**Assignment 1-Web Scraping Using BeautifulSoup for RPA**

**Objective:**

This assignment is designed to help you practice web scraping using Python's BeautifulSoup library. You will scrape data from a sample website, process the data, and save it in a structured format (CSV). The exercise also introduces the concepts of ethical web scraping.

**Assignment Tasks**

**Task 1: Setting Up the Environment**

1. **Install Required Libraries**:
   * Ensure you have Python installed on your system.
   * Install the required libraries using the following command:

pip install beautifulsoup4 requests pandas

1. **Create a Python Script**:
   * Create a new Python file named <name\_web\_scraping\_assignment>.py.
   * Extracted csv file name should be <name\_sol\_assignment>.csv.

**Task 2: Select the Website**

For this assignment, use the following website:

* **Books to Scrape**: **http://books.toscrape.com/**  
  This site is designed for practicing web scraping. You will scrape book titles, prices, and availability status.

**Task 3: Write the Python Script**

1. **Send an HTTP Request**:
   * Use the requests library to fetch the webpage content.
   * Check for a successful response using the status code.
2. **Parse the HTML Content**:
   * Use BeautifulSoup to parse the HTML content of the page.
   * Identify and extract the following details for each book:
     + **Title**: The name of the book.
     + **Price**: The price of the book.
     + **Availability**: Whether the book is in stock or not.
3. **Store Data**:
   * Store the extracted data in a structured format, such as a list of dictionaries.
4. **Save Data to CSV**:
   * Use the pandas library to save the scraped data into a CSV file

**Task 4: Test the Script**

1. Run your script to scrape data from the website.
2. Verify that the CSV file (books.csv) is created and contains the scraped data.

**Expected Output**

A CSV file named books.csv with the following columns:

* Title
* Price
* Availability

**Code:**

Here’s a code template to guide you:

import requests

from bs4 import BeautifulSoup

import pandas as pd

# URL of the website

url = "http://books.toscrape.com/"

# Send an HTTP request to the URL

response = requests.get(url)

# Check if the request was successful

if response.status\_code == 200:

# Parse the HTML content using BeautifulSoup

soup = BeautifulSoup(response.text, 'html.parser')

# Extract book information

books = []

for book in soup.find\_all('article', class\_='product\_pod'):

title = book.h3.a['title']

price = book.find('p', class\_='price\_color').text.strip()

availability = book.find('p', class\_='instock availability').text.strip()

books.append({

'Title': title,

'Price': price,

'Availability': availability

})

# Save to a CSV file

<Write two line of code here>

print("Book data has been saved to 'books.csv'")

else:

print(f"Failed to fetch the webpage. Status code: {response.status\_code}")

**Submission Requirements**

**Upload both file in following google drive link in corresponding folders:**

[**https://drive.google.com/drive/folders/1C5Txhyw-bSbuSY5BKIvXQ\_-lHdlXc3Np?usp=sharing**](https://drive.google.com/drive/folders/1C5Txhyw-bSbuSY5BKIvXQ_-lHdlXc3Np?usp=sharing)

1. **Python Script**: Submit the Python script (name\_web\_scraping\_assignment.py).
2. **Output File**: Submit the generated CSV file <name\_sol\_assignment>.csv