A PROJECT REPORT ON

WEB SCRAPING

Submitted by

MAYURESH R. PANDARE(19IT5011)
SHRUSHTI S. POLEKAR(18IT2021)
SHIVANI D.WAKDE(18IT1058)

Under the guidance of

Mrs. ANITA SENATHI

Assistant Professor

Department of Information Technology Ramrao Adik Institute of Technology Nerul ,Navi Mumbai

Certificate

This is to certify that the project entitled Web Scraping is being submitted to the Department of Information Technology, Ramrao Adik Institute of Technology, Navi Mumbai.

Project Guide

External

Examiner(Mrs. Anita Senathi)

( )

Acknowledgement

We would like to express our sincere thanks to many people who have supported us throughout this journey.

We would, firstly like to express our heartfelt gratitude towards our guide Anitha Senathi mam for her guidance, monitoring and constant encouragement throughout the course of this project. Not only our coding Skills but also our confidence is boosted in this process. Many workshops were conducted by mam for imparting knowledge which helped us to clear our concepts.

Lastly, we thank our friends and classmates who always helped us to solve our doubts and gave constant encouragement without doubting on our idea.

So contribution and encouragement of various people has made this project possible.

Introduction

Web Scraping is extracting large amount of data from different web sites. Web scraping a web page involves fetching it and extracting from it. Fetching is the downloading of a page (which a browser does when you view the page). Therefore, web crawling is a main component of web scraping, to fetch pages for later processing. Once fetched, then extraction can take place. The content of a page may be parsed, searched, reformatted, its data copied into a spreadsheet, and so on.

Proposed System

In this project ,Flipkart products have been Scraped. It includes scraping information regarding Redmi and Vivo mobile phones. Information such as mobile's title, price and ratings have been Scraped. The scraped information is then stored in csv file. This stored data can be used for further processing.

System Components

1. IDE : Pycharm
2. Libraries used: Beautiful Soup ,Requests
3. MS Excel

Pycharm:

PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language.It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django as well as Data Science with Anaconda. PyCharm provides API so that developers can write their own plugins to extend PyCharm features.

Beautiful Soup:

Beautiful Soup is a Python library for getting data out of HTML, XML, and other markup languages. Say you've found some webpages that display data relevant to your research, such as date or address information, but that do not provide any way of downloading the data directly. Beautiful Soup helps you pull particular content from a webpage, remove the HTML markup, and save the information. It is a tool for web scraping that helps you clean up and parse the documents you have pulled down from the web.

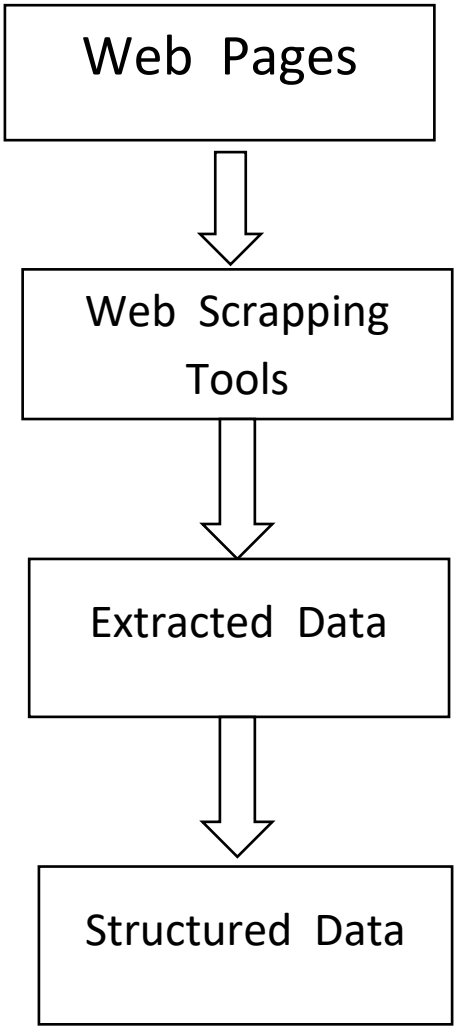
Requests:

The requests library is the de facto standard for making HTTP requests in Python. It abstracts the complexities of making requests behind a beautiful, simple API so that you can focus on interacting with services and consuming data in your application.
The urllib.requests module defines functions and classes which help in opening URLs (mostly HTTP) in a complex world — basic and digest authentication, redirections, cookies and more.

