

chap2-pandas

June 16, 2022

```
[1]: # import pandas library
import pandas as pd

# create empty DataFrame
df = pd.DataFrame()

# header of dataframe
df.head()
```

```
[1]: Empty DataFrame
Columns: []
Index: []
```

```
[2]: # create dictionary of list
data = {'Name': ['Leszek', 'Pawel', 'Jan', 'Monika'],
        'Age': [23, 24, 25, 26]}

# create the pandas DataFrame
df = pd.DataFrame(data)

# header of dataframe
df.head()
```

```
[2]:
```

| | Name | Age |
|---|--------|-----|
| 0 | Leszek | 23 |
| 1 | Pawel | 24 |
| 2 | Jan | 25 |
| 3 | Monika | 26 |

```
[4]: # creating pandas series using Dictionary
dict1 = {0: 'Ajay', 1: 'Jay', 2: 'Vijay'}

# create Pandas Series
series = pd.Series(dict1)

# show series
series
```

```
[4]: 0    Ajay
      1     Jay
      2    Vijay
      dtype: object
```

```
[5]: # load Pandas and NumPy libraries
import pandas as pd
import numpy as np

# create NumPy array
arr = np.array([30, 40, 30, 40])

# create Pandas Series
series = pd.Series(arr)
series
```

```
[5]: 0    30
      1    40
      2    30
      3    40
      dtype: int32
```

```
[6]: # create Pandas Series
series = pd.Series(10, index=[0, 1, 2, 3, 4, 5])
series
```

```
[6]: 0    10
      1    10
      2    10
      3    10
      4    10
      5    10
      dtype: int64
```

```
[7]: # load data using read_csv()
df = pd.read_csv("WHO_first9cols.csv")

# show initial 5 records
df.head()
```

```
[7]:
```

| | Country | CountryID | Continent | Adolescent fertility rate (%) | \ |
|---|-------------|-----------|-----------|-------------------------------|---|
| 0 | Afghanistan | 1 | 1 | 151.0 | |
| 1 | Albania | 2 | 2 | 27.0 | |
| 2 | Algeria | 3 | 3 | 6.0 | |
| 3 | Andorra | 4 | 2 | NaN | |
| 4 | Angola | 5 | 3 | 146.0 | |

| | Adult literacy rate (%) \ |
|---|---------------------------|
| 0 | 28.0 |
| 1 | 98.7 |
| 2 | 69.9 |
| 3 | NaN |
| 4 | 67.4 |

| | Gross national income per capita (PPP international \$) \ |
|---|---|
| 0 | NaN |
| 1 | 6000.0 |
| 2 | 5940.0 |
| 3 | NaN |
| 4 | 3890.0 |

| | Net primary school enrolment ratio female (%) \ |
|---|---|
| 0 | NaN |
| 1 | 93.0 |
| 2 | 94.0 |
| 3 | 83.0 |
| 4 | 49.0 |

| | Net primary school enrolment ratio male (%) \ |
|---|---|
| 0 | NaN |
| 1 | 94.0 |
| 2 | 96.0 |
| 3 | 83.0 |
| 4 | 51.0 |

| | Population (in thousands) total |
|---|---------------------------------|
| 0 | 26088.0 |
| 1 | 3172.0 |
| 2 | 33351.0 |
| 3 | 74.0 |
| 4 | 16557.0 |

```
[9]: # select a series
country_series = df['Country']

# check datatype of series
type(country_series)
```

[9]: pandas.core.series.Series

```
[10]: # show the shape of DataFrame
print("Shape:", df.shape)
```

Shape: (202, 9)

```
[11]: # check the column list of DataFrame
print("List of Columns:", df.columns)
```

```
List of Columns: Index(['Country', 'CountryID', 'Continent', 'Adolescent
fertility rate (%)',
      'Adult literacy rate (%)',
      'Gross national income per capita (PPP international $)',
      'Net primary school enrolment ratio female (%)',
      'Net primary school enrolment ratio male (%)',
      'Population (in thousands) total'],
      dtype='object')
```

```
[12]: # show the datatypes of columns
print("Data types: ", df.dtypes)
```

```
Data types: Country                                object
CountryID                                           int64
Continent                                           int64
Adolescent fertility rate (%)                      float64
Adult literacy rate (%)                          float64
Gross national income per capita (PPP international $) float64
Net primary school enrolment ratio female (%)      float64
Net primary school enrolment ratio male (%)        float64
Population (in thousands) total                   float64
dtype: object
```

```
[13]: # Pandas Series Slicing
country_series[-5:]
```

```
[13]: 197          Vietnam
      198  West Bank and Gaza
      199          Yemen
      200          Zambia
      201          Zimbabwe
      Name: Country, dtype: object
```

```
[ ]:
```