A PROJECT REPORT

ON

"IKEA (FRAME SECTION) WEB SCRAPING"

PROJECT FOR PYTHON MODULE

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

BY

SUSMITA SANDIP PAWAR

UNDER THE GUIDENCE OF
MR. SAMEER WARSOLKAR

MASTER IN DATA SCIENCE & ANALYTICS WITH AI

INSTITUTE OF IT VENDANT

BRANCH – THANE (3 to 5)

COURSE CODE - T305

DATE – 12-08-2023

Aim: Creating a web scraping project for the frame section of the IKEA website can be a useful task for various purposes, such as frame names, frame size, or market research on price and rating.

Here's a step-by-step guide to help you get started

1. **Description:**

    Ikea (frame section) web scraping project involves extracting data from websites by automating the process using scripts or software. This data can be used for various purposes, such as data analysis, research, reporting, or populating databases.

This database contains 4 tables:
    1. Frame name
    2. Frame size
    3. Frame price
    4. Frame rating

2 . **Choose a Programming Language:**

    Decide on a programming language for web scraping project. Python is a popular choice due to its rich ecosystem of libraries, including BeautifulSoup and Scrapy, which are excellent for web scraping.

3. Install Required Libraries:

Install the necessary libraries. I am using the pip to install BeautifulSoup and requests if /''decide to go with Python:

Output:

```python
import requests
from bs4 import BeautifulSoup
import pandas as pd
URL = "https://www.ikea.com/in/en/cat/frames-pictures-10757/"
htmlcontent = r.content
soup = BeautifulSoup(htmlcontent,"html.parser")
#print(soup)
soup.prettify
```

4. Identify the Target URL:

Determine the URL of the IKEA webpage that contains the frame section want to scrape. For example:

URL = https://www.ikea.com/in/en/cat/frames-pictures-10757/

5. Send HTTP Request:

Use the requests library to send an HTTP GET request to the target URL and retrieve the webpage's HTML content.

Output:

```
import requests
r = requests.get(URL)
htmlcontent = r.content
print(htmlcontent)
```

6. Parse the HTML:
Use BeautifulSoup to parse the HTML content and extract the relevant information. Inspect the HTML source code of the webpage to identify the elements want to scrape, such as frame names, prices, and size.

Output:

```
soup = BeautifulSoup(htmlcontent,"html.parser")
print(soup)
```

7. Extract Data:
Continue to extract the data you need, such as prices and rating. Be mindful of the website's terms of service and use rate-limiting to avoid overloading the server with requests.

8. Data Storage:
Decide how and where you will store the scraped data. Common options include CSV files, databases.

Ikea (frames section) web scraping coding command as follows:

- Access the title from ikea web for the frame section
  Output:

```
title=soup.title
print(title)
```

```
<title> Frames - Buy picture and frame online at affordable price in india.  - IKEA</title>
```

- Access the frame names from ikea web for the frame section by using the following class
  Output:

```
frame_name = soup.find_all(class_="pip-header-section__title--small notranslate")
print(frame_name)
```

```
[<span class="pip-header-section__title--small notranslate" translate="no">VÄXBO </span>, <span class="pip-header-section__titl
e--small notranslate" translate="no">SILVERHÖJDEN </span>, <span class="pip-header-section__title--small notranslate" translate
="no">YLLEVAD </span>, <span class="pip-header-section__title--small notranslate" translate="no">HIMMELSBY </span>, <span class
="pip-header-section__title--small notranslate" translate="no">LERBODA </span>, <span class="pip-header-section__title--small n
otranslate" translate="no">KÄNGSLEBODA </span>, <span class="pip-header-section__title--small notranslate" translate="no">FISKB
O </span>, <span class="pip-header-section__title--small notranslate" translate="no">YLLEVAD </span>, <span class="pip-header-s
ection__title--small notranslate" translate="no">GRÖNBY </span>, <span class="pip-header-section__title--small notranslate" tra
nslate="no">YLLEVAD </span>, <span class="pip-header-section__title--small notranslate" translate="no">KÄNGSLEBODA </span>, <sp
an class="pip-header-section__title--small notranslate" translate="no">DROPS </span>]
```

```
list1=[]
for i in range(0,11):
    list1.append(frame_name[i].get_text())
print(list1)
print(len(list1))
```

```
['VÄXBO ', 'SILVERHÖJDEN ', 'YLLEVAD ', 'HIMMELSBY ', 'LERBODA ', 'KÄNGSLEBODA ', 'FISKBO ', 'YLLEVAD ', 'GRÖNBY ', 'YLLEVAD ',
'KÄNGSLEBODA ']
11
```

- Access the frame size from ikea web for the frame section by using the following class
  Output:

```
frame_size = soup.find_all(class_="pip-header-section__description-measurement")
print(frame_size)
```

```
[<span class="pip-header-section__description-measurement" data-ga-action="measurements_header_click" data-ga-label='13x18 cm
(5x7 ")'>13x18 cm (5x7 ")</span>, <span class="pip-header-section__description-measurement" data-ga-action="measurements_header
_click" data-ga-label='13x18 cm (5x7 ")'>13x18 cm (5x7 ")</span>, <span class="pip-header-section__description-measurement" dat
a-ga-action="measurements_header_click" data-ga-label='13x18 cm (5x7 ")'>13x18 cm (5x7 ")</span>, <span class="pip-header-secti
on__description-measurement" data-ga-action="measurements_header_click" data-ga-label='10x15 cm (4x6 ")'>10x15 cm (4x6 ")</span
>, <span class="pip-header-section__description-measurement" data-ga-action="measurements_header_click" data-ga-label='16x16 cm
(6 ¼x6 ¼ ")'>16x16 cm (6 ¼x6 ¼ ")</span>, <span class="pip-header-section__description-measurement" data-ga-action="measurement
s_header_click" data-ga-label='20x20 cm (7 ¾x7 ¾ ")'>20x20 cm (7 ¾x7 ¾ ")</span>, <span class="pip-header-section__description-
measurement" data-ga-action="measurements_header_click" data-ga-label='21x30 cm (8 ¼x11 ¾ ")'>21x30 cm (8 ¼x11 ¾ ")</span>, <sp
an class="pip-header-section__description-measurement" data-ga-action="measurements_header_click" data-ga-label='21x30 cm (8 ¼x
11 ¾ ")'>21x30 cm (8 ¼x11 ¾ ")</span>, <span class="pip-header-section__description-measurement" data-ga-action="measurements_h
eader_click" data-ga-label='179x112 cm (70 ½x44 ")'>179x112 cm (70 ½x44 ")</span>, <span class="pip-header-section__description
-measurement" data-ga-action="measurements_header_click" data-ga-label='21x41 cm (8 ¼x16 ¼ ")'>21x41 cm (8 ¼x16 ¼ ")</span>, <s
pan class="pip-header-section__description-measurement" data-ga-action="measurements_header_click" data-ga-label='20x20 cm (7 ¾
x7 ¾ ")'>20x20 cm (7 ¾x7 ¾ ")</span>]
```

```
list2=[]
for i in range((len(frame_size))):
    list2.append(frame_size[i].get_text())
print(list2)
print(len(list2))
```

```
['13x18 cm (5x7 ")', '13x18 cm (5x7 ")', '13x18 cm (5x7 ")', '10x15 cm (4x6 ")', '16x16 cm (6 ¼x6 ¼ ")', '20x20 cm (7 ¾x7 ¾
")', '21x30 cm (8 ¼x11 ¾ ")', '21x30 cm (8 ¼x11 ¾ ")', '179x112 cm (70 ½x44 ")', '21x41 cm (8 ¼x16 ¼ ")', '20x20 cm (7 ¾x7 ¾
")']
11
```

- Access the frame price from ikea web for the frame section by using the following class
  Output:

```
frame_price = soup.find_all(class_="pip-price")
print(frame_price)
```

```
[<span class="pip-price"><span class="pip-price__currency-symbol pip-price__currency-symbol--leading">Rs.</span><span class="pi
p-price__integer">799</span></span>, <span class="pip-price"><span class="pip-price__currency-symbol pip-price__currency-symbol
--leading">Rs.</span><span class="pip-price__integer">249</span></span>, <span class="pip-price pip-price--highlight"><span cla
ss="pip-price__currency-symbol pip-price__currency-symbol--leading">Rs.</span><span class="pip-price__integer">59</span></span
>, <span class="pip-price"><span class="pip-price__currency-symbol pip-price__currency-symbol--leading">Rs.</span><span class
="pip-price__integer">249</span></span>, <span class="pip-price"><span class="pip-price__currency-symbol pip-price__currency-sy
mbol--leading">Rs.</span><span class="pip-price__integer">399</span></span>, <span class="pip-price"><span class="pip-price__cu
rrency-symbol pip-price__currency-symbol--leading">Rs.</span><span class="pip-price__integer">129</span></span>, <span class="p
ip-price"><span class="pip-price__currency-symbol pip-price__currency-symbol--leading">Rs.</span><span class="pip-price__intege
r">149</span></span>, <span class="pip-price pip-price--highlight"><span class="pip-price__currency-symbol pip-price__currency-
symbol--leading">Rs.</span><span class="pip-price__integer">119</span></span>, <span class="pip-price"><span class="pip-price__
currency-symbol pip-price__currency-symbol--leading">Rs.</span><span class="pip-price__integer">4,490</span></span>, <span clas
s="pip-price pip-price--highlight"><span class="pip-price__currency-symbol pip-price__currency-symbol--leading">Rs.</span><span
class="pip-price__integer">199</span></span>, <span class="pip-price"><span class="pip-price__currency-symbol pip-price__curren
cy-symbol--leading">Rs.</span><span class="pip-price__integer">129</span></span>, <span class="pip-price"><span class="pip-pric
e__currency-symbol pip-price__currency-symbol--leading">Rs.</span><span class="pip-price__integer">399</span></span>]
```

```
list3=[]
for i in range(0,11):
    list3.append(frame_price[i].get_text())
print(list3)
print(len(list3))
```

```
['Rs.799', 'Rs.249', 'Rs.59', 'Rs.249', 'Rs.399', 'Rs.129', 'Rs.149', 'Rs.119', 'Rs.4,490', 'Rs.199', 'Rs.129']
11
```

- Access the frame rating from ikea web for the frame section by using the following class
  Output:

```python
frame_rating = soup.find_all(class_="pip-rating pip-rating--small")
print(frame_rating)
```

```
[<span class="pip-rating pip-rating--small"><span aria-label="Review: 4.3 out of 5 stars. Total reviews: 70" role="img"><svg
aria-hidden="true" class="pip-svg-icon pip-rating__star pip-rating__star--filled" focusable="false" height="24" viewbox="0 0
24 24" width="24"><path d="m12 6 2.1245 3.6818 4.1255.9018-2.8125 3.1773L15.8627 18 12 16.2818 8.1373 18l.4252-4.2391L5.75 1
0.5836l4.1255-.9018L12 6z"></path></svg><svg aria-hidden="true" class="pip-svg-icon pip-rating__star pip-rating__star--fille
d" focusable="false" height="24" viewbox="0 0 24 24" width="24"><path d="m12 6 2.1245 3.6818 4.1255.9018-2.8125 3.1773L15.862
7 18 12 16.2818 8.1373 18l.4252-4.2391L5.75 10.5836l4.1255-.9018L12 6z"></path></svg><svg aria-hidden="true" class="pip-svg-i
con pip-rating__star pip-rating__star--filled" focusable="false" height="24" viewbox="0 0 24 24" width="24"><path d="m12 6 2.
1245 3.6818 4.1255.9018-2.8125 3.1773L15.8627 18 12 16.2818 8.1373 18l.4252-4.2391L5.75 10.5836l4.1255-.9018L12 6z"></path></
svg><svg aria-hidden="true" class="pip-svg-icon pip-rating__star pip-rating__star--filled" focusable="false" height="24" view
box="0 0 24 24" width="24"><path d="m12 6 2.1245 3.6818 4.1255.9018-2.8125 3.1773L15.8627 18 12 16.2818 8.1373 18l.4252-4.239
1L5.75 10.5836l4.1255-.9018L12 6z"></path></svg><svg aria-hidden="true" class="pip-svg-icon pip-rating__star pip-rating__star
--half-filled" focusable="false" height="24" viewbox="0 0 24 24" width="24"><path d="m12 6 2.1245 3.6818 4.1255.9018-2.8125 3.1773L15.8627 18 12 16.2818V6z" fill="r
5 10.5836l4.1255-.9018L12 6z"></path><path d="m12 6 2.1245 3.6818 4.1255.9018-2.8125 3.1773L15.8627 18 12 16.2818V6z" fill="r
gb(var(--colour-neutral-3, 223, 223, 223))"></path></svg></span><span class="pip-rating__label">(70)</span></span>, <span cla
ss="pip-rating pip-rating--small"><span aria-label="Review: 4.6 out of 5 stars. Total reviews: 180" role="img"><svg aria-hidd
en="true" class="pip-svg-icon pip-rating__star pip-rating__star--filled" focusable="false" height="24" viewbox="0 0 24 24" wi
dth="24"><path d="m12 6 2.1245 3.6818 4.1255.9018-2.8125 3.1773L15.8627 18 12 16.2818 8.1373 18l.4252-4.2391L5.75 10.5836l4.1
255-.9018L12 6z"></path></svg><svg aria-hidden="true" class="pip-svg-icon pip-rating__star pip-rating__star--filled" focusabl
e="false" height="24" viewbox="0 0 24 24" width="24"><path d="m12 6 2.1245 3.6818 4.1255.9018-2.8125 3.1773L15.8627 18 12 16.
```

```python
list4=[]
for i in range(0,11):
    list4.append(frame_rating[i].get_text())
print(list4)
print(len(list4))
```

```
['(70)', '(180)', '(84)', '(105)', '(131)', '(10)', '(783)', '(215)', '(27)', '(25)', '(13)']
11
```

