



# **5 Must-Know Python Libraries for Data Extraction**



# What is Web Scraping?

Web scraping is a method to extract data from websites. It's invaluable for analysts to gather unique datasets.

**Add Image:** Python logo intertwined with a spider web.



# 1. BeautifulSoup

Beginner-friendly & pairs well with requests.



```
from bs4 import BeautifulSoup
import requests
page = requests.get('URL')
soup = BeautifulSoup(page.content, 'html.parser')
```



## 2. Scrapy

Powerful, fast, and built for large-scale scraping.



```
import scrapy
class MySpider(scrapy.Spider):
    name = 'my_spider'
    start_urls = ['URL']
```





## 3. Selenium

Interact with web pages just like a human!



```
from selenium import webdriver  
driver = webdriver.Chrome()  
driver.get('URL')
```



## 4. PyQuery

jQuery syntax, making DOM manipulation easy.



```
from pyquery import PyQuery as pq  
d = pq(url='URL')
```



## 5. Requests-HTML

Combines the power of requests and HTML parsing.



```
from requests_html import HTMLSession
session = HTMLSession()
r = session.get('URL')
```



# Ethical Scraping

Always respect **robots.txt** of websites.  
Check the site's terms and conditions.  
Do not scrape aggressively!





# Dive In!

Web scraping opens up vast datasets.  
Explore these libraries and enhance your  
data collection skills!

MANOJ KUMAR





MANOJ KUMAR

**Join our Python course**  
to learn more.

Links are in  
the comments