

# Test 04

*4 September*  
*Artificial Intelligence*

*Mansoor Bukhari*



## MCQs

- 1) Requests
- 2) Parse Html Document
- 3) Get
- 4) Formats the parsed html for readability
- 5) Page not found
- 6) Knowledge of web frameworks
- 7) By Changing the url structure
- 8) You would get 404 error
- 9) `Df.to_csv('filename.csv', index=False )`
- 10) `Html.parser`
- 11) Find all elements with the specific tag or class
- 12) The text content of the element
- 13) To iterate through page until the certain condition is met
- 14) The loop would continue indefinitely
- 15) Both (a) and (b)
- 16) `Get_text()`
- 17) The attributes of the element
- 18) Pandas
- 19) Remove leading and trailing whitespaces
- 20) Identify the html tags and structure

## Project

```
import requests as r
from bs4 import BeautifulSoup
import pandas as pd

data = []

for page in range(1, 50):
    response =
    r.get(f'https://hashmicameras.pk/shop/page/{str(page)}')

    # Check if the request was successful
    if response.status_code == 200:
        # Parse the HTML content using BeautifulSoup
        soup = BeautifulSoup(response.content, 'html.parser')

        # Extract product information
        products = soup.find_all('h3', class_='wd-entities-title')

        for product in products:
            # Extract product link
            link = product.find('a')
```

```

        product_name = link.text.strip()
        product_url = link['href']

        # Find the price associated with the product
        parent_div = link.find_parent('div') # Adjust this based
on the HTML structure
        price_bdi = parent_div.find_next('bdi') if parent_div else
None
        product_price = price_bdi.text.strip() if price_bdi else
"Price not found"

        # Extract categories
        categories_div = soup.find('div', class_='wd-product-
cats')
        categories = []
        categories_url = []
        if categories_div:
            cat_links = categories_div.find_all('a')
            for category in cat_links:
                category_name = category.text.strip()
                category_url = category['href']
                categories.append(f"{category_name}")
                categories_url.append(f"{category_url}")
            else:
                categories.append("No categories found.")

        # Add product data to list
        data.append({
            'Product Name': product_name,
            'Product URL': product_url,
            'Product Price': product_price,
            'Categories': '; '.join(categories),
            'Categories Url': '; '.join(categories_url)
        })
    else:
        print(f'Failed to retrieve the page : {page}. Status code:
{response.status_code}')
        break

Failed to retrieve the page : 29. Status code: 404

# Convert list of dictionaries to DataFrame
df = pd.DataFrame(data)

# Export DataFrame to CSV
csv_file = 'products.csv'
df.to_csv(csv_file, index=False, encoding='utf-8')

print(f"Data successfully written to {csv_file}")

```

Data successfully written to products.csv

```
df = pd.read_csv('products.csv')
```

```
df.shape
```

```
(327, 5)
```

```
df.head()
```

	Product Name \
0	Angler BoomBox Octagonal Softbox with Bowens M...
1	Angler BoomBox Octagonal Softbox with Bowens M...
2	Aputure Amaran 200 XS Bi Color LED
3	Aputure Barndoors, Grid, and Gel Holder for LS...
4	Aputure Fresnel 2X Attachment

	Product URL	Product Price
0	https://hashmicameras.pk/product/angler-boombo...	Rs12,400.00
1	https://hashmicameras.pk/product/angler-boombo...	Rs92,000.00
2	https://hashmicameras.pk/product/aputure-amara...	Rs92,000.00
3	https://hashmicameras.pk/product/aputure-barnd...	Price not found
4	https://hashmicameras.pk/product/aputure-fresn...	Price not found

	Categories	Categories
Url		
0	Accessories; Softbox	
	https://hashmicameras.pk/product-category/acce...	
1	Accessories; Softbox	
	https://hashmicameras.pk/product-category/acce...	
2	Accessories; Softbox	
	https://hashmicameras.pk/product-category/acce...	
3	Accessories; Softbox	
	https://hashmicameras.pk/product-category/acce...	
4	Accessories; Softbox	
	https://hashmicameras.pk/product-category/acce...	