

In [1]:

```
pip install bs4
```

Requirement already satisfied: bs4 in c:\users\asus\anaconda3\lib\site-packages (0.0.1)

Requirement already satisfied: beautifulsoup4 in c:\users\asus\anaconda3\lib\site-packages (from bs4) (4.11.1)

Requirement already satisfied: soupsieve>1.2 in c:\users\asus\anaconda3\lib\site-packages (from beautifulsoup4->bs4) (2.3.1)

Note: you may need to restart the kernel to use updated packages.

In [2]:

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

In [4]:

```
url="https://books.toscrape.com/catalogue/category/books/fiction_10/page-2.html"
```

In [5]:

```
response=requests.get(url)
```

In [6]:

```
response=response.content
```

In [7]:

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

In [8]:

```
soup
```

Out[8]:

```
<!DOCTYPE html>
```

```
<!--[if lt IE 7]>      <html lang="en-us" class="no-js lt-ie9 lt-ie8 lt-
ie7"> <![endif]-->
```

```
<!--[if IE 7]>        <html lang="en-us" class="no-js lt-ie9 lt-ie8">
<![endif]-->
```

```
<!--[if IE 8]>        <html lang="en-us" class="no-js lt-ie9"> <![endi
f]-->
```

```
<!--[if gt IE 8]><!--> <html class="no-js" lang="en-us"> <!--<![endif]--
>
```

```
<head>
```

```
<title>
```

```
    Fiction |
```

```
    Books to Scrape - Sandbox
```

```
</title>
```

```
<meta content="text/html: charset=utf-8" http-equiv="content-type"/>
```

In [9]:

```
ol=soup.find('ol')
```

In [10]:

```
articles = ol.find_all('article', class_='product_pod')
```

In [11]:

```
print(articles)

[<article class="product_pod">
<div class="image_container">
<a href="../../eight-hundred-grapes_690/index.html"></a>
</div>
<p class="star-rating Four">
<i class="icon-star"></i>
<i class="icon-star"></i>
<i class="icon-star"></i>
<i class="icon-star"></i>
<i class="icon-star"></i>
</p>
<h3><a href="../../eight-hundred-grapes_690/index.html" title="Eight
Hundred Grapes">Eight Hundred Grapes</a></h3>
<div class="product_price">
<p class="price_color">£14.39</p>
<p class="instock availability">
<i class="icon-ok"></i>
```

In [12]:

```
# extract the things

for article in articles:
    image = article.find('img')
    title = image.attrs['alt']
    print(title)
```

```
Eight Hundred Grapes
Dear Mr. Knightley
Cometh the Hour (The Clifton Chronicles #6)
Balloon Animals
A Man Called Ove
The Silent Sister (Riley MacPherson #1)
The Dinner Party
The Improbability of Love
Mothering Sunday
Lies and Other Acts of Love
Daredevils
11/22/63
The Shack
The High Mountains of Portugal
Miller's Valley
Hystopia: A Novel
The Bourne Identity (Jason Bourne #1)
Sister Dear
Memoirs of a Geisha
Me Before You (Me Before You #1)
```

In [13]:

```
# find star_ratinng

for article in articles:
    star=article.find('p')
    star=star.attrs['class'][1]
    print(star)
```

Four
Five
Three
Three
One
Five
Two
One
Two
One
Three
Three
One
One
Two
Four
Four
Four
Three
One

In [14]:

```
# find the price of data

for article in articles:
    price=article.find('p',class_="price_color").text
    print(price)
```

£14.39
£11.21
£25.01
£17.03
£39.72
£46.29
£56.54
£59.45
£13.34
£45.14
£16.34
£48.48
£28.03
£51.15
£58.54
£21.96
£42.78
£40.20
£49.67
£19.02

In [15]:

```
# remove the douller simble
for article in articles:
    price=article.find('p',class_="price_color").text
    price=float(price[1:])
    print(price)
```

14.39
11.21
25.01
17.03
39.72
46.29
56.54
59.45
13.34
45.14
16.34
48.48
28.03
51.15
58.54
21.96
42.78
40.2
49.67
19.02

