

'Scraping Flipkart for Product Details using Python'



What is web scraping?

'Web Scraping' is the process of extracting data from the web automatically. Web scraping extracts underlying HTML code and the data stored in a database.

This extracted data can be saved in a structured format in the the form of a CSV or JSON file and also in other formats

1. Flipkart.

Flipkart Private Limited is an Indian e-commerce company, headquartered in Bengaluru, and incorporated in Singapore as a private limited company. The company initially focused on online book sales before expanding into other product categories such as consumer electronics, fashion, home essentials, groceries, and lifestyle products.



Objective:


To scrape the data of shoes by parsing the information from this website in the form of Tabular data.

If your footwear pinches or provides less space for your toes, your feet can develop various aches and blisters. A pair of shoes which provides a comfortable fit, shock-absorbing midsole cushioning will keep your feet comfortable for a long time. Some brands are known to consistently innovate and incorporate the latest technology to provide their diverse clientele, quality and comfortable footwear. Puma, SPARX, Red Tape, Denill, and Khadim's are some of the most acclaimed brands which endeavour to provide quality collections. A large number of the athletic variety is made with a sole of Phylon and Thermoplastic (TPR) material. Phylon provides ample flexibility to the footwear and makes the physical activity extremely comfortable. Thermoplastic rubber (TPR) complements the phylon to deliver incremental comfort to the feet. The house of Sparx is known to use this type of sole across its sports range. The rubber sole offers a sturdy grip and protects the person from slippery surfaces. The men's shoes (model-3M-678) from Sparx is ideal for running and the SL 164 model makes for a wonderful pair of women's shoes. The black and white kids' shoes (style code-SD0323B-BLACK-WHITE-4) can be worn by boys and girls. You can discover the wide range of these types of shoes online and make a selection accordingly. The information you are reading has been last updated on 04-Apr-23.

Shoes (Showing 1 – 40 products of 19,042 products)

Sort By **Relevance** Popularity Price -- Low to High Price -- High to Low Newest First

Ad



SFR

Casual Shoes For Men Sneakers F...

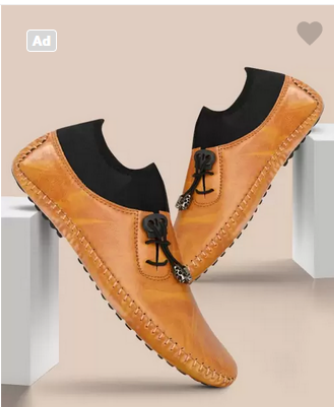
₹299

₹954

68% off

Hot Deal

Ad



K' Footlace

Loafers For Men


₹362

₹999

63% off

Price Drop

Ad



aadi

Mesh |Lightweight|Comfort|Sum...


₹229

₹999

77% off

Size 2, 6, 7, 8, 9, 10, 11, 12

Ad



BERSACHE

Bersache Casual Shoes For Men ...


₹664

₹2,995


77% off

Free delivery


Ad



Ad



Ad



Outline of the project:

- Installing required libraries.
- Download webpages using requests library,to parse and extract information use BeautifulSoup library and Accessing information such as shoe name, description, price and url.
- Storing the extracted data into a dictionary.
- Compiling all the data into a DataFrame using Pandas and saving the data into CSV file.

By the end of the project we'll create DataFrame in the following format:

	Title	price	description	link
0	density	₹1,999	Running Shoes For Men	https://flipkart.com/density-running-shoes-men...
1	CAMPUS	₹1,799	FIRST Running Shoes For Men	https://flipkart.com/campus-first-running-shoe...
2	SFR	₹998	FAST Trenddy Tainer Lace-ups Sporty Casuals Ru...	https://flipkart.com/sfr-fast-trenddy-tainer-l...
3	World Wear Footwear	₹998	Exclusive Affordable Collection of Trendy & St...	https://flipkart.com/world-wear-footwear-exclu...
4	SFR	₹715	2006 Trenddy Fashion Sporty Casuals Sneakers R...	https://flipkart.com/sfr-2006-trenddy-fashion-...
...
115	WOODLAND	₹4,195	Casuals For Men	https://flipkart.com/woodland-casuals-men/p/it...
116	CAMPUS	₹1,799	STREET-RUN Running Shoes For Men	https://flipkart.com/campus-street-run-running...
117	CAMPUS	₹3,499	REE-FLECT (N) Running Shoes For Men	https://flipkart.com/campus-ree-flect-n-runnin...
118	WAAN	₹1,999	Walking Shoes For Men	https://flipkart.com/waan-walking-shoes-men/p/...
119	Sparx	₹1,449	SM-484 Running Shoes For Men	https://flipkart.com/sparx-sm-484-running-shoe...

120 rows × 4 columns

1.Installing required libraries.

```
#Installing requests library.
```

```
!pip install requests --upgrade --quiet
```

```
import requests
```

```
# Installing pandas
```

```
!pip install pandas --quiet
```

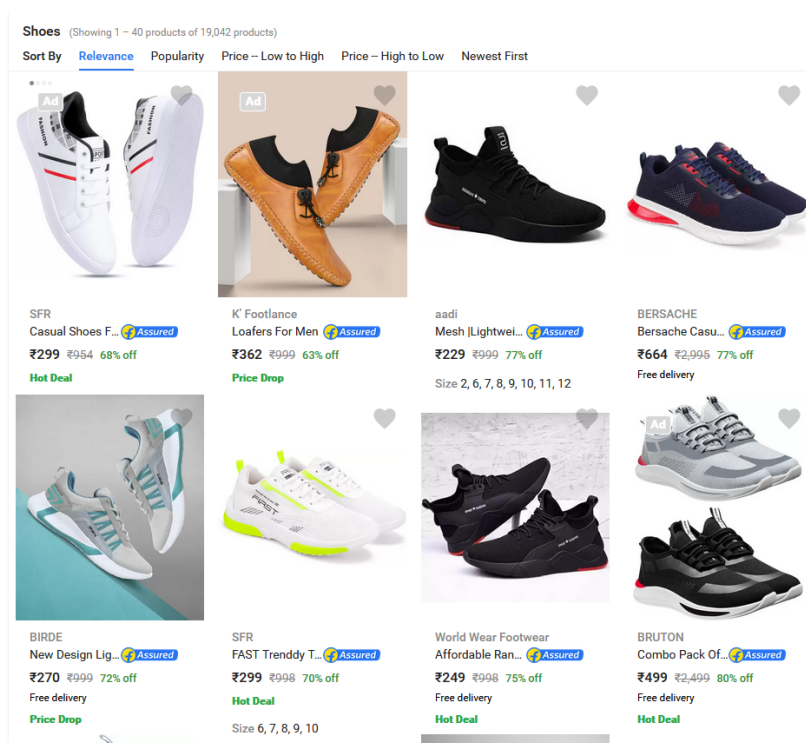
```
import pandas as pd
```

```
# Installing beautifulsoup4
```

```
!pip install beautifulsoup4 --upgrade --quiet
```

```
from bs4 import BeautifulSoup
```

2.Download webpage using requests library,to parse and extract information use BeautifulSoup library and Accessing information such as shoe name, description, price and url.



The screenshot shows a webpage titled 'Shoes (Showing 1 - 40 products of 19,042 products)'. It features a grid of shoe products with images, names, and prices. The products include SFR Casual Shoes, K' Footnace Loafers, aadi Mesh, BERSACHE Bersache Casu..., BIRDE New Design Lig..., SFR FAST Trendy T..., World Wear Footwear Affordable Ran..., and BRUTON Combo Pack Of... Each product listing includes a price, a discount percentage, and a 'Hot Deal' or 'Price Drop' badge.

```
<div class="_2ISqvP">
  <div class="_2uZuFa" target="_blank" rel="noopener noreferrer" href="/sfr-casual-shoes-m...
    <div data-
      <div class="_3yWGr_" style="padding-top: 120px;"
        <div class="_3l2yBx 5FzpgZ" style="padding-top: 120px;"
          
    <div class="_2HVSre _10ML35 -07Q4n">
      </div>
    </div>
  <div class="_28099V" style="transform: translate3d(0px, 0px, 0px);"
    <div class="_2WkVRV">
      <div class="1Rpw7a" title="Casual Shoes For Men Sneakers For Men" target="_blank" rel="
        href="/sfr-casual-shoes-men-sneakers/0/itm3c9ce7a5fc468?aid=SH000B-0t=s0800n=0r=producti
        ssid=log00mtf391ff281680618"
      </div>
      <div class="1a8UBa">
        <div class="3BPFwb" target="_blank" rel="noopener noreferrer" href="/sfr-casual-shoes-
        /0/itm3c9ce7a5fc468?aid=SH000B-0t=s0800n=0r=producti
        ssid=log00mtf391ff281680618"
      </div>
      <div class="_2ZdX08">
      </div>
      <div class="rcwEVK">
      </div>
    </div>
  </div>
```

The right side of the image shows a browser's developer tools window with the 'Rules' tab selected. It displays a list of CSS rules, with the rule for the selected element highlighted. The rule is for the class '_2WkVRV' and is located in the file 'app.chunk.8f4f44.css:1'. The rule includes properties for color, font-size, font-weight, box-sizing, border-box, margin, and padding.

```
def get_title(doc):
    product_name = []
    selection_class = '_2WkVRV'
    title_tags = doc.find_all('div',{'class' : selection_class})
    for i in title_tags:
        product_name.append(i.text)
    return product_name
```

```
def get_price(doc):
    product_price = []
    price_selector = '_3I9_wc'
```

```
price_tags = doc.find_all('div',{'class' : price_selector})
for i in price_tags:
    product_price.append(i.text)
return product_price
```

```
def get_description(doc):
    shoe_description = []
    desc_selector = 'IRpwTa'
    description_tags = doc.find_all('a',{'class' : desc_selector})
    for i in description_tags:
        shoe_description.append(i.text)
    return shoe_description
```

```
def get_link(doc):
    shoe_urls = []
    link_tags = doc.find_all('a',{'class' : '_2UzuFa' })
    for i in link_tags:
        base_url = 'https://flipkart.com'
        shoe_urls.append(base_url+ i['href'])
    return shoe_urls
```

3.Storing the extracted data into a dictionary.

```
def get_all_details(page_number):
    page_number = input('enter the number of pages')
    diction = {'title':[], 'price':[], 'desc':[], 'links':[]}
    for i in range(1,int(page_number)+1):
        #Url of the webpage that we are scrapping.
        url = 'https://www.flipkart.com/q/shoes?page=' + str(i)
        #We use requests.get to download the content from a webpage
        response = requests.get(url)
        page_content = response.text# Converting the page to BeautifulSoup document
        doc = BeautifulSoup( page_content, 'html.parser')
        #using get_title function to extract title
        diction['title'].extend(get_title(doc))
        #using get_price function to extract price
        diction['price'].extend(get_price(doc))
        #using get_description function to extract description
        diction['desc'].extend(get_description(doc))
        #using get_link function to extract link
        diction['links'].extend(get_link(doc))
    return diction
```

4.Compiling all the data into a DataFrame using Pandas and saving the data into CSV file.

```
shoe_df = pd.DataFrame(get_all_details(3))
shoe_df
```

enter the number of pages3

	title	price	desc	links
0	CAMPUS	₹1,799	FIRST Running Shoes For Men	https://flipkart.com/campus-first-running-shoe...
1	TR	₹1,499	Casual Shoes Caua Sneakers White Sneakers F...	https://flipkart.com/tr-casual-shoes-caual-sne...
2	aadi	₹999	Mesh Lightweight Comfort Summer Trendy Walkin...	https://flipkart.com/aadi-mesh-lightweight-com...
3	aadi	₹1,999	Synthetic Leather Lightweight Comfort Summer ...	https://flipkart.com/aadi-synthetic-leather-li...
4	BRUTON	₹1,299	Trendy Sports Running Running Shoes For Men	https://flipkart.com/bruton-trendy-sports-run...
...
115	ADIDAS	₹5,999	EvolveRun M Running Shoes For Men	https://flipkart.com/adidas-evolverun-m-runnin...
116	CAMPUS	₹1,699	NORTH PLUS Running Shoes For Men	https://flipkart.com/campus-north-plus-running...
117	BERSACHE	₹4,995	Bersache Sports Shoes For Men Light Grey For R...	https://flipkart.com/bersache-sports-shoes-men...
118	Asics	₹3,999	GEL-GALAXY 8B Running Shoes For Men	https://flipkart.com/asics-gel-galaxy-8b-runni...
119	aadi	₹1,999	Mesh Lightweight Comfort Summer Trendy Walkin...	https://flipkart.com/aadi-mesh-lightweight-com...

120 rows × 4 columns

```
shoe_df.to_csv('shoes'+'.csv', index=None)
```

Summary

- The Scraping was done using Python libraries such as Requests, beautifulsoup4 for extracting the data.
- Scraping the data of the shoes such as name, price, description and link.
- Parsed all the scraped data into a csv file containing 120 rows and 4 columns.

Future work

- Extracting more details of the shoe.
- Code optimization
- Improving the documentation part of the project.

References

- <https://www.youtube.com/watch?v=RKsLLG-bzEY>

- <https://www.flipkart.com/>
- <https://en.wikipedia.org/wiki/Flipkart>