

```
In [14]: import requests
from bs4 import BeautifulSoup
import pandas as pd
import numpy as np
```

```
In [15]: # Making a GET request
r = requests.get('https://www.drdata.in/list-doctors.php?search=Doctor&page=1')

# check status code for response received
# success code - 200
#print(r)

# Parsing the HTML
soup = BeautifulSoup(r.text, 'lxml')

table = soup.find('table', class_='table-bordered table-striped table-condensed cf')
column = table.find_all('th')

headers = []

for c in column:
    header = c.text
    headers.append(header)

df=pd.DataFrame(columns=headers)

df.to_csv('doctors_info.csv')

for i in range(1,25):
    url="https://www.drdata.in/list-doctors.php?search=Doctor&page="+str(i)
    r1=requests.get(url)
    soup = BeautifulSoup(r1.text, 'lxml')

    table = soup.find('table', class_='table-bordered table-striped table-condensed cf')

    rows = table.find_all('tr')

    for j in rows[1:]:
        data=j.find_all('td')
        row=[tr.text for tr in data]
        l=len(df)
        df.loc[l] =row

    df.to_csv('doctors_info.csv',mode='a', header=False)
```

```
In [16]: df.shape
```

Out[16]: (360, 6)

```
In [17]: df.drop('Details', axis=1, inplace=True)
```

```
In [18]: df = df.drop_duplicates()
```

```
In [19]: df.replace(['NA', 'null', '', 'None'], np.nan, inplace=True)

df.dropna(inplace=True)
```

```
In [20]: df.to_csv('doctors_info.csv', index=False)
```

```
In [21]: df.shape
```

Out[21]: (358, 5)

```
In [24]: df.tail()
```

Out[24]:

	Name of Doctor	Specialization	Degree	State	City
355	DR. AGRAWAL ANJANA	ENT SURGEON	MBBS, MS (ENT)	MADHYA PRADESH	UJJAIN
356	DR. AGRAWAL ANKIT	CARDIOLOGIST	MBBS, MD, DM (CARDIOLOGY)	MADHYA PRADESH	JABALPUR
357	DR. AGRAWAL ANKUR	ALLOPATHIC FAMILY PHYSICIAN	MBBS	MADHYA PRADESH	INDORE
358	DR. AGRAWAL ANKUR	OPHTHALMOLOGIST	MBBS, MS	MADHYA PRADESH	JABALPUR

In []: