WEB SCRAPPING

BEAUTIFULSOUP

libraries:

    `-->requests`

    `-->bs4.BeautifulSoup`

In [801]:

```python
import requests #to make request to the required page
import bs4 #contains beautifulSoup for pulling data out of HTML and XML files
import pandas as pd #to convert our data to a dataframe
```

In [802]:

```python
ourdata=requests.get('https://en.wikipedia.org/wiki/Artificial_intelligence')
```

In [808]:

```python
#ourdata.text #ourdata stores the information of the entire page
```

BeautifulSoup

Beautiful Soup is a Python library for pulling data out of HTML and XML files.

Read more of BeautifulSoup here: https://www.crummy.com/software/BeautifulSoup/bs4/doc/ (https://www.crummy.com/software/BeautifulSoup/bs4/doc/)

In [812]:

```python
 bsp=bs4.BeautifulSoup(ourdata.text,'lxml') #takes 2 parameters, your request va
riable and the second one is how you want to structure your data. e.g. html pars
er, lxml etc.

    #converting from object of type 'request to object of type beautiful soup'
```

In [815]:

```python
#bsp
```

In [ ]:

We can now do a lot of stuff with our beautifulsoup object

In [810]:

```python
titles=bsp.select('title')
```

In [816]:

```python
titles
```

Out[816]:

```
[<title>Artificial intelligence - Wikipedia</title>]
```

In [817]:

```python
titles[0].getText()
```

Out[817]:

```
'Artificial intelligence - Wikipedia'
```

Inspect the page to find which class contains the title headlines

In [818]:

```python
titles_class=bsp.select('.mw-headline') #extract the class that contains the tit
le, NB: use '.' for selecting class names,'#' for selecting ID
```

In [819]:

```
titles_class
```

Out[819]:

```
[<span class="mw-headline" id="History">History</span>,
 <span class="mw-headline" id="Definitions">Definitions</span>,
 <span class="mw-headline" id="Basics">Basics</span>,
 <span class="mw-headline" id="Challenges">Challenges</span>,
 <span class="mw-headline" id="Reasoning,_problem_solving">Reasonin
g, problem solving</span>,
 <span class="mw-headline" id="Knowledge_representation">Knowledge r
epresentation</span>,
 <span class="mw-headline" id="Planning">Planning</span>,
 <span class="mw-headline" id="Learning">Learning</span>,
 <span class="mw-headline" id="Natural_language_processing">Natural
language processing</span>,
 <span class="mw-headline" id="Perception">Perception</span>,
 <span class="mw-headline" id="Motion_and_manipulation">Motion and m
anipulation</span>,
 <span class="mw-headline" id="Social_intelligence">Social intellige
nce</span>,
 <span class="mw-headline" id="General_intelligence">General intelli
gence</span>,
 <span class="mw-headline" id="Approaches">Approaches</span>,
 <span class="mw-headline" id="Cybernetics_and_brain_simulation">Cyb
ernetics and brain simulation</span>,
 <span class="mw-headline" id="Symbolic">Symbolic</span>,
 <span class="mw-headline" id="Cognitive_simulation">Cognitive simul
ation</span>,
 <span class="mw-headline" id="Logic-based">Logic-based</span>,
 <span class="mw-headline" id="Anti-logic_or_scruffy">Anti-logic or
scruffy</span>,
 <span class="mw-headline" id="Knowledge-based">Knowledge-based</spa
n>,
 <span class="mw-headline" id="Sub-symbolic">Sub-symbolic</span>,
 <span class="mw-headline" id="Embodied_intelligence">Embodied intel
ligence</span>,
 <span class="mw-headline" id="Computational_intelligence_and_soft_c
omputing">Computational intelligence and soft computing</span>,
 <span class="mw-headline" id="Statistical_learning">Statistical lea
rning</span>,
 <span class="mw-headline" id="Integrating_the_approaches">Integrati
ng the approaches</span>,
 <span class="mw-headline" id="Tools">Tools</span>,
 <span class="mw-headline" id="Search_and_optimization">Search and o
ptimization</span>,
 <span class="mw-headline" id="Logic">Logic</span>,
 <span class="mw-headline" id="Probabilistic_methods_for_uncertain_r
easoning">Probabilistic methods for uncertain reasoning</span>,
 <span class="mw-headline" id="Classifiers_and_statistical_learning_
methods">Classifiers and statistical learning methods</span>,
 <span class="mw-headline" id="Artificial_neural_networks">Artificia
l neural networks</span>,
 <span class="mw-headline" id="Deep_feedforward_neural_networks">Dee
p feedforward neural networks</span>,
 <span class="mw-headline" id="Deep_recurrent_neural_networks">Deep
recurrent neural networks</span>,
 <span class="mw-headline" id="Evaluating_progress">Evaluating progr
ess</span>,
 <span class="mw-headline" id="Applications">Applications<span id="G
oals"></span></span>,
 <span class="mw-headline" id="Healthcare">Healthcare</span>,
 <span class="mw-headline" id="Automotive">Automotive</span>,
```

```
 <span class="mw-headline" id="Finance_and_economics">Finance and ec
onomics</span>,
 <span class="mw-headline" id="Cybersecurity">Cybersecurity</span>,
 <span class="mw-headline" id="Government">Government</span>,
 <span class="mw-headline" id="Law-related_professions">Law-related
professions</span>,
 <span class="mw-headline" id="Video_games">Video games</span>,
 <span class="mw-headline" id="Military">Military</span>,
 <span class="mw-headline" id="Hospitality">Hospitality</span>,
 <span class="mw-headline" id="Audit">Audit</span>,
 <span class="mw-headline" id="Advertising">Advertising</span>,
 <span class="mw-headline" id="Art">Art</span>,
 <span class="mw-headline" id="Philosophy_and_ethics">Philosophy and
ethics</span>,
 <span class="mw-headline" id="The_limits_of_artificial_general_inte
lligence">The limits of artificial general intelligence</span>,
 <span class="mw-headline" id="Potential_harm">Potential harm<span i
d="Potential_risks_and_moral_reasoning"></span></span>,
 <span class="mw-headline" id="Existential_risk">Existential risk</s
pan>,
 <span class="mw-headline" id="Devaluation_of_humanity">Devaluation
of humanity</span>,
 <span class="mw-headline" id="Social_justice">Social justice</span
>,
 <span class="mw-headline" id="Decrease_in_demand_for_human_labor">D
ecrease in demand for human labor</span>,
 <span class="mw-headline" id="Autonomous_weapons">Autonomous weapon
s</span>,
 <span class="mw-headline" id="Ethical_machines">Ethical machines</s
pan>,
 <span class="mw-headline" id="Artificial_moral_agents">Artificial m
oral agents</span>,
 <span class="mw-headline" id="Machine_ethics">Machine ethics</span
>,
 <span class="mw-headline" id="Malevolent_and_friendly_AI">Malevolen
t and friendly AI</span>,
 <span class="mw-headline" id="Machine_consciousness,_sentience_and_
mind">Machine consciousness, sentience and mind</span>,
 <span class="mw-headline" id="Consciousness">Consciousness</span>,
 <span class="mw-headline" id="Computationalism_and_functionalism">C
omputationalism and functionalism</span>,
 <span class="mw-headline" id="Strong_AI_hypothesis">Strong AI hypot
hesis</span>,
 <span class="mw-headline" id="Robot_rights">Robot rights</span>,
 <span class="mw-headline" id="Superintelligence">Superintelligence
</span>,
 <span class="mw-headline" id="Technological_singularity">Technologi
cal singularity</span>,
 <span class="mw-headline" id="Transhumanism">Transhumanism</span>,
 <span class="mw-headline" id="Economics">Economics</span>,
 <span class="mw-headline" id="In_fiction">In fiction</span>,
 <span class="mw-headline" id="See_also">See also</span>,
 <span class="mw-headline" id="Explanatory_notes">Explanatory notes
</span>,
 <span class="mw-headline" id="References">References</span>,
 <span class="mw-headline" id="AI_textbooks">AI textbooks</span>,
 <span class="mw-headline" id="History_of_AI">History of AI</span>,
 <span class="mw-headline" id="Other_sources">Other sources</span>,
 <span class="mw-headline" id="Further_reading">Further reading</spa
n>,
```

```
  <span class="mw-headline" id="External_links">External links</span>
>]
```

Select only the titles

In [820]:

```python
for i in titles_class:
    print(i.text) #to get only the text, NB: Find the difference between ".getTe
xt()" and ".text"
```

```
History
Definitions
Basics
Challenges
Reasoning, problem solving
Knowledge representation
Planning
Learning
Natural language processing
Perception
Motion and manipulation
Social intelligence
General intelligence
Approaches
Cybernetics and brain simulation
Symbolic
Cognitive simulation
Logic-based
Anti-logic or scruffy
Knowledge-based
Sub-symbolic
Embodied intelligence
Computational intelligence and soft computing
Statistical learning
Integrating the approaches
Tools
Search and optimization
Logic
Probabilistic methods for uncertain reasoning
Classifiers and statistical learning methods
Artificial neural networks
Deep feedforward neural networks
Deep recurrent neural networks
Evaluating progress
Applications
Healthcare
Automotive
Finance and economics
Cybersecurity
Government
Law-related professions
Video games
Military
Hospitality
Audit
Advertising
Art
Philosophy and ethics
The limits of artificial general intelligence
Potential harm
Existential risk
Devaluation of humanity
Social justice
Decrease in demand for human labor
Autonomous weapons
Ethical machines
Artificial moral agents
Machine ethics
Malevolent and friendly AI
Machine consciousness, sentience and mind
Consciousness
```

```
Computationalism and functionalism
Strong AI hypothesis
Robot rights
Superintelligence
Technological singularity
Transhumanism
Economics
In fiction
See also
Explanatory notes
References
AI textbooks
History of AI
Other sources
Further reading
External links
```

In [821]:

```
titles_class[7].getText()
```

Out[821]:

```
'Learning'
```

We can as well use the span to acheive the same results

In [822]:

```
titles_class1=bsp.select('h2 > span') #to select all the span instead of class
```

In [824]:

```
#titles_class1
```

We can also select the links

In [825]:

```
links=bsp.select('link')
```

In [827]:

```
#links
```

In [829]:

```
#  for link in bsp.find_all('link'):
#     print(link)
```

In [ ]:

```

```

In [830]:

```
imgs=bsp.select('img')
```

In [831]:

```
imgs
```

In [831]:

```
imgs
```

Out[831]:

