes específicas.

```

# Fonte de Dados - Opção 3: Web Scraping
# Dados do Site Oficial do National Art Gallery

import requests
from bs4 import BeautifulSoup

page = requests.get('https://web.archive.org/web/20121007172955,

soup = BeautifulSoup(page.text, 'html.parser')

last_links = soup.find(class_='AlphaNav')
last_links.decompose()

artist_name_list = soup.find(class_='BodyText')
artist_name_list_items = artist_name_list.find_all('a')

# Usar .contents para pegar as tags <a> filhas
for artist_name in artist_name_list_items:
    names = artist_name.contents[0]
    print(names)

```

Zabaglia, Niccola
Zaccone, Fabian
Zadkine, Ossip
Zaech, Bernhard
Zagar, Jacob
Zagroba, Idalia
Zaidenberg, A.
Zaidenberg, Arthur
Zaisinger, Matthäus
Zajac, Jack
Zak, Eugène
Zakharov, Gurii Fillipovich
Zakowortny, Igor
Zalce, Alfredo
Zalopany, Michele
Zammiello, Craig
Zammitt, Norman
Zampieri, Domenico
Zampieri, called Domenichino, Domenico
Zanartú, Enrique Antunez
Zanchi, Antonio
Zanetti, Anton Maria
Zanetti Borzino, Leopoldina
Zanetti I, Antonio Maria, conte
Zanguidi, Jacopo
Zanini, Giuseppe
Zanini-Viola, Giuseppe
Zanotti, Giampietro
Zao Wou-Ki