**38.How to scrap a website using requests and BeautifulSoup modules in python?**

**Objective:**

* To see how to scrap a website using requests and bs4 in python3.

**Process:**

* **Requests:**

The requests library will make a GET request to a web server.

It downloads the HTML contents of a given web page.

* **BeautifulSoup:**

This library was used to extract the text from the web page.

Used for parsing the HTML content in a web page.

**Input:**

* Wikipedia page.

**Output:**

* States and their capitals of India.

**Source code:**

#import libraries

import requests

from bs4 import BeautifulSoup

import pandas as pd

#base url which you want to scrap

base\_url = ('https://en.wikipedia.org/wiki/List\_of\_state\_and\_union\_

territory\_capitals\_in\_India')

#Make a get request to server

r = requests.get(base\_url)

#check the server response

print(r)

#initialize the soup object

soup = BeautifulSoup(r.text, 'html.parser')

#define the HTML table and class

table=soup.find('table', class\_='wikitable sortable plainrowheaders')

#declare empty lists

list1=[]

list2=[]

list3=[]

#make a simple loop

for row in table.findAll("tr"):

table\_data = row.findAll('td')

#to store second column data

table\_head = row.findAll('th')

#only extract table body not heading

if len(table\_data)==6:

list1.append(table\_data[0].find(text=True))

list2.append(table\_head[0].find(text=True))

list3.append(table\_data[1].find(text=True))

print("\n")

#store it in a data frame

df=pd.DataFrame(list1,columns=['Number'])

df['States/UT']=list2

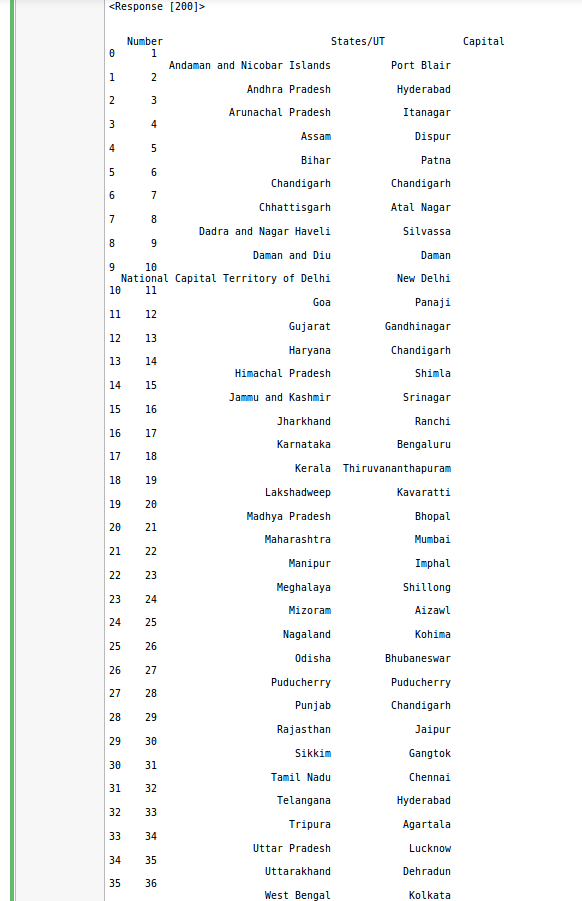
df['Capital']=list3

#print the data frame

print(df)

**Screen shots:**

****

****