[<img alt="Page protected with pending changes" data-file-height="51
2" data-file-width="512" decoding="async" height="20" src="//upload.
wikimedia.org/wikipedia/en/thumb/b/b7/Pending-protection-shackle.sv
g/20px-Pending-protection-shackle.svg.png" srcset="//upload.wikimedi
a.org/wikipedia/en/thumb/b/b7/Pending-protection-shackle.svg/30px-Pe
nding-protection-shackle.svg.png 1.5x, //upload.wikimedia.org/wikipe
dia/en/thumb/b/b7/Pending-protection-shackle.svg/40px-Pending-protec
tion-shackle.svg.png 2x" width="20"/>,
 <img alt="" class="thumbimage" data-file-height="1020" data-file-wi
dth="1020" decoding="async" height="220" src="//upload.wikimedia.or
g/wikipedia/commons/thumb/c/c8/Didrachm_Phaistos_obverse_CdM.jpg/220
px-Didrachm_Phaistos_obverse_CdM.jpg" srcset="//upload.wikimedia.or
g/wikipedia/commons/thumb/c/c8/Didrachm_Phaistos_obverse_CdM.jpg/330
px-Didrachm_Phaistos_obverse_CdM.jpg 1.5x, //upload.wikimedia.org/wi
kipedia/commons/thumb/c/c8/Didrachm_Phaistos_obverse_CdM.jpg/440px-D
idrachm_Phaistos_obverse_CdM.jpg 2x" width="220"/>,
 <img alt="" class="thumbimage" data-file-height="256" data-file-wid
th="377" decoding="async" height="149" src="//upload.wikimedia.org/w
ikipedia/commons/thumb/6/68/Overfitted_Data.png/220px-Overfitted_Dat
a.png" srcset="//upload.wikimedia.org/wikipedia/commons/thumb/6/68/O
verfitted_Data.png/330px-Overfitted_Data.png 1.5x, //upload.wikimedi
a.org/wikipedia/commons/6/68/Overfitted_Data.png 2x" width="220"/>,
 <img alt="" class="thumbimage" data-file-height="533" data-file-wid
th="800" decoding="async" height="147" src="//upload.wikimedia.org/w
ikipedia/commons/thumb/f/f5/D%C3%A9tection_de_personne_-_exemple_3.j
pg/220px-D%C3%A9tection_de_personne_-_exemple_3.jpg" srcset="//uploa
d.wikimedia.org/wikipedia/commons/thumb/f/f5/D%C3%A9tection_de_perso
nne_-_exemple_3.jpg/330px-D%C3%A9tection_de_personne_-_exemple_3.jpg
1.5x, //upload.wikimedia.org/wikipedia/commons/thumb/f/f5/D%C3%A9tec
tion_de_personne_-_exemple_3.jpg/440px-D%C3%A9tection_de_personne_-_
exemple_3.jpg 2x" width="220"/>,
 <img alt="" class="thumbimage" data-file-height="443" data-file-wid
th="469" decoding="async" height="208" src="//upload.wikimedia.org/w
ikipedia/commons/thumb/9/9e/GFO_taxonomy_tree.png/220px-GFO_taxonomy
_tree.png" srcset="//upload.wikimedia.org/wikipedia/commons/thumb/9/
9e/GFO_taxonomy_tree.png/330px-GFO_taxonomy_tree.png 1.5x, //upload.
wikimedia.org/wikipedia/commons/thumb/9/9e/GFO_taxonomy_tree.png/440
px-GFO_taxonomy_tree.png 2x" width="220"/>,
 <img alt="" class="thumbimage" data-file-height="539" data-file-wid
th="596" decoding="async" height="199" src="//upload.wikimedia.org/w
ikipedia/commons/thumb/7/76/Hierarchical-control-system.svg/220px-Hi
erarchical-control-system.svg.png" srcset="//upload.wikimedia.org/wi
kipedia/commons/thumb/7/76/Hierarchical-control-system.svg/330px-Hie
rarchical-control-system.svg.png 1.5x, //upload.wikimedia.org/wikipe
dia/commons/thumb/7/76/Hierarchical-control-system.svg/440px-Hierarc
hical-control-system.svg.png 2x" width="220"/>,
 <img alt="" class="thumbimage" data-file-height="120" data-file-wid
th="120" decoding="async" height="220" src="//upload.wikimedia.org/w
ikipedia/commons/thumb/6/6e/ParseTree.svg/220px-ParseTree.svg.png" s
rcset="//upload.wikimedia.org/wikipedia/commons/thumb/6/6e/ParseTre
e.svg/330px-ParseTree.svg.png 1.5x, //upload.wikimedia.org/wikipedi
a/commons/thumb/6/6e/ParseTree.svg/440px-ParseTree.svg.png 2x" width
="220"/>,
 <img alt="" class="thumbimage" data-file-height="471" data-file-wid
th="1415" decoding="async" height="73" src="//upload.wikimedia.org/w
ikipedia/commons/thumb/2/20/%C3%84%C3%A4retuvastuse_n%C3%A4ide.png/2
20px-%C3%84%C3%A4retuvastuse_n%C3%A4ide.png" srcset="//upload.wikime
dia.org/wikipedia/commons/thumb/2/20/%C3%84%C3%A4retuvastuse_n%C3%A4
ide.png/330px-%C3%84%C3%A4retuvastuse_n%C3%A4ide.png 1.5x, //upload.

```
wikimedia.org/wikipedia/commons/thumb/2/20/%C3%84%C3%A4retuvastuse_
n%C3%A4ide.png/440px-%C3%84%C3%A4retuvastuse_n%C3%A4ide.png 2x" widt
h="220"/>,
 <img alt="" class="thumbimage" data-file-height="1704" data-file-wi
dth="2272" decoding="async" height="165" src="//upload.wikimedia.or
g/wikipedia/commons/thumb/b/b8/Kismet_robot_at_MIT_Museum.jpg/220px-
Kismet_robot_at_MIT_Museum.jpg" srcset="//upload.wikimedia.org/wikip
edia/commons/thumb/b/b8/Kismet_robot_at_MIT_Museum.jpg/330px-Kismet_
robot_at_MIT_Museum.jpg 1.5x, //upload.wikimedia.org/wikipedia/commo
ns/thumb/b/b8/Kismet_robot_at_MIT_Museum.jpg/440px-Kismet_robot_at_M
IT_Museum.jpg 2x" width="220"/>,
 <img alt="" class="thumbimage" data-file-height="420" data-file-wid
th="560" decoding="async" height="165" src="//upload.wikimedia.org/w
ikipedia/commons/thumb/e/ec/ParticleSwarmArrowsAnimation.gif/220px-P
articleSwarmArrowsAnimation.gif" srcset="//upload.wikimedia.org/wiki
pedia/commons/thumb/e/ec/ParticleSwarmArrowsAnimation.gif/330px-Part
icleSwarmArrowsAnimation.gif 1.5x, //upload.wikimedia.org/wikipedia/
commons/thumb/e/ec/ParticleSwarmArrowsAnimation.gif/440px-ParticleSw
armArrowsAnimation.gif 2x" width="220"/>,
 <img alt="" class="thumbimage" data-file-height="309" data-file-wid
th="360" decoding="async" height="309" src="//upload.wikimedia.org/w
ikipedia/commons/6/69/EM_Clustering_of_Old_Faithful_data.gif" width
="360"/>,
 <img alt="" class="thumbimage" data-file-height="500" data-file-wid
th="560" decoding="async" height="196" src="//upload.wikimedia.org/w
ikipedia/commons/thumb/e/e4/Artificial_neural_network.svg/220px-Arti
ficial_neural_network.svg.png" srcset="//upload.wikimedia.org/wikipe
dia/commons/thumb/e/e4/Artificial_neural_network.svg/330px-Artificia
l_neural_network.svg.png 1.5x, //upload.wikimedia.org/wikipedia/comm
ons/thumb/e/e4/Artificial_neural_network.svg/440px-Artificial_neural
_network.svg.png 2x" width="220"/>,
 <img alt="" class="thumbimage" data-file-height="501" data-file-wid
th="400" decoding="async" height="276" src="//upload.wikimedia.org/w
ikipedia/commons/thumb/8/8b/Automated_online_assistant.png/220px-Aut
omated_online_assistant.png" srcset="//upload.wikimedia.org/wikipedi
a/commons/thumb/8/8b/Automated_online_assistant.png/330px-Automated_
online_assistant.png 1.5x, //upload.wikimedia.org/wikipedia/commons/
8/8b/Automated_online_assistant.png 2x" width="220"/>,
 <img alt="" class="thumbimage" data-file-height="1296" data-file-wi
dth="891" decoding="async" height="320" src="//upload.wikimedia.org/
wikipedia/commons/thumb/0/0d/Laproscopic_Surgery_Robot.jpg/220px-Lap
roscopic_Surgery_Robot.jpg" srcset="//upload.wikimedia.org/wikipedi
a/commons/thumb/0/0d/Laproscopic_Surgery_Robot.jpg/330px-Laproscopic
_Surgery_Robot.jpg 1.5x, //upload.wikimedia.org/wikipedia/commons/th
umb/0/0d/Laproscopic_Surgery_Robot.jpg/440px-Laproscopic_Surgery_Rob
ot.jpg 2x" width="220"/>,
 <img alt="" class="thumbimage" data-file-height="1827" data-file-wi
dth="1303" decoding="async" height="308" src="//upload.wikimedia.or
g/wikipedia/commons/thumb/5/5e/X-ray_of_a_hand_with_automatic_bone_a
ge_calculation.jpg/220px-X-ray_of_a_hand_with_automatic_bone_age_cal
culation.jpg" srcset="//upload.wikimedia.org/wikipedia/commons/thum
b/5/5e/X-ray_of_a_hand_with_automatic_bone_age_calculation.jpg/330px
-X-ray_of_a_hand_with_automatic_bone_age_calculation.jpg 1.5x, //upl
oad.wikimedia.org/wikipedia/commons/thumb/5/5e/X-ray_of_a_hand_with_
automatic_bone_age_calculation.jpg/440px-X-ray_of_a_hand_with_automa
tic_bone_age_calculation.jpg 2x" width="220"/>,
 <img alt="" data-file-height="399" data-file-width="512" decoding
="async" height="39" src="//upload.wikimedia.org/wikipedia/en/thumb/
9/99/Question_book-new.svg/50px-Question_book-new.svg.png" srcset
="//upload.wikimedia.org/wikipedia/en/thumb/9/99/Question_book-new.s
vg/75px-Question_book-new.svg.png 1.5x, //upload.wikimedia.org/wikip
```

```
edia/en/thumb/9/99/Question_book-new.svg/100px-Question_book-new.sv
g.png 2x" width="50"/>,
 <img alt="" class="thumbimage" data-file-height="316" data-file-wid
th="566" decoding="async" height="123" src="//upload.wikimedia.org/w
ikipedia/commons/thumb/8/87/Capek_play.jpg/220px-Capek_play.jpg" src
set="//upload.wikimedia.org/wikipedia/commons/thumb/8/87/Capek_play.
jpg/330px-Capek_play.jpg 1.5x, //upload.wikimedia.org/wikipedia/comm
ons/thumb/8/87/Capek_play.jpg/440px-Capek_play.jpg 2x" width="220"/
>,
 <img alt="icon" class="noviewer" data-file-height="1024" data-file-
width="896" decoding="async" height="28" src="//upload.wikimedia.or
g/wikipedia/commons/thumb/6/6f/Octicons-terminal.svg/24px-Octicons-t
erminal.svg.png" srcset="//upload.wikimedia.org/wikipedia/commons/th
umb/6/6f/Octicons-terminal.svg/37px-Octicons-terminal.svg.png 1.5x,
//upload.wikimedia.org/wikipedia/commons/thumb/6/6f/Octicons-termina
l.svg/49px-Octicons-terminal.svg.png 2x" width="24"/>,
 <img alt="closed access" data-file-height="1000" data-file-width="6
40" decoding="async" height="14" src="//upload.wikimedia.org/wikiped
ia/commons/thumb/0/0e/Closed_Access_logo_transparent.svg/9px-Closed_
Access_logo_transparent.svg.png" srcset="//upload.wikimedia.org/wiki
pedia/commons/thumb/0/0e/Closed_Access_logo_transparent.svg/14px-Clo
sed_Access_logo_transparent.svg.png 1.5x, //upload.wikimedia.org/wik
ipedia/commons/thumb/0/0e/Closed_Access_logo_transparent.svg/18px-Cl
osed_Access_logo_transparent.svg.png 2x" width="9"/>,
 <img alt="" data-file-height="391" data-file-width="391" decoding
="async" height="27" src="//upload.wikimedia.org/wikipedia/en/thumb/
0/06/Wiktionary-logo-v2.svg/27px-Wiktionary-logo-v2.svg.png" srcset
="//upload.wikimedia.org/wikipedia/en/thumb/0/06/Wiktionary-logo-v2.
svg/41px-Wiktionary-logo-v2.svg.png 1.5x, //upload.wikimedia.org/wik
ipedia/en/thumb/0/06/Wiktionary-logo-v2.svg/54px-Wiktionary-logo-v2.
svg.png 2x" style="vertical-align: middle" width="27"/>,
 <img alt="" data-file-height="1376" data-file-width="1024" decoding
="async" height="27" src="//upload.wikimedia.org/wikipedia/en/thumb/
4/4a/Commons-logo.svg/20px-Commons-logo.svg.png" srcset="//upload.wi
kimedia.org/wikipedia/en/thumb/4/4a/Commons-logo.svg/30px-Commons-lo
go.svg.png 1.5x, //upload.wikimedia.org/wikipedia/en/thumb/4/4a/Comm
ons-logo.svg/40px-Commons-logo.svg.png 2x" style="vertical-align: mi
ddle" width="20"/>,
 <img alt="" data-file-height="355" data-file-width="300" decoding
="async" height="27" src="//upload.wikimedia.org/wikipedia/commons/t
humb/f/fa/Wikiquote-logo.svg/23px-Wikiquote-logo.svg.png" srcset="//
upload.wikimedia.org/wikipedia/commons/thumb/f/fa/Wikiquote-logo.sv
g/35px-Wikiquote-logo.svg.png 1.5x, //upload.wikimedia.org/wikipedi
a/commons/thumb/f/fa/Wikiquote-logo.svg/46px-Wikiquote-logo.svg.png
2x" style="vertical-align: middle" width="23"/>,
 <img alt="" data-file-height="300" data-file-width="300" decoding
="async" height="27" src="//upload.wikimedia.org/wikipedia/commons/t
humb/f/fa/Wikibooks-logo.svg/27px-Wikibooks-logo.svg.png" srcset="//
upload.wikimedia.org/wikipedia/commons/thumb/f/fa/Wikibooks-logo.sv
g/41px-Wikibooks-logo.svg.png 1.5x, //upload.wikimedia.org/wikipedi
a/commons/thumb/f/fa/Wikibooks-logo.svg/54px-Wikibooks-logo.svg.png
2x" style="vertical-align: middle" width="27"/>,
 <img alt="" data-file-height="900" data-file-width="1000" decoding
="async" height="24" src="//upload.wikimedia.org/wikipedia/commons/t
humb/1/1b/Wikiversity-logo-en.svg/27px-Wikiversity-logo-en.svg.png"
srcset="//upload.wikimedia.org/wikipedia/commons/thumb/1/1b/Wikivers
ity-logo-en.svg/41px-Wikiversity-logo-en.svg.png 1.5x, //upload.wiki
media.org/wikipedia/commons/thumb/1/1b/Wikiversity-logo-en.svg/54px-
Wikiversity-logo-en.svg.png 2x" style="vertical-align: middle" width
="27"/>,
 <img alt="" data-file-height="590" data-file-width="1050" decoding
```

```
="async" height="15" src="//upload.wikimedia.org/wikipedia/commons/t
humb/f/ff/Wikidata-logo.svg/27px-Wikidata-logo.svg.png" srcset="//up
load.wikimedia.org/wikipedia/commons/thumb/f/ff/Wikidata-logo.svg/41
px-Wikidata-logo.svg.png 1.5x, //upload.wikimedia.org/wikipedia/comm
ons/thumb/f/ff/Wikidata-logo.svg/54px-Wikidata-logo.svg.png 2x" styl
e="vertical-align: middle" width="27"/>,
 <img alt="Socrates.png" class="noviewer" data-file-height="500" dat
a-file-width="326" decoding="async" height="28" src="//upload.wikime
dia.org/wikipedia/commons/thumb/c/cd/Socrates.png/18px-Socrates.png"
srcset="//upload.wikimedia.org/wikipedia/commons/thumb/c/cd/Socrate
s.png/27px-Socrates.png 1.5x, //upload.wikimedia.org/wikipedia/commo
ns/thumb/c/cd/Socrates.png/36px-Socrates.png 2x" width="18"/>,
 <img alt="Socrates.png" class="noviewer" data-file-height="500" dat
a-file-width="326" decoding="async" height="28" src="//upload.wikime
dia.org/wikipedia/commons/thumb/c/cd/Socrates.png/18px-Socrates.png"
srcset="//upload.wikimedia.org/wikipedia/commons/thumb/c/cd/Socrate
s.png/27px-Socrates.png 1.5x, //upload.wikimedia.org/wikipedia/commo
ns/thumb/c/cd/Socrates.png/36px-Socrates.png 2x" width="18"/>,
 <img alt="Nuvola apps kalzium.svg" class="noviewer" data-file-heigh
t="128" data-file-width="128" decoding="async" height="28" src="//up
load.wikimedia.org/wikipedia/commons/thumb/8/8b/Nuvola_apps_kalzium.
svg/28px-Nuvola_apps_kalzium.svg.png" srcset="//upload.wikimedia.or
g/wikipedia/commons/thumb/8/8b/Nuvola_apps_kalzium.svg/42px-Nuvola_a
pps_kalzium.svg.png 1.5x, //upload.wikimedia.org/wikipedia/commons/t
humb/8/8b/Nuvola_apps_kalzium.svg/56px-Nuvola_apps_kalzium.svg.png 2
x" width="28"/>,
 <img alt="Wikipedia book" data-file-height="185" data-file-width="1
80" decoding="async" height="16" src="//upload.wikimedia.org/wikiped
ia/commons/thumb/8/89/Symbol_book_class2.svg/16px-Symbol_book_class
2.svg.png" srcset="//upload.wikimedia.org/wikipedia/commons/thumb/8/
89/Symbol_book_class2.svg/23px-Symbol_book_class2.svg.png 1.5x, //up
load.wikimedia.org/wikipedia/commons/thumb/8/89/Symbol_book_class2.s
vg/31px-Symbol_book_class2.svg.png 2x" title="Wikipedia book" width
="16"/>,
 <img alt="Category" data-file-height="31" data-file-width="36" deco
ding="async" height="14" src="//upload.wikimedia.org/wikipedia/en/th
umb/4/48/Folder_Hexagonal_Icon.svg/16px-Folder_Hexagonal_Icon.svg.pn
g" srcset="//upload.wikimedia.org/wikipedia/en/thumb/4/48/Folder_Hex
agonal_Icon.svg/24px-Folder_Hexagonal_Icon.svg.png 1.5x, //upload.wi
kimedia.org/wikipedia/en/thumb/4/48/Folder_Hexagonal_Icon.svg/32px-F
older_Hexagonal_Icon.svg.png 2x" title="Category" width="16"/>,
 <img alt="Outline" data-file-height="200" data-file-width="130" dec
oding="async" height="16" src="//upload.wikimedia.org/wikipedia/comm
ons/thumb/4/41/Global_thinking.svg/10px-Global_thinking.svg.png" src
set="//upload.wikimedia.org/wikipedia/commons/thumb/4/41/Global_thin
king.svg/15px-Global_thinking.svg.png 1.5x, //upload.wikimedia.org/w
ikipedia/commons/thumb/4/41/Global_thinking.svg/21px-Global_thinkin
g.svg.png 2x" title="Outline" width="10"/>,
 <img alt="WikiProject" data-file-height="100" data-file-width="100"
decoding="async" height="16" src="//upload.wikimedia.org/wikipedia/c
ommons/thumb/3/37/People_icon.svg/16px-People_icon.svg.png" srcset
="//upload.wikimedia.org/wikipedia/commons/thumb/3/37/People_icon.sv
g/24px-People_icon.svg.png 1.5x, //upload.wikimedia.org/wikipedia/co
mmons/thumb/3/37/People_icon.svg/32px-People_icon.svg.png 2x" title
="WikiProject" width="16"/>,
 <img alt="Commons page" data-file-height="1376" data-file-width="10
24" decoding="async" height="16" src="//upload.wikimedia.org/wikiped
ia/en/thumb/4/4a/Commons-logo.svg/12px-Commons-logo.svg.png" srcset
="//upload.wikimedia.org/wikipedia/en/thumb/4/4a/Commons-logo.svg/18
px-Commons-logo.svg.png 1.5x, //upload.wikimedia.org/wikipedia/en/th
