

# Web Scraping Using Python - Mamata Garanayak

Imagine you have to pull a large amount of data from websites and you want to do it as quickly as possible. How would you do it without manually going to each website and getting the data? Well, "Web Scraping" is the answer. Web Scraping just makes this job easier and faster.

## Why Web Scraping

Web scraping is used to collect large information from websites. But why does someone have to collect such large data from websites? To know about this, let's look at the applications of web scraping:

### Price Comparison:

Services such as ParseHub use web scraping to collect data from online shopping websites and use it to compare the prices of products.

### Email address gathering:

Many companies that use email as a medium for marketing, use web scraping to collect email ID and then send bulk emails.

### Social Media Scraping:

Web scraping is used to collect data from Social Media websites such as Twitter to find out what's trending.

### Research and Development:

Web scraping is used to collect a large set of data (Statistics, General Information, Temperature, etc.) from websites, which are analyzed and used to carry out Surveys or for R&D.

### Job listings:

Details regarding job openings, interviews are collected from different websites and then listed in one place so that it is easily accessible to the user.

## How does Web Scraping work?

When you run the code for web scraping, a request is sent to the URL that you have mentioned. As a response to the request, the server sends the data and allows you to read the HTML or XML page. The code then, parses the HTML or XML page, finds the data and extracts it.

To extract data using web scraping with python, you need to follow these basic steps:

```
Find the URL that you want to scrape
Inspecting the Page
Find the data you want to extract
Write the code
Run the code and extract the data
Store the data in the required format
```

## Libraries used for Web Scraping

As we know, Python is used for various applications and there are different libraries for different purposes. In our further demonstration, we will be using the following libraries:

Selenium:

Selenium is a web testing library. It is used to automate browser activities.

BeautifulSoup:

Soup is a Python package for parsing HTML and XML documents. It creates parse trees that is helpful to extract the data easily.

Pandas:

Pandas is a library used for data manipulation and analysis. It is used to extract the data and store it in the desired format.

## Scraping "FLIPKART" Website

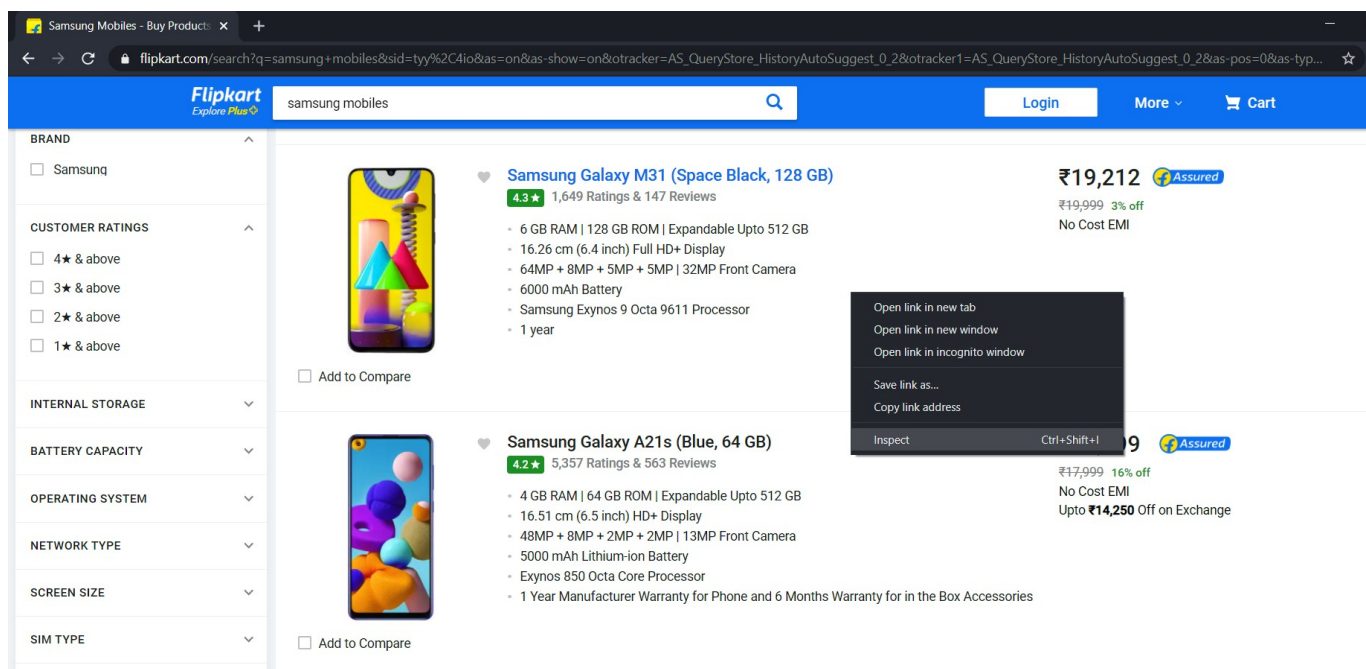
## Step 1: Find the URL that you want to scrape

For this example, we are going to scrape Flipkart website to extract the information about Samsung mobiles.

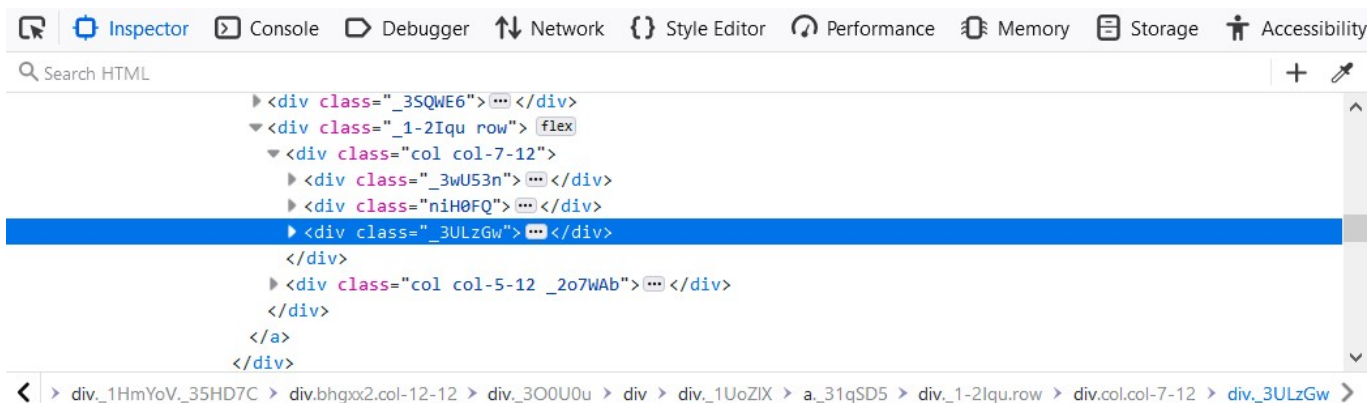
The URL for this page is [https://www.flipkart.com/search?q=samsung+mobiles&sid=tyy%2C4io&as=on&as-show=on&otracker=AS\\_QueryStore\\_HistoryAutoSuggest\\_0\\_2&otracker1=AS\\_QueryStore\\_HistoryAutoSuggest\\_0\\_2&as-pos=0&as-type=HISTORY&as-searchtext=samsung+mobiles](https://www.flipkart.com/search?q=samsung+mobiles&sid=tyy%2C4io&as=on&as-show=on&otracker=AS_QueryStore_HistoryAutoSuggest_0_2&otracker1=AS_QueryStore_HistoryAutoSuggest_0_2&as-pos=0&as-type=HISTORY&as-searchtext=samsung+mobiles)

## Step 2: Inspecting the Page

The data is usually nested in tags. So, we inspect the page to see, under which tag the data we want to scrape is nested. To inspect the page, just right-click on the element and click on "Inspect".



When you click on the "Inspect" tab, you will see a "Browser Inspector Box" open.



### Step 3: Find the data you want to extract

Let's extract the Price, Name, and Rating which is nested in the "div" tag respectively.

### Step 4: Write the code

First, let's create a Python file with .py extension.

### Install All the Required Libraries

```
In [1]: pip install bs4
```

```
In [2]: import pandas as pd
        from bs4 import BeautifulSoup as soup
        from urllib.request import urlopen as uReq
```

```
In [3]: my_url="https://www.flipkart.com/search?q=samsung+mobiles&sid=tyy%2C4io
        &as=on&as-show=on&otracker=AS_QueryStore_HistoryAutoSuggest_0_2&otracker1=AS_QueryStore_HistoryAutoSuggest_0_2&as-pos=0&as-type=HISTORY&as-searchtext=sa"
```

```
In [4]: uClient = uReq(my_url)
        page_html = uClient.read()
        uClient.close()
        page_soup = soup(page_html, "html.parser")
```

You can use the `find_all()` method of `soup` to extract useful html tags within a webpage. Examples of useful tags include `< a >` for hyperlinks, `< table >` for tables, `< tr >` for table rows, `< th >` for table headers, and `< td >` for table cells. The code below shows how to extract all the hyperlinks within the webpage.

```
In [5]: containers = page_soup.findAll("div", { "class": "_300U0u"})  
        print(len(containers))
```

24

```
In [6]: print(soup.prettify(containers[0]))
```

```
<div class="_300U0u">
  <div data-id="MOBFPNPS6QGTKBQB" style="width:100%">
    <div class="_1UoZlX">
      <a class="_3lqSD5" href="/samsung-galaxy-m31-space-black-128-gb/p/
itmeb29fb2c00580?pid=MOBFPNPS6QGTKBQB&lid=LSTM0BFPNPS6QGTKBQBLJM
AO&marketplace=FLIPKART&srno=s_1_1&otracker=AS_QueryStore
_HistoryAutoSuggest_0_2&otracker1=AS_QueryStore_HistoryAutoSugges
t_0_2&fm=organic&iid=3114eed4-32dc-45e6-9559-5bfbff40c68d.MOB
FPNPS6QGTKBQB.SEARCH&ssid=c4oxafmcsw0000001599206214916&qH=02
58c7d48242959a" rel="noopener noreferrer" target="_blank">
        <div class="_3SQWE6">
          <div class="_1OCn9C">
            <div>
              <div class="_3BTv9X" style="height:200px;width:200px">
                
              </div>
            </div>
            <div class="_2lesQu">
              <div class="_1O_CiZ">
                <span class="_1iHA1p">
                  <div class="_2kFyHg">
                    <label>
                      <input class="_3uUUD5" readonly="" type="checkbox"/>
                      <div class="_1p7h2j">
                        </div>
                    </label>
                  </div>
                </span>
                <label class="_10TB-Q">
                  <span>
                    Add to Compare
                  </span>
                </label>
              </div>
            </div>
            <div class="_3gDSOa _32A6AP">
              <div class="DsQ2eg">
                <svg class="_2oLiqr" height="16" viewBox="0 0 20 16" width="1
6" xmlns="http://www.w3.org/2000/svg">
                  <path class="_35Y7Yo" d="M8.695 16.682C4.06 12.382 1 9.536 1
6.065 1 3.219 3.178 1 5.95 1c1.566 0 3.069.746 4.05 1.915C10.981 1.74
5 12.484 1 14.05 1 16.822 1 19 3.22 19 6.065c0 3.471-3.06 6.316-7.695
10.617L10 17.897l-1.305-1.215z" fill="#2874F0" fill-rule="evenodd" op
acity=".9" stroke="#FFF">
                </path>
              </svg>
            </div>
          </div>
          <div class="_1-2Iqu row">
            <div class="col col-7-12">
              <div class="_3wU53n">
                Samsung Galaxy M31 (Space Black, 128 GB)
              </div>
            </div>
          </div>
        </a>
      </div>
    </div>
  </div>
```

```
</div>
<div class="niH0FQ">
  <span class="_2_KrJI" id="productRating_LSTMOBFPNPS6QGTKQBWLJ
MAO_MOBFPNPS6QGTKQB_">
    <div class="hGSR34">
      4.3
      
    </div>
  </span>
  <span class="_38sUEc">
    <span>
      <span>
        1,650 Ratings
      </span>
      <span class="_1VpSqZ">
        &
      </span>
      <span>
        148 Reviews
      </span>
    </span>
  </div>
  <div class="_3ULzGw">
    <ul class="vFw0gD">
      <li class="tVe95H">
        6 GB RAM | 128 GB ROM | Expandable Upto 512 GB
      </li>
      <li class="tVe95H">
        16.26 cm (6.4 inch) Full HD+ Display
      </li>
      <li class="tVe95H">
        64MP + 8MP + 5MP + 5MP | 32MP Front Camera
      </li>
      <li class="tVe95H">
        6000 mAh Battery
      </li>
      <li class="tVe95H">
        Samsung Exynos 9 Octa 9611 Processor
      </li>
      <li class="tVe95H">
        1 year
      </li>
    </ul>
  </div>
</div>
<div class="col col-5-12 _2o7WAb">
  <div class="_6BWGkk">
    <div class="_1uv9Cb">
      <div class="_1vC4OE _2rQ-NK">
        ₹19,212
      </div>
    </div>
  </div>
</div>
```



```

<div class="_3auQ3N _2GcJzG">
    ₹
    <!-- -->
    19,999
</div>
<div class="VGWI6T">
    <span>
        3% off
    </span>
</div>
</div>
<div class="_3n6o0t">
    
</div>
<div class="_2nE8_R">
    <div class="_3_G5Wj" style="color:#000000;font-size:14px;font-
family:inherit;font-weight:normal">
        No Cost EMI
    </div>
</div>
</div>
</div>
</a>
</div>
</div>
</div>

```

```

In [7]: container = containers[0]
        print(container.div.img["alt"])

```

Samsung Galaxy M31 (Space Black, 128 GB)

```

In [8]: price = container.findAll("div", {"class": "col col-5-12 _2o7WAb"})
        print(price[0].text)

```

₹19,212₹19,9993% offNo Cost EMI

```

In [9]: ratings = container.findAll("div", {"class": "niH0FQ"})
        print(ratings[0].text)

```

4.31,650 Ratings & 148 Reviews

```

In [10]: filename = "products.csv"
         f = open(filename, "w")

```

```

In [11]: headers = "Product_Name, Pricing, Ratings \n"
         f.write(headers)

```

Out[11]: 32

```
In [12]: for container in containers:
    product_name = container.div.img["alt"]
    price_container = container.findAll("div", {"class": "col col-5-12
_2o7WAb"})
    price = price_container[0].text.strip()

    rating_container = container.findAll("div", {"class": "niH0FQ"})
    rating = rating_container[0].text

    print("Product_Name:" + product_name)
    print("Price: " + price)
    print("Ratings:" + rating)

    #String parsing
    trim_price=''.join(price.split(','))
    rm_rupee = trim_price.split('₹')
    add_rs_price = "Rs." + rm_rupee[1]
    split_price = add_rs_price.split('E')
    final_price = split_price[0]

    split_rating = rating.split(" ")
    final_rating = split_rating[0]

    print(product_name.replace(",","|") + "," + final_price + "," + fina
l_rating + "\n")
    f.write(product_name.replace(",","|") + "," + final_price + "," + fi
nal_rating + "\n")
f.close()
```

Product\_Name:Samsung Galaxy M31 (Space Black, 128 GB)  
Price: ₹19,212₹19,993% offNo Cost EMI  
Ratings:4.31,650 Ratings & 148 Reviews  
Samsung Galaxy M31 (Space Black| 128 GB),Rs.19212,4.31,650

Product\_Name:Samsung Galaxy A21s (Blue, 64 GB)  
Price: ₹14,999₹17,99916% offNo Cost EMIUpto ₹14,250 Off on Exchange  
Ratings:4.25,362 Ratings & 564 Reviews  
Samsung Galaxy A21s (Blue| 64 GB),Rs.14999,4.25,362

Product\_Name:Samsung Galaxy A21s (Black, 64 GB)  
Price: ₹14,999₹17,99916% offNo Cost EMIUpto ₹14,250 Off on Exchange  
Ratings:4.25,362 Ratings & 564 Reviews  
Samsung Galaxy A21s (Black| 64 GB),Rs.14999,4.25,362

Product\_Name:Samsung Metro 313 Dual Sim  
Price: ₹2,130₹2,3408% offNo Cost EMI  
Ratings:4.216,650 Ratings & 1,928 Reviews  
Samsung Metro 313 Dual Sim,Rs.2130,4.216,650

Product\_Name:Samsung Metro 313 Dual Sim  
Price: ₹2,130₹2,3408% offNo Cost EMI  
Ratings:4.210,090 Ratings & 1,104 Reviews  
Samsung Metro 313 Dual Sim,Rs.2130,4.210,090