In []:

```
# scrapping all pages of Fiction
```

In [16]:

```
book=[]
for page in range(1, 5):
    url = f"https://books.toscrape.com/catalogue/page-{page}.html"
    response = requests.get(url)
    soup = BeautifulSoup(response.content, 'html.parser')
    ol = soup.find('ol')
    articles = ol.find_all('article', class_='product_pod')

    for article in articles:
        image = article.find('img')
        title = image.attrs['alt']
        star = article.find('p')
        star = star['class'][1]
        price = article.find('p', class_="price_color").text
        price = float(price[1:])
        book.append([title, star, price])

print(book)
```

```
[['A Light in the Attic', 'Three', 51.77], ['Tipping the Velvet', 'One', 53.74], ['Soumission', 'One', 50.1], ['Sharp Objects', 'Four', 47.82], ['Sa piens: A Brief History of Humankind', 'Five', 54.23], ['The Requiem Red', 'One', 22.65], ['The Dirty Little Secrets of Getting Your Dream Job', 'Four', 33.34], ['The Coming Woman: A Novel Based on the Life of the Infamous Feminist, Victoria Woodhull', 'Three', 17.93], ['The Boys in the Boat: Nine Americans and Their Epic Quest for Gold at the 1936 Berlin Olympics', 'Four', 22.6], ['The Black Maria', 'One', 52.15], ['Starving Hearts (Triangular Trade Trilogy, #1)', 'Two', 13.99], ["Shakespeare's Sonnets", 'Four', 20.66], ['Set Me Free', 'Five', 17.46], ["Scott Pilgrim's Precious Little Life (Scott Pilgrim #1)", 'Five', 52.29], ['Rip it Up and Start Again', 'Five', 35.02], ['Our Band Could Be Your Life: Scenes from the American Indie Underground, 1981-1991', 'Three', 57.25], ['Olio', 'One', 23.88], ['Mesaerion: The Best Science Fiction Stories 1800-1849', 'One', 37.59], ['Liber tarianism for Beginners', 'Two', 51.33], ["It's Only the Himalayas", 'Two', 45.17], ['In Her Wake', 'One', 12.84], ['How Music Works', 'Two', 37.32], ['Foolproof Preserving: A Guide to Small Batch Jams, Jellies, Pickles, Condiments, and More: A Foolproof Guide to Making Small Batch Jams, Jellies, Pickles, Condiments, and More', 'Three', 30.52], ['Chase Me (Paris Nights #2)', 'Five', 25.27], ['Black Dust', 'Five', 34.53], ['Birdsong: A Story in Pictures', 'Three', 54.64], ["America's Cradle of Quarterbacks: Western Pennsylvania's Football Factory from Johnny Unitas to Joe Montana", 'Three', 22.5], ['Aladdin and His Wonderful Lamp', 'Three', 53.13], ['Worlds Elsewhere: Journeys Around Shakespeare's Globe', 'Five', 40.3], ['Wall and Piece', 'Four', 44.18], ['The Four Agreements: A Practical Guide to Personal Freedom', 'Five', 17.66], ['The Five Love Languages: How to Express Heartfelt Commitment to Your Mate', 'Three', 31.05], ['The Elephant Tree', 'Five', 23.82], ['The Bear and the Piano', 'One', 36.89], ["Sophie's World", 'Five', 15.94], ['Penny Maybe', 'Three', 33.29], ['Maude (1883-1993): She Grew Up with the country', 'Two', 18.02], ['In a Dark, Dark Wood', 'One', 19.63], ['Behind Closed Doors', 'Four', 52.22], ["You can't bury them all: Poems", 'Two', 33.63], ['Slow States of Collapse: Poems', 'Three', 57.31], ['Reasons to Stay Alive', 'Two', 26.41], ['Private Paris (Private #10)', 'Five', 47.61], ['#HigherSelfie: Wake Up Your Life. Free Your Soul. Find Your Tribe.', 'Five', 23.11], ['Without Borders (Wanderlove #1)', 'Two', 45.07], ['When We Collided', 'One', 