

INTRODUÇÃO À CIÊNCIA DE DADOS

Introdução Web Scraping

WEB SCRAPING

- Vamos tratar de um tema que pode te levar a se sentir um verdadeiro hacker!
- Sim, realizar scraping é uma tarefa muito interessante, trata da prática de coletar dados automaticamente através de meios que não sejam uma API (Interface de Programação de Aplicações).
- O scraping consiste em uma aplicação capaz de consultar um servidor web, solicitar dados, analisar e extrair as informações necessárias.
- A prática de Web Scraping normalmente é uma solução quando não há outras maneiras de coletar dados necessários a um projeto, mas em muitos casos também é a prática de acelerar um processo que antes era feito manualmente.

WEB SCRAPING

- Muitas vezes bases de dados públicos não estão disponíveis para serem consumidas, entretanto os dados estão disponíveis por meio de acesso às páginas Web, o Scraping funciona muito bem para esse tipo de proposta.
- Lembre-se sempre que com a LGPD a coleta de dados é sempre algo que precisa ser pensado com muito cuidado. Nem todo dado que está público pode ser consumido e utilizado (como por exemplo os preços de um e-commerce).
- Ao pensar em Scraping, imagine um conjunto de dados que precisa coletar e juntar, e que um "robozinho" vai fazer isso pra você. Ahhh, o robzinho é você quem vai fazer!!

WEB SCRAPING

- Exemplos de dados a serem coletados!

TABELA

CLASSIFICAÇÃO		P	J	V	E	D	GP	GC	SG	%	ÚLT. JOGOS	
1	Atlético-MG	0	39	18	12	3	3	27	13	14	72.2	
2	Palmeiras	0	35	18	11	2	5	29	20	9	64.8	
3	Fortaleza	0	33	18	9	6	3	27	16	11	61.1	
4	Bragantino	0	32	18	8	8	2	29	19	10	59.3	
5	Flamengo	0	31	16	10	1	5	32	16	16	64.6	
6	Corinthians	0	27	18	7	6	5	17	15	2	50	
7	Atlético-GO	0	25	18	6	7	5	15	17	-2	46.3	
8	Ceará	0	24	18	5	9	4	19	19	0	44.4	
9	Athletico-PR	0	23	17	7	2	8	23	21	2	45.1	
10	Internacional	0	23	18	5	8	5	22	22	0	42.6	

Cotação

Dólar EUA	Compra (R\$)	Venda (R\$)
02/09 (PTAX)	5,1729	5,1735
02/09-13:00	5,1689	5,1695
Euro	Compra (R\$)	Venda (R\$)
02/09 (PTAX)	6,1325	6,1353
02/09-13:00	6,1277	6,1305

TradingView

Mercados · ações Brasil · Minerais Energéticos · Óleo Integrado



PETROBRAS PN N2

PETR4 BMFBOVESPA

26.95^{BRL} -0.09 (-0.33%)

MERCADO ABERTO (SET 02 14:32 UTC-3)

Visão geral

Ideias

S

Vacinômetro

Atualizado em: 02/09/2021 14:40

Doses aplicadas

52.128.408

Faça o seu pré-cadastro

Acompanhe a vacinação por município

96,36%

da população adulta (+18) de SP com pelo menos uma dose

50,48%

da população adulta (+18) de SP com esquema vacinal completo

34.180.518

Primeira dose

16.800.630

Segunda dose

1.147.260

Dose única*

76,32%

da população de SP com pelo menos uma dose

38,77%

da população de SP com esquema vacinal completo

*o imunizante de dose única disponível no país, no momento, é o da Janssen.

Tarifas de Pedágio

A CCR ViaOeste possui oito praças de pedágio instaladas pelas rodovias do Sistema Castello-Raposo. Seguindo os índices oficiais da Fundação Getúlio Vargas, os valores são ajustados anualmente em julho com base na variação do Índice Geral de Preços do Mercado (IGP-M), como estabelecido em contrato com o Governo do Estado de São Paulo. Consulte a tabela e confira onde estão localizadas as praças de pedágio no Sistema Castello-Raposo e o valor de cada uma das tarifas.

Raposo Tavares

SP-075

Castello Branco

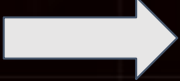
Raposo Tavares

Tipo de Veículo	Sentido Capital (Leste)	Sentido Interior (Oeste)
 Veículo de passeio	<ul style="list-style-type: none">Pedágio km 111 (Bidirecional) - Araçoiaba - R\$ 4.40Pedágio km 111 (Bidirecional) - AraçoiabaPedágio km 79 - Alumínio - R\$ 10.40Pedágio km 79 - Alumínio	<ul style="list-style-type: none">Pedágio km 111 (Bidirecional) - Araçoiaba - R\$ 4.40Pedágio km 111 (Bidirecional) - AraçoiabaPedágio km 46 - São Roque - R\$ 10.40Pedágio km 46 - São Roque
 Veículos comerciais (por eixo)	<ul style="list-style-type: none">Pedágio km 111 (Bidirecional) - Araçoiaba - R\$ 4.40Pedágio km 111 (Bidirecional) - AraçoiabaPedágio km 79 - Alumínio - R\$ 10.40Pedágio km 79 - Alumínio	<ul style="list-style-type: none">Pedágio km 111 (Bidirecional) - Araçoiaba - R\$ 4.40Pedágio km 111 (Bidirecional) - AraçoiabaPedágio km 46 - São Roque - R\$ 10.40Pedágio km 46 - São Roque

BIBLIOTECA BEAUTIFULSOUP

- A biblioteca BeautifulSoup para Python, é uma das bibliotecas mais utilizadas para fazer web scraping.
- Há muitas outras bibliotecas para scraping, como Selenium e Scrapy, que também têm funcionalidades parecidas. Cada uma delas tem suas vantagens e desvantagens, entretanto uma grande vantagem entre elas é que a lógica de pensamento para operá-las tem alguma similaridade.
- A biblioteca BeautifulSoup constrói uma árvore a partir de vários elementos de uma página e fornece uma interface simples para acessá-los.
- Para poder utilizar a biblioteca BeautifulSoup vamos precisar da ajuda da biblioteca Requests e de um parseador html.

BIBLIOTECA BEAUTIFULSOUP

- Iremos realizar uma demonstração do uso da biblioteca para coletar dados de um site de popularidade de linguagens de programação.
- Demonstração 

FINALIZANDO

- Em muitos projetos de DS, pode ocorrer parte da captação de dados via Scraping.
- Realizar o scraping é uma tarefa que tem um resultado imediato no que ela se propõe e quanto mais o desenvolvedor conhecer o conteúdo a ser coletado, mais rica é a experiência, além de mais eficaz.
- Abusem do uso do BeautifulSoup e quando possível experimentem conhecer outras bibliotecas como Selenium e Scrapy.

INTRODUÇÃO À CIÊNCIA DE DADOS

Introdução Web Scraping