NumPy (Numerical Python)

- Go to the website “numpy.org”
- Click “Documentation”
  - NumPy quickstart
  - NumPy: the absolute basics for beginners
  - NumPy fundamentals
- Array (N-dimensional array, ndarray)
  - 1-dimensional array (vector)
  - 2-dimensional array (matrix)
  - 3 or higher dimensional array (tensor)

Array from list

- import numpy as np
- np.array([1, 2, 3])
  - Vector
  - Has one axis (dimension)
  - Has 3 elements
  - Has a length of 3
- np.array([[1, 2, 3],
  [4, 5, 6]])
  - Matrix
  - Has two axes (dimensions)
  - First axis has a length of 2
  - Second axis has a length of 3

Attributes for numpy.ndarray

- .ndim
  - the number of axes (dimensions) of the array.
- .shape
  - the size of the array in each dimension.
- .size
  - the total number of elements of the array.
- .dtype
  - the type of the elements in the array.
- .itemsize
  - the size in bytes of each element of the array.

Array creation

```
np.zeros(3)
np.zeros((3, 4))
np.ones(3)
np.ones((3, 4))
np.identity(5)
np.arange(6), np.arange(6.0)
np.arange(2, 9)
np.arrange(2, 20, 3)
np.linspace(2, 3, 11)
np.arrange(12).reshape(3, 4)
```

Matrix operations

```
A * B   # elementwise product
A @ B  # matrix product
A.dot(B)  # matrix product
from numpy.linalg import inv
inv(A)
np.transpose(A)
A.T
```

Numpy functions

```
np.sum(B)
np.sum(B, axis=0)
np.sum(B, axis=1)
np.mean()
np.max()
np.argmax()
np.hstack((a, b))
np.vstack((a, b))
np.hsplit()
np.vsplit()
```

Pandas (Panel Data Analysis)

In 2008, *pandas* development began at AQR Capital Management.

Columns

Symbols	AAPL	NKE	GOOGL	AMZN	FB
Date					
2018-01-02	41.248272	61.186932	1073.209961	1189.010010	181.419998
2018-01-03	41.241089	61.177288	1091.520020	1204.199951	184.669998
2018-01-04	41.432659	61.138741	1095.760010	1209.589966	184.330002
2018-01-05	41.904385	61.659157	1110.290039	1229.140015	186.850006
2018-01-08	41.748737	62.208481	1114.209961	1246.869995	188.279999
...	...	...	...	...	...
2021-08-25	148.360001	169.560196	2841.580078	3299.179932	368.390015
2021-08-26	147.539993	166.645004	2828.810059	3316.000000	364.380005
2021-08-27	148.600006	167.580002	2880.080078	3349.629883	372.630005
2021-08-30	153.119995	168.029999	2891.810059	3421.570068	380.660004
2021-08-31	151.830002	164.740005	2893.949951	3470.790039	379.380005

Index

923 rows × 5 columns

Values

Dataframe and Series

- Dataframe (2-dimensions)
- Series (1-dimension)

```
df['AAPL']
```

```
Date
2018-01-02     41.248272
2018-01-03     41.241089
2018-01-04     41.432659
2018-01-05     41.904385
2018-01-08     41.748737
                   ...
2021-08-25    148.360001
2021-08-26    147.539993
2021-08-27    148.600006
2021-08-30    153.119995
2021-08-31    151.830002
Name: AAPL, Length: 923, dtype: float64
```

Index

Values

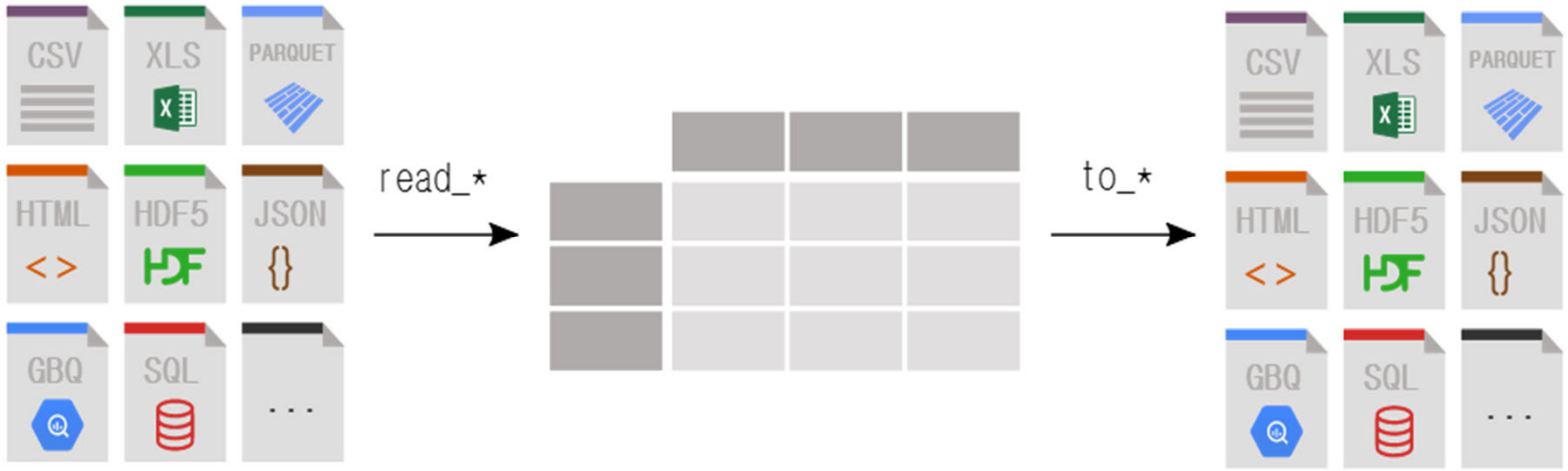
Functions, methods, and attributes

pd.Series()
pd.DataFrame()
pd.date_range()

s.index
s.values
s.abs()

df.index
df.columns
df.values
df.T
df.head()
df.tail()
df.describe()
df.sort_index()
df.sort_values()
df.loc[]
df.iloc[]

Reading and Writing




```
import seaborn as sns
df = sns.load_dataset('titanic')
df.to_excel('titanic.xlsx')
```

Additional methods

df.dtypes

df.info()

df.mean()

df.median()

df.groupby()

df.value_counts()

df.count()