Product\_Name:Samsung Galaxy A21s (White, 64 GB)  
Price: ₹14,999₹17,99916% offNo Cost EMIUpto ₹14,250 Off on Exchange  
Ratings:4.25,362 Ratings & 564 Reviews  
Samsung Galaxy A21s (White| 64 GB),Rs.14999,4.25,362

Product\_Name:Samsung Galaxy M21 (Midnight Blue, 128 GB)  
Price: ₹17,590₹17,9992% offNo Cost EMI  
Ratings:4.4327 Ratings & 23 Reviews  
Samsung Galaxy M21 (Midnight Blue| 128 GB),Rs.17590,4.4327

Product\_Name:Samsung Galaxy M31 (Ocean Blue, 64 GB)  
Price: ₹18,490₹18,9492% offNo Cost EMI  
Ratings:4.31,650 Ratings & 148 Reviews  
Samsung Galaxy M31 (Ocean Blue| 64 GB),Rs.18490,4.31,650

Product\_Name:Samsung Guru 1200  
Price: ₹1,405No Cost EMI  
Ratings:4.31,29,342 Ratings & 12,238 Reviews  
Samsung Guru 1200,Rs.1405No Cost ,4.31,29,342

Product\_Name:Samsung Galaxy M11 (Black, 64 GB)  
Price: ₹13,383  
Ratings:4.23,489 Ratings & 290 Reviews  
Samsung Galaxy M11 (Black| 64 GB),Rs.13383,4.23,489

Product\_Name:Samsung Galaxy M31 (Ocean Blue, 128 GB)  
Price: ₹19,387₹19,9993% offNo Cost EMI  
Ratings:4.31,650 Ratings & 148 Reviews  
Samsung Galaxy M31 (Ocean Blue| 128 GB),Rs.19387,4.31,650

Product\_Name:Samsung Guru FM Plus SM-B110E/D

Price: ₹1,662No Cost EMI

Ratings:4.349,622 Ratings & 5,282 Reviews

Samsung Guru FM Plus SM-B110E/D, Rs.1662No Cost , 4.349,622

Product\_Name:Samsung Galaxy A31 (Prism Crush Black, 128 GB)

Price: ₹20,999₹23,99912% offNo Cost EMIUpto ₹17,500 Off on Exchange

Ratings:4.34,184 Ratings & 432 Reviews

Samsung Galaxy A31 (Prism Crush Black| 128 GB), Rs.20999, 4.34,184

Product\_Name:Samsung Galaxy M11 (Metallic Blue, 64 GB)

Price: ₹13,980₹14,4703% off

Ratings:4.23,489 Ratings & 290 Reviews

Samsung Galaxy M11 (Metallic Blue| 64 GB), Rs.13980, 4.23,489

Product\_Name:Samsung Galaxy A51 (Prism Crush Black, 128 GB)

Price: ₹25,999₹29,99913% offNo Cost EMIUpto ₹17,500 Off on Exchange

Ratings:4.4736 Ratings & 54 Reviews

Samsung Galaxy A51 (Prism Crush Black| 128 GB), Rs.25999, 4.4736

Product\_Name:Samsung Galaxy A31 (Prism Crush Blue, 128 GB)

Price: ₹20,999₹23,99912% offNo Cost EMIUpto ₹17,500 Off on Exchange

Ratings:4.34,184 Ratings & 432 Reviews

Samsung Galaxy A31 (Prism Crush Blue| 128 GB), Rs.20999, 4.34,184

Product\_Name:Samsung Galaxy A51 (Prism Crush White, 128 GB)

Price: ₹25,999₹29,99913% offNo Cost EMIUpto ₹17,500 Off on Exchange

Ratings:4.4736 Ratings & 54 Reviews

Samsung Galaxy A51 (Prism Crush White| 128 GB), Rs.25999, 4.4736

Product\_Name:Samsung Galaxy A10s (Black, 32 GB)

Price: ₹10,999

Ratings:4.31,546 Ratings & 118 Reviews

Samsung Galaxy A10s (Black| 32 GB), Rs.10999, 4.31,546

Product\_Name:Samsung Galaxy M11 (Metallic Blue, 32 GB)

Price: ₹11,190₹11,8895% off

Ratings:4.25,747 Ratings & 487 Reviews

Samsung Galaxy M11 (Metallic Blue| 32 GB), Rs.11190, 4.25,747

Product\_Name:Samsung Galaxy A51 (Prism Crush Blue, 128 GB)

Price: ₹25,999₹29,99913% offNo Cost EMIUpto ₹17,500 Off on Exchange

Ratings:4.4736 Ratings & 54 Reviews

Samsung Galaxy A51 (Prism Crush Blue| 128 GB), Rs.25999, 4.4736

Product\_Name:Samsung Galaxy A51 (Prism Crush Blue, 128 GB)

Price: ₹23,999₹25,9997% offNo Cost EMIUpto ₹17,500 Off on Exchange

Ratings:4.45,228 Ratings & 670 Reviews

Samsung Galaxy A51 (Prism Crush Blue| 128 GB), Rs.23999, 4.45,228

Product\_Name:Samsung Galaxy A31 (Prism Crush White, 128 GB)

Price: ₹20,999₹23,99912% offNo Cost EMIUpto ₹17,500 Off on Exchange

Ratings:4.34,184 Ratings & 432 Reviews

Samsung Galaxy A31 (Prism Crush White| 128 GB), Rs.20999, 4.34,184

Product\_Name:Samsung Guru FM Plus SM-B110E/D

Price: ₹1,662No Cost EMI

Ratings:4.349,622 Ratings & 5,282 Reviews

Samsung Guru FM Plus SM-B110E/D, Rs.1662No Cost , 4.349,622

Product\_Name:Samsung Galaxy M21 (Raven Black, 64 GB)

Price: ₹15,407₹15,9903% offNo Cost EMI

Ratings:4.4996 Ratings & 68 Reviews

Samsung Galaxy M21 (Raven Black| 64 GB), Rs.15407, 4.4996

**Now You Will Get A CSV File on your Folder named "Products.csv" with Information that You used for Web Scrapping**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Product_Name	Pricing	Ratings												
2	Samsung Galaxy M31 (Space Black  128 GB)	Rs.19212	4.31	650											
3	Samsung Galaxy A21s (Blue  64 GB)	Rs.14999	4.25	359											
4	Samsung Galaxy M21 (Raven Black  64 GB)	Rs.15407	4.4995												
5	Samsung Galaxy M11 (Metallic Blue  32 GB)	Rs.11190	4.25	747											
6	Samsung Galaxy A21s (Black  64 GB)	Rs.14999	4.25	359											
7	Samsung Metro 313 Dual Sim	Rs.2130	4.216	650											
8	Samsung Metro 313 Dual Sim	Rs.2130	4.21	89											
9	Samsung Galaxy M01 (Blue  32 GB)	Rs.8890	4.26	843											
10	Samsung Galaxy M21 (Midnight Blue  128 GB)	Rs.17590	4.4327												
11	Samsung Galaxy A21s (White  64 GB)	Rs.14999	4.25	359											
12	Samsung Guru 1200	Rs.1405No Cost	4.31	29	341										
13	Samsung Galaxy M31 (Ocean Blue  64 GB)	Rs.18490	4.31	650											
14	Samsung Galaxy M11 (Black  64 GB)	Rs.13383	4.23	488											
15	Samsung Galaxy M31 (Ocean Blue  128 GB)	Rs.19387	4.31	650											
16	Samsung Guru FM Plus SM-B110E/D	Rs.1662No Cost	4.349	621											
17	Samsung Galaxy A21s (Blue  64 GB)	Rs.16499	4.22	192											
18	Samsung Galaxy A31 (Prism Crush Black  128 GB)	Rs.20999	4.34	184											
19	Samsung Guru Music 2	Rs.1937	4.31	45	711										
20	Samsung Metro 313 Dual Sim	Rs.2340No Cost	4.39	256											
21	Samsung Galaxy A31 (Prism Crush Blue  128 GB)	Rs.20999	4.34	184											
22	Samsung Galaxy A10s (Black  32 GB)	Rs.10999	4.31	546											
23	Samsung Galaxy A51 (Prism Crush Black  128 GB)	Rs.25999	4.4736												
24	Samsung Galaxy A51 (Prism Crush White  128 GB)	Rs.25999	4.4736												