MS Excel:

Microsoft Excel is a spreadsheet developed by Microsoft for Windows, macOS, Android and iOS. It features calculation, graphing tools, pivot tables , and a ,macro programming language called Visual Basic for applications. Microsoft Excel has the basic features of all spreadsheets,[7] using a grid of *cells* arranged in numbered *rows* and letter-named *columns* to organize data manipulations like arithmetic operations

Architecture



Working

Web scraping is an automated method used to extract large amounts of data from websites. The data on the websites are unstructured. Web scraping helps collect these unstructured data and store it in a structured form.In this project flipkart affiliated link is used to scrap mobiles information that are there for sell on flipkart.

When the code for web scraping is run, a request is sent to the URL that is mentioned(flipkart affiliated link). As a response to the request, the server sends the desired data(product price,review ,ratings) and allows us to read the HTML or XML page. The code then, parses the HTML or XML page, finds the data and extracts it. The data is then stored in csv form.This data can be then used for any kind of further processing.

Advantages of Web Scrapping

1. Web Scraping is mainly used by different companies for price monitoring.
2. Price monitoring enables companies to keep a track on opponents company's product prices ,reviews and ratings.
3. Analysis is done on the scraped data and after storing it , product prices are set accordingly to aren maximum profit.
4. Further this data can be used for any kind of different analysis also.
5. Web Scraping saves a lot of manual time and effort.
6. It is faster than manually copying & pasting the data.

Redmi.py

```
from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url="https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy,4io&otracker=nmenu_sub_Electronic
s_0_Mi"
uClient=uReq(my_url)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})
#print(len(containers))
#print(soup.prettify(containers[0]))

filename="redmi.csv"
f=open(filename,"w")

headers="Product_Name,Pricing,Rating\n"
f.write(headers)

for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()

  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text

  trim_price=''.join(pp.split(','))
  rm_rupee=trim_price.split("₹")
  add_rs_price="Rs."+rm_rupee[1]
  split_price=add_rs_price.split('E')
  final_price=split_price[0]

  split_rating=pr.split(" ")
  final_rating=split_rating[0]

  print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
  f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()


from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url2="https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy%2C4io&otracker=nmenu_sub_Elect
ronics_0_Mi&page=2"
uClient=uReq(my_url)
```

```
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})


for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()

  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text

  trim_price=''.join(pp.split(','))
  rm_rupee=trim_price.split("₹")
  add_rs_price="Rs."+rm_rupee[1]
  split_price=add_rs_price.split('E')
  final_price=split_price[0]

  split_rating=pr.split(" ")
  final_rating=split_rating[0]

  print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
  f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()

from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url3="https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy%2C4io&otracker=nmenu_sub_Electronics_0_Mi&page=3"
uClient=uReq(my_url)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})


for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()

  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text

  trim_price=''.join(pp.split(','))
  rm_rupee=trim_price.split("₹")
  add_rs_price="Rs."+rm_rupee[1]
  split_price=add_rs_price.split('E')
```

```
    final_price=split_price[0]

    split_rating=pr.split(" ")
    final_rating=split_rating[0]

    print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
    f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()

from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url3="https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy%2C4io&otracker=nmenu_sub_Electronics_0_Mi&page=4"
uClient=uReq(my_url)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})

for container in containers:
    prodname=container.div.img["alt"]

    print(prodname)
    prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
    pp=prodprice[0].text.strip()

    print(pp)
    prodrating=container.findAll("div",{"class":"niH0FQ"})
    pr=prodrating[0].text

    trim_price=''.join(pp.split(','))
    rm_rupee=trim_price.split("₹")
    add_rs_price="Rs."+rm_rupee[1]
    split_price=add_rs_price.split('E')
    final_price=split_price[0]

    split_rating=pr.split(" ")
    final_rating=split_rating[0]

    print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
    f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()

from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url3="https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy%2C4io&otracker=nmenu_sub_Electronics_0_Mi&page=6"
uClient=uReq(my_url)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})

f=open(filename,"a")
```

```
for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()

  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text

  trim_price=''.join(pp.split(','))
  rm_rupee=trim_price.split("₹")
  add_rs_price="Rs."+rm_rupee[1]
  split_price=add_rs_price.split('E')
  final_price=split_price[0]

  split_rating=pr.split(" ")
  final_rating=split_rating[0]

  print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
  f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()

from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url3="https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy%2C4io&otracker=nmenu_sub_Electronics_0_Mi&page=7"
uClient=uReq(my_url)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})

for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()

  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text

  trim_price=''.join(pp.split(','))
  rm_rupee=trim_price.split("₹")
  add_rs_price="Rs."+rm_rupee[1]
  split_price=add_rs_price.split('E')
  final_price=split_price[0]

  split_rating=pr.split(" ")
  final_rating=split_rating[0]
```

```
  print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
  f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()

from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url3="https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy%2C4io&otracker=nmenu_sub_Electronics_0_Mi&page=8"
uClient=uReq(my_url)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})

for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()

  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text

  trim_price=''.join(pp.split(','))
  rm_rupee=trim_price.split("₹")
  add_rs_price="Rs."+rm_rupee[1]
  split_price=add_rs_price.split('E')
  final_price=split_price[0]

  split_rating=pr.split(" ")
  final_rating=split_rating[0]

  print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
  f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()

from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url3="https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy%2C4io&otracker=nmenu_sub_Electronics_0_Mi&page=9"
uClient=uReq(my_url)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})


for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
```

```
    pp=prodprice[0].text.strip()

    print(pp)
    prodrating=container.findAll("div",{"class":"niH0FQ"})
    pr=prodrating[0].text

    trim_price=''.join(pp.split(','))
    rm_rupee=trim_price.split("₹")
    add_rs_price="Rs."+rm_rupee[1]
    split_price=add_rs_price.split('E')
    final_price=split_price[0]

    split_rating=pr.split(" ")
    final_rating=split_rating[0]

    print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
    f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()
from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url3="https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy%2C4io&otracker=nmenu_sub_Electronics_0_Mi&page=10"
uClient=uReq(my_url)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})

for container in containers:
    prodname=container.div.img["alt"]

    print(prodname)
    prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
    pp=prodprice[0].text.strip()

    print(pp)
    prodrating=container.findAll("div",{"class":"niH0FQ"})
    pr=prodrating[0].text

    trim_price=''.join(pp.split(','))
    rm_rupee=trim_price.split("₹")
    add_rs_price="Rs."+rm_rupee[1]
    split_price=add_rs_price.split('E')
    final_price=split_price[0]

    split_rating=pr.split(" ")
    final_rating=split_rating[0]

    print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
    f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()

from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq
```

```
my_url3="https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy%2C4io&otracker=nmenu_sub_Electronics_0_Mi&page=11"
uClient=uReq(my_url)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})
for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()

  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text

  trim_price=''.join(pp.split(','))
  rm_rupee=trim_price.split("₹")
  add_rs_price="Rs."+rm_rupee[1]
  split_price=add_rs_price.split('E')
  final_price=split_price[0]

  split_rating=pr.split(" ")
  final_rating=split_rating[0]

  print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
  f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()
```

vivo.py

```
from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url1="https://www.flipkart.com/mobiles/pr?sid=tyy%2C4io&p%5B%5D=facets.serviceability%5B%5D%3Dfalse&p%5B%5D=facets.brand%255B%255D%3DVivo&otracker=nmenu_sub_Electronics_0_Vivo"
uClient=uReq(my_url1)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})
filename="vivo.csv"
f=open(filename,"w")
headers="Product_Name,Pricing,Rating\n"
f.write(headers)
for container in containers:
  prodname=container.div.img["alt"]
  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()
  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text
  trim_price=''.join(pp.split(','))
  rm_rupee=trim_price.split("₹")
  add_rs_price="Rs."+rm_rupee[1]
  split_price=add_rs_price.split('E')
  final_price=split_price[0]
  split_rating=pr.split(" ")
  final_rating=split_rating[0]
  print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
  f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()
from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq
my_url2="https://www.flipkart.com/mobiles/pr?sid=tyy%2C4io&p%5B%5D=facets.serviceability%5B%5D%3Dfalse&p%5B%5D=facets.brand%255B%255D%3DVivo&otracker=nmenu_sub_Electronics_0_Vivo&page=2"
uClient=uReq(my_url2)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})
f=open(filename,"a")
for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()

  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text
```

```
    trim_price=''.join(pp.split(','))
    rm_rupee=trim_price.split("₹")
    add_rs_price="Rs."+rm_rupee[1]
    split_price=add_rs_price.split('E')
    final_price=split_price[0]
    split_rating=pr.split(" ")
    final_rating=split_rating[0]
    print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
    f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()
from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq
my_url2="https://www.flipkart.com/mobiles/pr?sid=tyy%2C4io&p%5B%5D=facets.serviceability%5B%5
D%3Dfalse&p%5B%5D=facets.brand%255B%255D%3DVivo&otracker=nmenu_sub_Electronics_0_Vivo&
page=3"
uClient=uReq(my_url2)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})
f=open(filename,"a")
for container in containers:
    prodname=container.div.img["alt"]
    print(prodname)
    prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
    pp=prodprice[0].text.strip()
    print(pp)
    prodrating=container.findAll("div",{"class":"niH0FQ"})
    pr=prodrating[0].text
    trim_price=''.join(pp.split(','))
    rm_rupee=trim_price.split("₹")
    add_rs_price="Rs."+rm_rupee[1]
    split_price=add_rs_price.split('E')
    final_price=split_price[0]
    split_rating=pr.split(" ")
    final_rating=split_rating[0]
    print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
    f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()
from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq
my_url2="https://www.flipkart.com/mobiles/pr?sid=tyy%2C4io&p%5B%5D=facets.serviceability%5B%5
D%3Dfalse&p%5B%5D=facets.brand%255B%255D%3DVivo&otracker=nmenu_sub_Electronics_0_Vivo&
page=4"
uClient=uReq(my_url2)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})
f=open(filename,"a")
for container in containers:
    prodname=container.div.img["alt"]

    print(prodname)
    prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
```

```
    pp=prodprice[0].text.strip()

    print(pp)
    prodrating=container.findAll("div",{"class":"niH0FQ"})
    pr=prodrating[0].text

    trim_price=''.join(pp.split(','))
    rm_rupee=trim_price.split("₹")
    add_rs_price="Rs."+rm_rupee[1]
    split_price=add_rs_price.split('E')
    final_price=split_price[0]

    split_rating=pr.split(" ")
    final_rating=split_rating[0]

    print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
    f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()

from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url2="https://www.flipkart.com/mobiles/pr?sid=tyy%2C4io&p%5B%5D=facets.serviceability%5B%5D%3Dfalse&p%5B%5D=facets.brand%255B%255D%3DVivo&otracker=nmenu_sub_Electronics_0_Vivo&page=5"
uClient=uReq(my_url2)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})
f=open(filename,"a")
for container in containers:
    prodname=container.div.img["alt"]

    print(prodname)
    prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
    pp=prodprice[0].text.strip()

    print(pp)
    prodrating=container.findAll("div",{"class":"niH0FQ"})
    pr=prodrating[0].text

    trim_price=''.join(pp.split(','))
    rm_rupee=trim_price.split("₹")
    add_rs_price="Rs."+rm_rupee[1]
    split_price=add_rs_price.split('E')
    final_price=split_price[0]

    split_rating=pr.split(" ")
    final_rating=split_rating[0]

    print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
    f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()

from bs4 import BeautifulSoup as soup
```

```
from urllib.request import urlopen as uReq

my_url2="https://www.flipkart.com/mobiles/pr?sid=tyy%2C4io&p%5B%5D=facets.serviceability%5B%5
D%3Dfalse&p%5B%5D=facets.brand%255B%255D%3DVivo&otracker=nmenu_sub_Electronics_0_Vivo&
page=6"
uClient=uReq(my_url2)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})
f=open(filename,"a")
for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()

  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text

  trim_price=''.join(pp.split(','))
  rm_rupee=trim_price.split("₹")
  add_rs_price="Rs."+rm_rupee[1]
  split_price=add_rs_price.split('E')
  final_price=split_price[0]

  split_rating=pr.split(" ")
  final_rating=split_rating[0]

  print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
  f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()

from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url2="https://www.flipkart.com/mobiles/pr?sid=tyy%2C4io&p%5B%5D=facets.serviceability%5B%5
D%3Dfalse&p%5B%5D=facets.brand%255B%255D%3DVivo&otracker=nmenu_sub_Electronics_0_Vivo&
page=7"
uClient=uReq(my_url2)
page_html=uClient.read()
uClient.close()
page_soup=soup(page_html,"html.parser")
containers=page_soup.findAll("div",{"class":"_3O0U0u"})
f=open(filename,"a")
for container in containers:
  prodname=container.div.img["alt"]

  print(prodname)
  prodprice=container.findAll("div",{"class":"col col-5-12 _2o7WAb"})
  pp=prodprice[0].text.strip()

  print(pp)
  prodrating=container.findAll("div",{"class":"niH0FQ"})
  pr=prodrating[0].text
```

```
    trim_price=''.join(pp.split(','))
    rm_rupee=trim_price.split("₹")
    add_rs_price="Rs."+rm_rupee[1]
    split_price=add_rs_price.split('E')
    final_price=split_price[0]

    split_rating=pr.split(" ")
    final_rating=split_rating[0]

    print(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
    f.write(prodname.replace(",","|")+","+final_price +","+final_rating+"\n")
print("*"*30)
f.close()
```

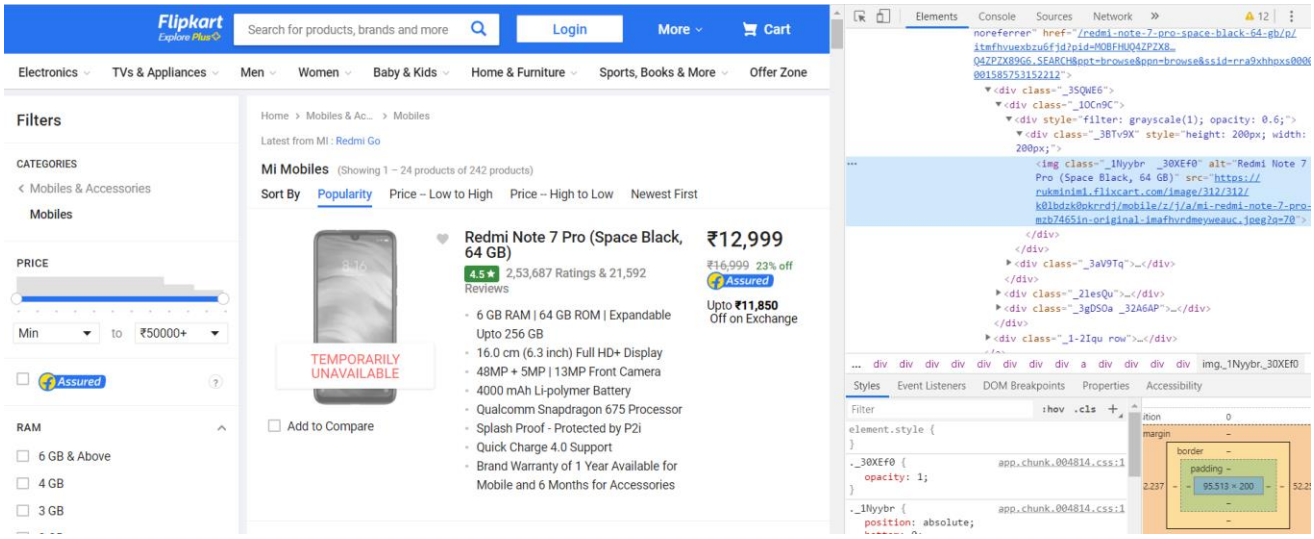
Implementation :

