Project :-FLIPKART MOBIL WEB SCRAPING

PRESENTED BY:- SUMIT SHARMA

HEY! Friend it to importent to fetch the data from any weblinks import some basic 3 libraries

step 1:- import basic library

```
In [ ]:  import pandas as pd
         import requests
         from bs4 import BeautifulSoup
```

step 2:- whtever data you wants to fetch use URL

```
In [ ]:  url="https://www.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&
```

for conformation with web link

```
In [ ]:  r=requests.get(url)
         print(r)
```

<Response [200]> its clear that we connect the server

step 3:- fetch html

```
In [ ]:  soup=BeautifulSoup(r.text,"lxml")
         print(soup)
```

```
In [ ]:  #next page link
         np=soup.find("a",class_= "_1LKTO3").get("href")
         print(np)
```

```
In [20]:  # complete next page link
          cnp="https://www.flipkart.com"+np
          print(cnp)
```

https://www.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&otr
acker1=search&marketplace=FLIPKART&as-show=on&as=off&page=2 (https://www.f
lipkart.com/search?q=mobil+under+50%2C000&otracker=search&otracker1=search
&marketplace=FLIPKART&as-show=on&as=off&page=2)

ites better we can use the loop

```
In [ ]:  soup=BeautifulSoup(r.text,"lxml")
         while True:
             np=soup.find("a",class_= "_1LKTO3").get("href")
             cnp="https://www.flipkart.com"+np
             print(cnp)
```

https://www.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&o
tracker1=search&marketplace=FLIPKART&as-show=on&as=off&page=2 (https://w
ww.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&otracker1=
search&marketplace=FLIPKART&as-show=on&as=off&page=2)
https://www.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&o
tracker1=search&marketplace=FLIPKART&as-show=on&as=off&page=2 (https://w
ww.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&otracker1=
search&marketplace=FLIPKART&as-show=on&as=off&page=2)
https://www.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&o
tracker1=search&marketplace=FLIPKART&as-show=on&as=off&page=2 (https://w
ww.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&otracker1=
search&marketplace=FLIPKART&as-show=on&as=off&page=2)
https://www.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&o
tracker1=search&marketplace=FLIPKART&as-show=on&as=off&page=2 (https://w
ww.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&otracker1=
search&marketplace=FLIPKART&as-show=on&as=off&page=2)
https://www.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&o
tracker1=search&marketplace=FLIPKART&as-show=on&as=off&page=2 (https://w
ww.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&otracker1=

TO GET LINK OF EACH PAGE

```
In [ ]:  # to get link of pages
         import pandas as pd
         import requests
         from bs4 import BeautifulSoup

         for i in range(2,10):
             url="https://www.flipkart.com/search?q=mobil+under+50%2C000&otracker=sea

             r=requests.get(url)
             soup=BeautifulSoup(r.text,"lxml")

             np=soup.find("a",class_= "_1LKTO3").get("href")
             cnp="https://www.flipkart.com"+np
             print(cnp)
```

```
In [ ]:  <div class="_4rR01T">APPLE iPhone 12 (Blue, 128 GB)</div>
```

step 4:- Get data of one page

In [ ]:
```python
import pandas as pd
import requests
from bs4 import BeautifulSoup

# what u wants to scape data from flipkart make a list

Product_name=[]
Prices=[]
Description=[]
Reviews=[]

url="https://www.flipkart.com/search?q=mobil+under+50%2C000&otracker=search&
r=requests.get(url)
soup=BeautifulSoup(r.text,'lxml')
# due to rating issue i am doing this

box=soup.find("div",class_="_1YokD2 _3Mn1Gg")


# extract the product name
names=box.find_all("div",class_="_4rR01T")

for i in names:
    name = i.text
    Product_name.append(name)
print(Product_name)

# check there all total product or not
print(len(Product_name))

# extract the product price
prices=box.find_all("div",class_="_30jeq3 _1_WHN1")

for i in prices:
    name = i.text
    Prices.append(name)
print(Prices)

# check there all total product or not
print(len(Prices))

# extract the Descriptiion of product name
desc=box.find_all("ul",class_="_1xgFaf")

for i in desc:
    name = i.text
    Description.append(name)
print(Description)

# check there all total product or not
print(len(Description))

# extract the Reviews of product name
revi=box.find_all("div",class_="_3LWZlK")

for i in revi:
    name = i.text
    Reviews.append(name)
print(Reviews)

# check there all total  product or not
```

```
# here we have only 24 product but i gotted 39 revies ? so for this i will t
print(len(Reviews))
```

step 5:- Get data as u need with the help of for loop

In [ ]:
```python
import pandas as pd
import requests
from bs4 import BeautifulSoup

# Create empty lists to store the data
Product_name = []
Prices = []
Description = []
Reviews = []

# Loop through multiple pages
for i in range(2, 12):
    url = "https://www.flipkart.com/search?q=mobil+under+50%2C000&otracker=s
    r = requests.get(url)
    soup = BeautifulSoup(r.text, 'lxml')

    box = soup.find("div", class_="_1YokD2 _3Mn1Gg")

    # extract the product name
    names = box.find_all("div", class_="_4rR01T")
    for name in names:
        Product_name.append(name.text)

    # extract the prices
    prices = box.find_all("div", class_="_30jeq3 _1_WHN1")
    for price in prices:
        Prices.append(price.text)

    # extract the description of the product
    desc = box.find_all("ul", class_="_1xgFaf")
    for d in desc:
        Description.append(d.text)

    # extract the reviews of the product
    revi = box.find_all("div", class_="_3LWZlK")
    for review in revi:
        Reviews.append(review.text)

df = pd.DataFrame({"Product_name": Product_name, "Prices": Prices, "Descript
print(df)
```

step 6:- convert the data into dataframe

In [ ]:
```python
df.to_csv("D:/Activity done by sumit/flipkart_mobiles_under_50000.csv")
```

In [ ]: