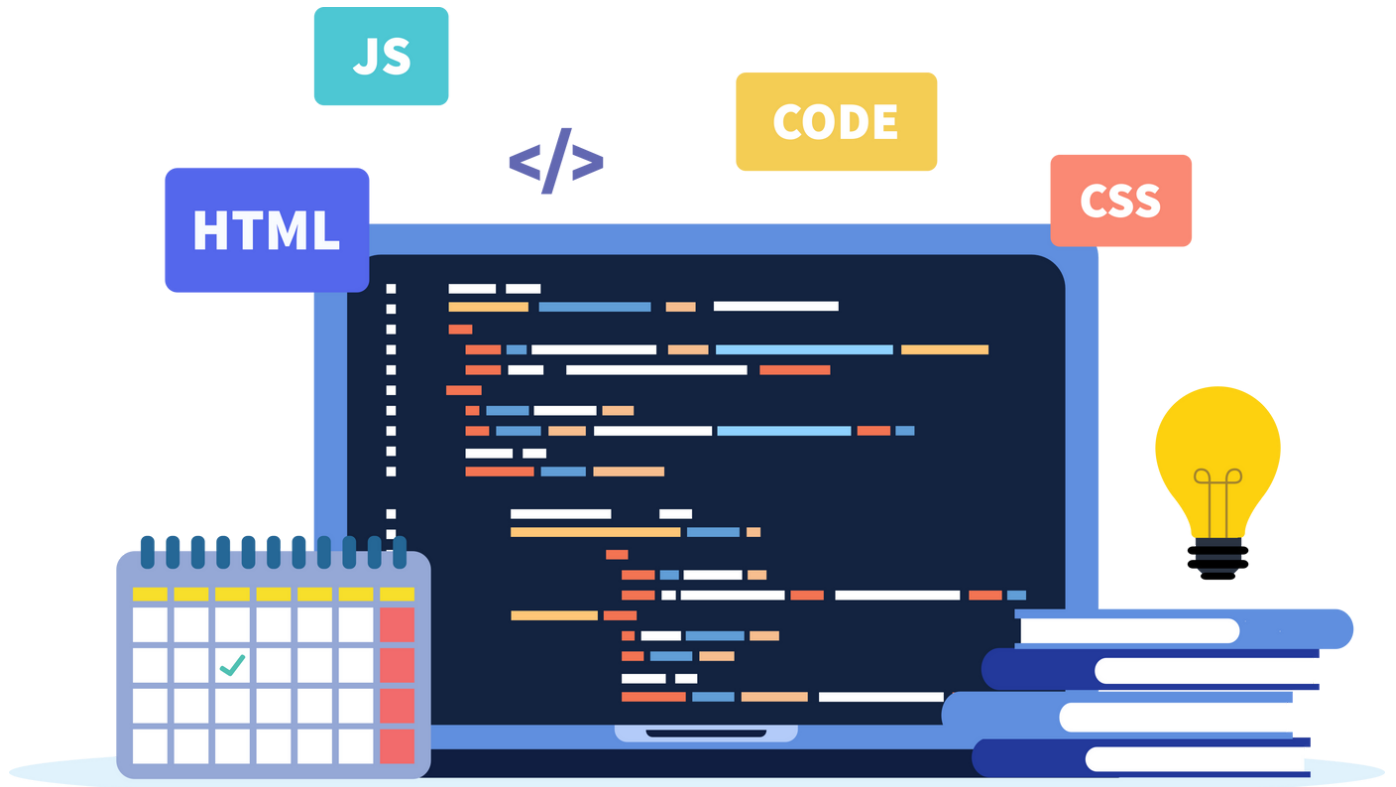# Web Scraping Cheatsheet

Fardeen A Khan

🔍 Installation

Install Beautiful Soup and Requests libraries using pip:

```
pip install beautifulsoup4
pip install requests
```

🌐 Import Libraries

Import the two libraries:

```
from bs4 import BeautifulSoup
import requests
```

📦 Send HTTP Request

Send a GET request to the URL:

```python
url = 'https://example.com'
response = requests.get(url)
```

🕵️ Create Soup Object

Create a BS object to parse HTML content:

```python
soup = BeautifulSoup(
response.text,'html.parser')
```

📝 Find Elements

Use find() and find_all() to locate HTML elements:

```html
<title> My Title </title>
<p class='para'> My Text1 </p>
<p class='para'> My Text2 </p>
```

```python
#get one title tag
title = soup.find('title')

#get all p tags with class as 'para'
paras = soup.find_all('p', class='para')
```

📊 Extract Data

Extract data from HTML elements using `.text`:

```html
<h1> My Heading </h1>
```

```python
#get heading tag
h_tag = soup.find('h1')

#get text from tag
h_text = h_tag.text

print(h_text)
#gives My Heading
```

📁 Navigate the DOM

Traverse the HTML DOM structure and get nested elements:

```html
<div id='parent'>
  <div class='child'>
  </div>
</div>
```

```python
#get parent
parent = soup.find('div', id='parent')
#get child from parent
child = parent.find('div', class ='child')
```

🔄 Loop Through Elements

Use loops to iterate through multiple elements:

```
<p> My para 1 </p>
<p> My para 2 </p>


#get all paragraphs
paras = soup.find_all('p')


#print text of all p elements
for para in paras:
    print(para.text)
```

FOLLOW FOR MORE CONTENT

FARDEEN AHMAD KHAN

https://linkedin.com/in/meetfardeen

https://github.com/I-Fardeen

Read my Technical Articles on:

Medium

https://fardeenk.medium.com