umb/4/4a/Commons-logo.svg/24px-Commons-logo.svg.png 2x" title="Commo
```

```
ns page" width="12"/>,
 <img alt="Category" data-file-height="31" data-file-width="36" deco
ding="async" height="14" src="//upload.wikimedia.org/wikipedia/en/th
umb/4/48/Folder_Hexagonal_Icon.svg/16px-Folder_Hexagonal_Icon.svg.pn
g" srcset="//upload.wikimedia.org/wikipedia/en/thumb/4/48/Folder_Hex
agonal_Icon.svg/24px-Folder_Hexagonal_Icon.svg.png 1.5x, //upload.wi
kimedia.org/wikipedia/en/thumb/4/48/Folder_Hexagonal_Icon.svg/32px-F
older_Hexagonal_Icon.svg.png 2x" title="Category" width="16"/>,
 <img alt="List-Class article" data-file-height="185" data-file-widt
h="180" decoding="async" height="16" src="//upload.wikimedia.org/wik
ipedia/en/thumb/d/db/Symbol_list_class.svg/16px-Symbol_list_class.sv
g.png" srcset="//upload.wikimedia.org/wikipedia/en/thumb/d/db/Symbol
_list_class.svg/23px-Symbol_list_class.svg.png 1.5x, //upload.wikime
dia.org/wikipedia/en/thumb/d/db/Symbol_list_class.svg/31px-Symbol_li
st_class.svg.png 2x" title="List-Class article" width="16"/>,
 <img alt="Shadow Hand Bulb large.jpg" data-file-height="3840" data-
file-width="2560" decoding="async" height="150" src="//upload.wikime
dia.org/wikipedia/commons/thumb/c/c5/Shadow_Hand_Bulb_large.jpg/100p
x-Shadow_Hand_Bulb_large.jpg" srcset="//upload.wikimedia.org/wikiped
ia/commons/thumb/c/c5/Shadow_Hand_Bulb_large.jpg/150px-Shadow_Hand_B
ulb_large.jpg 1.5x, //upload.wikimedia.org/wikipedia/commons/thumb/
c/c5/Shadow_Hand_Bulb_large.jpg/200px-Shadow_Hand_Bulb_large.jpg 2x"
width="100"/>,
 <img alt="Category" data-file-height="31" data-file-width="36" deco
ding="async" height="14" src="//upload.wikimedia.org/wikipedia/en/th
umb/4/48/Folder_Hexagonal_Icon.svg/16px-Folder_Hexagonal_Icon.svg.pn
g" srcset="//upload.wikimedia.org/wikipedia/en/thumb/4/48/Folder_Hex
agonal_Icon.svg/24px-Folder_Hexagonal_Icon.svg.png 1.5x, //upload.wi
kimedia.org/wikipedia/en/thumb/4/48/Folder_Hexagonal_Icon.svg/32px-F
older_Hexagonal_Icon.svg.png 2x" title="Category" width="16"/>,
 <img alt="Outline" data-file-height="200" data-file-width="130" dec
oding="async" height="16" src="//upload.wikimedia.org/wikipedia/comm
ons/thumb/4/41/Global_thinking.svg/10px-Global_thinking.svg.png" src
set="//upload.wikimedia.org/wikipedia/commons/thumb/4/41/Global_thin
king.svg/15px-Global_thinking.svg.png 1.5x, //upload.wikimedia.org/w
ikipedia/commons/thumb/4/41/Global_thinking.svg/21px-Global_thinkin
g.svg.png 2x" title="Outline" width="10"/>,
 <img alt="Category" data-file-height="31" data-file-width="36" deco
ding="async" height="14" src="//upload.wikimedia.org/wikipedia/en/th
umb/4/48/Folder_Hexagonal_Icon.svg/16px-Folder_Hexagonal_Icon.svg.pn
g" srcset="//upload.wikimedia.org/wikipedia/en/thumb/4/48/Folder_Hex
agonal_Icon.svg/24px-Folder_Hexagonal_Icon.svg.png 1.5x, //upload.wi
kimedia.org/wikipedia/en/thumb/4/48/Folder_Hexagonal_Icon.svg/32px-F
older_Hexagonal_Icon.svg.png 2x" title="Category" width="16"/>,
 <img alt="Socrates.png" class="noviewer" data-file-height="500" dat
a-file-width="326" decoding="async" height="28" src="//upload.wikime
dia.org/wikipedia/commons/thumb/c/cd/Socrates.png/18px-Socrates.png"
srcset="//upload.wikimedia.org/wikipedia/commons/thumb/c/cd/Socrate
s.png/27px-Socrates.png 1.5x, //upload.wikimedia.org/wikipedia/commo
ns/thumb/c/cd/Socrates.png/36px-Socrates.png 2x" width="18"/>,
 <img alt="Wikipedia book" data-file-height="185" data-file-width="1
80" decoding="async" height="16" src="//upload.wikimedia.org/wikiped
ia/commons/thumb/8/89/Symbol_book_class2.svg/16px-Symbol_book_class
2.svg.png" srcset="//upload.wikimedia.org/wikipedia/commons/thumb/8/
89/Symbol_book_class2.svg/23px-Symbol_book_class2.svg.png 1.5x, //up
load.wikimedia.org/wikipedia/commons/thumb/8/89/Symbol_book_class2.s
vg/31px-Symbol_book_class2.svg.png 2x" title="Wikipedia book" width
="16"/>,
 <img alt="Category" data-file-height="31" data-file-width="36" deco
ding="async" height="14" src="//upload.wikimedia.org/wikipedia/en/th
umb/4/48/Folder_Hexagonal_Icon.svg/16px-Folder_Hexagonal_Icon.svg.pn
```

```
g" srcset="//upload.wikimedia.org/wikipedia/en/thumb/4/48/Folder_Hex
agonal_Icon.svg/24px-Folder_Hexagonal_Icon.svg.png 1.5x, //upload.wi
kimedia.org/wikipedia/en/thumb/4/48/Folder_Hexagonal_Icon.svg/32px-F
older_Hexagonal_Icon.svg.png 2x" title="Category" width="16"/>,
 <img alt="Outline" data-file-height="200" data-file-width="130" dec
oding="async" height="16" src="//upload.wikimedia.org/wikipedia/comm
ons/thumb/4/41/Global_thinking.svg/10px-Global_thinking.svg.png" src
set="//upload.wikimedia.org/wikipedia/commons/thumb/4/41/Global_thin
king.svg/15px-Global_thinking.svg.png 1.5x, //upload.wikimedia.org/w
ikipedia/commons/thumb/4/41/Global_thinking.svg/21px-Global_thinkin
g.svg.png 2x" title="Outline" width="10"/>,
 <img alt="WikiProject" data-file-height="100" data-file-width="100"
decoding="async" height="16" src="//upload.wikimedia.org/wikipedia/c
ommons/thumb/3/37/People_icon.svg/16px-People_icon.svg.png" srcset
="//upload.wikimedia.org/wikipedia/commons/thumb/3/37/People_icon.sv
g/24px-People_icon.svg.png 1.5x, //upload.wikimedia.org/wikipedia/co
mmons/thumb/3/37/People_icon.svg/32px-People_icon.svg.png 2x" title
="WikiProject" width="16"/>,
 <img alt="Commons page" data-file-height="1376" data-file-width="10
24" decoding="async" height="16" src="//upload.wikimedia.org/wikiped
ia/en/thumb/4/4a/Commons-logo.svg/12px-Commons-logo.svg.png" srcset
="//upload.wikimedia.org/wikipedia/en/thumb/4/4a/Commons-logo.svg/18
px-Commons-logo.svg.png 1.5x, //upload.wikimedia.org/wikipedia/en/th
umb/4/4a/Commons-logo.svg/24px-Commons-logo.svg.png 2x" title="Commo
ns page" width="12"/>,
 <img alt="Category" data-file-height="31" data-file-width="36" deco
ding="async" height="14" src="//upload.wikimedia.org/wikipedia/en/th
umb/4/48/Folder_Hexagonal_Icon.svg/16px-Folder_Hexagonal_Icon.svg.pn
g" srcset="//upload.wikimedia.org/wikipedia/en/thumb/4/48/Folder_Hex
agonal_Icon.svg/24px-Folder_Hexagonal_Icon.svg.png 1.5x, //upload.wi
kimedia.org/wikipedia/en/thumb/4/48/Folder_Hexagonal_Icon.svg/32px-F
older_Hexagonal_Icon.svg.png 2x" title="Category" width="16"/>,
 <img alt="List-Class article" data-file-height="185" data-file-widt
h="180" decoding="async" height="16" src="//upload.wikimedia.org/wik
ipedia/en/thumb/d/db/Symbol_list_class.svg/16px-Symbol_list_class.sv
g.png" srcset="//upload.wikimedia.org/wikipedia/en/thumb/d/db/Symbol
_list_class.svg/23px-Symbol_list_class.svg.png 1.5x, //upload.wikime
dia.org/wikipedia/en/thumb/d/db/Symbol_list_class.svg/31px-Symbol_li
st_class.svg.png 2x" title="List-Class article" width="16"/>,
 <img alt="Edit this at Wikidata" data-file-height="20" data-file-wi
dth="20" decoding="async" height="10" src="//upload.wikimedia.org/wi
kipedia/en/thumb/8/8a/OOjs_UI_icon_edit-ltr-progressive.svg/10px-OOj
s_UI_icon_edit-ltr-progressive.svg.png" srcset="//upload.wikimedia.o
rg/wikipedia/en/thumb/8/8a/OOjs_UI_icon_edit-ltr-progressive.svg/15p
x-OOjs_UI_icon_edit-ltr-progressive.svg.png 1.5x, //upload.wikimedi
a.org/wikipedia/en/thumb/8/8a/OOjs_UI_icon_edit-ltr-progressive.svg/
20px-OOjs_UI_icon_edit-ltr-progressive.svg.png 2x" style="vertical-a
lign: text-top" width="10"/>,
 <img alt="" height="1" src="//en.wikipedia.org/wiki/Special:Central
AutoLogin/start?type=1x1" style="border: none; position: absolute;"
title="" width="1"/>,
 <img alt="Wikimedia Foundation" height="31" src="/static/images/wik
imedia-button.png" srcset="/static/images/wikimedia-button-1.5x.png
1.5x, /static/images/wikimedia-button-2x.png 2x" width="88"/>,
 <img alt="Powered by MediaWiki" height="31" src="/static/images/pow
eredby_mediawiki_88x31.png" srcset="/static/images/poweredby_mediawi
ki_132x47.png 1.5x, /static/images/poweredby_mediawiki_176x62.png 2
x" width="88"/>]
```

when analysing the webpage, you look for some kind of consistency so that you write a python code that loops through all the things

PERFORM SCRAPING ON BOOK STORE
http://books.toscrape.com/ (http://books.toscrape.com/)

What we need: pages prices ratings title urls(images)

In [879]:

```python
#creating empty lists
pages=[]
prices=[]
ratings=[]
title=[]
urls=[]

no_of_pages=1 #number of pages to be selected


#looping through the required pages and selecting the pages accordingly
for i in range(1,no_of_pages+1): #to include the last page
    url=('http://books.toscrape.com/catalogue/page-{}.html'.format(i))
    pages.append(url) #adding all the content of the respective pages to our cre
ated page list
print(pages)
```

```
['http://books.toscrape.com/catalogue/page-1.html']
```

In [880]:

```python
len(pages)
```

Out[880]:

```
1
```


Getting the data from the particular page url and converting from object of type 'request to object of type beautiful soup, making it more readable

In [881]:

```python
for item in pages:
    page=requests.get(item)
    soup=bs4.BeautifulSoup(page.text,'html.parser')

#print(soup)
```

In [ ]:

In an actual html tag, it includes indentations and nested tags that helps you to identify the actual relationship between and so forth. In order to get that indentation and relationship here, we have to use what is called **prettify()**

In [882]:

```
#print(soup.prettify())
```

Finding all the titles

In [883]:

```
#printing the title
for i in soup.findAll('h3'):
    titless=i.getText()
    title.append(titless)
```

In [884]:

```
# for i in soup.findAll('h3'):
#     print(i)
```

In [885]:

```
titless=i.getText()
```

In [886]:

```
titless
```

Out[886]:

```
"It's Only the Himalayas"
```

In [887]:

```
title
```

Out[887]:

```
['A Light in the ...',
 'Tipping the Velvet',
 'Soumission',
 'Sharp Objects',
 'Sapiens: A Brief History ...',
 'The Requiem Red',
 'The Dirty Little Secrets ...',
 'The Coming Woman: A ...',
 'The Boys in the ...',
 'The Black Maria',
 'Starving Hearts (Triangular Trade ...',
 "Shakespeare's Sonnets",
 'Set Me Free',
 "Scott Pilgrim's Precious Little ...",
 'Rip it Up and ...',
 'Our Band Could Be ...',
 'Olio',
 'Mesaerion: The Best Science ...',
 'Libertarianism for Beginners',
 "It's Only the Himalayas"]
```

In [ ]:

In [ ]:

Finding all the prices

In [888]:

```python
#find all the prices
for p in soup.find_all('p', class_='price_color'): #the 'p' tag is in the 'price
color' class. NB. Make sure to add underscore to class (class_) in order to diff
erentiate it from python inbuilt class
    price=p.getText()
    prices.append(price)
print(prices)
```

```
['Â£51.77', 'Â£53.74', 'Â£50.10', 'Â£47.82', 'Â£54.23', 'Â£22.65',
'Â£33.34', 'Â£17.93', 'Â£22.60', 'Â£52.15', 'Â£13.99', 'Â£20.66', 'Â
£17.46', 'Â£52.29', 'Â£35.02', 'Â£57.25', 'Â£23.88', 'Â£37.59', 'Â£5
1.33', 'Â£45.17']
```

In [ ]:

In [ ]:

Finding all the ratings

In [889]:

```python
#getting the ratings
for s in soup.find_all('p',class_='star-rating'):
    for k,v in s.attrs.items(): #to read about attr: https://www.attrs.org/en/st
able/
        star=v[1]
        ratings.append(star)
        print(star)
```

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

Finding all the url of the images

In [890]:

```python
divs=soup.find_all('div',class_='image_container') #fetching all the div tags in
the class called image_container
#print(divs)
for thumbs in divs:
    tagss=thumbs.find('img',class_='thumbnail')
    #print(tagss)
    links='http://books.toscrape.com/' + str(tagss['src'])
    newlinks=links.replace('..','')
    urls.append(newlinks)
```

In [891]:

```
  links
```

Out[891]:

```
'http://books.toscrape.com/../media/cache/27/a5/27a53d0bb95bdd88288e
af66c9230d7e.jpg'
```

In [892]:

```python
#print(urls)
```

In [893]:

```python
print(tagss)
```

```
<img alt="It's Only the Himalayas" class="thumbnail" src="../media/c
ache/27/a5/27a53d0bb95bdd88288eaf66c9230d7e.jpg"/>
```

In [894]:

```python
#urls
```

In [895]:

```python
web_data={'Title':title,'Price':prices,'Ratings':ratings,'URL':urls}
```

In [896]:

```
web_data
```

Out[896]:

```
{'Title': ['A Light in the ...',
  'Tipping the Velvet',
  'Soumission',
  'Sharp Objects',
  'Sapiens: A Brief History ...',
  'The Requiem Red',
  'The Dirty Little Secrets ...',
  'The Coming Woman: A ...',
  'The Boys in the ...',
  'The Black Maria',
  'Starving Hearts (Triangular Trade ...',
  "Shakespeare's Sonnets",
  'Set Me Free',
  "Scott Pilgrim's Precious Little ...",
  'Rip it Up and ...',
  'Our Band Could Be ...',
  'Olio',
  'Mesaerion: The Best Science ...',
  'Libertarianism for Beginners',
  "It's Only the Himalayas"],
 'Price': ['Â£51.77',
  'Â£53.74',
  'Â£50.10',
  'Â£47.82',
  'Â£54.23',
  'Â£22.65',
  'Â£33.34',
  'Â£17.93',
  'Â£22.60',
  'Â£52.15',
  'Â£13.99',
  'Â£20.66',
  'Â£17.46',
  'Â£52.29',
  'Â£35.02',
  'Â£57.25',
  'Â£23.88',
  'Â£37.59',
  'Â£51.33',
  'Â£45.17'],
 'Ratings': ['Three',
  'One',
  'One',
  'Four',
  'Five',
  'One',
  'Four',
  'Three',
  'Four',
  'One',
  'Two',
  'Four',
  'Five',
  'Five',
  'Five',
  'Three',
  'One',
  'One',
  'Two',
```

```
     'Two'],
  'URL': ['http://books.toscrape.com//media/cache/2c/da/2cdad67c44b00
2e7ead0cc35693c0e8b.jpg',
  'http://books.toscrape.com//media/cache/26/0c/260c6ae16bce31c8f8c9
5daddd9f4a1c.jpg',
  'http://books.toscrape.com//media/cache/3e/ef/3eef99c9d9adef34639f
510662022830.jpg',
  'http://books.toscrape.com//media/cache/32/51/3251cf3a3412f53f339e
42cac2134093.jpg',
  'http://books.toscrape.com//media/cache/be/a5/bea5697f2534a2f86a3e
f27b5a8c12a6.jpg',
  'http://books.toscrape.com//media/cache/68/33/68339b4c9bc034267e1d
a611ab3b34f8.jpg',
  'http://books.toscrape.com//media/cache/92/27/92274a95b7c251fea59a
2b8a78275ab4.jpg',
  'http://books.toscrape.com//media/cache/3d/54/3d54940e57e662c4dd1f
3ff00c78cc64.jpg',
  'http://books.toscrape.com//media/cache/66/88/66883b91f6804b2323c8
369331cb7dd1.jpg',
  'http://books.toscrape.com//media/cache/58/46/5846057e28022268153b
eff6d352b06c.jpg',
  'http://books.toscrape.com//media/cache/be/f4/bef44da28c98f905a3eb
ec0b87be8530.jpg',
  'http://books.toscrape.com//media/cache/10/48/1048f63d3b5061cd2f42
4d20b3f9b666.jpg',
  'http://books.toscrape.com//media/cache/5b/88/5b88c52633f53cacf162
c15f4f823153.jpg',
  'http://books.toscrape.com//media/cache/94/b1/94b1b8b244bce9677c2f
29ccc890d4d2.jpg',
  'http://books.toscrape.com//media/cache/81/c4/81c4a973364e17d01f21
7e1188253d5e.jpg',
  'http://books.toscrape.com//media/cache/54/60/54607fe8945897cdcced
0044103b10b6.jpg',
  'http://books.toscrape.com//media/cache/55/33/553310a7162dfbc2c6d1
9a84da0df9e1.jpg',
  'http://books.toscrape.com//media/cache/09/a3/09a3aef48557576e1a85
ba7efea8ecb7.jpg',
  'http://books.toscrape.com//media/cache/0b/bc/0bbcd0a6f4bcd81ccb10
49a52736406e.jpg',
  'http://books.toscrape.com//media/cache/27/a5/27a53d0bb95bdd88288e
af66c9230d7e.jpg']}
```

Convert our dictionary to a DataFrame

In [899]:

```
df=pd.DataFrame(web_data)
df.index+=1 #changing the index to start from 1 to 20 instead of 0 to 19
```

In [900]:

```
df
```

Out[900]:

	Title	Price	Ratings	URL
1	A Light in the ...	Â£51.77	Three	http://books.toscrape.com//media/cache/2c/da/2...
2	Tipping the Velvet	Â£53.74	One	http://books.toscrape.com//media/cache/26/0c/2...
3	Soumission	Â£50.10	One	http://books.toscrape.com//media/cache/3e/ef/3...
4	Sharp Objects	Â£47.82	Four	http://books.toscrape.com//media/cache/32/51/3...
5	Sapiens: A Brief History ...	Â£54.23	Five	http://books.toscrape.com//media/cache/be/a5/b...
6	The Requiem Red	Â£22.65	One	http://books.toscrape.com//media/cache/68/33/6...
7	The Dirty Little Secrets ...	Â£33.34	Four	http://books.toscrape.com//media/cache/92/27/9...
8	The Coming Woman: A ...	Â£17.93	Three	http://books.toscrape.com//media/cache/3d/54/3...
9	The Boys in the ...	Â£22.60	Four	http://books.toscrape.com//media/cache/66/88/6...
10	The Black Maria	Â£52.15	One	http://books.toscrape.com//media/cache/58/46/5...
11	Starving Hearts (Triangular Trade ...	Â£13.99	Two	http://books.toscrape.com//media/cache/be/f4/b...
12	Shakespeare's Sonnets	Â£20.66	Four	http://books.toscrape.com//media/cache/10/48/1...
13	Set Me Free	Â£17.46	Five	http://books.toscrape.com//media/cache/5b/88/5...
14	Scott Pilgrim's Precious Little ...	Â£52.29	Five	http://books.toscrape.com//media/cache/94/b1/9...
15	Rip it Up and ...	Â£35.02	Five	http://books.toscrape.com//media/cache/81/c4/8...
16	Our Band Could Be ...	Â£57.25	Three	http://books.toscrape.com//media/cache/54/60/5...
17	Olio	Â£23.88	One	http://books.toscrape.com//media/cache/55/33/5...
18	Mesaerion: The Best Science ...	Â£37.59	One	http://books.toscrape.com//media/cache/09/a3/0...
19	Libertarianism for Beginners	Â£51.33	Two	http://books.toscrape.com//media/cache/0b/bc/0...
20	It's Only the Himalayas	Â£45.17	Two	http://books.toscrape.com//media/cache/27/a5/2...

In [774]:

```
#df
```

Out[774]:

	Title	Price	Ratings	URL
1	A Light in the ...	Â£51.77	Three	http://books.toscrape.com//media/cache/2c/da/2...
2	Tipping the Velvet	Â£53.74	One	http://books.toscrape.com//media/cache/26/0c/2...
3	Soumission	Â£50.10	One	http://books.toscrape.com//media/cache/3e/ef/3...
4	Sharp Objects	Â£47.82	Four	http://books.toscrape.com//media/cache/32/51/3...
5	Sapiens: A Brief History ...	Â£54.23	Five	http://books.toscrape.com//media/cache/be/a5/b...
6	The Requiem Red	Â£22.65	One	http://books.toscrape.com//media/cache/68/33/6...
7	The Dirty Little Secrets ...	Â£33.34	Four	http://books.toscrape.com//media/cache/92/27/9...
8	The Coming Woman: A ...	Â£17.93	Three	http://books.toscrape.com//media/cache/3d/54/3...
9	The Boys in the ...	Â£22.60	Four	http://books.toscrape.com//media/cache/66/88/6...
10	The Black Maria	Â£52.15	One	http://books.toscrape.com//media/cache/58/46/5...
11	Starving Hearts (Triangular Trade ...	Â£13.99	Two	http://books.toscrape.com//media/cache/be/f4/b...
12	Shakespeare's Sonnets	Â£20.66	Four	http://books.toscrape.com//media/cache/10/48/1...
13	Set Me Free	Â£17.46	Five	http://books.toscrape.com//media/cache/5b/88/5...
14	Scott Pilgrim's Precious Little ...	Â£52.29	Five	http://books.toscrape.com//media/cache/94/b1/9...
15	Rip it Up and ...	Â£35.02	Five	http://books.toscrape.com//media/cache/81/c4/8...
16	Our Band Could Be ...	Â£57.25	Three	http://books.toscrape.com//media/cache/54/60/5...
17	Olio	Â£23.88	One	http://books.toscrape.com//media/cache/55/33/5...
18	Mesaerion: The Best Science ...	Â£37.59	One	http://books.toscrape.com//media/cache/09/a3/0...
19	Libertarianism for Beginners	Â£51.33	Two	http://books.toscrape.com//media/cache/0b/bc/0...
20	It's Only the Himalayas	Â£45.17	Two	http://books.toscrape.com//media/cache/27/a5/2...

