Data Structures in Pandas

1. Series
2. DataFrame

In [2]:

```
import numpy as np
import pandas as pd

import warnings
warnings.filterwarnings("ignore")
```

1. Seires

In [3]:

```
pd.Series()
```

Out[3]:

```
Series([], dtype: float64)
```

In [4]:

```
names = ["Jay", "Kumar", "Suraj", "Divya", "Raj"]

s = pd.Series(names)
s
```

Out[4]:

```
0      Jay
1    Kumar
2    Suraj
3    Divya
4      Raj
dtype: object
```

In [5]:

```
names
```

Out[5]:

```
['Jay', 'Kumar', 'Suraj', 'Divya', 'Raj']
```

In [6]:

```
s[2]
```

Out[6]:

```
'Suraj'
```

In [7]:

```
s = pd.Series(data=names, index=["a", "b", "c", "d", "e"])
s
```

Out[7]:

```
a      Jay
b    Kumar
c    Suraj
d    Divya
e      Raj
dtype: object
```

In [8]:

```
s["c"]
```

Out[8]:

```
'Suraj'
```

In [9]:

```
d = {
    "a":"Jay",
    "b": "Kumar",
    "c": "Suraj",
    "d": "Divya",
    "e": "Raj"
}

s = pd.Series(d)
s
```

Out[9]:

```
a      Jay
b    Kumar
c    Suraj
d    Divya
e      Raj
dtype: object
```

2. DataFrame

In [10]:

```
pd.DataFrame()
```

Out[10]:

—

In [11]:

```
names
```

Out[11]:

```
['Jay', 'Kumar', 'Suraj', 'Divya', 'Raj']
```

In [12]:

```
pd.DataFrame(names)
```

Out[12]:

	0
0	Jay
1	Kumar
2	Suraj
3	Divya
4	Raj

1. DataFrame
    - Name
    - Python
    - Machine Learning

In [14]:

```
data =[
    ["Pratik", 89, 90],
    ["Purvesh", 78, 59],
    ["Jay", 75, 89],
    ["Sandesh", 89, 78],
    ["Raj", 89, 76]
]

colname = ["Name", "Python", "Machine Learning"]

df = pd.DataFrame(data, columns=colname)
df
```

Out[14]:

	Name	Python	Machine Learning
0	Pratik	89	90
1	Purvesh	78	59
2	Jay	75	89
3	Sandesh	89	78
4	Raj	89	76

In [15]:

```
data = {
    "Name":["Pratik", "Purvesh", "Jay", "Sandesh", "Raj"],
    "Python": [89, 78, 75, 89, 89],
    "Machine Learning": [90, 59, 89, 78, 76]
}


df= pd.DataFrame(data)
df
```

Out[15]:

	Name	Python	Machine Learning
0	Pratik	89	90
1	Purvesh	78	59
2	Jay	75	89
3	Sandesh	89	78
4	Raj	89	76

In [17]:

```
df.Name=="Jay"
```

Out[17]:

```
0    False
1    False
2     True
3    False
4    False
Name: Name, dtype: bool
```

In [18]:

```
df[df.Name=="Jay"]
```

Out[18]:

	Name	Python	Machine Learning
2	Jay	75	89

In [19]:

```
df
```

Out[19]:

	Name	Python	Machine Learning
0	Pratik	89	90
1	Purvesh	78	59
2	Jay	75	89
3	Sandesh	89	78
4	Raj	89	76

In [20]:

```
data = {
    "Name":["Pratik", "Purvesh", "Jay", "Sandesh", "Raj"],
    "Python": [89, 78, 75, 89, 89],
    "Machine Learning": [90, 59, 89, 78, 76]
}


df= pd.DataFrame(data, index=["rank1", "rank2", "rank3", "rank4", "rank5"])
df
```

Out[20]:

	Name	Python	Machine Learning
rank1	Pratik	89	90
rank2	Purvesh	78	59
rank3	Jay	75	89
rank4	Sandesh	89	78
rank5	Raj	89	76

In [21]:

```
np.random.randn(10, 4)
```

Out[21]:

```
array([[ 0.13657276, -0.5279357 , -0.71585003, -0.3397155 ],
       [ 0.9333314 , -0.05500137, -0.92153107,  2.23896811],
       [-0.28605501, -0.43057789, -0.97153012, -2.27722152],
       [ 1.15586248,  0.90401362,  1.19035337, -0.13942774],
       [ 1.6178737 , -0.86576849,  1.28331416, -1.30382281],
       [ 0.43311574,  0.1104823 ,  2.29152108, -0.30038787],
       [-1.62854303, -1.13773677, -0.38385688,  0.15068384],
       [ 0.19352511, -0.51468128, -0.37173023,  0.86025596],
       [ 0.11637096,  1.0970359 , -0.49001749,  0.23982129],
       [ 1.33084557, -0.80664183,  0.3403219 ,  1.31136949]])
```

In [22]:

```
d = np.random.randn(10, 4)
df= pd.DataFrame(d, columns=["A", "B", "C", "D"])
df
```

Out[22]:

	A	B	C	D
0	0.686030	-1.183452	0.287314	1.133993
1	0.807829	0.096398	-1.640266	1.301005
2	-1.527376	2.233818	-0.530018	-0.371345
3	0.537792	0.454365	1.616601	-1.762229
4	0.154061	-1.592037	1.191109	-1.008560
5	0.365512	-0.249763	0.685206	0.204793
6	-1.092865	-0.726417	-1.747064	0.389714
7	0.835137	-0.247096	-1.044454	-1.341522
8	-0.685918	1.114754	-0.717764	0.330156
9	0.255090	0.999026	0.543548	0.679192

In [23]:

```
df.to_csv("random.csv", index=False)
```