# Pandas, Part 1

INTERMEDIATE PYTHON



**Hugo Bowne-Anderson**Data Scientist at DataCamp



## Tabular dataset examples

temperature	measured_at	location
76	2016-01-01 14:00:01	valve
86	2016-01-01 14:00:01	compressor
72	2016-01-01 15:00:01	valve
88	2016-01-01 15:00:01	compressor
68	2016-01-01 16:00:01	valve
78	2016-01-01 16:00:01	compressor

## Tabular dataset examples

temperature	measured_at	location
76	2016-01-01 14:00:01	valve
86	2016-01-01 14:00:01	compressor
72	2016-01-01 15:00:01	valve
88	2016-01-01 15:00:01	compressor
68	2016-01-01 16:00:01	valve
78	2016-01-01 16:00:01	compressor

row = observations column = variable



## Tabular dataset examples

temperature	measured_at	location
76	2016-01-01 14:00:01	valve
86	2016-01-01 14:00:01	compressor
72	2016-01-01 15:00:01	valve
88	2016-01-01 15:00:01	compressor
68	2016-01-01 16:00:01	valve
78	2016-01-01 16:00:01	compressor

row = observations column = variable

country	capital	area	population
Brazil	Brasilia	8.516	200.4
Russia	Moscow	17.10	143.5
India	New Delhi	3.286	1252
China	Beijing	9.597	1357
South	Pretoria	1.221	52.98



### Datasets in Python

- 2D Numpy array?
  - One data type

## Datasets in Python

country	capital	area	population
Brazil	Brasilia	8.516	200.4
Russia	Moscow	17.10	143.5
India	New Delhi	3.286	1252
China	Beijing	9.597	1357
South	Pretoria	1.221	52.98

float

float

### Datasets in Python

country	capital	area	population
Brazil	Brasilia	8.516	200.4
Russia	Moscow	17.10	143.5
India	New Delhi	3.286	1252
China	Beijing	9.597	1357
South	Pretoria	1.221	52.98
str	str	float	float

- pandas!
  - High level data manipulation tool
  - Wes McKinney
  - Built on Numpy
  - DataFrame

### **DataFrame**

brics

	country	capital	area	population
BR	Brazil	Brasilia	8.516	200.40
RU	Russia	Moscow	17.100	143.50
IN	India	New Delhi	3.286	1252.00
СН	China	Beijing	9.597	1357.00
SA	South Africa	Pretoria	1.221	52.98

### DataFrame from Dictionary

```
dict = {
    "country":["Brazil", "Russia", "India", "China", "South Africa"],
    "capital":["Brasilia", "Moscow", "New Delhi", "Beijing", "Pretoria"],
        "area":[8.516, 17.10, 3.286, 9.597, 1.221]
"population":[200.4, 143.5, 1252, 1357, 52.98] }
```

- keys (column labels)
- values (data, column by column)

```
import pandas as pd
brics = pd.DataFrame(dict)
```

## DataFrame from Dictionary (2)

brics

```
capital
                                population
                       country
  area
         Brasilia
 8.516
                        Brazil
                                    200.40
                                    143.50
17.100
                        Russia
           Moscow
                                   1252.00
 3.286 New Delhi
                         India
         Beijing
                                   1357.00
 9.597
                         China
                                     52.98
 1.221
        Pretoria South Africa
```

```
brics.index = ["BR", "RU", "IN", "CH", "SA"]
brics
```

```
capital
                       country population
  area
 8.516
         Brasilia
                        Brazil
                                     200.40
17.100
                                    143.50
          Moscow
                        Russia
 3.286
        New Delhi
                         India
                                   1252.00
 9.597
         Beijing
                                   1357.00
                         China
 1.221
        Pretoria South Africa
                                     52.98
```



#### DataFrame from CSV file

brics.csv

```
,country,capital,area,population
BR,Brazil,Brasilia,8.516,200.4
RU,Russia,Moscow,17.10,143.5
IN,India,New Delhi,3.286,1252
CH,China,Beijing,9.597,1357
SA,South Africa,Pretoria,1.221,52.98
```

• CSV = comma-separated values

#### DataFrame from CSV file

• brics.csv

```
,country,capital,area,population
BR,Brazil,Brasilia,8.516,200.4
RU,Russia,Moscow,17.10,143.5
IN,India,New Delhi,3.286,1252
CH,China,Beijing,9.597,1357
SA,South Africa,Pretoria,1.221,52.98
```

```
brics = pd.read_csv("path/to/brics.csv")
brics
```

```
Unnamed: 0
                              capital
                                               population
                   country
                                         area
                   Brazil
                            Brasilia
                                        8.516
                                                   200.40
          BR
0
                    Russia
                                                  143.50
                               Moscow 17.100
          RU
                    India New Delhi
                                                 1252.00
          IN
                                       3.286
                              Beijing
                                                  1357.00
          CH
                    China
                                        9.597
          SA South Africa
                            Pretoria
                                       1.221
                                                    52.98
```



