Nike Jordans WebScrapping

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

- requests library will give us the content from the given Url.
- By Using Beautiful Soup library we can extract and parse data from HTML files.
- Use requests and Beautiful Soup for scraping and parsing data from the Web.
- Pandas makes it easy to scrape a table (tag) on a web page.
- After obtaining it as a DataFrame, we can save it as an Excel file or csv file.

In [2]:
```
url = "https://www.nike.com/in/w/mens-jordan-shoes-37eefznik1zy7ok"
```

- we will use the above url for scrapping the data with the help of requests library.
- If we get '200' response then we can scrap the page

In [3]:
```
response = requests.get(url)
print(response)
```

```
<Response [200]>
```

In [4]:
```
soup = BeautifulSoup(response.text)
soup
```

Out[4]:
```
<!DOCTYPE html>
<html lang="en-IN"><head><script id="new-relic-browser-agent" type="text/javascript">window.NREUM||(NREUM={});NREUM.info =
{"agent":"","beacon":"bam.nr-data.net","errorBeacon":"bam.nr-data.net","licenseKey":"NRBR-26d48a12bc762b8fbd7","applicationI
D":"45012625","agentToken":null,"applicationTime":266.237751,"transactionName":"ZVdXbUtXXBIHVUNfXlwde1ZLW1MND0xSUmRAWxoT","qu
eueTime":0,"ttGuid":"5a383373cafe7ff1"}; (window.NREUM||(NREUM={})).init={privacy:{cookies_enabled:true},ajax:{deny_list:["ba
m-cell.nr-data.net"]},distributed_tracing:{enabled:true,allowed_origins:["https://insights-collector.newrelic.com"],cors_use_
newrelic_header:true,cors_use_tracecontext_headers:true}};(window.NREUM||(NREUM={})).loader_config={agentID:"45722989",accoun
tID:"714737",trustKey:"1631518",xpid:"UwcDVlVUGwcFVFBQAQMC",licenseKey:"NRBR-26d48a12bc762b8fbd7",applicationID:"4501262
5"};;/*! For license information please see nr-loader-spa-1.238.0.min.js.LICENSE.txt */
(()=>{"use strict";var e,t,r={5763:(e,t,r)=>{r.d(t,{P_:()=>f,Mt:()=>p,C5:()=>s,DL:()=>v,OP:()=>T,lF:()=>D,Yu:()=>y,Dg:()=>h,C
X:()=>c,GE:()=>b,sU:()=>_});var n=r(8632),i=r(9567);const o={beacon:n.ce.beacon,errorBeacon:n.ce.errorBeacon,licenseKey:void
0,applicationID:void 0,sa:void 0,queueTime:void 0,applicationTime:void 0,ttGuid:void 0,user:void 0,account:void 0,product:voi
d 0,extra:void 0,jsAttributes:{},userAttributes:void 0,atts:void 0,transactionName:void 0,tNamePlain:void 0},a={};function s
(e){if(!e)throw new Error("All info objects require an agent identifier!");if(!a[e])throw new Error("Info for ".concat(e," wa
s never set"));return a[e]}function c(e,t){if(!e)throw new Error("All info objects require an agent identifier!");a[e]=(0,i.
D)(t,o),(0,n.Qy)(e,a[e],"info")}var u=r(7056);const d=()=>{const e={blockSelector:"[data-nr-block]",maskInputOptions:{passwor
d:!0}};return{allow_bfcache:!0,privacy:{cookies_enabled:!0},ajax:{deny_list:void 0,block_internal:!0,enabled:!0,harvestTimeSe
conds:10},distributed_tracing:{enabled:void 0,exclude_newrelic_header:void 0,cors_use_newrelic_header:void 0,cors_use_traceco
ntext_headers:void 0,allowed_origins:void 0},session:{domain:void 0,expiresMs:u.oD,inactiveMs:u.Hb},ssl:void 0,obfuscate:void
```

- BeautifulSoup will parse the above document.

In [5]:
```
shoe_elements = soup.find("div", class_="product-grid")
shoe_elements
```

```
zy" width="400"/></noscript></div></a><div class="product-card__info disable-animations"><div class=""><div class="product-ca
rd__titles"><div class="product-card__title" id="Jordan Max Aura 5" role="link">Jordan Max Aura 5</div><div class="product-ca
rd__subtitle" role="link">Men's Shoes</div></div></div><div class="product-card__count-wrapper false false" data-testid="prod
uct-card__count-wrapper "><div class="product-card__count-item"><button aria-expanded="false" class="product-card__colorway-b
tn" data-testid="product-card__colorway-btn" type="button"><div aria-label="Available in 4 Colors" class="product-card__produ
ct-count font-override__body1">4 Colours</div></button></div></div><div class="product-card__animation_wrapper"><div class="p
roduct-card__price-wrapper"><div class="product-card__price" data-testid="product-card__price" role="link"><div class="produc
t-price__wrapper css-1q1feg0"><div class="product-price in__styling is--current-price css-11s12ax" data-testid="product-pric
e" role="link">MRP : ₹ 11 895.00</div></div></div></div></div></div></figure></div></div><div class="product-card product-gri
d__card css-1xl2eyj" data-product-position="2" data-testid="product-card"><div class="product-card__body" data-el-type="Card"
data-testid="product-card__body"><figure><a class="product-card__link-overlay" data-testid="product-card__link-overlay" href
="https://www.nike.com/in/t/jordan-retro-6-g-golf-shoes-TQq6kw/DV1376-100" tabindex="-1">Jordan Retro 6 G</a><a aria-label="J
ordan Retro 6 G" class="product-card__img-link-overlay" data-el-type="Hero" data-testid="product-card__img-link-overlay" href
="https://www.nike.com/in/t/jordan-retro-6-g-golf-shoes-TQq6kw/DV1376-100"><div class="wall-image-loader css-1la3v4n" data-te
stid="wall-image-loader"><img alt="Jordan Retro 6 G Men's Golf Shoes" class="product-card__hero-image css-1fxh5tw" loading="l
azy" sizes="" src="" srcset=""/><noscript><img
alt="Jordan Retro 6 G Men's Golf Shoes" class="product-card__hero-image css-1fxh5tw" height="400" loading="lazy" width="400"/
></noscript></div></a><div class="product-card__info disable-animations"><div class=""><div class="product-card__titles"><div
class="product-card__title" id="Jordan Retro 6 G" role="link">Jordan Retro 6 G</div><div class="product-card__subtitle" role
```

- We will find the div tag using soup.find and use "product-grid" class to get required data and store it in 'shoe_elements' variable.

```
In [6]: shoe_elements = soup.find_all('div', class_='product-card')
        shoe_elements
```

```
Out[6]: [<div class="product-card product-grid__card css-1xl2eyj" data-product-position="1" data-testid="product-card"><div class="pr
        oduct-card__body" data-el-type="Card" data-testid="product-card__body"><figure><a class="product-card__link-overlay" data-tes
        tid="product-card__link-overlay" href="https://www.nike.com/in/t/jordan-max-aura-5-shoes-ZBZ4Pz/DZ4353-160" tabindex="-1">Jor
        dan Max Aura 5</a><a aria-label="Jordan Max Aura 5" class="product-card__img-link-overlay" data-el-type="Hero" data-testid="p
        roduct-card__img-link-overlay" href="https://www.nike.com/in/t/jordan-max-aura-5-shoes-ZBZ4Pz/DZ4353-160"><div class="wall-im
        age-loader css-1la3v4n" data-testid="wall-image-loader"><img alt="Jordan Max Aura 5 Men's Shoes" class="product-card__hero-im
        age css-1fxh5tw" loading="lazy" sizes="" src=""
        srcset=""/><noscript><img alt="Jordan Max Aura 5 Men's Shoes" class="product-card__hero-image css-1fxh5tw" height="400" loadi
        ng="lazy" width="400"/></noscript></div></a><div class="product-card__info disable-animations"><div class=""><div class="prod
        uct-card__titles"><div class="product-card__title" id="Jordan Max Aura 5" role="link">Jordan Max Aura 5</div><div class="prod
        uct-card__subtitle" role="link">Men's Shoes</div></div></div><div class="product-card__count-wrapper false false" data-testid
        ="product-card__count-wrapper "><div class="product-card__count-item"><button aria-expanded="false" class="product-card__colo
        rway-btn" data-testid="product-card__colorway-btn" type="button"><div aria-label="Available in 4 Colors" class="product-card_
        _product-count font-override__body1">4 Colours</div></button></div></div><div class="product-card__animation_wrapper"><div cl
        ass="product-card__price-wrapper"><div class="product-card__price" data-testid="product-card__price" role="link"><div class
        ="product-price__wrapper css-1q1feg0"><div class="product-price in__styling is--current-price css-11s12ax" data-testid="produ
        ct-price" role="link">MRP : ₹ 11 895.00</div></div></div></div></div></figure></div></div>,
         <div class="product-card product-grid__card css-1xl2eyj" data-product-position="2" data-testid="product-card"><div class="pr
        oduct-card__body" data-el-type="Card" data-testid="product-card__body"><figure><a class="product-card__link-overlay" data-tes
```

- We will find the div tag using soup.find and use "product-card" class to get required data and store it in 'shoe_elements' variable.

```
In [7]: names = []
        prices = []

        for shoe in shoe_elements:
            name = shoe.find('div', class_='product-card__title')
            price = shoe.find('div', class_='product-price')

            names.append(name)
            prices.append(price)
```

- This code extracts the names and prices of shoes from a web page by searching for specific HTML elements with corresponding CSS classes and stores them in two separate lists.

```
In [8]: Shoes = {'Name': names,'Price': prices}
```

- A dictionary named Shoes is created, where the column names are used as keys, and the corresponding lists of values are used as values.

```
In [9]: data = pd.DataFrame(Shoes)
        data
```

Out[9]:

	Name	Price
0	[Jordan Max Aura 5]	[MRP : ₹ 11 895.00]
1	[Jordan Retro 6 G]	[MRP : ₹ 19 695.00]
2	[Air Jordan 1 Low]	[MRP : ₹ 8 995.00]
3	[Jordan Stay Loyal 2]	[MRP : ₹ 10 295.00]
4	[Air Jordan Low SE]	[MRP : ₹ 10 295.00]
5	[Air Jordan 1 Low OG]	[MRP : ₹ 12 795.00]
6	[Jordan One Take 4 PF]	[MRP : ₹ 8 695.00]
7	[Air Jordan 1 Mid SE]	[MRP : ₹ 12 295.00]
8	[Air Jordan 5 Retro SE Craft]	[MRP : ₹ 19 295.00]
9	[Air Jordan 6 Retro]	[MRP : ₹ 18 395.00]
10	[Air Jordan 7 Retro]	[MRP : ₹ 18 395.00]
11	[AJKO 1]	[MRP : ₹ 12 795.00]
12	[Air Jordan 1 Retro High OG]	[MRP : ₹ 16 995.00]
13	[Jordan Why Not .6 PF]	[₹ 13 297.00]
14	[Air Jordan I High G]	[MRP : ₹ 16 995.00]
15	[Air Jordan XXXVIII PF]	[MRP : ₹ 18 395.00]
16	[Air Jordan 2 Retro Low Titan]	[₹ 19 277.00]
17	[Air Jordan 1 Mid SE]	[MRP : ₹ 12 295.00]
18	[Air Jordan 1 Zoom CMFT 2]	[MRP : ₹ 13 295.00]
19	[Air Jordan XXXVIII 'FIBA' PF]	[MRP : ₹ 18 395.00]
20	[Air Jordan Legacy 312 Low]	[MRP : ₹ 12 995.00]
21	[Air Jordan 1 Mid SE]	[MRP : ₹ 12 295.00]
22	[Air Jordan 1 Low SE Craft]	[MRP : ₹ 10 295.00]
23	[Air Jordan 1 'Next Chapter']	[MRP : ₹ 18 395.00]

- The 'Shoes' dictionary is passed to the pd.DataFrame() function to create a Pandas DataFrame named 'data'.

n [10]:
```
csv_file = 'nike_jordan_shoes.csv'
data.to_csv(csv_file, index=False)
```

- Here, We will convert 'data' dataframe into csv file
- The index parameter is set to False to exclude the index column from the saved CSV file.