

Overview of Series

Print to PDF ►

Let us quickly go through one of the Pandas Data Structure - Series.

- Pandas Series is a one-dimensional labeled array capable of holding any data type.
- It is similar to one column in an excel spreadsheet or a database table.
- We can create Series by using dict.

```
d = {"JAN": 10, "FEB": 15, "MAR": 12, "APR": 16}
```

```
type(d)
```

```
dict
```

```
d
```

```
{'JAN': 10, 'FEB': 15, 'MAR': 12, 'APR': 16}
```

```
import pandas as pd  
s = pd.Series(d)
```

```
s
```

```
JAN    10  
FEB    15  
MAR     12  
APR     16  
dtype: int64
```

```
import pandas as pd  
s = pd.Series(d, name='val')
```

```
s
```

```
JAN    10  
FEB    15  
MAR     12  
APR     16  
Name: val, dtype: int64
```

```
s['FEB']
```

```
15
```

```
s[0]
```

```
10
```

```
s[1:3]
```

```
FEB    15  
MAR     12  
Name: val, dtype: int64
```

```
type(s)
```

```
pandas.core.series.Series
```

```
s.sum()
```

```
53
```

```
l = [10, 15, 12, 16]
```

```
l_s = pd.Series(l)
```

```
l_s
```

```
0    10
1    15
2    12
3    16
dtype: int64
```

```
l_s[0]
```

```
10
```

- When we fetch only one column from a Pandas Dataframe, it will be returned as Series.

Note

Don't worry too much about creating Data Frames yet, we are trying to understand how Data Frame and Series are related.

```
orders_path = "/data/retail_db/orders/part-00000"
```

```
orders_schema = [
    "order_id",
    "order_date",
    "order_customer_id",
    "order_status"
]
```

```
orders = pd.read_csv(orders_path,
    header=None,
    names=orders_schema
)
```

```
orders
```

	order_id	order_date	order_customer_id	order_status
0	1	2013-07-25 00:00:00.0	11599	CLOSED
1	2	2013-07-25 00:00:00.0	256	PENDING_PAYMENT
2	3	2013-07-25 00:00:00.0	12111	COMPLETE
3	4	2013-07-25 00:00:00.0	8827	CLOSED
4	5	2013-07-25 00:00:00.0	11318	COMPLETE
...
68878	68879	2014-07-09 00:00:00.0	778	COMPLETE
68879	68880	2014-07-13 00:00:00.0	1117	COMPLETE
68880	68881	2014-07-19 00:00:00.0	2518	PENDING_PAYMENT
68881	68882	2014-07-22 00:00:00.0	10000	ON_HOLD
68882	68883	2014-07-23 00:00:00.0	5533	COMPLETE

68883 rows × 4 columns

```
type(orders)
```

```
pandas.core.frame.DataFrame
```

```
orders.order_date
```

```
0      2013-07-25 00:00:00.0
1      2013-07-25 00:00:00.0
2      2013-07-25 00:00:00.0
3      2013-07-25 00:00:00.0
4      2013-07-25 00:00:00.0
...
68878   2014-07-09 00:00:00.0
68879   2014-07-13 00:00:00.0
68880   2014-07-19 00:00:00.0
68881   2014-07-22 00:00:00.0
68882   2014-07-23 00:00:00.0
Name: order_date, Length: 68883, dtype: object
```

```
type(orders.order_date)
```

```
pandas.core.series.Series
```

```
order_dates = orders.order_date
order_dates
```

```
0      2013-07-25 00:00:00.0
1      2013-07-25 00:00:00.0
2      2013-07-25 00:00:00.0
3      2013-07-25 00:00:00.0
4      2013-07-25 00:00:00.0
...
68878   2014-07-09 00:00:00.0
68879   2014-07-13 00:00:00.0
68880   2014-07-19 00:00:00.0
68881   2014-07-22 00:00:00.0
68882   2014-07-23 00:00:00.0
Name: order_date, Length: 68883, dtype: object
```

```
type(order_dates)
```

```
pandas.core.series.Series
```