

▼ HP Products Web Scraping

```
!pip install gazpacho
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting gazpacho
  Downloading gazpacho-1.1.tar.gz (7.9 kB)
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing metadata (pyproject.toml) ... done
Building wheels for collected packages: gazpacho
  Building wheel for gazpacho (pyproject.toml) ... done
  Created wheel for gazpacho: filename=gazpacho-1.1-py3-none-any.whl size=7482 sha256=64a36e81cd5f5282fb4b6a411de765ff8f1f719f07f42759d1d9480228a9
  Stored in directory: /root/.cache/pip/wheels/ec/45/e0/490eb5e25601b4f9425fcde4a0034601c492a29e82268be4d3
Successfully built gazpacho
Installing collected packages: gazpacho
Successfully installed gazpacho-1.1
```

```
from gazpacho import Soup
from requests import get
```

To import the URL into the object.

```
url = "https://www.hp.com/in-en/shop/all-products.html?filter-prnttech=ink&hp_facet_segment=Home"
```

To check the status code.

```
resp = get(url)
resp.status_code
```

```
200
```

Parsing this url into the text form.

```
product_list = Soup(resp.text)
```

Inspect the website to find the attribute of product names and prices.

```
product_list.find("h2", {'class': 'plp-h2-title'}, mode='first').strip()
```

```
'HP DesignJet T250 24 (61 cm) Compact Large Format Plotter Printer (up to A1 size) with Mobile Printing'
```

```
pdls = product_list.find("h2", {'class': 'plp-h2-title'})
```

```
clean_pdls = [pdl.strip() for pdl in pdls]
```

```
print(clean_pdls)
```

```
['HP DesignJet T250 24 (61 cm) Compact Large Format Plotter Printer (up to A1 size) with Mobile Printing', 'HP Smart Tank 530 Printer', 'HP Smart Tank 515 Wirele
```

```
product_list.find("div", {'class': 'price-box price-final_price'}, mode='first').strip()
```

```
'₹83,500 (incl. of all taxes) Includes GST ₹12,737'
```

```
prices = product_list.find("div", {'class': 'price-box price-final_price'})
```

```
clean_prices = [p.strip() for p in prices]
```

```
print(clean_prices)
```

```
['₹83,500 (incl. of all taxes) Includes GST ₹12,737', '₹19,299 (incl. of all taxes) Includes GST ₹2,944', '₹14,999 (incl. of all taxes) Includes GST ₹2,288', '₹4,299 (incl
```

To convert text file into the dataframe.

```
import pandas as pd
finished_pdls = pd.DataFrame({
```

```
"product": clean_pdfs,\n"price": clean_prices\n})\n\nfinished_pdfs.head(11)
```

	product	price	
0	HP DesignJet T250 24 (61 cm) Compact Large For...	₹83,500 (incl. of all taxes) Includes GST ₹12,737	
1	HP Smart Tank 530 Printer	₹19,299 (incl. of all taxes) Includes GST ₹2,944	
2	HP Smart Tank 515 Wireless All-in-One	₹14,999 (incl. of all taxes) Includes GST ₹2,288	
3	HP DeskJet 2332 All-in-One Printer	₹4,299 (incl. of all taxes) Includes GST ₹656	
4	HP Smart Tank 500 All-in-One	₹11,599 (incl. of all taxes) Includes GST ₹1,769	
5	HP Ink Tank 316	₹10,999 (incl. of all taxes) Includes GST ₹1,678	
6	HP DeskJet Ink Advantage 2776 All-in-One Printer	₹6,466 (incl. of all taxes) Includes GST ₹986	
7	HP Smart Tank 516 Wireless All-in-One	₹15,799 (incl. of all taxes) Includes GST ₹2,410	
8	HP DeskJet Ink Advantage 2338 All-in-One Printer	₹4,120 (incl. of all taxes) Includes GST ₹628	
9	HP Smart Tank 790 Wi Fi Duplexer All-in-One Pr...	₹26,599 (incl. of all taxes) Includes GST ₹4,057	
10	HP Smart Tank 670 All-in-One	₹18,299 (incl. of all taxes) Includes GST ₹2,791	

[Colab paid products](#) - [Cancel contracts here](#)

✓ 1s completed at 11:56 PM

