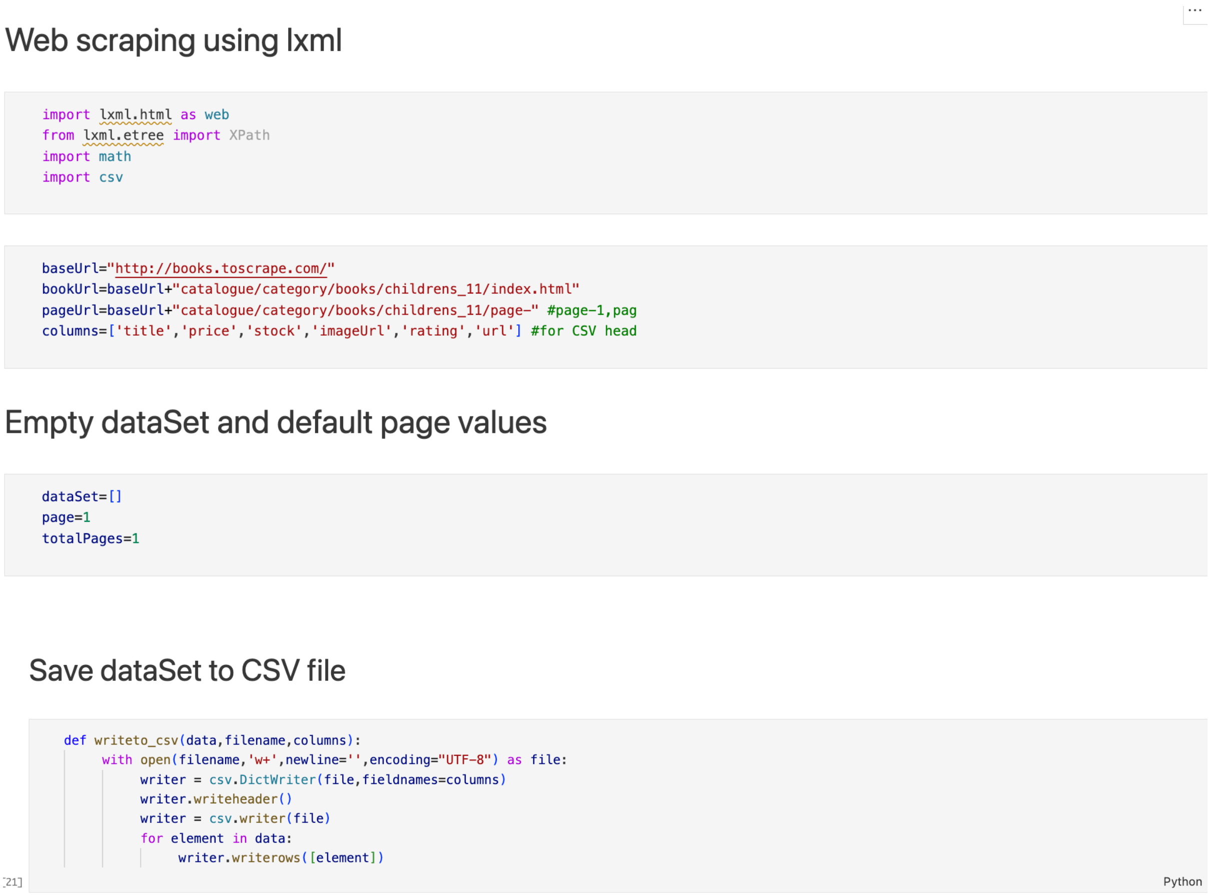


Web scraping using lxml

```python
import lxml.html as web
from lxml.etree import XPath
import math
import csv
```

```python
baseUrl="http://books.toscrape.com/"
bookUrl=baseUrl+"catalogue/category/books/childrens_11/index.html"
pageUrl=baseUrl+"catalogue/category/books/childrens_11/page-" #page-1,pag
columns=['title','price','stock','imageUrl','rating','url'] #for CSV head
```

