

Philips Website Scraping

A Project Submitted to the
IT Vedant Institute, Thane.

Data Science & Data Analytics With AI



Python-Web-Scrapping Project

BY

Vikrant.S.Matele

Under the Guidance of

Mr. Sameer Warsolkar

Philips Lights



USING PYTHON AND BEAUTIFULSOUP

DESCRIPTION

Lights web scrapping project involves extracting data from websites. This data can be used for various purposes, such as data analysis, research, reporting, or populating databases. We are going to use Lights to scrape site. To do so, we are going to use tools like Python, Requests, BeautifulSoup.

OUTLINE

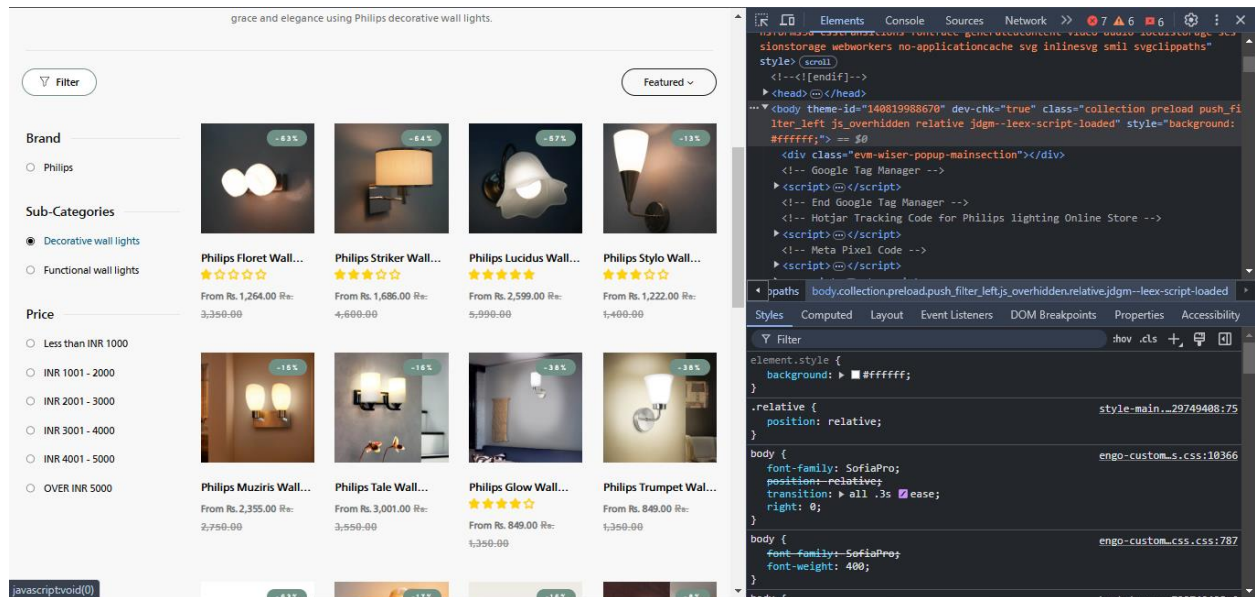
From this site, we are going to grab the following information:

- ☐ Lights Name
- ☐ Actual Price
- ☐ Discount
- ☐ Sale Price

STEPS:

Choose the Website and Webpage URL

Inspect the Website



Installing the important libraries

Python has several web scrapping libraries. We will use the following libraries:

- ☐ Requests
- ☐ BeautifulSoup
- ☐ Pandas

Write the Python source code

Exporting the extracted data

Philips lights Web Scraping Coding

Accessing the Website using requests and
BeautifulSoup Library , and Using html.parser

```
[1]: import requests
    from bs4 import BeautifulSoup
    import pandas as pd

[2]: gin=&gad_source=1&gclid=Cj0KCQiA57G5BhDUARIsACgCYnxK4uVcNMP7-00ruzQDB0teBtGPCXzjUkhGtB3ZwTdB22Xwan51srEaAkXoEALw_wcB&gclid=aw.ds

[3]: page

[3]: <Response [200]>

[4]: soup = BeautifulSoup(page.content, 'html.parser')
    soup

[4]: <!DOCTYPE html>

<!--[if IE 8]><html class="no-js lt-ie9" lang="en"> <![endif]-->
<!--[if IE 9 ]><html class="ie9 no-js"> <![endif]-->
<!--[if (gt IE 9)!!(IE)]><!--> <html class="no-js"> <!--<![endif]-->
<head>
```

```
[6]: print(soup.prettify())

<!DOCTYPE html>
<!--[if IE 8]><html class="no-js lt-ie9" lang="en"> <![endif]-->
<!--[if IE 9 ]><html class="ie9 no-js"> <![endif]-->
<!--[if (gt IE 9)!!(IE)]><!-->
<html class="no-js">
<!--<![endif]-->
<head>
  <!-- <script>
    window.onload = function() {
      window.scrollTo(0, 0);
    };
  </script> -->
<style>
  .custom-anim {
    display: inherit;
  }

  .info-product h4 a {
```

Now accessing the Lights name.

```
[5]: name = soup.find(class_="mobile_only").get_text()
    name

[5]: 'Philips Bling Wall light...'

[9]: name_tags = soup.select(".mobile_only")
    name_tags

[9]: [<a class="mobile_only" href="/collections/philips-festive-sale/products/philips-bling-wall-light?region_id=INNHO01">Philips
Bling Wall light...</a>,
      <a class="mobile_only" href="/collections/philips-festive-sale/products/philips-artizy-wall-light?region_id=INNHO01">Philips
Artizy Wall light...</a>,
      <a class="mobile_only" href="/collections/philips-festive-sale/products/philips-saturn-ceiling-light?region_id=INNHO01">Philips
Saturn LED Ceiling light...</a>,
      <a class="mobile_only" href="/collections/philips-festive-sale/products/philips-delight-wall-light?region_id=INNHO01">Philips
Delight Wall light...</a>,
      <a class="mobile_only" href="/collections/philips-festive-sale/products/philips-hexa-rimless-surface-light?region_id=INNHO01">Philips
Hexa Rimless LED Surface light...</a>,
      <a class="mobile_only" href="/collections/philips-festive-sale/products/philips-striker-floor-lamp?region_id=INNHO01">Philips
Striker Floor lamp...</a>,
      <a class="mobile_only" href="/collections/philips-festive-sale/products/philips-floret-wall-light?region_id=INNHO01">Philips
Floret Wall light...</a>]
```

```
[72]: light_title=[]

    for i in name_tags:
        a = i.get_text().replace("\n", " ")
        light_title.append(a)

    light_title = light_title[:16]
    light_title

[72]: ['Philips Bling Wall light...',
      'Philips Artizy Wall light...',
      'Philips Saturn LED Ceiling light...',
      'Philips Delight Wall light...',
      'Philips Hexa Rimless LED Surface light...',
      'Philips Striker Floor lamp...',
      'Philips Floret Wall light...',
      'Philips Striker Table Lamp...',
      'Philips Cosmos Desk light...',
      'Philips Lucent outdoor Wall washer...',
      'Philips Octave Wall light...',
      'Philips Shine LED Wall light...']
```

Accessing the Final price of every product.

```
[16]: light_price = soup.select(".price")
      light_price

[16]: [<span class="price">From Rs. 1,099.00</span>,
      <span class="price">From Rs. 999.00</span>,
      <span class="price">From Rs. 1,899.00</span>,
      <span class="price">From Rs. 999.00</span>,
      <span class="price">From Rs. 449.00</span>,
      <span class="price">From Rs. 2,499.00</span>,
      <span class="price">From Rs. 1,264.00</span>,
      <span class="price">From Rs. 1,299.00</span>,
      <span class="price">From Rs. 2,209.00</span>,
      <span class="price">From Rs. 1,199.00</span>,
      <span class="price">From Rs. 349.00</span>,
      <span class="price">From Rs. 449.00</span>,
      <span class="price">From Rs. 6,499.00</span>,
      <span class="price">From Rs. 2,798.00</span>,
      <span class="price">From Rs. 2,599.00</span>,
      <span class="price">From Rs. 1,659.00</span>,
      <label class="price product-price">
      <span class="compare-price dark opaci title14"></span>
      </label>]
```

```
•[13]: discount_price=[]
      for i in light_price:
          a=i.get_text().replace("\n", " ")
          discount_price.append(a)
      discount_price = discount_price[:16]
      discount_price
```

```
[13]: ['From Rs. 1,099.00',
      'From Rs. 999.00',
      'From Rs. 1,899.00',
      'From Rs. 999.00',
      'From Rs. 449.00',
      'From Rs. 2,499.00',
      'From Rs. 1,264.00',
      'From Rs. 1,299.00',
      'From Rs. 2,209.00',
      'From Rs. 349.00',
      'From Rs. 1,199.00',
      'From Rs. 449.00',
      'From Rs. 6,499.00',
      'From Rs. 2,798.00',
      'From Rs. 2,599.00',
      'From Rs. 1,659.00']
```

Accessing the Actual price of every product.

```
[12]: model_price = soup.select(".price-old")
      model_price
```

```
[12]: [<s class="price-old">Rs. 5,490.00</s>,
      <s class="price-old">Rs. 5,340.00</s>,
      <s class="price-old">Rs. 3,430.00</s>,
      <s class="price-old">Rs. 7,490.00</s>,
      <s class="price-old">Rs. 1,350.00</s>,
      <s class="price-old">Rs. 8,950.00</s>,
      <s class="price-old">Rs. 3,350.00</s>,
      <s class="price-old">Rs. 4,600.00</s>,
      <s class="price-old">Rs. 2,900.00</s>,
      <s class="price-old">Rs. 1,250.00</s>,
      <s class="price-old">Rs. 1,990.00</s>,
      <s class="price-old">Rs. 1,650.00</s>,
      <s class="price-old">Rs. 32,990.00</s>,
      <s class="price-old">Rs. 3,150.00</s>,
      <s class="price-old">Rs. 5,990.00</s>,
      <s class="price-old">Rs. 2,200.00</s>]
```

```
•[29]: old_price=[]
      for i in model_price:
          a=i.get_text()
          old_price.append(a)
      old_price
```

```
[29]: ['Rs. 5,490.00',
      'Rs. 5,340.00',
      'Rs. 3,430.00',
      'Rs. 7,490.00',
      'Rs. 1,350.00',
      'Rs. 8,950.00',
      'Rs. 3,350.00',
      'Rs. 4,600.00',
      'Rs. 2,900.00',
      'Rs. 1,990.00',
      'Rs. 1,250.00',
      'Rs. 1,650.00',
      'Rs. 32,990.00',
      'Rs. 3,150.00',
      'Rs. 5,990.00',
      'Rs. 2,200.00']
```


Accessing the Sale percentage of every product.

```
[31]: sale_percentage = soup.select(".sale-percent")
      sale_percentage
```

```
[31]: [<span class="sale-percent">-80%</span>,
      <span class="sale-percent">-82%</span>,
      <span class="sale-percent">-45%</span>,
      <span class="sale-percent">-87%</span>,
      <span class="sale-percent">-67%</span>,
      <span class="sale-percent">-73%</span>,
      <span class="sale-percent">-63%</span>,
      <span class="sale-percent">-72%</span>,
      <span class="sale-percent">-24%</span>,
      <span class="sale-percent">-40%</span>,
      <span class="sale-percent">-73%</span>,
      <span class="sale-percent">-73%</span>,
      <span class="sale-percent">-81%</span>,
      <span class="sale-percent">-12%</span>,
      <span class="sale-percent">-57%</span>,
      <span class="sale-percent">-25%</span>]
```

```
[32]: sale_per=[]

      for i in sale_percentage:
          a=i.get_text()
          sale_per.append(a)

      sale_per
```

```
[32]: ['-80%',
      '-82%',
      '-45%',
      '-87%',
      '-67%',
      '-73%',
      '-63%',
      '-72%',
      '-24%',
      '-40%',
      '-73%',
      '-73%',
      '-81%',
      '-12%',
```

