

Python Pandas working with HTML

In this Part we are going to learn about

1.read HTML (read_html)

2.TO HTML(to_html)

```
In [1]: import pandas as pd
```

```
In [2]: html=pd.read_html("https://en.wikipedia.org/wiki/Mobile_country_code")
```

```
In [3]: type(html)
```

```
Out[3]: list
```

```
In [8]: html
```

```
Out[8]: [  MCC  MNC Brand      Operator      Status Bands (MHz) \
0      1    1  TEST  Test network  Operational      any
1      1    1  TEST  Test network  Operational      any
2    999   99   NaN  Internal use  Operational      any
3    999  999   NaN  Internal use  Operational      any

                                References and notes
0                                     NaN
1                                     NaN
2  Internal use in private networks, no roaming[4]
3  Internal use in private networks, no roaming[4] ,
    Mobile country code                                Country ISO 31
66 \
0      289                                A Abkhazia    GE-
AB
1      412                                Afghanistan
AF
2      276                                Albania
AL
~      ~      ~      ~      ~      ~      ~      ~      ~      ~
```

In [9]: `html[0]`

Out[9]:

| | MCC | MNC | Brand | Operator | Status | Bands (MHz) | References and notes |
|---|-----|-----|-------|--------------|-------------|-------------|---|
| 0 | 1 | 1 | TEST | Test network | Operational | any | NaN |
| 1 | 1 | 1 | TEST | Test network | Operational | any | NaN |
| 2 | 999 | 99 | NaN | Internal use | Operational | any | Internal use in private networks, no roaming[4] |
| 3 | 999 | 999 | NaN | Internal use | Operational | any | Internal use in private networks, no roaming[4] |

In [10]: `html[1]`

Out[10]:

| | Mobile country code | Country | ISO 3166 | Mobile network codes | National MNC authority | Remarks |
|-----|---------------------|---|----------|---|------------------------|--------------------------|
| 0 | 289 | A Abkhazia | GE-AB | List of mobile network codes in Abkhazia | NaN | MCC is not listed by ITU |
| 1 | 412 | Afghanistan | AF | List of mobile network codes in Afghanistan | NaN | NaN |
| 2 | 276 | Albania | AL | List of mobile network codes in Albania | NaN | NaN |
| 3 | 603 | Algeria | DZ | List of mobile network codes in Algeria | NaN | NaN |
| 4 | 544 | American Samoa (United States of America) | AS | List of mobile network codes in American Samoa | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... |
| 247 | 452 | Vietnam | VN | List of mobile network codes in the Vietnam | NaN | NaN |
| 248 | 543 | W Wallis and Futuna | WF | List of mobile network codes in Wallis and Futuna | NaN | NaN |
| 249 | 421 | Y Yemen | YE | List of mobile network codes in the Yemen | NaN | NaN |
| 250 | 645 | Z Zambia | ZM | List of mobile network codes in Zambia | NaN | NaN |
| 251 | 648 | Zimbabwe | ZW | List of mobile network codes in Zimbabwe | NaN | NaN |

252 rows × 6 columns

In [11]: `type(html[1])` *## html convert the data into dataframe*Out[11]: `pandas.core.frame.DataFrame`

```
In [12]: html=pd.read_html("https://en.wikipedia.org/wiki/Mobile_country_code",match="M
```

```
In [15]: html[0]
```

```
Out[15]:
```

| | Mobile country code | Country | ISO 3166 | Mobile network codes | National MNC authority | Remarks |
|------------|---------------------------|---|-------------|---|------------------------------|-----------------------------|
| 0 | 289 | A Abkhazia | GE- AB | List of mobile network codes in Abkhazia | NaN | MCC is not listed by ITU |
| 1 | 412 | Afghanistan | AF | List of mobile network codes in Afghanistan | NaN | NaN |
| 2 | 276 | Albania | AL | List of mobile network codes in Albania | NaN | NaN |
| 3 | 603 | Algeria | DZ | List of mobile network codes in Algeria | NaN | NaN |
| 4 | 544 | American Samoa (United States of America) | AS | List of mobile network codes in American Samoa | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... |
| 247 | 452 | Vietnam | VN | List of mobile network codes in the Vietnam | NaN | NaN |
| 248 | 543 | W Wallis and Futuna | WF | List of mobile network codes in Wallis and Futuna | NaN | NaN |
| 249 | 421 | Y Yemen | YE | List of mobile network codes in the Yemen | NaN | NaN |
| 250 | 645 | Z Zambia | ZM | List of mobile network codes in Zambia | NaN | NaN |
| 251 | 648 | Zimbabwe | ZW | List of mobile network codes in Zimbabwe | NaN | NaN |

252 rows × 6 columns

```
In [16]: html=pd.read_html("https://en.wikipedia.org/wiki/New_York_City",match="Density
```

In [17]: `html[0]`

Out[17]: `.mw-parser-output .navbar-header{text-align:center;position:relative;white-space:nowrap}.mw-parse
.navbar-header .navbar{position:absolute;right:0;top:0;margin:0 5px}New York City's five boro`

| | Jurisdiction | | Population | Land area | | Density of population | | GDP † | U |
|---|--|--|--|--|--|--|--|--|---|
| | Borough | County | Census (2020) | square miles | squarekm | people/sq. mile | people/sq. km | billions(2012 US\$) 2 | U |
| 0 | The Bronx | Bronx | 1472654 | 42.2 | 109.3 | 34920 | 13482 | \$ 38.725 | |
| 1 | Brooklyn | Kings | 2736074 | 69.4 | 179.7 | 39438 | 15227 | \$ 92.230 | |
| 2 | Manhattan | New York | 1694263 | 22.7 | 58.8 | 74781 | 28872 | \$ 651.619 | |
| 3 | Queens | Queens | 2405464 | 108.7 | 281.5 | 22125 | 8542 | \$ 88.578 | |
| 4 | Staten Island | Richmond | 495747 | 57.5 | 148.9 | 8618 | 3327 | \$ 14.806 | |
| 5 | City of New York | City of New York | 8804190 | 302.6 | 783.8 | 29095 | 11234 | \$ 885.958 | |
| 6 | State of New York | State of New York | 20215751 | 47126.4 | 122056.8 | 429 | 166 | \$ 1,514.779 | |
| 7 | † GDP = Gross Domestic Product Sources: [161][1... | † GDP = Gross Domestic Product Sources: [161][1... | † GDP = Gross Domestic Product Sources: [161][1... | † GDP = Gross Domestic Product Sources: [161][1... | † GDP = Gross Domestic Product Sources: [161][1... | † GDP = Gross Domestic Product Sources: [161][1... | † GDP = Gross Domestic Product Sources: [161][1... | † GDP = Gross Domestic Product Sources: [161][1... | I |

In [18]: `type(html[0])`

Out[18]: `pandas.core.frame.DataFrame`

In [20]: `html[0].to_html('demo.html') ## to convert into html`

In []: