

```
In [27]: import pandas as pd
import requests
from bs4 import BeautifulSoup
url = 'http://www.worldometers.info/coronavirus/'
req = requests.get(url)
page = BeautifulSoup(req.content, 'html.parser')
table = page.find_all('table',id="main_table_countries_today")[0]
df=pd.read_html(str(table),displayed_only=False)[0]
df
```

Out[27]:

	#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	New Recovered	Active Cases	Serious, Critical	...	Total Tests	Tests/ 1M pop	Population	Continent	1 Case every X ppl	1 Death every X ppl	1 Test every X ppl	New Cases/1M pop	Deaths
0	NaN	North America	130661521	NaN	1667890.0	NaN	126081026.0	1500.0	2912605.0	7670.0	...	NaN	NaN	NaN	North America	NaN	NaN	NaN	NaN	
1	NaN	Asia	221369606	NaN	1552639.0	NaN	205498446.0	8499.0	14318521.0	14765.0	...	NaN	NaN	NaN	Asia	NaN	NaN	NaN	NaN	
2	NaN	Europe	252938203	NaN	2095607.0	NaN	248239760.0	2371.0	2602836.0	4615.0	...	NaN	NaN	NaN	Europe	NaN	NaN	NaN	NaN	
3	NaN	South America	69567479	NaN	1364659.0	NaN	66596108.0	NaN	1606712.0	8953.0	...	NaN	NaN	NaN	South America	NaN	NaN	NaN	NaN	
4	NaN	Oceania	14723934	NaN	32189.0	NaN	14540424.0	NaN	151321.0	55.0	...	NaN	NaN	NaN	Australia/Oceania	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
242	NaN	Total:	69567479	NaN	1364659.0	NaN	66596108.0	NaN	1606712.0	8953.0	...	NaN	NaN	NaN	South America	NaN	NaN	NaN	NaN	
243	NaN	Total:	14723934	NaN	32189.0	NaN	14540424.0	NaN	151321.0	55.0	...	NaN	NaN	NaN	Australia/Oceania	NaN	NaN	NaN	NaN	
244	NaN	Total:	12858372	NaN	258878.0	NaN	12089889.0	NaN	509605.0	529.0	...	NaN	NaN	NaN	Africa	NaN	NaN	NaN	NaN	
245	NaN	Total:	721	NaN	15.0	NaN	706.0	NaN	0.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
246	NaN	Total:	702119836	0.0	6971877.0	0.0	673046359.0	12370.0	22101600.0	36587.0	...	NaN	NaN	NaN	All	NaN	NaN	NaN	NaN	

247 rows × 22 columns

```
In [22]: df.head()
```

Out[22]:

	#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	New Recovered	Active Cases	Serious, Critical	...	Total Tests	Tests/ 1M pop	Population	Continent	1 Case every X ppl	1 Death every X ppl	1 Test every X ppl	New Cases/1M pop	Deaths
0	NaN	North America	130661521	NaN	1667890.0	NaN	126081026.0	1500.0	2912605.0	7670.0	...	NaN	NaN	NaN	North America	NaN	NaN	NaN	NaN	
1	NaN	Asia	221369606	NaN	1552639.0	NaN	205498446.0	8499.0	14318521.0	14765.0	...	NaN	NaN	NaN	Asia	NaN	NaN	NaN	NaN	
2	NaN	Europe	252938203	NaN	2095607.0	NaN	248239760.0	2371.0	2602836.0	4615.0	...	NaN	NaN	NaN	Europe	NaN	NaN	NaN	NaN	
3	NaN	South America	69567479	NaN	1364659.0	NaN	66596108.0	NaN	1606712.0	8953.0	...	NaN	NaN	NaN	South America	NaN	NaN	NaN	NaN	
4	NaN	Oceania	14723934	NaN	32189.0	NaN	14540424.0	NaN	151321.0	55.0	...	NaN	NaN	NaN	Australia/Oceania	NaN	NaN	NaN	NaN	

5 rows × 22 columns

```
In [23]: df.tail()
```

Out[23]:

	#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	New Recovered	Active Cases	Serious, Critical	...	Total Tests	Tests/ 1M pop	Population	Continent	1 Case every X ppl	1 Death every X ppl	1 Test every X ppl	New Cases/1M pop	Deaths
242	NaN	Total:	69567479	NaN	1364659.0	NaN	66596108.0	NaN	1606712.0	8953.0	...	NaN	NaN	NaN	South America	NaN	NaN	NaN	NaN	
243	NaN	Total:	14723934	NaN	32189.0	NaN	14540424.0	NaN	151321.0	55.0	...	NaN	NaN	NaN	Australia/Oceania	NaN	NaN	NaN	NaN	
244	NaN	Total:	12858372	NaN	258878.0	NaN	12089889.0	NaN	509605.0	529.0	...	NaN	NaN	NaN	Africa	NaN	NaN	NaN	NaN	
245	NaN	Total:	721	NaN	15.0	NaN	706.0	NaN	0.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
246	NaN	Total:	702119836	0.0	6971877.0	0.0	673046359.0	12370.0	22101600.0	36587.0	...	NaN	NaN	NaN	All	NaN	NaN	NaN	NaN	

5 rows × 22 columns

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 247 entries, 0 to 246
Data columns (total 22 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   #                                     231 non-null    float64
1   Country, Other                       246 non-null    object
2   Total Cases                          247 non-null    int64
3   New Cases                           2 non-null      float64
4   Total Deaths                        242 non-null    float64
5   New Deaths                          2 non-null      float64
6   Total Recovered                     198 non-null    float64
7   New Recovered                       16 non-null     float64
8   Active Cases                        199 non-null    float64
9   Serious, Critical                   68 non-null     float64
10  Tot Cases/ 1M pop                   231 non-null    float64
11  Deaths/ 1M pop                     226 non-null    float64
12  Total Tests                         213 non-null    float64
13  Tests/ 1M pop                       213 non-null    float64
14  Population                          229 non-null    float64
15  Continent                           243 non-null    object
16  1 Case every X ppl                  229 non-null    float64
17  1 Death every X ppl                 224 non-null    float64
18  1 Test every X ppl                  213 non-null    float64
19  New Cases/1M pop                    0 non-null      float64
20  New Deaths/1M pop                  0 non-null      float64
21  Active Cases/1M pop                 204 non-null    float64
dtypes: float64(19), int64(1), object(2)
memory usage: 42.6+ KB
```

```
In [25]: df.isnull().sum()
```

Out[25]:

#	16
Country, Other	1
Total Cases	0
New Cases	245
Total Deaths	5
New Deaths	245
Total Recovered	49
New Recovered	231
Active Cases	48
Serious, Critical	179
Tot Cases/ 1M pop	16
Deaths/ 1M pop	21
Total Tests	34
Tests/ 1M pop	34
Population	18
Continent	4
1 Case every X ppl	18
1 Death every X ppl	23
1 Test every X ppl	34
New Cases/1M pop	247
New Deaths/1M pop	247
Active Cases/1M pop	43
dtype:	int64

```
In [26]: dff=df.drop(columns=['New Cases', '#', 'New Deaths', 'New Recovered' , 'Serious, Critical', 'New Cases/1M pop', 'New Deaths/1M pop'])
dff.head()
```

Out[26]:

	Country, Other	Total Cases	Total Deaths	Total Recovered	Active Cases	Tot Cases/ 1M pop	Deaths/ 1M pop	Total Tests	Tests/ 1M pop	Population	Continent	1 Case every X ppl	1 Death every X ppl	1 Test every X ppl	Active Cases/1M pop
0	North America	130661521	1667890.0	126081026.0	2912605.0	NaN	NaN	NaN	NaN	NaN	North America	NaN	NaN	NaN	NaN
1	Asia	221369606	1552639.0	205498446.0	14318521.0	NaN	NaN	NaN	NaN	NaN	Asia	NaN	NaN	NaN	NaN
2	Europe	252938203	2095607.0	248239760.0	2602836.0	NaN	NaN	NaN	NaN	NaN	Europe	NaN	NaN	NaN	NaN
3	South America	69567479	1364659.0	66596108.0	1606712.0	NaN	NaN	NaN	NaN	NaN	South America	NaN	NaN	NaN	NaN
4	Oceania	14723934	32189.0	14540424.0	151321.0	NaN	NaN	NaN	NaN	NaN	Australia/Oceania	NaN	NaN	NaN	NaN

```
In [29]: dff.dropna(inplace=True)
```

```
In [30]: dff.isnull().sum()
```

Out[30]:

Country, Other	0
Total Cases	0
Total Deaths	0
Total Recovered	0
Active Cases	0
Tot Cases/ 1M pop	0
Deaths/ 1M pop	0
Total Tests	0
Tests/ 1M pop	0
Population	0
Continent	0
1 Case every X ppl	0
1 Death every X ppl	0
1 Test every X ppl	0
Active Cases/1M pop	0
dtype:	int64

```
In [33]: dff.to_csv('Corona_Dataset.csv')
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```