1.Find the url of the web page that is to be scraped. For eg.

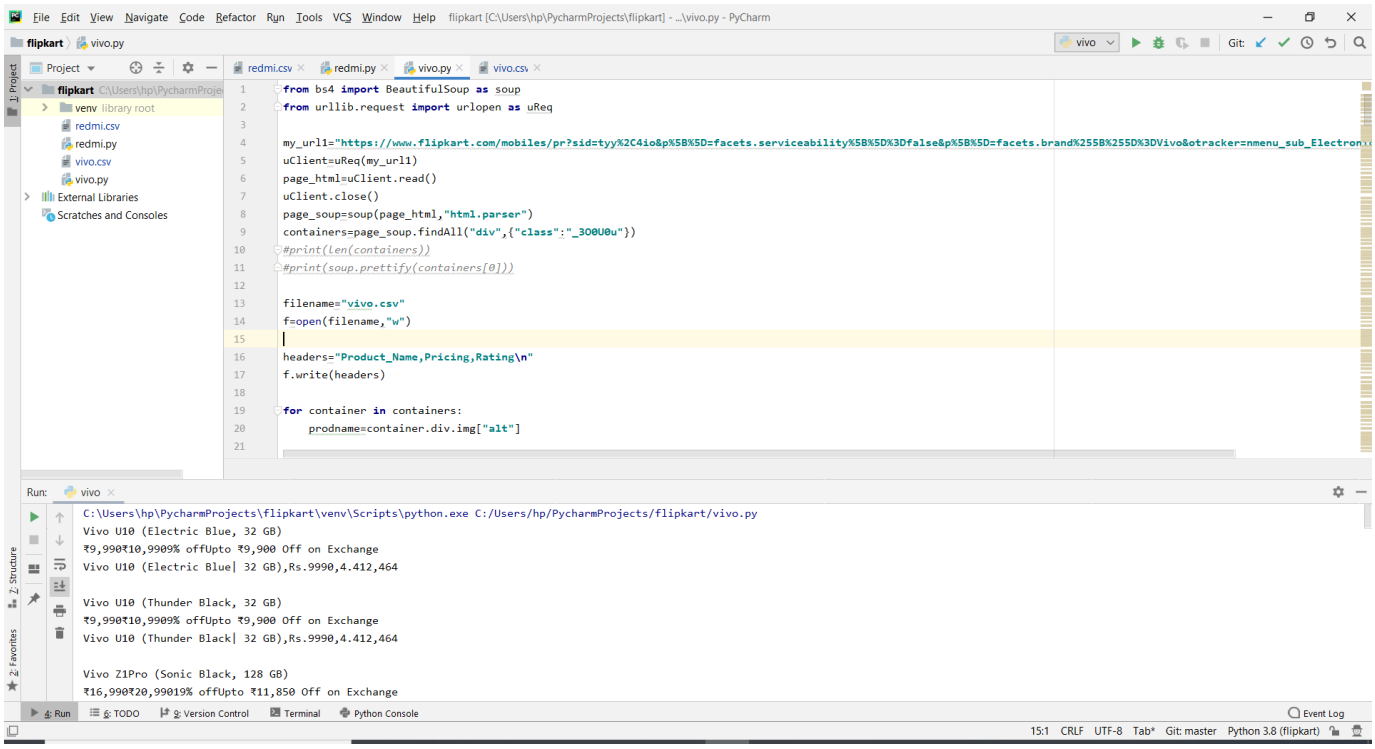
https://www.flipkart.com/mobiles/mi~brand/pr?sid=tyy,4io&otracker=nmenu_sub_Electronics_0_Mi.

2.Inspecting the web page

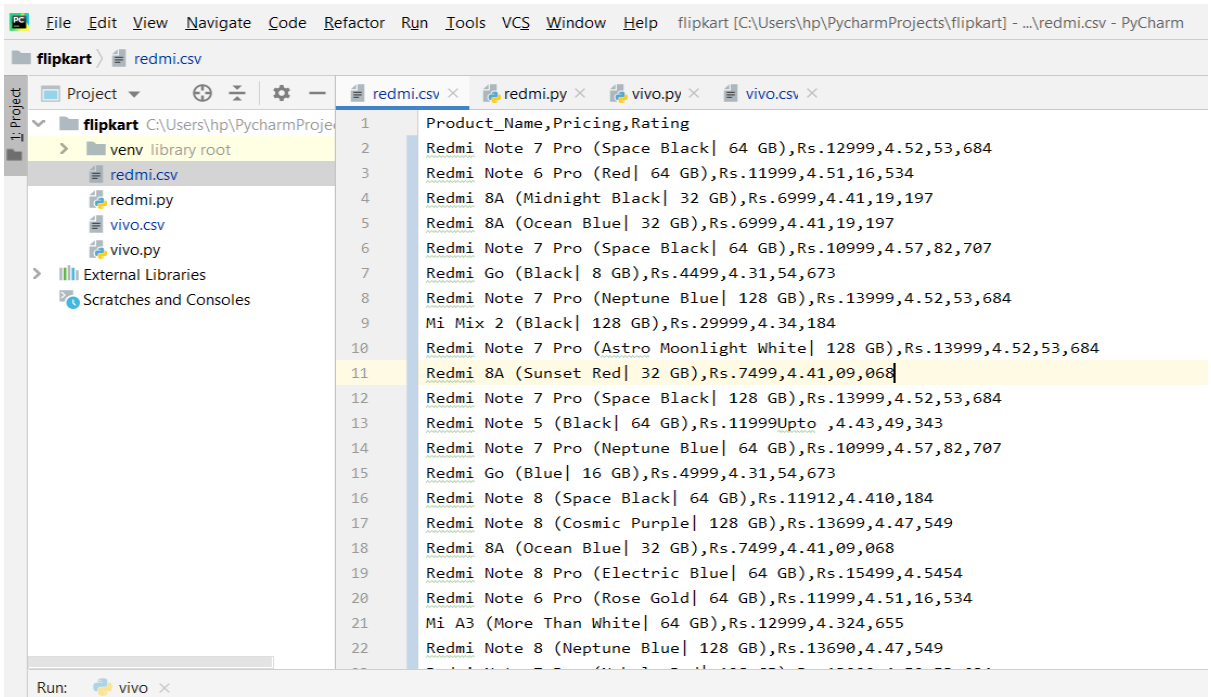
The data is usually nested in tags. So, we inspect the page to see, under which tag the data we want to scrape is nested. To inspect the page, just right click on the element and click on “Inspect”.



3.Write the code.Find the elements that are to be scrapped and the class and div of respective elements from the inspected code and include them in the source code.



4.Store the data in csv file.



5.The data stored in csv file can be used for any kind of processing.

Conclusion:

Web Scraping proves to be of great help in various fields such as e-commerce, travel and tourism, stock market etc.. Thus

Web Scraping has wide applications in various fields as it saves a lot of manual time and efforts.