Using Dataframe and pandas library.

```
[75]: import pandas as pd
file = pd.DataFrame(final_dict)
file
```

```
[75]:
```

	title	sale_price	actual_price	percentage_off
0	Philips Bling Wall light...	From Rs. 1,099.00	Rs. 5,490.00	-80%
1	Philips Artizy Wall light...	From Rs. 999.00	Rs. 5,340.00	-82%
2	Philips Saturn LED Ceiling light...	From Rs. 1,899.00	Rs. 3,430.00	-45%
3	Philips Delight Wall light...	From Rs. 999.00	Rs. 7,490.00	-87%
4	Philips Hexa Rimless LED Surface light...	From Rs. 449.00	Rs. 1,350.00	-67%
5	Philips Striker Floor lamp...	From Rs. 2,499.00	Rs. 8,950.00	-73%
6	Philips Floret Wall light...	From Rs. 1,264.00	Rs. 3,350.00	-63%
7	Philips Striker Table Lamp...	From Rs. 1,299.00	Rs. 4,600.00	-72%
8	Philips Cosmos Desk light...	From Rs. 2,209.00	Rs. 2,900.00	-24%
9	Philips Lucent outdoor Wall washer...	From Rs. 1,199.00	Rs. 1,990.00	-40%

Exporting all the data into a CSV File.

```
[19]: lights.to_csv("lights_report_scrapping.csv",index=False)

[20]: read=pd.read_csv("lights_report_scrapping.csv")
      read
```

[20]:

	title	sale_price	actual_price	percentage_off
0	Philips Bling Wall light...	From Rs. 1,099.00	Rs. 5,490.00	-80%
1	Philips Artizy Wall light...	From Rs. 999.00	Rs. 5,340.00	-82%
2	Philips Saturn LED Ceiling light...	From Rs. 1,899.00	Rs. 3,430.00	-45%
3	Philips Delight Wall light...	From Rs. 999.00	Rs. 7,490.00	-87%
4	Philips Hexa Rimless LED Surface light...	From Rs. 449.00	Rs. 1,350.00	-67%
5	Philips Striker Floor lamp...	From Rs. 2,499.00	Rs. 8,950.00	-73%
6	Philips Floret Wall light...	From Rs. 1,264.00	Rs. 3,350.00	-63%
7	Philips Striker Table Lamp...	From Rs. 1,299.00	Rs. 4,600.00	-72%
8	Philips Cosmos Desk light...	From Rs. 2,209.00	Rs. 2,900.00	-24%

1	title	sale_price	actual_price	percentage_off
2	Philips Bling Wall light...	From Rs. 1,099.00	Rs. 5,490.00	-80%
3	Philips Artizy Wall light...	From Rs. 999.00	Rs. 5,340.00	-82%
4	Philips Saturn LED Ceiling light...	From Rs. 1,899.00	Rs. 3,430.00	-45%
5	Philips Delight Wall light...	From Rs. 999.00	Rs. 7,490.00	-87%
6	Philips Hexa Rimless LED Surface light...	From Rs. 449.00	Rs. 1,350.00	-67%
7	Philips Striker Floor lamp...	From Rs. 2,499.00	Rs. 8,950.00	-73%
8	Philips Floret Wall light...	From Rs. 1,264.00	Rs. 3,350.00	-63%
9	Philips Striker Table Lamp...	From Rs. 1,299.00	Rs. 4,600.00	-72%
10	Philips Cosmos Desk light...	From Rs. 2,209.00	Rs. 2,900.00	-24%
11	Philips Octave Wall light...	From Rs. 349.00	Rs. 1,250.00	-73%
12	Philips Lucent outdoor Wall washer...	From Rs. 1,199.00	Rs. 1,990.00	-40%
13	Philips Shine LED Wall light...	From Rs. 449.00	Rs. 1,650.00	-73%
14	Philips Flagship suspended Chandelier...	From Rs. 6,499.00	Rs. 32,990.00	-81%
15	Philips Muziris Picture light...	From Rs. 2,798.00	Rs. 3,150.00	-12%
16	Philips Lucidus Wall light...	From Rs. 2,599.00	Rs. 5,990.00	-57%
17	Philips Orbit Desk light...	From Rs. 1,659.00	Rs. 2,200.00	-25%
18				

CONCLUSION:

In conclusion, the web scraping project focused on Philips lights has successfully achieved its objectives. The comprehensive data retrieval provided detailed specifications, product name, discount price, Actual price and Sale discount percentage of the product .

THANK

YOU

-viki matele