### DataFrame from CSV file

```
brics = pd.read_csv("path/to/brics.csv", index_col = 0)
brics
```

	country	population	area	capital
BF	R Brazil	200	8515767	Brasilia
Rl	J Russia	144	17098242	Moscow
11	N India	1252	3287590	New Delhi
CH	d China	1357	9596961	Beijing
SA	A South Africa	55	1221037	Pretoria



# Pandas, Part 2

INTERMEDIATE PYTHON



**Hugo Bowne-Anderson**Data Scientist at DataCamp



### brics

```
import pandas as pd
brics = pd.read_csv("path/to/brics.csv", index_col = 0)
brics
```

	country	capital	area	population
BR	Brazil	Brasilia	8.516	200.40
RU	Russia	Moscow	17.100	143.50
IN	India	New Delhi	3.286	1252.00
СН	China	Beijing	9.597	1357.00
SA	South Africa	Pretoria	1.221	52.98

### Index and select data

- Square brackets
- Advanced methods
  - o loc
  - iloc

```
country
                   capital
                                    population
                              area
BR
         Brazil
                  Brasilia
                             8.516
                                        200.40
         Russia
                    Moscow 17.100
                                      143.50
          India New Delhi
                                      1252.00
IN
                             3.286
CH
          China
                   Beijing
                            9.597
                                      1357.00
   South Africa
                  Pretoria
                            1.221
                                         52.98
```

brics["country"]

```
BR Brazil
RU Russia
IN India
CH China
SA South Africa
Name: country, dtype: object
```



```
capital
                                      population
         country
                                area
BR
                   Brasilia
                               8.516
                                          200.40
          Brazil
RU
                                          143.50
          Russia
                     Moscow
                              17.100
ΙN
           India
                  New Delhi
                               3.286
                                         1252.00
CH
           China
                              9.597
                                         1357.00
                    Beijing
SA
    South Africa
                               1.221
                   Pretoria
                                            52.98
```

```
type(brics["country"])
```

pandas.core.series.Series

• 1D labelled array

```
capital
                                     population
         country
                               area
BR
          Brazil
                   Brasilia
                              8.516
                                         200.40
RU
          Russia
                     Moscow
                             17.100
                                         143.50
          India
                  New Delhi
                             3.286
                                        1252.00
ΙN
CH
           China
                    Beijing
                             9.597
                                        1357.00
    South Africa
                   Pretoria
                              1.221
                                          52.98
```

```
brics[["country"]]
```

```
country

BR Brazil

RU Russia

IN India

CH China

SA South Africa
```



```
country
                   capital
                                    population
                              area
BR
         Brazil
                  Brasilia
                                        200.40
                             8.516
RU
         Russia
                    Moscow
                            17.100
                                      143.50
          India
                 New Delhi
                             3.286
                                       1252.00
ΙN
CH
          China
                   Beijing
                            9.597
                                       1357.00
    South Africa
                  Pretoria
                             1.221
                                         52.98
```

```
type(brics[["country"]])
```

pandas.core.frame.DataFrame



```
population
         country
                    capital
                               area
                   Brasilia
                                         200.40
BR
          Brazil
                             8.516
RU
          Russia
                     Moscow
                            17.100
                                       143.50
          India
                 New Delhi
                             3.286
                                        1252.00
ΙN
CH
          China
                    Beijing
                            9.597
                                        1357.00
    South Africa
                   Pretoria
                             1.221
                                          52.98
```

```
brics[["country", "capital"]]
```

```
capital
         country
BR
          Brazil
                   Brasilia
          Russia
RU
                     Moscow
           India
                  New Delhi
ΙN
                    Beijing
           China
CH
    South Africa
                   Pretoria
```



### Row Access []

```
population
         country
                    capital
                                area
BR
          Brazil
                   Brasilia
                              8.516
                                          200.40
RU
          Russia
                     Moscow
                              17.100
                                          143.50
           India
                  New Delhi
                              3.286
                                         1252.00
ΙN
CH
           China
                    Beijing
                              9.597
                                         1357.00
    South Africa
                   Pretoria
                               1.221
                                           52.98
```

```
brics[1:4]
```

```
country
              capital
                               population
                         area
   Russia
               Moscow 17.100
                                    143.5
    India
            New Delhi
                                   1252.0
IN
                        3.286
              Beijing
    China
                        9.597
                                   1357.0
```



### Row Access []

```
country
                    capital
                                     population
                               area
BR
                   Brasilia
                              8.516
                                         200.40
          Brazil
                                                   * 0 *
RU
          Russia
                     Moscow
                             17.100
                                       143.50
                                                   * 1 *
          India
                  New Delhi
                              3.286
                                        1252.00
                                                  * 2 *
ΙN
CH
           China
                    Beijing
                             9.597
                                        1357.00
                                                   * 3 *
    South Africa
                   Pretoria
                              1.221
                                          52.98
                                                   * 4 *
```

```
brics[1:4]
```

```
country
              capital
                               population
                         area
   Russia
               Moscow 17.100
                                    143.5
    India
            New Delhi
                                   1252.0
ΙN
                        3.286
              Beijing
    China
                        9.597
                                   1357.0
```



### Discussion []

- Square brackets: limited functionality
- Ideally
  - 2D Numpy arrays
  - o my\_array[rows, columns]
- pandas
  - loc (label-based)
  - iloc (integer position-based)

#### Row Access loc

```
population
        country
                 capital
                         area
BR
        Brazil
                Brasilia
                         8.516
                                    200.40
               Moscow 17.100
RU
        Russia
                                143.50
        India New Delhi
                         3.286
                                1252.00
ΙN
CH
         China
                Beijing
                        9.597
                                 1357.00
   South Africa
                Pretoria
                         1.221
                                     52.98
```

```
brics.loc["RU"]
```

```
country Russia
capital Moscow
area 17.1
population 143.5
Name: RU, dtype: object
```

• Row as pandas Series

#### Row Access loc

```
country
                    capital
                                     population
                               area
                   Brasilia
BR
          Brazil
                              8.516
                                         200.40
                                         143.50
          Russia
                     Moscow
                             17.100
RU
                  New Delhi
                              3.286
                                        1252.00
IN
           India
CH
           China
                    Beijing 9.597
                                        1357.00
    South Africa
                   Pretoria
                              1.221
                                           52.98
```

```
brics.loc[["RU"]]
```

```
country capital area population RU Russia Moscow 17.1 143.5
```

DataFrame

#### Row Access loc

```
country
                    capital
                                      population
                               area
BR
          Brazil
                   Brasilia
                              8.516
                                          200,40
RU
          Russia
                                          143.50
                     Moscow
                             17.100
ΙN
           India
                  New Delhi
                              3.286
                                         1252.00
           China
                    Beijing
                                         1357.00
CH
                              9.597
   South Africa
                   Pretoria
                              1.221
                                           52.98
```

```
brics.loc[["RU", "IN", "CH"]]
```

```
capital
                               population
  country
                         area
   Russia
               Moscow
                       17.100
                                    143.5
    India
           New Delhi
                        3.286
ΙN
                                   1252.0
    China
              Beijing
                                   1357.0
                        9.597
CH
```



```
country
                capital
                                population
                         area
BR
        Brazil
                Brasilia 8.516
                                    200,40
        Russia Moscow 17.100
                                143.50
RU
ΙN
        India New Delhi
                         3.286
                                   1252.00
         China
                 Beijing 9.597
                                   1357.00
CH
   South Africa
                Pretoria 1.221
                                     52.98
```

```
brics.loc[["RU", "IN", "CH"], ["country", "capital"]]
```

```
country capital
RU Russia Moscow
IN India New Delhi
CH China Beijing
```



```
population
        country
                  capital
                           area
                  Brasilia
                                      200.40
BR
         Brazil
                           8.516
RU
         Russia
                   Moscow
                          17.100
                                    143.50
         India New Delhi
                           3.286
                                    1252.00
ΙN
CH
          China
                  Beijing
                           9.597
                                     1357.00
   South Africa
                  Pretoria
                            1.221
                                        52.98
```

```
brics.loc[:, ["country", "capital"]]
```

```
capital
         country
BR
          Brazil
                   Brasilia
          Russia
RU
                     Moscow
          India
                  New Delhi
ΙN
                    Beijing
           China
CH
    South Africa
                   Pretoria
```



### Recap

- Square brackets
  - Column access brics[["country", "capital"]]
  - Row access: only through slicing brics[1:4]
- loc (label-based)
  - Row access brics.loc[["RU", "IN", "CH"]]
  - Column access brics.loc[:, ["country", "capital"]]
  - Row & Column access

```
brics.loc[
["RU", "IN", "CH"],
["country", "capital"]
]
```

#### Row Access iloc

```
population
        country
                  capital
                            area
                 Brasilia
                                     200.40
BR
         Brazil
                           8.516
                 Moscow 17.100
                                   143.50
RU
         Russia
         India New Delhi 3.286
                                   1252.00
IN
          China
                  Beijing 9.597
                                    1357.00
CH
   South Africa
                Pretoria 1.221
                                      52.98
```

brics.loc[["RU"]]

```
country capital area population
RU Russia Moscow 17.1 143.5
```

brics.iloc[[1]]

```
country capital area population RU Russia Moscow 17.1 143.5
```



#### Row Access iloc

```
population
         country
                   capital
                              area
                  Brasilia
                             8.516
                                        200.40
BR
         Brazil
RU
         Russia
                    Moscow
                            17.100
                                      143.50
          India
                 New Delhi
                            3.286
                                       1252.00
ΙN
CH
          China
                   Beijing
                            9.597
                                       1357.00
    South Africa
                  Pretoria
                             1.221
                                         52.98
```

```
brics.loc[["RU", "IN", "CH"]]
```

```
capital
                               population
  country
                         area
   Russia
              Moscow 17.100
                                   143.5
           New Delhi
                                  1252.0
IN
    India
                       3.286
             Beijing
    China
                       9.597
                                  1357.0
```



#### Row Access iloc

```
population
         country
                    capital
                               area
BR
          Brazil
                   Brasilia
                              8.516
                                         200.40
RU
          Russia
                     Moscow
                            17.100
                                       143.50
          India
                  New Delhi
                             3.286
                                        1252.00
ΙN
CH
           China
                    Beijing
                             9.597
                                        1357.00
    South Africa
                   Pretoria
                              1.221
                                          52.98
```

```
brics.iloc[[1,2,3]]
```

```
country
              capital
                               population
                         area
   Russia
               Moscow 17.100
                                    143.5
    India
           New Delhi
                                   1252.0
IN
                        3.286
              Beijing
    China
                        9.597
                                   1357.0
```



```
population
        country
                 capital
                          area
                 Brasilia
                         8.516
                                    200.40
BR
         Brazil
RU
         Russia
                Moscow 17.100
                                 143.50
        India New Delhi
                         3.286
                                  1252.00
ΙN
CH
         China
                Beijing
                         9.597
                                  1357.00
   South Africa
                 Pretoria
                          1.221
                                     52.98
```

```
brics.loc[["RU", "IN", "CH"], ["country", "capital"]]
```

```
country capital
RU Russia Moscow
IN India New Delhi
CH China Beijing
```



```
capital
                                    population
         country
                            area
BR
         Brazil
                  Brasilia
                             8.516
                                        200.40
                                      143.50
RU
         Russia
                    Moscow
                            17.100
          India
                 New Delhi
                            3.286
                                       1252.00
ΙN
CH
          China
                   Beijing
                            9.597
                                       1357.00
    South Africa
                  Pretoria
                             1.221
                                         52.98
```

```
brics.iloc[[1,2,3], [0, 1]]
```

```
country capital
RU Russia Moscow
IN India New Delhi
CH China Beijing
```



```
population
        country
                  capital
                           area
                  Brasilia
                                      200.40
BR
         Brazil
                           8.516
RU
         Russia
                   Moscow
                          17.100
                                    143.50
         India New Delhi
                           3.286
                                    1252.00
ΙN
CH
          China
                  Beijing
                           9.597
                                    1357.00
   South Africa
                  Pretoria
                            1.221
                                        52.98
```

```
brics.loc[:, ["country", "capital"]]
```

```
capital
         country
BR
          Brazil
                   Brasilia
          Russia
RU
                     Moscow
          India
                  New Delhi
ΙN
                    Beijing
           China
CH
    South Africa
                   Pretoria
```



```
population
         country
                   capital
                              area
         Brazil
                   Brasilia
                             8.516
                                        200.40
BR
                            17.100
RU
         Russia
                    Moscow
                                      143.50
          India
                 New Delhi
                             3.286
                                       1252.00
ΙN
CH
          China
                   Beijing
                            9.597
                                       1357.00
    South Africa
                   Pretoria
                             1.221
                                          52.98
```

```
brics.iloc[:, [0,1]]
```

```
country
                     capital
BR
          Brazil
                   Brasilia
          Russia
RU
                     Moscow
           India
                  New Delhi
ΙN
                    Beijing
           China
CH
    South Africa
                   Pretoria
```