Empty dataSet and default page values

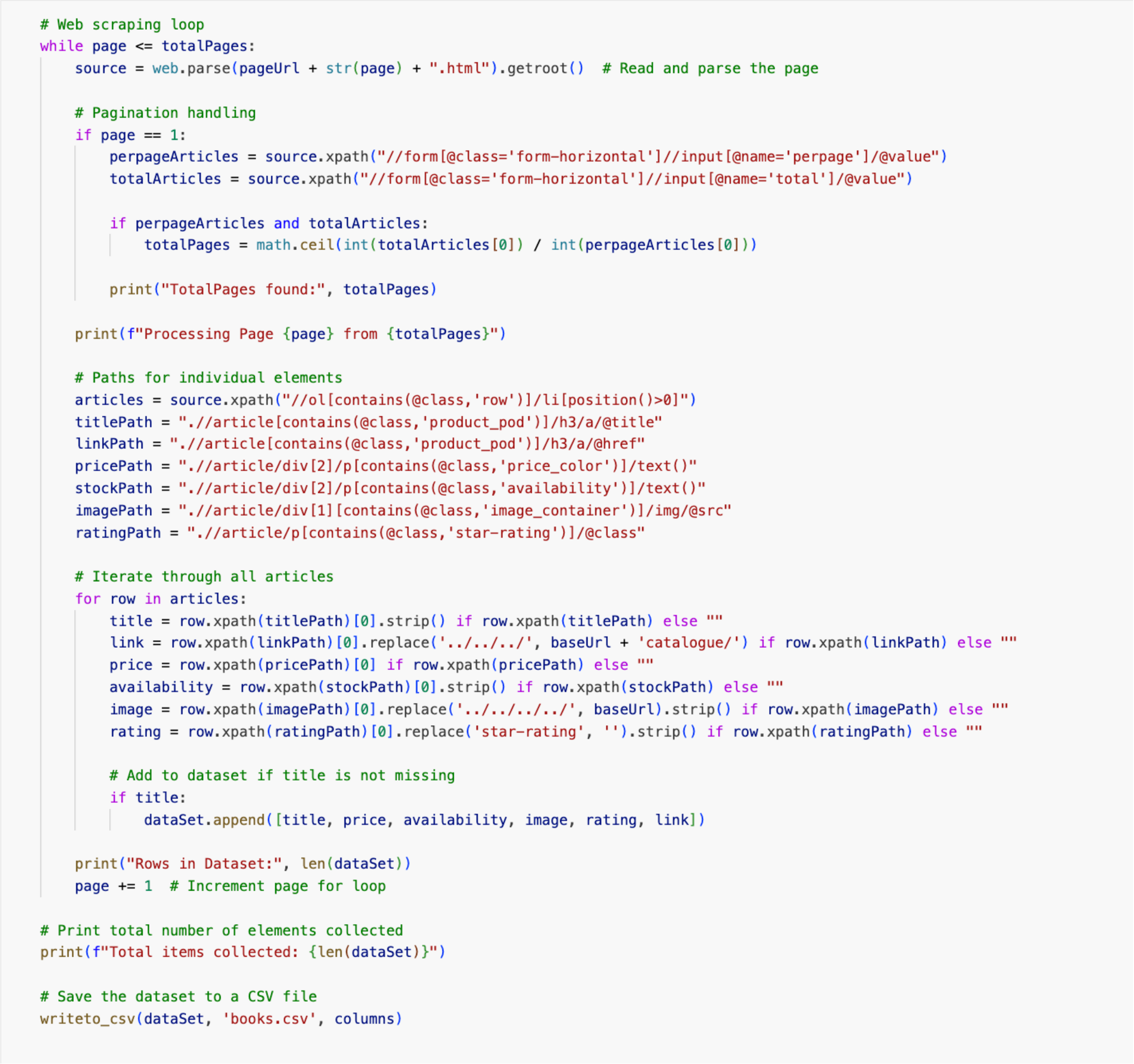
```python
dataSet=[]
page=1
totalPages=1
```