- Then arrange properly in column all above data by using import pandas library.
  Output:

```python
import pandas as pd
Frame = pd.DataFrame({
"Frame Name": list1,
"Frame Size": list2,
"Frame Price":list3,
"Frame Rating":list4

})
Frame
```

	Frame Name	Frame Size	Frame Price	Frame Rating
0	VÄXBO	13x18 cm (5x7 ")	Rs.799	(70)
1	SILVERHÖJDEN	13x18 cm (5x7 ")	Rs.249	(180)
2	YLLEVAD	13x18 cm (5x7 ")	Rs.59	(84)
3	HIMMELSBY	10x15 cm (4x6 ")	Rs.249	(105)
4	LERBODA	16x16 cm (6 ¼x6 ¼ ")	Rs.399	(131)
5	KÄNGSLEBODA	20x20 cm (7 ¾x7 ¾ ")	Rs.129	(10)
6	FISKBO	21x30 cm (8 ¼x11 ¾ ")	Rs.149	(783)
7	YLLEVAD	21x30 cm (8 ¼x11 ¾ ")	Rs.119	(215)
8	GRÖNBY	179x112 cm (70 ½x44 ")	Rs.4,490	(27)
9	YLLEVAD	21x41 cm (8 ¼x16 ¼ ")	Rs.199	(25)
10	KÄNGSLEBODA	20x20 cm (7 ¾x7 ¾ ")	Rs.129	(13)

- Then store above all data in CSV file and again open this file in application
  Output:

```python
Frame.to_csv("ikea_frame_Section.csv",index= False)
```

```python
frame = pd.read_csv("ikea_frame_Section.csv")
frame
```

	Frame Name	Frame Size	Frame Price	Frame Rating
0	VÄXBO	13x18 cm (5x7 ")	Rs.799	(70)
1	SILVERHÖJDEN	13x18 cm (5x7 ")	Rs.249	(180)
2	YLLEVAD	13x18 cm (5x7 ")	Rs.59	(84)
3	HIMMELSBY	10x15 cm (4x6 ")	Rs.249	(105)
4	LERBODA	16x16 cm (6 ¼x6 ¼ ")	Rs.399	(131)
5	KÄNGSLEBODA	20x20 cm (7 ¾x7 ¾ ")	Rs.129	(10)
6	FISKBO	21x30 cm (8 ¼x11 ¾ ")	Rs.149	(783)
7	YLLEVAD	21x30 cm (8 ¼x11 ¾ ")	Rs.119	(215)
8	GRÖNBY	179x112 cm (70 ½x44 ")	Rs.4,490	(27)
9	YLLEVAD	21x41 cm (8 ¼x16 ¼ ")	Rs.199	(25)
10	KÄNGSLEBODA	20x20 cm (7 ¾x7 ¾ ")	Rs.129	(13)

- Here the 2nd frame section from ikea web and the data frame is created
  Output:

```python
import pandas as pd
Frame1 = pd.DataFrame({
"Frame Name": list11,
"Frame Size": list12,
"Frame Price":list13,
"Frame Rating":list14

})
Frame1
```

	Frame Name	Frame Size	Frame Price	Frame Rating
0	VÄXBO	Collage frame for 8 photos, 13x18 cm (5x7 ")	Rs.799	(70)
1	SILVERHÖJDEN	Frame, 13x18 cm (5x7 ")	Rs.249	(180)
2	YLLEVAD	Frame, 13x18 cm (5x7 ")	Rs.59	(84)
3	HIMMELSBY	Frame, 10x15 cm (4x6 ")	Rs.249	(105)
4	LERBODA	Frame, 16x16 cm (6 ¼x6 ¼ ")	Rs.399	(131)
5	KÄNGSLEBODA	Wall decoration, 20x20 cm (7 ¾x7 ¾ ")	Rs.129	(10)
6	FISKBO	Frame, 21x30 cm (8 ¼x11 ¾ ")	Rs.149	(783)
7	YLLEVAD	Frame, 21x30 cm (8 ¼x11 ¾ ")	Rs.119	(215)
8	GRÖNBY	Picture, set of 9, 179x112 cm (70 ½x44 ")	Rs.4,490	(27)
9	YLLEVAD	Collage frame for 4 photos, 21x41 cm (8 ¼x16 ¼ ")	Rs.199	(25)
10	KÄNGSLEBODA	Wall decoration, 20x20 cm (7 ¾x7 ¾ ")	Rs.129	(13)
11	DROPS	Collage frame for 7 photos	Rs.399	(5)