# pandas Cheat Sheet

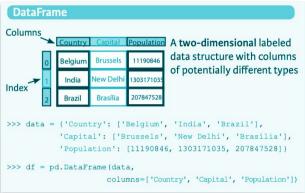
### **Pandas**

The Pandas library is built on NumPy and provides easy-to-use data structures and data analysis tools for the Python programming language.

### Use the following import convention:

>>> import pandas as pd





# Columns=['Country', 'Capital', 'Population'] Selection Also see NumPy Arrays Getting >>> s['b'] -5 >>> df[1:] Country Capital Population 1 India New Delhi 1303171035 2 Brazil Brasilia 207847528 Get subset of a DataFrame

```
Selecting, Boolean Indexing & Setting
By Position
                                         Select single value by row &
>>> df.iloc[[0],[0]]
 'Belgium'
                                         column
>>> df.iat([0],[0])
 'Belgium'
 Bv Label
                                         Select single value by row &
>>> df.loc[[0], ['Country']]
                                         column labels
  'Belgium'
>>> df.at([0], ['Country'])
  'Belgium'
 By Label/Position
                                         Select single row of
>>> df.ix[2]
                                         subset of rows
 Country
               Brazil
  Capital
            Brasília
  Population 207847528
                                         Select a single column of
>>> df.ix[:,'Capital']
                                         subset of columns
       Brussels
      New Delhi
       Brasilia
                                         Select rows and columns
>>> df.ix[1, 'Capital']
  'New Delhi'
 Boolean Indexing
>>> s[~(s > 1)]
                                         Series s where value is not >1
                                         s where value is <-1 or >2
>>> s[(s < -1) | (s > 2)]
>>> df[df['Population']>1200000000] Use filter to adjust DataFrame
 Setting
                                         Set index a of Series s to 6
>>> s['a'] = 6
```

## 1/0

### Read and Write to CSV

>>> pd.read\_csv('file.csv', header=None, nrows=5)

### >>> df.to\_csv('myDataFrame.csv')

### Read and Write to Excel

>>> pd.read\_excel('file.xlsx')

>>> df.to\_excel('dir/myDataFrame.xlsx', sheet\_name='Sheet1')

### Read multiple sheets from the same file

>>> xlsx = pd.ExcelFile('file.xls')

>>> df = pd.read excel(xlsx, 'Sheet1')

### Read and Write to SQL Query or Database Table

>>> from sqlalchemy import create\_engine

>>> engine = create\_engine('sqlite:///:memory:')

>>> pd.read\_sql("SELECT \* FROM my\_table;", engine)

>>> pd.read\_sql\_table('my\_table', engine)
>>> pd.read sql query("SELECT \* FROM my table;", engine)

read\_sql() is a convenience wrapper around read\_sql\_table() and
read\_sql query()

>>> df.to sql('myDf', engine)

### Retrieving Series/DataFrame Information

### **Basic Information**

### Summary

>>> df.sum()
>>> df.cumsum()
>>> df.min()/df.max()
>>> df.idxmin()/df.idxmax()
>>> df.idxmin()/df.idxmax()
>>> df.describe()
>>> df.mealn()
>>> df.median()

Sum of values
Cummulative sum of values
Minimum/Maximum values
Minimum/Maximum index value
Summary statistics
Median of values
Median of values