We can take away the currency

In [902]:

```
#[x.strip('Â£') for x in df.Price]
df['Price']=df['Price'].str.replace('Â£','')
```

In [903]:

```
df
```

Out[903]:

	Title	Price	Ratings	URL
1	A Light in the ...	51.77	Three	http://books.toscrape.com//media/cache/2c/da/2...
2	Tipping the Velvet	53.74	One	http://books.toscrape.com//media/cache/26/0c/2...
3	Soumission	50.10	One	http://books.toscrape.com//media/cache/3e/ef/3...
4	Sharp Objects	47.82	Four	http://books.toscrape.com//media/cache/32/51/3...
5	Sapiens: A Brief History ...	54.23	Five	http://books.toscrape.com//media/cache/be/a5/b...
6	The Requiem Red	22.65	One	http://books.toscrape.com//media/cache/68/33/6...
7	The Dirty Little Secrets ...	33.34	Four	http://books.toscrape.com//media/cache/92/27/9...
8	The Coming Woman: A ...	17.93	Three	http://books.toscrape.com//media/cache/3d/54/3...
9	The Boys in the ...	22.60	Four	http://books.toscrape.com//media/cache/66/88/6...
10	The Black Maria	52.15	One	http://books.toscrape.com//media/cache/58/46/5...
11	Starving Hearts (Triangular Trade ...	13.99	Two	http://books.toscrape.com//media/cache/be/f4/b...
12	Shakespeare's Sonnets	20.66	Four	http://books.toscrape.com//media/cache/10/48/1...
13	Set Me Free	17.46	Five	http://books.toscrape.com//media/cache/5b/88/5...
14	Scott Pilgrim's Precious Little ...	52.29	Five	http://books.toscrape.com//media/cache/94/b1/9...
15	Rip it Up and ...	35.02	Five	http://books.toscrape.com//media/cache/81/c4/8...
16	Our Band Could Be ...	57.25	Three	http://books.toscrape.com//media/cache/54/60/5...
17	Olio	23.88	One	http://books.toscrape.com//media/cache/55/33/5...
18	Mesaerion: The Best Science ...	37.59	One	http://books.toscrape.com//media/cache/09/a3/0...
19	Libertarianism for Beginners	51.33	Two	http://books.toscrape.com//media/cache/0b/bc/0...
20	It's Only the Himalayas	45.17	Two	http://books.toscrape.com//media/cache/27/a5/2...

We can now sort our price column to see the highest price

In [777]:

```
df.sort_values(by='Price',ascending=False)
```

Out[777]:

	Title	Price	Ratings	URL
16	Our Band Could Be ...	57.25	Three	http://books.toscrape.com//media/cache/54/60/5...
5	Sapiens: A Brief History ...	54.23	Five	http://books.toscrape.com//media/cache/be/a5/b...
2	Tipping the Velvet	53.74	One	http://books.toscrape.com//media/cache/26/0c/2...
14	Scott Pilgrim's Precious Little ...	52.29	Five	http://books.toscrape.com//media/cache/94/b1/9...
10	The Black Maria	52.15	One	http://books.toscrape.com//media/cache/58/46/5...
1	A Light in the ...	51.77	Three	http://books.toscrape.com//media/cache/2c/da/2...
19	Libertarianism for Beginners	51.33	Two	http://books.toscrape.com//media/cache/0b/bc/0...
3	Soumission	50.10	One	http://books.toscrape.com//media/cache/3e/ef/3...
4	Sharp Objects	47.82	Four	http://books.toscrape.com//media/cache/32/51/3...
20	It's Only the Himalayas	45.17	Two	http://books.toscrape.com//media/cache/27/a5/2...
18	Mesaerion: The Best Science ...	37.59	One	http://books.toscrape.com//media/cache/09/a3/0...
15	Rip it Up and ...	35.02	Five	http://books.toscrape.com//media/cache/81/c4/8...
7	The Dirty Little Secrets ...	33.34	Four	http://books.toscrape.com//media/cache/92/27/9...
17	Olio	23.88	One	http://books.toscrape.com//media/cache/55/33/5...
6	The Requiem Red	22.65	One	http://books.toscrape.com//media/cache/68/33/6...
9	The Boys in the ...	22.60	Four	http://books.toscrape.com//media/cache/66/88/6...
12	Shakespeare's Sonnets	20.66	Four	http://books.toscrape.com//media/cache/10/48/1...
8	The Coming Woman: A ...	17.93	Three	http://books.toscrape.com//media/cache/3d/54/3...
13	Set Me Free	17.46	Five	http://books.toscrape.com//media/cache/5b/88/5...
11	Starving Hearts (Triangular Trade ...	13.99	Two	http://books.toscrape.com//media/cache/be/f4/b...

Convert the ratings column to an integer instead of a string

In [904]:

```
df['Ratings']=df['Ratings'].replace({'Three':3,'One':1,'Two':2,'Four':4,'Five':5
})
```

`In [905]:`

```
df
```

`Out[905]:`

	Title	Price	Ratings	URL
1	A Light in the ...	51.77	3	http://books.toscrape.com//media/cache/2c/da/2...
2	Tipping the Velvet	53.74	1	http://books.toscrape.com//media/cache/26/0c/2...
3	Soumission	50.10	1	http://books.toscrape.com//media/cache/3e/ef/3...
4	Sharp Objects	47.82	4	http://books.toscrape.com//media/cache/32/51/3...
5	Sapiens: A Brief History ...	54.23	5	http://books.toscrape.com//media/cache/be/a5/b...
6	The Requiem Red	22.65	1	http://books.toscrape.com//media/cache/68/33/6...
7	The Dirty Little Secrets ...	33.34	4	http://books.toscrape.com//media/cache/92/27/9...
8	The Coming Woman: A ...	17.93	3	http://books.toscrape.com//media/cache/3d/54/3...
9	The Boys in the ...	22.60	4	http://books.toscrape.com//media/cache/66/88/6...
10	The Black Maria	52.15	1	http://books.toscrape.com//media/cache/58/46/5...
11	Starving Hearts (Triangular Trade ...	13.99	2	http://books.toscrape.com//media/cache/be/f4/b...
12	Shakespeare's Sonnets	20.66	4	http://books.toscrape.com//media/cache/10/48/1...
13	Set Me Free	17.46	5	http://books.toscrape.com//media/cache/5b/88/5...
14	Scott Pilgrim's Precious Little ...	52.29	5	http://books.toscrape.com//media/cache/94/b1/9...
15	Rip it Up and ...	35.02	5	http://books.toscrape.com//media/cache/81/c4/8...
16	Our Band Could Be ...	57.25	3	http://books.toscrape.com//media/cache/54/60/5...
17	Olio	23.88	1	http://books.toscrape.com//media/cache/55/33/5...
18	Mesaerion: The Best Science ...	37.59	1	http://books.toscrape.com//media/cache/09/a3/0...
19	Libertarianism for Beginners	51.33	2	http://books.toscrape.com//media/cache/0b/bc/0...
20	It's Only the Himalayas	45.17	2	http://books.toscrape.com//media/cache/27/a5/2...

Check the data type

`In [780]:`

```
df.dtypes
```

`Out[780]:`

```
Title      object
Price      object
Ratings     int64
URL        object
dtype: object
```

Convert the price column from object to float

In [907]:

```python
df['Price']=df.Price.astype(float)
```

In [908]:

```python
df.dtypes
```

Out[908]:

```
Title       object
Price      float64
Ratings      int64
URL         object
dtype: object
```

Is there any correlation between the ratings of the product and the price?

In [783]:

```python
df.corr()
```

Out[783]:

	Price	Ratings
Price	1.000000	-0.075857
Ratings	-0.075857	1.000000

we can as well visualise it

In [784]:

```python
import matplotlib.pyplot as plt
import seaborn as sns


plt.figure(figsize=(5,5))
sns.heatmap(df.corr())
```
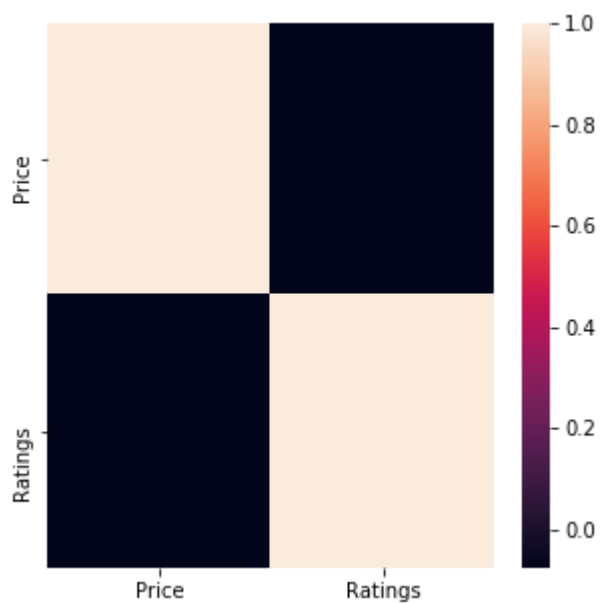
Out[784]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x121eceb70>
```


In [ ]: