Simple Scraping Data (Tanpa Melihat Element HTML)		
Praktikan	**Aslab**	
Nama: xxxx Nim: xxxx	Annur Hangga Prihadi	065001800028
	Faiz Kumara	065001800003

PRAKTIKUM 3

DATA SAINS DAN ANALITIK

Topik pertemuan praktikum ketiga adalah mengetahui cara sederhana untuk mengambil data di tabel yang berada di Website.

SOURCE CODE: https://github.com/hanggaa/PrakDSDA/blob/main/Prak%203.ipynb

Latihan 1

1. Memasang library yang dibutuhkan

```
In [1]: import sys
        !{sys.executable} -m pip install requests
        !{sys.executable} -m pip install lxml
        import requests
        import lxml.html as lh
        import pandas as pd
        %matplotlib inline
        import matplotlib.pyplot as plt
```

```
Requirement already satisfied: requests in c:\users\hangg\anaconda3\lib\site-packages (2.25.1)
Requirement already satisfied: chardet<5,>=3.0.2 in c:\users\hangg\anaconda3\lib\site-packages (from requests) (4.0.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\hangg\anaconda3\lib\site-packages (from requests) (1.26.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\hangg\anaconda3\lib\site-packages (from requests) (2020.12.5)
Requirement already satisfied: idna<3,>=2.5 in c:\users\hangg\anaconda3\lib\site-packages (from requests) (2.10)
Requirement already satisfied: lxml in c:\users\hangg\anaconda3\lib\site-packages (4.6.3)
```

2. Scrape Tabel

```
In [2]: url='http://pokemondb.net/pokedex/all'
        page = requests.get(url)
        doc = lh.fromstring(page.content)
        tr_elements = doc.xpath('//tr')
```

3. Mengetahui jumlah kolom tabel yang telah dikikis

```
In [3]: [len(Z) for Z in tr_elements[:8]]

Out[3]: [10, 10, 10, 10, 10, 10, 10, 10]
```

4. **Menguraikan tabel**

```
In [4]: tr_elements = doc.xpath('//tr')
        col=[]
        i=0
        for t in tr_elements[0]:
            i+=1
            name=t.text_content()
            print ('%d:"%s"'%(i,name))
            col.append((name,[]))
```

```
1:"#"
2:"Name"
3:"Type"
4:"Total"
5:"HP"
6:"Attack"
7:"Defense"
8:"Sp. Atk"
9:"Sp. Def"
10:"Speed"
```

5. **Mengubah tipe data rangkaian ke bentuk tuple**

```
In [5]: for j in range(1,len(tr_elements)):
            T=tr_elements[j]
            if len(T)!=10:
                break
            i=0

            for t in T.iterchildren():
                data=t.text_content()
                if i>0:
                    try:
                        data=int(data)
                    except:
                        pass
                col[i][1].append(data)
                i+=1
```

6. **Mengetahui jumlah baris**

```
In [6]: [len(B) for (title,B) in col]
Out[6]: [1045, 1045, 1045, 1045, 1045, 1045, 1045, 1045, 1045, 1045]
```

7. **Mengubah tipe data rangkaian ke dictionary lalu membuat dataframe dari dictionary**

```
In [7]: Dict={title:column for (title,column) in col}
        data=pd.DataFrame(Dict)
```

8. Tampilkan n data

```
In [8]: data.head()
```

Out[8]:

	#	Name	Type	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed
0	001	Bulbasaur	Grass Poison	318	45	49	49	65	65	45
1	002	Ivysaur	Grass Poison	405	60	62	63	80	80	60
2	003	Venusaur	Grass Poison	525	80	82	83	100	100	80
3	003	Venusaur Mega Venusaur	Grass Poison	625	80	100	123	122	120	80
4	004	Charmander	Fire	309	39	52	43	60	50	65

9. Mengetahui dimensi dataframe

```
In [9]: data.shape
Out[9]: (1045, 10)
```

10. Mengetahui deskripsi masing-masing kolom

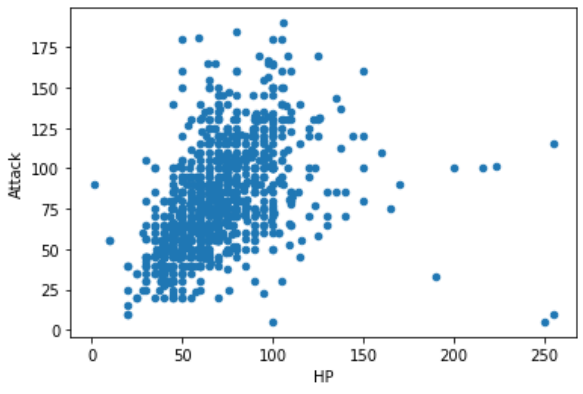
```
In [10]: data.describe()
```

Out[10]:

	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed
count	1045.000000	1045.000000	1045.000000	1045.000000	1045.000000	1045.000000	1045.000000
mean	439.314833	70.067943	80.466986	74.661244	73.022010	72.288995	68.807656
std	121.970701	26.671411	32.413665	31.237903	32.724797	28.074148	30.210094
min	175.000000	1.000000	5.000000	5.000000	10.000000	20.000000	5.000000
25%	330.000000	50.000000	55.000000	50.000000	50.000000	50.000000	45.000000
50%	458.000000	68.000000	77.000000	70.000000	65.000000	70.000000	65.000000
75%	515.000000	82.000000	100.000000	90.000000	95.000000	90.000000	90.000000
max	1125.000000	255.000000	190.000000	250.000000	194.000000	250.000000	200.000000

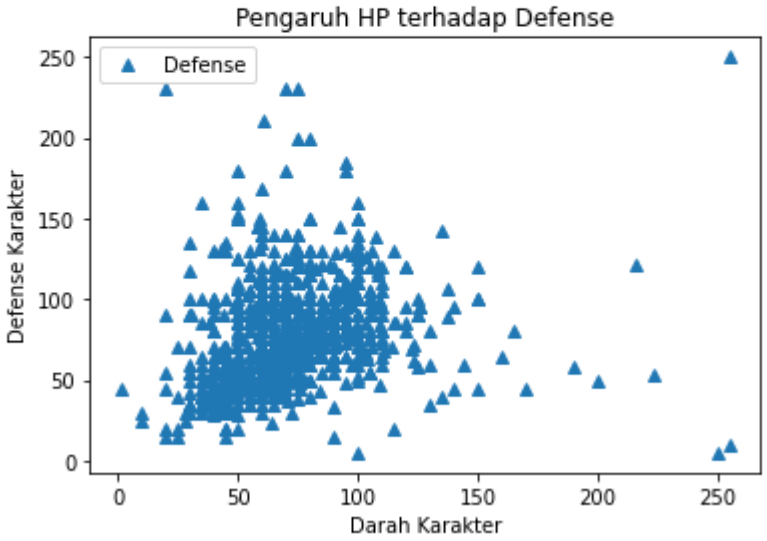
11. Visualisasi scatter plot antara HP dengan Attack

```
In [11]: data.plot.scatter('HP','Attack')
Out[11]: <AxesSubplot:xlabel='HP', ylabel='Attack'>
```



12. Visualisasi scatter plot antara HP dengan Defense

```
In [13]: data.plot(x = "HP", y = "Defense", style = "^")
         plt.title("Pengaruh HP terhadap Defense")
         plt.xlabel("Darah Karakter")
         plt.ylabel("Defense Karakter")
         plt.show()
```



Latihan 2

1. Visualisasikan scatter plot antara Attack dengan Defense
2. Visualisasikan scatter plot antara Attack dengan Speed
3. Visualisasikan scatter plot antara Attack dengan Sp. Atk
4. Visualisasikan scatter plot antara Attack dengan Sp. Def

Lampiran Screenshot hasil 1, 2, 3, dan 4

Input screenshot disini

Makna dari masing-masing hasil di atas!

Ketik makna disini