31.77], ['We Love You, Charlie Freeman', 'Five', 50.27], ['Untitled Collection: Sabbath Poems 2014', 'Four', 14.27], ['Unseen City: The Majesty of Pigeons, the Discreet Charm of Snails & Other Wonders of the Urban Wilderness', 'Four', 44.18], ['Unicorn Tracks', 'Three', 18.78], ['Unbound: How Eight Technologies Made Us Human, Transformed Society, and Brought Our World to the Brink', 'One', 25.52], ['Tsubasa: WORLD CHRONICLE 2 (Tsubasa WORLD CHRONICLE #2)', 'One', 16.28], ['Throwing Rocks at the Google Bus: How Growth Became the Enemy of Prosperity', 'Three', 31.12], ['This One Summer', 'Four', 19.49], ['Thirst', 'Five', 17.27], ['The Torch Is Passed: A Harding Family Story', 'One', 19.09], ['The Secret of Dreadwillow Carse', 'One', 56.13], ['The Pioneer Woman Cooks: Dinnertime: Comfort Classics, Freezer Food, 16-Minute Meals, and Other Delicious Ways to Solve Supper!', 'One', 56.41], ['The Past Never Ends', 'Four', 56.5], ['The Natural History of Us (The Fine Art of Pretending #2)', 'Three', 45.22], ['The Nameless City (The Nameless City #1)', 'Four', 38.16], ['The Murder That Never Was (Forensic Instincts #5)', 'Three', 54.11], ["The Most Perfect Thing: Inside (and Outside) a Bird's Egg", 'Four', 42.96], ['The Mindfulness and Acceptance Workbook for Anxiety: A Guide to Breaking Free from Anxiety, Phobias, and Worry Using Acceptance and Commitment Therapy', 'Four', 23.89], ['The Life-Changing Magic of Tidying Up: The Japanese Art of Decluttering and Organizing', 'Three', 16.77], ['The Inefficiency Assassin: Time Management Tactics for Working Smarter, Not Longer', 'Five', 20.59], ['The Gutsy Girl: Escapades for Your Life of Epic Adventure', 'One', 37.13], ['The Electric Pencil: Drawings from Inside State Hospital No. 3', 'One', 56.06], ['The Death of Humanity: and the Case for Life', 'Four', 58.11], ['The Bulletproof Diet: Lose up to a Pound a Day, Re
```

```
claim Energy and Focus, Upgrade Your Life', 'Three', 49.05], ['The Art Forger', 'Three', 40.76], ['The Age of Genius: The Seventeenth Century and the Birth of the Modern Mind', 'One', 19.73], ['The Activist's Tao Te Ching: Ancient Advice for a Modern Revolution', 'Five', 32.24], ['Spark Joy: An Illustrated Master Class on the Art of Organizing and Tidying Up', 'Four', 41.83], ['Soul Reader', 'Two', 39.58], ['Security', 'Two', 39.25], ['Saga, Volume 6 (Saga (Collected Editions) #6)', 'Three', 25.02], ['Saga, Volume 5 (Saga (Collected Editions) #5)', 'Two', 51.04], ['Reskilling America: Learning to Labor in the Twenty-First Century', 'Two', 19.83], ['Rat Queens, Vol. 3: Demons (Rat Queens (Collected Editions) #11-15)', 'Three', 50.4]]
```

In [17]:

```
df=pd.DataFrame(book,columns=['Title','Star','Price'])
```

In [18]:

```
df
```

Out[18]:

	Title	Star	Price
0	A Light in the Attic	Three	51.77
1	Tipping the Velvet	One	53.74
2	Soumission	One	50.10
3	Sharp Objects	Four	47.82
4	Sapiens: A Brief History of Humankind	Five	54.23
...
75	Security	Two	39.25
76	Saga, Volume 6 (Saga (Collected Editions) #6)	Three	25.02
77	Saga, Volume 5 (Saga (Collected Editions) #5)	Two	51.04
78	Reskilling America: Learning to Labor in the T...	Two	19.83
79	Rat Queens, Vol. 3: Demons (Rat Queens (Collec...	Three	50.40

80 rows × 3 columns

In [19]:

```
df.to_csv('internship_fiction.csv')
```

In []: