Webscraping-COVID19 Data-to-excel-file

June 8, 2020

```python
In [1]: #chek the directory where you want to save the file
        import os
        os.getcwd()

Out[1]: 'C:\\Users\\DELL'

In [2]: os.chdir(r"C:\Users\DELL\Projects")

In [3]: #import modules
        import requests
        import time
        from bs4 import BeautifulSoup
        import pandas as pd

        #copying the url and save it into variable as url

        url="https://www.worldometers.info/coronavirus/#countries"

        html_page=requests.get(url).text

        #print(html_page)

        #use BeautifulSoup to extract the data from html source code
        soup=BeautifulSoup(html_page,'lxml')
        get_table=soup.find("table",id ="main_table_countries_today")
        table_data=get_table.tbody.find_all("tr")
        #print(table_data)

        dic={}
        for i in range(len(table_data)):
            try:
                key=table_data[i].find_all("a",href=True)[0].string
            except:
                key=table_data[i].find_all("td")[0].string

            values=[j.string for j in table_data[i].find_all("td")]

            #print(key,values)
```

```python
        dic[key]=values

    #define the columns
    cols=['Total cases','New Cases','Total Deaths','New Deaths','Total Recovered','Active'

    df=pd.DataFrame(dic).iloc[2:,1:].T.iloc[:,:9]
    df.index_name="Country"

    df.columns=cols
    #print(dic)
```

In [4]: #Print the data
        df.head(10)

Out[4]:   Total cases New Cases Total Deaths New Deaths Total Recovered Active \
        USA   2,007,449   None   112,469   None   761,708   None
        Brazil   691,962   None   37,312   +813   302,084   None
        Russia   476,658   +8,985   5,971   +112   230,688 +3,957
        Spain   288,630   None   27,136   None   N/A   N/A
        UK   286,194   None   40,542   None   N/A   N/A
        India   258,090   +604   7,207   None   124,095   +247
        Italy   234,998   None   33,899   None   165,837   None
        Peru   196,515   None   5,465   None   86,219   None
        Germany   185,869   None   8,776   None   169,600   +500
        Iran   171,789   None   8,281   None   134,349   None

                     Serious Tot Cases/1M pop Tol Death/1M pop
        USA   1,133,272   16,923   6,067
        Brazil   352,566   8,318   3,257
        Russia   239,999   2,300   3,266
        Spain   N/A   617   6,173
        UK   N/A   604   4,217
        India   126,788   8,944   187
        Italy   35,262   287   3,886
        Peru   104,831   1,062   5,966
        Germany   7,493   568   2,219
        Iran   29,159   2,596   2,047

In [7]: #for storint=g the above data in excel sheet in required folder
        df.to_excel("COVID-19_DATA.xls")
        print("code worked and excel file is save din respected folder")

code worked and excel file is save din respected folder