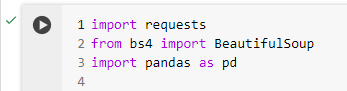
**Web Scraping (Book)**

# **Statement:**

Create a web scraping tool to extract essential book information, including titles, authors, genres, and publication dates, from a designated book-related website. Compile this data into a basic catalog, automating the process for regular updates. Provide a user-friendly interface for easy exploration and basic filtering by author, genre, or publication date. Adhere to legal and ethical guidelines throughout the scraping process and ensure timely updates to maintain a current and accessible book catalog.

# 

### **Importing Libraries**



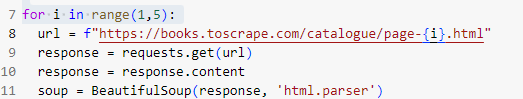
The required libraries are imported requests for making HTTP requests, BeautifulSoup for HTML parsing, and pandas for handling data in tabular form.

### **Initializing Book List**

### 

An empty list named books is created to store information about each book.

**Looping through Pages**



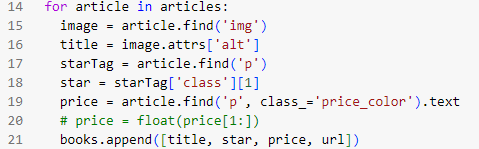
A loop is set up to iterate through pages 1 to 4 of the "[https://books.toscrape.com](https://books.toscrape.com/)" website. The requests.get() function is used to fetch the HTML content of each page, and BeautifulSoup is employed to parse the HTML.

### **Extracting Book Information**



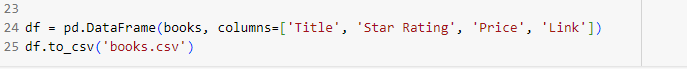
Within each page, it locates the 'ol' (ordered list) element containing book information and then finds all 'article' elements with the class 'product\_pod', which likely represents individual books.

### **Looping through Books on the Page**



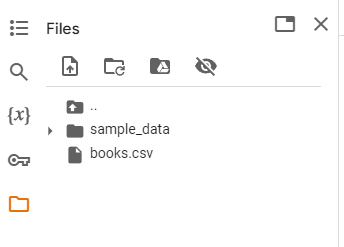
Within each book 'article', it finds the 'img' element to extract the book title from its 'alt' attribute, the 'p' element to extract the star rating from its 'class' attribute, and the 'p' element with the class 'price\_color' to extract the book price. The extracted information is then appended to the books list.

### **Creating DataFrame and Writing to CSV**



After the loop completes, the script creates a pandas DataFrame from the collected book information and writes it to a CSV file named 'books.csv'. The DataFrame columns are named 'Title', 'Star Rating', 'Price', and 'Link'. The CSV file will contain information about books from pages 1 to 4 of the specified website.

**Finally**



**Dataset after web scraping:**