Save dataSet to CSV file

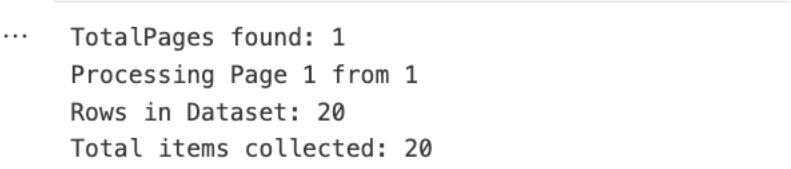
```python
def writeto_csv(data,filename,columns):
    with open(filename,'w+',newline='',encoding="UTF-8") as file:
        writer = csv.DictWriter(file,fieldnames=columns)
        writer.writeheader()
        writer = csv.writer(file)
        for element in data:
            writer.writerows([element])
```

[21]

Python

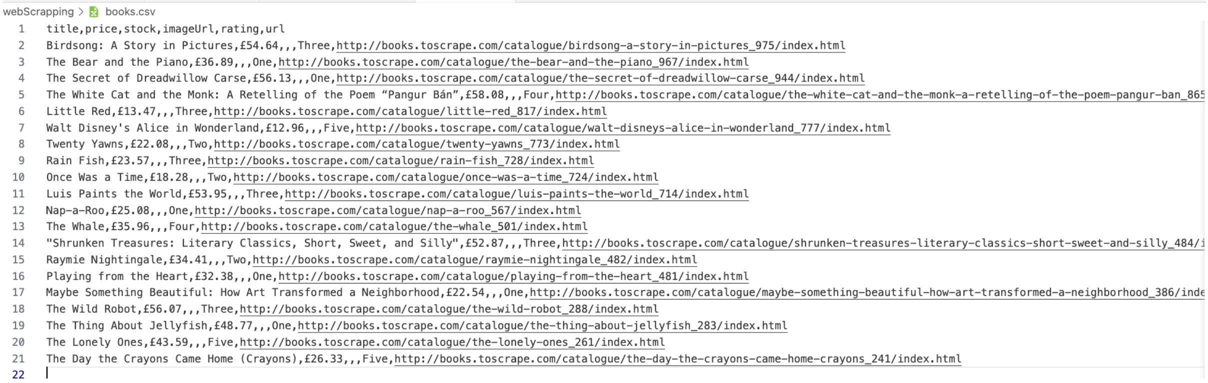


```python
# Web scraping loop
while page <= totalPages:
    source = web.parse(pageUrl + str(page) + ".html").getroot()  # Read and parse the page

    # Pagination handling
    if page == 1:
        perpageArticles = source.xpath("//form[@class='form-horizontal']//input[@name='perpage']/@value")
        totalArticles = source.xpath("//form[@class='form-horizontal']//input[@name='total']/@value")

        if perpageArticles and totalArticles:
            totalPages = math.ceil(int(totalArticles[0]) / int(perpageArticles[0]))

        print("TotalPages found:", totalPages)

    print(f"Processing Page {page} from {totalPages}")

    # Paths for individual elements
    articles = source.xpath("//ol[contains(@class,'row')]/li[position()>0]")
    titlePath = ".//article[contains(@class,'product_pod')]/h3/a/@title"
    linkPath = ".//article[contains(@class,'product_pod')]/h3/a/@href"
    pricePath = ".//article/div[2]/p[contains(@class,'price_color')]/text()"
    stockPath = ".//article/div[2]/p[contains(@class,'availability')]/text()"
    imagePath = ".//article/div[1][contains(@class,'image_container')]/img/@src"
    ratingPath = ".//article/p[contains(@class,'star-rating')]/@class"

    # Iterate through all articles
    for row in articles:
        title = row.xpath(titlePath)[0].strip() if row.xpath(titlePath) else ""
        link = row.xpath(linkPath)[0].replace('../../../', baseUrl + 'catalogue/') if row.xpath(linkPath) else ""
        price = row.xpath(pricePath)[0] if row.xpath(pricePath) else ""
        availability = row.xpath(stockPath)[0].strip() if row.xpath(stockPath) else ""
        image = row.xpath(imagePath)[0].replace('../../../../', baseUrl).strip() if row.xpath(imagePath) else ""
        rating = row.xpath(ratingPath)[0].replace('star-rating', '').strip() if row.xpath(ratingPath) else ""

        # Add to dataset if title is not missing
        if title:
            dataSet.append([title, price, availability, image, rating, link])

    print("Rows in Dataset:", len(dataSet))
    page += 1  # Increment page for loop

# Print total number of elements collected
print(f"Total items collected: {len(dataSet)}")

# Save the dataset to a CSV file
writeto_csv(dataSet, 'books.csv', columns)
```



```
TotalPages found: 1
Processing Page 1 from 1
Rows in Dataset: 20
Total items collected: 20
```

Book.csv



webScrapping > books.csv

```
title,price,stock,imageUrl,rating,url
Birdsong: A Story in Pictures,£54.64,,,Three,http://books.toscrape.com/catalogue/birdsong-a-story-in-pictures_975/index.html
The Bear and the Piano,£36.89,,,One,http://books.toscrape.com/catalogue/the-bear-and-the-piano_967/index.html
The Secret of Dreadwillow Carse,£56.13,,,One,http://books.toscrape.com/catalogue/the-secret-of-dreadwillow-carse_944/index.html
The White Cat and the Monk: A Retelling of the Poem “Pangur Bán”,£58.08,,,Four,http://books.toscrape.com/catalogue/the-white-cat-and-the-monk-a-retelling-of-the-poem-pangur-ban_865
Little Red,£13.47,,,Three,http://books.toscrape.com/catalogue/little-red_817/index.html
Walt Disney's Alice in Wonderland,£12.96,,,Five,http://books.toscrape.com/catalogue/walt-disneys-alice-in-wonderland_777/index.html
Twenty Yawns,£22.08,,,Two,http://books.toscrape.com/catalogue/twenty-yawns_773/index.html
Rain Fish,£23.57,,,Three,http://books.toscrape.com/catalogue/rain-fish_728/index.html
Once Was a Time,£18.28,,,Two,http://books.toscrape.com/catalogue/once-was-a-time_724/index.html
Luis Paints the World,£53.95,,,Three,http://books.toscrape.com/catalogue/luis-paints-the-world_714/index.html
Nap-a-Roo,£25.08,,,One,http://books.toscrape.com/catalogue/nap-a-roo_567/index.html
The Whale,£35.96,,,Four,http://books.toscrape.com/catalogue/the-whale_501/index.html
"Shrunken Treasures: Literary Classics, Short, Sweet, and Silly",£52.87,,,Three,http://books.toscrape.com/catalogue/shrunken-treasures-literary-classics-short-sweet-and-silly_484/i
Raymie Nightingale,£34.41,,,Two,http://books.toscrape.com/catalogue/raymie-nightingale_482/index.html
Playing from the Heart,£32.38,,,One,http://books.toscrape.com/catalogue/playing-from-the-heart_481/index.html
Maybe Something Beautiful: How Art Transformed a Neighborhood,£22.54,,,One,http://books.toscrape.com/catalogue/maybe-something-beautiful-how-art-transformed-a-neighborhood_386/inde
The Wild Robot,£56.07,,,Three,http://books.toscrape.com/catalogue/the-wild-robot_288/index.html
The Thing About Jellyfish,£48.77,,,One,http://books.toscrape.com/catalogue/the-thing-about-jellyfish_283/index.html
The Lonely Ones,£43.59,,,Five,http://books.toscrape.com/catalogue/the-lonely-ones_261/index.html
The Day the Crayons Came Home (Crayons),£26.33,,,Five,http://books.toscrape.com/catalogue/the-day-the-crayons-came-home-crayons_241/index.html

```